Analysis of Atlantic tropical cyclone landfall forecasts in coupled GCMs on seasonal and decadal timescales

Joanne Camp*[1] and Louis-Philippe Caron†[2]

[1]Met Office Hadley Centre, Exeter, UK

[2]Climate Forecasting Unit, Catalan Institute of Climate Sciences (IC3), Barcelona, Spain

February 17, 2016

Abstract

In this chapter we present advances in forecasting Atlantic tropical cyclone (TC) landfall statistics at both seasonal and multi-annual timescales using coupled global climate models. First, we demonstrate potential for forecasting TC landfall frequency on seasonal timescales using the Met Office seasonal forecast system, GloSea5, in some regions: statistically significant skill is found in the Caribbean and moderate skill is found for Florida. In contrast, low skill is found along the U.S. Coast as a whole. We show that the skill over the Caribbean is likely due to a good model response to ENSO forcing. Lack of skill along the U.S. Coast may be due to a weaker influence from ENSO compounded by a low bias in model storm tracks crossing the U.S. coastline. Secondly, we demonstrate that it is possible to construct reliable 4-year mean forecasts

*joanne.camp@metoffice.gov.uk

†louis-philippe.caron@ic3.cat

of landfalling hurricane numbers in the Atlantic using initialised global climate models to predict an index that relies on subpolar gyre temperature and subtropical sea level pressure, two quantities with links to hurricane activity. Furthermore, we give evidence that the forecast system anticipates large changes in at least one of the two components of this index, which suggests that the technique could be used to forecast shifts between active and inactive regimes of hurricane activity in the Atlantic.

1 Introduction

Atlantic tropical cyclone (TC) activity fluctuates on various timescales, ranging from intra-seasonal to decadal (and possibly longer). The El Niño–Southern Oscillation (ENSO) is the dominant climate mode influencing this variability on a year-to-year basis (Camargo et al., 2010), modulating both basin-wide activity as well as hurricane landfalls in the U.S. (e.g. Klotzbach, 2011a) and Caribbean (e.g. Tartaglione et al., 2003; Klotzbach, 2011b). Indeed, it was recognising the link between ENSO and Atlantic hurricane frequency which led to the first Atlantic hurricane seasonal forecasts in the 1980s (Gray, 1984a,b).

Since then, many other climate factors modulating Atlantic hurricane activity have been identified (see Caron et al. (2015a) for an overview) and to the extent that these climate influences are predictable, they can be used (individually or in combination) to estimate upcoming hurricane activity. This is done explicitly in statistical forecasts, which are constructed using past observations of seasonal hurricane activity and precursive climate indices. However, the short period of reliable observational data (typically since the 1970's with the advent of satellite data) can limit the sample size that can be used to develop such statistical models. Furthermore these methods rely on *past* observed relationships between climate predictors and TC activity, which may not remain stationary with changing climate (e.g. Klotzbach, 2007).

Over the last two decades there has been increasing interest in seasonal dynamical predictions of TC activity with initialised general circulation models (GCMs). These have been used to explicitly forecast TC statistics (e.g. Vitart et al., 1997; Camp et al., 2015), as well as predicted parameters (e.g. sea surface temperatures (SSTs)) within statistical forecasting models (e.g. Vecchi et al., 2011). It has long been recognised that even low-resolution climate models (grid spacing of 100 km or coarser at mid-latitudes) are capable of reproducing tropical cyclone-like vortices (Bengtsson et al., 1982, 1996; Walsh and Watterson, 1997; Vitart et al., 1997). These model representations of TCs tend to have lower wind speeds and larger diameters than observed storms, but have realistic climatology in terms of geographical distribution and season of occurrence (Camargo and Wing, 2015). The observed relationship between ENSO and Atlantic hurricane activity is also well captured (e.g. Vitart et al., 1997) and, as a result, the predictive skill of Atlantic TC frequency using such methods is competitive with statistical forecasts (e.g. Vitart et al., 2007; Vecchi et al., 2014). Today, dynamical seasonal predictions of Atlantic TC activity are issued by various centres, including the UK Met Office (Camp et al., 2015), European Centre for Medium-Range Weather Forecasts (ECMWF; Vitart and Stockdale, 2001), Florida State University (LaRow et al., 2010) and the National Oceanic and Atmospheric Administration's (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL; Vecchi et al., 2011, 2014). Such forecasts typically provide predictions of basin-wide numbers of TCs, hurricanes and accumulated cyclone energy (ACE) index (a measure of the combined strength and duration of TCs) during the season; dynamical forecasts of TS landfall are not currently available.

As dynamical models become more advanced in terms of resolution, physics and dynamical core, their ability to resolve characteristics of observed TCs—such as their geographical distribution, track, frequency, interannual variability, structure and intensity—also tend to improve (e.g. Zhao et al., 2009; Caron et al., 2010; Murakami et al., 2012; Strachan et al., 2013; Shaevitz et al., 2014; Roberts et al., 2015; Reed et al., 2015). However, such models

are computationally expensive to run, and therefore it has only been in recent years that increases in supercomputing power have enabled high-resolution GCMs (grid spacing of 50 km or finer at mid-latitudes) to be used for seasonal TC predictions (e.g. Zhao et al., 2010; Chen and Lin, 2011, 2013; Vecchi et al., 2014; Camp et al., 2015). The increase in resolution, combined with more realistic TC tracks, now provides an opportunity for seasonal forecasts of TC landfall risk to be explored.

In comparison, decadal or multi-annual hurricane forecasts are still in their experimental stage. Multi-annual forecasts are already being produced by some catastrophe modelling firms which typically construct them using a range of statistical forecasting models. Recently, it has also been shown (Smith et al., 2010; Vecchi et al., 2013; Caron et al., 2014) that it is possible to produce skilful predictions of Atlantic hurricane activity over lead times of several years using initialised coupled GCMs. For those vulnerable to losses from TC damage, forecasts covering multiple seasons may be more readily incorporated into planning and management strategies, because the lead time allows greater opportunity to integrate forecasts with fixed planning schedules. For example, TC forecasts covering a 5-year horizon can be used in the pricing of contracts by the insurance and reinsurance industry (Jewson et al., 2009), for which hurricane damage can be the leading cause of losses during a given year.

On longer (multi-decadal) timescales, many of the climate factors influencing interannual Atlantic hurricane variability, such as ENSO, tend to average out and prolonged periods of high and low hurricane activity are usually attributed to Atlantic multi-decadal variability (AMV) of SSTs, also referred to as the Atlantic Multi-decadal Oscillation (AMO) (Zhang and Delworth, 2006; Knight et al., 2006; Dunstone et al., 2013; Goldenberg et al., 2001). The link between the AMO/AMV and Atlantic hurricane activity is well documented (Zhang and Delworth, 2006; Vimont and Kossin, 2007; Kossin and Vimont, 2007) and has been shown

to operate through a modulation of various climate factors influencing cyclogenesis: SSTs, vertical wind shear, low-level convergence and low-level vorticity over the tropical Atlantic and a shift in the intertropical convergence zone. Periods of high (low) Atlantic TC activity have been associated with the positive (negative) phase of the AMO/AMV.

In this chapter, we present recent advances in seasonal and multi-annual forecasting of Atlantic hurricane activity using general circulation models, with a particular focus on the skill of these systems at predicting landfall statistics.

2 Seasonal Forecasting of Landfall Risk

Seasonal forecasts of TC landfall risk are not currently issued operationally from dynamical models. While such systems have been shown to have skill forecasting Atlantic basin-wide numbers of TCs (e.g. Vitart et al., 2007), the basin-wide frequency is not strongly correlated with U.S. hurricane landfalls and therefore has limited use as a proxy for forecasting TC landfall frequency. An alternative is to forecast landfall using the model TC tracks directly; however, this also presents problems: low-resolution GCMs—generally used for operational seasonal forecasting—typically tend to simulate larger TC vortices than observed, which can lead to biases in TC track (e.g. Camargo, 2013). Furthmore, observed TC tracks and landfall are also largely governed by weather conditions prior to landfall, which are only predictable on shorter range (e.g. 1–5 days) rather than seasonal timescales. Nevertheless, seasonal average preferences in TC track, such as those associated with ENSO (e.g. Wang et al., 2014), may be predictable on seasonal timescales. High-resolution GCMs, which can better represent observed TC track and geographical distribution, now provide the opportunity for seasonal forecasts of TC landfall risk to be explored further.

Two recent studies—Vecchi et al. (2014) and Camp et al. (2015)—examined the ability

of fully-coupled ocean-atmosphere GCMs with ∼50 km mid-latitude resolution to predict changes in landfall risk in the North Atlantic basin. Both studies, despite using different techniques and models, yielded similar results: significant skill was shown for regional TC landfall predictions in the Caribbean, whereas low skill was found for the U.S. Coast.

Here we expand the work of Camp et al. (2015) to more closely examine landfalls in the North Atlantic basin. The purpose of this assessment is twofold: to investigate potential causes of both the low predictive skill of TC landfalls along the U.S. Coast; and the comparatively higher skill in the Caribbean. To do this we look at various landfall statistics in GloSea5 and observations, including: frequency, interannual variability, genesis locations and the relationship with ENSO. We also examine landfalls over smaller regions of the U.S. (such as the U.S. Gulf Coast) and Caribbean, to identify whether GloSea5 has any sub-regional skill.

2.1 Models, data and analysis

The Met Office Global Seasonal Forecast System 5 (GloSea5; MacLachlan et al., 2014) is used in this study. GloSea5 is a fully coupled ocean-atmosphere GCM: the atmospheric component has a horizontal resolution of 0.83° longitude x 0.55° latitude (N216; ∼53 km at 55°N and ∼93 km at the equator) and 85 levels in the vertical; the ocean component has a horizontal resolution of 0.25° and 75 vertical levels.

The performance of GloSea5 is examined using retrospective forecasts (also known as hindcasts) over the 22-year period June–November 1992–2013. The hindcasts are initialised on three consecutive weeks centred around 1 May (25 April, 1 May and 9 May) using reanalyses from the ECMWF Interim Re-Analysis project (ERA-Interim; Dee et al., 2011). For each of the three weeks, ten ensemble members are run for each year, providing a total of 30 members per year.

TCs are detected and tracked in each ensemble member using a feature-based algorithm (TRACK; Hodges, 1995, 1996, 1999; Bengtsson et al., 2007), with the same configuration as described in Camp et al. (2015). Observed data for the North Atlantic basin are obtained from the National Hurricane Center's best-track Hurricane Database (HURDAT2; Landsea and Franklin, 2013). In this analysis the term "tropical cyclone" is used to describe all observed storms which reach a maximum intensity of tropical storm strength or higher; we have not included the contribution from subtropical storms, which make up a very small portion of the observed database (only 3 such storms were recorded during the study period and none of these made landfall).

A TC in observations and model data is considered to have made landfall when its track—generated from 6-hourly positions of mean-sea-level pressure minima—crosses a coastline. We consider landfalls across a total of 7 coastal regions: the U.S. Coast (further subdivided into the Gulf, Florida and East Coast) and the Caribbean (further subdivided into the eastern and western Caribbean). The coastline boundaries for each of these 7 regions are shown in Figure 1. The U.S. Coast and Caribbean regions are the same as those used in Camp et al. (2015); however, we reproduce the results again here for completeness and comparison. The counting method is as follows: for each region we simply count the number of TCs crossing the coastline. Each storm can only count towards the landfall total in each region once. Therefore, because a single storm can make landfall in more than one region during its lifetime, the total number of landfalling storms in the Caribbean, for example, may not equal the combined total for the eastern and western Caribbean; likewise, the landfall frequency along the U.S. Coast may not equal the combined total for the Gulf, Florida and East coasts.

To assess the relationship between TC landfall frequency and ENSO, observed SST anomalies (SSTA) for the equatorial Pacific Niño3.4 region (120°–170°W, 5°S–5°N) are ob-

tained from the NOAA Climate Prediction Center (CPC, 2015).

