Impact of Stochastic Physics and Model Resolution on the Simulation of Tropical Cyclones in Climate GCMs

Pier Luigi Vidale,[a] Kevin Hodges,[a,b] Benoit Vannière,[a] Paolo Davini,[c] Malcolm J. Roberts,[d]
Kristian Strommen,[e] Antje Weisheimer,[f,g] Elina Plesca,[a] and Susanna Corti[c]

[a] *NCAS-Climate, Department of Meteorology, University of Reading, Reading, United Kingdom*
[b] *Department of Meteorology, University of Reading, Reading, United Kingdom*
[c] *Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, Torino, Italy*
[d] *Met Office Hadley Centre, United Kingdom*
[e] *Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom*
[f] *NCAS, Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom*
[g] *ECMWF, Reading, United Kingdom*

ABSTRACT: The role of model resolution in simulating geophysical vortices with the characteristics of realistic tropical cyclones (TCs) is well established. The push for increasing resolution continues, with general circulation models (GCMs) starting to use sub-10-km grid spacing. In the same context it has been suggested that the use of stochastic physics (SP) may act as a surrogate for high resolution, providing some of the benefits at a fraction of the cost. Either technique can reduce model uncertainty, and enhance reliability, by providing a more dynamic environment for initial synoptic disturbances to be spawned and to grow into TCs. We present results from a systematic comparison of the role of model resolution and SP in the simulation of TCs, using EC-Earth simulations from project Climate-SPHINX, in large ensemble mode, spanning five different resolutions. All tropical cyclonic systems, including TCs, were tracked explicitly. As in previous studies, the number of simulated TCs increases with the use of higher resolution, but SP further enhances TC frequencies by ~30%, in a strikingly similar way. The use of SP is beneficial for removing systematic climate biases, albeit not consistently so for interannual variability; conversely, the use of SP improves the simulation of the seasonal cycle of TC frequency. An investigation of the mechanisms behind this response indicates that SP generates both higher TC (and TC seed) genesis rates, and more suitable environmental conditions, enabling a more efficient transition of TC seeds into TCs. These results were confirmed by the use of equivalent simulations with the HadGEM3-GC31 GCM.

KEYWORDS: Tropical cyclones; Numerical weather prediction/forecasting; Climate models; General circulation models; Stochastic models; Interannual variability

1. Introduction

The simulation of tropical cyclones (TCs) in contemporary climate models remains a challenge, with a systematic underestimation of both the number of TCs and their intensity (Shaevitz et al. 2014; Walsh et al. 2015; Roberts et al. 2020). While the importance of atmospheric model resolution in reducing these biases has long been recognized (Camargo and Wing 2016; Emanuel 2018) and increasing the horizontal resolution has been demonstrated to improve the representation of TCs in climate models (Jung et al. 2012; Murakami et al. 2012; Strachan et al. 2013; Murakami et al. 2014; Knutson et al. 2013; Manganello et al. 2014; Roberts et al. 2015; Bhatia et al. 2018; Bacmeister et al. 2018; Roberts et al. 2020), alternative stochastic approaches have been put forward in recent years [see Palmer (2019) for a motivational discussion]. Stochastic approaches account for unresolved processes and variability in models, and have the potential to complement some of the benefits of resolution. Moreover, these ideas explicitly acknowledge the multiscale nature of the atmosphere and the role of scale interactions for weather and climate variability. In this study we analyze the statistics of TCs in a systematic set of ensemble climate simulations, carried out at varying horizontal resolutions—with and without the inclusion of stochastic schemes—in order to quantify and compare their impact.

a. Stochastic physics and uncertainty in climate prediction

Traditional parameterization schemes rely on the assumption that the physics of the unresolved processes is uncoupled from the dynamics of the flow. However, in the presence of upscale energy cascades (e.g., through convection) the assumption clearly does not hold and nonlinear interactions between different scales [i.e., from meso-α to meso-γ (2–20 km)], comprising mesoscale convective systems to TCs, as well as organized convection like the Madden–Julian oscillation, cannot be adequately modeled. Stochastic physics (SP) parameterization schemes aim to account for the missing nonlinear interactions between unresolved processes, as well as their impacts on the large scale, and thus to account for some of the missing scale interactions in models (Palmer 2001; Williams 2005; Khouider et al. 2010; Slingo and Palmer 2011).

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

Corresponding author: Pier Luigi Vidale, p.l.vidale@reading.ac.uk

The development of SP schemes for weather and climate models was pioneered by the numerical weather prediction community and motivated by the need to explicitly represent uncertainty due to model errors on subgrid scales (Buizza et al. 1999; Shutts and Palmer 2007; Plant and Craig 2008; Teixeira and Reynolds 2008; Berner et al. 2009; Bengtsson et al. 2013; Sanchez et al. 2014). However, they are now commonly used in many operational forecasting centers; see Leutbecher et al. (2017) for a recent overview. In the context of atmospheric seasonal forecasts, Berner et al. (2012) argue that, for tropical precipitation biases and tropical variability, increasing the horizontal resolution has a rather small impact compared to the use of stochastic schemes or improved physical parameterizations.

SP schemes have also become increasingly relevant for climate simulations, mainly due to their potential role in modifying the mean state, via noise-induced drift processes, and thus reduce intransigent biases [see Berner et al. (2017) for a review and Palmer (2019) for a more recent comprehensive perspective]. A common theme in all these studies is the crucial role played by nonlinear interactions of the stochastic scheme with convective processes, which allow even a zero-mean perturbation to profoundly impact the model climate attractor. It is thus expected that SP also influences tropical storms in general and TCs in particular [see Stockdale et al. (2018) for an assessment of the SP impact on ECMWF's seasonal forecasts].

It should be stressed that the computational costs of adding SP to state-of-the-art climate models are very small indeed: while the doubling of resolution typically causes costs to rise by a factor of 8–10, stochastic perturbations such as those used in this study increase the costs by only ~5%.

b. TCs and prediction uncertainty

The simulation of TCs challenges all our current capabilities: their multiscale nature has justified the use of any and all affordable resolutions (e.g., Jung et al. 2012; Roberts et al. 2015, 2020). Their low annual frequency and large variability, from days to decades, require the use of ensembles and long simulations, including for predictions under climate change (Yoshida et al. 2017; Mei et al. 2019). Their sensitivity to the large-scale environment requires minimal model biases, which depends on the quality of model dynamics and physical parameterizations (LaRow 2013; Murakami et al. 2014; He and Posselt 2015; Kepert 2012; Camargo et al. 2020).

There are many challenges and opportunities in the simulation of large-scale TC drivers: the governing nature of El Niño–Southern Oscillation (ENSO) on TC location and frequency has been known for a long time (e.g., Gray 1984; Chan 1985; Lander 1994; Camargo et al. 2007b; Bell et al. 2014), but it is also increasingly evident that the presence of strong TC activity in the northwest Pacific can affect the Niño-3.4 index three months later (Wang et al. 2019). Thus the simulation of TCs could benefit from the improved simulation of ENSO using SP (e.g., Christensen et al. 2017); conversely, a better simulation of TC location, intensity, and frequency, particularly interannual variability, could benefit the prediction of ENSO in global seasonal and climate simulations [e.g., as implied by Wang et al. (2019)].

Prediction of TCs on decadal and climate time scales is also important to provide longer-term mitigation planning, in particular for coastal communities, which are particularly vulnerable to changes in TC track and intensity in a changing climate [see quantification of TC-related precipitation in Guo et al. (2017) and Franco-Díaz et al. (2019)].

There is consequently considerable uncertainty between climate models as to the climatology, variability, and changes in TCs with climate change (Camargo and Wing 2016; Knutson et al. 2020). Vecchi et al. (2019) have shown, in the context of the prediction of future changes in TCs, that pre-TC synoptic-scale disturbances (which they call "TC seeds") are the main drivers of the simulated response (see also Sugi et al. 2020; Yamada et al. 2021), where they also discuss changes in the large-scale environment governing their origins and development. Such disturbances are substantially weaker than the fully formed TCs, so that their representation in GCMs is even more uncertain [Slingo et al. 1994; see also the conclusions of Hodges et al. (2017) with regard to the skill in current reanalyses]. Other recent papers, such as the downscaling study of Emanuel (2021), use prescribed seeding, implicitly suggesting that this process is of secondary importance in the study of climate change and that what matters are changes in the TC environment.

c. TCs and GCM resolution

For the past 20 years the climate modeling community has been investigating the role of model resolution in the simulation of TCs; for example, Bengtsson et al. (1995), building on the work of Broccoli and Manabe (1990) and inspired by questions in Evans (1992), successfully simulated the climatology of TCs (then called tropical cyclone like vortices) with ECHAM3 at T_q106 truncation ($\approx$125-km Δx on a quadratic Gaussian grid). A comprehensive history of these studies has been traced in a review paper (Camargo and Wing 2016) that spans NWP and seasonal applications; Walsh et al. (2013) and Emanuel (2018) also summarize the impact of resolution on TCs in the context of climate simulations.

More recently, the modeling community has organized a number of intercomparison projects focusing on TC simulation [see, e.g., results from the U.S. CLIVAR Hurricane Working Group in Shaevitz et al. (2014)]. TCs are also one of the target phenomena in PRIMAVERA-HighResMIP project (Haarsma et al. 2016), which aims to systematically understand the role of model resolution in the context of climate simulations.

Based on previous studies, the current consensus is that realistic TCs "emerge" in climate models at sub-100-km resolution, and their track densities, as well as interannual variability by basin, start to look credible at about 20-km resolution (less so for the Atlantic basin), as shown for instance in Shaevitz et al. (2014) and Roberts et al. (2020). GCMs have also demonstrated increased skill in simulating interannual variability of TCs as resolution is increased (Zhao et al. 2009; Strachan et al. 2013; Roberts et al. 2015), but large intermodel and intra-ensemble uncertainties remain, so that ensemble sizes of at least $O(10)$ are required (Yoshida et al. 2017; Mei et al. 2019; Roberts et al. 2020). Credible simulations of TC intensity remains elusive (for surface winds, but some GCMs simulate TCs that are too deep; Manganello et al. 2012). Roberts et al. (2020) demonstrate that only one of the PRIMAVERA GCMs (CNRM) presents a credible pressure–wind relationship, and even a few

TABLE 1. The experimental configuration in SPHINX.

Truncation	Equivalent resolution (m)	No. of ensemble members (base + stochastic)	Label	Prefix
T_L159	125 150	10 + 10	COARSE	CA
T_L255	80 600	10 + 10	LOW	LA
T_L511	40 200	6 + 6	MEDIUM	MA
T_L799	25 150	3 + 3	HIGH	HA
T_L1279	16 100	1 + 1	ULTRA-HIGH	UA

of the DYAMOND (i.e., Dynamics of the Atmospheric General Circulation Modeled on Non-hydrostatic Domains) GCMs (Stevens et al. 2019), with an average resolution of 5 km, struggle with this aspect [see further details in Judt et al. (2021)].

SP thus seems to be particularly well suited for the simulation of TCs because it targets unresolved variability, delivering some of the benefits of resolution, while enabling the use of large ensembles by being parsimonious. The focus of the present study is to investigate the relative impact of enhancing resolution and/or using SP on the representation of TCs in climate simulations.

The main questions of this study are the following: Can SP be an efficient surrogate for model resolution, and, specifically, can it be beneficial for the simulation of tropical cyclones? Therefore two hypotheses will be considered with regards to the role of SP:

1) SP injects flow-dependent noise into the simulations, increasing the initial number of (cyclonic) tropical disturbances that can possibly grow into TCs [i.e., TC seeds, as in Vecchi et al. (2019)].
2) SP creates a more favorable large-scale environment for genesis and development of TCs, revealed as increased likelihood of TC seeds transitioning into TCs.

It is also possible that both hypotheses are valid in tandem, and we use explicit tracking, as well as an empirical approach (index-based), to investigate both. The specific aspects under investigation are the climatology of TC numbers and geographical distribution, as well as their interannual variability; TC intensity, while briefly analyzed in this study, will be the focus of a follow-on paper, including the analysis of Climate-SPHINX (i.e., Stochastic Physics High Resolution Experiments) coupled (ocean–atmosphere) and climate change simulations.

The paper continues with a description of the models and analysis methods in section 2, and the results are presented in sections 3–5, spanning means and variability of TCs, including its drivers. A discussion is given in section 6 and a summary and final conclusions in section 7.

2. Models and methods

a. EC-Earth configuration, and the Climate-SPHINX project campaign

In the present work, EC-Earth version 3.1 is used, exploiting a set of atmosphere-only simulations carried out within the Climate-SPHINX project [see Davini et al. (2017), which also provides details of the scientific configuration].

All the simulations span the period from 1979 to 2008 (30 years) where well-mixed GHGs, stratospheric ozone, and volcanic aerosol concentrations follow the CMIP5 (phase 5 of the Coupled Model Intercomparison Project) protocol (historic forcing ending in 2005, then RCP8.5; Moss et al. 2010). Sea surface temperatures (SSTs) and sea ice concentration used as boundary conditions have been obtained from the HadISST2.1.1 dataset (Titchner and Rayner 2014), modified to provide daily increments suitable for high resolution (Kennedy et al. 2017, the same used in HighResMIP).

The hydrostatic EC-Earth atmospheric component is based on the Integrated Forecasting System (IFS) and has been tuned and improved for climate purposes by the EC-Earth Consortium (Hazeleger et al. 2010). It is important to stress that, for a clear interpretation of the sensitivity studies in this paper, model tuning has been performed only for the T255 deterministic (hereafter referred to as BASE) configuration, and all other simulations are performed without retuning, in order to enable a clear understanding of the role of resolution and SP. Therefore, energy budgets at resolutions different from T255 present small biases [see Davini et al. (2017) for details].

The Climate-SPHINX simulations comprise several ensemble members over a range of five resolutions from T_L159 (≈125 km) to T_L1279 (≈16 km; see also Table 1) but retain the same vertical grid configuration (L91), as hybrid sigma levels with the last full level at 0.01 hPa. For each resolution—which is defined by the spectral truncation—half of the ensemble members have the stochastic physics parameterizations activated (STOC) and half are run with only the BASE configuration. The prefixes for each experiment are shown in the last column of Table 1, and the suffix B indicates BASE experiments, while S stands for STOC, so that "CAB" is the BASE experiment at COARSE resolution and "CAS" is the corresponding STOC experiment. The number of ensemble members starts at 10 for the lowest resolutions, for each of BASE and STOC, and decreases to 1 for the highest resolution (see Table 1), due to computational costs. The stochastic parameterizations used within EC-Earth comprise two different schemes: the stochastically perturbed parameterization tendencies (SPPT) and the stochastic kinetic energy backscatter (SKEB) scheme (Berner et al. 2009; Palmer et al. 2009; Davini et al. 2017). Both schemes are always used together in all STOC experiments in SPHINX; the SPPT configuration is exactly the same across all model resolutions, while SKEB uses different backscatter ratios at each resolution.

The SPPT scheme focuses on the uncertainty arising from the existing subgrid parameterization schemes (including radiation, clouds, convection, turbulence and boundary layer processes, and gravity wave drag) using a multiplicative noise

approach. The SPPT scheme perturbs the total diabatic tendencies for T, U, V, and q at each time step, using the same perturbation field e, which is the sum of three independent random fields with horizontal correlation scales of 500, 1000, and 2000 km. These fields are evolved in time using an autoregressive process with lag 1 [AR(1)] on time scales of 6 h, 3 days, and 30 days, with fields standard deviations of 0.52, 0.18, and 0.06, respectively (Leutbecher et al. 2017).

The SKEB scheme (Berner et al. 2009) was developed for the ECMWF IFS model, which is a spectral model, and computes the backscatter of kinetic energy based on the dissipation rates from deep convection, numerical dissipation, and gravity mountain wave drag. This upscale transfer of energy is observed in the real atmosphere, albeit absent in traditional (deterministic) climate simulations. The kinetic energy lost in the model at the smallest scales, due to dissipation, is scattered upscale through perturbation of the streamfunction at the largest scales. The SKEB scheme streamfunction perturbations are modulated using the same stochastic spectral pattern as for SPPT, except that the perturbations vary in height as well as in space.

The relative contributions and total dissipation rates are shown in Berner et al. (2012) in their Fig. 1. The dominant contributor is deep convection in the tropics. Numerical dissipation is the second largest contributor. Berner et al. (2012) pointed out a weak resolution dependence of these calculations and suggested, based on the apparent underdispersion of the forecast ensembles, that there are additional sources of model uncertainty that are not captured by the scheme.

The analysis of initialized seasonal forecasts with the IFS model in Weisheimer et al. (2014) revealed that the SKEB scheme had almost no impact, while a positive impact was found for the SPPT scheme, in particular in regions where deep convection plays a major role, such as the western tropical Pacific. In the same study, SPPT also shows a positive impact on the Madden–Julian oscillation (MJO) amplitude distribution and ENSO forecast quality.

b. The unified model

The configuration of the global coupled model HadGEM3-GC3.1, submitted to the CMIP6 HighResMIP (Haarsma et al. 2016), is described in Roberts et al. (2019): it incorporates a global atmosphere–land configuration called GA/GL7.1 (Walters et al. 2019), forced by the same SSTs and sea ice used for the SPHINX experiments. The HighResMIP protocol was followed, which recommends the use of the MACv2-SP scheme (Stevens et al. 2017) for simplified and standardized aerosol forcing. This specifies the change of anthropogenic aerosol optical properties over time, enabling easier comparison between different models, while retaining the model's own aerosol mean background climatology and therefore requiring little or no additional tuning. It is used here in place of the prognostic GLOMAP-mode scheme (Mulcahy et al. 2018).

The nonhydrostatic atmospheric model uses a regular latitude–longitude grid, and has 85 levels extending to 85 km. The HadGEM3 simulation analyzed in this study has a $\Delta x \approx$ 60 km at midlatitudes (N216). Compared to the spectral IFS, this formulation brings with it substantial differences in

numerics, the effective resolution (see Klaver et al. 2020), and in particular the dissipation rates (convective and numerical dissipation only).

