

# *Seasonal prediction of tropical cyclones over the North Atlantic and Western North Pacific: dependence on model, resolution and stochastic physics*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Befort, D. J., Hodges, K. I. ORCID: <https://orcid.org/0000-0003-0894-229X> and Weisheimer, A. (2022) Seasonal prediction of tropical cyclones over the North Atlantic and Western North Pacific: dependence on model, resolution and stochastic physics. *Journal of Climate*, 35 (5). pp. 1385-1397. ISSN 1520-0442 doi: 10.1175/JCLI-D-21-0041.1 Available at <https://centaur.reading.ac.uk/96742/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1175/JCLI-D-21-0041.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Seasonal Prediction of Tropical Cyclones over the North Atlantic and Western North Pacific

DANIEL J. BEFORT,<sup>a</sup> KEVIN I. HODGES,<sup>b,c</sup> AND ANTJE WEISHEIMER<sup>d,e</sup>

<sup>a</sup> *Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom*

<sup>b</sup> *National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom*

<sup>c</sup> *Department of Meteorology, University of Reading, Reading, United Kingdom*

<sup>d</sup> *National Centre for Atmospheric Science, University of Oxford, Oxford, United Kingdom*

<sup>e</sup> *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*

(Manuscript received 25 January 2021, in final form 22 October 2021)

**ABSTRACT:** In this study, tropical cyclones (TCs) over the western North Pacific (WNP) and North Atlantic (NA) basins are analyzed in seasonal forecasting models from five European modeling centers. Most models are able to capture the observed seasonal cycle of TC frequencies over both basins; however, large differences for numbers and spatial track densities are found. In agreement with previous studies, TC numbers are often underestimated, which is likely related to coarse model resolutions. Besides shortcomings in TC characteristics, significant positive skill (deterministic and probabilistic) in predicting TC numbers and accumulated cyclone energy is found over both basins. Whereas the predictions of TC numbers over the WNP basin are mostly unreliable, most seasonal forecast provide reliable predictions for the NA basin. Besides positive skill over the entire NA basin, all seasonal forecasting models are skillful in predicting the interannual TC variability over a region covering the Caribbean and North American coastline, suggesting that the models carry useful information, including for adaptation and mitigation purposes ahead of the upcoming TC season. However, skill in all forecast models over a smaller region centered along the Asian coastline is smaller compared to their skill in the entire WNP basin.

**KEYWORDS:** Tropical cyclones; Forecast verification/skill; Seasonal forecasting

## 1. Introduction

Every year, tropical cyclones (TCs) are associated with casualties and high economic costs, due to the high wind speeds, storm surges, and large rainfall amounts associated with them (Pant and Cha 2019). This is especially the case for highly populated regions of the Caribbean, North America, and the coastline of Asia. For example, between 1984 and 2015 China experienced economic losses of roughly 6.9 billion dollars (USD) per year on average related to tropical cyclones (Wang et al. 2016).

Given the high socioeconomic impact of these events, there is a large interest and demand in skillful predictions of TC numbers and their intensities ahead of the upcoming season. Several studies have investigated the ability of dynamical models to predict TC interannual variability on seasonal time scales (e.g., Chen and Lin 2013; LaRow et al. 2010; Zhao et al. 2010). One of the earliest attempts was made by Vitart and Stockdale (2001) using the ECMWF model. They found high

correlations between observed and forecasted seasonal TC numbers over the western North Pacific (WNP) and North Atlantic (NA) basins. However, in their study the hindcast period was limited to 1991–99. More recently TCs have been analyzed in the U.K. Met Office seasonal forecast systems. For GloSea5 data from 1996 until 2009, Camp et al. (2015) found significant positive correlation between observed and simulated interannual TC frequency and accumulated cyclone energy (ACE; Saunders and Lea 2005) in the North Atlantic, western Pacific, Australian region, and South Pacific.

The well-known advantages of using multimodel ensembles (MMEs) over single-model ensembles has motivated further studies of TCs in seasonal forecasting MMEs. Vitart (2006) assessed the representation of TCs in seven coupled ocean–atmosphere models from the DEMETER (Palmer et al. 2004) project and found for the period from 1987 to 2001 significant positive correlations for TCs over the NA, the eastern North Pacific (ENP), and WNP ocean basins for some models. Using three models for which extended data from 1959 to 2001 were available they also showed that the skill for TC numbers varies throughout the hindcast period, which the author argues may be associated with changes in ENSO variability and improving ocean initial conditions through the period. However, recent studies using coupled and atmosphere-only century long seasonal hindcasts suggest that multidecadal variability in skill could also be physically based as found for the North Atlantic Oscillation and ENSO (Weisheimer et al. 2017, 2020). Multidecadal variability of tropical cyclone frequencies has been further discussed in Fink et al. (2010) and Caron et al. (2015). Vitart et al. (2007) analyzed the skill of

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-21-0041.s1>.

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

Corresponding author: Daniel J. Befort, [daniel.befort@physics.ox.ac.uk](mailto:daniel.befort@physics.ox.ac.uk)

DOI: 10.1175/JCLI-D-21-0041.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

three seasonal forecast models within the European Seasonal to Interannual Prediction (EUROSIP) system (Stockdale 2013) in simulating TC variability over the North Atlantic during 1993–2006 using deterministic and probabilistic skill measures. Besides confirming earlier findings of high correlations between observed and predicted TC numbers over the NA basin, they also reported that the EUROSIP system was able to successfully distinguish the exceptional hurricane season of 2005 and the average TC season of 2006. More recently, an extensive study of the predictability of tropical cyclones over the North Atlantic using four seasonal forecast models from the North American Multi-Model Ensemble (NMME)-Phase II has been carried out by Manganello et al. (2019). While confirming results from Vitart (2006) and Vitart et al. (2007) on generally moderate to high skill over the North Atlantic basin, they show that the respective models are most skillful over the western tropical North Atlantic and Caribbean. Furthermore, high fluctuations in skill are found on decadal time scales, which supports results presented in Vitart (2006).

One main shortcoming of current and past seasonal forecasting systems is their relatively coarse horizontal model resolution, which is usually on the order of 50 km or larger. The effect of increased model resolution on tropical cyclone properties has been investigated in various studies (e.g., Roberts et al. 2020; Murakami et al. 2015; Manganello et al. 2012), which have shown that enhanced resolution results in improvements of TC frequencies and intensities. The impact of resolution in seasonal forecast models has been investigated by Manganello et al. (2016) using a series of hindcasts performed with ECMWF's coupled model system at different resolutions. In general, it was found that skill for ACE, and to a lesser degree also for TC numbers, increases with resolution. However, in another study using the U.K. Met Office GloSea5 seasonal prediction model, increased model resolution did not lead to improvements in forecast skill (Scaife et al. 2019).

In addition to the impact of increased model resolution, TC characteristics and their predictability are also affected by other model dependencies. For example, the study of Camp et al. (2019) used two versions of the U.K. Met Office's seasonal forecast model GloSea5 with the same resolution but different dynamical cores and differences in the physics schemes. While they found significant skill for TCs over the western North Pacific for one of the model versions (GloSea5-GA3), no such skill was found for the other model version (GloSea5-GC2). Feng et al. (2020) showed that the limited skill of GloSea5-GC2 for TCs in the northeast WNP is related to the overestimation of the negative TC–ENSO teleconnection in this model.

In the past, several different schemes have been developed and applied to objectively identify tropical cyclones in various datasets (e.g., Befort et al. 2020; Camargo 2013; Walsh et al. 2013; Ullrich and Zarzycki 2017). Studies using more than one detection scheme have found large differences in TC numbers, suggesting that the methods used to identify TCs in gridded datasets can introduce another source of uncertainty (e.g., Bell et al. 2019; Horn et al. 2014; Murakami 2014).

In Europe, several national weather centers provide operational seasonal forecasts, including forecasts of TC activity. Furthermore, a multimodel based tropical cyclone forecast system for the North Atlantic can be found at <https://seasonalhurricanepredictions.bsc.es/>. Studies on the level of skill of European models in simulating tropical cyclones either have focused on a single model (e.g., Camp et al. 2015, 2019) or are based on multimodel ensembles using older seasonal forecast systems (e.g., Vitart 2006; Vitart et al. 2007). Recently, the Copernicus Climate Change Service (C3S) multimodel seasonal forecasting system has been established (Brookshaw 2017), succeeding the EUROSIP system. C3S currently provides seasonal forecasts from five different European institutions, as well as forecasts from four other forecasting systems.

In this study, we analyze Northern Hemisphere tropical cyclones with an emphasis on the western North Pacific and North Atlantic in six European seasonal forecasting systems from five different centers: ECMWF, the U.K. Met Office, Météo-France, the German Weather Service [Deutscher Wetterdienst (DWD)], and the Euro-Mediterranean Center on Climate Change [Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)]. TCs are detected using a state-of-the-art detection scheme for the common 22-yr hindcast period from 1993 until 2014 and the skill is assessed against five different reanalyses and IBTrACS observations using deterministic and probabilistic skill measures. The short hindcast period provides further motivation to use a multimodel ensemble. Similar skill patterns in different single model ensembles increases the confidence of the findings, as seen for skill in extratropical cyclones over the North Atlantic in Befort et al. (2019).