2.2 Results

2.2.1 U.S. and Caribbean landfall frequency

Over the 22-year period June–November 1992–2013, a total of 294 observed TCs formed in the North Atlantic (average 13.4 per year) and of these 152 (52%) made landfall in the U.S. and Caribbean (average 6.9 landfalls per year). Along the U.S. Coast, the Gulf and East Coast experienced the greatest average number of landfalls per year (2.0 and 1.8, respectively) and Florida the fewest (1.3); and in the Caribbean, the eastern Caribbean experienced more landfalls (2.7) than the western Caribbean (2.3).

In GloSea5, an average of 6.1 TCs formed in the Atlantic basin per member per year. Of these, an average of 1.8 TCs (29%) made landfall in the U.S. and Caribbean: just over half the percentage of landfalling storms per year compared to observations. Along the U.S. Coast, GloSea5 simulates the greatest average number of landfalls per year over Florida (0.4) and the fewest along the Gulf (0.2) and East (0.3) coasts, which is the opposite pattern to that seen in observations. However, we note that the landfall rates for these regions are low and the sample of 22 years is relatively small, thus the difference may not be statistically significant. In the Caribbean, GloSea5 simulates a greater average number of landfalls per year in the eastern Caribbean (1.1) compared to the western Caribbean (0.4), as seen in observations, although the difference is more pronounced.

2.2.2 U.S. and Caribbean landfall track density

The track density of all TCs that made landfall in the U.S. and Caribbean during June–November 1992–2013 is shown for GloSea5 and observations in Figure 2. A corresponding track density difference (GloSea5 minus observations) is also provided.

In observations, the greatest frequency of landfalling TC tracks is concentrated in the western half of the basin: in the western Atlantic hurricane Main Development Region (MDR; 10–20°N, 20–60°W), Caribbean Sea, Gulf of Mexico and to the east of Florida (Figure 2a). This pattern is well simulated by GloSea5 (Figure 2b); however, the frequency of landfalling TC tracks is much lower than observed, particularly in the western Caribbean Sea, Gulf of Mexico and along the U.S. Gulf Coast (Figure 2c). This may in part be due to a defecit in the total *basin-wide* TC track density in these regions, which was highlighted in Camp et al. (2015) and also seen in other GCMs (e.g. Strazzo et al., 2013; Vecchi et al., 2014; Mei et al., 2014) as well as in regional climate models (Caron and Jones, 2011). Deficits in landfall track density are also seen along the eastern U.S. Coast, suggesting that model storms may not reach higher latitudes as frequently as observed storms and/or that too few storms recurve towards the eastern U.S. Coast from the Atlantic MDR and Caribbean.

The higher frequency of landfalls in Florida and the eastern Caribbean in GloSea5 are likely due to the mean location of model storm tracks, which take a preferential path from the Atlantic MDR towards the eastern Caribbean and Florida (Figure 2b). In contrast, too few storms move from the MDR into the Caribbean Sea and Gulf of Mexico and into higher latitudes along the U.S. East Coast, therefore resulting in a lower landfall frequency in these regions.

2.2.3 Genesis locations

In addition to the low frequency of landfalling TC tracks passing through the Caribbean Sea, Gulf of Mexico and close to the U.S. Coast, it is possible that the number of landfalling storms *forming* in these regions is also too low. To investigate this, we examine the genesis location of all landfalling TCs in the U.S. and Caribbean from June–November 1992–2013.

In order to aid comparison between GloSea5 and observations we divide the genesis density for each year by the total number of landfalling storms. Results are presented in Figure 3.

In observations, the majority of U.S. and Caribbean landfalling storms form in the western tropical Atlantic MDR or close to the coastline in the western Caribbean Sea and Gulf of Mexico. An additional minor peak in TC genesis is found in the western subtropical Atlantic close to the U.S. East Coast. This pattern of genesis is generally well captured by GloSea5; however, the frequency of genesis, relative to the total landfall frequency, is too high in the central MDR and too low in the Gulf of Mexico and off the U.S. East Coast.

In GloSea5, the high frequency of genesis in the MDR, combined with the preference in tracks to move towards the eastern Caribbean and Florida, is the likely cause of the high frequency of landfalls in the latter regions. It may also be the primary reason why the greatest number of landfalls along the U.S. Coast occur in Florida in GloSea5, compared to the U.S. Gulf and East Coast as seen in observations. It could also account for the greater difference in landfall frequency between the eastern and western Caribbean in GloSea5 compared to observations. Conversely, the low genesis frequency in the Gulf of Mexico is the likely reason why the Gulf Coast sees the fewest landfalls in GloSea5, since many observed TCs that form in the Caribbean Sea and Gulf of Mexico later make landfall here (Lyons, 2004). Finally, the low frequency of genesis off the U.S. East Coast, combined with the low frequency of TC tracks from the Atlantic MDR that reach higher latitudes, is the likely cause of the low landfall frequency along the U.S. East Coast in GloSea5, compared to observations.

We can also speculate that the low proportion of landfalling TCs in GloSea5 compared to the total basin-wide count may also be due to preferred regions of TC genesis and tracks. For example, many of the storms in GloSea5 form in the Atlantic MDR; some of these make landfall in the eastern Caribbean and Florida, however many recurve without making

landfall (not shown). In contrast, very few storms form in the Gulf of Mexico, which, in observations, are likely to make landfall along the U.S. Gulf Coast.

2.2.4 Interannual variability

Timeseries of TC landfalls along the U.S. Coast, Caribbean and each of the individual sub-regions are shown for observations and GloSea5 over the period June–November 1992–2013 in Figure 4. Corresponding linear correlations between observations and the GloSea5 ensemble-mean landfall count are shown in Table 1.

Overall, we find moderate, but significant skill (at the 5% level) for predictions of TC landfall frequency in the Caribbean (linear correlation of 0.69), eastern Caribbean (0.52), Florida (0.41) and western Caribbean (0.39). Along the U.S. Coast we find low skill (0.22, not significant) and along the U.S. Gulf and East coasts we find no skill (-0.01 and -0.06, respectively). However, it is worth noting that the landfall regions assessed here are small and the frequency of landfalling storms is low (both in observations and GloSea5), thus small differences between observed and model predicted landfalls, particularly over this short time period, can have a large impact on correlation scores.

2.2.5 Relationship with ENSO

One of the most important factors influencing TC landfall variability on intraseasonal timescales is ENSO. In the tropical North Atlantic, hurricane landfalls in the U.S. and Caribbean are reduced during El Niño (EN) events, whereas they are enhanced during La Niña (LN) events (e.g. Bove et al., 1998; Lyons, 2004; Tartaglione et al., 2003; Larson et al., 2005; Klotzbach, 2011a).

To investigate whether this relationship is represented in GloSea5, we examine the difference in landfall track density in the U.S. and Caribbean between EN and LN events and

compare these to observations. EN and LN events are here defined as years in which the 3-month (August–October) averaged Niño3.4 SSTA is $\geq$0.5°C and $\leq$-0.5°C, respectively. Over the period 1992–2013, five years are classified as EN: 1997, 2002, 2004, 2006 and 2009; and seven years as LN: 1995, 1998, 1999, 2000, 2007, 2010 and 2011. The track density of landfalling TCs during EN and LN events, as well as the track density difference (EN minus LN), is shown for observations and GloSea5 in Figure 5.

In observations, we find a reduction in landfalling TC tracks throughout the MDR, western Atantic and far western Caribbean Sea and Gulf of Mexico during EN events compared to LN events (Figure 5e). However, in contrast to the literature (e.g. Xie et al., 2002; Smith et al., 2007; Klotzbach, 2011a), we do not find a reduction in TC landfalls in the central Caribbean or along the U.S. Coast. This could be due to differences in the classification of EN events (for example, Klotzbach (2011a) do not include 2004, which was a marginal EN year but had strongly favourable conditions for TC development) as well as the small sample size of observed EN and LN years used. Further investigation reveals that either removing 2004 from the analysis or by using observations over a longer period (we used 1950–2014 as this is the longest period for which CPC SST data are available; see Figure 6) provides results that are in better agreement with the literature. We note, however, that the two main regions of lower track density in the western Atlantic and in the western Caribbean Sea and Gulf of Mexico during EN years relative to LN years are evident in both Figures 5e and 6.

In GloSea5 there is a clear reduction in TC landfall frequency throughout the Caribbean and tropical Atlantic during EN events relative to LN events (Figure 5f). Indeed, in the western MDR, GloSea5 shows a statistically significant reduction in track density, which matches the long-term observed response well (Figure 6). Similar observed regions of low track density are also evident over the shorter period 1992–2013 (Figure 5e), although these

are not statistically significant. The good response of TC landfall track frequency to ENSO forcing is the likely reason for relatively high skill of interannual predictions of TC landfall in the Caribbean in GloSea5.

Around the U.S. Coast, GloSea5 shows reduced activity (not statistically significant) along the Gulf and Florida coastline, as seen in Klotzbach (2011a). However, GloSea5 shows a significant reduction in TC landfall frequency around the southern tip of Florida, which is not present in observations (Figures 5e and 6). Along the U.S. East Coast, we find no significant difference in TC landfall frequency between EN and LN events in GloSea5. In observations, we find a small decrease in TC landfalls during EN events relative to LN events over the period 1950–2014 (Figure 6) and slightly enhanced activity during the period 1992–2013 (Figure 5e), although neither of these are statistically significant. The absence of a strong (i.e. statistically significant) response of TC landfalls to ENSO in observations is likely one of the main reasons why skilful seasonal predictions of TC landfall risk along the U.S. Coast are difficult to provide.

2.2.6 Conclusion

TC landfalls in the North Atlantic basin are assessed using a fully-coupled GCM over the period June–November 1992–2013. Overall, we find significant skill for predictions of TC landfall in the Caribbean and moderate skill (not statistically significant) over Florida; however, low skill is found along the U.S. Coast as a whole. In the Caribbean, the GCM shows a realistic response to ENSO forcing and the interannual variability in landfall rates is well simulated for both the eastern and western Caribbean. In contrast, along the U.S. Coast, we find a deficit of landfalling TCs, particularly along the Gulf and East coasts. This is due to a combination of too few storms forming and tracking through the Caribbean and Gulf of Mexico as well as too few storms reaching higher latitudes from the tropical Atlantic MDR. In addition, we find no significant difference in TC landfall rates along the U.S. Coast

between El Niño and La Niña events in either GloSea5 or observations. The absence of a strong relationship between ENSO and U.S. landfalls is one possible reason why skilful seasonal predictions of landfall risk along the U.S. Coast are not presently available. However, the relationship between ENSO and U.S. landfalls is stronger for more intense storms (e.g. Bove et al., 1998), therefore investigating this relationship in GloSea5 would be worthwhile.

3 Decadal Forecasting of Landfall Risk

While tracking tropical cyclone-like disturbances directly (as in the previous section) has the merit of being very intuitive, it is also very computationally expensive and can be difficult to apply in the context of multi-model ensemble decadal forecast studies. This is in large part due to the large volume of data required wherein a set of retrospective decadal predictions typically requires running 5200 years of simulations (10 members, 52 start dates, 10 forecast years) combined with the fact that tracking storms requires high frequency ($\sim$12 hours or less) of surface and multiple levels of atmospheric data. As such, it is of interest to devise alternative techniques by which TC activity can easily be estimated using large-scale fields that are readily available by way of international efforts such as CMIP. Such a first attempt was made by Vecchi et al. (2013), which investigated the use of a statistical relationship between tropical SSTs and Atlantic hurricane activity, however they concluded that most of the skill they had obtained originated from persisting the initial conditions. Of particular interest was the failure to predict the upward shift in hurricane activity that occurred in the mid-1990s. Skilful predictions of such consequential climate shifts are a prerequisite for decadal forecasts to be considered useful.