We use simulations with and without SP: the three ensemble members taken as control correspond to the official GA/GL7.1 scientific configuration, using SP, while three sensitivity experiments disable SP entirely or else use an individual SP scheme at a time. The GA/GL7.1 scientific configuration encodes SP schemes similar to EC-Earth, albeit following entirely independent implementations, including their parameter settings: stochastic perturbation of parameters (SPT) and SKEB-type [SKEB2; see Sanchez et al. (2016) for details]. Tennant et al. (2011) describe the SKEB2 system as developed for the Unified Model (UM; the parent of HadGEM3). Differences in the dissipation rates of the UM, compared to the IFS, are ascribed to differences in the convection schemes between the models. Another noteworthy difference between the schemes is that the SKEB2 scheme not only provides a forcing for the streamfunction (as in the IFS), but also acts as a forcing to the velocity potential [see Fig. 1 in Sanchez et al. (2016) for details], which represents an important difference from the SKEB scheme used in Climate-SPHINX.

The same Climate-SPHINX period was extracted from each for the purpose of TC tracking in this paper.

c. Tracking and TC identification

TCs are tracked and identified on an annual basis (January–December in the NH and July–June in the SH), using 6-hourly data from each individual simulation using the same methodology as in Hodges et al. (2017), there applied to reanalyses, and previously used in several other model-based studies of TCs (e.g., Manganello et al. 2012; Strachan et al. 2013; Roberts et al. 2015). Initially all cyclonic systems are tracked using the vertical average of relative vorticity in the layer 850–600 hPa, at a common resolution of T63, between 60°S and 60°N. This initial step, bringing all model fields to a common low-resolution truncation (here T63), has been used for the past 20 years in order to enable a fair comparison between models with different formulations and resolutions, and has been shown to reduce noise and to produce more coherent tracks and more complete TC life cycles, including their very early stages. Cyclonic disturbances (TC precursors) are initially identified as grid point maxima (NH) or minima (SH), exceeding a magnitude of $5 \times 10^{-6} \, s^{-1}$ at each time step (scaled by -1 in the SH), and the off-grid locations are then obtained using B-spline interpolation and steepest ascent maximization, resulting in smoother tracks. Tracks are initialized using a nearest neighbor method and are then refined by minimizing a cost function for track smoothness, subject to adaptive constraints for track smoothness and displacement distance within a time step. Following the tracking, tracks are filtered to retain those that last longer than 2 days. To allow warm core criteria to be applied, the T63 vorticity maxima (minima in the SH) are recursively added to the 2-day tracks at levels 850, 700, 600, 500, 400, 300, and 200 hPa, using the B-spline interpolation and maximization to assign a value if the maximum is within a 5° radius of the center at the previous level. Also added to the tracks are the full-resolution MSLP minima and 10-m

maximum wind speeds. For MSLP, B-splines and steepest descent are used using the 850-hPa vorticity center as the starting point for the minimization and assigning minima if they are within a 5° radius of the tracked vorticity center. For the 10-m winds a direct search for the maximum wind of the grid points within 6° radius of the tracked center is performed.

TCs are identified from among all the tracked cyclonic systems (TC precursors) using incremental criteria:

1) tracks must start within a latitudinal band (30°S–30°N) and last for at least 2 days (TC seeds);
2) the T63 relative vorticity at 850 hPa must attain a threshold of at least $6 \times 10^{-5}\,\mathrm{s}^{-1}$;
3) the difference in vorticity between 850 and 250 hPa (at T63 spectral truncation) must be greater than $6 \times 10^{-5}\,\mathrm{s}^{-1}$, to provide evidence of a warm core (via thermal wind balance);
4) the T63 vorticity center must exist at each level between 850 and 250 hPa for a coherent vertical structure; and
5) criteria 2–4 must be jointly attained for a minimum of four consecutive time steps (one day) and only apply over the oceans.

Note that the initial vorticity thresholds are very low by design [e.g., two orders of magnitude smaller than what is used in Hsieh et al. (2020)], as the TRACK approach to identification is in the use of the structural criteria described above. A common identification procedure based on a small set of criteria is also beneficial, in that it reduces the number of subjective choices to be made, such as resolution-dependent thresholds, and removes the identification as an uncertainty in comparing the results from different models. Some of these aspects, including comparison to other tracking methods, are covered in Roberts et al. (2020).

Following the tracking, spatial statistics including the track and genesis densities are computed using spherical kernel estimators (Hodges 1996).

The benefit of this approach is that the full life cycles of TCs are identified, including the precursor stage (e.g., easterly waves; Thorncroft and Hodges 2001; Serra et al. 2010; Yang et al. 2019), as well as other precursors (Fine et al. 2016) and post-TC stages (Sainsbury et al. 2020; Baker et al. 2021). After all TCs have been identified with the criteria above, TC precursors are further split, via an additional analysis stage, and only for the specific goal of the tables in section 3b, into tropical (30°S and 30°N) and extratropical (poleward of 30°S or 30°N). TRACK follows a Lagrangian approach that identifies each cyclonic feature explicitly, and is complementary to the work on TC seeds in Vecchi et al. (2019), who used instead bandpass-filtered variances, with a period of 3–10 days. A more recent paper by Hsieh et al. (2020), continuing previous studies of TC seeds, used explicit tracking, as in this study, albeit performed on the original model grid.

d. The genesis potential index and its terms

To complement the understanding of what controls TC formation, distribution, and lifetime, both in terms of climate means and variability, we have computed the genesis potential index (GPI; see e.g., Emanuel and Nolan 2004; Camargo and Wing 2016) as

$$\mathrm{GPI} = |10^5 \eta|^{3/2} \left(\frac{H}{50}\right)^3 \left(\frac{V_{\mathrm{pot}}}{70}\right)^3 (1 + 0.1 V_{\mathrm{shear}})^{-2},$$

where η is the absolute vorticity at 850 hPa, H is the 600-hPa relative humidity, V_{pot} is the potential intensity, and V_{shear} is the magnitude of 850–200-hPa wind shear. All GPI terms were computed by following the methodology of Bister and Emanuel (2002), as implemented in the Python script provided online.[1] All relevant variables, as monthly means for each ensemble member, were transferred from CINECA (Bologna, Italy), and climatologies, as well as basin time series, were built for each experiment. The same analysis was carried out by applying the algorithm to the ERA5 reanalysis, in order to provide an observational foundation.

3. Mean climatology of TC frequencies, distributions, and their response to the large-scale environment

a. Basin statistics

Figure 1 shows the number of tropical depressions, tropical storms (combined), and tropical cyclones, by category, identified by TRACK in each hemisphere. We have used a classification based on MSLP, as in Klotzbach et al. (2020), because of their strong arguments with regard to the study of the impact of TCs. This classification also enables us to identify TD+TS in IBTrACS, and at the same time it avoids the peculiarities of surface layer extrapolation of model level winds, a problem that has been particularly evident in recent intercomparisons (see, e.g., Roberts et al. 2020). Figure 1 shows the different SPHINX simulations (BASE with B suffix and STOC with S suffix), organized with increasing resolution (from left to right) and each SP experiment (S) inserted between each pair of resolutions (B). The height of the bars for each category suggests that SP is equivalent to an in-between model resolution when it comes to how many TCs are produced in each hemisphere. The results compare well to past studies, indicating that, as model resolution is increased, more TCs are produced in the climate simulation, with an initial steep increase for the weaker categories, up to T255, then tending to reach a plateau. As in previous GCM intercomparison studies (e.g., Roberts et al. 2020), a realistic simulation of the frequency of TC categories 4 and 5 is still elusive, even at the highest resolution (T1279), which nonetheless demonstrates a tendency for resolution and SP to both be beneficial. A recent intercomparison study by Judt et al. (2021), in which the PRIMAVERA-HighResMIP IFS model participated, albeit with convection explicitly represented, indicates that the full spectrum of TC intensities (both winds and central pressure) does become more realistic at 5-km resolution for that particular model configuration.

The climatology of the spatial distribution for the cyclonic disturbances counted in Fig. 1 confirms that these are indeed tropical cyclones, with spatial signatures that bear strong resemblance to observations. Figure 2 shows the track density

---

[1] See ftp://texmex.mit.edu/pub/emanuel/TCMAX/.

FIG. 2. TC track densities from the IBTrACS dataset for the Climate-SPHINX simulation period (1979–2008). Track density units are number per month per unit area, where the unit area is a 5° spherical cap.
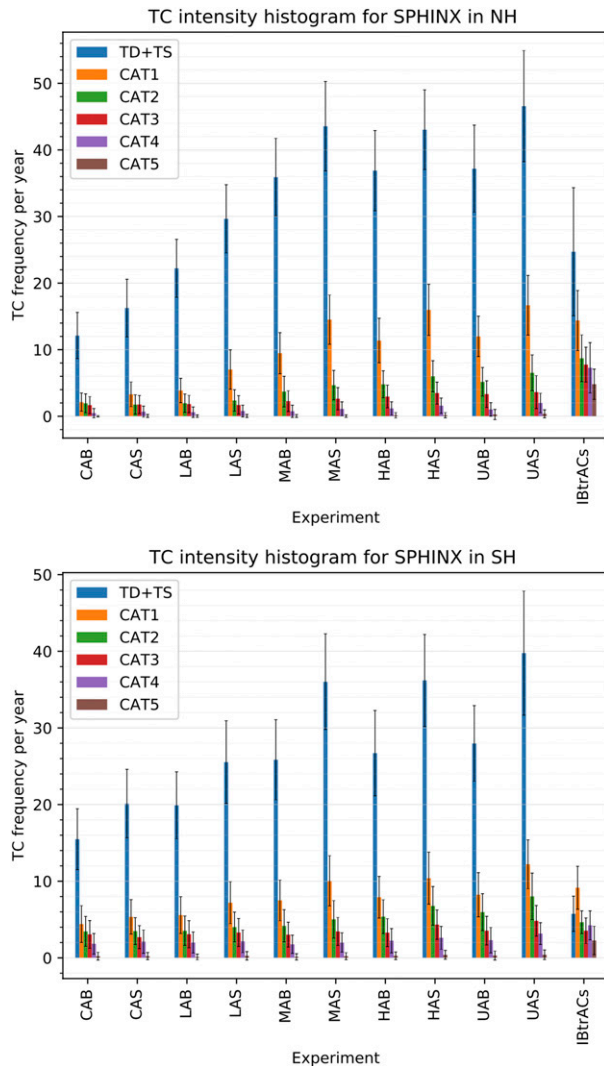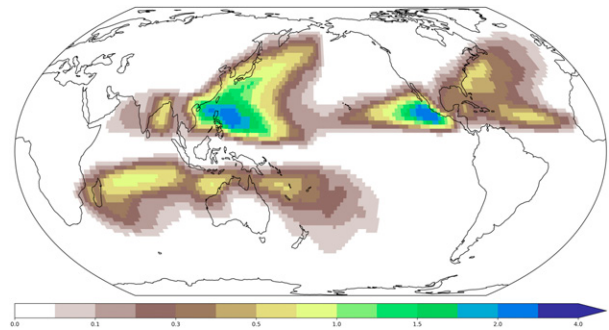
FIG. 1. The (top) NH and (bottom) SH TC counts per category, for the SPHINX period, 1979–2008, together with the corresponding IBTrACS observations (shown at the far right). Units are number per year. The vertical black lines indicate the ensemble spread.

climatologies from IBTrACS observations (Knapp et al. 2010), for the 1979–2008 period, and is provided as the reference against which all simulated TC density maps should be assessed. To verify this, Fig. 3 shows the (ensemble mean) track density climatologies simulated in Climate-SPHINX for each resolution (rows) and for the two experiments BASE and STOC (first and second columns). The information is presented in the form of TC track densities (color mesh) and supplemented in each basin by average storm counts per year [domains defined as in Hodges et al. (2017)]. The coarsest resolution in each experiment is shown at the top.

Tables 2 and 3 also provide some overall statistics that are useful to interpret the information in the track density maps (Fig. 3), as well as the ensemble spread and the systematic

differences between resolutions, which are significant, given the small ensemble spread.

The comparison of Figs. 2 and 3 demonstrates two things: first, similar to previous studies (Zhao et al. 2009; Strachan et al. 2013; Roberts et al. 2015), enhancing resolution increases the number of simulated TCs, which is reflected in the enhanced track density and better agreement with observations. The results in Fig. 3 are quite unique in terms of their robustness, afforded by the large ensemble and the five different resolutions. The resolution effect is apparent in each of the major TC basins and is consistent in both sets of simulations, BASE and STOC, as each column is inspected individually. For the southwest Pacific domain, it is obvious that all models at moderate to high resolution, independent of the use of SP, overestimate the track densities and counts, as compared to observations. This is a common feature in GCMs, particularly for AGCMs, but also partially reflects observational uncertainty, due to how TCs are counted in different basins, as discussed in Hodges et al. (2017). Similar errors occur over the north Indian and South Atlantic domains, and have been amply discussed in previous publications (e.g., Hodges et al. 2017; Roberts et al. 2020).

Additionally, and similar to the resolution effect, SP enhances the number of simulated TCs, which is revealed by comparing the first and second columns, as well as their differences in the third column. A striking resemblance between the STOC plot in each row and its companion BASE plots in the left column (same row and one row down) supports a strong and systematic finding for TC simulation: the use of SP is equivalent to an increase in resolution, albeit not quite equivalent to a full doubling of resolution. By inspection of the panels in Fig. 3, the model response is spatially coherent across resolutions, with the track densities tending to increase rather than shift, with few local exceptions. The individual ensemble members are also coherent with each other (not shown). Further, the differences between STOC and BASE are shown to be robust in the TC-active regions, as indicated by the Welch test (Wilk 2011).

The SP effect is most pronounced and consistent in the northwest Pacific, and presents a local maximum in the southwest Pacific, to the east of Australia. In other basins (e.g., the North Atlantic, where most GCMs still struggle to simulate a
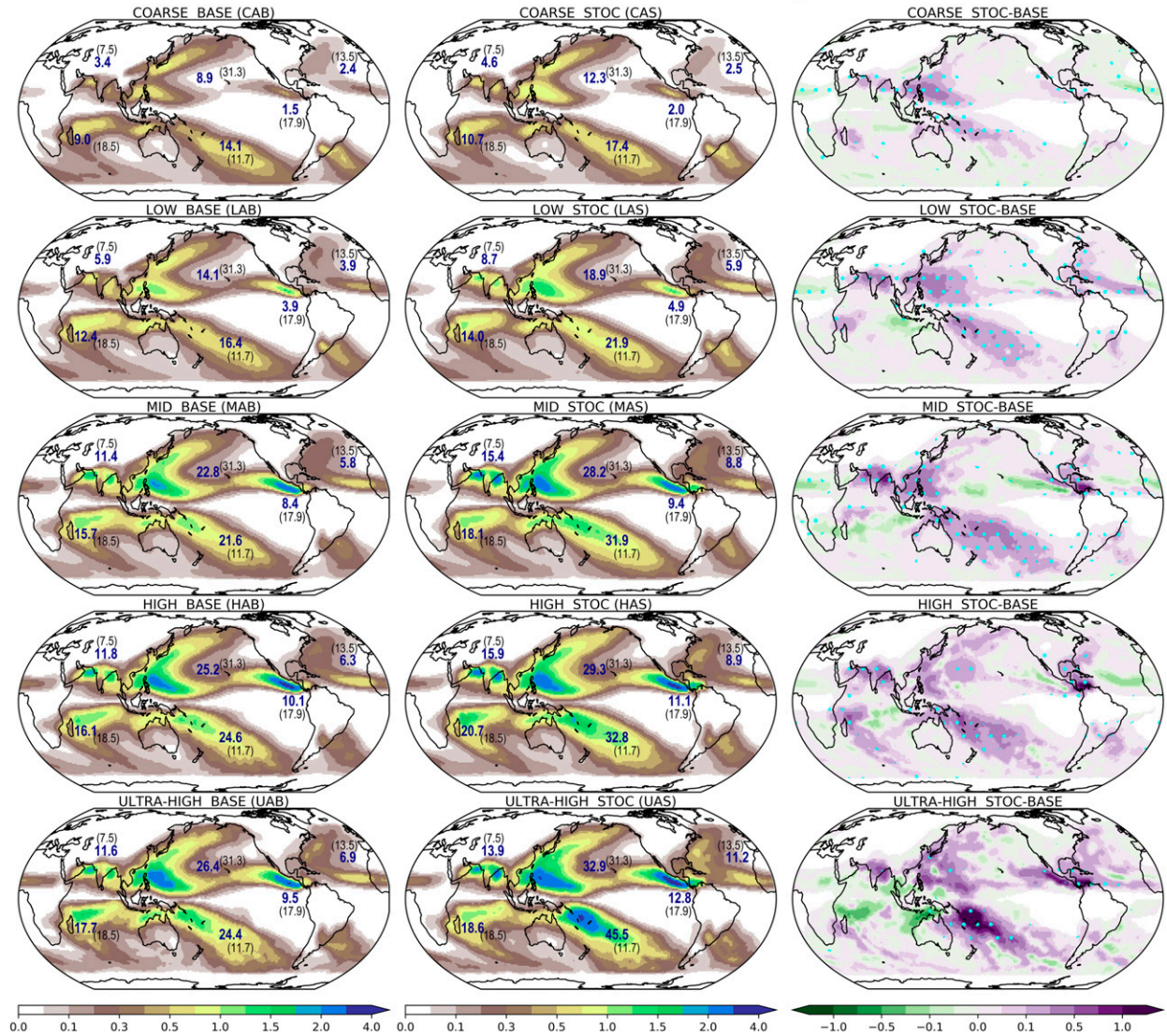
FIG. 3. TC track densities for the EC-Earth (IFS) model, Climate-SPHINX campaign. Blue numbers in each panel indicate the number of storm transit per months for each sector, while black numbers in the top row are the IBTrACS observations. Stippling indicates the significance at the 5% level; track density units are number per month per unit area, where the unit area is a 5° spherical cap.

realistic number of TCs), the response to SP is not consistent at all resolutions, with a decrease at COARSE, then an increase at LOW, and a dipole response, reflecting differences in the eastern versus western North Atlantic, for MEDIUM, HIGH, and ULTRA-HIGH resolutions. In terms of significance, there is an interplay between information provided by the large ensemble sizes at COARSE resolution and the stronger signal at HIGH and ULTRA-HIGH resolutions. In other words, taking the

TABLE 2. All NH TC precursors, TC seeds, TCs, number per year.