The manuscript is structured as following: the datasets and methods are described in section 2. Results for TCs in C3S seasonal forecast models are presented in section 3 and summarized in section 4.

## 2. Data and methods

### a. Data

The TC identification is performed for six different seasonal forecast systems: the ECMWF SEAS5 (hereafter referred to as ECMWF-SEAS5; Johnson et al. 2019), U.K. Met Office GloSea5-GC2 (hereafter UKMO-GloSea5-GC2; MacLachlan et al. 2015), Météo-France System 5 (hereafter Météo-France-S5; Météo-France 2015), Météo-France System 6 (hereafter Météo-France-S6; Dorel et al. 2017), DWD GCFS2.0 (hereafter DWD-GCFS2.0; Fröhlich et al. 2021), and CMCC SPS3 (hereafter CMCC-SPS3; Gualdi et al. 2020). Apart from structural differences between the models (e.g., physical parameterizations), these seasonal forecast systems differ with regards to horizontal and vertical resolution (atmosphere and ocean) and ensemble size (Table 1). Besides these differences, all the seasonal forecasts provide 12-hourly data on pressure levels at a  $1^\circ \times 1^\circ$  horizontal resolution and are available for a common 22-yr hindcast period from 1993 until 2014. The main focus is on the North Atlantic (NA) and

TABLE 1. Datasets used in this study.

Model	Initial dates	Atmosphere resolution	Ocean resolution	Ensemble size
ECMWF SEAS5 (ECMWF-SEAS5)	May, June	TCO319/L91 $\approx$ 32 km	0.25°/L75	25
GloSea5-GC2 (UKMO-GloSea5-GC2)	May, June	N216/L85 $\approx$ 90 km	0.25°/L75	28
Météo-France S5 (Météo-France-S5)	May, June	TL255/L91 $\approx$ 80 km	1°/42 levels	15
Météo-France S6 (Météo-France-S6)	May, June	TL359/L91 $\approx$ 50 km	1°/75 levels	25
DWD (DWD-GCFS2.0)	May, June	T127/L91 $\approx$ 100 km	0.4°/L40	30
CMCC (CMCC-SPS3)	May, June	$\approx$ 110 km/L46	0.25°/L50	40

western North Pacific (WNP) basins as well as two subbasins centered on the North American and Asian coastlines (Fig. 1). Due to the differences in the seasonal cycle of TCs over both basins, the active TC season for the WNP is taken as June–October (JJASO) and for the NA as July–October (JASO). To ensure we use the same lead time for both basins, forecasts initialized in May for the WNP basin and in June for the NA basin are analyzed.

Results from the seasonal forecast models are verified against IBTrACS v4.0 observations (Knapp et al. 2010) and a set of five different reanalyses: ERA-Interim (Dee et al. 2011), ERA5 (Hersbach et al. 2020), NCEP-CFSR (Saha et al. 2010), MERRA-2 (Gelaro et al. 2017), and JRA-55 (Kobayashi et al. 2015) (see Table ST1 in the online supplemental material for details). In this study, results for the seasonal forecasts will be primarily compared to the mean of these five reanalyses. The advantage of reanalyses over observations for validation is that the TCs are identified in the same way as for the forecasts while for observations operational procedures are different. It has been previously shown that reanalyses have limitations to represent TCs (Hodges et al. 2017). However, even though the limitations found in reanalyses might to a large extent be due to their resolution, there is also uncertainty in the observations of especially weaker tropical storms. The limitations in both current reanalysis and IBTrACS mean that the record of tropical cyclones during the past has uncertainties in both TC intensity and frequency. The use of multiple reanalyses is thought to obtain more robust TC number statistics compared to using a single reanalysis.

### b. Tropical cyclone identification

In this study, we used an objective tracking algorithm based on vorticity fields to identify tropical cyclones over the Northern Hemisphere (Hodges et al. 2017). This is based on the vertical average of vorticity at 850 and 700 hPa spectrally filtered to remove the large-scale background ( $n \leq 5$ ) and truncated to T63 spectral resolution with spectral tapering applied to remove the small-scale noise and allow for more reliable tracking (in previous studies the 600-hPa level is used as well; however, data on this pressure level were not available for the seasonal forecast systems). Initially all vorticity maxima are tracked that exceed a threshold in the filtered vorticity field of  $5 \times 10^{-6} \text{ s}^{-1}$ . The tracking is performed by first initializing a set of tracks based on a nearest neighbor method, which are then refined by minimizing a cost function for track smoothness subject to adaptive constraints (Hodges 1995, 1999), with the constraints chosen to be suitable for the 12-hourly time steps that are available in the archived

seasonal forecast data. Tracking is performed in the Northern Hemisphere in the latitude band  $0^\circ$ – $40^\circ\text{N}$ . Following the tracking the tracks are filtered to retain those with lifetimes  $\geq 2$  days (four time steps) and additional fields are added to the tracks, including the T63 vorticity (no background removal, no tapering) maxima at all levels between 850 and 200 hPa (850, 700, 500, 400, 300, 200 hPa) using B-spline interpolation and steepest ascent maximization within a  $5^\circ$  (geodesic) radius of the tracked center. Also added to the tracks are the mean sea level pressure (MSLP) minima within the  $5^\circ$  radius using the interpolation and steepest descent minimization and the maximum 10-m wind speeds within a  $6^\circ$  radius using a direct search of the grid point values.

The identification of TCs in the reanalyses differs slightly from that used for the seasonal models as 6-hourly input data have been used. However, to make all datasets comparable to each other, the reanalysis tracks are subsampled by only using the 0000 and 1200 UTC time steps.

The tropical cyclones are identified from among all 2-day and longer tracks using the same criteria used in similar studies (Feng et al. 2020; Hodges et al. 2017; Roberts et al. 2020), namely:

- 1) the T63 relative vorticity at 850 hPa must attain a threshold of at least  $6 \times 10^{-5} \text{ s}^{-1}$ ;
- 2) the difference in vorticity between 850 and 200 hPa (at T63 resolution) must be greater than  $6 \times 10^{-5} \text{ s}^{-1}$  to provide evidence of a warm core;
- 3) the T63 vorticity center must exist at each level between 850 and 200 hPa for a coherent vertical structure;
- 4) criteria 1–3 must be jointly attained for a minimum of two consecutive time steps (one day) and only apply over the oceans; and
- 5) tracks must start within  $0^\circ$ – $30^\circ\text{N}$ .

Spatial statistics for track and genesis densities and mean intensities are computed from the tracks using spherical kernel estimators (Hodges 1996).

### c. Verification metrics

The individual model skill in simulating the interannual variability of TC numbers over the different ocean basins is measured using two deterministic scores: linear correlation coefficients and root-mean-square error (RMSE). For these skill measures, confidence intervals are estimated by randomly sampling over years with replacement. If not otherwise stated 10th and 90th percentiles of the resulting bootstrapped distribution are shown. Significant positive correlations are those for which the 10th percentile of the bootstrap

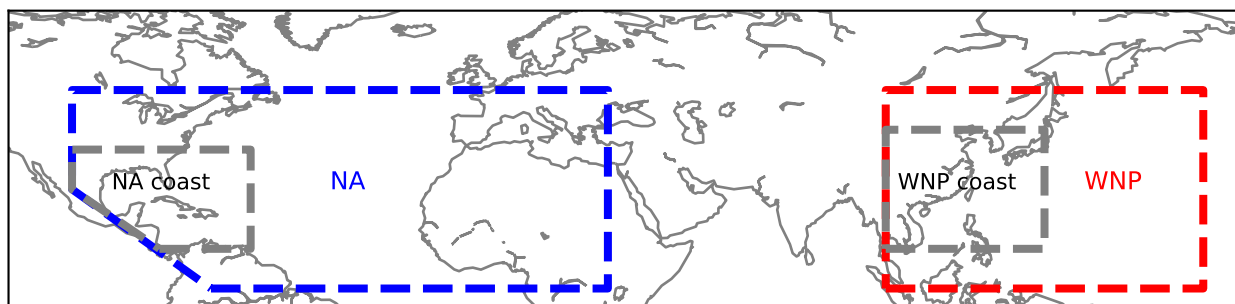


FIG. 1. The five basins used in this study, based on the IBTrACS definition; WNP is western North Pacific and NA is North Atlantic. Additionally, two subregions centered on the coastlines (but including ocean areas) are used: WNPcoast and NAcoast.

distribution is larger than 0; for RMSE significant values are those for which the 90th percentile of the bootstrap distribution for the respective model is smaller than the RMSE of a climatological forecast.

Besides these deterministic metrics, the value of the prediction systems is assessed using the relative operating characteristic area under curve (ROC-AUC) score, which measures the model's ability to discriminate between an event and a nonevent. In this study events are defined as seasons with TC occurrences above the upper tercile (active season) and seasons below the lower tercile (inactive seasons), respectively. The confidence of the ROC-AUC values is assessed by randomly sampling over years (analog to correlation coefficients and RMSE). The short hindcast period of only 22 years introduces some problems as tercile boundaries might not be uniquely defined due to the discrete nature of TC numbers. This means that the number of active (inactive) seasons differs if they are defined as seasons with counts larger (smaller) than the percentile or with counts larger than or equal to (smaller than or equal to) the percentile. Here, we use whatever method (larger/smaller vs larger or equal/smaller or equal) provides a value closest to a tercile frequency (33.3%). For each bootstrap sample of observations, active seasons are defined as the 7 years with highest TC counts and inactive seasons as those with the lowest seven TC counts.

