Replies to Anonymous Reviewer #2

Dear Reviewer,

Thank you for your comments and observations, which helped us to considerably improve the quality of the manuscript. We made several minor adjustments and a few major changes, which are listed here below:

- We widely revised our introduction and conclusion part - together with the section describing the stochastic physics - in order to better frame our work in the context of other existing stochastic methods in earth system modeling.
- We added two figures aimed at describing the changes in the mean state in a series of selected simulations.
- We improved the discussion of the results to better characterize the discussion of the climate variability.

You may find a detailed reply to the reviewer's comments in the following pages. We reported for completeness below each point in grey also the part of text that has been strongly modified. However, since many changes has been made extensively throughout the text, we attach to this reply also the new version of the manuscript.

Comments:
In the following I will address how the evaluation section could be improved and extended.

0. I would have liked to see some mean field comparison of temperature, wind, precipitation and in particular their variability.
The reviewer is right, the presentation of the results in the previous version was not sufficient. However, we must keep in mind that we cannot include in the present GMD manuscript the full description of the mean state: we have 5 different resolutions, decades of variables, stochastic physics schemes and so on. This would touch several topics and therefore makes the manuscript unnecessary long.
Hoping to provide a reasonable equilibrium between the length of the text and the usefulness of the information, we thus decided to add a section introducing the results where we show in a 6-panel plot the sensitivity of precipitation and upper tropospheric zonal wind (200 hpa) to resolution and stochastic schemes for a selected high resolution (T799) and a low resolution version (T255).
The two figures are reported for completeness here below, while the full description is now present in the new version of the manuscript.
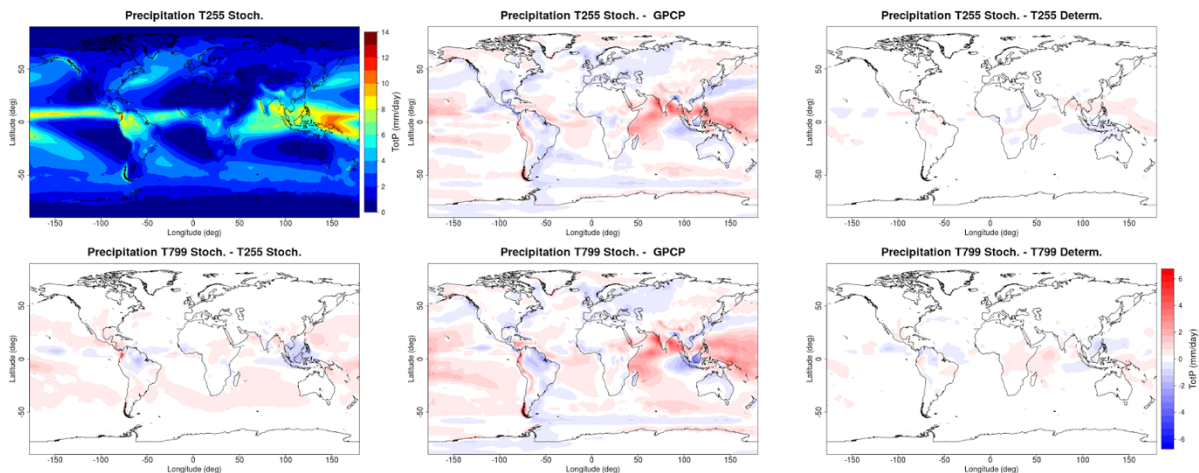

Figure R1. Upper left: Climatological ensemble mean precipitation for the PDA experiments (1979-2008) for T255 with stochastic physics. Lower left: T799 stochastic minus T255 stochastic precipitation. Central panels: T255 (upper) and T799 (lower) precipitation bias with respect to GPCP. Right panels: Stochastic minus deterministic climatological precipitation for T255 (upper) and T799 (lower).
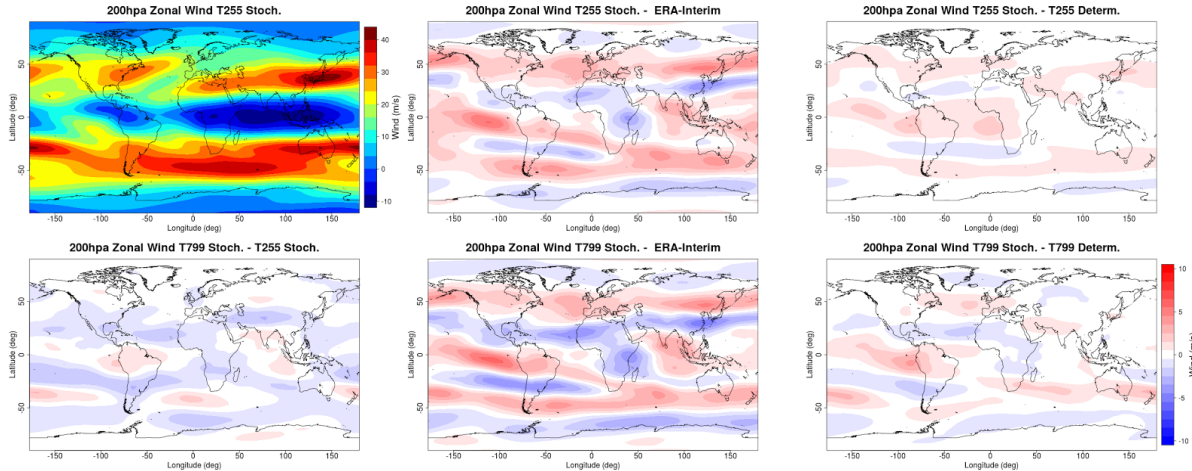
Figure R2: Same as R1 but for zonal wind at 200 hPa. Here bias is evaluated against ERA-Interim reanalysis.

The new discussion in the manuscript is here reported:

Although a detailed analysis of the mean climate in all the simulations performed would be excessively long to be included in the present work, we introduce a couple of figures showing the sensitivity to resolution and stochastic physics parameterization of the climatology of precipitation (Figure 5) and 200 hpa zonal wind (Figure 6). We compare the ensemble mean average fields of a low resolution version (T255) with a high resolution one (T799), in both its deterministic and stochastic configurations. Data have been interpolated on common 2.5°x2.5°grid.

The precipitation model bias - shown in Figure 5, with respect to Global Precipitation Climatology Project (GPCP) (Huffman et al., 2001) dataset - is especially strong in Indian Monsoon region, with an excess of precipitation that ranges from the Indian Ocean to the Western Pacific. More generally, EC-Earth tends to underestimate the precipitation over the continents and overestimate it over the oceans. When the comparison is carried out between stochastic and deterministic configurations, it is possible to see that SPPT and SKEB do neither improve nor deteriorate the climatology at both T255 and T799 resolutions.

Conversely, a slight more evident change is seen comparing the high and low resolutions: here T799 shows a widespread increase of the extratropical precipitation. But again, when it is evaluated against the model bias such changes are minor.

Impacts on the upper-tropospheric zonal wind field are clearer and they are shown in Figure 6. The T255 version - compared against ECMWF ERA-Interim reanalysis (Dee et al., 2011) - shows too strong jets in both the hemispheres. The subtropical jet over Asia and Pacific is also poleward displaced, while equatorial easterly jets are too weak. Again, stochastic physics bring minor changes, with a slightly stronger Atlantic jet, more penetrating over Europe. Conversely, the higher resolution leads to an overall weakening of the upper tropospheric winds: this is especially true over North America and the Tibetan Plateau, suggesting that this change may be induced by the stronger surface drag caused by the higher orography.

More generally, in these and other climatological fields (not shown) the impact of stochastic physics and resolution appears to be small if compared to the model bias. Indeed, larger benefits from increasing resolution and stochastic physics are expected more in terms of variability rather than in terms of mean state.

Therefore, in the following sections we will focus on a few selected features of climate variability. We will investigate the improvements and/or deteriorations following resolution increases and including stochastic parameterisation of three different phenomena: the distribution of the intensity of tropical rainfall, the tropical variability related to the Madden-Julian Oscillation and the mid-latitude variability associated with atmospheric blocking.

Tropical Rain:

1. The manuscript interprets the difference between the TRMM and GPCP datasets as observational uncertainty. The TRMM data should be added to the right panels displaying the frequency fractions with regard to GPCP.

We added in the Figure 5 TRMM data too. The new figures are reported here below.

Frequency distribution of daily-mean total precip
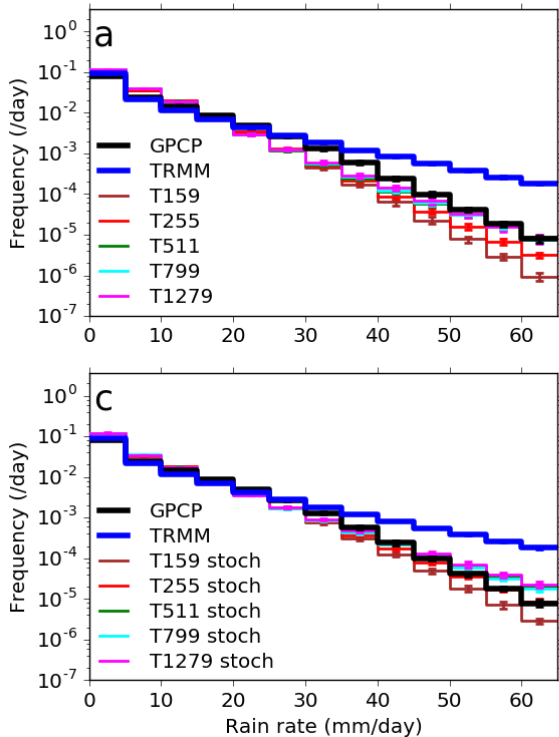
Frequency as fraction of that in GPCP
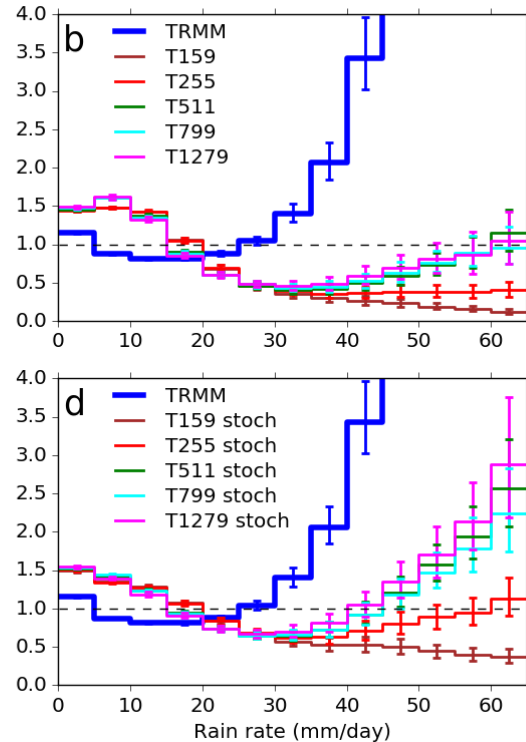
2. Since in the observational datasets the differences in the rainfall rates of 40mm/day are larger than the differences between the model simulations, a crude significance test would help with the interpretation of the results. This could be as simple as displaying rainfall rates for the first and second half of the simulation period.
In order to evaluate the uncertainties, we calculated 95% confidence intervals using a bootstrap method. These confidence intervals have been added on in Figure 5 to show the sampling uncertainty of the frequencies. The method is explained in the text:

Vertical bars in figure 7 show the 95% confidence intervals of the frequencies associated with sampling uncertainty. These were calculated using a bootstrap method. For each dataset, a surrogate dataset was created by randomly sampling individual years of data with replacement. The frequency distribution of the surrogates and their frequency ratios with respect to the GPCP surrogate were calculated. This was repeated 1000 times to produce the distribution of the calculations associated with sampling uncertainty, from which the confidence intervals were derived

3. I would suggest a second figure showing of rainfall rates and frequency fractions for the small rainfall rates only, so that the differences in the curves can be seen better.
We appreciate the suggestion, but in our view this would lead to there being too many figures, making the paper excessively long. This is especially true considering the new figures on the mean state. We have made sure that important details of the results for low rain rates are spelt out in the text.

MJO
4. Figure 6: Please specify how the patterns for the 4 phases of the MJO are computed. Are they based on all data, or are they calculated for the observations only? It seems the comparison of the frequency of occurrence and the amplitude of the MJO should

be done in a common basis system. Otherwise a case must be made, why different basis system are used and the patterns for each experiments (or at least some) shown and the change in patterns discussed.

Yes, the MJO index was computed for the period 1980-2001. We have mentioned this in the manuscript now. The observations are also for same period. The MJO is essentially a mode of internal variability in the atmospheric system and hence, our MJO analysis in this study is mainly focused on statistical differences between the model MJO variability and the observational variability based on the same period of evaluation.

Figure 8 shows the frequency of occurrence vs. the mean amplitude of the MJO in the four different regions around the Tropics for all the different runs (colours) and for ECMWF ERA-Interim Reanalysis (grey, Dee et al., 2011) over the 1980-2001 period.

5. Please refer to Weisheimer et al. 2014, who report that in particular SPPT improves MJO in System 4.

We thank the reviewer for this comment. We have now included a reference to the Weisheimer et al. 2014 result suggesting that the SPPT scheme helps improve MJO representation in the seasonal forecasting system.

6. Are these results significant? Again a dot for each first half and second half of the simulations would be sufficient.

