Call: H2020-SC5-2014-two-stage

Topic: SC5-01-2014

PRIMAVERA

Grant Agreement 641727

PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment

Deliverable D*4.1*

Quantification of the relative cost/performance of different approaches to going beyond simple parameterisation

Deliverable Title	*Quantification of the relative cost/performance of different approaches to going beyond simple parameterisation*
Brief Description	*Quantify the relative cost/performance of the different approaches to going beyond simple parameterisation (making use of the simulated historical period and comparing with observations), the relative strengths and weaknesses of each approach, the consequences for representing particular climate extremes, their impacts on climate risks and sectors considered in WP10,11*
WP number	4
Lead Beneficiary	*Thomas Arsouze (BSC)*
Contributors	*Thomas Arsouze, Miguel Castrillo, Mario Acosta (BSC)* *Oliver Gutjahr, Dian Putrasahan, Katja Lohmann (MPG)* *Malcolm Roberts (Met Office)* *Kristian Strommen, Hannah Christensen, Tim Palmer (UOXF)* *Pier Luigi Vidale (UREAD)* *William David Cabos Narvaez (AWI)*
Creation Date Version Number Version Date	03.07.2020 1 03.07.2020
Deliverable Due Date Actual Delivery Date	31.03.2020
Nature of the Deliverable	*R – Report X* *P - Prototype* *D - Demonstrator* *O - Other*
Dissemination Level/ Audience	*PU – Public X* *PP - Restricted to other programme participants, including the Commission services* *RE - Restricted to a group specified by the consortium, including the Commission services* *CO - Confidential, only for members of the consortium, including the Commission services*

Version	**Date**	**Modified by**	**Comments**
1.0	03/07/2020	Thomas Arsouze	First complete draft

Table of Contents

1. Executive Summary

Within the WP4 Frontiers of Climate work, we have investigated several ways in which we can represent the small scale processes in a better way than with deterministic parameterisations. Atmospheric convection and the ocean mesoscale (e.g. eddies, boundary currents) can be explicitly represented if the model resolution is high enough (above 10km), while stochastic parameterisations can improve the simulated mean state and variance.

The stochastic schemes only add a small amount to runtime (~4%), while increasing the atmosphere or ocean resolution to 10km (from 25km in the high resolution standard models) multiples the simulation cost by ~16x.

Improvements are found in simulation of some key mean state variables when including stochastic physics schemes, sometimes being competitive with changes due to resolution. However, as expected, if the focus is on improved representation of mesoscale ocean processes and large-scale circulation, then model resolution is very important. Attempting to represent convection explicitly at 10km generally provided a worse simulation than parameterised convection, apart from the phase of the diurnal cycle. This may mean that the model needs extensive retuning for such a large change in physics.

There may be key resolution thresholds (e.g. Agulhas) that stochastic cannot represent.

Overall, we would recommend that these state-of-the-art stochastic schemes be included by default in model simulations, due to their low cost and assessed benefits. We can see clear improvements in simulation when resolution reaches 10km (particularly in the ocean), but currently such resolutions are not feasible for CMIP6-style experiments (i.e. >600's years of spinup and pre-industrial control), but may become affordable in the near future.

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, Part B Section 1.1) WP numbers are in brackets:

No.	Objective	Yes	No
A	To develop a new generation of global high-resolution climate models. *(3, 4, 6)*		X
B	To develop new strategies and tools for evaluating global high-resolution climate models at a process level, and for quantifying the uncertainties in the predictions of regional climate. *(1, 2, 5, 9, 10)*		X
C	To provide new high-resolution protocols and flagship simulations for the World Climate Research Programme (WCRP)'s Coupled Model Intercomparison Project (CMIP6) project, to inform the Intergovernmental Panel on Climate Change (IPCC) assessments and in support of emerging Climate Services. *(4, 6, 9)*		X
D	To explore the scientific and technological frontiers of capability in global climate modelling to provide guidance for the development of future generations of prediction systems, global	X	

	climate and Earth System models (informing post-CMIP6 and beyond). *(3, 4)*		
E	To advance understanding of past and future, natural and anthropogenic, drivers of variability and changes in European climate, including high impact events, by exploiting new capabilities in high-resolution global climate modelling. *(1, 2, 5)*	X	
F	To produce new, more robust and trustworthy projections of European climate for the next few decades based on improved global models and advances in process understanding. *(2, 3, 5, 6, 10)*		X
G	To engage with targeted end-user groups in key European economic sectors to strengthen their competitiveness, growth, resilience and ability by exploiting new scientific progress. *(10, 11)*		X
H	To establish cooperation between science and policy actions at European and international level, to support the development of effective climate change policies, optimize public decision making and increase capability to manage climate risks. *(5, 8, 10)*	X	

3. Detailed Report

3.1 Increasing horizontal resolution

A list of all simulations performed in the framework of PRIMAVERA WP4 is provided in Table 1. Each model defined a set of experiments with differing horizontal resolution of either, or both, the atmospheric or oceanic model, with the aim of reaching an eddy-resolving resolution for the ocean (~10km) and as close as possible from convection permitting for the atmosphere (~10km also).

3.1.1 Models Costs

Table 1 provides the cost in terms of computation and storage for each model and each configuration.

