AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2017JD026492

Key Points:

- Statistical models replicate hurricane activity well over a 135 year period
- Trade wind speed is the environmental field which best replicates long-term hurricane activity
- A dropout in replication skill centered on the 1940s is linked to increased data uncertainty

Supporting Information:

- Supporting Information S1

Correspondence to:
M. A. Saunders,
m.saunders@ucl.ac.uk

Citation:
Saunders, M. A., P. J. Klotzbach, and A. S. R. Lea (2017), Replicating annual North Atlantic hurricane activity 1878–2012 from environmental variables, *J. Geophys. Res. Atmos.*, *122*, doi:10.1002/2017JD026492.

Received 13 JAN 2017
Accepted 24 MAY 2017
Accepted article online 30 MAY 2017

Replicating annual North Atlantic hurricane activity 1878–2012 from environmental variables

Mark A. Saunders[1], Philip J. Klotzbach[2], and Adam S. R. Lea[1]

[1]Mullard Space Science Laboratory, University College London, Surrey, UK, [2]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA

Abstract Statistical models can replicate annual North Atlantic hurricane activity from large-scale environmental field data for August and September, the months of peak hurricane activity. We assess how well the six environmental fields used most often in contemporary statistical modeling of seasonal hurricane activity replicate North Atlantic hurricane numbers and Accumulated Cyclone Energy (ACE) over the 135 year period from 1878 to 2012. We find that these fields replicate historical hurricane activity surprisingly well, showing that contemporary statistical models and their seasonal physical links have long-term robustness. We find that August–September zonal trade wind speed over the Caribbean Sea and the tropical North Atlantic is the environmental field which individually replicates long-term hurricane activity the best and that trade wind speed combined with the difference in sea surface temperature between the tropical Atlantic and the tropical mean is the best multi-predictor model. Comparing the performance of the best single-predictor and best multi-predictor models shows that they exhibit little difference in hindcast skill for predicting long-term ACE but that the best multipredictor model offers improved skill for predicting long-term hurricane numbers. We examine whether replicated real-time prediction skill 1983–2012 increases as the model training period lengthens and find evidence that this happens slowly. We identify a dropout in hurricane replication centered on the 1940s and show that this is likely due to a decrease in data quality which affects all data sets but Atlantic sea surface temperatures in particular. Finally, we offer insights on the implications of our findings for seasonal hurricane prediction.

Plain Language Summary Many universities, government agencies and private companies issue seasonal outlooks for North Atlantic hurricane activity. However, the longer-term historical robustness of these models and their skill is unknown. Clarity on this matter is desirable because current seasonal hurricane outlooks are built on data which extend back, at best, only to the 1950s, and because predictors which are identified from data which span only a few decades can sometimes later fail. Here we assess how well annual North Atlantic hurricane activity is replicated over an extended 135-year period from 1878 to 2012; this by using statistical models and the large-scale environmental fields used most often in contemporary statistical modeling of seasonal hurricane activity. We find that these environmental fields replicate historical hurricane activity surprisingly well, showing that contemporary statistical models and their seasonal physical links have long-term robustness. We find that trade wind speed over the Caribbean Sea and the tropical North Atlantic is the environmental field which individually replicates long-term hurricane activity the best. We identify a dropout in hurricane replication centered on the 1940s and show that this is likely due to a decrease in data quality which affects all data sets but Atlantic sea surface temperatures in particular.

1. Introduction

Statistical modeling of North Atlantic hurricane activity was initiated by William Gray at Colorado State University (CSU) in 1984 [*Gray*, 1984b]. Gray's pioneering work established the field of seasonal hurricane forecasting. Now many universities, government agencies, and private companies issue seasonal outlooks for North Atlantic hurricane activity. Most of the models which these outlooks employ are statistical in form. An assessment of the skill of the statistical seasonal hurricane outlooks issued publicly in real time by three organizations between 2003 and 2014 shows that moderate-to-good forecast skill is achieved by the start of the main hurricane season in early August and that forecast skill extends back to early April for the best performing model [*Klotzbach et al.*, 2017]. Retrospective forecasts by a statistical-dynamical model for the 1982–2009 period indicate that skillful seasonal hurricane predictions may extend back to November of the previous year [*Vecchi et al.*, 2011].

Despite the skill shown by contemporary statistical seasonal hurricane forecast models, the longer-term historical robustness of these models and their skill is unknown. Clarity on this matter is desirable because current models are built on data extending back, at best, only to the 1950s, and predictors which are identified from data which span only a few decades can sometimes later fail. Examples of such predictor failures are the stratospheric quasi-biennial oscillation extrapolated to September (from the prior November) and west African monsoon rainfall from the prior August–November. These predictors had statistically significant links to hurricane activity between the 1950s and 1980s [*Gray et al.*, 1992; *Landsea and Gray*, 1992] but failed thereafter [*Klotzbach and Gray*, 2004; *Camargo and Sobel*, 2010]. The recent availability of global climate reanalysis data [*Compo et al.*, 2011] extending back into the 1800s combined with improvements in the historical quality of the North Atlantic hurricane database [*Landsea and Franklin*, 2013] provide the opportunity for the long-term robustness of contemporary statistical seasonal hurricane forecast predictors to be assessed.

Two principles underpin the ability of statistical models to forecast seasonal hurricane activity. First, a large proportion of the variance in this activity can be replicated by using large-scale environmental fields from the tropical North Atlantic and Caribbean Sea for August and September, the months of peak hurricane activity [*Gray*, 1984a; *Gray et al.*, 1993; *Gray et al.*, 1994; *Saunders and Harris*, 1997; *Klotzbach*, 2007; *Saunders and Lea*, 2008; *Vecchi et al.*, 2008; *Klotzbach*, 2011; *Ramsay and Sobel*, 2011; *Vecchi et al.*, 2011]. Second, these environmental fields are predictable months into the future. Our study concerns the validity of the first principle which we explore by examining the long-term strength, significance, and stability of the links between environmental fields and annual hurricane activity for the 135 year period from 1878 to 2012.

The primary aim of our study is to assess how well annual North Atlantic hurricane activity is replicated over an extended 135 year period by using statistical models and the large-scale environmental fields used most often in contemporary statistical modeling of seasonal hurricane activity. Our work also offers insights for steering future statistical seasonal hurricane modeling by (1) examining whether the best multi-predictor model offers better skill than the best single-predictor model, (2) examining the influence of training period length on prediction skill, and (3) examining the cause of a dropout in hurricane activity replication skill centered on the 1940s.

The manuscript is structured as follows. Section 2 describes the six large-scale environmental fields that we use in replicating historical hurricane activity. Section 3 describes the data sets that we employ for the hurricane and environmental field data, and section 4 describes the methods which underpin our analyses. Section 5 describes our main findings on how well long-term hurricane activity is replicated using statistical models and the large-scale environmental fields, and also compares the performance of single-predictor and multi-predictor models. Section 6 describes our findings relating to the influence of training period length on seasonal prediction skill. Section 7 addresses the nature and cause of temporal changes in the level of historical hurricane activity replication. Section 8 provides insights on whether the availability of longer time period data will improve current seasonal hurricane prediction skill. Section 9 describes the main conclusions.

2. Environmental Variables

Table 1 lists the six environmental fields and their associated regions that we examine in replicating seasonal North Atlantic hurricane activity between 1878 and 2012. These environmental fields are selected because they are used regularly in contemporary statistical modeling of seasonal hurricane activity and because they are available back to 1878 underpinned by observations. The six fields divide into three classes which comprise two atmospheric fields, two sea surface temperature (SST) fields, and two fields related to El Niño–Southern Oscillation (ENSO). Our study examines August–September conditions for all fields because these months replicate slightly more variance in North Atlantic hurricane activity than do August–September–October conditions for these same fields.