Smith et al. (2010) and Dunstone et al. (2011) argue that much of the long-term hurricane predictability that they identified in their GCM could be traced back to the North Atlantic

sub-polar gyre (SPG), a region where initialised climate simulations show a high level of skill at the multi-annual timescale. They showed that changes in surface temperature over the SPG could be linked to changes in the atmospheric circulation over the Atlantic MDR, more specifically the ascending branch of the Hadley circulation, which in turn impacted TC formation in their climate simulations. Caron et al. (2015b) suggested taking advantage of the relatively high skill displayed by decadal forecast systems over the northern North Atlantic by using a proxy index of the AMV (Klotzbach and Gray, 2008) to produce multi-annual forecasts of Atlantic TC activity. Constructed using SSTs over the SPG as well as tropical and extra-tropical Atlantic sea level pressure (SLPs), this index has been shown to vary with Atlantic hurricane activity over the course of the 20th century (ibid.) and is defined such that

$$AMV_{index} = SSTA - SLPA \tag{1}$$

where $SSTA$ is the standardised annual mean SST anomaly over the North Atlantic sub-polar gyre ($50°N$, $60°N$, $50°W$, $10°W$) and $SLPA$ is the standardised annual mean sea level pressure anomaly over the tropical and extra-tropical North Atlantic ($0°$, $50°N$, $70°W$, $10°W$; see boxes in figure 8). Caron et al. (2015b) showed that an ensemble of initialised forecasts showed skill at predicting this index over a 5-year horizon, which led to useful information on multi-annual levels of accumulated cyclone energy (ACE). Here, we expand on this work by investigating whether this technique can further be used to make reliable forecasts of hurricane landfall statistics over multiple seasons.

3.1 Methodology

3.1.1 Models and Observational Data

Our analysis relies on two different multi-model ensembles (MMEs) of 10-year long simulations performed within the context of the CMIP5 project (GFDL-CM2.1 (Delworth et al.,

2006); HadCM3 (Gordon et al., 2000); MIROC5 (Watanabe et al., 2010)) and SPECS project (MPI-ESM (Matei et al., 2012)), for a total of four forecast systems, each with start dates available every year from 1961 to 2010. The first ensemble is constructed using simulations initialised with contemporaneous observations, thus aligning the simulated natural variability with the observed variability. The multi-model ensemble-mean anomalies are computed by giving an equal weight to each model mean, regardless of the number of ensemble members available for a particular model. This ensemble is referred to as Ini. The second ensemble is composed of 10-year long simulations constructed using the CMIP5 historical and RCP 4.5 scenario (Meinshausen et al., 2011) simulations. This second ensemble is referred to as NoIni. The difference in skill between Ini and NoIni is a measure of the added-value of initialisation. The number of members for each ensemble is shown in table 2. We note that we have added one start date and increased the ensemble size using new simulations that have become available since Caron et al. (2015b).

In both ensembles, external forcings (greenhouse gases, solar activity, stratospheric aerosols associated with volcanic eruptions and anthropogenic aerosols) are taken from observations for the period 1961–2005 and the RCP 4.5 scenario afterwards. Because any significant unexpected changes in external forcing (e.g. large volcanic eruption) cannot be taken into account in a true forecast, the skill obtained by using observed forcings is somewhat overestimated. However, the difference in skill between Ini and NoIni should not be affected. Unlike Caron et al. (2015b) which showed the skill of 5-year mean forecasts, here we show results of 4-year mean forecasts, using only forecast years 2–5, which is a standard procedure when evaluating decadal forecasts (Goddard et al., 2013).

Finally, reference data for SSTAs are taken from NOAA's Extended Reconstructed Sea Surface Temperatures (ERSST; version 3b) (Smith et al., 2008), the SLPAs are taken from the JRA-55 reanalyses (Kobayashi et al., 2015) and the hurricane data are taken from HUR-

DAT2 (Landsea and Franklin, 2013). The number of landfall events is calculated as the number of times hurricane tracks (linearly interpolated to 1-hour) cross over onto a land mass. A storm is considered to have made multiple landfalls if it continued back over a large body of open water (e.g. Gulf of Mexico, Atlantic ocean) after a first landfall (e.g. a storm passing over Florida, the Gulf of Mexico and Louisiana would be considered as two landfalls). Storms were required to be of hurricane strength upon landfall to be considered, but could have transitioned to extra-tropical status. Figure 7 shows all of the hurricane landfall locations for the period 1962–2014.

3.1.2 Forecast Skill Assessment

Different measures are used to evaluate the skill of the forecasts: the Anomaly Correlation Coefficient (ACC), the Root Mean Square Skill Score (RMSSS) and the Brier Skill Score (BSS).

ACCs are computed by correlating the MME anomalies with the observed anomalies across the start date dimension, using both standard Pearson's correlation and Kendall's rank correlation. The RMSSS shows the improvement relative to a climatological forecast and is defined as

$$RMSSS = 1 - \frac{RMSE_{for}}{RMSE_{clim}} \tag{2}$$

where $RMSE_{for}$ and $RMSE_{clim}$ are, respectively, the root mean square error of our forecast and of the reference forecast. RMSSS = 1 shows a perfect forecast and RMSSS $\leq$0, a forecast with no improvement over the benchmark (in this case, climatology). In both cases, a t-test, after a Fisher-Z transformation, is performed to assess the significance level. Autocorrelation in the various timeseries is accounted for by following Von Storch and Zwiers (2001).

Finally, the BSS measures the improvement of a probabilistic forecast relative to a bench-

mark, in this case either climatology or NoIni. It is defined as

$$BSS = 1 - \frac{BS_{for}}{BS_{ref}} \tag{3}$$

where BS_{for} and BS_{ref} are the Brier scores of the actual and reference forecast, respectively. The Brier score itself is defined as the average of the squared differences between each forecast probability and the corresponding binary observations (1 if the event is observed, 0 if it is not). Our binary probabilities are constructed as follows: for each 4-year period we calculate, using forecast years 2–5 of each model, the probability that the index will be positive based on the number of members for which the index is positive. We then average the model probabilities to obtain the probability of the MMEs, giving an equal weight to each model, regardless of the number of ensemble members available for that particular model. This result is then compared to the observed 4-year mean number of hurricane landfalls.

3.2 Multiannual Forecast Skill of AMV index

Figure 8 shows the ACCs between the Ini MME and observations for 4-year mean SSTAs (8a) and SLPAs (8b). For SSTA, high and significant skill is found over almost the entire domain, but a large portion of this skill originates from the radiatively-driven upward trend in SSTs captured by the MME. This can be seen in figure 8c, which shows the difference in ACCs between the Ini and NoIni MMEs. However, for a large part of the region considered in the construction of the AMV index (black box), the skill of the Ini MME is significantly higher than the NoIni MME. In comparison, skill for SLPA is more modest (8b), but subtracting the ACC from the NoIni MME (8d) has a much lesser impact than for SSTA since the NoIni MME shows very poor skill at predicting SLPA. In effect, figure 8 shows that initialised CGMs have skill at capturing multi-annual variations in the large-scale fields that are used in the construction of the AMV index.

Figure 9 shows the timeseries for standardised 4-year mean SSTA (9a) and SLPA (9b) for both the Ini (red) and the NoIni (blue) MME as well as the observational reference (black). Both time series show significant ACCs for the Ini ensemble, but only the SSTA forecasts show a RMSSS greater than 0. Figure 9c shows the value of 4-year mean forecasted index for both ensembles, and for observations. For the Ini ensemble both the ACC and the RMSSS are statistically significant, with values of 0.79 and 0.37 respectively. The timing of the downward shift (1969–1970) towards negative values is fairly well captured while the upward shift of the mid 1990's occurs somewhat too early, the latter being driven by the early drop in simulated SLPA in the Ini MME.

Figure 10 (top) compares the 4-year mean index forecasted by the Ini MME and the 4-year mean number of hurricane landfalls (expressed as anomalies) observed during the corresponding period. Both the linear (0.65) and the ranked (0.48) correlation coefficients are statistically significant, confirming the skill of our MME at forecasting multi-annual hurricane landfall numbers. In fact, the index explains ∼40% of the variability in total hurricane landfalls in the Atlantic basin. We also evaluate the skill of our forecasts using a binary probability forecast verification technique, wherein a high ($> 50\%$) probability of a positive index is suggestive of above normal hurricane landfall numbers. The result is shown in the form of a reliability diagram (figure 10, bottom), which compares the forecasted frequencies with the actual observed frequencies. The reliability diagram is constructed by grouping the probabilistic forecasts into three bins (0–33%, 34%–66%, 67%–100%) on the horizontal axis according to the probability derived from the MME (as described in the methodology section). For perfect reliability, the forecast probability and the frequency of occurrence should be equal and the predictions should align along the diagonal (solid line in the figure). However, due to the finite number of predictions, a forecast system may still be deemed reliable even if its predictions do not lie precisely along the diagonal. To address this is-

sue, we include consistency bars, showing the 5% and 95% quantiles. A histogram with the distribution of the forecasts within the different bins is shown in the bottom right of the diagram and the BSSs (using both climatology and NoIni MME as a benchmark) are shown in the upper-left corner.

The forecasts are reliable, with all points of the Ini MME lying close to the diagonal, within the consistency bars. Furthermore, the Ini ensemble is capable of predicting relatively high or low probabilities of 4-year mean hurricane landfalls, as evidenced by the fairly even number of predictions in each of the three forecast categories. In effect, this graph shows that, above average periods of activity tend to occur when the MME forecasts return high probabilities (>66%) that such high activity will occur. Similarly, low activity periods tend to occur when the MME returns low probabilities (<33%) of high activity. Finally, BSSs are both positive and significant, thus further confirming the skill of our MME forecasts.

3.3 Persistence or Predictability?

As stated earlier, for decadal forecasts to be truly useful, their skill must not originate solely from the persistence of anomalies introduced at the initialisation stage. Particularly desirable (arguably, necessary) is the ability of the MME to predict the shifts between active and inactive regimes. For the period covered here, such a downward shift was observed in 1969–1970 while an upward shift was observed in 1994–1995. We note that these shifts in activity were matched by similar shifts in the index itself (see figure 9c and 10). Both shifts in the AMV index were driven by simultaneous changes in SLPA and SSTA (figure 9a,b): the first shift is driven by a decrease in SSTA and an increase in SLPA (the opposite is true for the second shift). Thus, for forecasts initialised in the years leading up to the 1969–1970 shift, we would expect, on average, the SSTA (SLPA) of the later years to be smaller (larger) than that of the earlier years. For the upward shift of the mid-90's, we would expect the opposite.

In figure 11, we show the distribution of the difference between the average of forecast years 2–5 and that of forecast year 1 for both SSTA and SLPA using i) hindcasts from all start dates (gray), ii) only hindcasts initialised in the years leading up to the downward shift (1965–1968; blue) and iii) only hindcasts initialised in the years leading up to the upward shift (start dates: 1990–1993; red). For SLPA (11b), the distribution, when all the hindcasts are considered, is centered around zero. When only the years leading up to the shifts are considered, small differences in the distribution are observed, but these differences are not statistically significant. This suggests that most of the skill in SLPA results from persistence, since there is no tendency towards higher (lower) SLPAs in the years leading up to a decrease (increase) in AMV index value. However, for SSTA, a clear shift in the distribution towards lower (1969–1970 shift) and higher (1994–1995) values can be seen. A Kolmogorov-Smirnov test confirms that these differences are significant at the 1% level. This suggests that the MME has some predictive power at the multi-annual level for SSTs over this particular region. This skill in initialised MME has been identified in previous studies (Robson et al., 2012, 2014) and has been linked to the ability of the CGMs at capturing the ocean dynamics of the Atlantic Meridional Overturning Circulation (AMOC). This latter result suggests that this technique could indeed be used to predict shifts in prolonged periods of high or low hurricane activity in the Atlantic.