Model name	CMIP6 resolution (atmos-ocean) km	Initial condition	Total years (spinup years)	Cray XC40 Nodes (atmos - ocean)	Max turnaround (years per day)	Output per year (TB)

HadGEM3-GC31						
LL	250-100	LL-spinup (30 years)	1130 (30)	12-2	4	0.13
MM	100-25	MM-spinup (30 years)	680 (30)	50-24	1.3	0.73
HM	50-25	MM-spinup (30 years)	117(0)	90-24	0.5	2.8
MH	100-8	MH-spinup (30 years)	205 (30)	34-171	0.45	2.0
HH	50-8	MH-spinup (30 years)	100 (0)	90-171	0.4	4.5
HadGEM3-GC31 (atmosphere only)						
LM	130	-	-	12	4	-
MM	60	-	-	72	2.3	-
HM	25	-	-	86	0.6	-
UH	10	-	-	202	0.12	-
EC-Earth3P						
SR	60 - 100	SR-spinup (50 years)	500 (50)	10-5	18	0.1
HR	30 - 25	HR-spinup (50 years)	200 (50)	19-29	3.1	0.6
VHR	15 - 10	VHR-spinup	100 (50)	27-67	0.45	1

		(50 years)				
VHR-CFB	15-10	VHR-spinup (10 years)	25 (0)	27-67	0.45	1
MPI-ESM1-2						
HR	100 - 45	HR-spinup (50 years)	200 (50)	110	12	0.15
XR	50 - 45	ER-spinup (50 years)	200 (50)	110	8	0.2
ER	100 - 11	ER-spinup (50 years)	130 (50)	240	1	0.25
AWI-CM						
LR/T63	200-50	LR-spinup (50 years)	200 (50)	24-20	20	0.25
HR/T127	100-25	HR-spinup (50 years)	200 (50)	48-100	8	1

Table 1: For each HadGEM3-GC31, EC-Eart3P and MPI-ESM1-2 model, we provide resolution for the control-1950 simulation, the initial condition, total simulated years, and costs of various model resolutions (Cray XC40 with 36 cores per node for HadGEM3-GC31, Lenovo with 48 cores per node for EC-Earth3P, Bull/Atos with 36 cores per nodes for MPI-ESM1-2 and AWI-CM), together with raw model output volumes.

In order to resolve the mesoscale to a reasonable extent at mid-high latitudes, 1/12° ocean at least is required. Table 1 shows the costs in both HPC time for simulation, and also storage volumes, for different resolution models. For HadGEM3-GC31 model, moving from a standard CMIP6 type model (LL) to eddy-rich, with a reasonable atmosphere resolution (HH) comes at an increased HPC cost of 180 times (18 times more nodes and 10 times slower), and increased storage cost of 34. For EC-Earth3P, the increase from standard (CMIP6-type) resolution to very-high resolution (eddy-resolving in the ocean, close to convection-permitting in the atmosphere) induces an increase in HPC cost of 250 times (~7 times more nodes and ~40 times slower)

For MPI-ESM1-2 model, switching from HR to ER (i.e. only an increase in ocean resolution to eddy-resolving resolution, keeping the same atmospheric resolution) leads to an increased HPC cost of 24. The increase in terms of storage is much more moderate as outputs (in particular at high frequency) have been reduced.

For AWI-CM, in which the ocean is formulated in an unstructured mesh, going to an eddy resolving resolution in the ocean and a doubling in atmospheric resolution (from LR/T63 to HR/T127) leads to an increase in HPC cost of only 8. This improvement is attained by an efficient design of the oceanic grid, which is based on the Rossby radius and/or the observed variability pattern. This mesh design strategy allows for eddy-resolving resolution where necessary, although tends to coarsen the resolution in tropical and partly subtropical latitudes compared to the regular meshes with a similar number of nodes. The increase in storage cost, a factor of 4, is also significantly lower.

From the atmosphere perspective only (with HadGEM3-GC31 model), moving from a 130km model to around 10km causes an increase in model cost of 566 times in HPC resources and turnaround time. The output comparison has not been done, as it was not possible to configure the same diagnostics in this high resolution model. This is a global future challenge at such resolutions: to reduce the output data volumes while retaining the variables needed for ever more sophisticated analyses compared to observations, in both space and time dimensions, and leveraging the tools and computational resources needed for handling such heavy datasets.

Overall, the cost of switching to high resolution varies among models but is of the order of several ten-fold for switching from CMIP6-type resolution to eddy-rich resolution in the ocean, and several hundred-fold from a CMIP6-type resolution for reaching a resolution of about 10km for the atmosphere. The cost of storage for high resolution is several ten-fold, but strongly depends on scientific questions tackled. However, a proper use of the advantages of unstructured meshes in AWI-CM allows for a significant reduction of the cost with respect to the models with regular grid oceanic components, both in computational resources and storage capacity. From a software and hardware perspective, handling such type of data for scientific analysis requires adaptation as the most common tools used for CMIP6 analysis are found to be inadequate. Developing efficient online diagnostics and solutions for high performance data analysis on supercomputers are needed to help users for post-treatment.

3.1.2 Validation of simulations

A comparison between *MPI-ESM1-2* HR and ER control simulations were made and published in Gutjahr et al., (2019).

Compared to the UK Met Office EN4 dataset (averaged over 1950-1954), MPI-ESM1-2-HR shows a relatively cooler SST distribution (Fig. 1a & 1b), with approximately -7°C cooler in the warming hole region. This is likely due to a too-zonal North Atlantic Current that causes fresh and cold Labrador Sea water to intrude much further south, thus weakening the northward ocean heat transport by AMOC in the subtropical, inter-gyre and subpolar latitudes. There is also a cold bias of about -2°C along the ACC and hints of warm bias (~1°C) in major upwelling regions. In MPI-ESM1-2-ER, much of the cold bias in SST is reduced (Fig. 1c). For example, in the Southern Ocean, the presence of eddies acts to flatten and shift outcropping isopycnals southward, thereby ameliorating the cold bias along the ACC. However, in the upwelling regions, warm bias seems to have increased slightly. These warm biases are likely related to the wind system, which are not much improved by resolving ocean eddies.