TRUNC	All TC precursors			TC seeds			TCs		
	BASE	STOC	% DIFF	BASE	STOC	% DIFF	BASE	STOC	% DIFF
T_L159	9283 ± 158	$11\,800 \pm 185$	27	2895 ± 57	3603 ± 69	24	18 ± 4	24 ± 5	30
T_L255	9958 ± 165	$12\,291 \pm 189$	23	3201 ± 57	3837 ± 68	20	31 ± 5	42 ± 6	36
T_L511	$10\,111 \pm 191$	$13\,053 \pm 182$	29	3375 ± 66	4120 ± 64	22	52 ± 7	67 ± 8	28
T_L799	$10\,145 \pm 158$	$12\,791 \pm 176$	26	3399 ± 59	4079 ± 76	20	57 ± 7	70 ± 6	22
T_L1279	$10\,186 \pm 223$	$12\,809 \pm 212$	26	3419 ± 62	4062 ± 62	19	59 ± 7	76 ± 8	29

TABLE 3. All SH TC precursors, TC seeds, TCs, number per year.

TRUNC	All TC precursors			TC seeds			TCs		
	BASE	STOC	% DIFF	BASE	STOC	% DIFF	BASE	STOC	% DIFF
T_L159	9509 ± 168	$12\,045 \pm 178$	27	3016 ± 53	3720 ± 58	23	28 ± 5	34 ± 5	19
T_L255	$10\,334 \pm 171$	$12\,801 \pm 198$	24	3374 ± 50	4041 ± 63	20	34 ± 6	43 ± 6	24
T_L511	$10\,671 \pm 174$	$13\,502 \pm 193$	27	3611 ± 54	4311 ± 65	19	42 ± 6	57 ± 7	34
T_L799	$10\,716 \pm 193$	$13\,295 \pm 204$	24	3644 ± 63	4277 ± 60	17	46 ± 7	61 ± 7	33
T_L1279	$10\,693 \pm 178$	$13\,193 \pm 205$	23	3665 ± 53	4290 ± 77	17	48 ± 6	68 ± 9	42

ensemble mean of the larger ensembles at COARSE to MEDIUM resolution plays less of a role than the stronger signal found in HIGH and ULTRA-HIGH, which is further supported by the larger numbers seen in Tables 2 and 3.

An additional question arises as to whether increased track density is due to changes in the TC to area ratio (e.g., larger number of TCs in the same area) or to unchanged numbers, albeit with longer-lasting tracks. Examining the genesis density statistics in Fig. 4, as well as the information on interannual and seasonal variability shown in later sections, it is clear that the use of SP enhances genesis by up to 30%, thus reflecting the first mechanism. However, the genesis density change is not uniform in all basins: the northwest Pacific and north Indian basins show increases that are consistent across resolutions and rather widespread. The eastern North Atlantic also shows a consistent increase at the first three resolutions, at the location where easterly waves encounter the ocean (Thorncroft and Hodges 2001), as well as the western North Atlantic region just north of Colombia, which is important for Gulf TC tracks and east Pacific TCs (Serra et al. 2010). Other locations, such as the eastern Pacific, see a reduction in genesis at MEDIUM and HIGH resolution, which could be a response to the representation of the mountains in Central America, which have increased height and complexity at high resolution, affecting the propagation of atmospheric waves from the Caribbean.

b. Statistic of transition from precursors to TC seeds and to TCs

We now turn to addressing the two complementary hypotheses on the role of seeding and the role of environments favorable for cyclogenesis. Table 2 shows the TC precursors, TC seeds, and full warm-core TC counts for the Northern Hemisphere (including ensemble spreads) and their differences, as percentages; Table 3 shows the counts for the Southern Hemisphere. In both hemispheres we start with around 10 000 TC precursors per year [an equivalent number of extratropical cyclone (ETC) precursors are found poleward of 30°]; next, around 3000 TC seeds (2-day systems) are found, finally ending up with less than 100 TCs, in broad agreement with observations. The ensemble spread is very small at each resolution, around 2%, and independent of the use of SP.

Both tables show that TC precursors and seeds increase systematically with resolution in both BASE and STOC, but the increase from BASE to STOC is larger than the increase caused by enhancing resolution (even across all five resolutions); this is true in both hemispheres. Further, the relative increase in TCs caused by the use of SP is seen both in the

number of precursors, compatible with the first hypothesis, as well as in the last transition, to full warm-core TCs, supporting the second hypothesis.

Comparing the numbers in terms of percentages, the SP-induced increase in precursor generation is between 23% and 29% (23%–27% in the SH), with lower-resolution simulations showing a larger percentage increase. However, the last column of data shows that the extra transition into full TCs is in the range of 22%–36% (19%–42% in the SH). The impact of SP on the final transition to warm-core TC further increases the initial probability of generating a TC from precursors, and is less systematically dependent on resolution, although it appears to increase overall in the SH. This spatially dependent resolution response (hemispheric here, and basin-dependent in later sections) is counterintuitive, as the (dominant) SPPT scheme perturbations are the same at all resolutions, and this requires careful interpretation (see section 6).

TRANSITION FROM PRECURSORS TO 2-DAY CYCLONES IN THE EXTRATROPICS

Given the fact that SP has been shown in Watson et al. (2017) to preferentially impact the tropics in terms of moisture availability, and given previous studies on the effects of SP on ETCs, it is interesting to contrast TC and ETC transition rates in the context of a single study. The statistics for ETCs, shown in Tables 4 and 5, indicate that we start from a number of precursors almost identical to what has been found in the tropical region, still accompanied by small ensemble spreads. However, for the ETC precursors, differences between 10% (12% in SH) for COARSE and 7% (5% in the SH) for ULTRA-HIGH were found in terms of sensitivity to SP; for the (2-day) ETCs, the differences are even smaller (1%–4%), indicating that SP plays some role in seeding, albeit only a limited role in conditioning the ETC environment, as ETC dynamics is different from TC dynamics, and less dependent on convection.

c. Probability of cyclogenesis and its environmental control

The probability of cyclogenesis [PC; similar to survival rate (SR) in Yamada et al. (2021, their Eq. (1)] is the local ratio between the frequency of TCs and that of TC seeds: PC = TC/TC_{seed}.

We computed the spatial distribution of PC by taking the ratio of track density for TCs and TC seeds at each grid point. All PCs are computed separately for each of the 60 simulations and then combined as ensemble and multiyear averages, in order to illustrate the regional impacts of SP (columns) and resolution (rows).

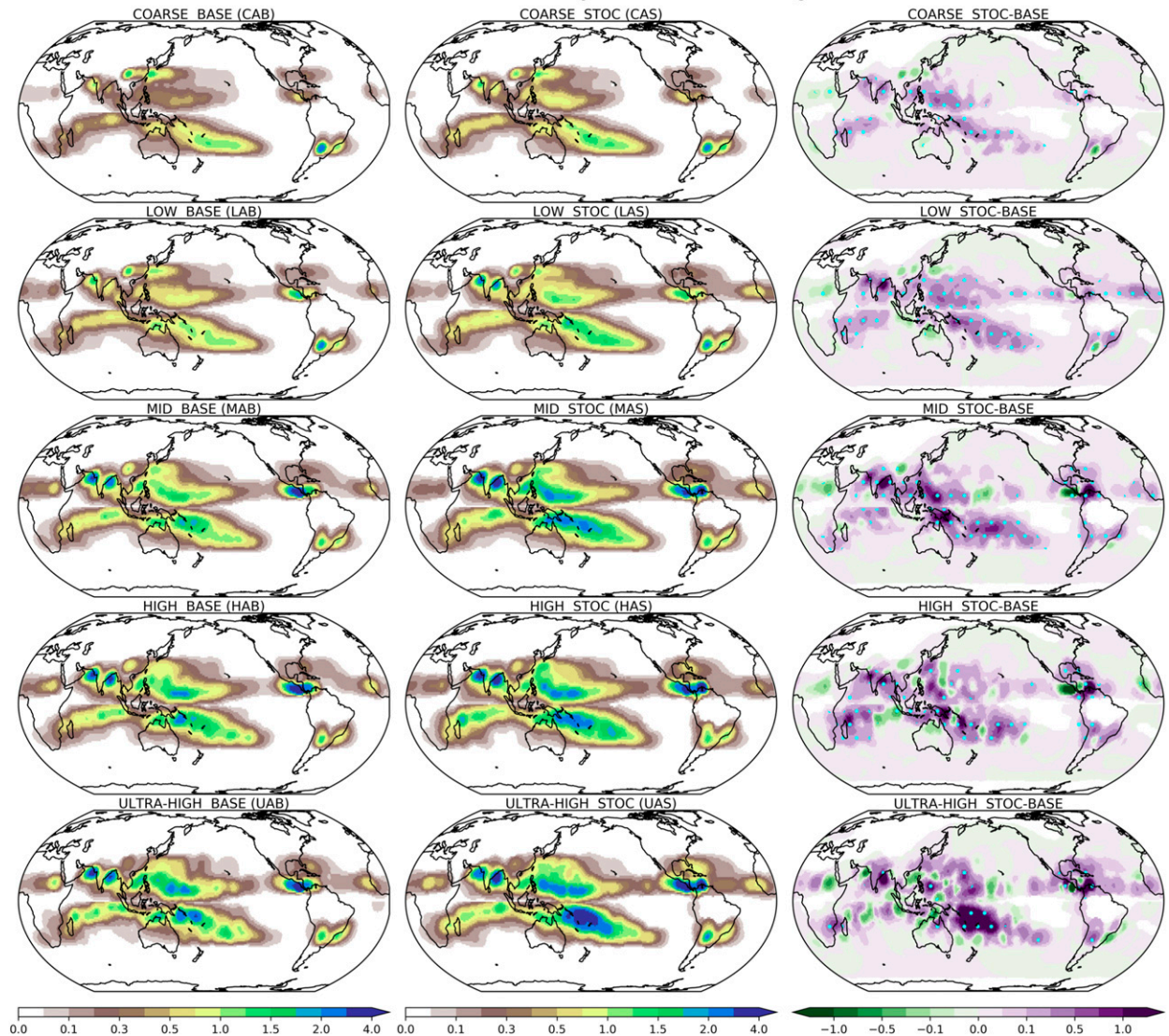PRESENT Genesis Density: annual climatologies

FIG. 4. TC genesis densities for the EC-Earth (IFS) model, Climate-SPHINX campaign. Stippling indicates the significance at the 5% level. Genesis density units are number per month per unit area, where the unit area is a 5° spherical cap.

Figure 5 shows that PC for track density is as high as 30% in large regions of TC activity: the highest values are found in the northwest Pacific and the largest changes correspond to the change in resolution, with SP showing a similar (albeit smaller) response. A few cases of reduction in PC when using SP are found, as large as −10%, as seen for instance in the eastern Pacific. These results are comparable with those in studies considering developing versus nondeveloping tropical synoptic disturbances identified in reanalyses in both the Atlantic and northwest Pacific (Fu et al. 2012; Peng et al. 2012; Brammer and Thorncroft 2015; Hankes et al. 2015), where PC can be as high as 30% or more.

Figure 5 also shows the GPI, computed from the Bister–Emanuel algorithm (Bister and Emanuel 2002) as shaded colors. The GPI fields shown here indicate that the mean TC

environment responds less to the introduction of SP, or increase of resolution, than the PC itself. The figure also indicates that GPI peaks at the point of entry of the most active TC regions [e.g., in the North Atlantic main development region (MDR)]. The subtler aspects of the response of GPI to SP and resolution will be explored in a later section, particularly for variability.

d. Location and relative magnitude of seeding and cyclogenesis

To better understand the SP response of TCs seen so far, ΔTC, we break it into two terms at each resolution: (i) extra formation of TCs due to an increase in the number of TC seeds, versus (ii) extra transition from TC seed to TCs, due to cyclogenesis. These terms are directly related to our two hypotheses and we call them the seeding term and the cyclogenesis term:

TABLE 4. All NH ETC precursors and ETCs, number per year.

TRUNC	All ETC precursors			ETCs		
	BASE	STOC	% DIFF	BASE	STOC	% DIFF
T_L159	8909 ± 101	9825 ± 98	10	2664 ± 37	2746 ± 38	3
T_L255	9947 ± 103	$10\,835 \pm 112$	9	2818 ± 40	2904 ± 39	3
T_L511	$10\,608 \pm 109$	$11\,559 \pm 116$	9	2909 ± 42	3011 ± 41	4
T_L799	$10\,760 \pm 110$	$11\,452 \pm 105$	6	2940 ± 42	3017 ± 39	3
T_L1279	$10\,786 \pm 102$	$11\,558 \pm 136$	7	2939 ± 47	3035 ± 39	3

$$\Delta TC = TC \times \frac{\Delta TC_{seed}}{TC_{seed}} + TC \times \frac{\Delta PC}{PC}.$$

Ideally, the two rightmost terms should add to the leftmost exactly, but in practice this is an approximation, as the seeding term will tend to affect the genesis stage of TCs, while the cyclogenesis term applies to their entire lifetime, so that the contributions from the two terms cannot be expected to be exactly collocated, even for relatively large regions, as resolved by the COARSE simulations. For the large sample size used in this paper, however, the approximation holds well (also verified by plotting the difference between left-hand side and the sum of the right-hand side terms; not shown). Figure 6 shows the ΔTC in response to SP at each resolution for the ensemble means and the breakdown into the contribution of TC seeding term versus the contribution of the cyclogenesis term. The seeding response to SP is mostly positive and largest in areas where TCs spend most of their lifetime, as well as slightly sensitive to resolution. The cyclogenesis term has a stronger response to SP, and it can result in either an increase or decrease of TCs in each region, most notably with a decrease in the eastern Pacific and over most of the Indian Ocean. The cyclogenesis term also responds to resolution more vigorously, and in an apparently linear way.

It can be said in summary that SP acts to enhance TC seeding overall, and this effect is slightly sensitive to resolution; however, cyclogenesis is more important for transition, and more sensitive to both resolution and SP, able to cause both an increase and a decrease of TCs simulated in a region. This justifies turning our attention to the TC environment simulated by the SPHINX models.

e. GPI terms and their control on the climatology of TCs in each basin

To understand systematically the location and magnitude of changes due to resolution or SP, a number of difference fields

for TCs and their environment are computed. We have organized the plots so that the leftmost column in each of the four panels of Fig. 7 shows the response to the introduction of SP at each resolution (STOC − BASE), while the rightmost column shows the response to one increment of resolution (e.g., LAB − CAB, MAB − LAB, and so on), while retaining the BASE model formulation (no SP is used).

The ΔPC values are plotted as contours overlaid onto changes in environmental conditions as used in the calculation of GPI (shear, relative humidity, vorticity, potential intensity), shown in Fig. 7. The convention is that signs for GPI terms are presented exactly as shown by the GPI equation, so that, for instance, a large GPI shear term corresponds to a small magnitude of vertical wind shear and a positive change in the GPI shear term corresponds to a decrease in vertical wind shear. For potential intensity, we have also computed all terms, in particular the thermodynamic efficiency. We plotted GPI and the three most significant GPI terms in Fig. 7: for GPI itself, SP acts to reduce PC in the eastern Pacific and around the Maritime Continent, while acting to increase PC in the North Atlantic MDR, in the SPCZ, in the Indian Ocean, and in the northwest Pacific. Increasing the model resolution acts to increase PC in the eastern Pacific, as well as in all main TC activity regions. At LOW to MEDIUM resolution, the spatial patterns for the regions where PC increases are very similar to each other.

Considering the GPI terms individually, SP acts to moisten the environment in the western part of each basin, mostly where PC increases, while drying the environment in the eastern Pacific, where PC decreases (e.g., for MEDIUM and HIGH). Model resolution appears to cause more of a drying effect at LOW to MEDIUM resolution, particularly around the Maritime Continent, with the exception of the Bay of Bengal and the northwest Pacific at MEDIUM resolution and the eastern Pacific at LOW and MEDIUM resolution.

The use of SP causes a stronger shear term, which is clear in the eastern Pacific and to the east of the Maritime Continent,

TABLE 5. All SH ETC precursors and ETCs, number per year.