3.4 Discussion

The results presented here suggest that initialised GCMs offer an opportunity to develop reliable forecasts of multi-annual ($\sim$5 years) hurricane landfall statistics in the Atlantic basin and could predict shifts between prolonged periods of high and low hurricane activity. There are caveats to this approach however. First, the skill at predicting one of the components (SLPA) of the AMV index originates from persistence and, as such, shows no true predictive skill. This might change with improvements in initialisation procedures and climate models,

but it might also be the case that changes in SLPA over the tropical and extra-tropical Atlantic are not predictable at the multiannual timescale. Furthermore, the forecasted quantity is an index as opposed to an estimated storm number. In a climate service context, the latter is much easier to use (e.g. insurance loss models expect a storm number as input, not an index). Both issues could potentially be addressed by using a multiple regression technique (similar to that of Vecchi et al. (2011)), where SLPA and SSTA are used as predictors for hurricane landfall numbers and wherein the variable with the best skill (SSTA) is given more weight than the one with more limited skill (SLPA). This approach is likely to improve on the results shown here, and such work is currently underway. Nonetheless, even in its current form, this technique appears useful in estimating upcoming hurricane activity levels and could be used to shed some light on whether we have indeed entered a new era of lower Atlantic TC activity (Klotzbach et al., 2015).

4 Concluding Remarks

In the last few years, significant progress has been made both in improving global climate models and their representation of TCs and in understanding the TC-climate connection. This has led to the development of dynamical forecasts of TC activity at various timescales, with two such examples given here. Skill of dynamical seasonal forecasts over the entire Atlantic is already fairly high for basin-wide activity (Zhao et al., 2010; Chen and Lin, 2011, 2013; Vecchi et al., 2014; Camp et al., 2015); however, operational seasonal forecasts of TC landfall are not presently available. In this study we show that a high-resolution GCM shows promising skill for predictions of TC landfall in the Caribbean region, likely due to a good model response of TC landfall frequency to ENSO forcing. However, along the U.S. Coast, the absence of a strong observed response of TC landfalls to ENSO, combined with a defecit of landfalling storms in the model, limits the skill of seasonal TC landfall forecasts for the U.S. Coast using the present forecasting system. Nevertheless, improvements in GCM reso-

lution and physics offer the opportunity for seasonal TC forecasts of landfall risk along the U.S. Coast to be investigated further.

Recently, multi-annual predictions of Atlantic TC activity using GCMs have been developed and show promising skill (e.g. Smith et al., 2010; Vecchi et al., 2013; Caron et al., 2014, 2015b). In the present study we also show that an initialised GCM offers the opportunity to provide reliable forecasts of hurricane landfall statistics on ∼5 year timescales. Such forecasts maybe of use to those vulnerable to TC damage and losses, but require decisions to be made at lead times longer than presently available from seasonal forecasts (such as in the insurance and reinsurance industry). Furthermore, with the continuous increase in computing power and the improvement in CGCMs, multi-model ensembles of dynamical forecasts should become possible, further improving the skill of such forecasts.

Acknowledgements

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We would also like to thank the Earth System Research Laboratory (NOAA) and the Japan Meteorological Agency for making their data available, and Katherine Barrett for proofreading this manuscript. JC acknowledges financial support from the UK Public Weather Service, NSF of China (NSFC) grants (40805028) and China Meteorological Special Project (GYHY201506013). LPC acknowledges financial support from the Ministerio de Economía y Competitividad (MINECO; project CGL2014-55764-R), Risk Prediction Initiative at BIOS (grant number RPI2.0-2013-CARON) and from

584 the EU-funded SPECS project (grant number 308378). Both authors wish to thank Dr
585 Philip Klotzbach and one anonymous reviewer for their valuable comments for improving
586 the manuscript.

References

Bengtsson, L., Bottger, H., and Kanamitsu, M. (1982). Simulation of hurricane-type vortices in a general circulation model. *Tellus A*, 34:440–457.

Bengtsson, L., Botzet, M., and Esch, M. (1996). Will greenhouse gas-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes? *Tellus A*, 48:57–73.

Bengtsson, L., Hodges, K. I., Esch, M., Keenlyside, N., Kornblueh, L., Luo, j.-J., and Yamagata, T. (2007). How May Tropical Cyclones Change in a Warmer Climate? *Tellus A*, 59(4):539–561.

Bove, M. C., O'Brien, J. J., Eisner, J. B., Landsea, C. W., and Niu, X. (1998). Effect of El Niño on U.S. Landfalling Hurricanes, Revisited. *Bull. Amer. Meteor. Soc.*, 79(11):2477–2482.

Camargo, S. J. (2013). Global and Regional Aspects of Tropical Cyclone Activity in the CMIP5 Models. *J. Climate*, 26(24):9880–9902.

Camargo, S. J., Sobel, A. H., Barnston, A. G., and Klotzbach, P. J. (2010). *The Influence of Natural Climate Variability on Tropical Cyclones, and Seasonal Forecasts of Tropical Cyclone Activity*, volume 4, pages 325–360. World Scientific.

Camargo, S. J. and Wing, A. A. (2015). Tropical cyclones in climate models. *WIREs Clim Change*, page n/a.

Camp, J., Roberts, M., MacLachlan, C., Wallace, E., Hermanson, L., Brookshaw, A., Arribas, A., and Scaife, A. A. (2015). Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Q.J.R. Meteorol. Soc.*, page n/a.

Caron, L.-P., Boudreault, M., and Bruyère, C. L. (2015a). Large-scale control of Atlantic

610 tropical cyclone activity as a function of the Atlantic Multi-decadal Oscillation phase.
611 *Climate Dynamics*, 44:1801–1821.

612 Caron, L.-P., Hermanson, L., and Doblas-Reyes, F. J. (2015b). Multiannual forecasts of At-
613 lantic U.S. tropical cyclone wind damage potential. *Geophysical Research Letters*, 42:2417–
614 2425.

615 Caron, L.-P. and Jones, C. G. (2011). Understanding and simulating the link between African
616 easterly waves and Atlantic tropical cyclones using a regional climate model: the role of
617 domain size and lateral boundary conditions. *Climate Dynamics*, 39(1-2):113–135.

618 Caron, L.-P., Jones, C. G., and Doblas-Reyes, F. (2014). Multi-year prediction skill of
619 Atlantic hurricane activity in CMIP5 decadal hindcasts. *Climate Dynamics*, 42:2675–
620 2690.

621 Caron, L.-P., Jones, C. G., and Winger, K. (2010). Impact of resolution and downscaling
622 technique in simulating recent Atlantic tropical cylone activity. *Climate Dynamics*, 37(5-
623 6):869–892.

624 Chen, J.-H. and Lin, S.-J. (2011). The remarkable predictability of inter-annual variability
625 of Atlantic hurricanes during the past decade. *Geophys. Res. Lett.*, 38(11):L11804+.

626 Chen, J.-H. and Lin, S.-J. (2013). Seasonal Predictions of Tropical Cyclones Using a 25-km-
627 Resolution General Circulation Model. *J. Climate*, 26(2):380–398.

628 CPC (2015). Historical El Nino/ La Nina episodes (1950-present). CPC. [Available online at
629 http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml].

630 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae,
631 U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de
632 Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J.,
633 Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler,

M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F. (2011). The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System. *Q.J.R. Meteorol. Soc.*, 137(656):553–597.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K. A., Durachta, J. W., Findell, K. L., Ginoux, P., Gnanadesikan, A., Gordon, C. T., Griffies, S. M., Gudgel, R., Harrison, M. J., Held, I. M., Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhorst, A. R., Lee, H.-C., Lin, S.-J., Lu, J., Malyshev, S. L., Milly, P. C. D., Ramaswamy, V., Russell, J., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J. J., Spelman, M. J., Stern, W. F., Winton, M., Wittenberg, A. T., Wyman, B., and Zeng, F. (2006). GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. *J. Climate*, 19:643–674.

Dunstone, N. J., Smith, D. M., Booth, B. B. B., Hermanson, L., and Eade, R. (2013). Anthropogenic aerosol forcing of Atlantic tropical storms. *Nature Geoscience*, 6(7):1–6.

Dunstone, N. J., Smith, D. M., and Eade, R. (2011). Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophysical Research Letters*, 38:L14701.

Goddard, L., Kumar, a., Solomon, a., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., Mason, S. J., Kirtman, B. P., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C. a. T., Stephenson, D. B., Meehl, G. a., Stockdale, T., Burgman, R., Greene, a. M., Kushnir, Y., Newman, M., Carton, J., Fukumori, I., and Delworth, T. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40(1-2):245–272.

Goldenberg, S. B., Landsea, C. W., Mestas-Nunez, A. M., and Gray, W. M. (2001).

659 The recent increase in Atlantic hurricane activity: Causes and implications. *Science*,
660 293(5529):474–479.

661 Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J., Johns, T., Mitchell, J., and
662 Wood, R. (2000). The simulation of SST, sea ice extents and ocean heat transports in a
663 version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*,
664 16:147–168.

665 Gray, W. M. (1984a). Atlantic Seasonal Hurricane Frequency. Part I: El Niño and 30 mb
666 Quasi-Biennial Oscillation influences. *Monthly Weather Review*, 112:1649–1668.

667 Gray, W. M. (1984b). Atlantic Seasonal Hurricane Frequency. Part II: Forecasting its Vari-
668 ability. *Monthly Weather Review*, pages 1669–1683.

669 Hodges, K. I. (1995). Feature Tracking on the Unit Sphere. *Mon. Wea. Rev.*, 123(12):3458–
670 3465.

671 Hodges, K. I. (1996). Spherical Nonparametric Estimators Applied to the UGAMP Model
672 Integration for AMIP. *Mon. Wea. Rev.*, 124(12):2914–2932.

673 Hodges, K. I. (1999). Adaptive Constraints for Feature Tracking. *Mon. Wea. Rev.*,
674 127(6):1362–1373.

675 Jewson, S., Bellone, E., Khare, S., Laepple, T., Lonfat, M., Shay, A. O., Penzer, J., and
676 Coughlin, K. (2009). 5-Year Prediction of the number of Hurricanes which make U.S.
677 Landfall. In Elsner, J. B. and Jagger, T. H., editors, *Hurricanes and Climate Change*,
678 pages 73–99. Springer US.

679 Klotzbach, P., Gray, W., and Fogarty, C. (2015). Active Atlantic hurricane era at its end?
680 *Nature Geosci*, 8(10):737–738.

681 Klotzbach, P. J. (2007). Revised Prediction of Seasonal Atlantic Basin Tropical Cyclone
682 Activity from 1 August. *Wea. Forecasting*, 22(5):937–949.

683 Klotzbach, P. J. (2011a). El Niño-Southern Oscillation's Impact on Atlantic Basin Hurricanes
684 and U.S. Landfalls. *J. Climate*, 24(4):1252–1263.

685 Klotzbach, P. J. (2011b). The Influence of El Niño-Southern Oscillation and the Atlantic
686 Multidecadal Oscillation on Caribbean Tropical Cyclone Activity. *J. Climate*, 24(3):721–
687 731.

688 Klotzbach, P. J. and Gray, W. M. (2008). Multidecadal Variability in North Atlantic Tropical
689 Cyclone Activity. *Journal of Climate*, 21(15):3929–3935.

690 Knight, J. R., Folland, C. K., and Scaife, A. A. (2006). Climate impacts of the Atlantic
691 Multidecadal Oscillation. *Geophysical Research Letters*, 33(17):L17706.

692 Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori,
693 H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K. (2015). The JRA-55 Reanal-
694 ysis: General Specifications and Basic Characteristics. *J. Meteor. Soc. Japan*, 93:5–48.

695 Kossin, J. P. and Vimont, D. J. (2007). A More General Framework for Understanding At-
696 lantic Hurricane Variability and Trends. *Bulletin of the American Meteorological Society*,
697 88(11):1767–1781.

698 Landsea, C. W. and Franklin, J. L. (2013). Atlantic Hurricane Database Uncertainty and
699 Presentation of a New Database Format. *Mon. Wea. Rev.*, 141(10):3576–3592.