We have now included error bars in this figure by computing the same statistic for periods half the length of the original analysis as suggested.
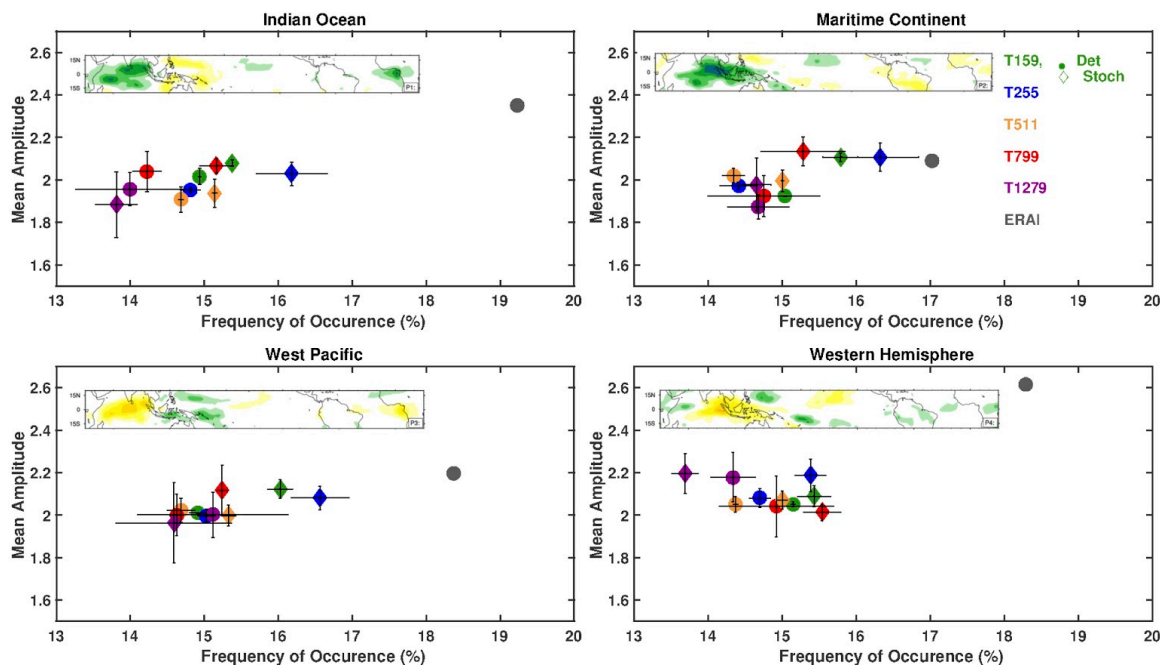

Figure R3: New version of MJO figure (new figure 8) from the manuscript.,

Blocking

7. Dawson et al. 2015 is in the Bibliography, but I couldn't find a reference in the text. This paper should be clearly cited in the blocking section.

Dawson and Palmer (2015) has been cited in the introduction. In addition, since as stated by the reviewer is relevant for the discussion on atmospheric blocking, it is cited also in the blocking and conclusion sections.

8. Other work has suggested a positive impact of stochastic physics on blocking. Why is this not the case here? Maybe it only helps at horizontal resolutions of less than T159? Is there really no benefit even at resolutions of T159?

It is hard to say it without a direct comparison of the diagnostics used: at this stage we can only provide some hypothesis.

A first reason may be exactly as stated by the reviewer, i.e. stochastic physics providing benefits for blocking at lower resolution than T159.

A second one could be related to the metrics adopted: for instance, blocking diagnosed from weather regimes (Dawson and Palmer, 2015) is not always associated with blocking as evaluated directly from blocking indices. In this sense, for instance, the instantaneous blocking index here reported does not take into account the duration of the events.

The last possibility can be deducted from Figure R4 here attached. It shows the full behaviour of the blocking diagnostic in all the ensemble members from Climate SPHINX. Dashed lines are the stochastic runs and solid ones are the deterministic ones. Bold lines mark the ensemble mean, suggesting that the two versions are extremely similar, indistinguishable from a statistical point of view. Indeed, the natural variability is extremely evident in this diagnostic. It is easy to find a pair of members where a deterministic run outperforms a stochastic one or vice versa.

It may be possible that some of the studies (e.g. Berner et al., 2012, Dawson and Palmer, 2015) reporting an improvement of blocking with stochastic schemes may be affected by sampling of the internal variability: one of the things that Climate SPHINX clearly highlights is that a very large number of ensemble members is needed if we want to draw any conclusion on the sensitivity of mid-latitude climate to the adoption of stochastic schemes.

We mentioned that in the manuscript now, including a comment in the blocking section and in the conclusions.

Conversely, in the mid-latitudes, where atmospheric blocking frequencies were analysed, no statistical difference is found between stochastic and deterministic runs. Previous works (Berner et al., 2012; Dawson and Palmer, 2015) suggest that block- ing regimes can benefit from stochastic schemes. We note that the simulations presented here are at a higher resolution than Berner et al. (2012) and have been analysed using a different metric to Dawson and Palmer (2015). Nevertheless, we must observe that our blocking diagnostic shows strong variability in the different ensemble members, suggesting that a single realization may be not enough to capture the real sensitivity of this diagnostic to stochastic schemes. On the other hand, we found that increased horizontal resolution seems extremely important to decrease the blocking bias: in agreement with other recent works (Davini and D'Andrea, 2016; Schiemann et al., 2016) this is true especially over the Euro-Atlantic sector - where the T799 resolution (~25 km) reduces it to negligible values - but not evident over the Pacific.
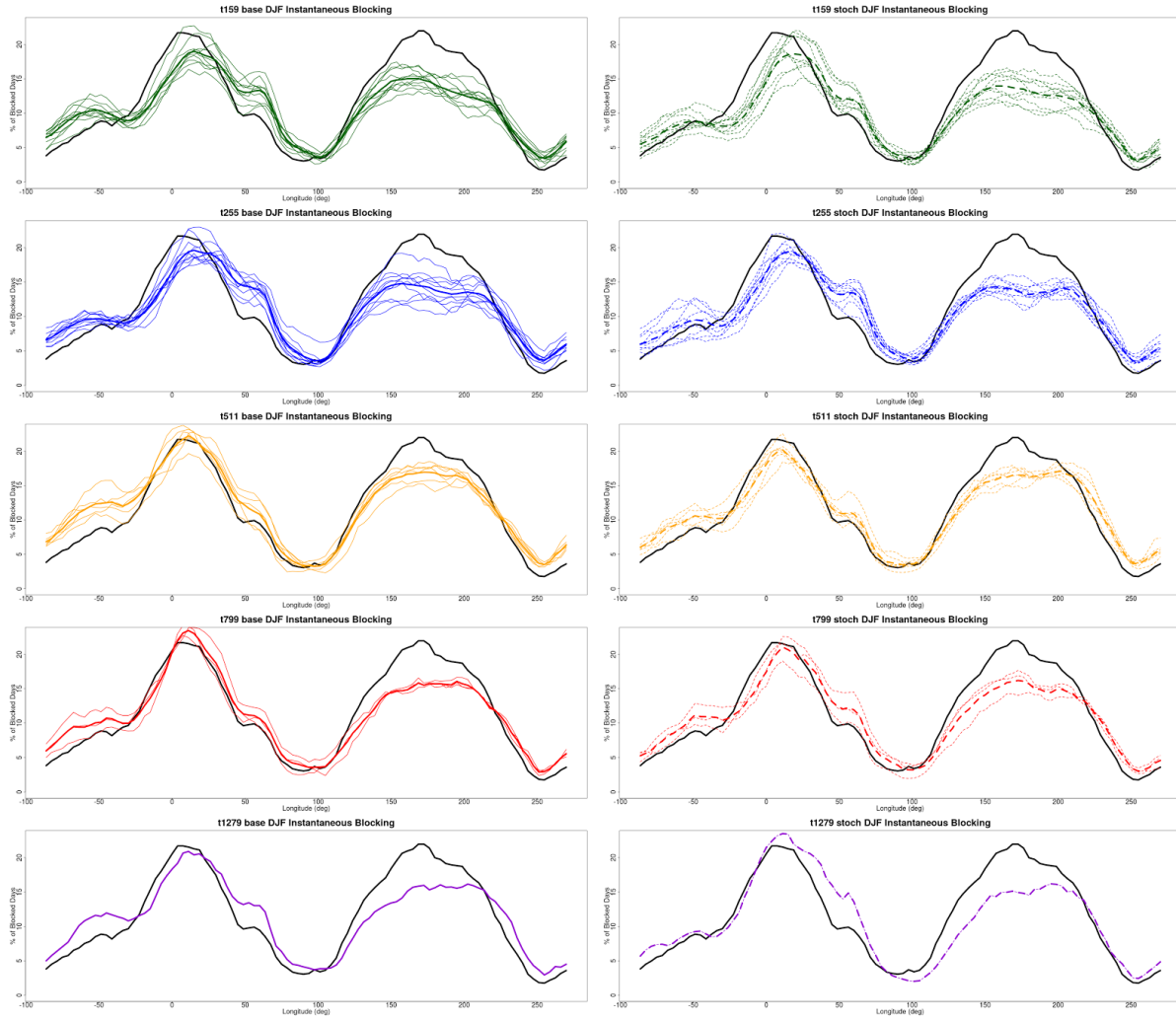
Figure R4: Winter locking climatologies for all the PDA simulations. Dashed lines are stochastic simulations, solid ones deterministic ones. Bold lines are ensemble means. Black line is the Reanalysis.

Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth Global Climate Model

Paolo Davini[1,2], Jost von Hardenberg[2], Susanna Corti[3], Hannah M. Christensen[4], Stephan Juricke[4], Aneesh Subramanian[4], Peter A.G. Watson[4], Antje Weisheimer[4,5], and Tim N. Palmer[4]

[1]Laboratoire de Météorologie Dynamique / IPSL, Ecole Normale Supérieure, PSL Research University, CNRS, Paris, France
[2]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Torino, Italy
[3]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy
[4]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK
[5]National Centre for Atmospheric Science (NCAS), University of Oxford, Oxford, UK

Correspondence to: Paolo Davini (pdavini@lmd.ens.fr)

Abstract. The Climate SPHINX (Stochastic Physics HIgh resolutioN eXperiments) project is a comprehensive set of ensemble simulations aimed at evaluating the sensitivity of present and future climate to model resolution and stochastic parameterisation. The EC-Earth Earth-System Model is used to explore the impact of stochastic physics in a large ensemble of 30-year climate integrations at five different atmospheric horizontal resolutions (from 125km up to 16km). The project includes more than 120 simulations in both a historical scenario (1979-2008) and a climate change projection (2039-2068), together with coupled transient runs (1850-2100). A total of 20.4 million core hours have been used, made available from a single year grant from PRACE (the Partnership for Advanced Computing in Europe), and close to 1.5 PBytes of output data have been produced on SuperMUC IBM Petascale System at the Leibniz Supercomputing Center (LRZ) in Garching, Germany. About 140 TBytes of post-processed data are stored on the CINECA supercomputing center archives and are freely accessible to the community thanks to an EUDAT Data Pilot project. This paper presents the technical and scientific setup of the experiments, including the details on the forcing used for the simulations performed, defining the SPHINX v1.0 protocol. In addition, an overview of preliminary results is given: an improvement in the simulation of Euro-Atlantic atmospheric blocking following resolution increases is observed. It is also shown that including stochastic parameterisation in the low resolution runs helps to improve some aspects of the tropical climate - specifically the Madden-Julian Oscillation and the tropical rainfall variability. These findings show the importance of representing the impact of small scale processes on the large scale climate variability either explicitly (with high resolution simulations) or stochastically (in low resolution simulations).

1 Introduction

The simulation and prediction of Earth's climate is one of the scientific and computational grand challenges. In order to make quantitative projections of future climate, it is necessary to use climate models that simulate all the important processes governing the evolution of the climate system. Over the past few decades, climate models have developed considerably - increasing both in complexity and resolution - as computational power has increased. Yet there is a notable difference in the horizontal

resolution of models used in operational numerical weather prediction (NWP) and those used for climate simulations in the fifth Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012). Atmospheric horizontal resolutions in operational NWP are in the range 16-40 km, whereas the resolution of CMIP5 climate models is (on average) coarser than 120 km.

It is well known that a typical climate model is unable to represent many sub-synoptic-scale systems, and only poorly represents smaller baroclinic features. Typically climate models underestimate the number of observed storms (Zappa et al., 2013) and poorly simulate the statistics of atmospheric midlatitude blocking (Anstey et al., 2013). In fact it has been shown (e.g. van Oldenborgh et al., 2012) that at standard (low) climate resolution, forecast systems have pervasive systematic errors, which impact on quasi-persistent weather regimes (Dawson et al., 2012) and, more generally, on temporal variability and regional patterns of the leading modes of variability (Delworth et al., 2012; Kinter III et al., 2013). On the other hand, recent experiments have shown that high-resolution climate models are significantly better at simulating important physical processes such as the global water cycle (Demory et al., 2014), as well as relevant features of the large scale atmospheric circulation such as the jet stream (Lu et al., 2015), the Euro-Atlantic blocking (Jung et al., 2012) and the Madden-Julian Oscillation (MJO, Peatman et al., 2015).

The fact that enhanced horizontal resolution in climate models can positively impact some aspects of the simulated large scale atmospheric circulation is further evidence of the role that small scale processes play in "shaping" large scale motions. However, it is unlikely that climate integrations at very high resolution (i.e. at the resolution used in NWP), will be feasible in the near future. There are numerous other areas of climate model development that compete for the given computing resources: for example, the need for ensembles of integrations, the need to integrate over century and longer time-scales, and the need to incorporate additional Earth System complexity. In addition, parameterisations which have been developed for coarse scales may need retuning or to be substituted with alternative parameterisations at higher resolutions, requiring a consistent development effort.

Instead of explicitly resolving small-scale processes by increasing the resolution of climate models, a possible alternative is to use stochastic parameterisation schemes. There has been significant progress in developing stochastic schemes over the last decade, primarily for use in medium-range and seasonal ensemble forecasts (e.g. Plant and Craig, 2008; Khouider et al., 2010; Bengtsson et al., 2013; Grell and Freitas, 2013; Dorrestijn et al., 2016; Sakradzija et al., 2016; Ollinaho et al., 2016). These schemes introduce an element of randomness into physical parameterisation schemes to account for the impact of uncertain, unresolved processes on the resolved scale flow (Palmer, 2012). Stochastic schemes have been shown to improve the reliability of probabilistic forecasts on medium range and seasonal timescales, as well as improving biases in the mean state.