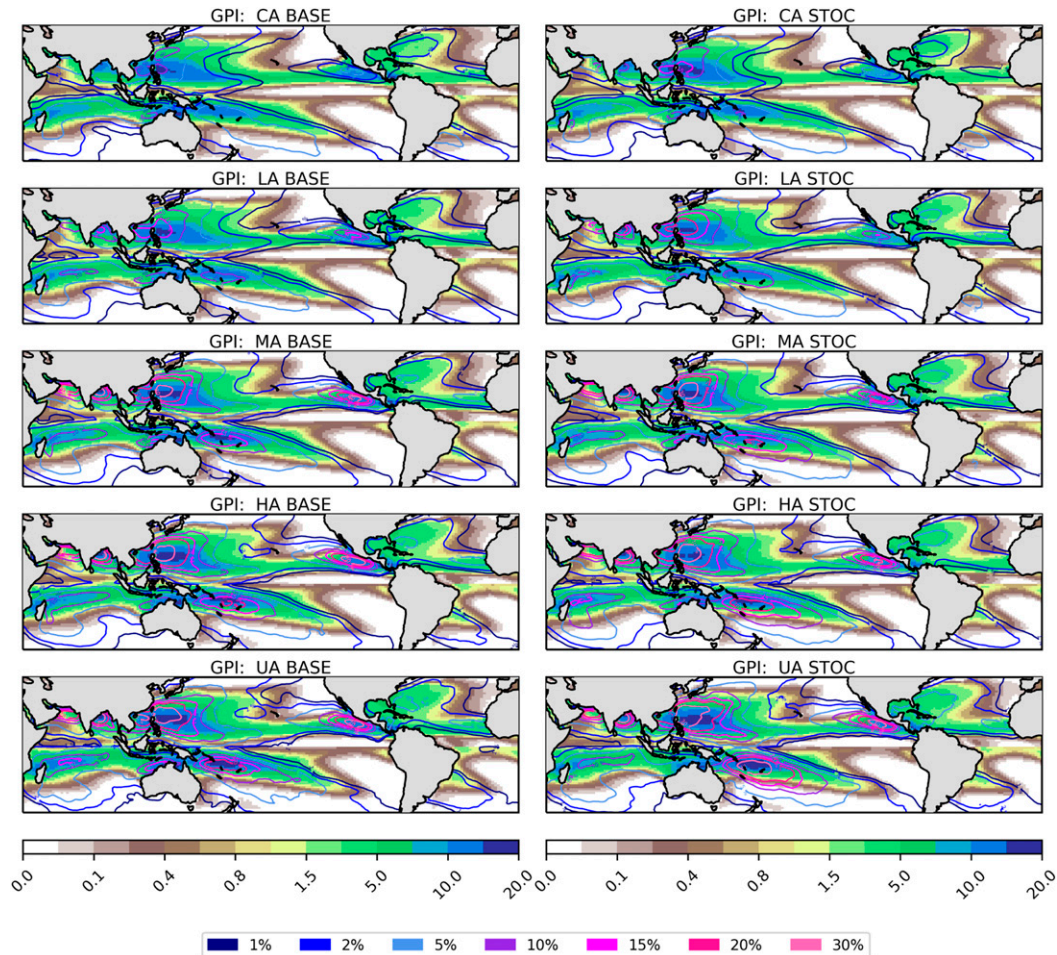| TRUNC	All pre-ETC features			ETCs		
	BASE	STOC	% DIFF	BASE	STOC	% DIFF
T_L159	7282 ± 112	8168 ± 125	12	2167 ± 33	2236 ± 37	3
T_L255	8515 ± 114	9313 ± 133	9	2365 ± 34	2413 ± 38	2
T_L511	9573 ± 124	$10\,259 \pm 127$	7	2515 ± 40	2559 ± 38	2
T_L799	9807 ± 128	$10\,329 \pm 134$	5	2564 ± 45	2586 ± 36	1
T_L1279	9940 ± 147	$10\,396 \pm 109$	5	2573 ± 30	2603 ± 47	1

FIG. 5. The SPHINX models GPI (shaded color mesh) and the probability of cyclogenesis (PC %) from TC seeds to TCs (contours), as a function of resolution (rows, with COARSE at top and ULTRA-HIGH at bottom) and SP treatment (columns: BASE on the left and STOC on the right). Contour levels for PC are shown in the lower box.

corresponding to reduced PC, and, to a lesser extent, a weakened shear in the northwest Pacific, increasing PC. The resolution enhancement reduces shear overall, and this is beneficial to PC, particularly in the eastern Pacific.

The potential intensity maps show that the use of SP tends to decrease the simulated PI, while an increase in resolution tends to increase PI. The effect is rather widespread and, while located predominantly in the tropics, it not show any particular spatial correlation with the regions where PC changes.

Emanuel et al. (2013) identify three thermodynamic variables useful to understand the formation of TCs: 1) the presence of moist convection, 2) midtropospheric humidity, and 3) potential intensity (PI). We have focused so far on the last two variables, part of GPI, and their role in the conversion of TC seeds; our findings are supportive of the conclusions in Vecchi et al. (2019) and Yamada et al. (2021). The role of moist convection may not be revealed by GPI, but can be investigated by the examination of the OLR fields in the EC-Earth model. Results from Davini et al. (2017) (their Table 2), indicate that top-of-the-atmosphere emitted longwave radiation is 238.71 W m^{-2} for COARSE

BASE and 241.74 W m^{-2} for ULTRA-HIGH STOC, with a pattern of nearly linear increase. Examination of the OLR field (not shown) reveals that both resolution and SP act, instead, to reduce radiative loss at the locations where TCs are present, and particularly so where PC rates are high. The overall model behavior is thus consistent with deep convection being present at the locations of TC activity, and strongest at the locations of maximum PC. The tropical impact of SP is in fact consistent with the broad moistening of the tropical atmosphere shown by Watson et al. (2017) for the same model. These two aspects of large-scale versus local-scale response point to a potential feedback between TCs and their environment, which may be enhanced or hindered by SP, such as via the multiplicative nature of the SPPT scheme.

4. Interannual to seasonal variability and environmental controls on TCs

Linking the interannual variability (IV) of TCs to the simulated environment in each experiment can be used to understand what the main drivers of the response to resolution and SP are, as well as
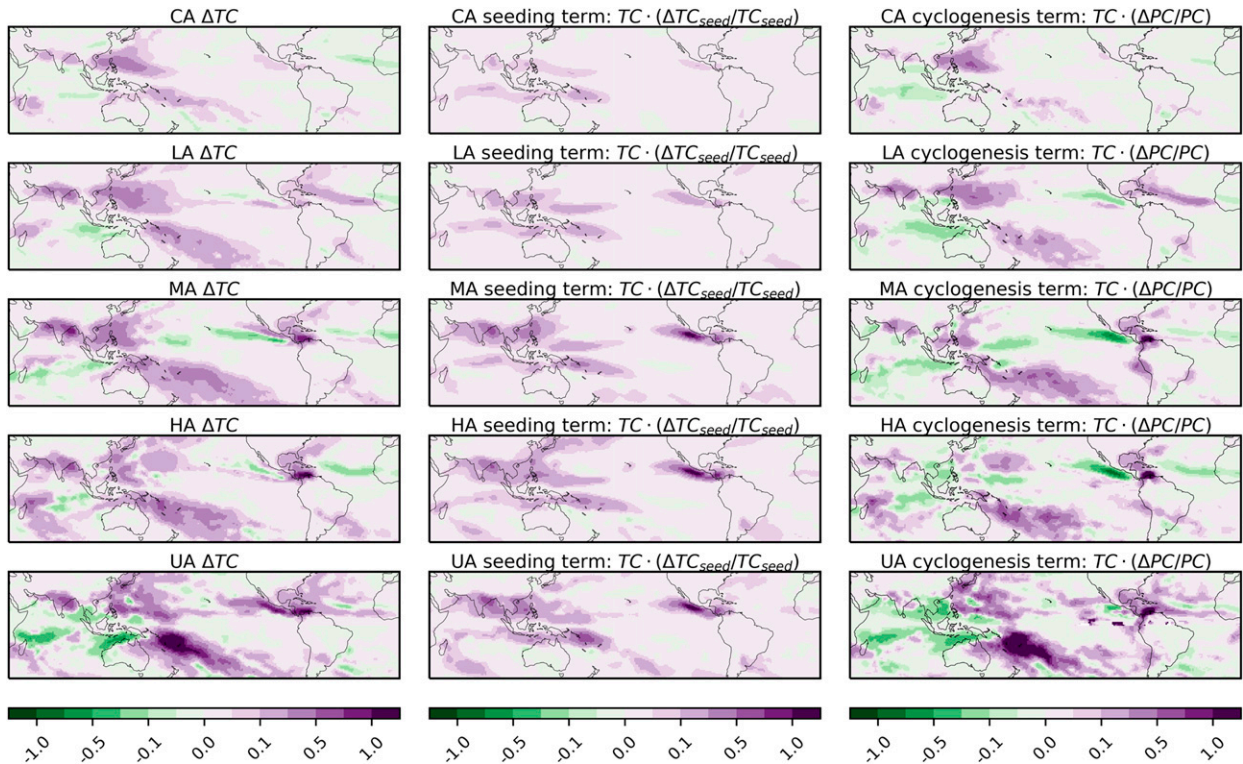
FIG. 6. The relative roles of seeding and cyclogenesis in creating a TC response to SP at each resolution: (left) total count, (center) role of seeding, and (right) role of cyclogenesis. Units are as in Fig. 3.

uncovering any improvements in predictability, defined as skill in predicting year-to-year changes in TC frequency. The other reason for focusing on IV is that predictive skill at this level provides an important degree of trust in a model's dynamics and physics, as it is harder to tune for variability than for climate means, and IV provides a stricter test for a model to be deemed "fit for purpose" in terms of climate change applications [see discussion in Vidale et al. (2003)].

In this section, statistics for the interannual variability of TC counts for the entire period are presented by basin, for all simulations. First, Table 6 demonstrates how the use of SP increases interannual variability across resolutions and basins, with few exceptions, mostly representing an improvement, with the notable exception of the SWPAC, where IV is overestimated at truncations of T255 (LAB) and above. This is an important result, as GCMs, particularly at low resolution (as seen in COARSE BASE), tend to underestimate variability, and it is not clear a priori how an ensemble of AGCMs, all driven by the same observed SSTs and sea ice, might react to the addition of SP. It remains to be seen whether this additional variability originates from signals contained in the SST/sea ice and radiative (aerosols) forcings, potentially a sign of climate predictability, or whether it is just internal variability (noise).

It is of particular interest and value to the community to improve the prediction of interannual changes in TC frequencies by basin, based on previous evidence of potential predictability (see Zhao et al. 2009; Roberts et al. 2015, 2020). To this end, we focus on four traditional basins: the North Atlantic (Fig. 8a), the northwest Pacific (Fig. 8b), the south Indian Ocean (Fig. 8c), and the southwest Pacific

(Fig. 8d), showing the time series of TC counts at each resolution, as well as correlations with observed TC statistics (IBTrACS).

The counts and variability at COARSE resolution are both underestimated but increase as we move to the bottom of each panel, toward ULTRA-HIGH. In the North Atlantic the correlation between the IBTrACS counts and the number of TCs identified in the Climate-SPHINX models increases from 0.23 to 0.35 (COARSE to HIGH), but is then reduced to 0.19 at ULTRA-HIGH. The correlations in STOC are mostly lower, except for LOW resolution. It is particularly noticeable how all models entirely miss the exceptional TC season of 2005, but are able to capture some La Niña responses (e.g., in 1995). For the northwest Pacific, correlations are once again modest, and slightly higher as we increase the resolution (range from 0.15 to 0.28); the addition of SP in STOC would seem beneficial in COARSE, LOW, and MEDIUM. There is moderate skill in the south Indian Ocean, with correlations from 0.25 to 0.37, with some improvement from COARSE to HIGH (also reflected in the means). SP improves the IV correlations at COARSE to MEDIUM, but this decays with HIGH and ULTRA-HIGH. BASE and STOC also appear to be distinctly different at HIGH resolution in the last 10 years of simulation, also based on the small ensemble spread. For the southwest Pacific, it is clear that there is a significant overestimate of the number of TCs, which had also been observed in reanalyses (Hodges et al. 2017), and made worse by SP; the representation of IV is similar to that seen in the North Atlantic in the range COARSE to
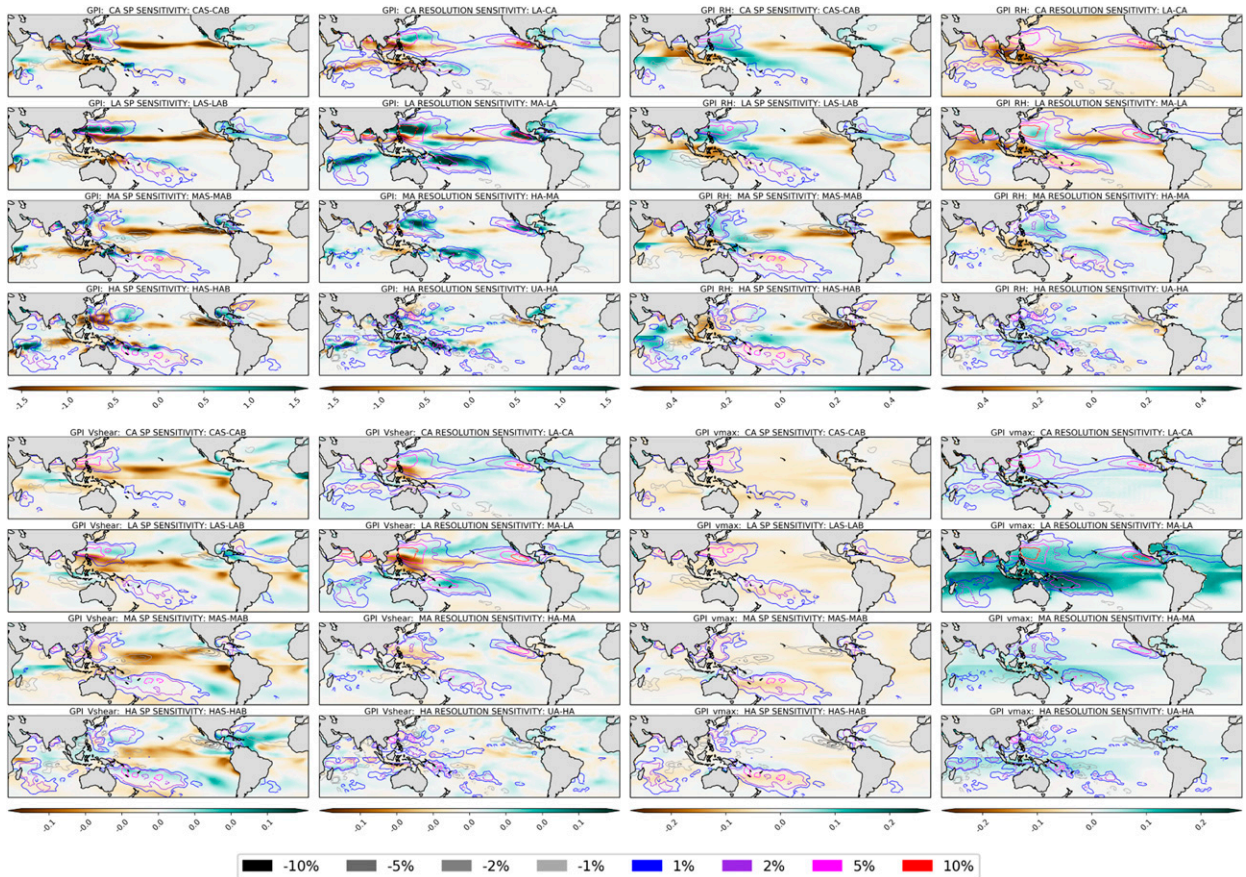
FIG. 7. The environment governing TC seed transition in the NH: (top left) GPI and its components (top right) RH, (bottom left) vertical wind shear, and (bottom right) potential intensity. All are scaled according to the GPI equation. Superposed are contours of the change in the probability of cyclogenesis (ΔPC). Contour levels are shown in the lower box.

MEDIUM, but decaying after that. SP seems to make no significant improvement, or even to give a negative correlation at HIGH resolution.

Overall, the interannual variability skill is poor when compared to other models (see section 6); moreover, by using a two-tailed Kolmogorov–Smirnov test ($p = 0.05$) we cannot

state that any of the differences encountered, in any of the basins, are significant, despite the sizable ensemble.

a. Seasonal variability

Figure 9 shows the mean annual cycle for the North Atlantic (top) and northwest Pacific basins (bottom), defined as in

TABLE 6. TC variability (standard deviation of annual means) in selected basins: NATL = North Atlantic, NWPAC = northwest Pacific; EPAC = eastern Pacific; NIND = northern Indian Ocean; SIND = southern Indian Ocean, and SWPAC = southwest Pacific. Units are TCs per year.

| Truncation/experiment	NATL	NWPAC	EPAC	NIND	SIND	SWPAC
T_L159 CAB	1.60	3.15	1.21	1.61	2.85	4.11
CAS	1.63	3.90	1.30	1.78	2.87	4.24
T_L255 LAB	1.88	3.99	2.01	2.07	3.43	4.59
LAS	2.51	5.13	2.29	2.48	3.82	5.15
T_L511 MAB	2.29	5.25	2.88	2.82	4.12	5.89
MAS	2.79	5.44	2.81	3.44	4.10	5.57
T_L799 HAB	2.61	5.42	3.18	3.06	4.29	6.33
HAS	3.00	5.59	2.81	3.27	4.47	5.55
T_L1279 UAB	2.46	6.17	3.30	2.72	5.77	6.86
UAS	3.72	6.00	3.78	3.75	3.89	7.06
OBS	5.41	8.11	4.81	3.41	4.29	4.10
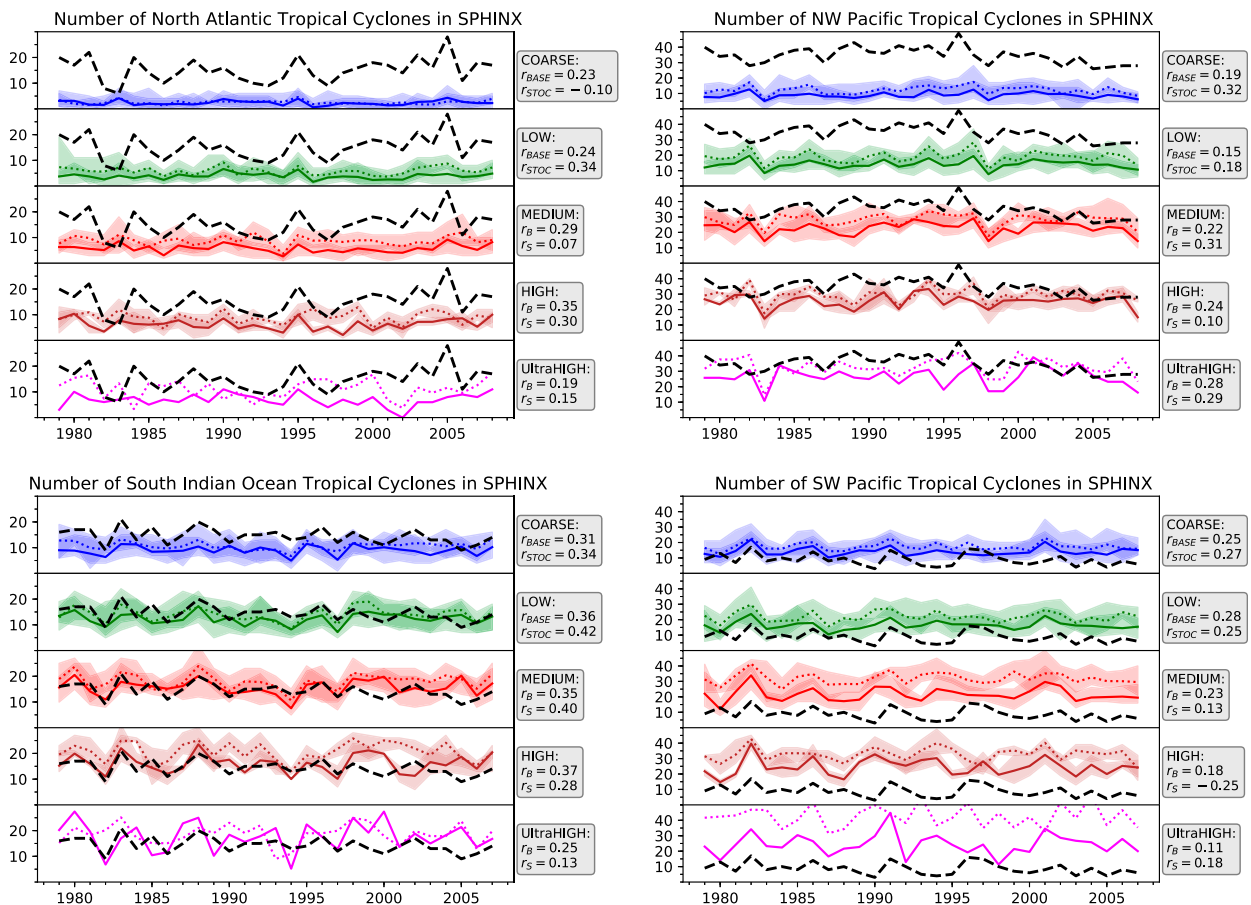
FIG. 8. TC interannual variability for the (top left) North Atlantic, (top right) northwest Pacific, (bottom left) south Indian Ocean, and (bottom right) southwest Pacific for each SPHINX simulation. BASE experiments are shown as continuous lines, while STOC experiments are shown by the dotted lines. The ensemble spread, defined by adding shading the area between ensemble minimum and ensemble maximum for each year, is shown by the shaded areas. IBTrACS observations are shown by the thick dashed lines. The yearly Pearson correlation between ensemble mean and IBTrACS is reported for both BASE (r_B) and STOC (r_S) to the right of each panel.