The two atmospheric fields assessed are the anomaly in the low-level zonal trade wind speed, u_T, at 950 hPa over the Caribbean Sea and tropical North Atlantic [*Saunders and Lea*, 2008; *Klotzbach*, 2011], and the anomaly in sea level pressure (SLP) over the tropical North Atlantic (region 10–20°N, 20–60°W; termed here the hurricane main development region (MDR)) [*Klotzbach*, 2007]. The zonal trade wind influences cyclonic vorticity and vertical wind shear over the main hurricane track region [*Saunders and Lea*, 2008] and the size of the Atlantic Warm Pool [*Wang and Lee*, 2007], which is linked to levels of hurricane activity. This environmental field has been employed in seasonal hurricane forecasts by Tropical Storm Risk (TSR) since 2002 and by

CSU since 2012 [*Klotzbach et al*., 2017]. The SLP field over the MDR is associated with anomalous atmospheric stability, vertical motion, and midlevel moisture [*Knaff*, 1997; *Klotzbach*, 2007]; conditions which influence levels of seasonal hurricane activity [*Gray*, 1968]. This environmental field has been employed by CSU to replicate seasonal hurricane activity since 2006 [*Klotzbach*, 2007].

The two SST fields examined are the anomaly in the MDR SST [*Saunders and Harris*, 1997; *Saunders and Lea*, 2008] and the anomaly in the SST difference between the tropical Atlantic and the tropical global mean (also termed the Relative SST) [*Vecchi et al.*, 2008; *Ramsay and Sobel*, 2011]. A warmer tropical North Atlantic provides heat and moisture to help power the development of storms within the MDR [*Shapiro and Goldenberg*, 1996; *Saunders and Lea*, 2008] and to increase the maximum potential intensity that these storms can achieve [*Emanuel*, 1988]. A sensitivity analysis indicates that after removal of the influence of u_T, a 0.5°C increase in August–September MDR SST is linked to a ~40% increase in hurricane frequency/activity [*Saunders and Lea*, 2008]. MDR SST has been employed in seasonal hurricane forecasting by TSR since 2002. Relative SST provides an alternative interpretation between SST and seasonal hurricane activity. The parameter may be viewed as an index of anomalous Walker circulation between the tropical Pacific and tropical Atlantic caused by an anomalous longitudinal difference in SST. This interpretation is supported by the strong correlation between Relative SST and August–September u_T at 950 hPa (r_{rank} = 0.78; 1950–2012 data). Relative SST is also associated with changes in vertical wind shear and atmospheric circulation over the tropical North Atlantic [*Vecchi et al.*, 2008].

The two August–September fields related to ENSO that we assess are the Niño 3.4 SST and the Southern Oscillation Index (SOI). The SOI [*Troup*, 1965] is included in order to allay potential concerns over the historical accuracy of Niño 3.4 SST data. August–September Multivariate ENSO Index (MEI) data [*Wolter and Timlin*, 1998, 2011] are also examined but are not displayed. The MEI is a comprehensive ENSO index which incorporates multiple atmospheric and SST ENSO components [*Wolter and Timlin*, 2011]. The influence of ENSO on Atlantic hurricane activity has been recognized since the pioneering paper by *Gray* [1984a]. ENSO causes an anomalous Walker circulation which, in turn, leads to anomalous upper tropospheric zonal winds and

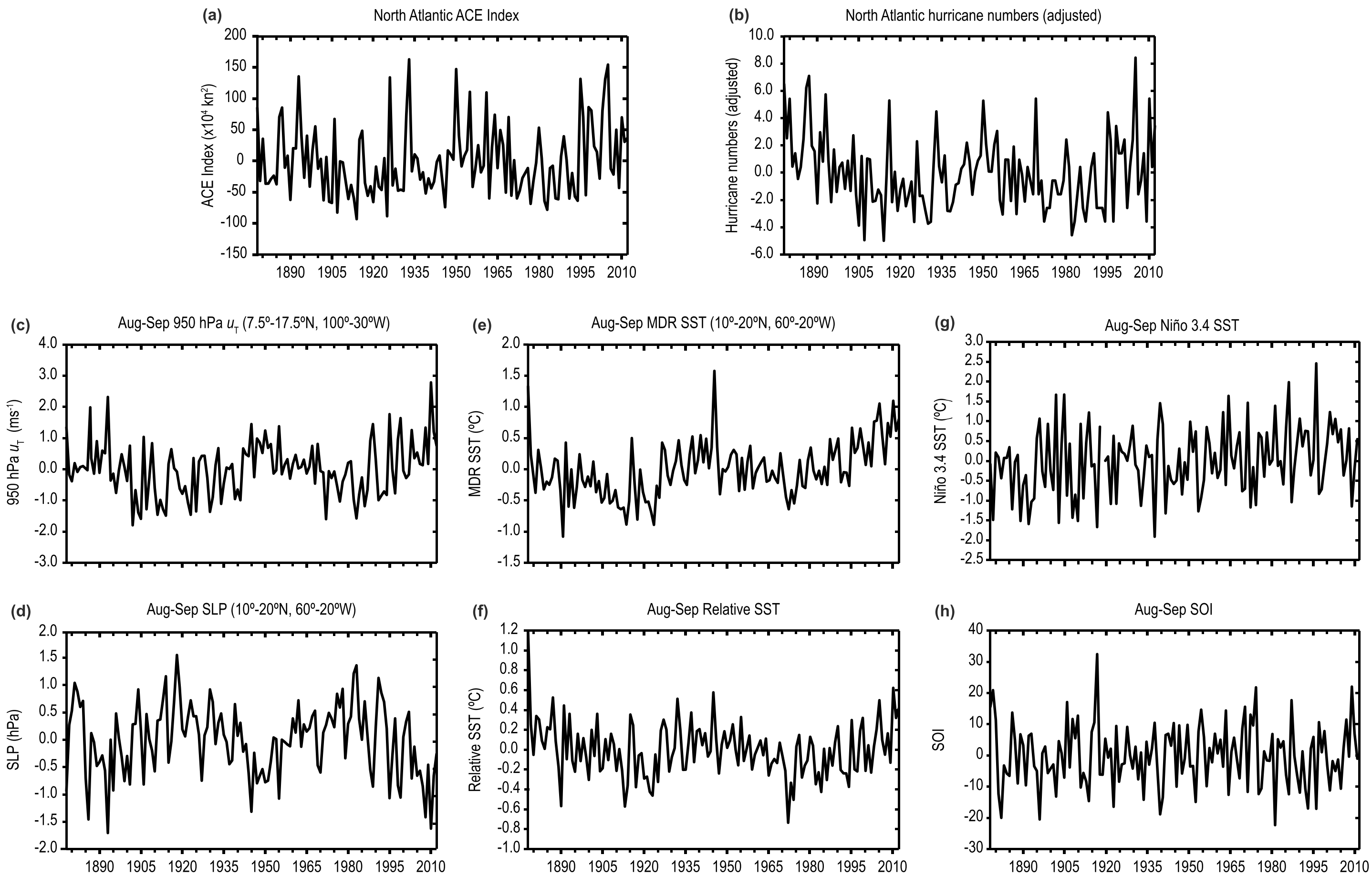

Figure 1. Nature of the data time series between 1878 and 2012. The panels display (a) the anomalies in the North Atlantic ACE index, (b) the anomalies in the adjusted number of North Atlantic hurricanes, and (c–h) the anomalies in each of the six environmental fields used to replicate annual North Atlantic hurricane activity. The data anomalies are relative to the following 1878–2012 climate norms: 106×10^4 kts^2 (annual ACE), 6.6 (annual adjusted hurricane numbers), -4.19 ms^{-1} (August–September 950 hPa u_T), 1013.35 hPa (August–September SLP), 0.04°C (August–September MDR SST), 0.12°C (August–September Relative SST), -0.19°C (August–September Niño 3.4 SST), and -0.19 (August–September SOI).