The effect of eddies on temperature bias are even more evident at deeper depths, such as at 740m (Fig. 1d-f). HR presents a general warm bias at this depth, with particularly warm waters in the Atlantic (Fig. 2b). In fact, the general warm bias spans most intermediate depths from 100-1500m (Fig. 1h). In the Atlantic Ocean, around 40°S, much of the warm and even salty bias (Fig. 1e) is related to Agulhas leakage. Since HR uses a ~0.4° grid resolution, the Agulhas Retroflection is not sufficiently captured, thus allowing excess amounts of relatively warmer and saltier waters from the Indian Ocean to leak into the Atlantic Ocean. With ER, the Agulhas Retroflection and Agulhas leakage are much better represented, thereby providing more appropriate amounts of transport from the Indian Ocean to the Atlantic Ocean, which results in significantly reduced temperature and salinity biases

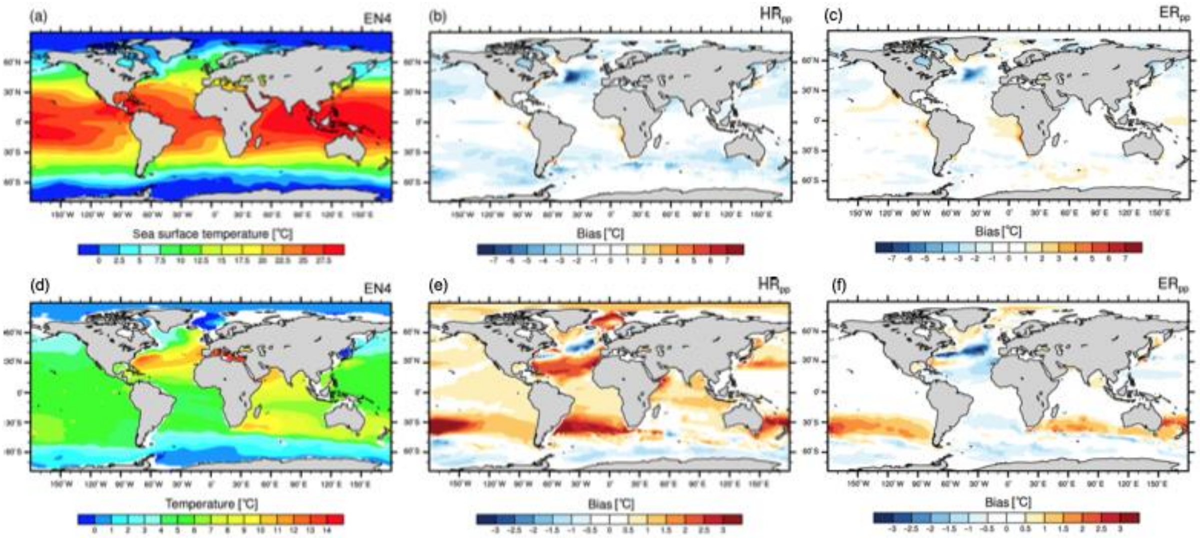

Figure 1: a) SST distribution averaged over 1950-1954 from UK MetOffice EN4 dataset.; b) SST bias of HR relative to EN4.; c) SST bias of ER relative to EN4.; d), e) and f): Same as a), b), and c) respectively, except for seawater potential temperature at 740m depth.

In the HR configuration, another maximum temperature and salinity bias can be seen at intermediate depths of around 800-1000m at about 30°N (Fig. 2b & 2e). These biases are related to the transport of Mediterranean Outflow water (MOW) and its spreading into the Atlantic. MOW is too warm and saline in HR configuration and is related to the representation of the bathymetry of the Strait of Gibraltar. In reality, this strait is ~12km wide with a sill depth of ~300m. For HR, the coarse resolution resulted with a width of ~54km, compared to ER representing a width of ~24km for said strait. Both however, had ~100m shallower sill depths compared to reality. Increased outflow of warm and saline MOW water through the straits in HR resulted in the larger biases, while ER produced more realistic properties (Fig. 2c & 2f).

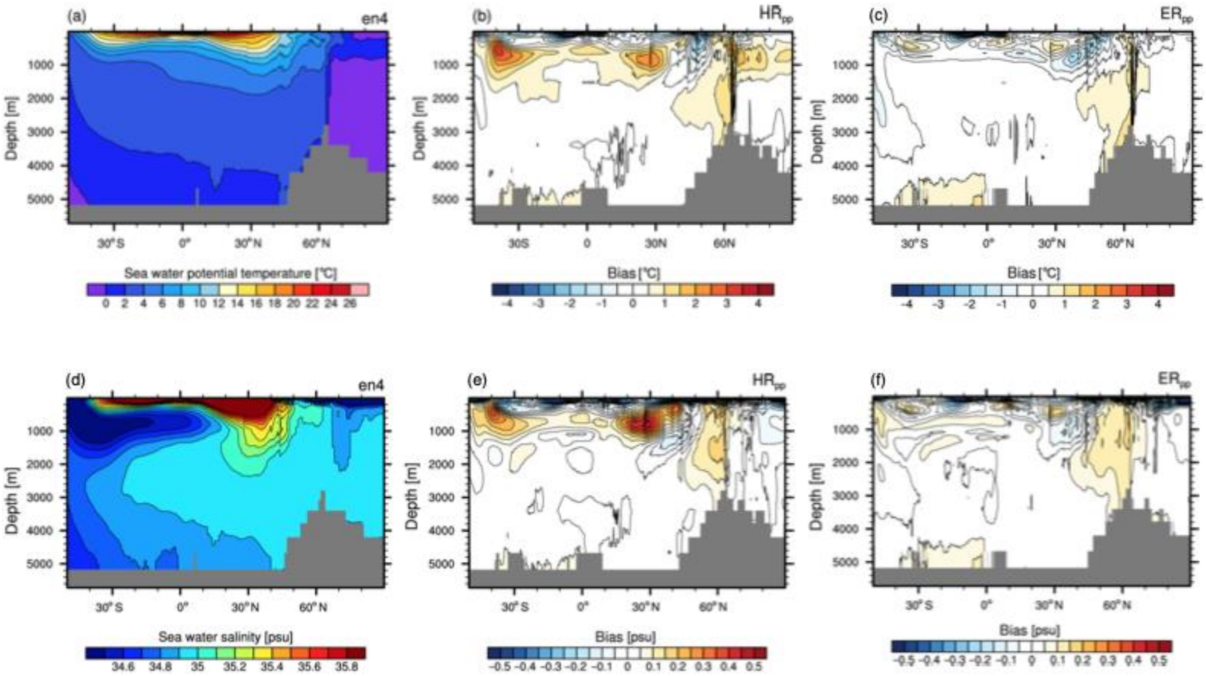

Figure 2: a) Zonally-averaged seawater potential temperature averaged over 1950-1954 from UK MetOffice EN4 dataset.; b) Bias of HR relative to EN4.; c) Bias of ER relative to EN4.; d), e) and f): Same as a), b), and c) respectively, except for seawater salinity.