Hodges et al. (2017), for each of the SPHINX experiments, averaged over the respective ensemble (see Table 1 for details). For comparison, the IBTrACS mean annual cycle for the 1979–2008 period is added on the right-hand side.

Figure 9 shows that, for the North Atlantic, the number of TCs simulated in each month increases with resolution, as well as with the use of SP, resembling more and more the IBTrACS observations; further, for MEDIUM, HIGH, and ULTRA-HIGH, the seasonal peak is reached in August without SP, but in September using SP, providing a better match with observations. The start and end of the TC seasons are less sensitive to the use of SP. For the northwest Pacific, all experiments show a delay in the peak of the season, and a too late end, but this error is mitigated by the use of the highest resolutions, combined with SP.

A separate analysis (not shown) of North Atlantic basin 850–600-hPa humidity and 200–850-hPa vertical wind shear shows that, in the August–September transition, both are slightly more favorable in this basin, but no coherent spatial pattern (e.g., associated with the positioning or magnitude of

the steering anticyclone) is found. The nature of this seasonal response appears thus to be associated with individual TCs.

b. GPI terms and the modulation of the annual number of TCs in each basin

We extended our GPI analysis to its temporal behavior, considering monthly means of GPI terms in each basin, using, wherever possible, the same domain definitions as in Wing et al. (2015), albeit altering the seasons to July–November for the Northern Hemisphere, and November–March for the Southern Hemisphere, for two reasons: 1) these new definitions of season provide significantly stronger correlations, when compared to those shown in Wing et al. (2015); and 2) a longer, more homogeneous seasonality, also closer to operational criteria, makes it possible to retain the same definitions for the analysis of climate change experiments.

All GPI terms were computed for each ensemble member at each grid point, and the results are shown as basin and ensemble means, together with an envelope that shows the ensemble spread. The figures contain both the seasonal evolution and the annual means for the particular months considered in

PRESENT, North Atlantic
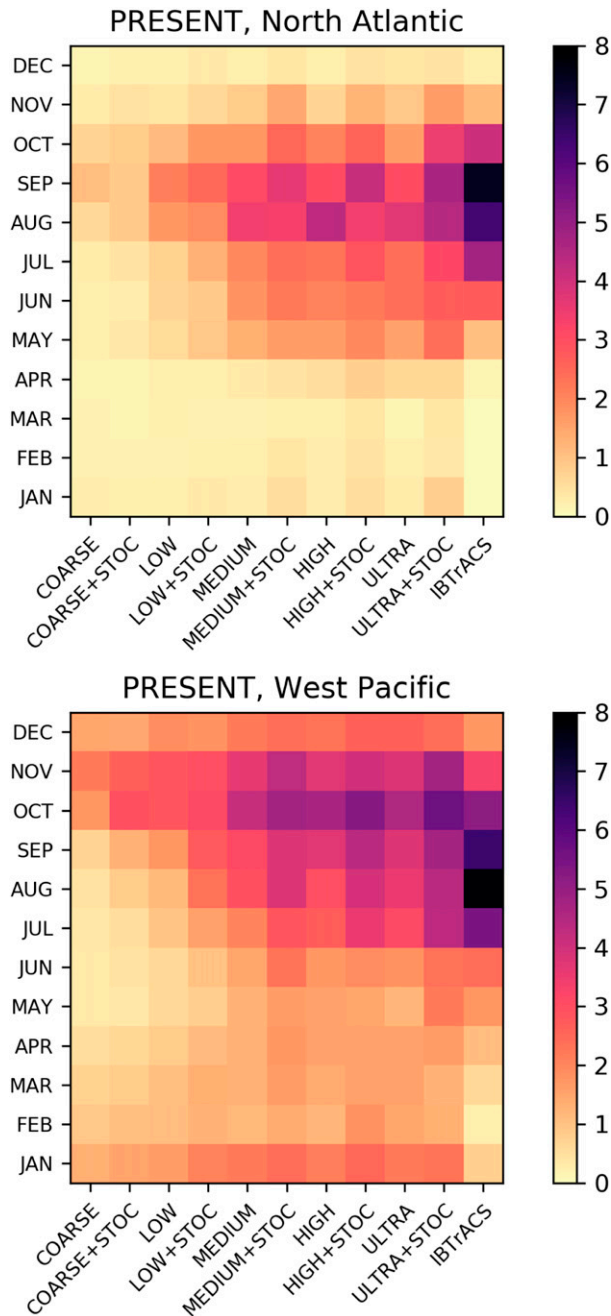

PRESENT, West Pacific

FIG. 9. Area averaged TC seasonal variability by month (ensemble climatologies), with IBTrACS observations shown in the right column of each panel, for (top) the Atlantic basin and (bottom) the northwest Pacific. With regard to the monthly split, a tropical cyclone is attributed to a specific month according to its date of genesis.

each basin. To the right of each time series, four correlations are shown: (i) model environment to model TCs (top left), (ii) model environment to IBTrACS TCs (top right), (iii) model environment to ERA5 environment (bottom left), and (iv) model TCs to IBTrACS TCs (bottom right).

Figure 10 shows the temporal evolution of the GPI term for each resolution in two basins: the North Atlantic and northwest Pacific, superposed on the ERA5 estimates. All experiments do a reasonable job of reproducing GPI, including seasonal to multiannual temporal signatures. The SP experiments have a small localized tendency to overshoot GPI at the peak of each season in the North Atlantic (end of season in the northwest Pacific), but overall it is hard to distinguish any strong response to SP or resolution. This is also shown by the interannual variability statistics on the right-hand side of each panel, which show that all models have high and comparable skill in simulating the IV of the environment, which is then reflected in high correlations between the models GPI and the models TCs in the North Atlantic (and south Indian Ocean, not shown), as well as providing a reasonably good predictor for observed TCs (*r* from 0.3 to 0.6). For the northwest Pacific, the correlation between simulated and ERA5 GPI is high, and so is the correlation between simulated GPI and TCs, but GPI is a poor predictor for observed TCs; the same applies in the southwest Pacific (not shown).

In terms of the PI terms, the thermodynamic efficiency provides an interesting case. Previous studies [see discussion in Camargo and Wing (2016) and Camargo et al. (2007a,b)] include mention of PI and interannual variability of TCs. Emanuel et al. 2013, Wing et al. 2015, and Bengtsson et al. (1995) computed the thermodynamic efficiency term of PI, $(T_s - T_o)/T_o$, where T_s is the surface temperature and T_o is the temperature of the outflow layer,[2] which was shown (Emanuel et al. 2013; Wing et al. 2015) to explain up to 30% of the PI trend in the North Atlantic, although virtually no signal was found in the northwest Pacific. The time series presented in both Emanuel et al. (2013) and Wing et al. (2015) contain, in fact, evidence of substantial interannual variability. In terms of seasonal variability of PI, Gilford et al. (2017) [also Gilford et al. (2019) for along-track PI] found that, in most ocean basins, the air–sea enthalpy disequilibrium (part of PI) drives seasonal variability, but in the western North Pacific, the only basin in which outflow levels are above the tropopause throughout the seasonal cycle, the seasonal cycle of lower-stratospheric temperatures influences outflow temperatures (and thus thermodynamic efficiency) and damps the seasonality of PI.

Inspired by the discussion in Wing et al. (2015)—"Investigating the contribution of tropical tropopause layer temperatures to interannual variability in maximum intensity, rather than trends, may therefore be a valuable extension of this work" (p. 8676)—we turn to the analysis of the interannual variability of the thermodynamic efficiency term. In so doing, the ultimate focus of this paper is the explicitly simulated TCs, rather than just the indices, which are used in this study as guidance in the investigation of potential predictors contained in the TC environment. The thermodynamic efficiency in the SPHINX simulations is governed, in terms of sensitivity to resolution

---

[2] Emanuel and Rotunno (2011) indicate that the temperature of tropical cyclone outflow is approximately equivalent to the ambient tropopause temperature.
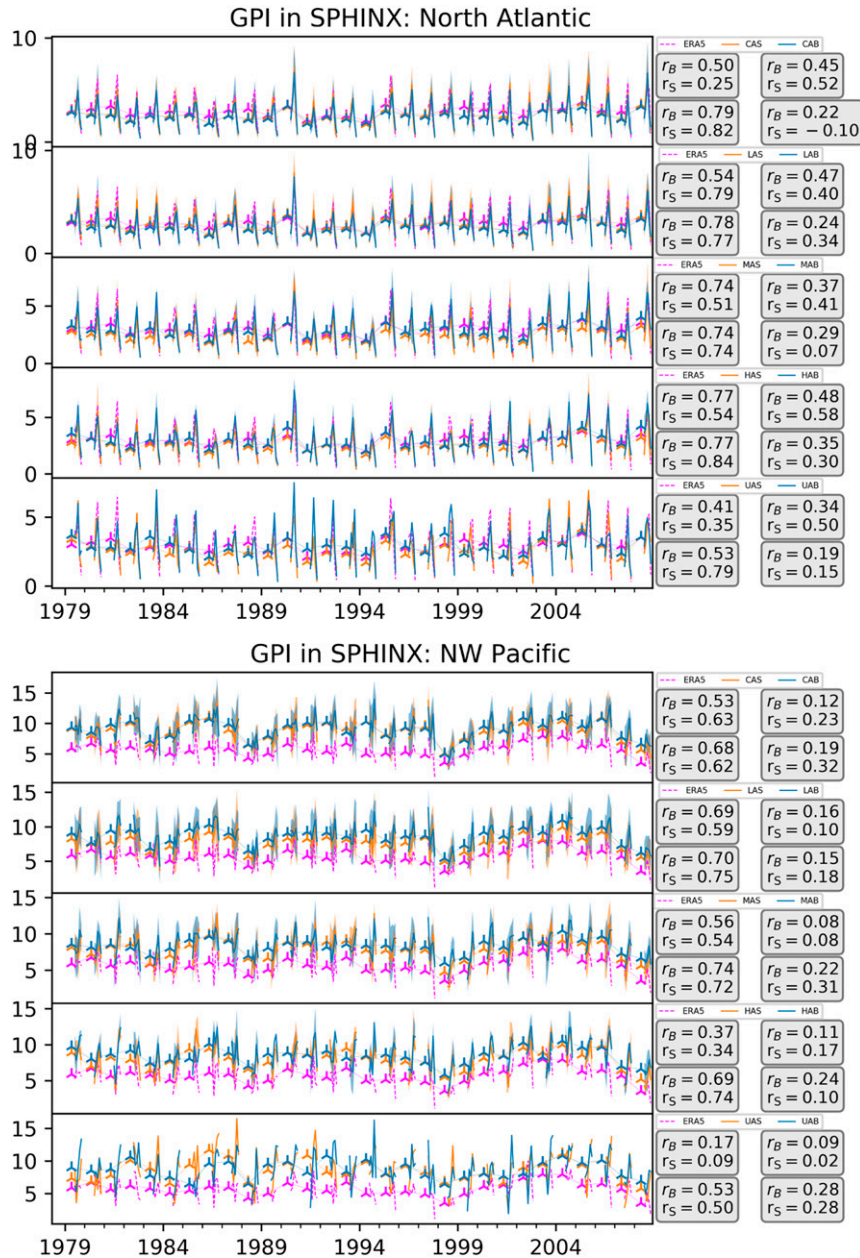
FIG. 10. (left) Time series of GPI and (right) annual correlations with observed environment and TCs, shown for the (top) North Atlantic and (bottom) northwest Pacific, with lowest resolution at the top of each panel and highest resolution at the bottom. A symbol to the left of each season in the monthly time series indicates the annual mean, for each experiment and for ERA5 observations; the shaded interval is a measure of the ensemble spread. Correlations are reported for both BASE (r_B) and STOC (r_S) to the right of each panel, as four boxes: (i) model environment to model TCs (top left); (ii) model environment to IBTrACS TCs (top right); (iii) model environment to ERA5 environment (bottom left), and (iv) model TCs to IBTrACS TCs (bottom right).

or SP, by T_o variability, since SSTs are prescribed, and could inject predictability at the interannual level.

Figure 11 shows the temporal evolution of the thermodynamic efficiency term, based on monthly means. This term seems to be overestimated for all SPHINX experiments, and

this error increases systematically as we reach the highest resolution. The reason for this is a negative bias in the T_o term, which is increasingly made worse by model resolution in all domains; compared to this, the SP sensitivity is insignificant. The seasonal (and even intraseasonal) signal is, however,
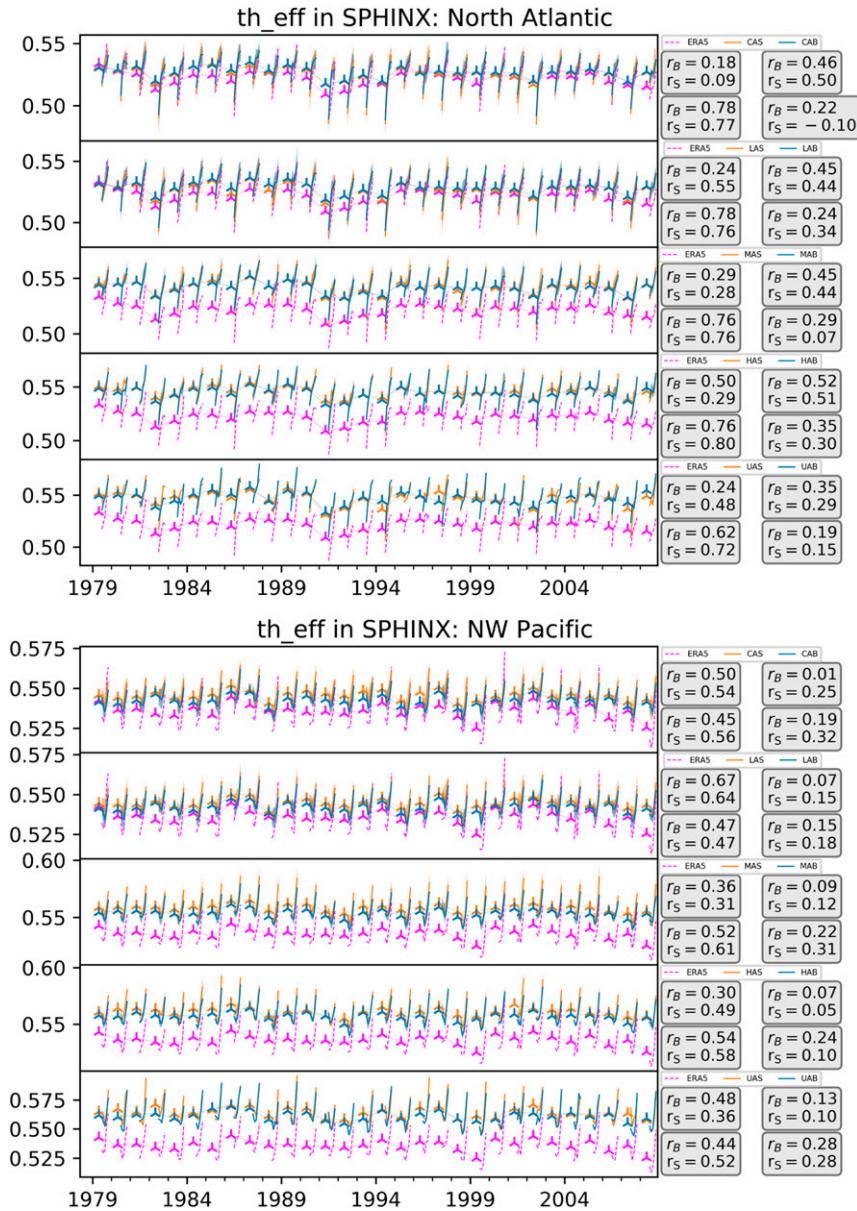
FIG. 11. As in Fig. 11, but for the GPI thermodynamic efficiency term (part of PI).

captured rather well, and, by comparison to ERA5, the skill at simulating interannual variability is the highest in the entire set, reaching correlations of up to 0.8. The correlation with the model TCs also shows that this is often the most important term in GPI for this particular model formulation.