There is mounting evidence that stochastic parameterisations can also prove beneficial for climate simulations (e.g. Lin and Neelin, 2000, 2003; Arnold et al., 2013). Berner et al. (2012) show that including stochastic physics can reduce systematic biases in the model's mean climate, comparable to improvements gained by increasing the model resolution. Several recent papers have also demonstrated that the variability of a climate model can significantly improve with the introduction of a stochastic physics scheme, with improvements observed in the representation of the Madden-Julian Oscillation (MJO) (Deng et al., 2015; Wang et al., 2016), the El Nino-Southern Oscillation (Christensen et al., 2016), and extra-tropical flow regimes

(Weisheimer et al., 2014; Dawson and Palmer, 2015; Christensen et al., 2015). As was the case for the mean state, the observed improvements can be similar to that observed on increasing the resolution of the model (Dawson et al., 2012).

These results highlight the influence of small-scale processes on large-scale climate variability, and indicate that although simulating variability at small scales is a necessity, it may not be necessary to represent the small-scales accurately, or even explicitly, in order to improve the simulation of large-scale climate. This issue is important in the light of the next CMIP6 project. In fact it seems quite unrealistic that in the near future, climate simulations at NWP resolution could be affordable. However resolutions around 40 km might be more feasible and indeed they are planned within the HighResMIP project (Haarsma et al., in review, 2016).

In the coordinated project Climate SPHINX (Climate Stochastic Physics HIgh resolutioN eXperiments), we use the EC-Earth Earth System Model (Hazeleger et al., 2010, 2012, http://www.ec-earth.org) to investigate the sensitivity of climate simulations to model resolution and stochastic parameterisations. A key aim of the study is to investigate the degree to which stochastic parametrisation schemes can be used as a computationally cheaper alternative to increased model resolution.

The experiments follow one historical and one scenario projection following CMIP5 specifications in AMIP configuration (i.e. atmosphere-only integrations forced with observed - for the past - and simulated - for the future - sea surface temperatures). The AMIP integrations have been carried out keeping constant the vertical resolution (91 levels) and exploring five different horizontal resolutions: (i) low ($\sim$125 km), (ii) moderate ($\sim$80 km), (iii) intermediate ($\sim$40 km), (iv) high ($\sim$25 km) and (v) very high ($\sim$16 km). Each integration is repeated with the joint implementation of two stochastic parameterizations: the Stochastically Perturbed Parameterisation Tendencies (SPPT) scheme (Palmer et al., 2009) and the Stochastic Kinetic Energy Backscatter (SKEB) scheme (Shutts, 2005; Palmer et al., 2009). In order to sample the natural variability, several ensemble members are produced for each configuration. The simulations (ii-iv) aim to investigate whether configurations with intermediate horizontal resolution are able to partially bridge the gap between very high and low resolution. In other words, a systematic comparison is useful to understand if a gradual increase in resolution will lead to a similar gradual improvement of climate simulations or whether there is a true passing threshold in resolution, which is required to get acceptable simulations of the main climate features.

By comparing integrations carried out at different resolutions we evaluate the impact of increased atmospheric horizontal resolution on the simulation of key climate processes and of climate variability. By comparing experiments with and without the implementation of stochastic physics we evaluate the impact of stochastic physics on the simulation of key climate process and of the associated climate variability when the model resolution is the same. By comparing experiments with the implementation of stochastic physics with experiments carried out without stochastic physics, but at higher resolutions, we assess to what extent the stochastic representation of the sub-grid processes can compete with a more refined horizontal resolution. The results of this project integrate with several other efforts currently underway (e.g. the EU-H2020 PRIMAVERA project). In particular this study complements groundbreaking past initiatives in pioneering the use of HPC for climate simulations such as the UPSCALE (Mizielinski et al., 2014) and the ATHENA (Kinter III et al., 2013) projects.

Climate SPHINX was made possible by a considerable amount of computing time provided by PRACE (the Partnership for Advanced Computing in Europe) and of data storage from EUDAT (the collaborative Pan-European infrastructure providing

research data services). We were granted 20 million core hours during a single year at SuperMUC, the IBM Petascale System at the Leibniz-Rechenzentrum (LRZ) in Garching near Munich, Germany. Storage of data produced by Climate SPHINX is secured by the EUDAT Pilot Project DATA SPHINX (DATA Storage and Preservation of High resolution climate eXperiments), which provides a widely accessible archive for medium term storage to facilitate data access and discovery. DATA SPHINX is managed by CINECA (the largest Italian computing centre) and at present hosts 140 TB of data generated by Climate SPHINX.

In this paper we describe in detail the important technical aspects of this project and highlight some preliminary scientific results on the impact of increased resolution and stochastic parameterisations on climate simulations. Model configuration and tuning are presented in Section 2, while the experimental setup is described in Section 3. Section 4 is devoted to detail the technical configuration. An overview of results and concluding remarks are reported in Sections 5 and 6.

2 The EC-Earth Global Climate Model

We use version 3.1 of the state-of-art atmosphere-ocean Earth System Model EC-Earth (Hazeleger et al., 2010, 2012). EC-Earth is based on the Integrated Forecast System (IFS, cycle 36r4) (ECWMF, 2009) atmospheric circulation model, developed by the ECMWF. In order to represent land surface dynamics, IFS integrates the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (H-TESSEL) land surface scheme (Balsamo et al., 2009).

When used in coupled mode, the NEMO 3.3.1 oceanic circulation model (Madec, 2008) which includes the LIM3 sea ice model (Vancoppenolle et al., 2012) is used. The atmospheric and oceanic components are coupled through OASIS3 (Valcke, 2013).

2.1 The IFS atmospheric model

IFS is the current atmospheric model developed at ECMWF, and given the predominance of atmosphere-only simulations run within SPHINX, it is the core of the project. The version that is part of EC-Earth is derived from cycle 36r4 and has been tuned and improved for climate purposes by the EC-Earth Consortium. IFS uses a combination of spectral and reduced Gaussian grids (where, in the latter, the number of longitudinal grid points decreases towards the poles). Physical parameterisations and advection are computed on the reduced Gaussian grid and then, using the spectral transform, semi-implicit time stepping is performed in the spectral space.

Traditionally the spectral harmonic at which truncation occurs defines the horizontal resolution: IFS uses a linear triangular truncation for which for a specified number of N harmonics retained corresponds to 2(N+1) grid points in the longitudes along the Equator. If the resolution is T255, this means that post-processed output will have 512x256 grid points on a regular Gaussian grid, which corresponds to a resolution of about 80 km at the equator.

For Climate SPHINX, the horizontal resolutions T159, T255, T511, T799 and T1279 have been explored. Conversely, the number of levels is fixed in all the configurations at 91: these are hybrid levels with the last full level at 0.01 hPa. The description of the main parameterisation schemes can be found in Beljaars et al. (2004): the parameterisations are in general independent

4

of resolution, with the only exception of the convective adjustment time, which decreases with increasing resolution as reported in Table 1.

2.2 The stochastic physics parameterisation schemes

We consider two complementary approaches to stochastic parameterisation, both developed at ECMWF for use in the IFS. The two schemes considered here are used operationally at weather and seasonal forecasting centres worldwide (Palmer et al., 2009; Yonehara and Ujiie, 2011; Bouttier et al., 2012; Sanchez et al., 2016), and so have been extensively tested in a range of models on medium range and seasonal timescales. Following the "seamless prediction" paradigm, we choose to test these schemes here on climate timescales.

The first approach is the Stochastically Perturbed Parameterisation Tendencies (SPPT) scheme (Palmer et al., 2009), which uses multiplicative noise to represent model uncertainty due to the parameterisation process. The use of multiplicative noise has been motivated through several coarse-graining studies (Shutts and Palmer, 2007; Shutts and Pallarès, 2014). SPPT can be expressed as

$$\frac{\partial X}{\partial t} = D + K + (1 + \mu e) \sum_i P_i, \tag{1}$$

where $\frac{\partial X}{\partial t}$ is the modelled total tendency in X. This is the sum of D, the dynamical tendency, K, the horizontal diffusion, and each P_i term, with P_i being the tendency from the ith physics scheme. The zero mean random perturbation, e, is constant in the vertical, but μ tapers the perturbation to zero close to the surface and in the stratosphere. The scheme acts on the tendencies of the physical fields (i.e. temperature, winds and specific humidity) resulting from the five main parameterisation schemes: radiation, turbulence and gravity wave drag, non-orographic gravity wave drag, convection, and large scale water processes. All variables are perturbed with the same random number.

The perturbation, e, is generated using a spectral pattern generator (Berner et al., 2009), ensuring that it smoothly varies in space, while the patterns evolve in time following an AR(1) process. The perturbation at each timestep is the sum of three independent random fields which represent uncertainties on different temporal and spatial scales. The fields have horizontal correlations of 500 km, 1000 km and 2000 km and temporal decorrelations of 6 h, 3 days and 30 days respectively, with associated standard deviations of 0.52, 0.18 and 0.06. The magnitude of the perturbation has been motivated through coarse-graining high-resolution model simulations (Shutts and Pallarès, 2014), and a recent coarse graining study has also provided justification for the noise temporal and spatial correlation scales (Christensen et al., 2017). While the smallest scale (500 km and 6 h) dominates on weather forecasting timescales, it is expected that the larger scales will also be important on climate time scales. The SPPT scheme requires no retuning with changing horizontal resolution: the same noise characteristics are used operationally at ECMWF across model resolutions. This is because the multiplicative nature of the scheme applied to the total of all parameterised physical tendencies results in perturbations to the model that scale automatically as the parametrised tendencies scale with resolution. The resolution dependence of the individual contributions from the parametrisation schemes is implicitly dealt with at the (deterministic) parametrisation level.

In contrast to SPPT, the Stochastic Kinetic Energy Backscatter (SKEB) scheme aims to represent a physical process that is otherwise absent from the model (Shutts, 2005; Palmer et al., 2009). Kinetic energy loss is common in models due to numerical integration schemes and the parameterisation process (Berner et al., 2009). To counteract this, the SKEB scheme represents upscale transfer of kinetic energy by randomly perturbing the streamfunction.

5   Similar to SPPT, the SKEB scheme uses a spectral pattern generator to generate a spatially and temporally correlated perturbation field which is added at each timestep to the deterministic streamfunction tendency,

$$\dot{\psi_p}(\phi, \lambda, t) = \dot{\psi_{dyn}}(\phi, \lambda, t) + f(\phi, \lambda, t),\tag{2}$$

where $\dot{\psi_p}$ is the streamfunction tendency after the perturbation, $\dot{\psi_{dyn}}$ is the tendency before perturbation and f is the additive perturbation field. The perturbation field is expressed in spherical harmonics, and each coefficient is evolved separately in

10   time following an AR(1) process. A tuning parameter for the SKEB scheme, the backscatter ratio, is set to increase following resolution increase (see Table 1, following practise at ECMWF) in order to improve the slope of the kinetic energy spectrum.

     The standard SPPT and SKEB schemes were designed for use at NWP timescales. When implemented in the EC-Earth climate model, it was found that the SPPT scheme resulted in a large imbalance involving the water cycle, with a negative P-E that was ten times larger than in the deterministic model associated with an anomalous latent heat flux and a negative

15   net surface flux of about 2 W/m^2. This arises because the SPPT scheme was not designed specifically to conserve water vapour, and resulted in a water vapour sink in the atmosphere. A fix has been implemented, requiring that the global average of the tendencies (i.e. winds, temperature and more importantly specific humidity) before and after the SPPT perturbation are conserved. The new scheme removes the imbalance in P-E, which is now equal to that in the run where the SPPT scheme is disabled. This fix has subsequently been implemented at ECMWF (Leutbecher et al., 2017).

20   Hereafter, SPHINX simulations where stochastic parameterisation is operational will be defined as "stochastic" runs, while simulation where the scheme is deactivated will be mentioned as "deterministic" runs.

2.3   Model tuning

With respect to the previous version (v3.0.1), EC-Earth 3.1 shows a reduced radiative imbalance and an improved hydrological cycle. However it still exhibits a cold bias in both its atmosphere-only and coupled configuration and a small imbalance in

25   precipitation minus evaporation (P-E). In a "present day" AMIP configuration, the 3.1 version is too cold, extracting heat from the underlying sea surface temperatures (SST) by about 1.5 W/m^2 and showing unrealistically high values for net SW and LW fluxes at TOA (around 243-244 W/m^2). Thus the first goal of the tuning was to provide reasonable radiative fluxes at TOA and at the surface for the standard deterministic version of the model (T255, see next paragraph for further description on the configurations adopted).

30   To improve the radiation budget, some of the convection and microphysical parameters from a more recent version of IFS (cy40r1) were retrieved. In addition to this, two standard tuning parameters have been modified (see Mauritsen et al., 2012): the entrainment rate for organised convection (ENTRORG) was reduced from $1.75 \cdot 10^{-4}$ to $1.5 \cdot 10^{-4}$, and the rate of conversion of liquid water to rain (RPRCON) was reduced from $1.4 \cdot 10^{-3}$ to $1.2 \cdot 10^{-3}$.

The optimal choice of tuning parameters provides reasonable fluxes at the TOA (around 240 W/m^2) and a positive flux at surface of about 0.6 W/m^2, in accordance with the best estimates from observations (Wild et al., 2013). It is important to note that the tuning of the radiative fluxes has been carried out only for the T255 deterministic model version: the radiative balance has not been tuned in the higher resolution or stochastic models. This ensures a clean comparison between simulations at different resolutions and with and without stochastic physics. If the model were retuned for each run, it is not possible to determine whether it is changing the tuning parameters, changing the resolution, or including stochastic physics that is responsible (Haarsma et al., in review, 2016). Including the SPPT scheme led to a negative bias in the surface heat fluxes of about 0.8 W/m^2, likely caused by a different distribution of the clouds.

The main radiative fluxes resulting after the complete tuning procedure are reported in Table 2. As shown in this table, the radiative balance of the model at higher resolution (and with stochastic physics) shows larger TOA SW and LW with increasing resolution. Net surface fluxes are highly variable, with higher values for coarser resolutions.