Increased resolution does not necessarily improve biases everywhere. For instance, around the northern flank of the Atlantic subtropical gyre, a colder bias is seen in ER relative to HR (Fig. 1e & 1f). In this case, while the Gulf stream separates earlier from the eastern US coast in ER, the flow path of the North Atlantic Current is still too zonal, resulting in the cold bias of the warming hole around 50°N (Fig. 1c).

Last but not least, the warm bias in HR over the Arctic Ocean at depths 200-1000m (Fig. 2b) is abated in the ER configuration (Fig. 2c). This may be due to spurious numerical mixing of the advection operator that affects the vertical mixing strength. In ER, reduced vertical mixing decreases diffusion of Atlantic water into the Arctic, thereby removing the warm biases.

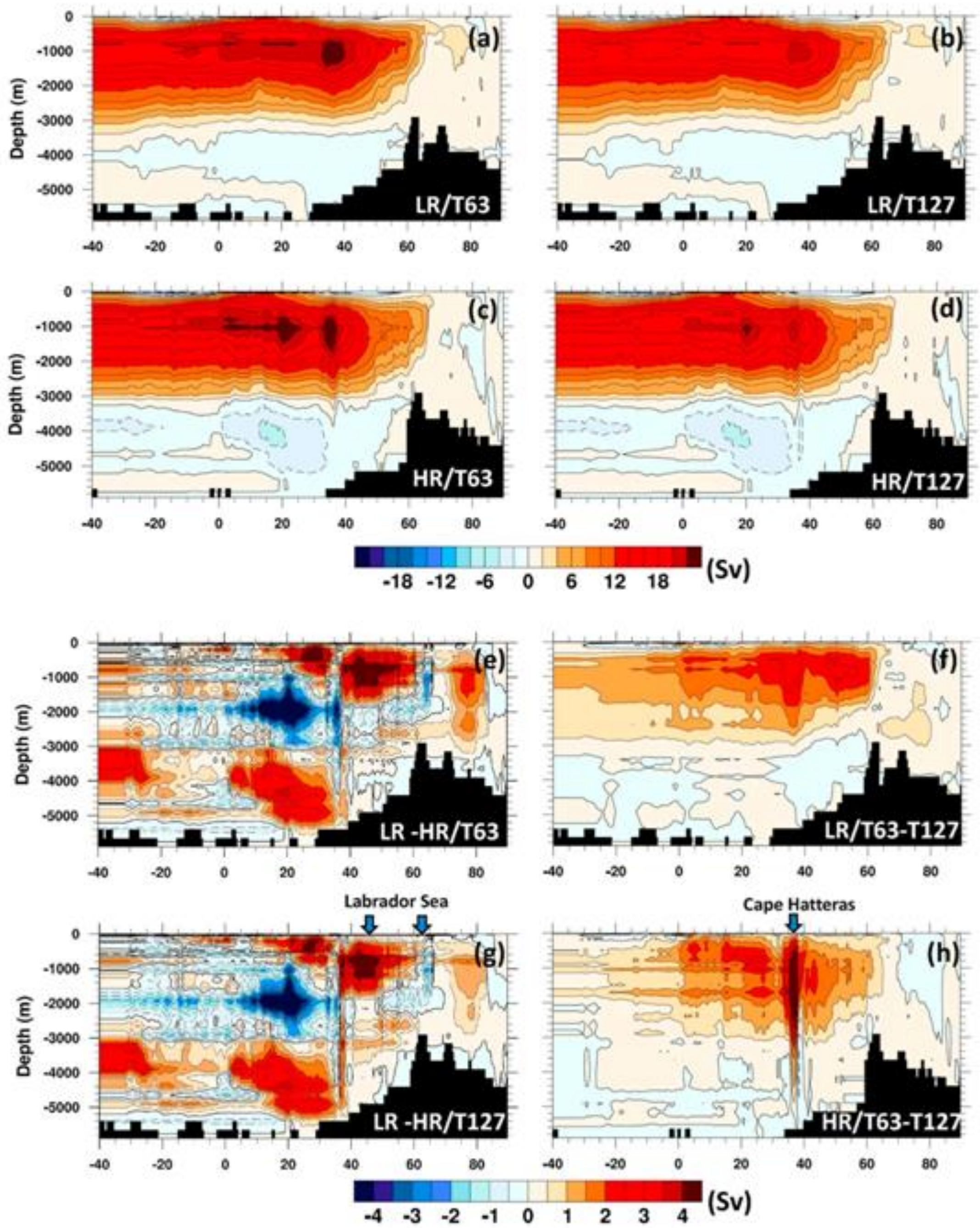

Figure 3: (a–d) Mean (1980–2005) Atlantic meridional overturning circulation for different setups and the difference between setups with (f and h) the same atmosphere but different ocean and (e and g) the same ocean but different atmospheric resolution

The impact of resolution on the climate simulated by AWI-CM has been studied systematically using two complementary HR/T63 and LR/T127 simulations. The panels 3.e and 3.g illustrate the main effects of increased oceanic resolution on the simulated AMOC. The strengthening of the deep cell in HR (3,000- to 5,000-m range),which is already apparent in the full AMOC patterns (Figures 3.a –d) could be related to the change in the frontal structure and the slope of outcropping isopycnal layers in the Southern Ocean. The local eddy resolution in the Agulhas retroflection and Brazil-Malvinas Confluence regions in HR allows FESOM to simulate the eddy-induced transport which compensates the tendency to steepening of the isopycnal slopes by the winds.

The deeper southward flow of the upper cell in HR at depths about 2,000 m (from about 0° to 30°N) can be mainly related to a better representation of the bottom topography at the latitudes of the Caribbean basin. The coarseness of the LR mesh makes part of the Deep Western Boundary Current (DWBC) to flow along the submerged topographic features there. Furthermore, the higher resolution in HR affects the representation of the DWBC: it is sharper (but weaker) in HR close to the slope and its core is located deeper than in LR.