The correlations between the annual mean count of TCs, using IBTrACS data, and the annual mean thermodynamic efficiency term have been computed and are shown in Fig. 11. The correlations for the ensemble mean and for each experiment (BASE, STOC) are also shown in the right-hand boxes. In general, correlations between the thermodynamic efficiency term and the observed TC count for the North Atlantic are surprisingly higher (~0.5) than have been found for Fig. 8 for

the correlation between simulated and observed TCs, albeit nearly zero in the northwest Pacific, as also found, albeit limited to the trend, in Wing et al. (2015). For the North Atlantic there is a slight increase in the correlations going from BASE to STOC, except for the ULTRA-HIGH resolution; STOC has higher correlations in only two cases, COARSE and MEDIUM.

A survey of all possible plot combinations, both time series and scatterplots (total of 108), indicates that the top governing environmental variables for predicting TC IV in the North Atlantic and south Indian Ocean are wind shear, thermodynamic efficiency, PI, and GPI (in that order, but nearly with the same weights for wind shear and thermodynamic efficiency). For the eastern Pacific, RH is the most important factor, followed by

GPI and then PI and wind shear. For the northwest Pacific GPI plays the largest role, while for the southwest Pacific it is vorticity, but both basins have such low skill at representing IV that this is hardly worth mentioning.

Once again, applying a Kolmogorov test, as in Fig. 8 reveals that the differences between the different experiments are not significant, but pooling all the results shown in Fig. 11 for the Atlantic and northwest Pacific basin and contrasting with those in Fig. 8 reveals that the improvement in skill provided by the thermodynamic efficiency term is in fact significant.

c. Scatterplots of IV correlations

A metric of IV correlation is introduced in order to better uncover any evidence of potential predictability, and to summarize the relative importance of SP and resolution for each field computed in the GPI diagnostic. While all correlations of all GPI variables have been computed in each basin, only two notable examples will be shown.

Figure 12 shows a summary of the correlations in Figs. 10 and 11. The set of correlations for BASE is plotted against the set of correlations for STOC, providing information on each resolution and on the potential of the GPI and its terms to provide interannual predictability for the number of TCs in each basin. It is clear that GPI is well simulated in both the Atlantic and the northwest Pacific, and that neither resolution nor SP makes a difference in the simulation of the interannual variability of GPI. Further, GPI is important for predicting the number of simulated TCs in a given year, albeit not a good predictor of observed TCs in the North Atlantic. For the northwest Pacific, while GPI is not a very good predictor of simulated TCs, PI is, and it is also in better agreement with the PI estimates in ERA5. The thermodynamic efficiency term is one of the two most important terms in the North Atlantic, but it is not well simulated in the northwest Pacific. PI is in fact not a good predictor for observed TCs in the northwest Pacific, but is a reasonably good predictor of simulated TCs (a moderately better predictor in this basin is in fact the GPI vorticity term; not shown).

5. Investigation of SP sensitivity with an independent GCM, HadGEM3-GC31

With regard to the reproducibility of the results based on the Climate-SPHINX simulations, it can be useful to resort to analyzing the simulations based on another model, with a completely independent dynamical core and set of physical parameterizations. Sensitivity experiments to test the impact of SP were designed using an independent model, HadGEM3-GC31 (see section 2) with comparable scientific configuration (except for SKEB2 instead of SKEB), forcing, and resolution of N216 ($\Delta x \approx 60$ km at 50°N), comparable to the LOW resolution in Climate-SPHINX. Contrary to the results from the Climate-SPHINX approach, however, HadGEM3-GC31 was tuned with the SP schemes enabled; therefore disabling SP is bound to produce slightly worse results overall (e.g., in terms of radiative balance).

Three ensemble members (control experiment, with full SP) are compared in Figs. 13a–c to one ensemble member using no SP (corresponding to the SPHINX BASE configuration; Fig. 13d) and two further ensemble members (Figs. 13e,f), which used only a single SP package: SKEB2 in one case and SPT in the second case.

The results shown in Fig. 13 are fully consistent with Climate-SPHINX: by comparison of the panels, the use of SP increases the mean counts by up to 30% and does so in regions that are directly comparable to those discussed earlier in this study (see, e.g., Fig. 3). The sensitivity to the use of SP is clearly larger than the ensemble spread, which can be estimated by comparing Figs. 13a, 13b, and 13c. The fact that the response to SP is far larger than the internal variability of the model (spawned by perturbing initial conditions) also matches SPHINX results well.

In terms of location and extent, the areas of strong sensitivity also match what was found in the SPHINX experiments (e.g., for the North Atlantic, the northwest Pacific, and the southwest Pacific, east of Australia). The sensitivity in the north Indian Ocean is far less, but there is some sensitivity in the south Indian Ocean, which partially matches SPHINX (at least around Madagascar). The sensitivity in South America is a well-known characteristic of HadGEM models, for instance as a response to resolution enhancement, and has been shown in multiple papers in the course of the last 10 years (Strachan et al. 2013; Roberts et al. 2015). With that, these South Atlantic systems tend to be tropical depressions or hybrid systems that are included in observational counts in some other basins, albeit not all. In fact, the South Atlantic Ocean is not officially classified as a tropical cyclone basin by the World Meteorological Organization and does not have a designated regional specialized meteorological center (RSMC).

These independent model results confirm our first finding: that the use of stochastic physics reduces the mean error in the simulation of TC track densities.

6. Discussion

The richness of the Climate-SPHINX ensembles enables a robust comparative study of the effects of resolution and the effects of SP, individually and combined, in the simulation of TCs. Cyclonic disturbances from 3600 model years have been extracted, with 10 000 TC precursors identified each year in each hemisphere. As a result, we were able to study the characteristics and evolution of 36 million cyclonic disturbances, from TC precursors to warm-core TCs. Climate-SPHINX confirms other studies (Zhao et al. 2009; Strachan et al. 2013; Roberts et al. 2015; Shaevitz et al. 2014) on the role of resolution for TC simulation, but also offers important further insights. For instance, the more recent PRIMAVERA simulations (Roberts et al. 2020) took a multimodel heterogeneous ensemble approach, with individual centers submitting three ensemble members each, and this was enough to show robust intermodel agreement on the TC response to resolution. The small ensemble spread found in Climate-SPHINX, for means and variability of TCs, also supports the robustness of previous studies. This finding should, however, not undermine the value of a large ensemble, which will come into play once we start to analyze extratropical transition, landfall, and most of all the response of intensity to imposed climate change.
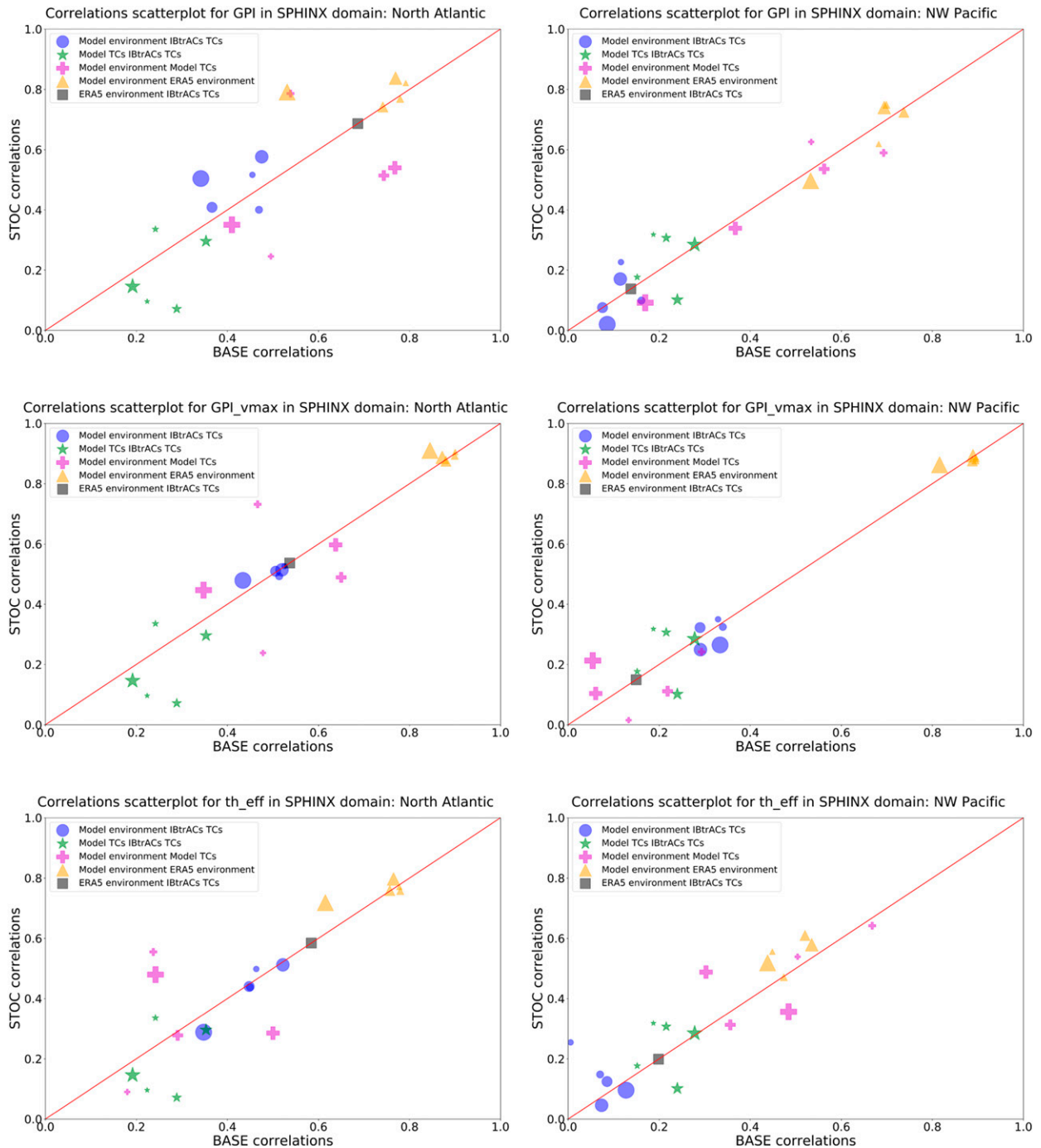
FIG. 12. Scatterplots of the IV correlations, corresponding to the boxes in Figs. 10 and 11, for (top) GPI, (middle) PI, and (bottom) the PI thermodynamic term for the (left) North Atlantic and (right) northwest Pacific basins. The size of the symbols corresponds to the different resolutions, with the smallest symbols identifying COARSE and the largest symbols identifying ULTRA-HIGH.

The novelty of this study lies in the robust comparison of the impact of the use of SP for the simulation of TCs, versus the traditional focus on just increasing resolution, under the premise that the two are to some extent equivalent (see Palmer 2019). The results of this study have indeed uncovered sensitivity to SP that mimics the effects of increased resolution

(albeit with smaller amplitude) and is overall beneficial. This claim is supported, for instance, by comparing the maps of TC track and genesis density in both EC-Earth and HadGEM3-GC31. The linear response to SP at each resolution in Climate-SPHINX, and for each basin, as well as the remarkable ensemble coherence, add to the robustness of our results.
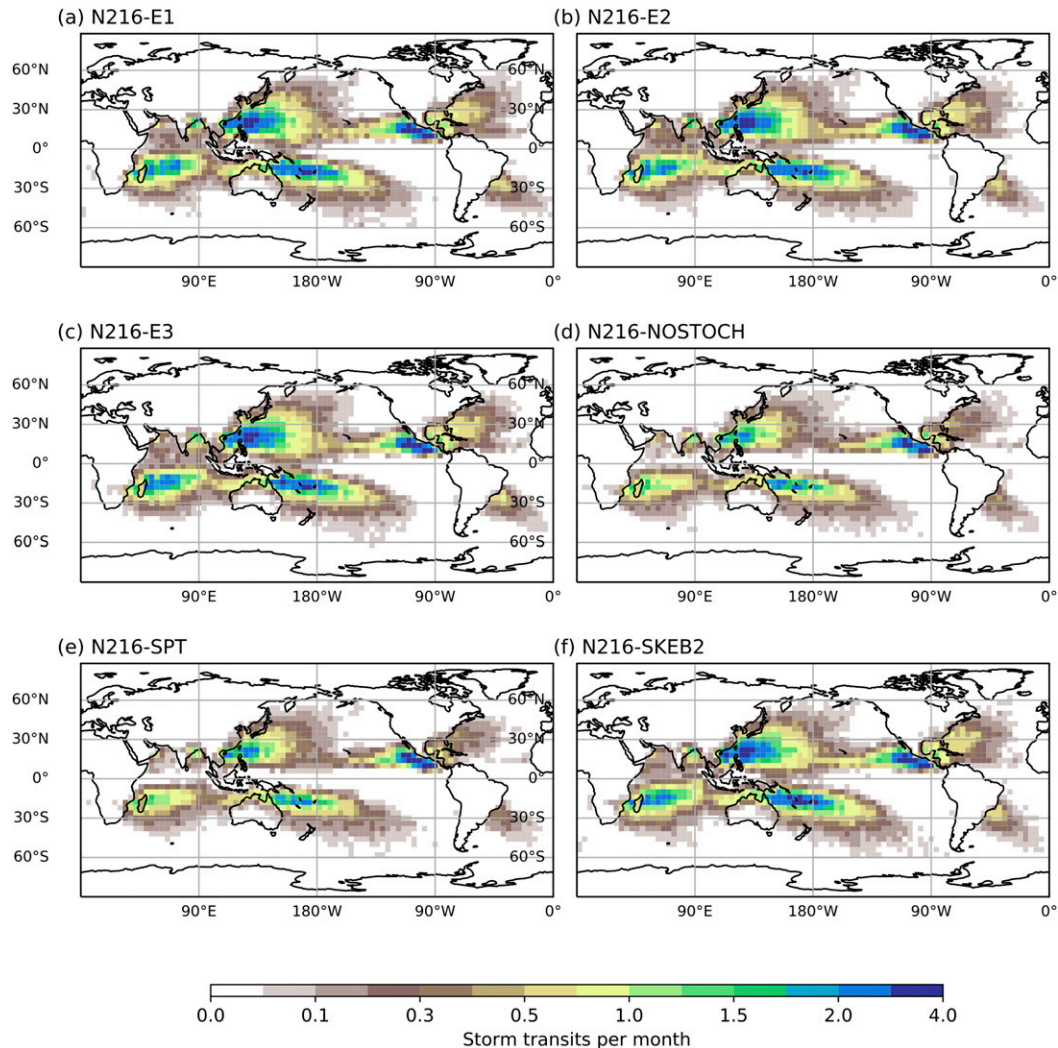
FIG. 13. TC track densities for the HadGEM3-GC31 model, as configured and run for the PRIMAVERA/ HighResMIP campaign. Shown are (a)–(c) three ensemble simulations with full SP, and simulations with (d) SP fully disabled, (e) SKEB2 disabled (SPT only), and (f) SPT disabled (SKEB2 only). Track density units are number per month per unit area, where the unit area is a 5° spherical cap.

There is, however, some localized evidence (east and southwest Pacific, despite substantial observational uncertainty; see Hodges et al. 2017) that the increase of the number of TCs for the higher resolutions may be excessive, and made worse by SP. SP has also been shown to have a far larger impact on the simulation of TCs than on the simulation of ETCs, which is important for the configuration of models used in operational prediction: Sanchez et al. (2014) for instance also found a larger impact in NWP simulations of the tropics versus extratropics.

Analysis of the relative roles of seeding and of the environment, via the cyclogenesis term, for the mean number and distribution of TCs (e.g., in Fig. 7) indicates that the latter is more important and more strongly responding to SP and resolution, but can lead to excessive transition, especially clear in the southwest Pacific basin, seen as very large values of PC. The smaller role played by the seeding term, and its lack of

sensitivity to resolution or SP, indicate that the relatively smooth seeding used by, for instance, Emanuel (2021), is likely adequate to study TC formation, as long as the significantly variable nature of environmental controls of cyclogenesis are then credibly taken into account. The finding, however, cannot be reconciled immediately with the importance of seeding put forward in Vecchi et al. (2019) in the context of climate change. A follow-on study, in preparation, will therefore present results from an identical Climate-SPHINX ensemble, forced by an RCP8.5 scenario, as well as century-long coupled simulations at MEDIUM resolution.

In terms of the interannual variability of TCs frequency, the correlation between simulations and IBTrACS observations for the EC-Earth configuration used in SPHINX is limited to 0.33 (HIGH-BASE) in the North Atlantic, which is a rather poor number when compared to past findings, for example

between 0.6 and 0.7 as in Roberts et al. (2015) for the UPSCALE campaign, or Roberts et al. (2020) for the HighResMIP experiments, using the HadGEM3-GC31 climate model. The range of correlations in HighResMIP is 0.3–0.7, and EC-Earth, in a configuration nearly identical to Climate-SPHINX, with the same SST and sea ice (Kennedy et al. 2017), albeit with some different forcings (e.g., MAC aerosols; Mulcahy et al. 2018), shows correlation of IV between observed and simulated TCs to be nearly identical to that in Fig. 8.

The correlation of observed IV to TCs found in the northwest Pacific is at most 0.27 (ULTRA-HIGH BASE); in the east Pacific it is 0 at best (MEDIUM-BASE), and mostly negative. Roberts et al. (2020) found in PRIMAVERA-HighResMIP, using HadGEM3-G31 at 25-km resolution, a correlation of 0.3 in the northwest Pacific and 0.5 in the eastern Pacific. However, at comparable resolution, HadGEM3-GC31 has substantially more TCs than EC-Earth (and even too many in the northwest Pacific), as seen in the multimodel intercomparison in Roberts et al. (2020), which was already found, to a lesser extent, in an earlier version of HadGEM3 at the same resolution (Roberts et al. 2015) that did not use SP.