700 LaRow, T. E., Stefanova, L., Shin, D.-W., and Cocke, S. (2010). Seasonal Atlantic Tropical
701 Cyclone Hindcasting/Forecasting Using Two Sea Surface Temperature Datasets. *Geophys.*
702 *Res. Lett.*, 37(2):L02804+.

703 Larson, J., Zhou, Y., and Higgins, R. W. (2005). Characteristics of Landfalling Tropical
704 Cyclones in the United States and Mexico: Climatology and Interannual Variability. *J.*
705 *Climate*, 18(8):1247–1262.

706 Lyons, S. W. (2004). U.S. Tropical Cyclone Landfall Variability: 1950-2002. *Wea. Forecast-*
707 *ing*, 19(2):473–480.

708 MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A.,
709 Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec,
710 G. (2014). Global Seasonal Forecast System Version 5 (GloSea5): a High-Resolution
711 Seasonal Forecast System. *Q.J.R. Meteorol. Soc.*, page n/a.

712 Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., Haak, H., and Marotzke, J. (2012). Two
713 Tales of Initializing Decadal Climate Prediction Experiments with the ECHAM5/MPI-OM
714 Model. *Journal of Climate*, 25(24):8502–8523.

715 Mei, W., Xie, S.-P., and Zhao, M. (2014). Variability of Tropical Cyclone Track Density in the
716 North Atlantic: Observations and High-Resolution Simulations. *J. Climate*, 27(13):4797–
717 4814.

718 Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque,
719 J.-F., Matsumoto, K., Montzka, S. a., Raper, S. C. B., Riahi, K., Thomson, a., Velders,
720 G. J. M., and Vuuren, D. P. (2011). The RCP greenhouse gas concentrations and their
721 extensions from 1765 to 2300. *Climatic Change*, 109(1-2):213–241.

722 Murakami, H., Wang, Y., Yoshimura, H., Mizuta, R., Sugi, M., Shindo, E., Adachi, Y.,
723 Yukimoto, S., Hosaka, M., Kusunoki, S., Ose, T., and Kitoh, A. (2012). Future Changes
724 in Tropical Cyclone Activity Projected by the New High-Resolution MRI-AGCM*. *J.*
725 *Climate*, 25(9):3237–3260.

726 Reed, K. A., Bacmeister, J. T., Rosenbloom, N. A., Wehner, M. F., Bates, S. C., Lauritzen,
727 P. H., Truesdale, J. E., and Hannay, C. (2015). Impact of the dynamical core on the direct
728 simulation of tropical cyclones in a high-resolution global model. *Geophys. Res. Lett.*,
729 42(9):2015GL063974+.

730 Roberts, M. J., Vidale, P. L., Mizielinski, M. S., Demory, M.-E., Schiemann, R., Strachan, J.,
731 Hodges, K., Bell, R., and Camp, J. (2015). Tropical Cyclones in the UPSCALE Ensemble
732 of High-Resolution Global Climate Models. *J. Climate*, 28(2):574–596.

733 Robson, J., Sutton, R., Lohmann, K., Smith, D., and Palmer, M. D. (2012). Causes of
734 the Rapid Warming of the North Atlantic Ocean in the Mid-1990s. *Journal of Climate*,
735 25(12):4116–4134.

736 Robson, J., Sutton, R., and Smith, D. (2014). Decadal predictions of the cooling and
737 freshening of the North Atlantic in the 1960s and the role of ocean circulation. *Climate*
738 *Dynamics*, 42(9-10):2353–2365.

739 Shaevitz, D. A., Camargo, S. J., Sobel, A. H., Jonas, J. A., Kim, D., Kumar, A., LaRow,
740 T. E., Lim, Y.-K., Murakami, H., Reed, K. A., Roberts, M. J., Scoccimarro, E., Vidale,
741 P. L., Wang, H., Wehner, M. F., Zhao, M., and Henderson, N. (2014). Characteristics of
742 tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth*
743 *Syst.*, 6(4):1154–1172.

744 Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H., and
745 Scaife, A. A. (2010). Skilful multi-year predictions of Atlantic hurricane frequency. *Nature*
746 *Geoscience*, 3(12):846–849.

747 Smith, S. R., Brolley, J., O'Brien, J. J., and Tartaglione, C. A. (2007). ENSO's Impact on
748 Regional U.S. Hurricane Activity. *J. Climate*, 20(7):1404–1414.

749 Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J. (2008). Improvements
750 to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006).
751 *Journal of Climate*, 21(10):2283–2296.

752 Strachan, J., Vidale, P. L., Hodges, K., Roberts, M., and Demory, M.-E. (2013). Investi-
753 gating Global Tropical Cyclone Activity with a Hierarchy of AGCMs: The Role of Model
754 Resolution. *J. Climate*, 26(1):133–152.

Strazzo, S., Elsner, J. B., LaRow, T., Halperin, D. J., and Zhao, M. (2013). Observed versus GCM-Generated Local Tropical Cyclone Frequency: Comparisons Using a Spatial Lattice. *J. Climate*, 26(21):8257–8268.

Tartaglione, C. A., Smith, S. R., and O'Brien, J. J. (2003). ENSO Impact on Hurricane Landfall Probabilities for the Caribbean. *J. Climate*, 16(17):2925–2931.

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., Zeng, F., Anderson, W., Balaji, V., Dixon, K., Jia, L., Kim, H. S., Krishnamurthy, L., Msadek, R., Stern, W. F., Underwood, S. D., Villarini, G., Yang, X., and Zhang, S. (2014). On the Seasonal Forecasting of Regional Tropical Cyclone Activity. *J. Climate*, 27(21):7994–8016.

Vecchi, G. A., Msadek, R., Anderson, W., Chang, Y.-S., Delworth, T., Dixon, K., Gudgel, R., Rosati, A., Stern, B., Villarini, G., Wittenberg, A., Yang, X., Zeng, F., Zhang, R., and Zhang, S. (2013). Multi-year Predictions of North Atlantic Hurricane Frequency: Promise and limitations. *Journal of Climate*, 26:5337–5357.

Vecchi, G. A., Zhao, M., Wang, H., Villarini, G., Rosati, A., Kumar, A., Held, I. M., and Gudgel, R. (2011). Statistical-Dynamical Predictions of Seasonal North Atlantic Hurricane Activity. *Mon. Wea. Rev.*, 139(4):1070–1082.

Vimont, D. J. and Kossin, J. P. (2007). The Atlantic Meridional Mode and hurricane activity. *Geophysical Research Letters*, 34(7):L07709.

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

Vitart, F., Huddleston, M. R., Déqué, M., Peake, D., Palmer, T. N., Stockdale, T. N., Davey, M. K., Ineson, S., and Weisheimer, A. (2007). Dynamically-Based Seasonal Forecasts of Atlantic Tropical Storm Activity Issued in June by EUROSIP. *Geophys. Res. Lett.*, 34(16):L16815+.

780 Vitart, F. and Stockdale, T. N. (2001). Seasonal Forecasting of Tropical Storms Using
781 Coupled GCM Integrations. *Mon. Wea. Rev.*, 129(10):2521–2537.

782 Von Storch, H. and Zwiers, F. (2001). *Statistical Analysis in Climate Research*. Cambridge
783 University Press, Cambridge, UK.

784 Walsh, K. and Watterson, I. G. (1997). Tropical cyclone-like vortices in a limited area model:
785 Comparison with observed climatology. *Journal of Climate*, 10(9):2240–2259.

786 Wang, H., Long, L., Kumar, A., Wang, W., Schemm, J.-K. E., Zhao, M., Vecchi, G. A.,
787 Larow, T. E., Lim, Y.-K., Schubert, S. D., Shaevitz, D. A., Camargo, S. J., Henderson,
788 N., Kim, D., Jonas, J. A., and Walsh, K. J. E. (2014). How Well Do Global Climate
789 Models Simulate the Variability of Atlantic Tropical Cyclones Associated with ENSO? *J.*
790 *Climate*, 27(15):5673–5692.

791 Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T.,
792 Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa,
793 T., Hasumi, H., Tatebe, H., and Kimoto, M. (2010). Improved climate simulation by
794 MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, 23:6312–6335.

795 Xie, L., Pietrafesa, L., and Wu, K. (2002). Interannual and decadal variability of landfalling
796 tropical cyclones in the southeast coastal states of the united states. 19(4):677–686.

797 Zhang, R. and Delworth, T. L. (2006). Impact of Atlantic multidecadal oscillations on
798 India/Sahel rainfall and Atlantic hurricanes. *Geophysical Research Letters*, 33(17):L17712.

799 Zhao, M., Held, I. M., Lin, S.-J., and Vecchi, G. A. (2009). Simulations of Global Hurricane
800 Climatology, Interannual Variability, and Response to Global Warming Using a 50-km
801 Resolution GCM. *J. Climate*, 22(24):6653–6678.

802 Zhao, M., Held, I. M., and Vecchi, G. A. (2010). Retrospective Forecasts of the Hurricane

Season Using a Global Atmospheric Model Assuming Persistence of SST Anomalies. *Mon. Wea. Rev.*, 138(10):3858–3868.

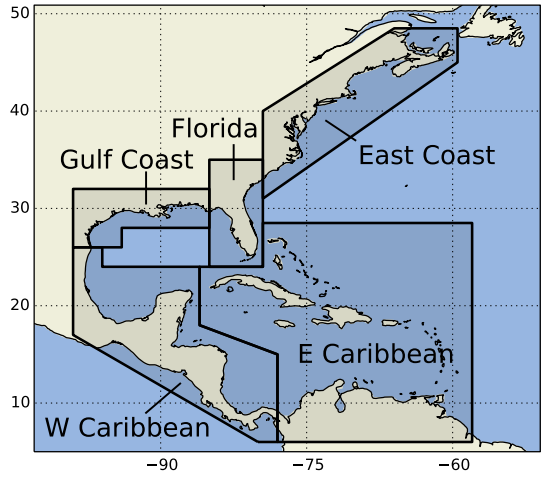

Figure 1: Tropical cyclone landfall regions: U.S. Gulf Coast, Florida, East Coast (including Nova Scotia), western Caribbean and eastern Caribbean. The "U.S. Coast" encompasses the U.S. Gulf, Florida and East Coast; and the "Caribbean" encompasses the eastern and western Caribbean. Note a TC is considered to have made landfall when its track crosses the coastline within a region boundary. Both the U.S. Coast and Caribbean regions are the same as those defined in Camp et al. (2015).