In general, AMOC differences in the upper 3,000 m between HR and LR are local and do not propagate further south, hence do not result in the changes in the total water transport between the hemispheres. Most of these differences are related to horizontal and vertical redistribution of flows due to topography or model dynamics and have no or little effect on net transport. The AMOC differences below 3,000 m are partly related to the Southern Ocean dynamics, in particular, to resolving (HR)/parameterizing (LR) of eddies.

Atmospheric resolution also has an impact on the simulated AMOC. Higher atmospheric resolution leads to weaker AMOC in response to weaker mean winds associated with higher cyclone activity. The change in atmospheric resolution affects the mid-depth cell of the AMOC. The likely reason for the observed AMOC decrease with increased atmospheric resolution is the change in the amount of mechanical energy that is transferred from the atmosphere to the ocean. The higher resolution atmospheric model is expected to better reproduce details of medium-scale atmospheric circulation including better representation of cyclones. In general, changes between runs with the same ocean but different atmosphere show that storm activity in T63 runs is noticeably weaker over the NA compared to T127 runs. Details of the spatial distribution of storm track differences depend on ocean resolution. Indeed, difference in ocean resolutions affects the spatial distribution of oceanic SST fronts and hence influences position of storm tracks.

3.1.3 The ocean-atmosphere coupling

Although generally much weaker than winds, surface oceanic currents' effect on atmospheric stress influences both the atmosphere and the ocean (“current feedback”, referred to as CFB). By reducing the energy input from the atmosphere to the ocean, the current feedback slows down the mean oceanic currents, but also induces a dampening of the mesoscale activity, with a sink of energy from eddies to the atmosphere. Historically, CFB effect was generally not implemented in climate models because of the coarse resolution used and the slow oceanic surface currents generated. This is the case of EC-Earth3P configuration, as well as some other climate models used in the CMIP exercise. With EC-Earth3P-VHR configuration, where oceanic mesoscale activity is explicitly reproduced, this hypothesis has been revisited. A companion simulation EC-Earth3P-VHR-CFB (cf. Table 1) that includes

CFB has been run to evaluate for the first time on a long time frame and at eddy-resolving resolution the role of CFB on the ocean and on climate.

We first described a proper framework to implement the CFB in climate coupled models (Renault et al., 2019). This must be done in two distinct steps: 1) calculation of air-sea fluxes using relative winds, i.e. the difference between the near-surface winds and the surface oceanic currents, instead of absolute winds, 2) use of relative winds also involves a modification of the tridiagonal problem associated with the discretization of the vertical turbulent viscosity.

This simple but however important parameterization of air-sea exchanges (at negligible computational cost) has important effects on ocean dynamics, particularly on western boundary currents (WBCs) areas like the Gulf Stream (North Atlantic), the Kuroshio (North Pacific), the Agulhas (South Africa) and to a lesser extent in Eastern Australia or the Malvinas' region (Figure 4). The order of magnitude of SST changes induced by CFB is the same as for a change between EC-Earth3P-VHR and EC-Earth3P-HR (not shown).

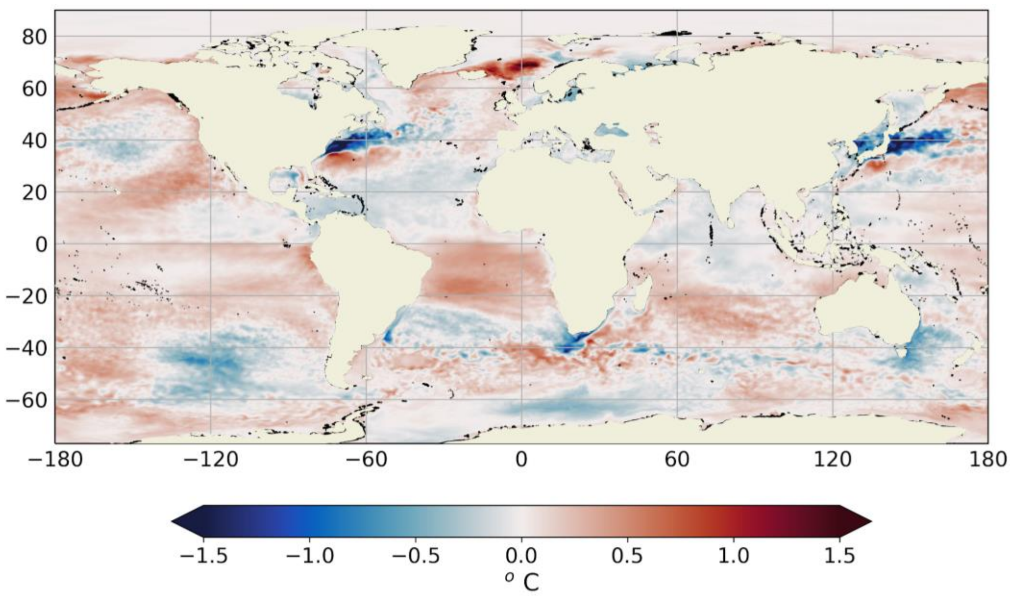

Figure 4: SST difference over 25 years between EC-Earth3P-VHR and EC-Earth3P-VHR-CFB.

These dynamical changes have further implications in a climate perspective. In particular, they induce important local precipitation changes in the WBC areas (not shown), but also a shift from North to South in the precipitations in the ITCZ (Figure 5). The mechanisms involved in these modifications are currently under investigation.

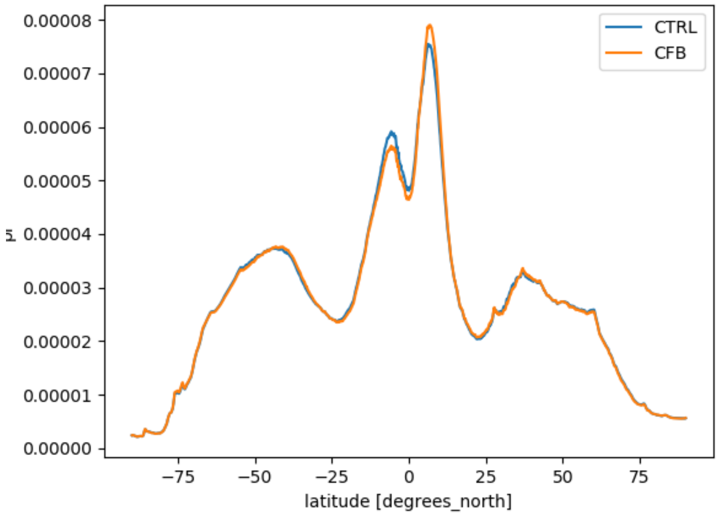

Figure 5: Zonally integrated precipitations (in kg.m-2.s-1) over 25 years between EC-Earth3P-VHR (CTRL) and EC-Earth3P-VHR-CFB (CFB).