The small spread in the time series of TC counts in each basin (e.g., Fig. 8) and the inability to accurately reproduce observed variability suggests that the EC-Earth model, as configured for the Climate-SPHINX campaign, is more skillful at predicting itself than at predicting observed TC counts in each year. This is also evidenced by the SNR results (Fig. A1) and overall characterizes the models in Climate-SPHINX as overconfident. These results are also in agreement with what has been shown in Camargo and Barnston (2009) in terms of model overconfidence, in the case of initialized seasonal forecasts, and stresses the fact that prediction of TC IV continues to constitute a challenge at the basic level of GCM formulation, which neither resolution nor SP is able to fully overcome.

SP enables a better representation of the seasonal cycle of TC frequency, particularly at the higher range of resolutions. The North Atlantic end of season response to SP is small, albeit robust, as even a small change near the end of August would impact the monthly means; this should be further investigated in (larger) ensembles of seasonal forecasts. To explain the seasonal skill, the simplest hypothesis would be that the response is due to the multiplicative nature of SPPT: the impact of the scheme should be larger when the variance of the deterministic tendencies is higher. It is reasonable to expect, a priori, for this to coincide with the peak of the TC season, when the variance is presumably higher, than at the tails of the season. This asymmetry might project onto the mean state changes to humidity, with a greater expected change at the peak of the seasonal cycle. For the northwest Pacific, however, the evolution of the seasonal anticyclone will also play a strong role at the end of the season, when TCs track northward, transporting much water vapor in the process, supplying other precipitation systems inland (see Guo et al. 2017).

In terms of environmental controls of TC interannual variability, the extensive review in Camargo and Wing (2016) pointed out that there is a stronger relationship between model GPI and observed TC variability than with the model TC variability. These conclusions still apply to the present study, where we showed that the environment is in general very well simulated, as compared to ERA5, and that the model GPI has correlations to observed TCs interannual frequency often twice as large as the correlation between simulated and observed TCs. The relative roles of the individual terms of the GPI are, however, moderately sensitive to resolution and to SP, with some important distinctions, depending on the basin.

The role of shear for TC development is well established, and the dependence of shear on model resolution in this context has been discussed in Bell et al. (2013) and Roberts et al. (2015), among others. The same relationship is apparent in SPHINX, and vertical shear appears to respond to resolution and SP in terms of climate means. For IV, it is an important driver in the North Atlantic and south Indian Ocean, although not so across the rest of the globe, which could be related to the overall problems in storm structure (EC-Earth TCs are too large compared to observations and other models; see Vanniere et al. 2020).

The role of atmospheric moisture appears to be comparatively more prominent for simulations in which SP has been activated; this sensitivity has also been pointed out in previous studies by Bell et al. (2013) and Camargo et al. (2020). More specifically, in the context of SP studies, this is consistent with the idea put forward in Strommen et al. (2019b) that, by broadening the distribution of humidity tendencies, SPPT has a particularly strong impact on humidity, due to the nonlinearities associated with condensation. For example, for a parcel of air close to saturation, a tendency perturbation in one direction may trigger condensation, increasing the total liquid water, while the opposite perturbation will not. Changes in available moisture might therefore be expected to be magnified in a model with SPPT turned on, and for the specific problem of TC simulation, with potential for local feedbacks. A cautionary note is needed here with regard to lack of moisture conservation when using SPPT, which required a fix [see discussion in Davini et al. (2017)]. The full impact of the moisture fix is not known, since the model will not run without it, and yet sensitivity to SP in SPHINX simulations shows a stronger impact of the moisture field on TCs, while ETCs are less impacted, possibly due the moisture response being limited to the tropics. RH is, however, not a useful predictor in terms of IV.

The thermodynamic efficiency term shows promise in explaining the different TC responses for the North Atlantic and south Indian Ocean domains. This is particularly evident in the lower range of resolutions, in which SP is more active in terms of changes in the probability of cyclogenesis (e.g., see Fig. 7). The thermodynamic efficiency term shows correlations with observed TC frequencies (IBTrACS) that are nearly twice as large as the correlations between observed and model-produced TC frequencies. It is reasonable to expect that this in part reflects the use of observed SSTs and that the importance of this term would be diminished in AOGCM simulations. Future work on existing Climate-SPHINX data will in fact exploit the coupled atmosphere–ocean experiments, with a special focus on TC intensity, for which there was no space in this paper, as well as the 2070–2100 (FUTURE) atmosphere-

only simulations, expanding the investigation of TC predictability.

While it is notable that thermodynamic efficiency is found to play an important role in these simulations (and this reflects the importance of the outflow temperature variability), the SPHINX models develop a substantial cold bias as resolution is increased, seen when plotting the tropical tropopause layer temperature time series (TTL; not shown), which is the reason for the errors seen in the PI terms. This can in part be attributed to the choice of not retuning the models at each resolution, so it is not a final statement of models being "fit for purpose" in terms of climate change. In fact, this is very important for our trust in future projections, as Emanuel et al. (2013) (in the context of decadal to interannual TTL signals) suggest that "the failure of most GCMs to capture this cooling must be addressed before such models can be used to project future changes in tropical cyclone activity" (p. 2300). It should not be too hard to adapt EC-Earth to more reliably simulate the future of TCs: to improve the overall simulated environment, tuning TOA radiation is normally easier than tuning surface layer similarity (often done for improving TC wind–pressure relationships and air-sea fluxes). Given the computational costs involved, exploiting IV for carefully selected periods is more economical than running long climate simulations.

The apparent skill in GPI, particularly the thermodynamic efficiency term in the North Atlantic and south Indian Ocean provides a paradox, pointing to other types of model formulation uncertainty, unaffected by resolution or SP, that seem to prevent the model from exploiting its own potential source of predictability. The fixed roles of dynamics and physics could be unsuitable for such a large range of resolutions [see discussion in Vanniere et al. (2020)], affecting their interaction with the large-scale environment. Recent development work with the IFS (e.g., Magnusson et al. 2019), on surface friction, seems to have already improved the EC-Earth model with respect to TC simulation, and some of these advances are in fact evidenced by the results in Roberts et al. (2020).

In terms of future directions, the resources used to produce the T1279-BASE simulation could have enabled a further four to five ensemble members with T799-STOC, bringing it to eight, or enabling more sensitivity experiments. For instance, as a general recommendation, we endorse Strommen et al. (2019a) in pointing out that model tuning performed on SP configurations may benefit the prediction of TCs in an operational setting; this may require SP and even resolution-based tuning in the case of scale-aware SP parameterizations, such as SKEB2. This requires further research, as SKEB2 has been shown to have a larger impact than SPT (as exemplified by Fig. 13). It is important to remember that EC-Earth is a spectral hydrostatic model, while HadGEM3 is a grid point nonhydrostatic model: horizontal advection and wave propagation are, for instance, substantially different, and so is the representation of tropical convection, so that the superposed effects of SP cannot be expected to be the same. Moreover, the type of SP formulation (SKEB2 instead of SKEB), as well as the individual parameter settings in HadGEM3-GC31 are different from what has been used in Climate-SPHINX, as well as different from the parameter settings for EC-Earth in PRIMAVERA-HighResMIP [the models analyzed by Roberts et al. (2020) and Vanniere et al. (2020)], all of which points to more fundamental aspects of model formulation to explain the different responses, and a large number of individual sensitivity studies would be needed for robustness.

All of the above does not detract from the strong finding that the overall sensitivity of TC simulation to resolution and SP is consistent across two entirely independent CMIP6 GCMs.

7. Summary and conclusions

Statistics of TCs identified in Climate-SPHINX confirm past and current findings that increasing model resolution systematically improves the simulated climatology—numbers and distribution—in both hemispheres. The use of stochastic physics further increases the number of TCs, by $\approx 30\%$, when compared to the base simulations, in a spatially realistic way, thus representing a surrogate for resolution.

Analysis of the impact of SP as a cause of additional TC seeding, versus increasing the probability of transition of TC seeds to TCs, through modification of the TC environment, points to the latter being the prevalent effect. Further analysis, focusing on the interannual variability of TC numbers per year indicates that it is larger overall, and more realistic, when applying SP and/or enhancing resolution. Unfortunately, the increased IV of simulated TCs does not translate into significantly enhanced skill in terms of predicting the annual number of observed TCs: if anything, SP seems to be adding more noise than signal to the problem of predicting the annual number of TCs. The representation of the seasonal cycle of TC frequency is however improved by the use of SP at the higher range of resolutions.

From the point of view of TC predictors, the realistic performance of the thermodynamic efficiency term (part of the potential intensity calculation) applied to the EC-Earth simulations provides a stringent test to strengthen our trust in the use of EC-Earth (or similar GCMs) for the prediction of TC responses to climate variability and change, even in the case of models at moderate resolution. However, for this generation of EC-Earth specifically, other aspects of simulation fidelity need to be improved before this potential predictability can be exploited. The parsimonious nature of SP can in fact be exploited for further model development.
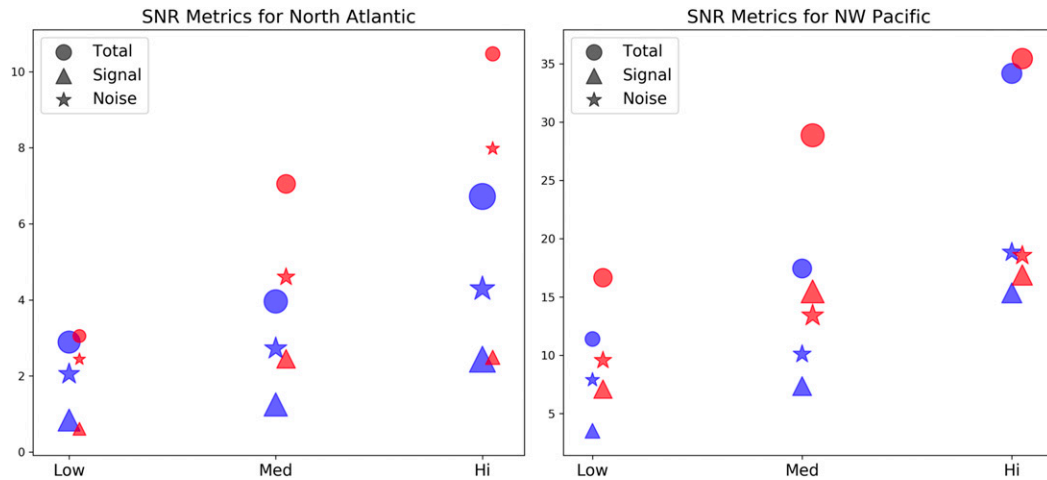
FIG. A1. TC variances (total, signal, and noise) based on the ensemble interannual variability for different domains in Climate-SPHINX. The symbol sizes are proportional to the signal-to-noise ratio. Units are TCs per year.
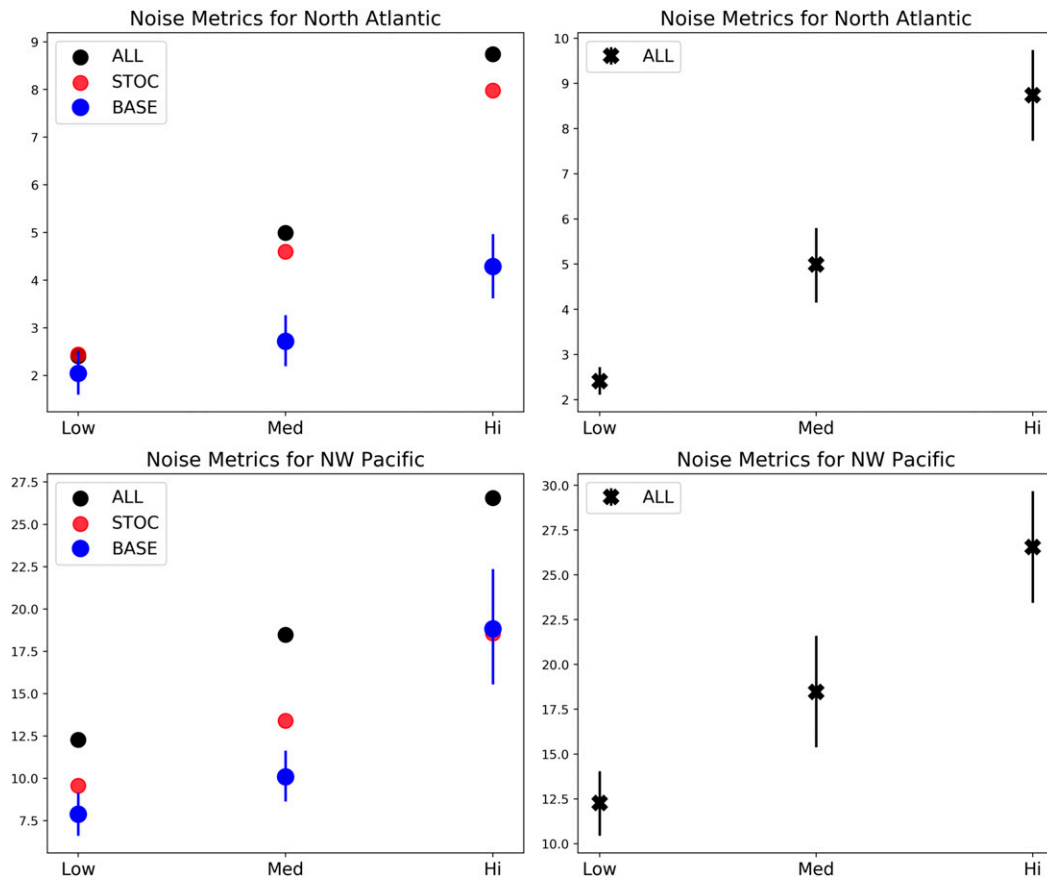
FIG. A2. The noise component in the TC variance calculations, based on the ensemble interannual variability for different domains in Climate-SPHINX. Units are TCs per year.

APPENDIX

The Signal-to-Noise Ratio of Simulated TC Interannual Variability

As a way to further interpret Fig. 8, we computed the signal-to-noise ratio (SNR): $\sigma^2_{\text{signal}}/\sigma^2_{\text{noise}}$. where σ^2_{signal} is the signal variance of the model ensemble mean in time, while σ^2_{noise} is the variance of the ensemble members about the ensemble mean [the spread,[A1] as in Eade et al. (2014)]. The analysis was made more robust by resampling with bootstrapping, to get 1000 artificially generated ensembles (for each of the stochastic/deterministic ensembles at each resolution). The 95% confidence interval obtained in this way is used to test statistical significance. Additionally, the three higher resolutions were combined, in order to create an ensemble size comparable to that of COARSE and LOW.

Figure A1 summarizes the findings so far, with STOC in red and BASE in blue. In terms of signal, STOC is significantly higher than BASE for the MEDIUM resolution, but otherwise not distinguishable, with no significant differences overall. Across the resolution it is hard to distinguish any two adjacent resolutions from each other, but the (combined) high resolutions are in fact significantly higher than the lowest resolutions, suggesting that the increase is robust. For SNR, STOC is significantly lower for the (combined) HIGH experiment, and just barely not significantly lower for the COARSE. No significant difference in SNR were found for MEDIUM and no distinguishable change was found across all resolutions.

The clearest and most robust change across both resolution and stochastic physics is in the noise component, shown in Fig. A2. A significant increase in noise is seen for both basins, with each successive increase of resolution, when pooling both BASE and STOC (right column of Fig. A2). SP significantly increases the noise in four of the six cases, with a particularly pronounced impact in the North Atlantic. Because the SNR metric does not experience a similarly robust increase, the data suggest that the first-order impact of increased resolution, and, to a lesser extent, SP, is mainly to increase internal variability. Additional model tuning is likely required in order to obtain comparable increases in the signal and associated improvements in predictability.

[A1] Also the difference $\sigma^2_{\text{total}} - \sigma^2_{\text{signal}}$, where σ^2_{total} is the average variance of the individual members.

REFERENCES

Bacmeister, J. T., K. A. Reed, C. Hannay, P. Lawrence, S. Bates, J. E. Truesdale, N. Rosenbloom, and M. Levy, 2018: Projected changes in tropical cyclone activity under future warming scenarios using a high-resolution climate model. *Climatic Change*, **146**, 547–560, https://doi.org/10.1007/s10584-016-1750-x.

Baker, A. J., K. I. Hodges, R. K. H. Schiemann, and P. L. Vidale, 2021: Historical variability and lifecycles of North Atlantic midlatitude cyclones originating in the tropics. *J. Geophys. Res. Atmos.*, https://doi.org/10.1029/2020JD033924, in press.

Bell, R., J. Strachan, P. L. Vidale, K. Hodges, and M. Roberts, 2013: Response of tropical cyclones to idealized climate change experiments in a global high-resolution coupled general circulation model. *J. Climate*, **26**, 7966–7980, https://doi.org/10.1175/JCLI-D-12-00749.1.

——, K. Hodges, P. L. Vidale, J. Strachan, and M. Roberts, 2014: Simulation of the global ENSO–tropical cyclone teleconnection by a high-resolution coupled general circulation model. *J. Climate*, **27**, 6404–6422, https://doi.org/10.1175/JCLI-D-13-00559.1.

Bengtsson, L., M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus*, **47A**, 175–196, https://doi.org/10.3402/tellusa.v47i2.11500.

——, M. Steinheimer, P. Bechtold, and J. F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Quart. J. Roy. Meteor. Soc.*, **139**, 1533–1543, https://doi.org/10.1002/qj.2108.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

——, T. Jung, and T. N. Palmer, 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25**, 4946–4962, https://doi.org/10.1175/JCLI-D-11-00297.1.

——, and Coauthors, 2017: Stochastic parameterizations: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

Bhatia, K., G. Vecchi, H. Murakami, S. Underwood, and J. Kossin, 2018: Projected response of tropical cyclone intensity and intensification in a global climate model. *J. Climate*, **31**, 8281–8303, https://doi.org/10.1175/JCLI-D-17-0898.1.

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, https://doi.org/10.1029/2001JD000776.