The models' reliability is assessed using the spread-over-error (SoE) metric, defined by the ratio of the square root of the average variance of the ensemble over all years and the RMSE of the ensemble mean (Fortin et al. 2014). A value of 1 indicates a perfectly reliable ensemble, and values below 1 an overconfident and values above 1 an underconfident ensemble.

Several studies investigating tropical cyclones in models and observations also analyze the accumulated cyclone energy (ACE) in addition to tropical cyclone numbers (Saunders and Lea 2005). Here, ACE is computed for the different regions using the formula

$$ACE = 2 \times 10^{-4} \sum_i \sum_n V_{\max}^2,$$

where  $V_{\max}$  is the maximum 10-m wind speed (in knots;  $1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$ ) and the summations,  $i$  and  $n$ , are over all tracks and track points within a region. ACE is scaled so as to replicate 6-hourly sampling.

Over the different basins and subbasins, TC numbers are determined for those storms with at least one time step in the domain. For ACE only wind values for time steps for which the storm is located within the domain are taken into account.

### 3. Results on tropical cyclones in seasonal forecasts

#### a. Climatological features

The observed (IBTrACS) seasonal cycle of tropical cyclone genesis over the WNP basin is shown in Fig. 2a. Over this basin, most tropical cyclones are observed during June to October. The observed seasonal cycle is well captured by the reanalyses and thus by the multimodel reanalyses (MMRs), with a slight overestimation of the TC numbers over this region. In contrast, differences are large for the six seasonal forecast models initialized in May. Too many TCs are detected in UKMO-GloSea5-GC2 and in Météo-France-S5, whereas the numbers of TCs are underestimated in the CMCC-SPS3 and DWD-GCFS2.0 forecasts. The models' ability to capture the seasonal cycle, measured by the Spearman rank correlation coefficient to the MMR mean from May to October, indicates that despite their large biases, CMCC-SPS3 and UKMO-GloSea5-GC2 perform best alongside ECMWF-SEAS5. Both Météo-France models show the lowest correlations, which is due to the large number of TC events wrongly detected in October. Note however that these correlations might be sensitive to small changes as only six data points are considered in calculating the correlation statistics.

Over the NA basin most tropical cyclones are observed during July–October (Fig. 2b), which is again very well captured by the ensemble of reanalyses (MMR). All seasonal forecast models initialized in June are able to capture the observed seasonal cycle with (rank) correlations above 0.8. Besides this good agreement in the seasonal cycle, UKMO-GloSea5-GC2, ECMWF-SEAS5, and both Météo-France models are also able to reasonably well simulate the number of TCs over this basin. Similar to results over the WNP, CMCC-SPS3 and DWD-GCFS2.0 substantially underestimate the number of TCs over the NA. It is worth mentioning that both these models have relatively coarse resolutions (see Table 1) compared to the other systems, which might contribute to the underestimation of TCs as shown by previous



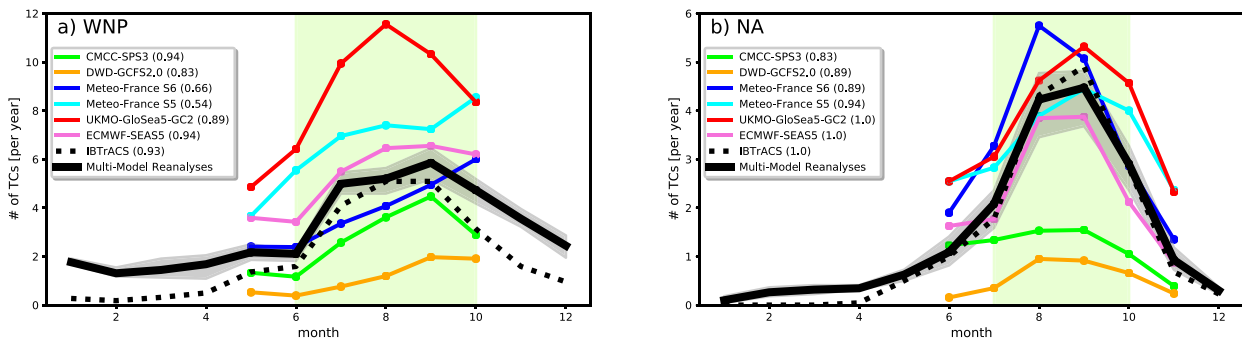


FIG. 2. Tropical cyclone seasonal cycle (genesis months) over (a) the WNP basin and (b) the NA basin. Colored lines indicate the different seasonal forecast models, whereas the dotted line shows IBTrACS observations, and the solid black line shows the five-model multi-model reanalysis (MMR) mean. The spread of the reanalysis is shown as gray shading. The verification months are also indicated by the green shading. Temporal correlations between models and MMR over the verification months are displayed in parentheses in the legend.

studies (e.g., Roberts et al. 2020; Murakami et al. 2015). These results for the NA basin are similar to those found for climate simulations carried out with similar model versions presented in Roberts et al. (2020) (their Fig. 9). The UKMO-GloSea5-GC2 model simulates too many TCs in October compared to observations, whereas the DWD/MPI-ESM1-2 model simulates too few events during the entire season. However, a direct comparison is impossible as the model setup is not identical between the simulations used in this study and those used in Roberts et al. (2020).

The spatial TC track density pattern for the MMR over the WNP during June–October shows that most TCs are found over the South China Sea and west of the Philippines (Fig. 3a). Interestingly all three models that overestimate the number of TCs over the WNP (UKMO-GloSea5-GC2, Météo-France-S5, and ECMWF-SEAS5) show a similar track density difference pattern to the MMR (Figs. 3c,d,f). These models tend to simulate too many TCs over the subtropical western Pacific Ocean (which is most pronounced in the UKMO-GloSea5-GC2 model), whereas north and south of this area TCs are generally underestimated by the models compared to reanalysis. In contrast, the DWD-GCFS2.0, CMCC-SPS3, and Météo-France-S6 models underestimate TCs over almost the entire WNP basin, with the largest biases found in the DWD-GCFS2.0 forecast system. Over the North Atlantic basin, most models exhibit smaller biases and their spatial distribution is heavily model dependent. Comparing results for the MMR mean and IBTrACS observations over the NA basin shows a similar pattern though the maximum found west of the African coast in the MMR mean is much reduced in the observations. This is probably related to the fact that the precursor parts of the TC tracks are identified by the objective algorithm in the reanalyses and forecasts used in this study, whereas only that part of the TC life cycle that exceeds tropical storm intensity is usually included in observations.

The intensity distributions for both basins based on the TC's minimum MSLP and maximum 10-m wind speeds (WIND10M) are shown in Fig. 4. As discussed in Hodges et al. (2017), current reanalyses datasets do not fully capture the observed characteristics of tropical cyclones, which is partly related to their resolution so that strong TC intensities are

underestimated compared to observations. This is also the case for all the seasonal forecast models, with all of them underestimating maximum intensity even compared to reanalyses. This is in line with all models having resolutions, which are often coarser than the reanalysis (and could also be affected as the C3S forecast model data are only available on a  $1^\circ \times 1^\circ$  horizontal grid). For MSLP over the WNP (Fig. 4c) all models show similar intensity distributions with more frequent weak TCs (with a minimum MSLP of about 1000 hPa), a better fit to observations for TCs with minima in the range of 970–990 hPa, and an underestimation of TCs with higher intensities. Only the DWD-GCFS2.0 model is different, which shows a substantial overestimation of TCs with lower intensity up to about 990 hPa and a substantial underestimation of stronger TCs (in line with DWD-GCFS2.0 having the coarsest resolution). Over the NA the underestimation is similar in magnitude for DWD-GCFS2.0 and CMCC-SPS3, whereas the other four seasonal forecast models simulate a higher frequency of TCs with stronger intensities (Fig. 4d). For WIND10M the model distributions differ more drastically between each other. Whereas CMCC-SPS3 and DWD-GCFS2.0 show the largest underestimation over both basins (in accordance to results found for MSLP), Météo-France-S6 performs best for TC maximum 10-m wind speeds, especially over the North Atlantic. Over the WNP, ECMWF-SEAS5 also simulates a relatively high number of intense TCs compared to the rest of the seasonal forecast systems.

Overall, results on climatological characteristics of TCs in the analyzed seasonal multimodel ensemble indicate large differences between the single model ensembles. Biases in TC numbers and intensities are likely to be related to the model resolutions but might also be linked to other model components (see section 4).