For the atmospheric wind and sea level pressure environmental fields, we employ monthly data from the NOAA 20th Century Global Reanalysis Version 2c (20CR V2c) data set [*Compo et al.*, 2011]. This reanalysis assimilates surface observations of synoptic pressure with monthly SST and sea ice distribution data to produce global atmospheric data fields. We also examine the same atmospheric fields from the 20CR V2 and ERA-20C [*European Centre for Medium-Range Weather Forecasts*, 2014] twentieth century global atmospheric reanalyses for the period 1900–2010 to reaffirm that the findings obtained with 20CR V2c are not sensitive to the atmospheric reanalysis used. For the two SST fields and for the ENSO Niño 3.4 SST index we employ monthly data from the HadSST3 data set [*Kennedy et al.*, 2011a, 2011b]. HadSST3 attempts to minimize the effects of changes in instrumentation over the historical record. We obtain monthly data for the SOI—defined as the standardized anomaly of the mean sea level pressure difference between Tahiti and Darwin—from the *Australian Bureau of Meteorology* [2014].

Although the HURDAT2, 20CR V2c, HadSST3, and SOI historical data all commence before 1878, we begin our analysis in 1878 as this is the first year for which *Vecchi and Knutson* [2011] provide adjusted hurricane numbers. The year 1878 is when the US Signal Corps began the systematic tracking of hurricanes [*Fernández-Partagás and Diaz*, 1996]. Our analysis ends in 2012 as the 20th Century Reanalysis data were only available up to this year when our main work was conducted. The HURDAT2, 20CR V2c, 20CR V2, ERA-20C, and HadSST3 data that we employ were all accessed in December 2015.

Figure 1 shows the data time series from 1878 to 2012 for the two measures of annual North Atlantic hurricane activity and for the six environmental fields examined herein. These time series are displayed as anomalies relative to each data set's 1878–2012 climate norm. The figure clarifies the anticipated quality and temporal variability of these eight data sets over the 135 year study period. The time series all exhibit substantial year-to-year variability, and several also show multidecadal variability. There are a few outlier data records which appear questionable and are suggestive of incorrect data; notably the 1.58°C and 1.33°C anomalies present in the August–September MDR SST in 1945 and 1878, respectively, the 1.19°C anomaly present in the August–September Relative SST in 1878, and the 32.3 anomaly present in the SOI in 1917. Also, the two measures of North Atlantic hurricane activity appear inconsistent before ~1895 with hurricane numbers (adjusted) showing a higher level of activity than indicated by ACE. These issues show that the data quality is not perfect and suggest that uncertainties in the data likely increase as one goes back in time. Nevertheless the visual impression provided by Figure 1 is that the eight time series appear, in general, to be reasonable and useable.

4. Methods

The strength and significance of the links between the six environmental fields in Table 1 and North Atlantic hurricane activity between 1878 and 2012 are computed as follows. The strength of each association is given as the running 30 year Spearman rank correlation, r_{rank}, value computed from the annual time series of each environmental field and each hurricane activity measure. The significance of each association is given by the r_{rank} p value computed using the methods described in *Saunders and Lea* [2005] (see their methods subsections therein headed "serial autocorrelation" and "significances").

The replication of seasonal hurricane activity is made using linear regression models. These models use either single or multiple environmental fields from Table 1 as predictors and satisfy the validity assumptions for ordinary least squares linear regression [*Freund and Wilson*, 1998; *Wilks*, 2011]. Because ACE, the number of hurricanes, and the adjusted number of hurricanes all have a positively skewed distribution, their distributions do not pass the test for normality for any time period examined. To ensure that observations are drawn from a normal distribution and the hindcast errors are normally distributed with a mean of zero (both requirements of linear regression modeling), we transform both distributions to a normal distribution using the square root function. We test for normality using the Kolmogorov-Smirnov test. The linear regression modeling is performed on these transformed data to produce hindcasts, which are then transformed back before the hindcast skill is computed.

Hindcasts of hurricane activity are made in two ways. The first method is cross validation with block elimination as described in *Saunders and Lea* [2005] (see their methods section, "cross-validated hindcasts" subsection). This hindcast skill is assessed and compared for three periods: 1878–2012, 1878–1949, and 1950–2012. The split of the two shorter periods is chosen on the basis that hurricane replication skill has not been assessed before 1950 to our knowledge. The second hindcast method is called replicated real-time forecasts [*Lloyd-Hughes et al.*, 2004]. In this method linear regression modeling is performed only on data prior to the year being hindcast and the training period increases 1 year at a time as each hindcast is made. Replicated real-time forecast skill is assessed for the 30 year period 1983–2012 using three different lengths of model training: T1878 which uses training data initiated in 1878, T1950 which uses training data initiated in 1950, and T15yr which uses a rolling prior 15 year trained model. Thus, taking T1878, for example, one uses 1878–1982 data to predict 1983, 1878–1983 data to predict 1984, through to 1878–2011 data to predict 2012.

Hindcast skill is assessed using three skill measures: the mean square error (MSE) between the hindcast and observed values, the correlation (r_{rank}) between the hindcast and observed values, and the mean square skill score (MSSS) defined as the percentage reduction in mean square error of the model hindcasts compared to hindcasts made with the period mean or climatology value. MSSS is the skill metric recommended by the World Meteorological Organization for verification of deterministic seasonal forecasts [*World Meteorological Organization,* 2002]. Hindcast significance (p value) is computed from bootstrapped estimates of r_{rank} as described in *Saunders and Lea* [2005] (see their methods section, "significances" subsection, paragraph two).

5. Replicating Historical Seasonal Hurricane Activity

This section quantifies how well historical seasonal hurricane activity is replicated by using single environmental fields (section 5.1) and by using a combination of environmental fields (section 5.2). The section then

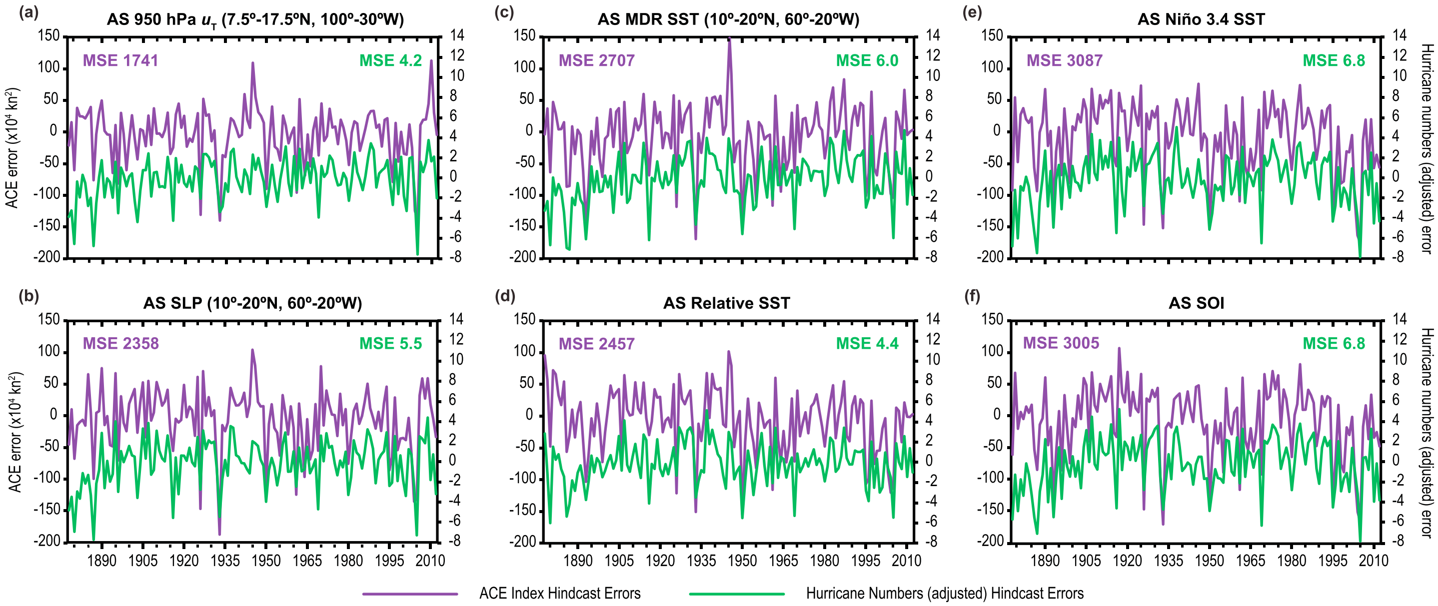

Figure 2. The hindcast error time series between 1878 and 2012 obtained with each environmental field. (a–f) The six panels display by environmental field the hindcast errors in the North Atlantic ACE index (purple lines) and the hindcast errors in the adjusted number of North Atlantic hurricanes (green lines). The mean square error (MSE) of the 135 year hindcast errors for each hurricane activity measure is included in the top corners of each panel.

compares the performance of the best single-predictor model with the best multipredictor model (section 5.3). These assessments are made for the three data periods of 1878–2012, 1878–1949, and 1950–2012.