3.1.4 Explicitly representing the ocean mesoscale and enabling comparison with observations

3.1.4.1 Large-scale circulation

The introduction of the Gent-McWilliams parameterisation scheme (Gent and McWilliams, 1990, henceforth GM90) into low resolution ocean models transformed their performance, since it enabled removal of cross-density surface mixing in the ocean interior, which caused many model biases. The scheme removes available potential energy (APE) by slumping density slopes adiabatically, with this energy removed from the system. This mimics part of the real world process of conversion of APE into mean and eddy kinetic energy (KE) which happens at the ocean mesoscale and can energise both the small and large scales of the flow. Such processes are particularly important near ocean boundary currents, along the Antarctic Circumpolar Current (ACC) and in ocean eddies. Such schemes have been used in ocean models for more than 20 years and hence are well understood and well tuned.

Such schemes have drawbacks if we want to represent the real ocean, which has an energetic peak in the mesoscale, with ubiquitous eddies and energy transfers between different scales and processes. One can try and compare the large-scale mean state of a low resolution model with long term mean observations, but more process-oriented analysis is very difficult.

As we increase ocean model resolution to eddy-present (~1/4°) and eddy-rich (1/10° and above), and begin to explicitly represent mesoscale processes, we have to choose what to do with such parameterisations. Many groups switch off GM90 and assume that these processes are now sufficiently resolved. We can see from Bock et al. (2020) that typically this causes problems in eddy-present models with simulation in the Southern Ocean at latitudes where the mesoscale becomes important but is now poorly resolved and not parameterised. This is also illustrated in Figure 6 (taken from Hewitt et al. 2020) showing the simulation of the Antarctic Circumpolar Current (ACC) and the Northward Heat Transport (NHT) in the Atlantic at 26.5°N from a range of models using HighResMIP and the CMIP6 OMIP experiment. As we increase ocean resolution, the ACC transport typically moves away from the observed range, while the NHT tends to stay the same or move towards observations.

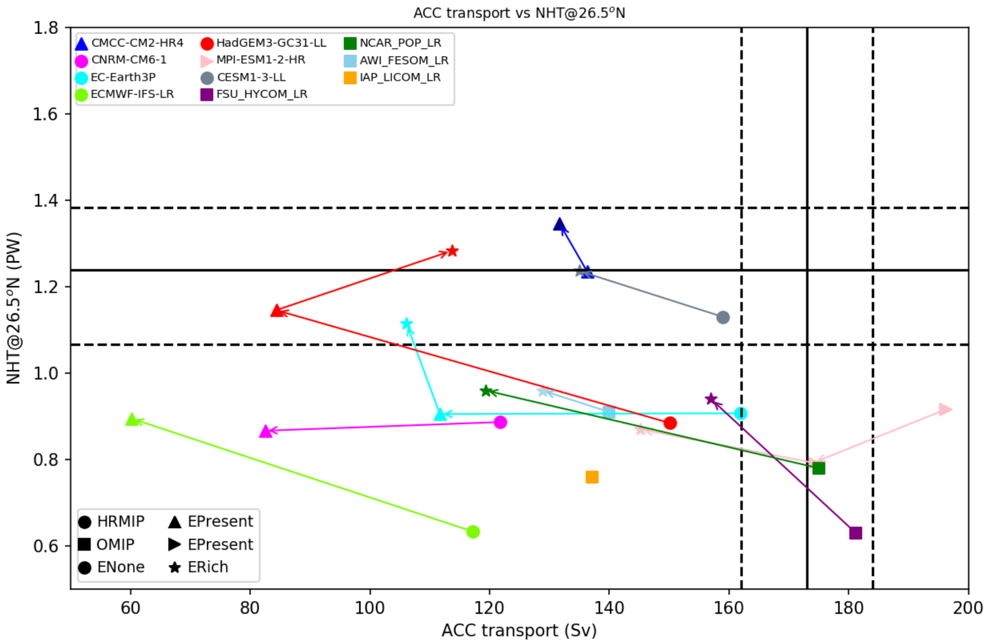

Figure 6: The long term mean values of the ACC transport and the NHT transport at 26.5°N from a range of models and resolutions, both from the control-1950 HighResMIP simulation (starting from circles) and from OMIP (using JRA55 forcing, starting from squares). Different ocean resolutions use different symbols, and different models use different colours

When we reach eddy-rich resolutions, we tend to better represent the mesoscales and the eddy kinetic energy in the models further increases. Figure 6 shows that in general the ACC weakens slightly more (HadGEM3-GC31-HH being an exception here) while the NHT tends to further increase. Roberts et al. (2020,) showed that, for the Atlantic Meridional Overturning Circulation (AMOC), properties typically improve compared to RAPID-MOCHA observations at 26.5°N as model resolution is increased, but that further north in the subpolar gyre any improvement is less clear (Jackson et al. 2020). Ongoing work suggests that, perhaps particularly for NEMO models, excessive heat and salt gets into the Labrador Sea and causes excessively deep mixing there. This is exposing limitations of our understanding of model numerics and performance at resolutions where we have less experience compared to typical CMIP-type models.