Brammer, A., and C. D. Thorncroft, 2015: Variability and evolution of African easterly wave structures and their relationship with tropical cyclogenesis over the eastern Atlantic. *Mon. Wea. Rev.*, **143**, 4975–4995, https://doi.org/10.1175/MWR-D-15-0106.1.

Broccoli, A. J., and S. Manabe, 1990: Can existing climate models be used to study anthropogenic changes in tropical cyclone climate? *Geophys. Res. Lett.*, **17**, 1917–1920, https://doi.org/10.1029/GL017i011p01917.

Buizza, R., M. Miller, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Camargo, S. J., and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491, https://doi.org/10.1175/2008WAF2007099.1.

——, and A. A. Wing, 2016: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, https://doi.org/10.1002/wcc.373.

——, A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007a: Tropical cyclone genesis potential index in climate models. *Tellus*, **59**, 428–443, https://doi.org/10.1111/j.1600-0870.2007.00238.x.

——, K. A. Emanuel, and A. H. Sobel, 2007b: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834, https://doi.org/10.1175/JCLI4282.1.

——, and Coauthors, 2020: Characteristics of model tropical cyclone climatology and the large-scale environment. *J. Climate*, **33**, 4463–4487, https://doi.org/10.1175/JCLI-D-19-0500.1.

Chan, J. C. L., 1985: Tropical cyclone activity in the northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. *Mon. Wea. Rev.*, **113**, 599–606, https://doi.org/10.1175/1520-0493(1985)113<0599:TCAITN>2.0.CO;2.

Christensen, H. M., J. Berner, D. R. Coleman, and T. N. Palmer, 2017: Stochastic parameterization and El Niño–Southern Oscillation. *J. Climate*, **30**, 17–38, https://doi.org/10.1175/JCLI-D-16-0122.1.

Davini, P., and Coauthors, 2017: Climate SPHINX: Evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geosci. Model Dev.*, **10**, 1383–1402, https://doi.org/10.5194/gmd-10-1383-2017.

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, https://doi.org/10.1002/2014GL061146.

Emanuel, K., 2018: 100 years of progress in tropical cyclone research. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial, Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0016.1.

——, 2021: Response of global tropical cyclone activity to increasing CO_2: Results from downscaling CMIP6 models. *J. Climate*, **34**, 57–70, https://doi.org/10.1175/JCLI-D-20-0367.1.

——, and D. S. Nolan, 2004: Tropical cyclone activity and the global climate system. *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 10A.2, https://ams.confex.com/ams/26HURR/techprogram/paper_75463.htm.

——, and R. Rotunno, 2011: Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. *J. Atmos. Sci.*, **68**, 2236–2249, https://doi.org/10.1175/JAS-D-10-05024.1.

——, S. Solomon, D. Folini, S. Davis, and C. Cagnazzo, 2013: Influence of tropical tropopause layer cooling on Atlantic hurricane activity. *J. Climate*, **26**, 2288–2301, https://doi.org/10.1175/JCLI-D-12-00242.1.

Evans, J. L., 1992: Comment on "Can existing climate models be used to study anthropogenic changes in tropical cyclone climate?". *Geophys. Res. Lett.*, **19**, 1523–1524, https://doi.org/10.1029/92GL01120.

Fine, C. M., R. H. Johnson, P. E. Ciesielski, and R. K. Taft, 2016: The role of topographically induced vortices in tropical cyclone formation over the Indian Ocean. *Mon. Wea. Rev.*, **144**, 4827–4847, https://doi.org/10.1175/MWR-D-16-0102.1.

Franco-Díaz, A., N. P. Klingaman, P. L. Vidale, L. Guo, and M.-E. Demory, 2019: The contribution of tropical cyclones to the atmospheric branch of middle America's hydrological cycle using observed and reanalysis tracks. *Climate Dyn.*, **53**, 6145–6158, https://doi.org/10.1007/s00382-019-04920-z.

Fu, B., M. S. Peng, T. Li, and D. E. Stevens, 2012: Developing versus nondeveloping disturbances for tropical cyclone for-

mation. Part II: Western North Pacific. *Mon. Wea. Rev.*, **140**, 1067–1080, https://doi.org/10.1175/2011MWR3618.1.

Gilford, D. M., S. Solomon, and K. A. Emanuel, 2017: On the seasonal cycles of tropical cyclone potential intensity. *J. Climate*, **30**, 6085–6096, https://doi.org/10.1175/JCLI-D-16-0827.1.

——, ——, and ——, 2019: Seasonal cycles of along-track tropical cyclone maximum intensity. *Mon. Wea. Rev.*, **147**, 2417–2432, https://doi.org/10.1175/MWR-D-19-0021.1.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668, https://doi.org/10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2.

Guo, L., N. P. Klingaman, P. L. Vidale, A. G. Turner, M.-E. Demory, and A. Cobb, 2017: Contribution of tropical cyclones to atmospheric moisture transport and rainfall over East Asia. *J. Climate*, **30**, 3853–3865, https://doi.org/10.1175/JCLI-D-16-0308.1.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Hankes, I., Z. Wang, G. Zhang, and C. Fritz, 2015: Merger of African easterly waves and formation of Cape Verde storms. *Quart. J. Roy. Meteor. Soc.*, **141**, 1306–1319, https://doi.org/10.1002/qj.2439.

Hazeleger, W., and Coauthors, 2010: EC-Earth: A seamless Earth-system prediction approach in action. *Bull. Amer. Meteor. Soc.*, **91**, 1357–1364, https://doi.org/10.1175/2010BAMS2877.1.

He, F., and D. J. Posselt, 2015: Impact of parameterized physical processes on simulated tropical cyclone characteristics in the Community Atmosphere Model. *J. Climate*, **28**, 9857–9872, https://doi.org/10.1175/JCLI-D-15-0255.1.

Hodges, K., 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.

——, 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.

Hsieh, T.-L., G. A. Vecchi, W. Yang, I. M. Held, and S. T. Garner, 2020: Large-scale control on the frequency of tropical cyclones and seeds: A consistent relationship across a hierarchy of global atmospheric models. *Climate Dyn.*, **55**, 3177–3196, https://doi.org/10.1007/s00382-020-05446-5.

Judt, F., and Coauthors, 2021: Tropical cyclones in global storm-resolving models. *J. Meteor. Soc. Japan*, https://doi.org/10.2151/jmsj.2021-029, in press.

Jung, T., and Coauthors, 2012: High-resolution global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate, and seasonal forecast skill. *J. Climate*, **25**, 3155–3172, https://doi.org/10.1175/JCLI-D-11-00265.1.

Kennedy, J., H. Titchner, N. Rayner, and M. Roberts, 2017: input4mips.mohc.sstsandseaice.highresmip.mohc-hadisst-2-2-0-0-0. Earth System Grid Federation, accessed 2020, https://doi.org/10.22033/ESGF/input4MIPs.1221.

Kepert, J. D., 2012: Choosing a boundary layer parameterization for tropical cyclone modeling. *Mon. Wea. Rev.*, **140**, 1427–1445, https://doi.org/10.1175/MWR-D-11-00217.1.

Khouider, B., J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216, https://doi.org/10.4310/CMS.2010.v8.n1.a10.

Klaver, R., R. Haarsma, P. L. Vidale, and W. Hazeleger, 2020: Effective resolution in high-resolution global atmospheric models for climate studies. *Atmos. Sci. Lett.*, **21**, e952, https://doi.org/10.1002/asl.952.

Klotzbach, P. J., M. M. Bell, S. G. Bowen, E. J. Gibney, K. R. Knapp, I. Schreck, and J. Carl, 2020: Surface pressure a more skillful predictor of normalized hurricane damage than maximum sustained wind. *Bull. Amer. Meteor. Soc.*, **101**, E830–E846, https://doi.org/10.1175/BAMS-D-19-0062.1.

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). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Knutson, T. R., and Coauthors, 2013: Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *J. Climate*, **26**, 6591–6617, https://doi.org/10.1175/JCLI-D-12-00539.1.

——, and Coauthors, 2020: Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, **101**, E303–E322, https://doi.org/10.1175/BAMS-D-18-0194.1.

Lander, M. A., 1994: An exploratory analysis of the relationship between tropical storm formation in the western North Pacific and ENSO. *Mon. Wea. Rev.*, **122**, 636–651, https://doi.org/10.1175/1520-0493(1994)122<0636:AEAOTR>2.0.CO;2.

LaRow, T. E., 2013: The impact of SST bias correction on North Atlantic hurricane retrospective forecasts. *Mon. Wea. Rev.*, **141**, 490–498, https://doi.org/10.1175/MWR-D-12-00152.1.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

Magnusson, L., and Coauthors, 2019: ECMWF activities for improved hurricane forecasts. *Bull. Amer. Meteor. Soc.*, **100**, 445–458, https://doi.org/10.1175/BAMS-D-18-0044.1.

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, 2014: Future changes in the western North Pacific tropical cyclone activity projected by a multidecadal simulation with a 16-km global atmospheric GCM. *J. Climate*, **27**, 7622–7646, https://doi.org/10.1175/JCLI-D-13-00678.1.

Mei, W., Y. Kamae, S.-P. Xie, and K. Yoshida, 2019: Variability and predictability of North Atlantic hurricane frequency in a large ensemble of high-resolution atmospheric simulations. *J. Climate*, **32**, 3153–3167, https://doi.org/10.1175/JCLI-D-18-0554.1.

Moss, R., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, https://doi.org/10.1038/nature08823.

Mulcahy, J. P., and Coauthors, 2018: Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1. *J. Adv. Model. Earth Syst.*, **10**, 2786–2805, https://doi.org/10.1029/2018MS001464.

Murakami, H., and Coauthors, 2012: Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM. *J. Climate*, **25**, 3237–3260, https://doi.org/10.1175/JCLI-D-11-00415.1.

——, P.-C. Hsu, O. Arakawa, and T. Li, 2014: Influence of model biases on projected future changes in tropical cyclone frequency of occurrence. *J. Climate*, **27**, 2159–2181, https://doi.org/10.1175/JCLI-D-13-00436.1.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, https://doi.org/10.1002/qj.49712757202.

——, 2019: Stochastic weather and climate models. *Nat. Rev. Phys.*, **1**, 463–471, https://doi.org/10.1038/s42254-019-0062-2.

——, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 44 pp., www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf.

Peng, M. S., B. Fu, T. Li, and D. E. Stevens, 2012: Developing versus nondeveloping disturbances for tropical cyclone formation. Part I: North Atlantic. *Mon. Wea. Rev.*, **140**, 1047–1066, https://doi.org/10.1175/2011MWR3617.1.

Plant, R., and C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, https://doi.org/10.1175/2007JAS2263.1.

Roberts, M. J., and Coauthors, 2015: Tropical cyclones in the upscale ensemble of high-resolution global climate models. *J. Climate*, **28**, 574–596, https://doi.org/10.1175/JCLI-D-14-00131.1.

——, and Coauthors, 2019: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, **12**, 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019.

——, 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.

Sainsbury, E. M., R. K. H. Schiemann, K. I. Hodges, L. C. Shaffrey, A. J. Baker, and K. T. Bhatia, 2020: How important are post-tropical cyclones for European windstorm risk? *Geophys. Res. Lett.*, **47**, e2020GL089853, https://doi.org/10.1029/2020GL089853.

Sanchez, C., K. D. Williams, G. Shutts, and M. Collins, 2014: Impact of a stochastic kinetic energy backscatter scheme across time-scales and resolutions. *Quart. J. Roy. Meteor. Soc.*, **140**, 2625–2637, https://doi.org/10.1002/qj.2328.

——, ——, and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models. *Quart. J. Roy. Meteor. Soc.*, **142**, 147–159, https://doi.org/10.1002/qj.2640.

Serra, Y. L., G. N. Kiladis, and K. I. Hodges, 2010: Tracking and mean structure of easterly waves over the Intra-Americas Sea. *J. Climate*, **23**, 4823–4840, https://doi.org/10.1175/2010JCLI3223.1.

Shaevitz, D. A., and Coauthors, 2014: Characteristics of tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth Syst.*, **6**, 1154–1172, https://doi.org/10.1002/2014MS000372.

Shutts, G., and T. Palmer, 2007: Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, **20**, 187–202, https://doi.org/10.1175/JCLI3954.1.

Slingo, J., and T. Palmer, 2011: Uncertainty in weather and climate prediction. *Philos. Trans. Roy. Soc. London*, **369A**, 4751–4767, https://doi.org/10.1098/rsta.2011.0161.

——, and Coauthors, 1994: Mean climate and transience in the tropics of the UGAMP GCM: Sensitivity to convective parametrization. *Quart. J. Roy. Meteor. Soc.*, **120**, 881–922, https://doi.org/10.1002/qj.49712051807.

Stevens, B., S. Fiedler, S. Kinne, K. Peters, S. Rast, J. Müsse, S. J. Smith, and T. Mauritsen, 2017: MACv2-sp: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geosci. Model Dev.*, **10**, 433–452, https://doi.org/10.5194/gmd-10-433-2017.

——, and Coauthors, 2019: DYAMOND: The dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Prog. Earth Planet. Sci.*, **6**, 61, https://doi.org/10.1186/s40645-019-0304-z.

Stockdale, T., and Coauthors, 2018: SEAS5 and the future evolution of the long-range forecast system. ECMWF Tech. Memo. 835, 83 pp., http://ecmwf.int/sites/default/files/elibrary/2018/18750-seas5-and-future-evolution-long-range-forecast-system.pdf.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts, and M.-E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Climate*, **26**, 133–152, https://doi.org/10.1175/JCLI-D-12-00012.1.

Strommen, K., H. M. Christensen, D. MacLeod, S. Juricke, and T. N. Palmer, 2019a: Progress towards a probabilistic Earth system model: Examining the impact of stochasticity in the atmosphere and land component of EC-Earth v3.2. *Geosci. Model Dev.*, **12**, 3099–3118, https://doi.org/10.5194/gmd-12-3099-2019.

——, P. A. G. Watson, and T. N. Palmer, 2019b: The impact of a stochastic parameterization scheme on climate sensitivity in EC-Earth. *J. Geophys. Res. Atmos.*, **124**, 12 726–12 740, https://doi.org/10.1029/2019JD030732.

Sugi, M., Y. Yamada, K. Yoshida, R. Mizuta, M. Nakano, C. Kodama, and M. Satoh, 2020: Future changes in the global frequency of tropical cyclone seeds. *SOLA*, **16**, 70–74, https://doi.org/10.2151/sola.2020-012.

Teixeira, J., and C. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496, https://doi.org/10.1175/2007MWR1870.1.

Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206, https://doi.org/10.1175/2010MWR3430.1.

Thorncroft, C., and K. Hodges, 2001: African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J. Climate*, **14**, 1166–1179, https://doi.org/10.1175/1520-0442(2001)014<1166:AEWVAI>2.0.CO;2.

Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, https://doi.org/10.1002/2013JD020316.

Vannière, B., and Coauthors, 2020: The moisture budget of tropical cyclones in HighResMIP models: Large-scale environmental balance and sensitivity to horizontal resolution. *J. Climate*, **33**, 8457–8474, https://doi.org/10.1175/JCLI-D-19-0999.1.

Vecchi, G. A., and Coauthors, 2019: Tropical cyclone sensitivities to CO_2 doubling: Roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dyn.*, **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Vidale, P. L., D. Lüthi, C. Frei, S. I. Seneviratne, and C. Schär, 2003: Predictability and uncertainty in a regional climate model. *J. Geophys. Res.*, **108**, 4586, https://doi.org/10.1029/2002JD002810.

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.

——, and Coauthors, 2015: Hurricanes and climate: The U.S. CLIVAR working group on hurricanes. *Bull. Amer. Meteor. Soc.*, **96**, 997–1017, https://doi.org/10.1175/BAMS-D-13-00242.1.

Walters, D., and Coauthors, 2019: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.*, **12**, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019.

Wang, Q., and Coauthors, 2019: Tropical cyclones act to intensify El Niño. *Nat. Commun.*, **10**, 3793, https://doi.org/10.1038/s41467-019-11720-w.

Watson, P. A. G., J. Berner, S. Corti, P. Davini, J. von Hardenberg, C. Sanchez, A. Weisheimer, and T. N. Palmer, 2017: The impact of stochastic physics on tropical rainfall variability in global climate models on daily to weekly timescales. *J. Geophys. Res. Atmos.*, **122**, 5738–5762, https://doi.org/10.1002/2016JD026386.

Weisheimer, A., S. Corti, T. Palmer, and F. Vitart, 2014: Addressing model error through atmospheric stochastic physical parametrizations: Impact on the coupled ECMWF seasonal forecasting system. *Philos. Trans. Roy. Soc.*, **372**, 20130290, https://doi.org/10.1098/rsta.2013.0290.

Wilk, D., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 704 pp.

Williams, P., 2005: Modelling climate change: The role of unresolved processes. *Philos. Trans. Roy. Soc. London*, **363A**, 2931–2946, https://doi.org/10.1098/rsta.2005.1676.

Wing, A. A., K. Emanuel, and S. Solomon, 2015: On the factors affecting trends and variability in tropical cyclone potential intensity. *Geophys. Res. Lett.*, **42**, 8669–8677, https://doi.org/10.1002/2015GL066145.

Yamada, Y., and Coauthors, 2021: Evaluation of the contribution of tropical cyclone seeds to changes in tropical cyclone frequency due to global warming in high-resolution multi-model ensemble simulations. *Prog. Earth Planet. Sci.*, **8**, 11, https://doi.org/10.1186/s40645-020-00397-1.

Yang, C., H. Christensen, S. Corti, J. von Hardenberg, and P. Davini, 2019: The impact of stochastic physics on the El Niño southern oscillation in the EC-Earth coupled model. *Climate Dyn.*, **53**, 2843–2859, https://doi.org/10.1007/s00382-019-04660-0.

Yoshida, K., M. Sugi, R. Mizuta, H. Murakami, and M. Ishii, 2017: Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res. Lett.*, **44**, 9910–9917, https://doi.org/10.1002/2017GL075058.

Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678, https://doi.org/10.1175/2009JCLI3049.1.