5.1. Using Single Environmental Fields

We first introduce the performance of the six environmental fields in replicating North Atlantic hurricane activity and follow this with more detailed analyses and results. Figure 2 displays the hindcast error time series for ACE and adjusted hurricane numbers obtained statistically with each environmental field in Table 1 for 1878–2012. These hindcasts are made by using linear regression modeling with cross validation (section 4). The 135 year hindcast skills are displayed in each panel in terms of MSE. Careful inspection of Figure 2 reveals differences in performance with certain environmental fields exhibiting greater hindcast errors than other fields. Trade wind speed is the environmental field which replicates long-term hurricane activity the best, as it has the lowest MSE for both ACE and adjusted hurricane numbers. The two ENSO fields individually perform worst in replicating long-term hurricane activity. In general, the hindcast error time series appear to be temporally stable. However, there are periods common to several environmental fields where the hindcast errors are larger, for example, around World War 2 and before ~1890 for adjusted hurricane numbers.

Figure 3 and Table 2 quantify in more detail than Figure 2 how well the six environmental fields individually replicate historical seasonal hurricane activity. Figure 3 displays the strength (r_{rank}) and significance (p value) of the links between each environmental field and the two measures of hurricane activity—ACE and the "adjusted" number of hurricanes—for running 30 year intervals across the 135 year period. Table 2 shows the predictive (cross-validated hindcast) skill and significance for these same measures of hurricane activity computed from each environmental field for each of the three data periods. Using "raw" hurricane numbers instead of adjusted hurricane numbers produces similar results to Figure 3 over the 135 year period (see Figure S1 in the supporting information). There are four main findings evident from Figure 3 and Table 2:

1. The six environmental fields provide good replication of North Atlantic hurricane numbers and ACE over the extended period from 1878 to 2012. This finding is clearest for the two atmospheric fields and occurs despite an anticipated decrease in data quality as one goes back in time. Each environmental field provides significant hindcast predictive skill (p value ≤ 0.05) for ACE and adjusted hurricane numbers over the full 135 year period. There is also significant predictive skill provided by the 950 hPa u_T, SLP, and Relative SST fields for both of the shorter periods of 1878–1949 and 1950–2012.

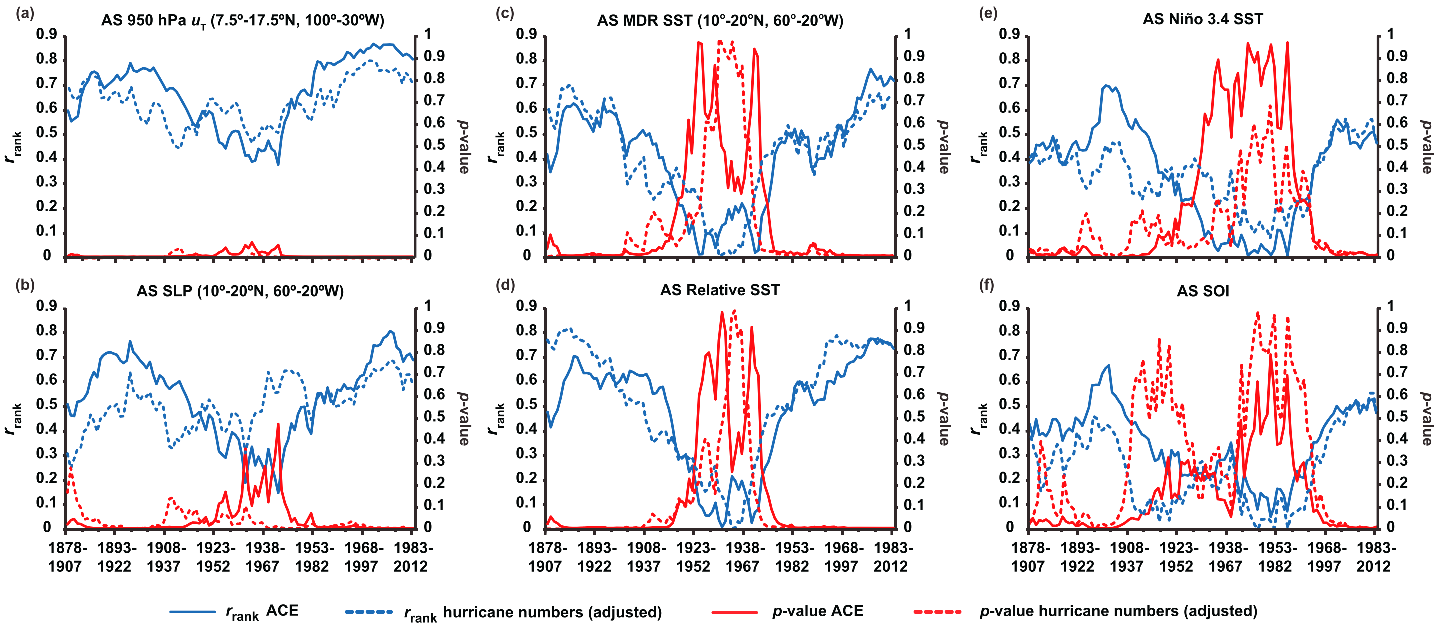

Figure 3. The strength, significance, and stability of the links between the six environmental fields in Table 1 and North Atlantic hurricane activity between 1878 and 2012. The environmental fields are grouped by type into three columns: (a, b) atmospheric fields (left column), (c, d) SST fields (center column), and (e, f) ENSO fields (right column). The two measures of hurricane activity examined are ACE (solid lines) and the number of adjusted hurricanes (dashed lines). The strength and significance of the associations are shown respectively by rank correlations (blue lines) and by *p* values (red lines). Values are computed for running 30 year periods. "AS" signifies August–September.

2. Historical hurricane activity is replicated best by the August–September zonal trade wind speed over the Caribbean Sea and tropical North Atlantic. This environmental field gives more predictive skill for ACE and adjusted hurricane numbers than any other environmental field in Table 1. This result is found for the 135 year period and for both shorter periods therein. For the 135 year period the trade

Table 2. Predictive Skill for the North Atlantic ACE Index and Adjusted Number of Hurricanes From Single Environmental Fields[a]

		ACE Hindcast Skill and Significance			Number of Adjusted Hurricanes Hindcast Skill and Significance		
Predictor	Period	MSSS (%)	r_{rank}	*p* value	MSSS (%)	r_{rank}	*p* value
950 hPa u_T	1878–2012	41	0.69	<0.001	41	0.63	<0.001
	1878–1949	32	0.62	<0.001	39	0.61	<0.001
	1950–2012	41	0.77	<0.001	42	0.69	<0.001
SLP	1878–2012	25	0.52	<0.001	24	0.47	<0.001
	1878–1949	19	0.47	<0.001	11	0.32	0.02
	1950–2012	29	0.61	<0.001	31	0.58	<0.001
MDR SST	1878–2012	14	0.42	<0.001	15	0.37	0.007
	1878–1949	0	0.29	0.02	4	0.30	0.08
	1950–2012	29	0.53	0.002	31	0.50	0.002
Relative SST	1878–2012	22	0.51	<0.001	38	0.61	<0.001
	1878–1949	17	0.43	<0.001	30	0.54	<0.001
	1950–2012	35	0.63	<0.001	48	0.66	<0.001
Niño 3.4 SST	1878–2012	1	0.25	0.004	5	0.30	<0.001
	1878–1949	2	0.35	0.003	0	0.26	0.03
	1950–2012	0	0.19	0.16	2	0.27	0.07
SOI	1878–2012	4	0.34	<0.001	4	0.24	0.006
	1878–1949	0	0.32	0.01	0	0.11	0.39
	1950–2012	0	0.29	0.04	1	0.25	0.08

[a]August–September field values are used for the 'predictors'. Hindcast predictive skill is given by the mean square skill score (MSSS) and by the rank correlation between the predicted and actual time series (r_{rank}). The data period gives the climatology for MSSS.