#### b. Interannual predictability

After assessing the degree to which current seasonal forecast models are able to simulate observed climatological characteristics of tropical cyclones, their ability to predict the observed interannual variability of TCs is analyzed. Due to

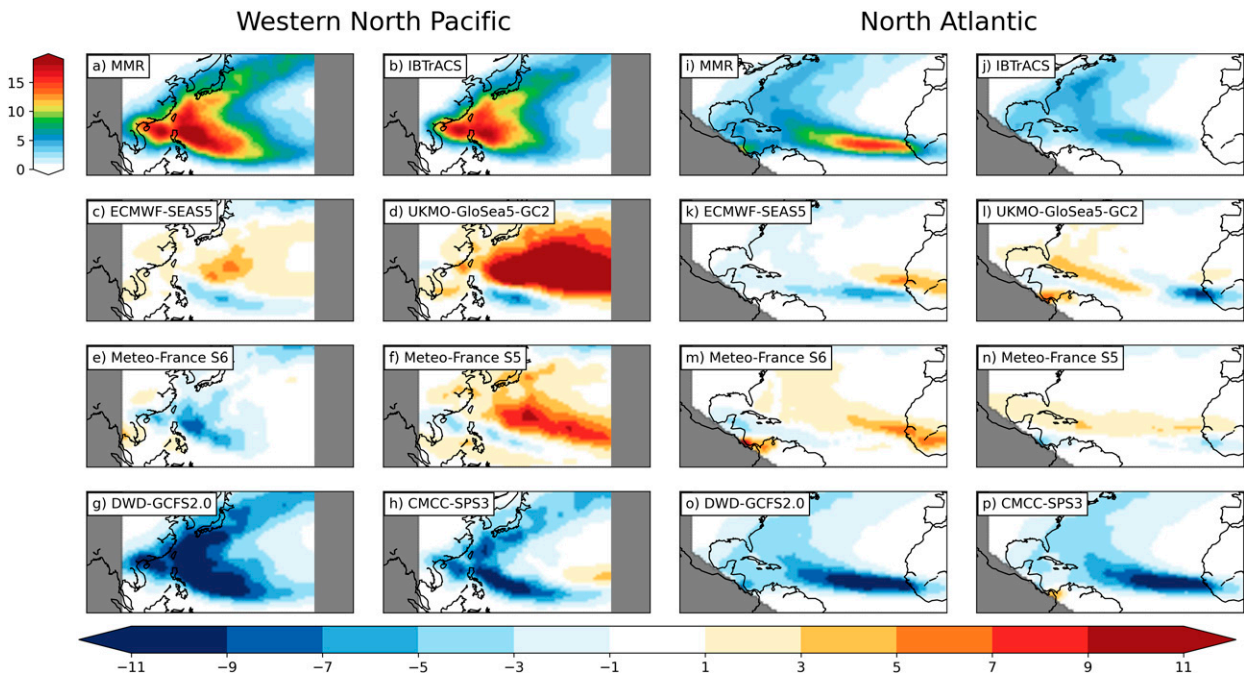


FIG. 3. Absolute TC track density for (a),(i) a mean of five reanalyses (MMR) and (b),(j) IBTrACS. Also shown are (c)–(h) differences between each C3S seasonal model and the MMR over the WNP basin for JJASO (May start dates) and (k)–(p) differences between each C3S seasonal model and the MMR over the NA basin for JASO (June start dates). Track densities are given in units of number per season per ensemble member per unit area (unit area equivalent to a  $5^\circ$  spherical cap). The MMR represents the average over the five individual reanalysis datasets.

the limited sample size of only 22 years, several skill scores (deterministic and probabilistic) are considered (see section 2b). Figure 5a shows the linear correlation coefficients for tropical cyclone numbers between each seasonal forecast system and the MMR mean. Correlation coefficients for TC numbers over the WNP region for the JJASO season are moderate with values around 0.6 for all forecast models. However, skill decreases drastically for the region centered on the western North Pacific coast (WNPcoast). Correlations for the best models (DWD-GCFS2.0, CMCC-SPS3, Météo-France-S6) are around 0.4, whereas for the remaining models (ECMWF-SEAS5, UKMO-GloSea5-GC2, Météo-France-S5) skill is even lower. For UKMO-GloSea5-GC2 these results are in line with those from Feng et al. (2020), who found major model deficiencies in simulating the observed TC–ENSO teleconnections over the WNP and also over a region similar to the WNPcoast region used in this study. For the North Atlantic basin during the JASO season, skill measured by the linear correlation coefficient is around 0.5 for most models, with the DWD-GCFS2.0 model showing the lowest score (0.4). In contrast to the WNP basin, skill for all models is also significantly positive for the NAcoast subregion, which includes the Caribbean and parts of the North American coastline. This is especially the case for ECMWF-SEAS5 with a correlation of about 0.7 for the NAcoast, which is the highest single-model score for the NA. This suggests that seasonal forecasts could provide valuable information in predicting TC numbers for the upcoming season. However, the interannual TC variability of the ensemble mean shows large differences between the

models, with all of them underestimating the observed variability, which is especially pronounced in the DWD-GCFS2.0 model (Figs. S1–S4 in the online supplemental material).

Further to the skill of the single seasonal forecast systems, the skill of the combined multimodel ensemble (MME) is derived. Even if not always providing the highest skill (as, e.g., over the NAcoast region), it is found that the MME provides significantly positive correlation coefficients for all regions analyzed here. However, it should be kept in mind that correlation coefficients are sensitive to ensemble size (Kharin et al. 2001), which differs between the systems (see Table 1) and is much larger for the MME than for the individual systems (163 members). Hence, the skill of the MME is not directly comparable to skill derived for the individual seasonal forecast systems.

Several studies use the TC ACE per season instead of or in addition to TC numbers (e.g., Camp et al. 2015). It is found that using ACE a similar skill pattern to that using TC numbers is found with the largest correlation coefficients found for the WNP, NA, and NAcoast regions and lower skill over the WNPcoast (Fig. 5b; see Figs. S5–S8 in the online supplemental material for individual time series for different basins). Over the WNP the correlation coefficients for ACE tend to be higher than for TC counts, which could potentially be linked to a stronger ENSO–ACE relationship in the WNP [see Camp et al. (2015) and references therein for more information]. The difference in skill between using numbers and ACE is probably partly related to the fact that ACE depends



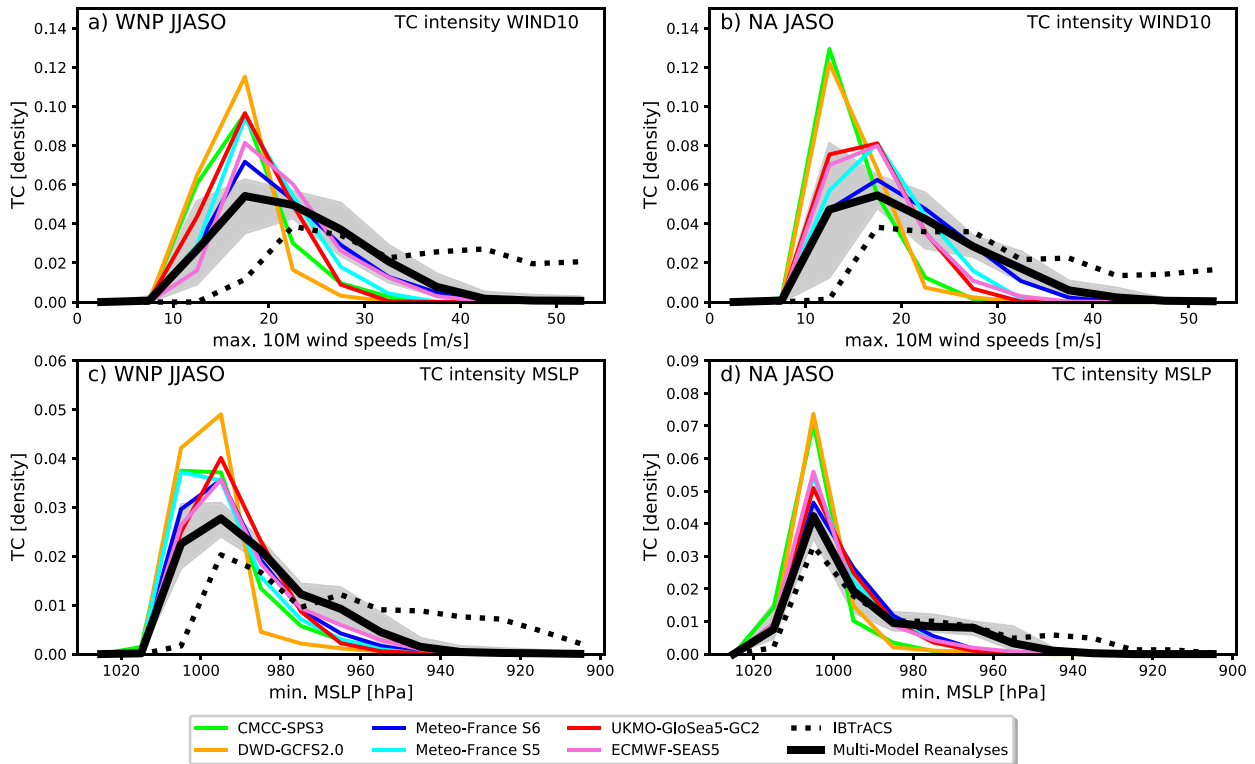


FIG. 4. Intensity distributions for the MMR (solid black; range given as gray shading), IBTrACS observations (black dotted), and six different seasonal forecast models (colored lines). Results are shown for (a),(c) the WNP during JJASO using May start dates and (b),(d) the NA during JASO using June start dates. Intensity is measured using (top) 10-m wind speeds (WIND10M) and (bottom) mean sea level pressure (MSLP).

on both the length of the tracks and the intensity of TCs, for which major differences between the models are found (Fig. 4). However, it must be kept in mind that the uncertainty of correlation coefficients for ACE and TC numbers is large (see uncertainty range in Figs. 5a,b) and that compared to these uncertainties the differences between correlations found for ACE and TC numbers for individual systems are small. However, given the short hindcast period from 1993 until 2014 (22 years) it is reassuring that the different skill metrics derived for TC numbers and ACE provide a similar picture, which increases the trustworthiness and robustness of these results.