Further tuning has been performed in order to produce a realistic Quasi-Biennial Oscillation (QBO) at higher resolutions. The EC-Earth 3.1 non-orographic gravity waves scheme is characterised by a momentum flux that is continuously launched at in the mid-troposphere to simulate the effect of gravity waves. The latitudinal profile of this momentum flux governs the correct parameterisation of gravity waves: a too high amplitude of the momentum flux will disturb the QBO in equatorial zones, particularly at high resolutions, while a too low value will lead to unrealistic eddy-driven jets, especially in the southern hemisphere, where orographically induced wave drag is low. With the current latitudinal profile, the QBO was simulated only at standard resolution (T255 with 91 vertical levels). Following advice from ECMWF staff, a resolution-dependent parameterisation of non-orographic gravity wave drag replaced the version-dependent parameterisation present in EC-Earth 3.1 (an ad-hoc parametrization developed for the ECMWF System4 seasonal forecast system). Namely, instead of using a low momentum flux average value (GFLUXLAUN=0.02) with a positive Gaussian peak at 50°S, we use a higher value (GFLUXLAUN=0.0375) which is reduced with a Gaussian shape at the equator. This negative peak is slightly deeper for stochastic runs than for deterministic simulations to compensate the effect of the stochastic noise. The average value of the momentum flux was further reduced with increasing resolution (starting from T799) according to the ECMWF specification for IFS cy40r1 (see Table 1).

The new non-orographic gravity wave scheme - a standard in the current operational forecast ECMWF model - allows now the simulation of the QBO at all the resolutions explored in Climate SPHINX, without deteriorating the jet streams. Given these positive results, the new parameterisation will be implemented in the upcoming EC-Earth 3.2 version.

3 Science configuration: the SPHINX v1.0 protocol

The following sections describe the scientific configuration, including the simulations performed, the initial and boundary conditions, the Sea Surface Temperatures (SST) and Sea Ice Concentration (SIC) used that together define the SPHINX v1.0 protocol.

3.1 Climate SPHINX simulations

The SPHINX project comprises more than 120 different simulations, the majority of them were carried out in an atmosphere-only configuration. The experiments are run at five different horizontal resolutions: T159 (~125km), T255 (~80 km), T511 (~40 km), T799 (~25 km) and T1279 (~16 km), all with the same vertical configuration with 91 levels (L91). The different simulations are grouped into three main blocks: Present-day AMIP (PDA), Future Scenario AMIP (FSA) and Past-to-Future Coupled (PFC).

PDA and FSA are atmosphere-only simulations: 20 ensemble members are run at T159, 20 at T255, 12 at T511, six at T799 and two at T1279 for both PDA and FSA experiments. For each resolution, half of the ensemble members have the stochastic physics parameterisations activated. The number of ensemble members run and their resolution is also reported in Table 1.

The atmosphere-only experiments extend for 30 consecutive years, from 1979 up to 2008 for PDA, while FSA experiments are run from 2039 up to 2068.

PFC simulations are run with IFS at the T255L91 configuration, coupled with NEMO using the ORCA1 grid (a tripolar grid with resolution of 1°longitudinally and refinement to 1/3° at the Equator) with 46 vertical levels. Six ensemble members are run, three with the stochastic parameterisation active and three control members, from 1850 up to 2100.

3.2 Initial conditions

The Initial Conditions (ICs) in both the PDA and FSA experiments are taken from the ECMWF ERA-Interim Reanalysis (Dee et al., 2011) for 01/01/1979. A first experiment is run at each resolution for a few days, and it is used to create the ICs for the other experiments. For instance, for T255 experiments, the ICs for the 10 ensemble members are extracted using the midnight values (00:00) from each of the first 10 days respectively, and then reassigned to the 1st of January.

The same ICs are used also for FSA: in order to account for the land-surface adjustment to the new forcing, a 1-year spin up has been carried out for FSA (which is therefore starting from 2038).

For PFC simulations, given the different expected climatologies of integrations with/without stochastic physics, two 320-year spin-ups are carried out in coupled mode to equilibrate the ocean to the atmospheric forcing. Having spun up, three oceanic states - from spin-up year 300, 310 and 320 - are coupled with three different atmospheric ICs: these are run in coupled mode for a further ten years with fixed greenhouse gas (GHG) forcing for the year 1850. In this way the phase space distance between the simulations is 20 years and the atmosphere and land surface have had enough time to adjust to the new oceanic state.

3.3 Forcing and boundary conditions

Well mixed greenhouse gases (GHGs), stratospheric ozone and volcanic aerosol concentrations have been set according to the CMIP5 protocol (Moss et al., 2010; Taylor et al., 2012). Historical forcing is used for PDA experiments, whereas for the FSA experiments the high emission scenario (Representative Concentration Pathway 8.5, RCP8.5) is adopted. PFC simulations use the historical CMIP5 specification from year 1850 up to year 2005 included: after that, the forcing is taken from the RCP8.5 scenario. Albedo, land use and vegetation patterns are set using the standard configuration of EC-Earth 3.1, which uses a

MODIS-derived fixed climatological seasonal cycle for snow-free albedo and the Leaf Area Index. The average yearly solar irradiance was set at 1368.2 W/m^2 with intrannual variations, following the standard EC-Earth 3.1 setup. All the simulations of PDA and FSA experiments use this setup. For the PFC simulations interannual variations following CMIP5 prescriptions (i.e. the 11-year solar cycle) have been added.

3.4 Present-day SST and SIC

Given that both FSA and PDA simulations are atmosphere-only runs, a special effort has been taken to provide reliable SSTs in order to fully exploit the high resolution.

For PDA, SSTs have been obtained from the daily SST and sea ice concentration (SIC) HadISST2.1.1, a pentad based dataset with resolution of 0.25°x0.25° for SSTs (Kennedy et al., 2016) and 1°x 1° for SIC (Titchner and Rayner, 2014). These are bilinearly interpolated onto the required reduced Gaussian grid for each resolution: climatologies for SST and SIC for the 1979-2008 period can be seen in the upper panels of Figure 1.

A number of inconsistencies are found between the land-sea mask of IFS and of HadISST2.1.1: these are due to slightly different coastlines and a different representation of the lakes. For the different coastlines, linear extrapolation from HadISST2.1.1 has been performed. For the interior (i.e. lakes), a methodology similar to the one used in ERA20CM dataset (Hersbach et al., 2015) has been adopted: one-month lagged 2m temperatures from the ERA-Interim monthly climatology of 1979-2008 are used as SST. Where the temperature is below zero, SIC is set to one, otherwise it is left at zero. This is interpolated in time on a daily basis and in space on the needed grid to create a smoothed seasonal cycle for lakes.

3.5 Future scenario SST and SIC

The creation of SST and SIC for the FSA experiment is more complex. We would like to account for the high-resolution variability provided by the HadISST2.1.1 but also consider the mean change and trend seen in a climate model. Therefore the SST for the future scenario have been obtained as a combination of HadISST2.1.1 variability and the CMIP5 EC-Earth simulations ensemble mean.

Firstly, the 1979-2008 HadISST2.1.1 SST has been detrended point by point to provide a set of anomalies with realistic variability. Secondly, the monthly seasonal cycle of the difference between the CMIP5 EC-Earth RCP8.5 ensemble mean over 2038-2068 (10 members) and the CMIP5 EC-Earth Historical ensemble mean over 1979-2008 (10 members) has been computed. This provides for each grid point the average expected SST increase from the present day to the future period according to a GCM, as a function of calendar month. To account for changes in SST during the FSA period, for each grid point the trend in SST from the CMIP5 EC-Earth RCP8.5 integration for 2038-2068 was also extracted. All CMIP5 EC-Earth data were bilinearly interpolated in space on the HadISST2.1.1 grid and linearly in time to daily frequency.

Finally, a new Step1HadISST dataset has been created combining the detrended HadISST2.1.1 (expressing the high-resolution daily variability), the average daily change of CMIP5 EC-Earth (from RCP8.5 and Historical, expressing the expected average temperature increase) and the linear trend of the CMIP5 EC-Earth RCP 8.5 (expressing the expected future trend in SST). The

9

methodology used here, that shares the main characteristics with the method developed by Mizuta et al. (2008), is sketched in Figure 2.

However, the Step1HadISST reconstruction misses an important element: there is no information on the sea ice cover in the future. To account for this, we took data from the CMIP5 EC-Earth simulations as a reference for SIC. CMIP5 EC-Earth simulations show a considerable cold bias in SST with respect to HadISST2.1.1, but they show good ice coverage, especially for the Northern Hemisphere (see average NH and SH hemisphere SIC in Figure 3).

Considering that an ensemble mean would be unrealistic, especially for a field with a large spatial variance as sea ice, we select a single ensemble member representative of the ensemble. Member "r8i1p1" has been chosen to characterise the ensemble, since its climatology shows the smallest SIC Root Mean Square Error (RMSE) when compared to the ensemble mean climatology in the time window 2038-2068. Clearly, using RMSE is only one of the possible metric to perform such selection: our main goal is to pick an ensemble member that is not an outlier when compared to the other EC-Earth CMIP5 ensemble members.

As a last step, we must evaluate SST for points where SIC coverage has disappeared in the future scenario. The lack of information about historical SST under sea ice results in undefined SST at these points using the methodology outlined above. We define these as "bare point". For bare points, we want to make use of the model variability, but we do not want to have inconsistent SST at the boundaries (i.e. where bare points border the Step1Hadisst dataset).

Initially, we perform a linear extrapolation for bare points for Step1HadISST SSTs - which gives us a measure of the average SSTs at the bare points. However, these extrapolated values are missing a realistic spatial variability. We then mask the bare points also in the SST field of the CMIP5 EC-Earth ensemble member "r8i1p1", and we subsequently linearly extrapolate new values. We then subtract from the original field of CMIP5 EC-Earth ensemble member "r8i1p1" these new extrapolated values, in order to obtain an anomaly field which includes the spatial and temporal variability of the SST field over the bare points given by CMIP5 EC-Earth "r8i1p1". This final field is then added to the linearly extrapolated Step1HadISST SST.

Hence, for each day, the SSTs for bare points are given by the EC-Earth CMIP5 RCP8.5 ensemble member "r8i1p1" SST minus extrapolated EC-Earth CMIP5 RCP8.5 SSTs plus extrapolated Step1HadiSST. The methodology to obtain this specific SST reconstruction is illustrated in Figure 4.

This provides a pattern of SSTs physically consistent with SICs: indeed, it avoids unrealistic values of SSTs in the proximity of the polar cap during winter and - using the CMIP5 EC-Earth data - it provides a reasonable distribution of SSTs in summer, where in the future scenario the ice coverage in the NH often disappears. Moreover, there is no discontinuity at the border with Step1HadiSST. The new dataset is defined as FutureHadiSST2.1.1.

The same methodology used for the PDA simulations has been adopted also for FSA runs in order to solve the issues of the lakes and the different land sea-mask: however, in this case we must account for the estimated temperature change over land. We consider the difference between the one-month lagged 2m surface temperature from CMIP5 EC-Earth RCP8.5 and the one-month lagged 2m surface temperature from CMIP5 EC-Earth historical ensemble (averaged over 8 members). We then add this to the one month lagged 2m surface temperature ERA-Interim monthly climatology of 1979-2008. This, analogous to what done for SSTs, accounts for climate change.

The SST and SIC changes between FutureHadiSST2.1.1 and HadISST2.1.1 are reported in Figure 1: as expected larger warming and sea ice retreat is seen in the Northern Hemisphere high latitude. Figure 3 reports the timeseries and trends for SST (between 45°S and 45°N) and SIC (for both Northern and Southern Hemisphere) for both FutureHadiSST2.1.1 and HadISST2.1.1.

4 Technical configuration

4.1 High-Performance Computing details

Simulations have been run on the 6.8-Petaflop SuperMUC IBM Petascale System at LRZ. The initial setup and configurations have been performed on the Supermuc-I platform, based on Sandy Bridge-EP Xeon E5-2680 8C processors. For processor decomposition the Message Passage Interface (MPI) parallelism paradigm has been used. EC-Earth allows also for OpenMP/Shared memory parallelisation, which has been tested without showing any significant computational benefit.

An accurate scaling of the performance was performed during the first months of the simulations. However, a conservative choice has been undertaken, after considering that the walltime needed to run the simulations was not the main concern for the project success. The number of cores assigned to each experiments have been selected following the resolution of the model considered. Although stochastic physics experiments showed about 5-10% decrease in performance (according to different resolutions), the same number of cores has been retained.

In summer 2015 a new Supermuc-II platform based on Haswell Xeon Processor E5-2697 v3 processors was made available by the LRZ. The new HPC granted a reduction of about 5% of the total core hours used, without affecting the walltime. About 75% of the simulations have been run using the Haswell nodes. Details on the processor decomposition, computational costs and data outputs are reported in Table 3.

4.2 Data output and postprocessing

In Climate SPHINX, IFS has been set up to provide output in GRIB format every 3 hours: however, the four T1279 simulations alone sum up to about 200 TB of raw data output. Summing together the restarts files and the output of all experiments the total amount of space occupied at the peak of the project (February 2016) reached about 1 PB. In order to reduce the size of the output and to increase the data accessibility to a larger audience, automatic post-processing routines have been implemented. At the end of each simulation leg, a script aimed at post-processing is launched: the script handles both the spectral and reduced Gaussian data from IFS and extracts and converts the requested variables from the default ECMWF format to a user-friendly, CMOR-like format on regular Gaussian grid. With this automatic procedure, more than 140 TB of post-processed data has been produced. A significant reduction of the data volume was obtained making use of the NetCDF-4 Zip format (HDF5).

Monthly (MON), daily (DAY), 6-hour (6HRS) data for different subsets of variables has been produced. More than 50 fields have been stored at monthly frequency. In order to further reduce the space requirements, daily and 6-hour 3D fields have been degraded to the spectral resolution of T255. Additional data at 3-hour frequency have been stored for the Euro-Cordex domain

(3HRS-CDX) and for a sub-domain including India Tibet and Pakistan (3HRS-ITP). Total precipitation has been saved also over the global grid at full resolution with 3 hours frequency (3HRS). Finally, synoptic monthly means have been stored for the main radiative variables (SMON). A few fields that are nonlinear functions of the output (e.g. specific humidity) have been computed from the original 3-hour output and then averaged at the required frequency in order to record them accurately. In

5 addition to the atmospheric data, about 10 TB of oceanic output have been stored for PFC simulations. Data at daily and pentad frequency have been retained.