Table 3. Predictive Skill for the North Atlantic ACE Index and Adjusted Number of Hurricanes From Combinations of Two and Three Environmental Fields[a]

		ACE Hindcast Skill and Significance			Number of Adjusted Hurricanes Hindcast Skill and Significance		
Predictors	Period	MSSS (%)	r_{rank}	p value	MSSS (%)	r_{rank}	p value
950 hPa u_T and SLP	1878–2012	40	0.69	<0.001	40	0.63	<0.001
	1878–1949	30	0.63	<0.001	37	0.58	<0.001
	1950–2012	39	0.75	<0.001	39	0.66	<0.001
950 hPa u_T and MDR SST	1878–2012	40	0.69	<0.001	41	0.64	<0.001
	1878–1949	28	0.61	<0.001	38	0.60	<0.001
	1950–2012	33	0.75	<0.001	41	0.68	<0.001
950 hPa u_T and Relative SST	1878–2012	43	0.70	<0.001	50	0.70	<0.001
	1878–1949	36	0.63	<0.001	48	0.65	<0.001
	1950–2012	38	0.77	<0.001	47	0.71	<0.001
SLP and MDR SST	1878–2012	28	0.60	<0.001	29	0.54	<0.001
	1878–1949	14	0.51	<0.001	13	0.42	0.009
	1950–2012	27	0.64	<0.001	34	0.61	<0.001
SLP and Relative SST	1878–2012	33	0.61	<0.001	44	0.65	<0.001
	1878–1949	28	0.60	<0.001	34	0.58	<0.001
	1950–2012	36	0.67	<0.001	48	0.68	<0.001
950 hPa u_T, SLP and MDR SST	1878–2012	39	0.69	<0.001	40	0.63	<0.001
	1878–1949	26	0.61	<0.001	37	0.58	<0.001
	1950–2012	31	0.74	<0.001	38	0.66	<0.001
950 hPa u_T, SLP and Relative SST	1878–2012	42	0.70	<0.001	50	0.69	<0.001
	1878–1949	34	0.63	<0.001	47	0.65	<0.001
	1950–2012	36	0.74	<0.001	46	0.68	<0.001

[a]All parameters are defined as in Table 2.

wind's predictive skill for ACE is r_{rank} = 0.69; a value which we anticipate would increase if data errors could be removed.

3. The strength of the links between the environmental variables and historical seasonal hurricane activity appears, in general, to be nonconstant over the 135 year period. The SST fields in particular show a significant dropout in r_{rank} for 30 year periods between 1913–1942 and 1948–1977 (Figures 3c and 3d). A smaller dropout in r_{rank} is seen in the two atmospheric variables across these same periods (Figures 3a and 3b).
4. The significant ENSO link to hurricane activity that is widely documented to occur in recent decades becomes weaker and insignificant during 30 year periods between 1913–1942 and 1963–1992 (Figures 3e and 3f). MEI data (not shown) confirm the same historical dropout in the strength and significance of the ENSO link to hurricane activity.

5.2. Using Combinations of Environmental Fields

Table 3 quantifies in the same format as Table 2 how well combinations of the six environmental fields in Table 1 replicate historical seasonal hurricane activity. Predictive skills are displayed for five combinations of two environmental fields and for two combinations of three environmental fields. These combinations are selected as they offer the highest hindcast skills. The two main findings evident from Table 3 are as follows:

1. The best multipredictor model is AS 950 hPa zonal trade wind speed combined with the difference in AS SST between the tropical Atlantic and the tropical global mean. This is the case for the 1878–2012 full period and for the 1878–1949 historical period. However, for the more recent 1950–2012 period, other multifield combinations offer similar predictive skill.
2. Using more than two environmental fields does not lead to any noticeable further increase in hindcast skill. This outcome may be expected because the best multi-predictor model already exhibits sizeable skill, and with the six environmental fields not being independent, there is little additional independent skill that may be added.

5.3. Comparing Best Single-Predictor and Best Multi-Predictor Models

Figure 4 compares the performance of the best single-predictor model (AS 950 hPa u_T) in replicating North Atlantic hurricane activity 1878–2012 with the performance of the best multi-predictor model (AS 950 hPa u_T and AS Relative SST). A comparison of Tables 2 and 3 is also relevant. The two main findings are as follows:

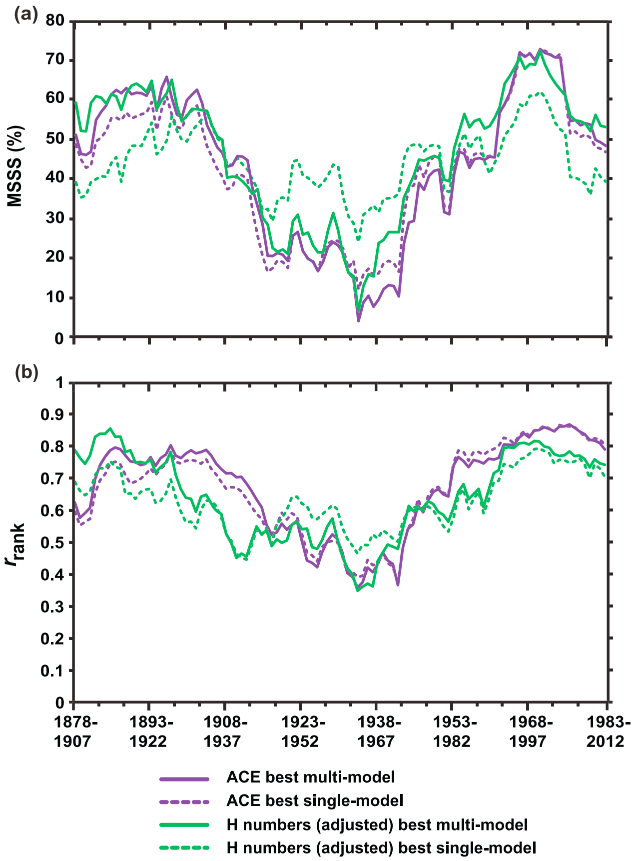

Figure 4. Comparison of the performance of the best multi-predictor statistical model (AS 950 hPa u_T and AS Relative SST) and the best single-predictor statistical model (AS 950 hPa u_T) in replicating North Atlantic hurricane activity 1878–2012. The two measures of hurricane activity examined are ACE (purple lines) and number of adjusted hurricanes (green lines). The replication skill is assessed using mean square skill score (Figure 4a) and rank correlation (Figure 4b). Values are computed for running 30 year periods.

1. For long-term ACE there is little difference in the replication skill and significance between using the best single-predictor model and the best multi-predictor model. Taking the MSSS (%) skill metric, the mean difference in MSSS between the two models for the three periods (1878–2012, 1878–1949, and 1950–2012) is just 1%.
2. For long-term adjusted hurricane numbers the best multi-predictor model offers improved replication skill compared to the best single-predictor model. Here the mean difference in MSSS between the two models for the three periods is 8%. This result arises because AS Relative SST individually has more hindcast skill for adjusted hurricane numbers than for ACE (Table 2).

In contrast, the five other environmental predictors in Table 2 have comparable hindcast skill for ACE and adjusted hurricane numbers. A further finding of note is that while there is little difference in the temporal stability of the ACE replication skill across the 135 year historical period between using the best single-predictor model and the best multi-predictor model, this is not the case for adjusted hurricane numbers. For the latter measure of hurricane activity, the dropout in replication skill for 30 year periods between 1913–1942 and 1948–1977 noted in Figures 3c and 3d is more prominent for the best multi-predictor model than for the best single-predictor model. This is evident by comparing the solid and dashed green lines in Figure 4. An explanation for this finding is discussed in section 7.