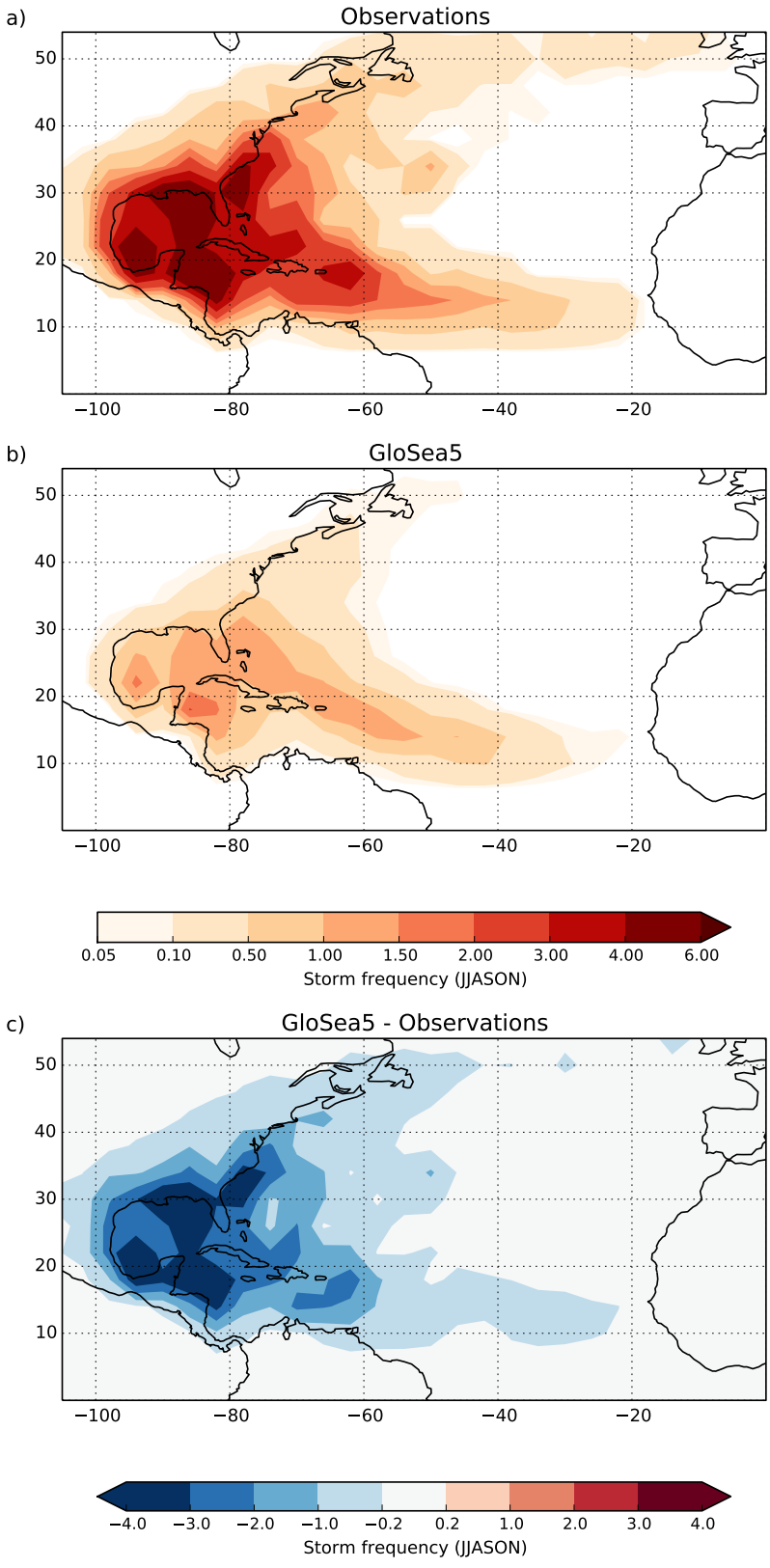

Figure 2: Average seasonal track density (transits per 4°x4° box) of U.S. and Caribbean landfalling TCs in a) observations, b) GloSea5 and c) GloSea5 minus observations over the period June–November (JJASON) 1992–2013. GloSea5 results are averaged over all ensemble members.

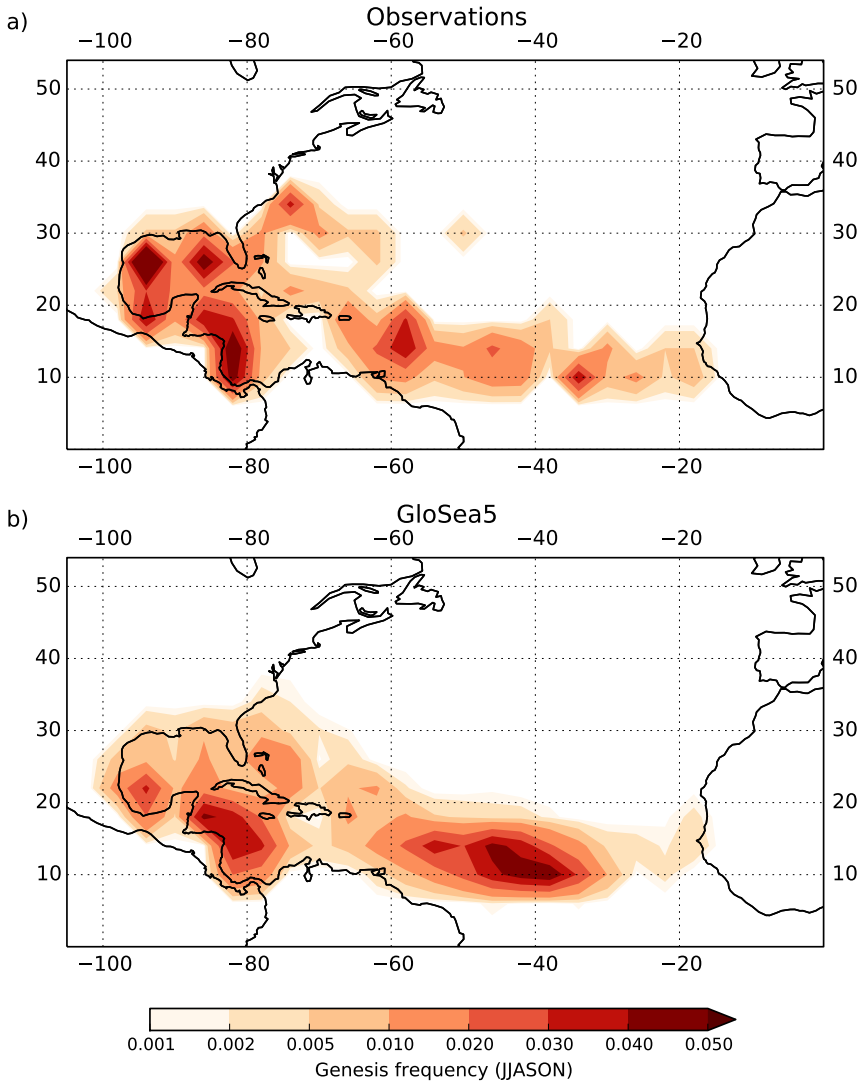

Figure 3: Average seasonal genesis density (per 4°x4° box) of all U.S. and Caribbean landfalling TCs, measured as a proportion of the total number of landfalling U.S. and Caribbean TCs, in a) observations and b) GloSea5 (all ensemble members) over the period June–November (JJASON) 1992–2013.

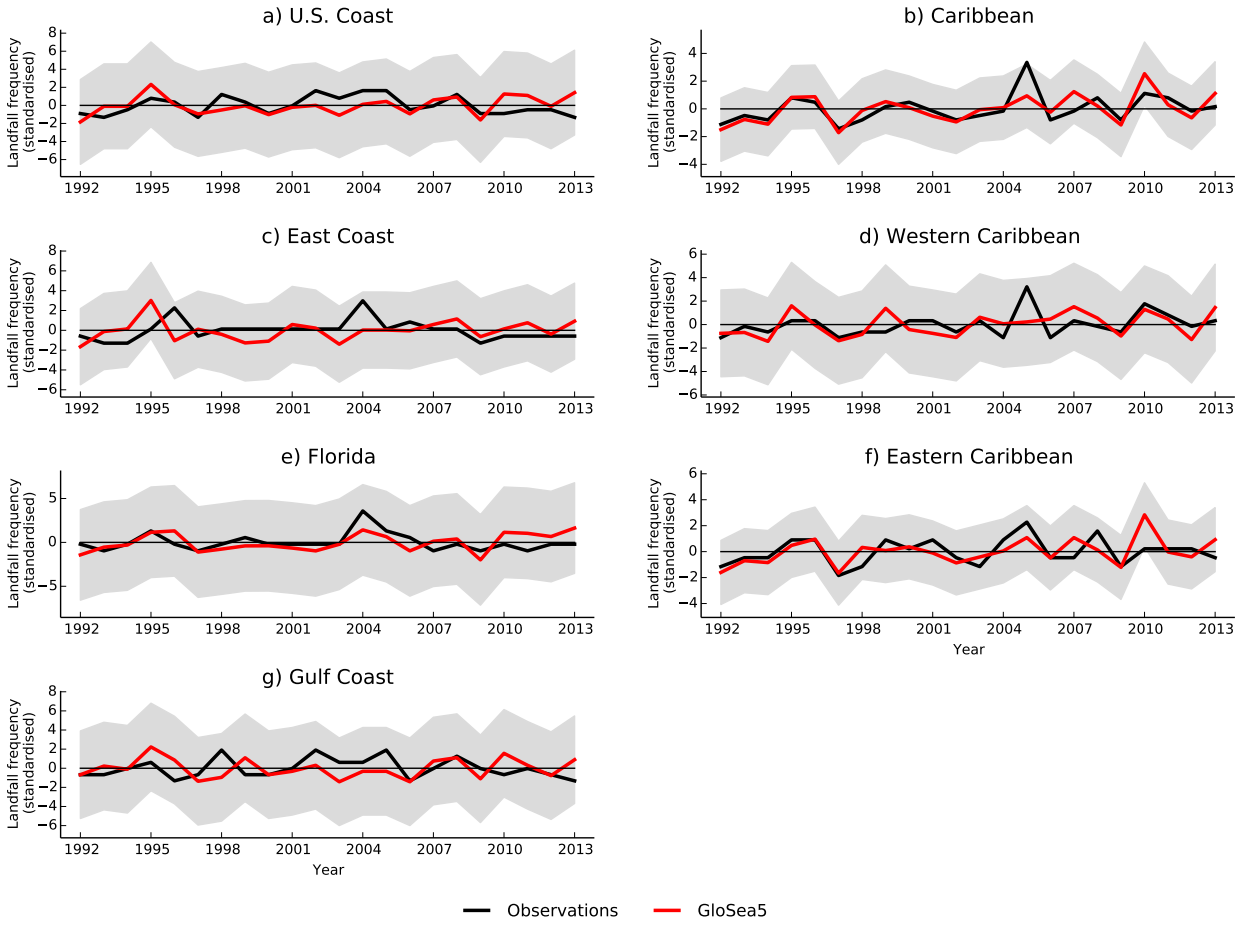

Figure 4: Standardised TC landfall frequency in each of the regions defined in Figure 1 over the period June–November 1992–2013. Observations are in black; GloSea5 ensemble mean is in red. The grey shading illustrates the model ensemble spread measured as ± 1 standard deviation about the ensemble mean for each year. Results are standardised for each year by subtracting the mean of the whole ensemble for all years and then dividing by the standard deviation of the ensemble means.