The RMSE of the ensemble mean for TC numbers of each forecast system compared to the MMR is shown in Fig. 6. For the WNP it is found that all forecast systems have significantly lower RMSE values compared to a climatological forecast, with the lowest errors found for the MME. In accordance with low correlation coefficients, RMSEs are larger for the WNPcoast region, with no model showing significantly lower RMSE values than climatology. For the NA basin and NAcoast region, about half of the models show significantly lower errors compared to climatology, especially ECMWF-SEAS5 over the NAcoast and Météo-France-S5 over the whole NA basin. It is worth mentioning that the results

depend on the reference used, as we generally find larger RMSE values for all basins if we compare against IBTrACS instead of the MMR.

The ability of a seasonal forecast system to discriminate between active and inactive seasons can be measured by the ROC-AUC score. Here, an active season is defined as an upper tercile and an inactive season is defined as a lower tercile event. Consistent with results for the correlation score and RMSE, it is found that the ROC-AUC scores are highest for the WNP, NA and NAcoast regions, whereas the ROC-AUC scores are smaller for the WNPcoast region (Fig. 7). The ROC-AUC scores are particularly high for active seasons over the NA and NAcoast regions, whereas the ROC-AUC scores are lower for inactive seasons, which is especially the case for UKMO-GloSea5-GC2 over the NAcoast. To understand differences between active and inactive seasons, we analyze the time series of TC counts over the NAcoast region for each model (Fig. S2). In reanalyses the most active seasons are 1995, 1998, 1999, 2005, 2008, 2010, and 2011 and the most inactive seasons are 1993, 1994, 1997, 2006, 2007, 2009, and 2013. Here we focus on the representation of active and inactive seasons in UKMO-GloSea5-GC2 and ECMWF-SEAS5. For both models the percentage of ensemble members correctly predicting an (observed) active season is on average larger compared to the average percentage of

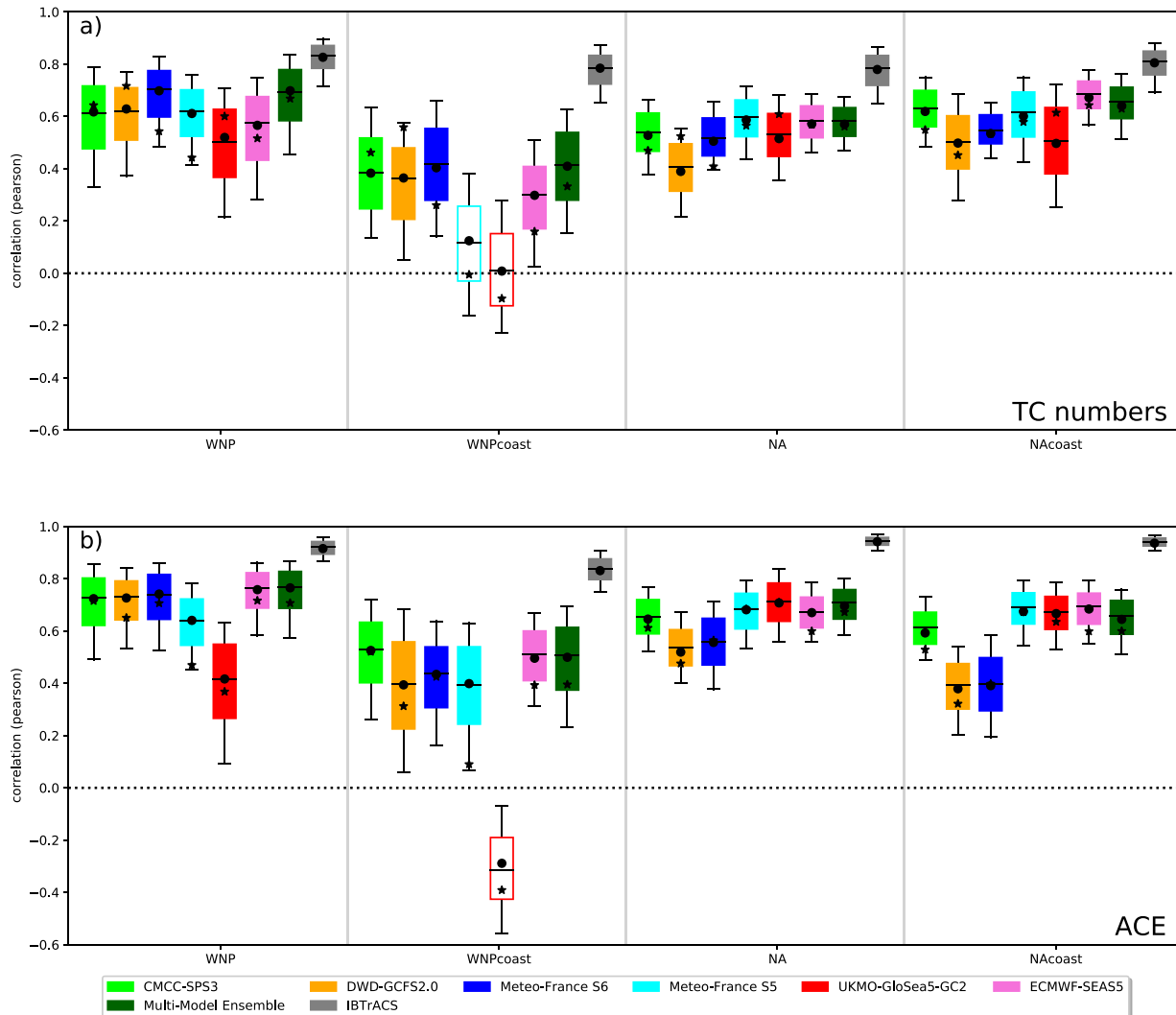


FIG. 5. (a) Anomaly correlation coefficients (ACC) for TC numbers over WNP, NA, WNPcoast, and NAcoast. Significance is tested using a 1000 sample bootstrap, where whiskers indicate the 10th and 90th percentile. Filled box-and-whisker plots indicate models with significant positive correlations. The dot indicates the linear correlation coefficient between the respective model and the MMR, whereas the star indicates the linear correlation between IBTrACS and the seasonal forecast model. (b) As in (a), but for accumulated cyclone energy (ACE).

ensemble members correctly predicting an (observed) inactive season. The UKMO-GloSea5-GC2 model especially fails to predict the inactive TC seasons of 2013 and 2006, whereas the inactive seasons in 2006 and 2007 are not well predicted by ECMWF-SEAS5.

Next, the reliability of the different seasonal forecast systems is assessed using the spread-over-error statistic. As explained in section 2, for a perfectly reliable ensemble the average ensemble spread matches the error of the ensemble mean. An overdispersive (underconfident) forecast is thus a forecast with a too large spread compared to the error, whereas an underdispersive (overconfident) forecast is characterized by a too small spread compared to its error. For TC

numbers, we find that over the WNP most models are overdispersive (underconfident) except DWD-GCFS2.0 and CMCC-SPS3, with the latter being the only model which is not significantly unreliable over the WNP and WNPcoast regions (Fig. 8a). A different result is found over the NA region, where the reliability for all models except CMCC-SPS3 and DWD-GCFS2.0 cannot be distinguished from a perfect value of 1. Reliability for the ACE is different compared to that found for TC numbers, as all models tend to be overconfident (underdispersive) over the WNP and NA regions (Fig. 8b). The ECMWF-SEAS5 and Météo-France-S6 systems over the WNP and WNPcoast are the only forecasts whose reliability measured by the SoE cannot be distinguished from the

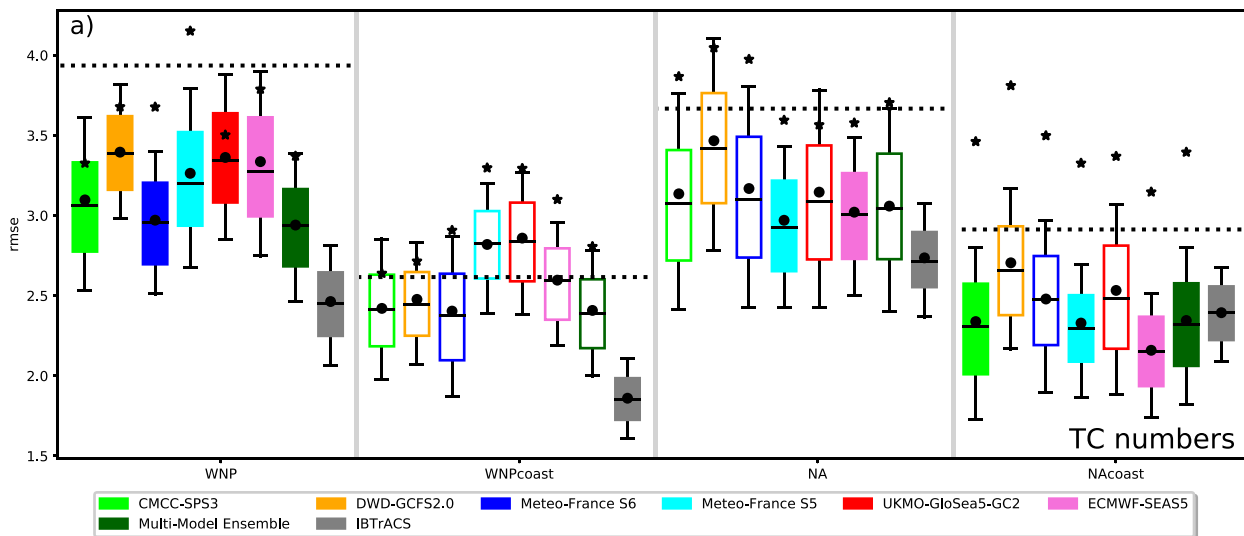


FIG. 6. RMSE for WNP, NA, WNPcoast, and NAcoast. Regions, seasons, and confidence intervals as in Fig. 5. Filled box-and-whisker plots indicate models with RMSE values, which are significantly lower than the RMSE value of a climatological forecast (dotted line). The dot indicates the RMSE between the respective model and the MMR, whereas the star indicates the RMSE between IBTrACS and the seasonal forecast model.

reliability of a perfect system. It should be noted that reliability can potentially be improved by applying calibration methods as discussed in Camp et al. (2018).