6. Influence of Model Training Period Length on Replicated Real-Time Skill 1983–2012

Our use of 135 years of historical data permits examination of the influence of model training period length on skill. Although it is physically reasonable to anticipate that model skill will grow as the training period length increases, this outcome has not been examined before due to the shortness of the previously studied data records. We examine this outcome by testing whether the replicated real-time prediction skill for the 1983–2012, 30 year period increases as the model training period lengthens. Our assessment is made for the three training period lengths of T1878, T1950, and T15yr defined in section 4 and for the best single-predictor and best multipredictor models identified in section 5. Our replicated real-time forecast methodology is described in section 4. Table 4 displays our findings. Overall, we find evidence that real-time prediction skill increases, albeit slowly, as the model training period length increases from 15 years to ≥105 years. Considering the MSSS (%) skill metric and taking the mean of the ACE and hurricane number hindcast skills for both models, there is a 6% mean improvement in going from T15yr to T1950, and a 10% mean improvement in going from T15yr to T1878. However, this increase is more pronounced for the ACE hindcast skill than

Table 4. Assessment of the Influence of Model Training Period Length on Predictive Skill[a]

Predictor(s)	Model Training	ACE Hindcast Skill and Significance			Number of Adjusted Hurricanes Hindcast Skill and Significance		
		MSSS (%)	r_{rank}	p value	MSSS (%)	r_{rank}	p value
950 hPa u_T	T1878	45	0.81	<0.001	39	0.71	<0.001
	T1950	41	0.81	<0.001	38	0.70	<0.001
	T15yrs	33	0.74	<0.001	34	0.62	0.001
950 hPa u_T and Relative SST	T1878	47	0.80	<0.001	53	0.75	<0.001
	T1950	40	0.79	<0.001	49	0.73	<0.001
	T15yrs	27	0.74	<0.001	52	0.72	<0.001

[a]The assessment is made for the best single-predictor model and the best multi-predictor model. The influence is examined by using three different lengths of model training and computing the replicated real-time skill for ACE and adjusted number of hurricanes for the 30 year period 1983–2012. The types/lengths of model training examined are defined in section 4. The other parameters are defined as in Tables 2 and 3. The 1983–2012 period is used throughout as the climatology for computing MSSS.

for the hurricane number hindcast skill. The same skill increase is less evident when using the correlation (r_{rank}) skill measure. A factor which complicates this assessment and its interpretation is the variable quality of the historical data (see section 7).

7. Temporal Changes in the Level of Replication

The three most noteworthy temporal changes in the level of historical hurricane activity replication from August–September environmental fields (Figure 3) are the following: (1) A dropout in replication centered on the 1940s (associated in particular with the SST fields); (2) a general slow decrease in the strength of replication associated with most environmental fields as time extends back to 1878; and (3) a dropout in replication extending through the 1950s and 1960s associated with the ENSO fields. We examine the cause(s) of these temporal changes in hurricane activity replication to clarify whether they likely arise from a decrease in data quality or reflect a change in physical mechanism. This is achieved by first quantifying with more precision when the r_{rank} between each environmental field and hurricane activity measure was the lowest during the historical period. Establishing this sharpness in timing is not possible from r_{rank} values computed over running 30 year (or other large multiyear) periods as in Figure 3.

Figure 5 gives improved temporal precision on the strength of the links between environmental fields and hurricane activity. This is achieved by plotting the percent of occurrences by year where the running 30 year r_{rank} is in the lowest quartile of all 30 year running r_{ranks} 1878–2012. A minimum of 12 realizations is chosen in computing the percent of occurrences, thereby enabling a plot range from 1890 to 2000. Years between 1907 and 1983 employ 30 realizations. The use of running 20 year r_{ranks} and quintiles were also assessed but these combinations gave a less clear outcome than in Figure 5. The figure shows that the r_{rank} between the SST and atmospheric fields and annual hurricane activity was the lowest in the middle to late 1940s; namely, during and immediately after World War 2 (WW2). In contrast, the r_{rank} between the ENSO fields and hurricane activity was the lowest between the middle 1950s and 1970. These results are similar for ACE and numbers of hurricanes.

To further clarify the cause(s) of the temporal changes in hurricane activity replication we compute historical area average SST uncertainties in the HadSST3 data set. This is achieved by applying the *Kennedy et al.* [2011a, 2011b] uncertainty analysis—which defines the full SST uncertainty in global and hemispheric HadSST3 data—to the smaller MDR and Niño 3.4 area average regions. The methodology for computing the three constituent SST uncertainties and the full SST uncertainty for an area average region is described in Text S1 in the supporting information [*Kennedy et al.*, 2011a, 2011b; *Rayner et al.*, 2003]. Figure 6 displays our computed uncertainties in August–September MDR SST and August–September Niño 3.4 SST between 1878 and 2012. The MDR SST uncertainty peaks around WW2 (years 1942–1946 in particular), slowly increases from the late 1930s back to the 1870s, and is approximately constant between 1950 and 2012. The uncertainty in Niño 3.4 SST is also the highest around WW2, has a secondary peak around WW1, and displays elevated uncertainties prior to ~1960. In both regions the total SST uncertainty is dominated by the measurement and sampling uncertainty, with measurement error likely exceeding the sampling error (J. J. Kennedy,

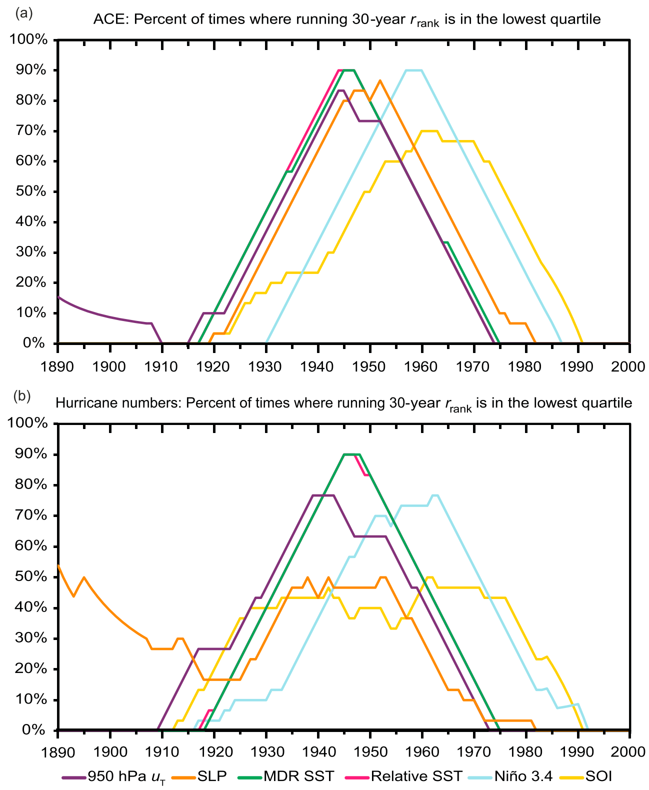

Figure 5. Quantification of when the rank correlation between each environmental field and North Atlantic hurricane activity was the lowest between 1890 and 2000. The assessment is made for (a) ACE and for (b) raw hurricane numbers. The six environmental fields are color coded.

personal communication, 2015). SST uncertainties computed from the new monthly extended reconstructed sea surface temperature version 4 data set (ERSST.v4) also show an uncertainty maximum around WW2 for the MDR and and Niño 3.4 regions [*Liu et al.*, 2015; see their Figure 2].