In collaboration with the iHESP group (US groups at NCAR and Texas A&M University, and Qingdao, China), we have also begun to look at the processes of vertical heat transport in models with parameterised and resolved eddies. The total vertical heat transport is:

$$
\begin{aligned}
Total &= \overline{wT} + bolus\ (wT) + diffusion\ terms \\
&= \bar{w}\bar{T} + \overline{w'T'} + bolus(wT) + diffusion\ terms
\end{aligned}
$$

where we break down the total advection term $\overline{wT}$ into the mean advection, and the time-varying eddy component, which adds to the parameterised component (bolus term, only in LL), and the diffusion terms. The diffusion terms were not saved from model output, and so for now we assume they are not too different at high and low resolution. Figure 7 shows these terms at different resolutions in HadGEM3-GC31 control-1950 simulations. On the right, one can see that the upwards heat transport near the surface (red) is much stronger at eddy-rich HH resolution, and this is due to the strong eddy term (stronger and nearer the surface than the parameterised bolus term in LL. The downward heat transport at larger depths is also smaller in HH. This may explain some of the differences in deep ocean temperature bias in Roberts et al. (2019). Ongoing work suggests that this may also cause enhanced future near-surface warming in higher resolution models, with potential consequences for global surface temperature increase, and hence that explicitly representing these processes may be important for future projections.

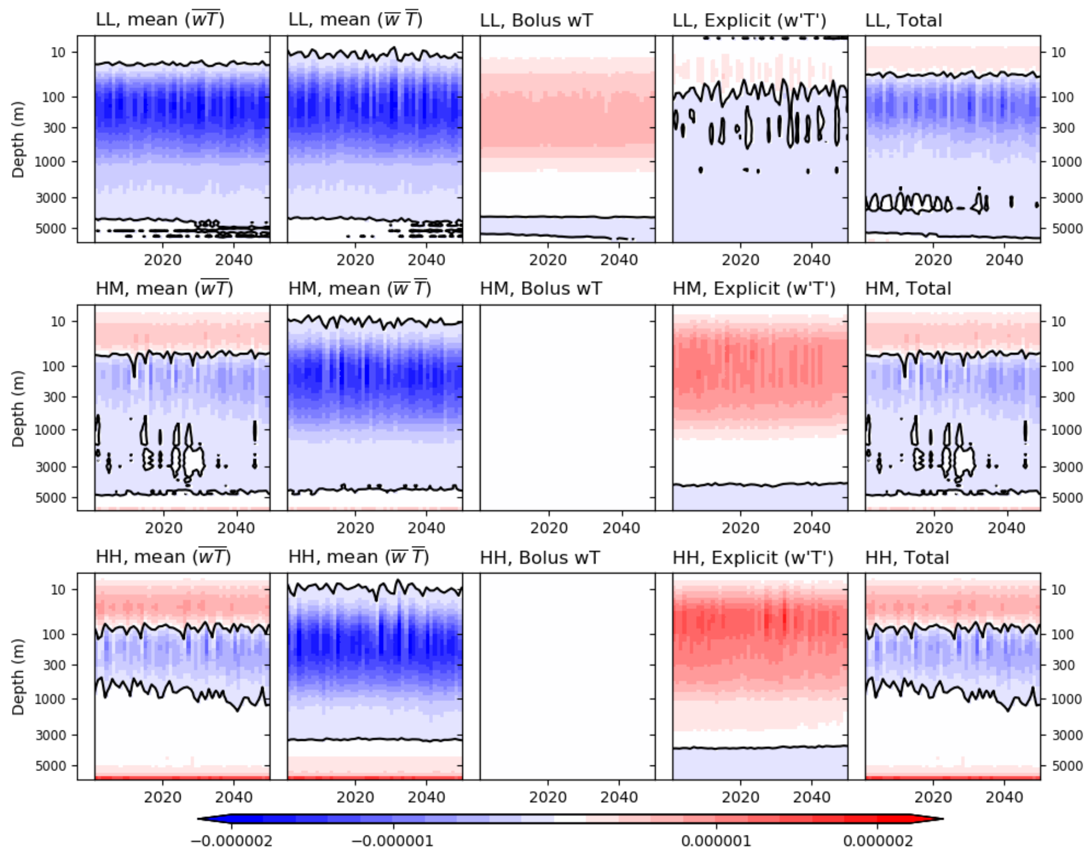

Figure 7: Vertical heat transport terms (columns) from the HadGEM3-GC31 model at different resolutions (rows).

3.1.4.2 Evaluation of eddy properties

It goes without saying that in order to compare the ocean eddy properties with observations, it is necessary to represent the eddies. Moreton et al. (2020) tracked ocean eddies in both eddy-present and eddy-rich simulations of HadGEM3-GC31 and compared them to observations. The eddy-rich simulations produce around 40% more eddies than the eddy-present, with more genesis in active regions and longer lifetimes. Eastern boundary currents in particular have more activity, and these are important regions for transporting heat and nutrients into the nutrient-poor open ocean. There are regions of the ocean, such as South Africa and the flow of Agulhas Rings into the South Atlantic, whose processes cannot be represented by local parameterisations but have potential to be important for long term climate variability and change.

Further work (Moreton et al., in prep) will look at composites of eddy properties together with air-sea interactions, to begin to understand the role that eddies play in the climate system. Results so far suggest that the surface damping of the eddies is too low if the atmosphere resolution is much less than the ocean resolution, with potential implications for eddy lifetimes, heat and salt transport.

3.1.4.3 Summary

Moving from typical CMIP6-type resolutions to cutting-edge eddy-rich modelling causes model costs to increase some hundreds-fold, with storage costs increasing ten-fold. This clearly means that we can perform less simulations at such resolutions, less ensemble members, less future scenarios and less model tuning of the mean climate (important since we have less experience at these resolutions).

However, moving to these resolutions enables us to actually compare processes with observations and better understand model bias and mixing or poorly represented processes. Such resolutions can already demonstrate improved model bias and other performance aspects, but clearly not all the changes are improvements. Given that low resolution models have been developed over more than 20 years, it seems clear that the scope for improved simulation is much larger in these higher resolution models. It is possible that such models will produce different future projections due to the extra processes represented, but more models are needed to see how robust this is.

3.1.2 Approach for atmosphere and explicit convection

3.1.2.1 Explicit convection in global models

Representation of convective processes in the atmosphere are of fundamental importance to the whole climate system, both at local and large scales. Such processes happen at the scale of metres and below. However, typical global climate models operate with grid boxes of 100km or above. In such models there is no alternative than to parameterise the convective processes – this has remained a great challenge, and such parameterisations typically suffer from biases such as a poor diurnal cycle of precipitation, inadequate precipitation extremes.