#### 4. Summary and discussion

In this study, the ability of six European seasonal prediction systems (DWD-GCFS2.0, CMCC-SPS3, ECMWF-SEAS5, UKMO-GloSea5-GC2, Météo-France-S5, and Météo-France-S6) to represent observed characteristics of tropical cyclones over the NA and WNP has been assessed for the common hindcast period from 1993 until 2014. These characteristics include climatological aspects such as temporal and spatial variability and TC intensities as well as the skill of the models to simulate the interannual TC frequency and accumulated cyclone energy. Due to differences in the observed seasonal cycle characteristics, the season from June to October (JJASO) is used for the WNP, whereas July–October (JASO) is used for the NA basin. To compare the level of skill over both ocean basins, May start dates are used for verification of TCs over the WNP and June start dates for the NA. Thus, we use a lead time of 1 month for each basin. Besides the main NA and WNP basins, smaller subregions along the coastlines named NAcoast and WNPcoast are also used. The seasonal forecasts are verified against the multi-reanalysis mean consisting of five different reanalyses (ERA5, ERA-Interim, NCEP-CFSR, JRA-55, and MERRA-2) as well as IBTrACS observations.

We find large differences between the seasonal forecasting systems in representing the climatological properties of tropical cyclones. The UKMO-GloSea5-GC2 and Météo-France-S5 models overestimate TC occurrences over the NA and WNP basins, whereas DWD-GCFS2.0 and CMCC-SPS3 underestimate TCs over these two basins. Besides large differences

between the models with regards to seasonal mean TC numbers, the ability of the models to simulate the observed seasonal cycle varies considerably between the models, with the UKMO-GloSea5-GC2 and ECMWF-SEAS5 models showing the best performance. Apart from structural differences between the models and resolutions, they also differ with regard to the initialization techniques used. For example, the UKMO-GloSea5-GC2 model uses a lagged initialization technique, in contrast to, say, the ECMWF model. The impact of the lagged start dates is not the focus of this study; however, lead time-dependent biases have been analyzed using hindcasts initialized in May, June, July, and August for all forecast systems. These show rather similar biases in terms of track densities for all the initializations as those already shown (see Figs. S9 and S10 in the online supplemental material), indicating that the different initialization techniques only contribute in a limited way to the mean seasonal TC biases found in the models (given the relative proximity in time of the different lagged start dates; e.g., a maximum of 3 weeks apart for UKMO-GloSea5-GC2).

All seasonal forecast systems have difficulties in representing the observed intensity distribution over the WNP and NA basins, with too few strong TC events simulated. Intensity biases are largest for the two models with the lowest resolution, CMCC-SPS3 and DWD-GCFS2.0, which is in line with findings from previous studies (e.g., Roberts et al. 2020). However, ECMWF-SEAS5 and UKMO-GloSea5-GC2 also include stochastic schemes to represent model uncertainty. Stochastic schemes have been found to have a positive impact on tropical SSTs in seasonal predictions (e.g., Befort et al. 2020; Weisheimer et al. 2014), and a recent study by Vidale et al. (2021) suggests that such schemes can have a similar effect on the frequency of TCs to increasing the horizontal

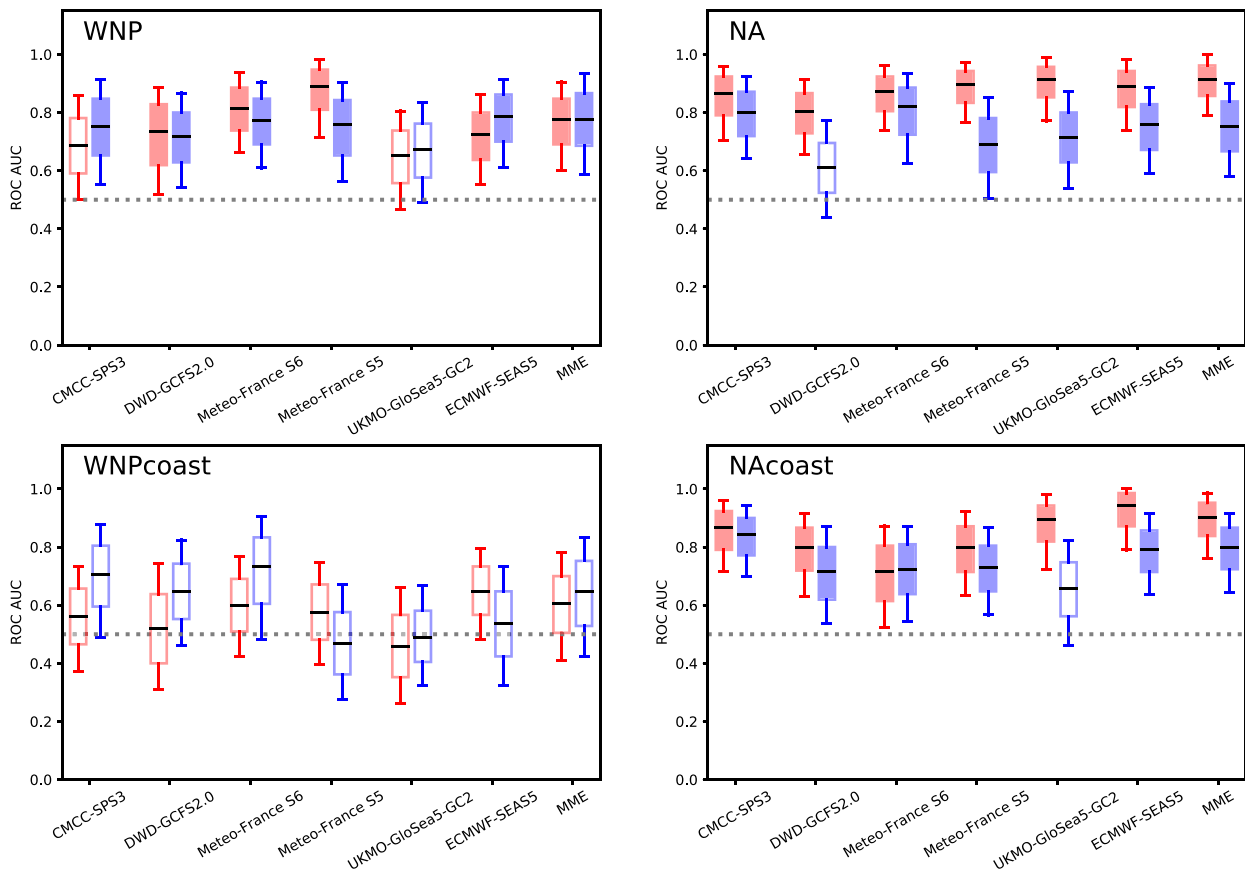


FIG. 7. ROC scores for active (upper tercile; red) and inactive season (lower tercile; blue) for the WNP, WNPcoast, NA, and NAcoast regions. Seasons and confidence intervals as in Fig. 5. Filled box-and-whisker plots indicate models with ROC scores, which are significantly larger than using a climatological forecast (dotted line).

resolution. The low computational costs of such stochastic schemes provides motivation for further research to assess if such schemes improve TC characteristics in seasonal forecasts.

We assess the skill of all six seasonal forecast systems in predicting TCs using several deterministic and probabilistic metrics. The use of several measures is motivated by the short common hindcast period of only 22 years as it is assumed that positive significant skill in several metrics is an indicator of more robust skill results. Significant positive correlations for both tropical cyclone counts and ACE are found over the NA and WNP basins for all models, in line with previous studies (e.g., Camp et al. 2015; Manganello et al. 2019). We find high correlations over the NAcoast region, including the Caribbean, which agrees with results from Manganello et al. (2019), who found that the MME consisting of models from the NMME-Phase II show significant positive correlations for TC counts over the Caribbean. Moderate correlation coefficient values are also found for the WNP basin. However, in contrast to results for the NA basin, skill for all models is largely reduced for the WNPcoast regions compared to the whole WNP basin. This is especially pronounced for the UKMO-GloSea5-GC2 model, in agreement with the results presented in Camp et al. (2019) and discussed in Feng et al. (2020), who showed that the limited skill of GloSea5-GC2 for TCs in the

northeast WNP is related to the overestimation of the negative TC–ENSO teleconnection. Even though our analysis has revealed large differences for TC numbers and intensities between the different models, this seems not to translate into differences in skill. However, as intensities and numbers are largely biased in the models, calibration techniques are necessary to obtain meaningful TC counts and intensities.