The historical uncertainty in annual hurricane activity is also relevant to deducing the cause(s) of temporal changes in hurricane activity replication. Published uncertainties for the number of North Atlantic tropical cyclones and hurricanes [*Vecchi and Knutson*, 2008, 2011] show an uncertainty maximum around WW2, a slightly smaller uncertainty maximum around WW1, and a general slow increase in uncertainty as time extends backward from the middle 1960s to 1878. Published uncertainties for ACE do not exist but one would reasonably expect them to be similar in temporal form to the uncertainties in the number of hurricanes. Since 'raw' and 'adjusted' hurricane numbers exhibit comparable temporal changes in the level of historical hurricane activity replication (see Figure S1), we deduce that the uncertainties in annual hurricane numbers may be of secondary importance to the uncertainties in the environmental fields in influencing the replication of historical hurricane activity.

The outcomes in Figures 5 and 6, combined with the information on historical hurricane uncertainty, allow informed deductions to be made about the likely cause(s) of the noteworthy temporal changes in the level of annual hurricane activity replication displayed in Figure 3. First, the dropout in replication centered on the 1940s (associated in particular with the SST fields) seems to be likely due to a decrease in data quality. This is because the timing of this dropout (Figures 3c and 5) matches very well with when the MDR SST uncertainty is the greatest (Figure 6a) and when the uncertainty in hurricane activity is the greatest. It would be an unlikely coincidence if the physical influence of MDR SST and Relative SST on ACE somehow happened to weaken only at the same time as when the data quality degraded. Similarly, it would seem likely that the smaller decrease in hurricane activity replication associated with the atmospheric fields SLP and u_T around the 1940s (Figures 3b and 3a) is also due to a decrease in data quality. Second, the general slow decrease in the strength of replication associated with most environmental fields as time extends back to 1878 also

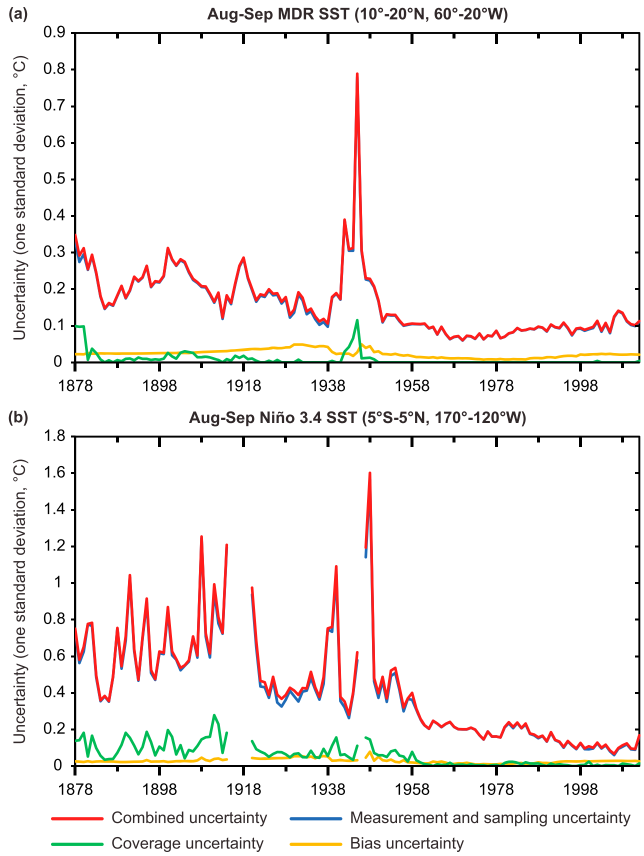

Figure 6. The uncertainty (a) in August–September hurricane main development region SST and (b) in August–September Niño 3.4 SST. Each combined uncertainty is also broken down into its three constituent parts [*Kennedy et al.,* 2011a, 2011b]. Years with no uncertainty value indicate an absence of SST observations.

seems to likely reflect a decrease in data quality. This is because the uncertainty in the SST environmental fields (Figure 6), the uncertainty in hurricane numbers [*Vecchi and Knutson*, 2008, 2011], and, presumably, the uncertainty in atmospheric fields, all slowly increase with time back to the 1870s.

The dropout in hurricane activity replication associated with the ENSO fields (Figures 3e and 3f) is more challenging to explain. The dropout peaks between the mid-1950s and 1970 (Figure 5b). Temporal changes in the Niño 3.4 SST uncertainty (Figure 6b) and in the uncertainty of hurricane numbers cannot explain much of this dropout. This suggests that the physical influence of ENSO on North Atlantic hurricane activity may have weakened during this period.

Finally, Figure 4 offers an explanation for why the best multi-predictor model has less replication skill than the best single-predictor model for adjusted hurricane numbers for 30 year periods between 1913–1942 and 1948–1977. This is likely due to a decrease in the data quality of the Relative SST environmental field during this period. Relative SST and 950 hPa u_T constitute the best multi-predictor model.

8. Implications for Contemporary Seasonal Hurricane Prediction

What are the insights and implications which our work can offer for contemporary seasonal hurricane prediction? We address this by considering two questions: (Q1) Does the availability of longer time period data improve the skill of current seasonal hurricane predictions?; (Q2) Do multi-predictor models provide a significant skill advantage over using the best single-predictor model? Our answers are predicated upon our analysis of the six environmental fields (Table 1) used most often in contemporary statistical modeling of seasonal hurricane activity. However, we consider it is unlikely that further environmental fields will be discovered that would significantly change our current answers to Q1 and Q2 below.

Our answer to Q1 is “yes but that the skill increases only slowly as the training period length increases beyond 15 years.” We find only a 10% mean improvement in MSSS (%) in going from T15yr to T1878 (section 6). Although the decrease in data quality which affects historical environmental field and hurricane data (section 7) would act to offset the increase in skill that arises from using a longer training period, it would appear that the latter skill improvement is likely modest compared to the skill obtained from only using a prior 15 year training period. Having said that, we would still recommend that current seasonal forecasting efforts train their models on the full data period from 1878 in order to maximize their potential skill.

Our answer to Q2 is “probably no.” As noted in section 5.3, for ACE there is little difference in the replication skill between using the best single-predictor model and the best multi-predictor model. Although for adjusted hurricane numbers, the best multi-predictor model offers improved replication skill compared to the best single-predictor model, this improvement is more pronounced for the 1878–1949 historical period than for the more recent 1950–2012 period where data quality is best. Furthermore, if the “Relative SST” environmental field was used as the best single-predictor model for adjusted hurricane numbers instead of "950 hPa u_T" there would be no difference in the model skill 1950–2012 between the best multi-predictor and best single-predictor models. A physical explanation for why little additional skill is added by including a second predictor comes from the facts that the best single-predictor model already exhibits sizeable skill and the environmental fields examined here are not independent.

9. Conclusions

Annual hurricane activity over the 135 year period from 1878 to 2012 is replicated surprisingly well by large-scale environmental fields. This finding gives confidence that there is long-term robustness in the physical links between these fields and annual hurricane activity. From the six environmental fields and data examined, August–September trade wind speed over the Caribbean Sea and tropical North Atlantic replicates long-term hurricane activity the best, achieving predictive skill for ACE of $r_{\text{rank}} = 0.69$ for the 135 year period. Multi-predictor models offer no improvement over the single trade wind predictor for replicating ACE, but trade wind speed combined with the difference in SST between the tropical Atlantic and the global tropical mean offers modest improvement for replicating adjusted hurricane numbers. The skill of current seasonal hurricane predictions will slowly increase by incorporating longer time period data (out to 135 years) into model training, but by only ~10% in MSSS compared to using a prior 15 year trained model. Temporal changes occur in the level of historical hurricane replication, most notably around WW2. These changes occur at times when there is increased uncertainty in the environmental field data and hurricane data. If the quality of the historical data could be improved further, we anticipate that the environmental fields studied herein would replicate historical hurricane activity even better, and that modest additional improvements in the skill of current seasonal hurricane predictions would ensue.

Acknowledgments
We thank the late William M. Gray for helpful comments on different aspects of this work and John J. Kennedy for help in computing area average SST uncertainties from the HadSST3 data set. Support for the Twentieth Century Reanalysis Project version 2c data set is provided by the U.S. Department of Energy, the Office of Biological and Environmental Research, and by the National Oceanic and Atmospheric Administration Climate Program Office. All the data necessary to evaluate and build upon the work in this paper are available in the cited references, figures, tables, and supporting information. We acknowledge helpful comments from three referees.