All the data, including raw output, post-processed data and restart files, have been archived on the tape archives of the Tivoli Storage Management Infrastructure (TSM) of the LRZ.

5 Results overview

10 In this section we present a brief overview of the preliminary findings of the Climate SPHINX project. Considering that the number of diverse climate aspects that could be analyzed in such a large dataset is large, we decided to present hereafter only a few selected features of the climate and its variability. For all the results presented hereafter - if not specified differently - the complete set of ensemble members available for the present-day climate (i.e. PDA experiments) has been used.

5.1 Mean climate

15 Although a detailed analysis of the mean climate in all the simulations performed would be excessively long to be included in the present work, we introduce a couple of figures showing the sensitivity to resolution and stochastic physics parameterization of the climatology of precipitation (Figure 5) and 200 hpa zonal wind (Figure 6). We compare the ensemble mean average fields of a low resolution version (T255) with a high resolution one (T799), in both its deterministic and stochastic configurations. Data have been interpolated on common 2.5°x2.5°grid.

20 The precipitation model bias - shown in Figure 5, with respect to Global Precipitation Climatology Project (GPCP) dataset (Huffman et al., 2001) - is especially strong in Indian Monsoon region, with an excess of precipitation that ranges from the Indian Ocean to the Western Pacific. More generally, EC-Earth tends to underestimate the precipitation over the continents and overestimate it over the oceans. When the comparison is carried out between stochastic and deterministic configurations, it is possible to see that SPPT and SKEB do neither improve nor deteriorate the climatology at both T255 and T799 resolutions.

25 Conversely, a slight more evident change is seen comparing the high and low resolutions: here T799 shows a widespread increase of the extratropical precipitation. But again, when it is evaluated against the model bias such changes are minor.

Impacts on the upper-tropospheric zonal wind field are clearer and they are shown in Figure 6. The T255 version - compared against ECMWF ERA-Interim reanalysis (Dee et al., 2011) - shows too strong jets in both the hemispheres. The subtropical jet over Asia and Pacific is also poleward displaced, while equatorial easterly jets are too weak. Again, stochastic physics bring

30 minor changes, with a slightly stronger Atlantic jet, more penetrating over Europe. Conversely, the higher resolution leads to an overall weakening of the upper tropospheric winds: this is especially true over North America and the Tibetan Plateau, suggesting that this change may be induced by the stronger surface drag caused by the higher orography.

More generally, in these and other climatological fields (not shown) the impact of stochastic physics and resolution appears to be small if compared to the model bias. Indeed, larger benefits from increasing resolution and stochastic physics are expected more in terms of variability rather than in terms of mean state.

Therefore, in the following sections we will focus on a few selected features of climate variability. We will investigate the improvements and/or deteriorations following resolution increases and including stochastic parameterisation of three different phenomena: the distribution of the intensity of tropical rainfall, the tropical variability related to the Madden-Julian Oscillation and the mid-latitude variability associated with atmospheric blocking.

5.2 Tropical rainfall variability

Climate models generally have too little tropical variability on timescales of several days (e.g. Hung et al., 2013). One aspect of the variability of particular interest is the occurrence of heavy precipitation events, which can result in flooding, affect disease incidence and reduce crop yields (IPCC, 2015). Changes in the frequency of these events can also affect trends in total precipitation due to non-linearity in land surface processes (Saeed et al., 2013).

Figure 7a shows the frequency distribution of daily-mean precipitation rates averaged over $2.5°$x$2.5°$ grid boxes between $10°$S-$10°$ N over the period 1998-2008 in data from GPCP, data from the Tropical Rainfall Measuring Mission (TRMM) 3B42 Version 7 product (Huffman et al., 2007) and one ensemble member for each PDA run. Figure 7b shows the ratio of the frequency in each rain rate interval as a fraction of that in GPCP for each resolution.

Vertical bars in figure 7 show the 95% confidence intervals of the frequencies associated with sampling uncertainty. These were calculated using a bootstrap method. For each dataset, a surrogate dataset was created by randomly sampling individual years of data with replacement. The frequency distribution of the surrogates and their frequency ratios with respect to the GPCP surrogate were calculated. This was repeated 1000 times to produce the distribution of the calculations associated with sampling uncertainty, from which the confidence intervals were derived.

At all resolutions, rain rates below 15 $\mathrm{mm/day}$ occur too often in the model data, by about 50%, and rain rates between 20-60 $\mathrm{mm/day}$ occur too infrequently compared to both observational datasets. At rain rates near 30 $\mathrm{mm/day}$, the simulated frequencies are between about 35-50% of the frequency in GPCP. At higher rain rates, the frequency differences between TRMM and GPCP become comparable in size to or larger than the differences between the modelled frequencies and the observational datasets. We do not know of a reason to strongly prefer one dataset over the other, so we consider the model bias to be uncertain at these rain rates. The frequency of rain rates above 30 $\mathrm{mm/day}$ in the T159 and T255 models are below about 40% of that in GPCP. At T511, T799 and T1279 the relative frequency difference compared to GPCP and TRMM decreases as the rain rate increases, and becomes comparable to that in GPCP in the 60-65 $\mathrm{mm/day}$ interval, though still much smaller than that in TRMM. Therefore increasing the model resolution from T159 to T511 improves the simulated frequency of heavy rainfall events compared to observational datasets, with the further improvements caused by increasing the resolution to T799 or T1279 being considerably smaller.

Figures 7c,d show the same data for the stochastic PDA runs. Stochastic physics has a similar effect at all resolutions. Frequencies of rain rates between 5-15mm/day are reduced by about 10% compared to those in the deterministic models,

reducing the model bias. Frequencies above about ~20 mm/day are substantially increased, by a larger factor at larger rain rates, up to a factor of ~2.5 at rain rates around 60 mm/day. This reduces the difference from GPCP and TRMM up to rain rates of 45 mm/day at all resolutions.

The higher resolution stochastic models have rain rate frequencies between those of GPCP and TRMM at rates above 45 mm/day, so they seem consistent with the observations given the observational uncertainty. The T255 stochastic model has rain rate frequencies closer to those in GPCP than any of the deterministic models in all but two of the 5 mm/day rain rate intervals shown. One hypothesis to explain this effect is that the stochastic perturbations sometimes increase the moistening tendency of the air, so that it occurs more often that there is a high amount of water vapour in the air and heavier rain events can occur, and there is a compensating decrease in the frequency of moderate rain events.

Therefore stochastic physics brings this aspect of the simulations into better agreement with observations, suggesting that including a representation of unresolved variability and model error is important for simulating the statistics of extreme tropical precipitation events.

5.3 The Madden-Julian Oscillation variability

The Madden-Julian Oscillation (MJO) is the dominant mode of variability in the tropical region on sub-seasonal timescales (Madden and Julian, 1994). It is characterised by a strong interaction between tropical convection and the large-scale environment, manifest as a coherent eastward propagating pattern of precipitation followed by subsequent rainfall suppression (Khouider et al., 2011; Zhang et al., 2013; Kim et al., 2014; Raymond et al., 2015). It is a challenge for the current generation of global climate and weather models to represent the dynamics and thermodynamics of the MJO realistically (Slingo et al., 1996; Lin et al., 2006; Kim et al., 2009; Sperber et al., 2011; Klingaman et al., 2015).

We here use the Wheeler and Hendon (2004) technique to identify the dominant modes of variability in zonal winds and Outgoing Longwave Radiation (OLR) in these model runs. Combined Empirical Orthogonal Functions (CEOFs) of intraseasonal OLR, U850 (zonal winds at 850 hPa) and U200 (zonal winds at 200 hPa) are computed for each of the runs. The first two leading modes (RMM1 and RMM2) correspond to MJO signatures in the tropical wind field and OLR. The amplitude A of the MJO is defined as:

$$A = \sqrt{(RMM1)^2 + (RMM2)^2} \tag{3}$$

and the phase Φ of the MJO is defined as:

$$\Phi = \tan^{-1}(\frac{RMM2}{RMM1}) \tag{4}$$

MJO occurrence is defined when the MJO A is > 1. Conventionally, eight phases of the MJO are defined (Gottschalck et al., 2010). We reduce the eight phases to four phases corresponding respectively to the MJO being active in the Indian Ocean, Maritime Continent, West Pacific and Western Hemisphere. We note that the Wheeler Hendon Realtime Multivariate MJO

(RMM) index has been shown to be deficient in detecting MJO events when largescale circulation signals of the MJO are missing (Straub, 2013). Figure 8 shows the frequency of occurrence vs. the mean amplitude of the MJO in the four different regions around the Tropics for all the different runs (colours) and for ECMWF ERA-Interim Reanalysis (grey, Dee et al., 2011) over the 1980-2001 period.

5   Overall the frequency of occurrence of the MJO in the different regions in the Tropics for the different model resolutions is underestimated with respect to that of ERA-Interim. The MJO amplitude in the model simulations is lower than reanalysis over the Indian Ocean and the Western Hemisphere.

More importantly, increasing horizontal resolution does not seem to improve the representation of the phenomenon significantly. This may be explained considering that the simulation of the MJO in GCMs is influenced primarily by the representation
10 of mesoscale dynamics and of convection. The simulation of mesoscale dynamics can be helped by increasing the resolution, while improvements in convection are driven by changes in physical parameterisations of the model. Yet, the coupling between the mesoscale dynamics and the convection is key for convectively coupled waves in the tropics (Raymond et al., 2015). Therefore, increasing resolution alone may not be sufficient to improve the simulation of the MJO.

Conversely, the stochastic physics parameterisation improves the MJO frequency in all regions at all resolutions but T1279.
15 It must be noted that this latter run was done with only one ensemble member compared to the other runs with 3 or more ensemble members over the same period, therefore a sampling error due to natural variability should be considered. Above all, the best results are obtained for the T255 with stochastic physics, suggesting that the tuning of the mean state of the model might play a relevant role for a better MJO simulation.

Additionally, the stochastic physics climate runs show an improvement in the representation of the MJO propagation over
20 the Maritime continent (not shown). The lack of propagation of the MJO over the Maritime continent into the West Pacific region is a known problem in GCMs (Zhang et al., 2013). An improvement in the MJO propagation past the Maritime continent due to SPPT has also been seen in the ECMWF seasonal forecasting system 4 (Weisheimer et al., 2014). In this study, they show that the ECMWF system 4 has increased number of MJO active days in each phase of the MJO when SPPT is active as compared to runs without stochastic physics but still has less total number of active MJO days than the observations. Such
25 improved propagation in stochastic runs indicates either that there is an impact of the stochastic physics on the mean state in the region or that the variability in the region helps maintain the intraseasonal signal. The reasons for the change in MJO representation due to stochastic physics will be explored further in a more detailed future study by the authors.

5.4 Mid-latitude atmospheric blocking variability

One of the most important challenges for the current generation of climate models is the simulation of atmospheric blocking
30 (Anstey et al., 2013; Masato et al., 2013; Dunn-Sigouin and Son, 2013; Davini and D'Andrea, 2016). Blocking is a recurrent weather pattern typically occurring at the exit of the Atlantic and Pacific jet stream, more frequently during the winter season but observed throughout the year (Rex, 1950; Tibaldi and Molteni, 1990). It is characterised by a high-pressure, long-lasting low vorticity anomaly that "blocks" the mid-latitude westerly flow, diverting synoptic disturbances poleward or equatorward

(Tyrlis and Hoskins, 2008; Davini et al., 2012). A blocking event can last several days or even weeks, and it may be associated with cold spells in winter and heat waves in summer (Sillmann et al., 2011; Dole et al., 2011).

Blocking here is diagnosed using the simple index introduced by D'Andrea et al. (1998), an extension of the more known Tibaldi and Molteni (1990) index. This 1-d blocking index detects the reversal of the zonal flow measuring the geopotential height gradient at 500 hPa at 60°N, providing a binary blocking timeseries for each longitude. Although there is some evidence (e.g. Berner et al., 2012; Dawson and Palmer, 2015) that stochastic physics may improve the blocking simulation, with the current diagnostic no statistically significant difference emerges - even at low resolution - when comparing deterministic and stochastic simulations. Therefore the two simulations are combined together to provide an unique ensemble.

The upper panel of Figure 9 shows the blocking frequency for the ERA-Interim Reanalysis (black) and the ensemble mean of the different horizontal resolutions (colours) of PDA experiments over the DJF period. The common negative bias over the Atlantic and Pacific basins is clearly evident. Increasing the horizontal resolution leads to benefits over both the basins, with marked improvements especially for the Atlantic: here T799 and T1279 runs show values comparable to the reanalysis. The largest improvement is however seen upgrading from T255 to T511, where the bias - measured as the relative difference between the blocking frequency averaged between 10°W and 30°E - is reduced from the 18% to 3%.

Those clear improvements in blocking frequency are interestingly reflected by a change in the mean state. A simple way to represent the flow variability is to highlight a few isopleths of geopotential height, as done in the lower panel of Figure 9. Indeed, the higher resolution models show a strengthened pattern of the dominant Northern Hemisphere planetary waves, with marked ridges over the Rockies and Europe. Especially the former over the Rockies (Brayshaw et al., 2009) suggests an important role of orography resolution in the representation of the eddy-driven jet stream, and indirectly, of Euro-Atlantic blocking frequencies.

Indeed, the reduction of the bias following resolution increase for winter Atlantic blocking (and not for Pacific blocking) seems to be a common feature of several GCMs (Davini and D'Andrea, 2016; Schiemann et al., 2016). Such improvements have been associated with both better resolved transient eddy activity - that should sustain the blocking persistence (Shutts, 1983; Berckmans et al., 2013) - and with higher orography variance - which affects the mean state through planetary waves shaping (Jung et al., 2012; Berckmans et al., 2013). Conversely, Pacific blocking has been shown to be phenomenologically different (Pelly and Hoskins, 2003; Davini et al., 2012) and to be strongly affected by tropical dynamics (e.g. Renwick and Wallace, 1996), therefore it is not surprising that the latter would be less affected by horizontal resolution changes.

A more detailed analysis of blocking and mid-latitude variability will be carried out by the authors in future studies.