In addition to correlation coefficients, RMSEs are calculated for each region and each forecast system. These results support those for correlation coefficients, with significantly lower RMSEs for all models than for a climatological forecast over the WNP and for some models over the NA and NAcoast regions. Over the WNPcoast region RMSEs for all models are similar to the RMSE from a climatological forecast. Besides these deterministic skill metrics, the ability of the models to discriminate between active and inactive seasons has been assessed. Again, these results indicate that the models are skillful over the NA, NAcoast, and WNP but less skillful for the WNPcoast region. Our results for the NA are in line with those from Vitart et al. (2007), who showed the EUROSIP ensemble was able to discriminate between the hurricane seasons of 2005 and 2006. Interestingly, some of the models analyzed here are less skillful to simulate TC activity during 2005

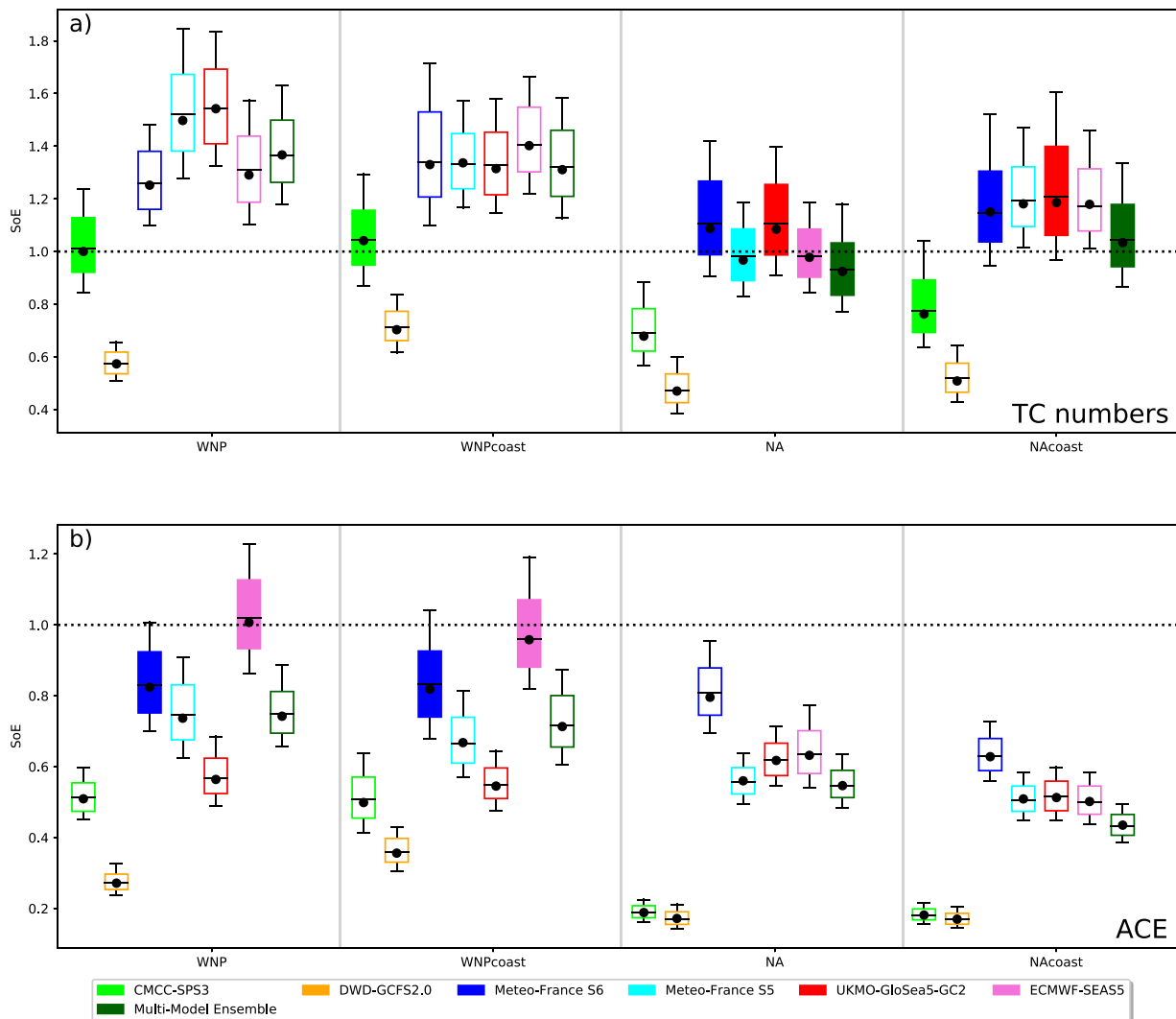


FIG. 8. (a) Spread-over-error statistic (SoE) for TC numbers over WNP, NA, WNPcoast, and NAcoast. Regions, seasons, and confidence intervals are as in Fig. 5. Filled box-and-whisker plots indicate models with an SoE value that cannot be distinguished from 1 (perfect reliability). (b) As in (a), but for ACE.

and 2006 (see Fig. S1). In this study, especially the low TC activity during 2006 is not well captured by the models. However, results presented in Vitart et al. (2007) are based on 2 years only. Here the full common hindcast period from 1993 until 2014 has been analyzed, which provides a much more robust assessment of the models ability to discriminate between active and inactive seasons.

The reliability of the different seasonal forecast ensembles has been assessed using the spread-over-error (SoE) statistic. For TC counts, it is found that most models over the NA and NAcoast are reliable in a statistical sense, meaning that the average ensemble spread matches the error of the ensemble mean (Fortin et al. 2014). In contrast, most models are underconfident over the WNP and WNPcoast regions. For ACE most models are overconfident meaning that ACE values from reanalyses are often outside the predicted range of the

models. Exceptions are the ECMWF-SEAS5 and Météo-France-S6 models over the WNP and WNPcoast regions with these ensembles being reliable.

Overall, these results suggest that the six seasonal forecasting models used in this study provide useful information, especially over the North Atlantic basin but also over a subregion centered over the Caribbean. All skill scores suggest that seasonal forecasts over this region may be an useful tool for potential end users and for decision making processes. In contrast, skill over the WNPcoast region is much smaller for all systems.

One shortcoming of this study is the short time period analyzed, which is due to the common hindcast period of all models. Previous studies have shown large decadal-scale variability of seasonal forecast models with regard to TC frequency (e.g., Manganello et al. 2019). Recently developed century-long seasonal hindcasts (Weisheimer et al. 2017, 2020) might prove to



be a useful tool to evaluate long-term variability in skill and also how far this variability is related to changes in observation density and/or in predictive skill of atmospheric and oceanic large-scale conditions. This study has not investigated the causes for the presence or nonexistence of skill. On seasonal time scales ENSO is known to heavily affect TC intensity and frequency over the WNP and NA basins (e.g., Camargo et al. 2010). Thus, future research on this topic should also assess the ability of the models to simulate ENSO events as well as the teleconnections to TCs in both basins [as was done for the UKMO-GloSea5-GC2 model in, e.g., Camp et al. (2015) and Feng et al. (2020)].

**Acknowledgments.** This study received support from the European Union's Horizon 2020 EUCP project (Grant GA 776613). Kevin Hodges acknowledges the support of the Natural Environment Research Council (NERC). The authors thank the three anonymous reviewers for their valuable comments.

**Data availability statement.** All seasonal forecast data from ECMWF SEAS5, Met Office GloSea5, CMCC, DWD, and both Météo-France models can be retrieved via the C3S webpage (<https://climate.copernicus.eu/>).