References

Australian Bureau of Meteorology (2014), Southern Oscillation Index (SOI) archives. [Available at http://www.bom.gov.au/climate/current/soihtm1.shtml, accessed 2014-17-10.]

Bell, G. D., M. S. Halpert, R. C. Schnell, R. W. Higgins, J. Lawrimore, V. E. Kousky, R. Tinker, W. Thiaw, M. Chelliah, and A. Artusa (2000), Climate assessment for 1999, *Bull. Am. Meteorol. Soc.*, *81*, S1–S50.

Camargo, S. J., and A. H. Sobel (2010), Revisiting the influence of the quasi-biennial oscillation on tropical cyclone activity, *J. Clim.*, *23*, 5810–5825, doi:10.1175/2010JCLI3575.1.

Compo, G. P., et al. (2011), The twentieth century reanalysis project, *Q. J. R. Meteorol. Soc.*, *137*, 1–28, doi:10.1002/qj.776.

DeMaria, M. (1996), The effect of vertical shear on tropical cyclone intensity change, *J. Atmos. Sci.*, *53*, 2076–2087.

Emanuel, K. (1988), The maximum intensity of hurricanes, *J. Atmos. Sci.*, *45*, 1143–1155.

European Centre for Medium-Range Weather Forecasts (2014), ERA-20C project (ECMWF atmospheric reanalysis of the 20th century), Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. [Available at doi:10.5065/D6VQ30QG, accessed 2014-02-10.]

Fernández-Partagás, J., and H. F. Diaz (1996), Atlantic hurricanes in the second part of the nineteenth century, *Bull. Am. Meteorol. Soc.*, *77*, 2899–2906.

Freund, R. J., and W. J. Wilson (1998), *Regression Analysis: Statistical Modeling of a Response Variable*, 444 pp., Academic Press, San Diego, Calif.

Gray, W. M. (1968), Global view of the origin of tropical disturbances and storms, *Mon. Weather Rev.*, *96*, 669–700.

Gray, W. M. (1984a), Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences, *Mon. Weather Rev.*, *112*, 1649–1668.

Gray, W. M. (1984b), Atlantic seasonal hurricane frequency. Part II: Forecasting its variability, *Mon. Weather Rev.*, *112*, 1669–1683.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry (1992), Predicting Atlantic seasonal hurricane activity 6-11 months in advance, *Weather Forecasting*, *7*, 440–455.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry (1993), Predicting Atlantic basin seasonal tropical cyclone activity by 1 August, *Weather Forecasting*, *8*, 73–86.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry (1994), Predicting Atlantic basin seasonal tropical cyclone activity by 1 June, *Weather Forecasting*, *9*, 103–115.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011a), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850, part 1: Measurement and sampling errors, *J. Geophys. Res.*, *116*, D14103, doi:10.1029/2010JD015218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011b), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850, part 2: Biases and homogenisation, *J. Geophys. Res.*, *116*, D14104, doi:10.1029/2010JD015220.

Klotzbach, P. J. (2007), Revised prediction of seasonal Atlantic basin tropical cyclone activity from 1 August, *Weather Forecasting*, *22*, 937–949, doi:10.1175/WAF1045.1.

Klotzbach, P. J. (2011), The influence of El Niño–Southern Oscillation and the Atlantic multi-decadal oscillation on Caribbean tropical cyclone activity, *J. Clim.*, *24*, 721–731, doi:10.1175/2010JCLI3705.1.

Klotzbach, P. J., and W. M. Gray (2004), Updated 6–11 month prediction of Atlantic basin seasonal hurricane activity, *Weather Forecasting*, *19*, 917–934.

Klotzbach, P. J., M. A. Saunders, G. D. Bell, and E. S. Blake (2017), North Atlantic seasonal hurricane prediction: Underlying science and an evaluation of statistical models, in *Climate Extremes: Patterns and Mechanisms*, *Geophys. Monogr. Ser*, vol. 226, edited by S.-Y. Wang et al., pp. 315–328, AGU, Washington, D. C.

Knaff, J. A. (1997), Implication of summertime sea level pressure anomalies in the tropical Atlantic region, *J. Clim.*, *10*, 789–804.

Landsea, C. W., and W. M. Gray (1992), The strong association between western Sahel monsoon rainfall and intense Atlantic hurricanes, *J. Clim.*, *5*, 435–453.

Landsea, C. W., and J. L. Franklin (2013), Atlantic hurricane database uncertainty and presentation of a new database format, *Mon. Weather Rev.*, *141*, 3576–3592, doi:10.1175/MWR-D-12-00254.1.

Liu, W., et al. (2015), Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations, *J. Clim.*, *28*, 931–951, doi:10.1175/JCLI-D-14-00007.1.

Lloyd-Hughes, B., M. A. Saunders, and P. Rockett (2004), A consolidated CLIPER model for improved August-September ENSO prediction skill, *Weather Forecasting*, *19*, 1089–1105.

Ramsay, H. A., and A. H. Sobel (2011), Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model, *J. Clim.*, *24*, 183–193, doi:10.1175/2010JCLI3690.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.

Saunders, M. A., and A. R. Harris (1997), Statistical evidence links exceptional 1995 Atlantic hurricane season to record sea warming, *Geophys. Res. Lett.*, *24*, 1255–1258, doi:10.1029/97GL01164.

Saunders, M. A., and A. S. Lea (2005), Seasonal prediction of hurricane activity reaching the coast of the United States, *Nature*, *434*, doi:10.1038/nature03454.

Saunders, M. A., and A. S. Lea (2008), Large contribution of sea surface warming to recent increase in Atlantic hurricane activity, *Nature*, *451*, 557–560, doi:10.1038/nature06422.

Shapiro, L. J., and S. B. Goldenberg (1996), Atlantic sea surface temperatures and tropical cyclone formation, *J. Clim.*, *11*, 578–590.

Tang, B. H., and J. D. Neelin (2004), ENSO influence on Atlantic hurricanes via tropospheric warming, *Geophys. Res. Lett.*, *31*, L24204, doi:10.1029/2004GL021072.

Troup, A. J. (1965), The Southern Oscillation, *Q. J. R. Meteorol. Soc*, *91*, 490–506.

Vecchi, G. A., K. L. Swanson, and B. J. Soden (2008), Wither hurricane activity?, *Science*, *322*, 687–689, doi:10.1126/science.1164396.

Vecchi, G. A., and T. R. Knutson (2008), On estimates of historical North Atlantic tropical cyclone activity, *J. Clim.*, *21*, 3580–3600, doi:10.1175/2008JCLI2178.1.

Vecchi, G. A., and T. R. Knutson (2011), Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878–1965) using ship track density, *J. Clim.*, *24*, 1736–1746, doi:10.1175/2010JCLI3810.1.

Vecchi, G. A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R. Gudgel (2011), Statistical-dynamical predictions of seasonal North Atlantic hurricane activity, *Mon. Weather Rev.*, *139*, 1070–1082, doi:10.1175/2010MWR3499.1.

Wang, C., and S.-K. Lee (2007), Atlantic warm pool, Caribbean low-level jet, and their potential impact on Atlantic hurricanes, *Geophys. Res. Lett.*, *34*, L02703, doi:10.1029/2006GL028579.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, 704 pp., Academic Press, Oxford.

Wolter, K., and M. S. Timlin (1998), Measuring the strength of ENSO events—How does 1997/98 rank?, *Weather*, *53*, 315–324.

Wolter, K., and M. S. Timlin (2011), El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.Ext), *Int. J. Climatol.*, *31*, 1074–1087, doi:10.1002/joc.2336.

World Meteorological Organization (2002), Standardised verification system (SVS) for long-range forecasts (LRF), New Attachment II-9 to the Manual on the GDPS (WMO-No. 485), vol. 1, Geneva, Switzerland.