6 Conclusions

In the present work we have described the scientific configuration and technical setup/tuning of the EC-Earth Earth System Model used for the Climate SPHINX project, which defines the SPHINX v1.0 protocol. More than 120 climate simulations have been produced making use of more than 20 million core hours and generating about 140 TB of post-processed data. Climate SPHINX includes both present day (PDA simulations, 1979-2008) and future scenario (FSA simulations, 2038-2068)

atmosphere-only simulations according respectively to CMIP5 historical and RCP8.5 forcing. These have been run at five different horizontal resolutions - spanning from 125 km up to 16 km - with several ensemble members. Furthermore, a smaller set of transient coupled simulations (PFC simulations, 1850-2100) at T255 ORCA1 (∼80 km for the atmosphere and about 1°for the ocean) has been run.

5   Each deterministic experiment included in Climate SPHINX has a counterpart where the sub-grid unresolved scales have been parameterised with two different stochastic physics schemes (namely the SPPT and SKEB schemes). This makes Climate SPHINX the first climate dataset which includes a large number of ensemble members with a stochastic parameterisation at different horizontal resolution: along with other high-resolution simulation campaigns such as UPSCALE (Mizielinski et al., 2014) or ATHENA (Kinter III et al., 2013), this demonstrates the ability of the climate community to exploit the more recent 10   HPC machines.

Details on the tuning procedure (aimed at providing a correct radiation budget in the standard configuration T255) have been presented. Moreover, a comprehensive description of the methodology adopted for the creation of the present day and future scenario SST and SIC (starting from the HadISST 2.1.1 dataset) has been described. A novel method aimed at estimating SST where SIC have disappeared in future climate simulations has been introduced.

15   More importantly, Climate SPHINX post-processed outputs are freely accessible to the climate community. This has been possible thanks to an EUDAT pilot project which makes available a THREDDS server operational at CINECA from which data can be easily downloaded.

Preliminary results show the importance of both resolution and stochastic perturbations on the representation of the climate variability, although different phenomena show different sensitivities. Tropical rainfall variability seems to benefit from both 20   increased horizontal resolution and stochastic parameterisation, whereas the Madden-Julian Oscillation shows improvements only when the stochastic perturbations are added. In general - in the tropics - applying stochastic schemes at low resolution leads to interesting improvements: on the other hand, increasing resolution beyond T511 does not seem to further improve the tropical variability.

Conversely, in the mid-latitudes, where atmospheric blocking frequencies were analysed, no statistical difference is found 25   between stochastic and deterministic runs. Previous works (Berner et al., 2012; Dawson and Palmer, 2015) suggest that blocking regimes can benefit from stochastic schemes. We note that the simulations presented here are at a higher resolution than Berner et al. (2012) and have been analysed using a different metric to Dawson and Palmer (2015). Nevertheless, we must observe that our blocking diagnostic shows strong variability in the different ensemble members, suggesting that a single realization may be not enough to capture the real sensitivity of this diagnostic to stochastic schemes. On the other hand, we found 30   that increased horizontal resolution seems extremely important to decrease the blocking bias: in agreement with other recent works (Davini and D'Andrea, 2016; Schiemann et al., 2016) this is true especially over the Euro-Atlantic sector - where the T799 resolution (∼25 km) reduces it to negligible values - but not evident over the Pacific.

To summarise, the best improvements are observed on upgrading from T255 to T511, whereas minor improvements are observed using higher resolutions. This may be associated with the absence of specific tuning for both deterministic and 35   stochastic higher resolution configurations, which can affect the mean climate and consequently partially deteriorate the climate

17

variability. Indeed, such tuning does not involve only the surface and TOA radiative fluxes, but also dynamical components of the climate model. Some schemes, e.g. deep and shallow convection parameterisations, may be satisfactory at coarse resolutions but may perform poorly at finer ones.

Given the similarities between the dynamical cores of climate models , since they are all based on a controlled discretization of the same governing equations, we hope that the resolution sensitivity aspect of Climate SPHINX will be useful to the whole climate modelling community. On the other hand, several promising stochastic schemes exist, and the sensitivity of EC-Earth to SPPT and SKEB described here cannot be easily extrapolated to these alternative approaches. Nevertheless, considering that Climate SPHINX is the first large experiment where stochastic schemes are used massively on the climate time range, we hope that this work paves the way for other climate-oriented simulations aimed at investigating the impact of different stochastic schemes on climate variability.

Furthermore, Climate SPHINX focuses attention on the controversial choice between increasing resolution or increasing the size of ensembles - whilst keeping the same computing time available (e.g. Buizza et al., 1998). Indeed, running 30 years of one member at T1279 on the SuperMUC Petascale System costs about 1.4 million core hours; with the same amount of time it would be possible to run 9-10 simulations at T511. However, the benefits of the two pathways may be different: while a single member with 16-km resolution can provide local information at a topographic scale, useful for instance for hydrological models - particularly in areas with complex topography - in contrast many ensemble members at 40-km resolution can provide a correct assessment of the natural variability, a key element for instance for mid-latitude climate (Deser et al., 2012; Kay et al., 2015). However, we must keep in mind that the computational constraints would become particularly relevant for coupled simulations, in which the computing time devoted to the oceanic model and - above all - to the spin up of the coupled system will inflate considerably the number of core hours needed. Stochastic physics parameterisations, especially at lower resolution, seem able to provide an interesting alternative to tackle such controversy, improving model performance without increasing the nominal resolution and the overall computational cost.

7 Data availability

Post-processed data have been transferred from LRZ to CINECA via GridFTP, where they have been permanently stored. More important, free data accessibility to the climate user community is granted through a dedicated THREDDS Web Server hosted by CINECA (https://sphinx.hpc.cineca.it/thredds/sphinx.html), where it is possible to browse and directly download SPHINX data. Details on the the data accessibility and on the Climate SPHINX project itself are available on the website of the project (http://www.to.isac.cnr.it/sphinx/).

The set up of this infrastructure for data sharing has been possible thanks to DATA SPHINX, an EUDAT Data Pilot project, which will allow long-term storage and sharing among a wide scientific user community of high-resolution climate model output data. DATA SPHINX aims at building a repository serving the climate change impact modelling community, providing selected variables at high temporal and spatial resolution, with a focus on climate extremes and the hydrological cycle in areas with complex orography.

Table 1. Resolution dependent scientific configuration for EC-Earth in the Climate SPHINX experiments. The same number of ensemble members has been run for PDA and FSA experiments. T255C is the coupled configuration used for PFC simulations. Resolution is estimated at the Equator. The number of members indicate the deterministic and stochastic members. Backscatter ratio (tuning parameter for SKEB stochastic scheme), convective adjustment time (tuning parameter for deep convection) and momentum launch (tuning parameter for non-orographic gravity waves) are unit-less.

Truncation	Resolution	# members	Timestep	Backscatter ratio	Conv. adj. time	Mom. launch
T159	125.2 km	10+10	3600s	0.032	2.6	0.00375
T255	78.3 km	10+10	2700s	0.040	2.0	0.00375
T511	39.1 km	6+6	900s	0.085	1.5	0.00375
T799	25.0 km	3+3	720s	0.095	1.3	0.00368
T1279	15.7 km	1+1	600s	0.095	1.2	0.00334
T255C	78.3 km	3+3	2700s	0.040	2.0	0.00375

Table 2. Radiative fluxes expressed in W/m^2 for the reference experiment (i.e. the first simulation run) at different resolution for PDA simulations. D stands for deterministic simulation, S for stochastic. Fluxes have been tuned for T255D.

Simulation	Net Sfc	Net TOA	TOA SW	TOA LW
T159D	1.57	1.22	239.93	238.71
T159S	0.75	0.33	239.32	238.99
T255D	0.67	0.41	240.23	239.82
T255S	-0.16	-0.49	239.65	240.14
T511D	0.16	1.05	241.50	240.44
T511S	-0.75	0.19	241.07	240.88
T799D	-0.12	1.16	242.10	240.94
T799S	-0.82	0.47	241.78	241.31
T1279D	-0.09	1.44	242.58	241.14
T1279S	-1.09	0.41	242.16	241.74

Table 3. Resolution dependent technical details for EC-Earth in the Climate SPHINX experiments. T255C is the coupled configuration used for PFC simulations. Walltime has been measured on the Supermuc-II Haswell platform, and it is evaluated for deterministic simulations; stochastic simulations walltime is about the 5% higher.

Truncation	# Cores	Walltime (per year)	Leg length	Output data (per year)	Post-proc data (per year)
T159	224	52 min	1 year	26 GB	9.7 GB
T255	588	1h12 min	1 year	64 GB	24 GB
T511	840	6h10 min	6 months	249 GB	35 GB
T799	1120	14 h	2 months	605 GB	57 GB
T1279	1540	30 h	1 month	1.6 TB	111 GB
T255C	588	1h35 h	1 year	38 GB	30 GB
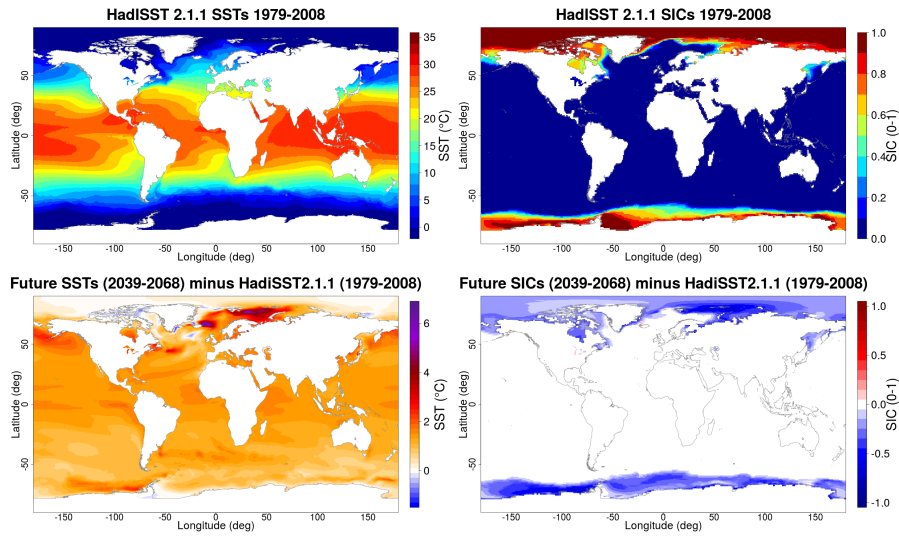
Figure 1. Upper panels: HadiSST2.1.1 climatology for the SSTs (left) and SIC (right) for the 1979-2008 period. Lower panels: climatological changes between the FutureHadiSST 2.1.1 dataset and the HadiSST 2.1.1 dataset for SST (left) and SIC (right).
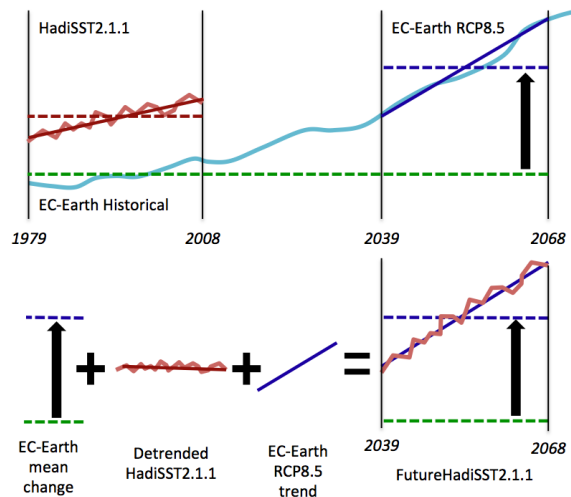

Figure 2. Scheme representing the methodology adopted to create the FutureHadiSST 2.1.1. The new dataset is a combination of detrended daily variability from HadiSST 2.1.1, CMIP5 EC-Earth mean change and CMIP5 EC-Earth RCP8.5 trend.
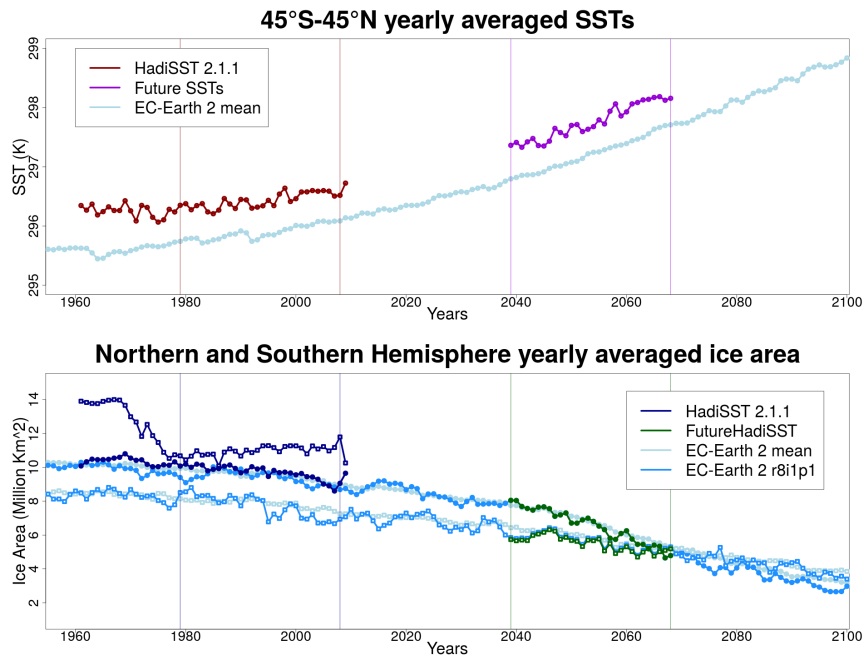
Figure 3. Upper panel: timeseries for 45°S-45°N yearly averaged SST for present-day HadISST 2.1.1 (red), FutureHadiSST 2.1.1 (violet) and CMIP5 EC-Earth ensemble mean (light blue). Lower panel: timeseries for Northern Hemisphere (filled circles) and Southern Hemisphere (empty circles) yearly averaged sea ice area for HadiSST 2.1.1 (dark blue), FutureHadiSST 2.1.1 (green), CMIP5 EC-Earth ensemble member "r8i1p1" (light blue) and the CMIP5 EC-Earth ensemble mean (faint blue).
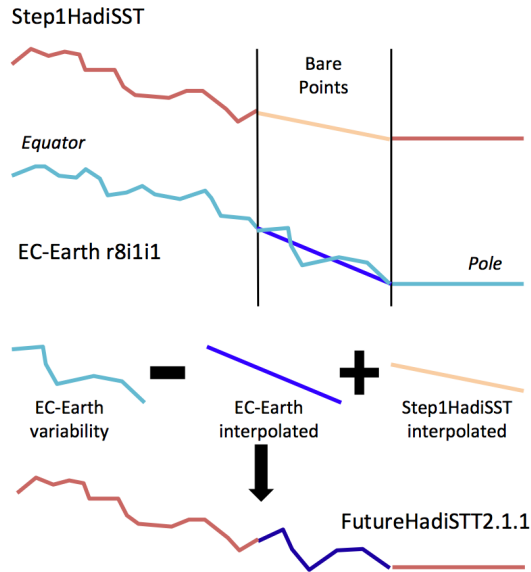
Figure 4. Scheme representing the methodology adopted to fill the "bare points", i.e. the points where sea ice have retreated in the CMIP5 EC-Earth RCP8.5 simulation. Each line represent a SST profile from the equator to the pole.
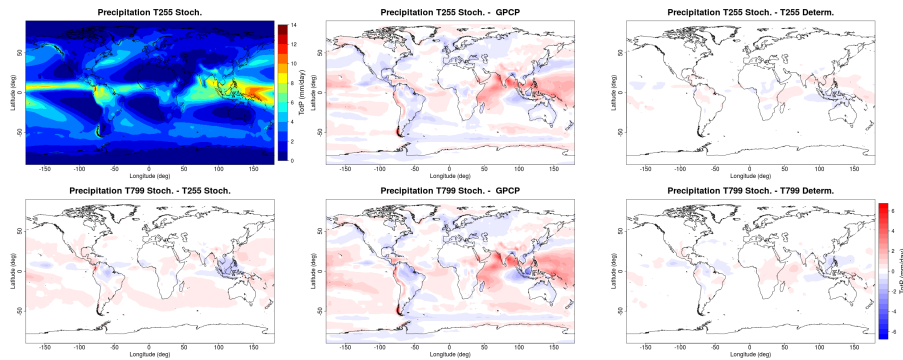