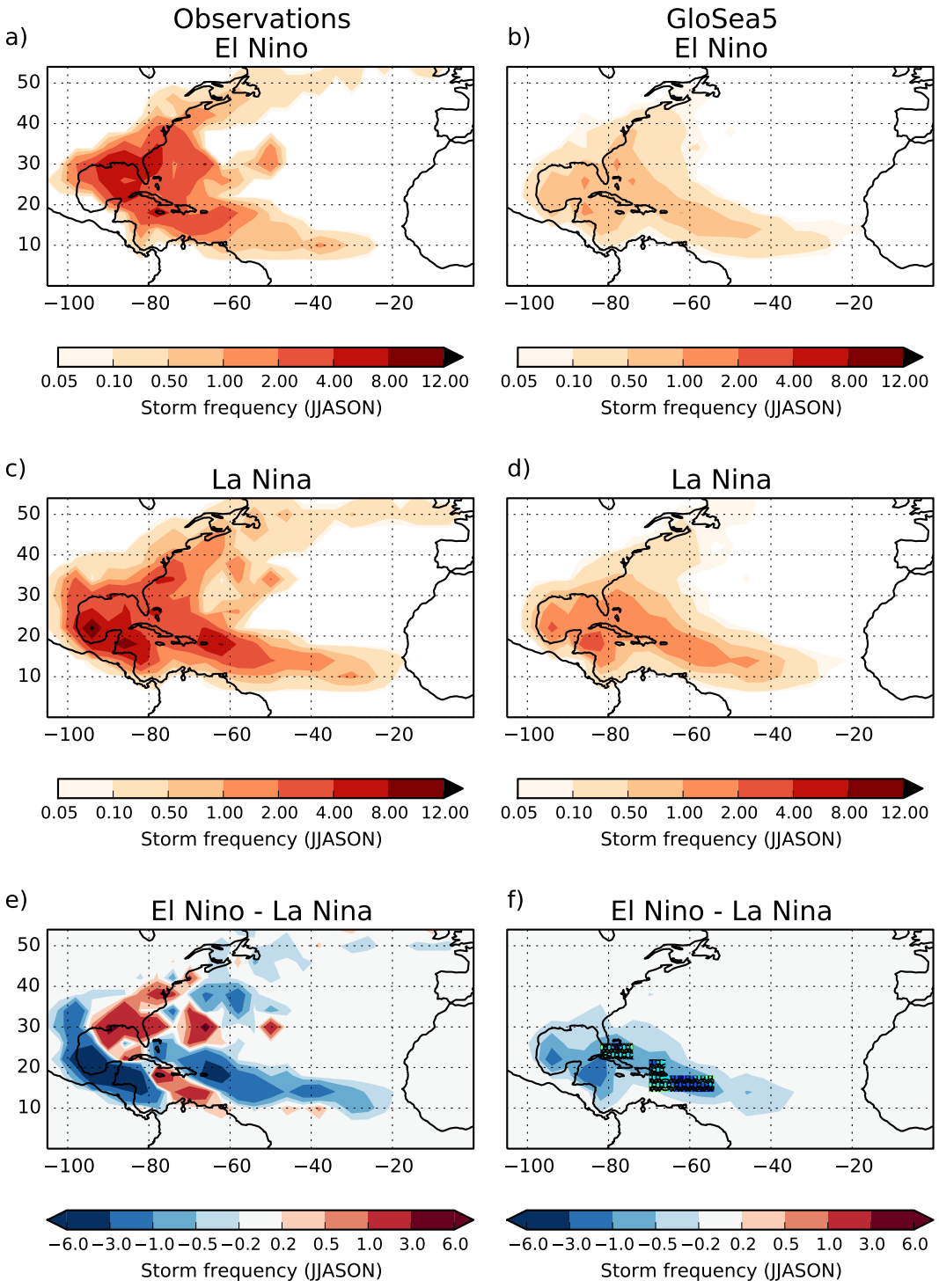

Figure 5: Observed (left) and GloSea5 ensemble-mean (right) storm track density (number of storm tracks per 4°x4° box) for all landfalling TCs along the U.S. and Caribbean coast during (a, b) El Niño and (c, d) La Niña events during June–November (JJASON). (e, f) show the track density difference (El Niño minus La Niña events): red (blue) regions show where the track density is enhanced (reduced) during El Niño relative to La Niña events. Black boxes show where changes have a p-value < 0.1 using a two-tailed Student's t-test. El Niño years: 1997, 2002, 2004, 2006 and 2009; La Niña years: 1995, 1998, 1999, 2000, 2007, 2010 and 2011.

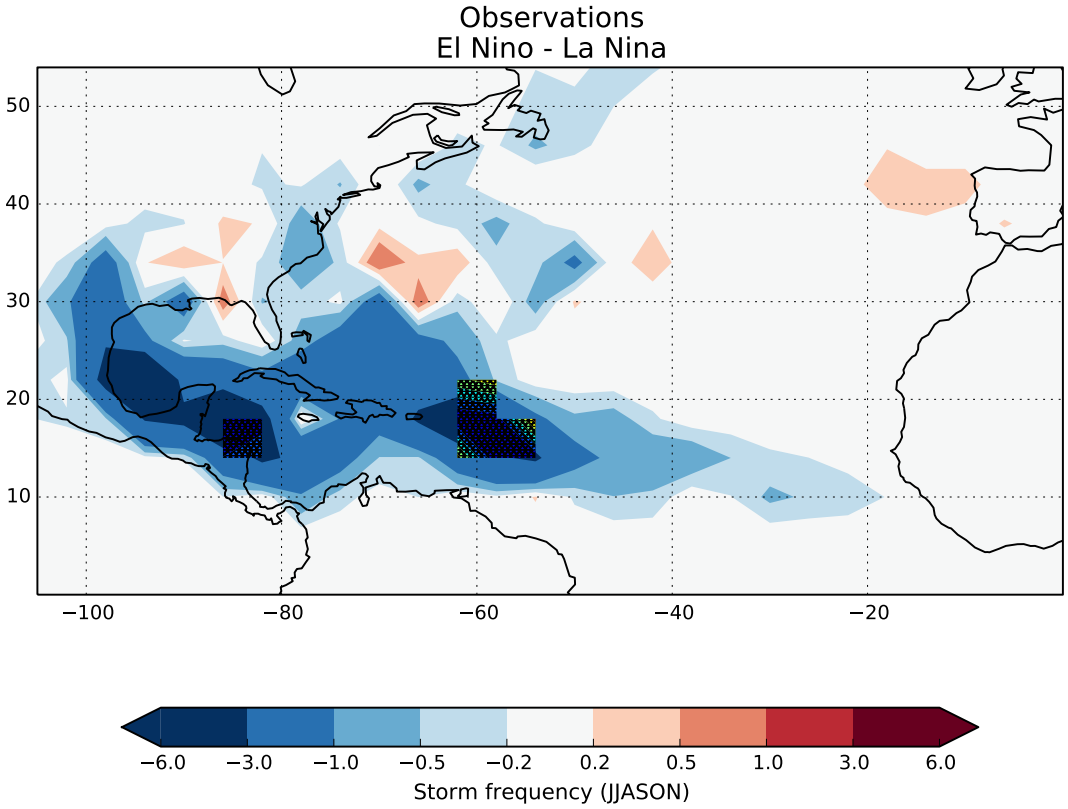

Figure 6: As Figure 5e, but for the U.S. and Caribbean landfall track density difference between EN and LN events averaged over the longer period 1950–2014. There are 18 EN and 16 LN events in total. Black boxes show where changes have a p-value < 0.1 using a two-tailed Student's t-test. El Niño years: 1951, 1953, 1957, 1963, 1965, 1969, 1972, 1976, 1977, 1982, 1986, 1987, 1991, 1997, 2002, 2004, 2006 and 2009; La Niña years: 1950, 1954, 1955, 1964, 1970, 1971, 1973, 1975, 1988, 1995, 1998, 1999, 2000, 2007, 2010 and 2011.

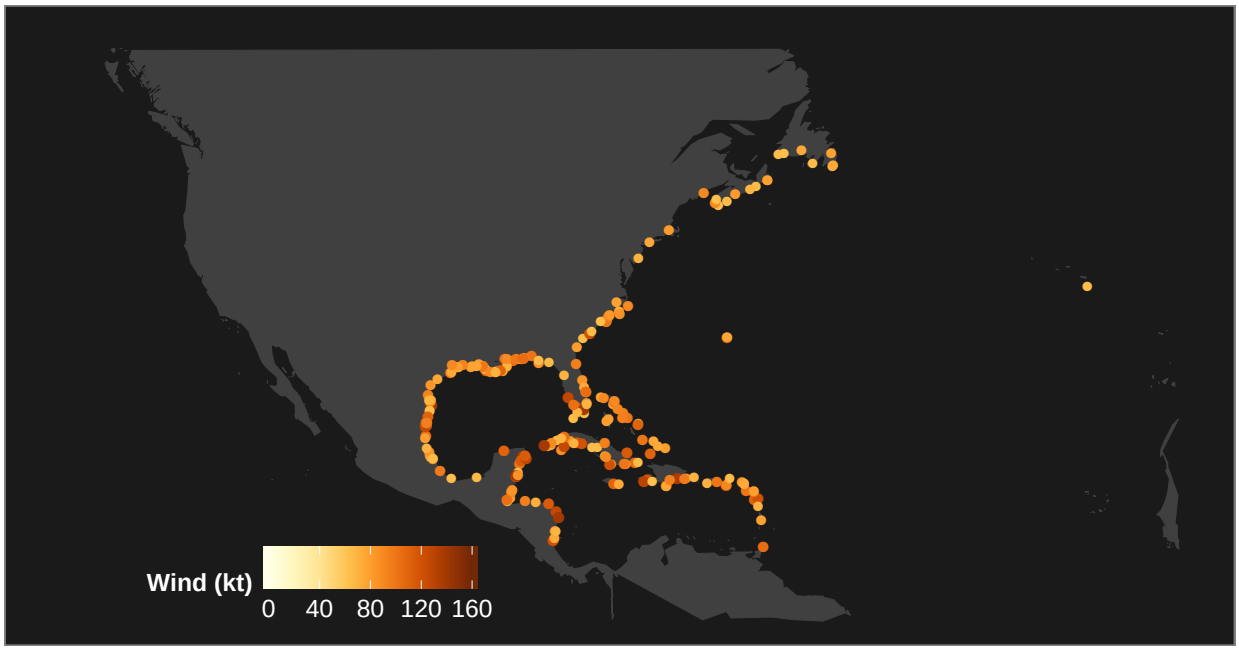

Figure 7: Landfalling hurricane locations for the 1962–2014 period.

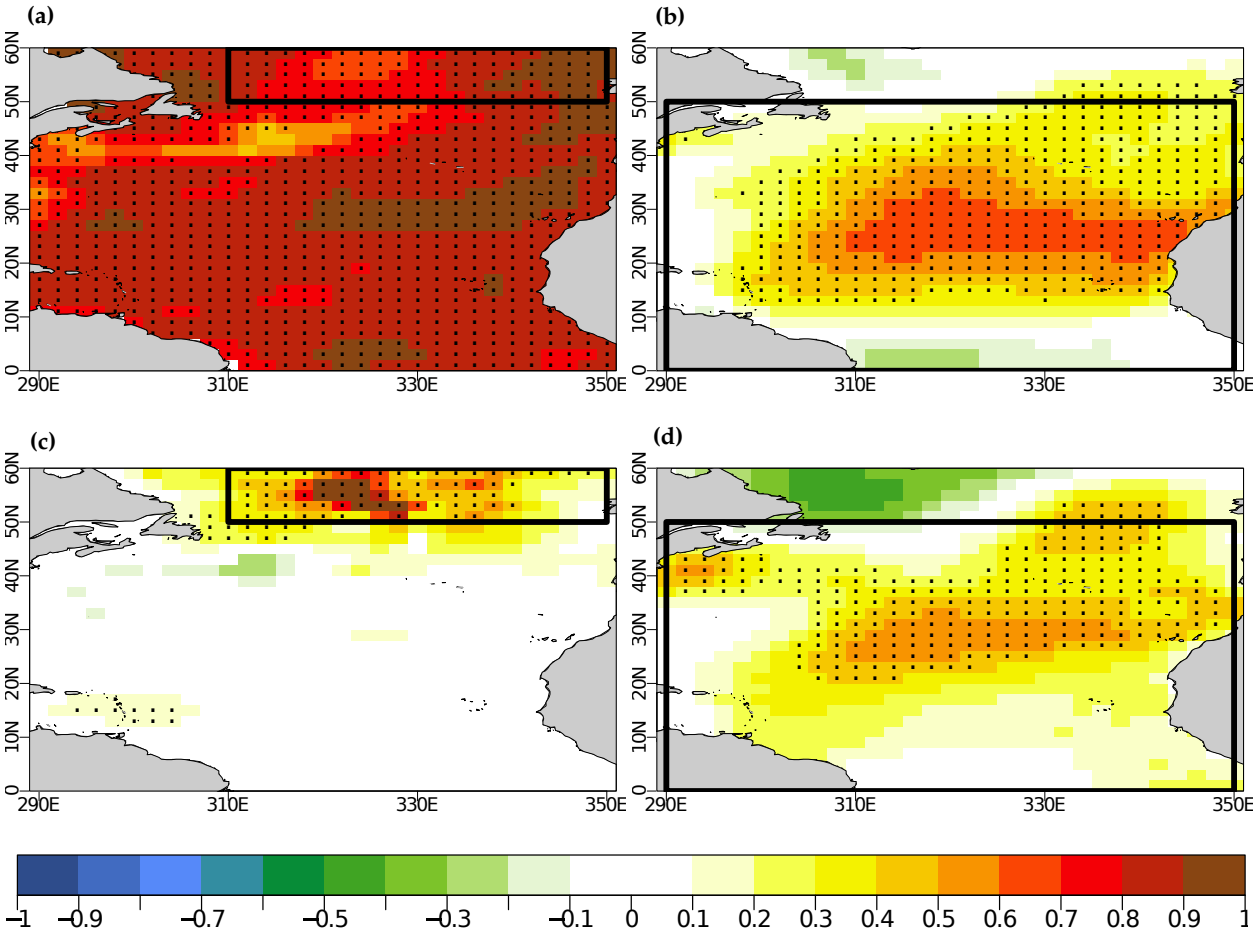

Figure 8: First row: Anomaly correlation coefficients (ACCs) for 4-year mean a) SSTA and b) SLPA in Ini MME forecasts. Second row: Difference in ACCs between the Ini MME and NoIni MME for c) SSTA and d) SLPA. The black dots indicate the regions where the results are significant at the 5% level and the black boxes indicate the area considered in the construction of the index.