## REFERENCES

- Beaufort, D. J., and Coauthors, 2019: Seasonal forecast skill for extratropical cyclones and windstorms. *Quart. J. Roy. Meteor. Soc.*, **145**, 92–104, <https://doi.org/10.1002/qj.3406>.
- , T. Kruschke, and G. C. Leckebusch, 2020: Objective identification of potentially damaging tropical cyclones over the Western North Pacific. *Environ. Res. Commun.*, **2**, 031005, <https://doi.org/10.1088/2515-7620/ab7b35>.
- Bell, S. S., S. S. Chand, S. J. Camargo, K. J. Tory, C. Turville, and H. Ye, 2019: Western North Pacific tropical cyclone tracks in CMIP5 models: Statistical assessment using a model-independent detection and tracking scheme. *J. Climate*, **32**, 7191–7208, <https://doi.org/10.1175/JCLI-D-18-0785.1>.
- Brookshaw, A., 2017: C3S trials seasonal forecast service. *ECMWF Newsletter*, No. 150, ECMWF, Reading, United Kingdom, <https://www.ecmwf.int/en/newsletter/150/news/c3s-trials-seasonal-forecast-service>.
- Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, <https://doi.org/10.1175/JCLI-D-12-00549.1>.
- , A. H. Sobel, A. G. Barnston, and P. J. Klotzbach, 2010: The influence of natural climate variability on tropical cyclones, and seasonal forecasts of tropical cyclone activity. *Global Perspectives on Tropical Cyclones*, J. C. L. Chan and J. D. Kepert, Eds., World Scientific, 325–360, [https://doi.org/10.1142/9789814293488\\_0011](https://doi.org/10.1142/9789814293488_0011).
- Camp, J., M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 2206–2219, <https://doi.org/10.1002/qj.2516>.
- , and Coauthors, 2018: Skillful multiweek tropical cyclone prediction in ACCESS-S1 and the role of the MJO. *Quart. J. Roy. Meteor. Soc.*, **144**, 1337–1351, <https://doi.org/10.1002/qj.3260>.
- , and Coauthors, 2019: The western Pacific subtropical high and tropical cyclone landfall: Seasonal forecasts using the Met Office GloSea5 system. *Quart. J. Roy. Meteor. Soc.*, **145**, 105–116, <https://doi.org/10.1002/qj.3407>.
- Caron, L.-P., M. Boudreault, and C. L. Bruyère, 2015: Changes in large-scale controls of Atlantic tropical cyclone activity with the phases of the Atlantic multidecadal oscillation. *Climate Dyn.*, **44**, 1801–1821, <https://doi.org/10.1007/s00382-014-2186-5>.
- Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, <https://doi.org/10.1175/JCLI-D-12-00061.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dorel, L., C. Ardilouze, L. Batté, M. Déqué, and J.-F. Guérémy, 2017: Documentation of the Météo-France pre-operational seasonal forecasting system. Météo-France, 32 pp., <https://www.umr-cnrm.fr/IMG/pdf/system6-technical.pdf>.
- Feng, X., N. P. Klingaman, K. I. Hodges, and Y.-P. Guo, 2020: Western North Pacific tropical cyclones in the Met Office Global Seasonal Forecast System: Performance and ENSO teleconnections. *J. Climate*, **33**, 10489–10504, <https://doi.org/10.1175/JCLI-D-20-0255.1>.
- Fink, A. H., J. M. Schrage, and S. Kotthaus, 2010: On the potential causes of the nonstationary correlations between West African precipitation and Atlantic hurricane activity. *J. Climate*, **23**, 5437–5456, <https://doi.org/10.1175/2010JCLI3356.1>.
- Fortin, V., M. Abaza, F. Anctil, and R. Turcotte, 2014: Why should ensemble spread match the RMSE of the ensemble mean? *J. Hydrometeorol.*, **15**, 1708–1713, <https://doi.org/10.1175/JHM-D-14-0008.1>.
- Fröhlich, K., and Coauthors, 2021: The German Climate Forecast System: GCFS. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002101, <https://doi.org/10.1029/2020MS002101>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gualdi, S., and Coauthors, 2020: The new CMCC Operational Seasonal Prediction System. CMCC Tech. Rep., 34 pp., <https://doi.org/10.25424/CMCC/SPS3.5>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hodges, K. I., 1995: Feature tracking on the unit-sphere. *Mon. Wea. Rev.*, **123**, 3458–3465, [https://doi.org/10.1175/1520-0493\(1995\)123<3458:FTOTUS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2).
- , 1996: Spherical nonparametric estimators applied to the UGAMP model integration for AMIP. *Mon. Wea. Rev.*, **124**, 2914–2932, [https://doi.org/10.1175/1520-0493\(1996\)124:2914:SNEATT.2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124:2914:SNEATT.2.0.CO;2).
- , 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373, [https://doi.org/10.1175/1520-0493\(1999\)127<1362:ACFFT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2).
- , A. Cobb, and P. L. Vidale, 2017: How well are tropical cyclones represented in reanalysis datasets? *J. Climate*, **30**, 5243–5264, <https://doi.org/10.1175/JCLI-D-16-0557.1>.
- Horn, M., and Coauthors, 2014: Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *J. Climate*, **27**, 9197–9213, <https://doi.org/10.1175/JCLI-D-14-00200.1>.

- Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>.
- Kharin, V. V., F. W. Zwiers, and N. Gagnon, 2001: Skill of seasonal hindcasts as a function of the ensemble size. *Climate Dyn.*, **17**, 835–843, <https://doi.org/10.1007/s003820100149>.
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, <https://doi.org/10.1175/2009BAMS2755.1>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- LaRow, T. E., L. Stefanova, D.-W. Shin, and S. Cocke, 2010: Seasonal Atlantic tropical cyclone hindcasting/forecasting using two sea surface temperature datasets. *Geophys. Res. Lett.*, **37**, L02804, <https://doi.org/10.1029/2009GL041459>.
- MacLachlan, C., and Coauthors, 2015: Global Seasonal Forecast System version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Manganello, J. V., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, <https://doi.org/10.1175/JCLI-D-11-00346.1>.
- , and Coauthors, 2016: Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *J. Climate*, **29**, 1179–1200, <https://doi.org/10.1175/JCLI-D-15-0531.1>.
- , B. A. Cash, K. I. Hodges, and J. L. Kinter, 2019: Seasonal forecasts of North Atlantic tropical cyclone activity in the North American Multi-Model Ensemble. *Climate Dyn.*, **53**, 7169–7184, <https://doi.org/10.1007/s00382-017-3670-5>.
- Météo-France, 2015: Météo-France seasonal forecast system 5 for Eurosip. Météo-France Tech. Rep., 38 pp., <https://www.umr-cnrm.fr/IMG/pdf/system5-technical.pdf>.
- Murakami, H., 2014: Tropical cyclones in reanalysis data sets. *Geophys. Res. Lett.*, **41**, 2133–2141, <https://doi.org/10.1002/2014GL059519>.
- , and Coauthors, 2015: Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **28**, 9058–9079, <https://doi.org/10.1175/JCLI-D-15-0216.1>.
- Palmer, T. N., and Coauthors, 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872, <https://doi.org/10.1175/BAMS-85-6-853>.
- Pant, S., and E. J. Cha, 2019: Wind and rainfall loss assessment for residential buildings under climate-dependent hurricane scenarios. *Struct. Infrastruct. Eng.*, **15**, 771–782, <https://doi.org/10.1080/15732479.2019.1572199>.
- Roberts, M. J., and Coauthors, 2020: Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *J. Climate*, **33**, 2557–2583, <https://doi.org/10.1175/JCLI-D-19-0639.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Saunders, M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature*, **434**, 1005–1008, <https://doi.org/10.1038/nature03454>.
- Scaife, A. A., and Coauthors, 2019: Does increased atmospheric resolution improve seasonal climate predictions? *Atmos. Sci. Lett.*, **20**, e922, <https://doi.org/10.1002/asl.922>.
- Stockdale, T., 2013: The EUROSIP system—A multi-model approach. *Seminar on Seasonal Prediction: Science and Applications*, ECMWF, 257–268, <https://www.ecmwf.int/node/12429>.
- Ullrich, P. A., and C. M. Zarzycki, 2017: TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geosci. Model Dev.*, **10**, 1069–1090, <https://doi.org/10.5194/gmd-10-1069-2017>.
- Vidale, P. L., and Coauthors, 2021: Impact of stochastic physics and model resolution on the simulation of tropical cyclones in climate GCMs. *J. Climate*, **34**, 4315–4341, <https://doi.org/10.1175/JCLI-D-20-0507.1>.
- Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteor. Soc.*, **132**, 647–666, <https://doi.org/10.1256/qj.05.65>.
- , and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537, [https://doi.org/10.1175/1520-0493\(2001\)129<2521:SFOTSU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2).
- , and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EURO-SIP. *Geophys. Res. Lett.*, **34**, L16815, <https://doi.org/10.1029/2007GL030740>.
- Walsh, K., S. Lavender, E. Scoccimarro, and H. Murakami, 2013: Resolution dependence of tropical cyclone formation in CMIP3 and finer resolution models. *Climate Dyn.*, **40**, 585–599, <https://doi.org/10.1007/s00382-012-1298-z>.
- Wang, Y., S. Wen, X. Li, F. Thomas, B. Su, R. Wang, and T. Jiang, 2016: Spatiotemporal distributions of influential tropical cyclones and associated economic losses in China in 1984–2015. *Nat. Hazards*, **84**, 2009–2030, <https://doi.org/10.1007/s11069-016-2531-6>.
- Weisheimer, A., S. Corti, T. Palmer, and F. Vitart, 2014: Addressing model error through atmospheric stochastic physical parameterizations: Impact on the coupled ECMWF seasonal forecasting system. *Philos. Trans. Roy. Soc.*, **372A**, 20130290, <https://doi.org/10.1098/rsta.2013.0290>.
- , N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer, 2017: Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quart. J. Roy. Meteor. Soc.*, **143**, 917–926, <https://doi.org/10.1002/qj.2976>.
- , D. J. Bafort, D. MacLeod, T. Palmer, C. O'Reilly, and K. Strømme, 2020: Seasonal forecasts of the twentieth century. *Bull. Amer. Meteor. Soc.*, **101** (8), E1413–E1426, <https://doi.org/10.1175/BAMS-D-19-0019.1>.
- Zhao, M., I. M. Held, and G. A. Vecchi, 2010: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868, <https://doi.org/10.1175/2010MWR3366.1>.