Figure 5. Upper left: Climatological ensemble mean precipitation for the PDA experiments (1979-2008) for T255 with stochastic physics. Lower left: T799 stochastic minus T255 stochastic precipitation. Central panels: T255 (upper) and T799 (lower) precipitation bias with respect to GPCP. Right panels: Stochastic minus deterministic climatological precipitation for T255 (upper) and T799 (lower).
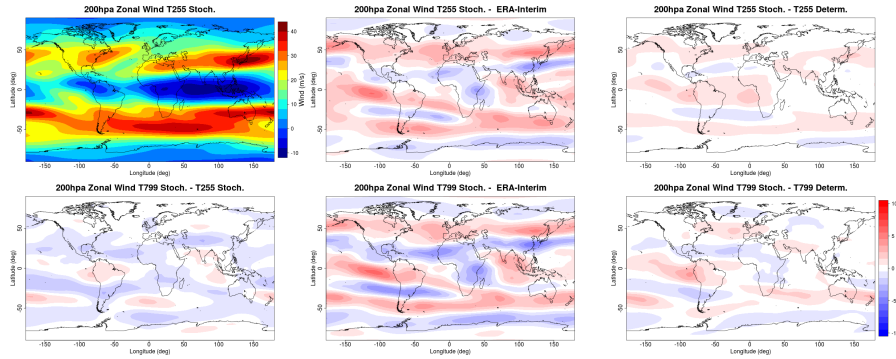
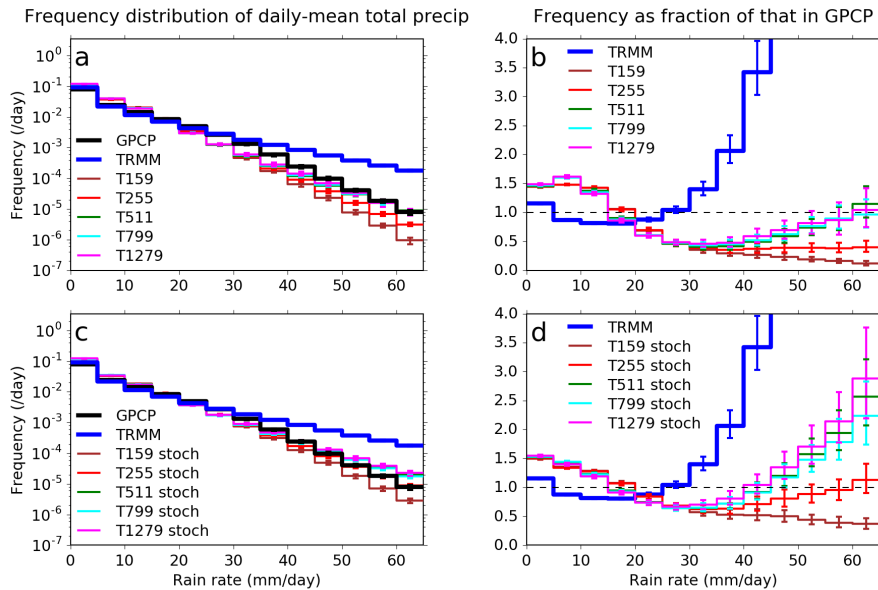**Figure 6.** Same as Fig. 5 but for zonal wind at 200 hPa. Here bias is evaluated against ERA-Interim reanalysis.

Figure 7. (a) and (c) show the frequency of occurrence of daily-mean rain rates averaged over 2.5°x2.5°grid boxes between 10°S-10°N in different datasets in 5 mm/day intervals, with rates below 0.1 mm/day omitted. (a) shows data for GPCP, TRMM and the deterministic SPHINX PDA simulations and (c) shows the same for the PDA simulations with stochastic physics. Note that the vertical axis is logarithmic. (b) and (d) show the the rain rates in each simulation and TRMM as a fraction of that in GPCP for the deterministic and stochastic runs respectively. Horizontal dashed lines indicate a fraction of 1, which would correspond to perfect agreement with GPCP. Vertical bars indicate the 95% confidence intervals. The frequency in (a) and (c) corresponds to that for an individual grid box, if all grid boxes were statistically equivalent. Data are shown for 1998-2008, the time period common to all datasets, for all ensemble members.
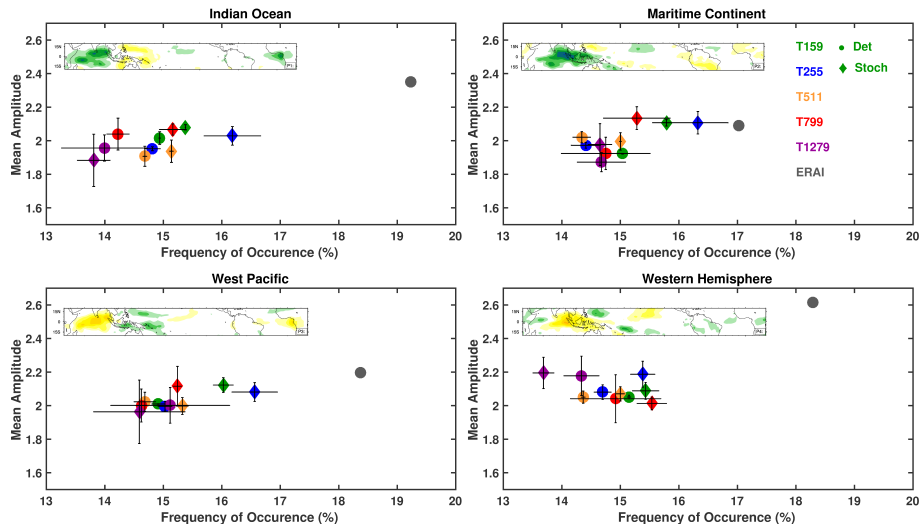
Figure 8. MJO frequency of occurrence vs. mean amplitude for the PDA experiments in the four different phases given the MJO amplitude to be > 1. The four phases are classified as Indian Ocean, Maritime Continent, West Pacific and Western Hemisphere, and their geographical location is shown by the boxes at the bottom of each panel with anomalous positive/negative precipitation patterns (green/yellow regions). Colours indicate the ensemble mean of the different resolutions as shown in the legend, where the circles are the deterministic runs and the diamonds the stochastic runs. ERA-Interim is reported in grey. Statistics are shown for the period 1980-2001. Error bars show the uncertainty range by providing the same statistic for periods half the length of the analysis.
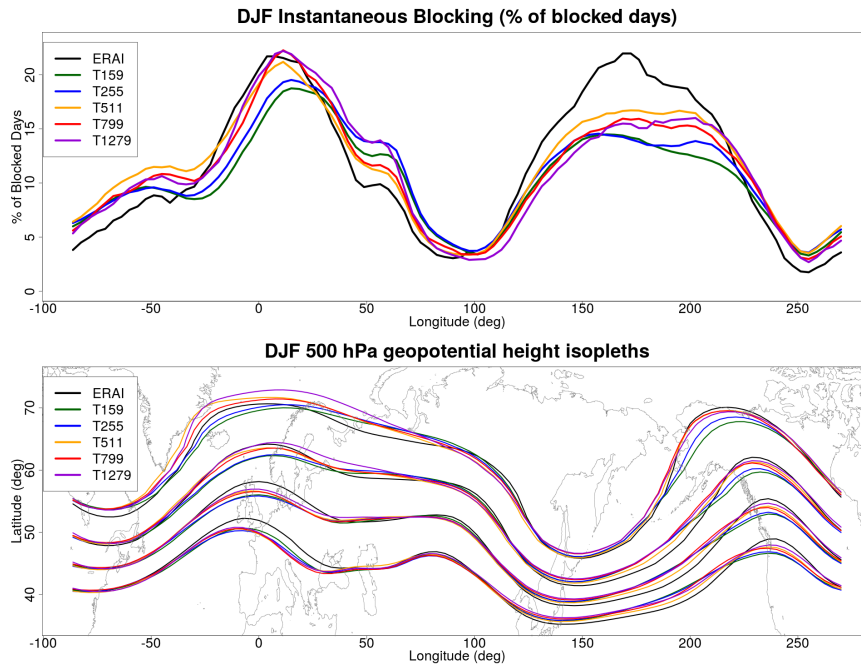
Figure 9. Upper panel: ensemble mean blocking frequencies following D'Andrea et al. (1998) for the different PDA experiments. Members of deterministic and stochastic experiments have been combined together for each resolution. ERA-Interim for the 1979-2008 period is shown as comparison in black. Lower panel: DJF climatological mean for geopotential height at 500 hPa for the ensemble mean of PDA experiments. Only 5200, 5300, 5400 and 5500m isopleths are reported.

References

Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S. C., Osprey, S. M., and Yang, S.: Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution, Journal of Geophysical Research: Atmospheres, 118, 3956–3971, 2013.

5   Arnold, H., Moroz, I., and Palmer, T.: Stochastic parametrizations and model uncertainty in the Lorenz'96 system, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 371, 20110 479, 2013.

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., and Scipal, K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, Journal of Hydrometeorology, 10, 2009.

10   Beljaars, A., Bechtold, P., Köhler, M., Morcrette, J.-J., Tompkins, A., Viterbo, P., and Wedi, N.: The numerics of physical parametrization, in: Proc. of ECMWF Seminar on Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling, ECMWF, Reading, UK, 2004.

Bengtsson, L., Steinheimer, M., Bechtold, P., and Geleyn, J.-F.: A stochastic parametrization for deep convection using cellular automata, Quarterly Journal of the Royal Meteorological Society, 139, 1533–1543, 2013.

15   Berckmans, J., Woollings, T., Demory, M.-E., Vidale, P.-L., and Roberts, M.: Atmospheric blocking in a high resolution climate model: influences of mean state, orography and eddy forcing, Atmospheric Science Letters, 14, 34–40, 2013.

Berner, J., Shutts, G., Leutbecher, M., and Palmer, T.: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, Journal of the Atmospheric Sciences, 66, 603–626, 2009.

Berner, J., Jung, T., and Palmer, T.: Systematic model error: the impact of increased horizontal resolution versus improved stochastic and

20   deterministic parameterizations, Journal of Climate, 25, 4946–4962, 2012.

Bouttier, F., Vié, B., Nuissier, O., and Raynaud, L.: Impact of stochastic physics in a convection-permitting ensemble, Monthly Weather Review, 140, 3706–3721, 2012.

Brayshaw, D., Hoskins, B., and Blackburn, M.: The basic ingredients of the North Atlantic storm track. Part I: land?sea contrast and orography, Journal of the Atmospheric Science, 66, 2539–2558, 2009.

25   Buizza, R., Petroliagis, T., Palmer, T., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., and Wedi, N.: Impact of model resolution and ensemble size on the performance of an ensemble prediction system, Quarterly journal of the royal meteorological society, 124, 1935–1960, 1998.

Christensen, H., Berner, J., Coleman, D. R., and Palmer, T.: Stochastic parameterisation and the El Niño-Southern Oscillation, Journal of Climate, 2016.

30   Christensen, H. M., Moroz, I. M., and Palmer, T. N.: Simulating Weather Regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model, Climate Dynamics, 44, 2195–2214, 2015.

Christensen, H. M., Dawson, A., and Palmer, T.: Constraining Stochastic Parametrisation Schemes using High-resolution Model Simulations, in preparation, 2017.

D'Andrea, F., Tibaldi, S., Blackburn, M., Boer, G., Deque, M., Dix, M. R., Dugas, B., Ferranti, L., Iwasaki, T., Kitoh, A., Pope, V., Randall,

35   D., Roeckner, E., Straus, D., Stern, W., den Dool, H. V., and Williamson, D.: Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979-1988, Climate Dynamics, 14, 383–407, 1998.

Davini, P. and D'Andrea, F.: Northern Hemisphere atmospheric blocking representation in Global Climate Models: twenty years of improvements?, Journal of Climate, 2016.