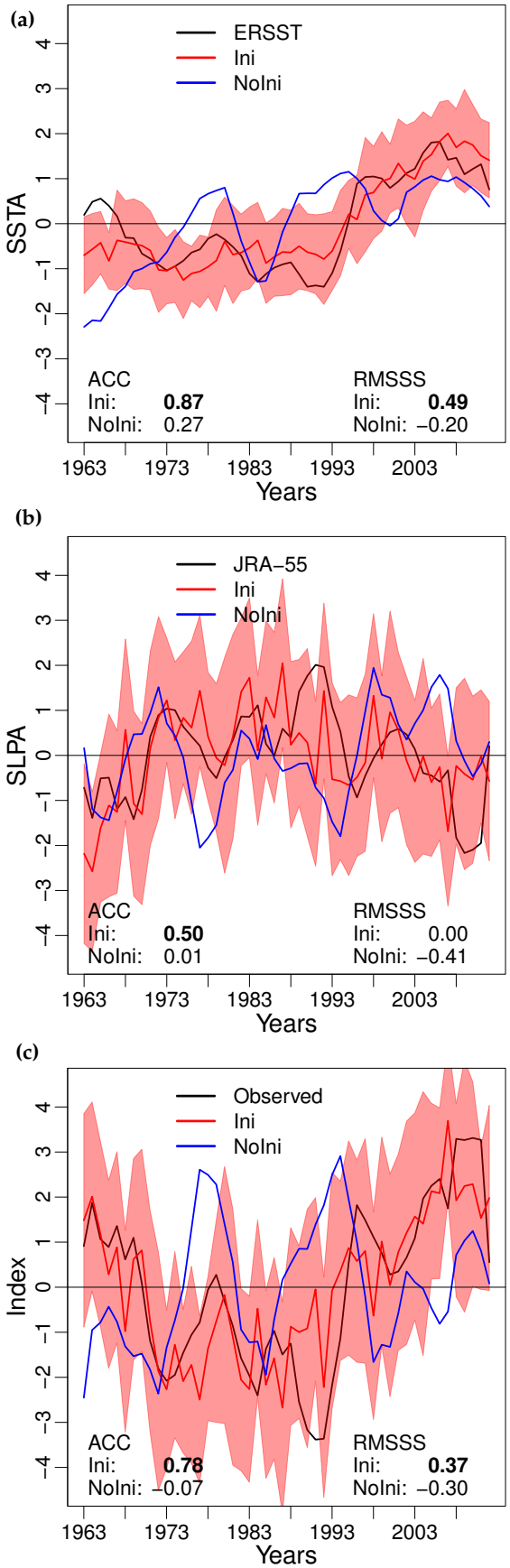

Figure 9: 4-year mean a) standardised SSTA, b) standardised SLPA and c) AMV index. Observations are in black, the Ini MME in red and the NoIni MME in blue. The red shading represents the 95% confidence interval. ACCs and RMSSSs are shown in the bottom left and right corners, respectively. Significant values are in bold.

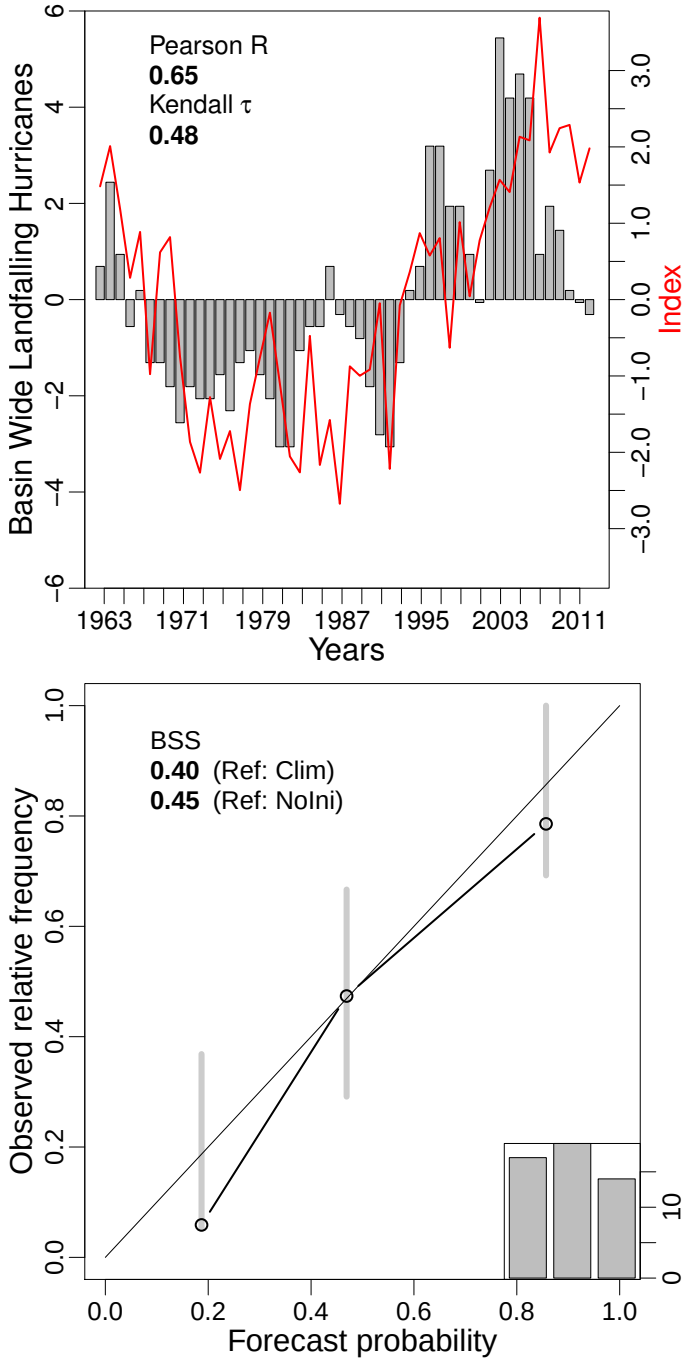

Figure 10: Top: Time series of 4-year mean hurricane landfall numbers (bars) and 4-year mean Ini MME forecasted index (forecast year 2–5; red line). Hurricane landfalls are expressed as anomalies with respect to the 1962–2014 average. Pearson's linear correlation and Kendall's ranked correlation indices are indicated in the upper-left corner. Both are statistically significant at the 5% level. Bottom: Reliability diagrams of 4-year mean Ini MME forecasts for landfalling hurricanes. Brier Skill Scores using both climatology and NoIni MME as benchmark are shown in the upper left corner. Both are statistically significant at the 5% level. The probabilistic forecasts represent the probability that the activity level will be above the climatological average.

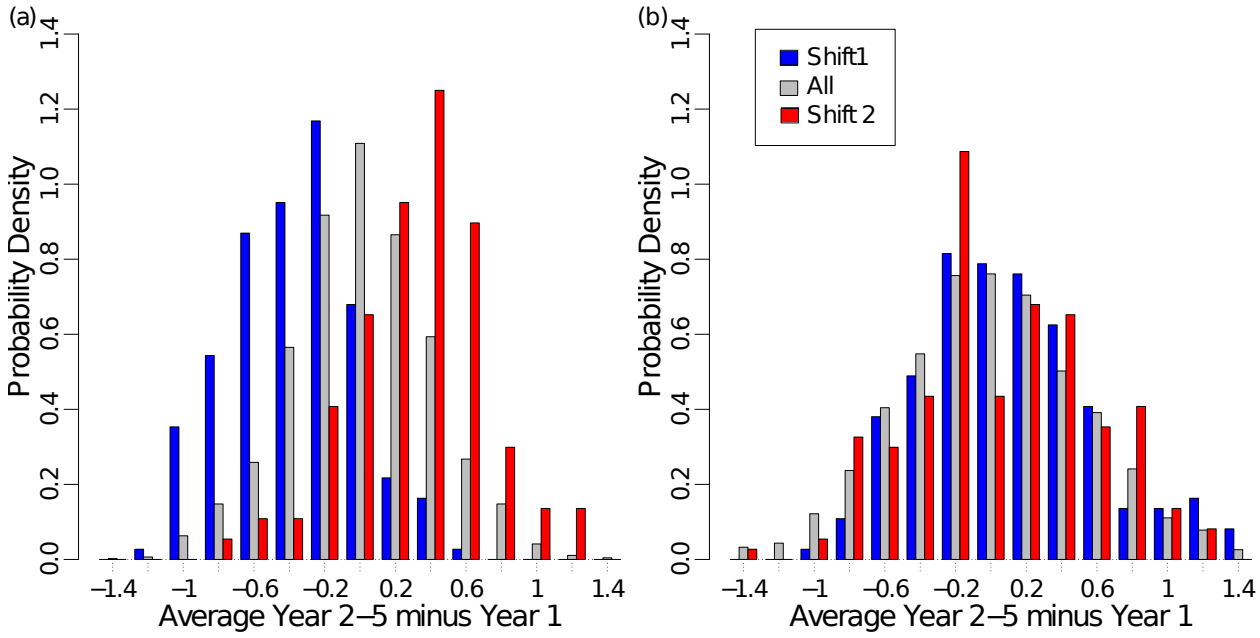

Figure 11: Empirical probability density function (PDF) for changes in Ini MME forecasted a) SSTA and b) SLPA. The three distributions are constructed using all the start dates (gray) and the four start dates preceding the 1969–1970 (blue) and 1994–1995 (red) shifts. The PDF is based on the differences between the mean values of forecast year 2–5 and forecast year 1 values.

Table 1: Pearson linear correlations (top) and p-values (bottom) between observed and the GloSea5 ensemble-mean TC landfall frequency along the U.S. Coast, East Coast, Florida, Gulf Coast, Caribbean (Carib), western Caribbean (W Carib) and eastern Caribbean (E Carib) over the period June–November 1992–2013. Bold implies statistical significance at the 95% confidence level.

	U.S. Coast	East Coast	Florida	Gulf Coast	Carib	W Carib	E Carib
r	0.22	-0.06	**0.41**	-0.01	**0.69**	**0.39**	**0.52**
p-value	0.16	0.40	0.03	0.49	0.00	0.04	0.01

Table 2: Models (and their respective number of members) used to construct the ensembles used in this study.

Model I.D.	No. Ini members	No. NoIni members
GFDL-CM2.1	10	10
HadCM3	20	10
MIROC5	6	3
MPI-ESM	10	3

Table 3: Frequently used acronyms.

ACC	anomaly correlation coefficient
AMO	Atlantic multi-decadal oscillation
AMV	Atlantic multi-decadal variability
BSS	Brier skill score
CGCM	coupled global climate model
EN	El Niño
ENSO	El Niño Southern Oscillation
GCM	global climate model
LN	La Niña
MDR	main development region
MME	multi-model ensemble
RMSSS	root mean square skill score
SLP(A)	sea level pressure (anomaly)
SPG	sub polar gyre
SST(A)	sea surface temperature (anomaly)
TC	tropical cyclone