Davini, P., Cagnazzo, C., Gualdi, S., and Navarra, A.: Bidimensional diagnostics, variability and trends of Northern Hemisphere blocking, Journal of Climate, 25, 6996–6509, 2012.

Dawson, A. and Palmer, T.: Simulating weather regimes: impact of model resolution and stochastic parameterization, Climate Dynamics, 44, 2177–2193, 2015.

Dawson, A., Palmer, T., and Corti, S.: Simulating regime structures in weather and climate prediction models, Geophysical Research Letters, 39, 2012.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., Dixon, K., Griffies, S. M., Lee, H.-C., Pacanowski, R. C., et al.: Simulated climate and climate change in the GFDL CM2. 5 high-resolution coupled climate model, Journal of Climate, 25, 2755–2781, 2012.

Demory, M.-E., Vidale, P. L., Roberts, M. J., Berrisford, P., Strachan, J., Schiemann, R., and Mizielinski, M. S.: The role of horizontal resolution in simulating drivers of the global hydrological cycle, Climate Dynamics, 42, 2201–2225, 2014.

Deng, Q., Khouider, B., and Majda, A. J.: The MJO in a coarse-resolution GCM with a stochastic multicloud parameterization, Journal of the Atmospheric Sciences, 72, 55–74, 2015.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: The role of internal variability, Climate Dynamics, pp. 1–20, 2012.

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., and Murray, D.: Was there a basis for anticipating the 2010 Russian heat wave?, Geophysical Research Letters, 38, 2011.

Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., Jonker, H. J., and Selten, F.: Stochastic convection parameterization with Markov Chains in an intermediate-complexity GCM, Journal of the Atmospheric Sciences, 73, 1367–1382, 2016.

Dunn-Sigouin, E. and Son, S.-W.: Northern Hemisphere blocking frequency and duration in the CMIP5 models, Journal of Geophysical Research: Atmospheres, 118, 1179–1188, 2013.

ECMWF: IFS cycle36r1, http://www.ecmwf.int/research/ifsdocs/CY36r1/, European Center for Medium Range Forecast, 2009.

Gottschalck, J., Wheeler, M., Weickmann, K., Vitart, F., Savage, N., Lin, H., Hendon, H., Waliser, D., Sperber, K., Nakagawa, M., et al.: A framework for assessing operational Madden-Julian oscillation forecasts: a CLIVAR MJO working group project, Bulletin of the American Meteorological Society, 91, 1247, 2010.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmos. Chem. Phys. Discuss, 13, 23 845–23 893, 2013.

Haarsma, R. J., Roberts, M., V., L., P., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fuckar, N. S., V., G., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., , and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP), Geosci. Model Dev. Discuss., in review, 2016.

Hazeleger, W., Severijns, C., Semmler, T., ŞŞtefăănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., et al.: EC-Earth: A Seamless Earth-System Prediction Approach in Action., Bulletin of the American Meteorological Society, 91, 2010.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al.: EC-Earth V2.2: description and validation of a new seamless Earth system prediction model, Climate dynamics, 39, 2611–2629, 2012.

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., and Dee, D.: ERA-20CM: a twentieth-century atmospheric model ensemble, Quarterly Journal of the Royal Meteorological Society, 141, 2350–2375, 2015.

5 Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multisatellite observations, Journal of Hydrometeorology, 2, 36–50, 2001.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of Hydrometeorology, 8, 38–55, 2007.

10 Hung, M.-P., Lin, J.-L., Wang, W., Kim, D., Shinoda, T., and Weaver, S. J.: MJO and convectively coupled equatorial waves simulated by CMIP5 climate models, Journal of Climate, 26, 6185–6214, 2013.

IPCC: Climate change 2014: impacts, adaptation, and vulnerability. Part B: regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2015.

Jung, T., Miller, M., Palmer, T., Towers, P., Wedi, N., Achuthavarier, D., Adams, J., Altshuler, E., Cash, B., Kinter Iii, J., et al.: High-resolution
15 global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate, and seasonal forecast skill, Journal of Climate, 25, 3155–3172, 2012.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, Bulletin of the American Meteorological Society, 96, 1333–1349, 2015.

20 Kennedy, J., Reyner, N., Millington, S. C., and Saunby, M.: The Met Office Hadley Centre sea ice and sea-surface temperature data set, version 2, part 2: Sea surface temperature analysis, in preparation, 2016.

Khouider, B., Biello, J., Majda, A. J., et al.: A stochastic multicloud model for tropical convection, Communications in Mathematical Sciences, 8, 187–216, 2010.

Khouider, B., St-Cyr, A., Majda, A. J., and Tribbia, J.: The MJO and convectively coupled waves in a coarse-resolution GCM with a simple
25 multicloud parameterization, Journal of the Atmospheric Sciences, 68, 240–264, 2011.

Kim, D., Sperber, K., Stern, W., Waliser, D., Kang, I.-S., Maloney, E., Wang, W., Weickmann, K., Benedict, J., Khairoutdinov, M., et al.: Application of MJO simulation diagnostics to climate models, Journal of Climate, 22, 6413–6436, 2009.

Kim, D., Xavier, P., Maloney, E., Wheeler, M., Waliser, D., Sperber, K., Hendon, H., Zhang, C., Neale, R., Hwang, Y.-T., et al.: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection, Journal of Climate, 27, 5379–5395, 2014.

30 Kinter III, J., Cash, B., Achuthavarier, D., Adams, J., Altshuler, E., Dirmeyer, P., Doty, B., Huang, B., Jin, E., Marx, L., et al.: Revolutionizing climate modeling with Project Athena: A multi-institutional, international collaboration, Bulletin of the American Meteorological Society, 94, 231–245, 2013.

Klingaman, N. P., Woolnough, S. J., Jiang, X., Waliser, D., Xavier, P. K., Petch, J., Caian, M., Hannay, C., Kim, D., Ma, H.-Y., et al.: Vertical structure and physical processes of the Madden-Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening,
35 Journal of Geophysical Research: Atmospheres, 120, 4690–4717, 2015.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H. M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R. M., Goddard, J., Haiden, T., Hogan, R. J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson,

L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P. K., Subramanian, A., Vitart, F., Wedi, N., and Weisheimer, A.: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision, in preparation Q J Roy Meteor Soc., 2017.

Lin, J.-L., Kiladis, G. N., Mapes, B. E., Weickmann, K. M., Sperber, K. R., Lin, W., Wheeler, M. C., Schubert, S. D., Del Genio, A., Donner, L. J., et al.: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals, Journal of Climate, 19, 2665–2690, 2006.

Lin, J. W.-B. and Neelin, J. D.: Influence of a stochastic moist convective parameterization on tropical climate variability, Geophysical research letters, 27, 3691–3694, 2000.

Lin, J. W.-B. and Neelin, J. D.: Toward stochastic deep convective parameterization in general circulation models, Geophysical research letters, 30, 2003.

Lu, J., Chen, G., Leung, L. R., Burrows, D. A., Yang, Q., Sakaguchi, K., and Hagos, S.: Toward the dynamical convergence on the jet stream in aquaplanet AGCMs, Journal of Climate, 28, 6763–6782, 2015.

Madden, R. A. and Julian, P. R.: Observations of the 40-50-day tropical oscillation-A review, Monthly Weather Review, 122, 814–837, 1994.

Madec, G.: NEMO ocean engine, Tech. rep., Institut Pierre-Simon Laplace (IPSL), 2008.

Masato, G., Hoskins, B. J., and Woollings, T.: Winter and Summer Northern Hemisphere blocking in CMIP5 models, Journal of Climate, 26, 7044–7059, 2013.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., et al.: Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, 2012.

Mizielinski, M., Roberts, M., Vidale, P., Schiemann, R., Demory, M.-E., Strachan, J., Edwards, T., Stephens, A., Lawrence, B., Pritchard, M., et al.: High-resolution global climate modelling: the UPSCALE project, a large-simulation campaign, Geoscientific Model Development, 7, 1629–1640, 2014.

Mizuta, R., Adachi, Y., Yukimoto, S., and Kusunoki, S.: Estimation of future distribution of sea surface temperature and sea ice using CMIP3 multi-model ensemble mean, Technical Report of the Meteorological Research Institute, p. 28pp, 2008.

Moss, R., Edmonds, J., Hibbard, K., Manning, M., Rose, S., van Vuuren, D., Carter, T., Emori, S., Kainuma, M., Kram, T., et al.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, 2010.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J., and Sandu, I.: Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble, Quarterly Journal of the Royal Meteorological Society, 2016.

Palmer, T.: Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction, Quarterly Journal of the Royal Meteorological Society, 138, 841–861, 2012.

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., and Weisheimer, A.: Stochastic parametrization and model uncertainty, European Centre for Medium-Range Weather Forecasts, 2009.

Peatman, S. C., Matthews, A. J., and Stevens, D. P.: Propagation of the Madden–Julian Oscillation and scale interaction with the diurnal cycle in a high-resolution GCM, Climate Dynamics, 45, 2901–2918, 2015.

Pelly, J. and Hoskins, B.: A New Perspective on Blocking, Journal of the Atmospheric Science, 60, 743–755, 2003.

Plant, R. and Craig, G. C.: A stochastic parameterization for deep convection based on equilibrium statistics, Journal of the Atmospheric Sciences, 65, 87–105, 2008.

Raymond, D., Fuchs, Ž., Gjorgjievska, S., and Sessions, S.: Balanced dynamics and convection in the tropical troposphere, Journal of Advances in Modeling Earth Systems, 7, 1093–1116, 2015.

Renwick, J. and Wallace, J.: Relationships between North Pacific wintertime blocking, El Niño, and the PNA pattern, Monthly weather review, 124, 2071–2076, 1996.

Rex, D.: Blocking action in the middle troposphere and its effect upon regional climate: I. An aerological study of blocking action, Tellus, 2, 196–211, 1950.

5   Saeed, F., Haensler, A., Weber, T., Hagemann, S., and Jacob, D.: Representation of extreme precipitation events leading to opposite climate change signals over the Congo basin, Atmosphere, 4, 254–271, 2013.

Sakradzija, M., Seifert, A., and Dipankar, A.: A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone, Journal of Advances in Modeling Earth Systems, 2016.

Sanchez, C., Williams, K. D., and Collins, M.: Improved stochastic physics schemes for global weather and climate models, Quarterly Journal
10 of the Royal Meteorological Society, 142, 147–159, 2016.

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F., and Jung, T.: The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models, Journal of Climate, 2016.

Shutts, G.: The propagation of eddies in diffluent jetstreams: eddy vorticity forcing of blocking flow fields., Quarterly Journal of the Royal
15 Meteorological Society, 109, 737–761, 1983.

Shutts, G.: A kinetic energy backscatter algorithm for use in ensemble prediction systems, Quarterly Journal of the Royal Meteorological Society, 131, 3079–3102, 2005.

Shutts, G. and Pallarès, A. C.: Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 372, 20130 284,
20 2014.

Shutts, G. and Palmer, T.: Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem, Journal of climate, 20, 187–202, 2007.

Sillmann, J., Croci-Maspoli, M., Kallache, M., and Katz, R. W.: Extreme Cold Winter Temperatures in Europe under the Influence of North Atlantic Atmospheric Blocking, Journal of Climate, 24, 5899–5913, 2011.

25 Slingo, J., Sperber, K., Boyle, J., Ceron, J.-P., Dix, M., Dugas, B., Ebisuzaki, W., Fyfe, J., Gregory, D., Gueremy, J.-F., et al.: Intraseasonal oscillations in 15 atmospheric general circulation models: results from an AMIP diagnostic subproject, Climate Dynamics, 12, 325–357, 1996.

Sperber, K., Slingo, J., and Inness, P.: Modeling intraseasonal variability, Intraseasonal Variability of the Atmosphere-Ocean Climate System,, p. 613, 2011.

30 Straub, K. H.: MJO initiation in the real-time multivariate MJO index, Journal of Climate, 26, 1130–1151, 2013.

Taylor, K., Stouffer, R., and Meehl, G.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485, 2012.

Tibaldi, S. and Molteni, F.: On the operational predictability of blocking, Tellus, 42A, 343–365, 1990.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice
35 concentrations, Journal of Geophysical Research: Atmospheres, 119, 2864–2889, 2014.

Tyrlis, E. and Hoskins, B.: The Morphology of Northern Hemisphere Blocking, Journal of Atmospheric Science, 65, 1653–1665, 2008.

Valcke, S.: The OASIS3 coupler: a European climate modelling community software, Geoscientific Model Development, 6, 2013.

van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., and Hazeleger, W.: Decadal prediction skill in a multi-model ensemble, Climate Dynamics, 38, 1263–1280, 2012.

Vancoppenolle, M., Bouillon, S., Fichefet, T., Goosse, H., Lecomte, O., Morales Maqueda, M., and Madec, G.: LIM, The Louvain-la-Neuve sea Ice Model, Notes du Pôle de modélisation, 2012.

5 Wang, Y., Zhang, G. J., and Craig, G. C.: Stochastic convective parameterization improves the simulation of tropical precipitation variability in the NCAR CAM5, Geophysical Research Letters, 2016.

Weisheimer, A., Corti, S., Palmer, T., and Vitart, F.: Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 372, 20130 290, 2014.

10 Wheeler, M. C. and Hendon, H. H.: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, Monthly Weather Review, 132, 1917–1932, 2004.

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E. G., and König-Langlo, G.: The global energy balance from a surface perspective, Climate dynamics, 40, 3107–3134, 2013.

Yonehara, H. and Ujiie, M.: Stochastic physics scheme for model uncertainties in the JMA one-week ensemble prediction system, WMO
15 CAS/JSC WGNE, 2011.

Zappa, G., Shaffrey, L. C., and Hodges, K. I.: The ability of CMIP5 models to simulate north atlantic extratropical cyclones, Journal of Climate, 26, 5379–5396, 2013.

Zhang, C., Gottschalck, J., Maloney, E. D., Moncrieff, M. W., Vitart, F., Waliser, D. E., Wang, B., and Wheeler, M. C.: Cracking the MJO nut, Geophysical Research Letters, 40, 1223–1230, 2013.