Over the last 20 years some global modelling groups (e.g. NICAM in Japan) have attempted to increase global model resolution finer than 15km to make it possible to switch off the convective parameterisation and explicitly represent convection. More recently the DYAMOND project (Stevens et al. 2019) showed results from 8 different global climate models at sub-5km resolution run over 40 days of simulation, including both ECMWF and HadGEM3-GC31 models used in PRIMAVERA.

In order to try and study the impact of explicitly representing convection on climate, longer simulations are required. Currently this requires a compromise in resolution. The HadGEM3-GC31 model was run at around 10km (mid-latitude) resolution for about 22 years in total with various configurations of convection (parameterised, no convective parameterisation at all, and only the shallow and mis-level convection parameterised). On the other hand, EC-Earth3P-VHR model has been run 100 years with a resolution of 15km at mid-latitude.

Representing convection explicitly enables the addition of prognostic graupel (soft hail) into the model, which can be used as an input to calculate lightning frequency (Field et al. 2018). Over land the explicit convection model captures much of the spatial distribution of lightning, and its diurnal cycle. Over the ocean the lightning scheme is less successful, with too high frequency for reasons not yet understood. There are open questions about changes in lightning frequency in the future, but this has not been investigated at this time because the mean state of the model without the convection scheme is not considered adequate to justify future simulations.

A large problem in investigation of the representation of convection is that, being such a key part of a model simulation, the complete model tuning and performance is intimately entwined with it. This means that the model may need to be completely retuned for a different convective setup, making it difficult to analyse the processes. This is illustrated in Figure 8 which shows time series of radiation components in three simulations which have had just the convection parameterisation completely or partially removed. Such a large change to radiative balance will clearly have big impacts on the mean climate, which in general is worse compared to the control – at least in a global sense, there are some regional improvements.

Despite the global biases, what the explicit convection is able to improve upon is the phase of the diurnal cycle of precipitation (Birch et al. 2014). The standard parameterisation has peak precipitation in the tropics around noon, rather than typically later in the afternoon as seen in observations. Similarly, in summer over central and southern Europe, as shown in Figure 9, the observations generally show a peak in mid-late afternoon. This is much better captured by the explicit convection simulation. More details of the science performance are included in D4.4.

3.1.2.2 Summary

Going beyond the parameterisation of convection is currently very challenging. With no parameterisation, each model grid square must be forced to explicitly convect, but since model resolutions beyond ~10km are very expensive, this is a very large amount of mass to convect. Regional models now routinely move into the convection permitting modelling space (CPM) using resolutions of a few kilometres. Such resolutions are clearly an improvement on the global, though such models have their own biases and errors that need to be addressed as experience with such simulations grows.

Aspects of global model simulation can be improved or enabled when going beyond standard parameterisation. The phase of the diurnal cycle of precipitation is improved, which may well be important for radiation budgets, and perhaps land surface properties given that rainfall happens later in the day. The simulation of lightning is enabled, and over land the frequency can be well-represented in a 10km resolution model. However, the mean state and biases of such explicit convection models are generally much worse than the parameterised convection, with the models likely to need complete retuning given how important convection is to the global climate system. Further experience is needed to improve such explicit convection models for them to be as good or better than parameterised models for

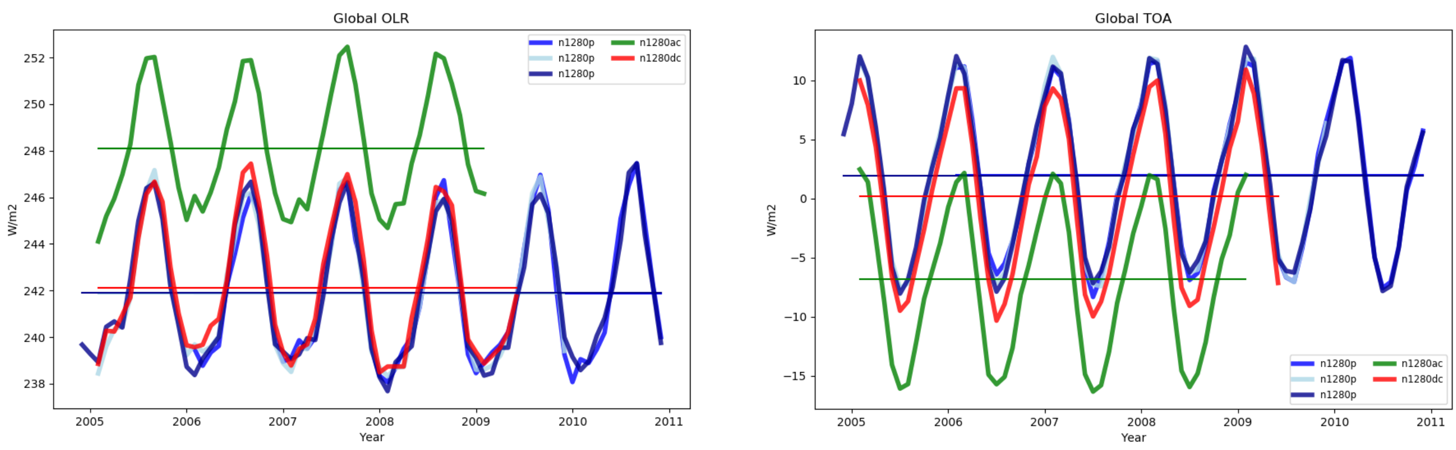

Figure 8: Time Series of the monthly (top) outgoing Longwave (OLR) and (bottom) Top of Atmosphere radiation (TOA) from the three N1280 simulations with different convection representations. n1280p is standard parameterised, n1280ac is all explicit convection, and n1280dp is explicit deep convection and parameterised shallow and mid-level. The straight lines are the mean values over the timeseries. Observed values from CERES-EBAF suggest OLR ~ 240 W/m2 and TOA ~ 0.5 W/m2.

the most important climate variables.

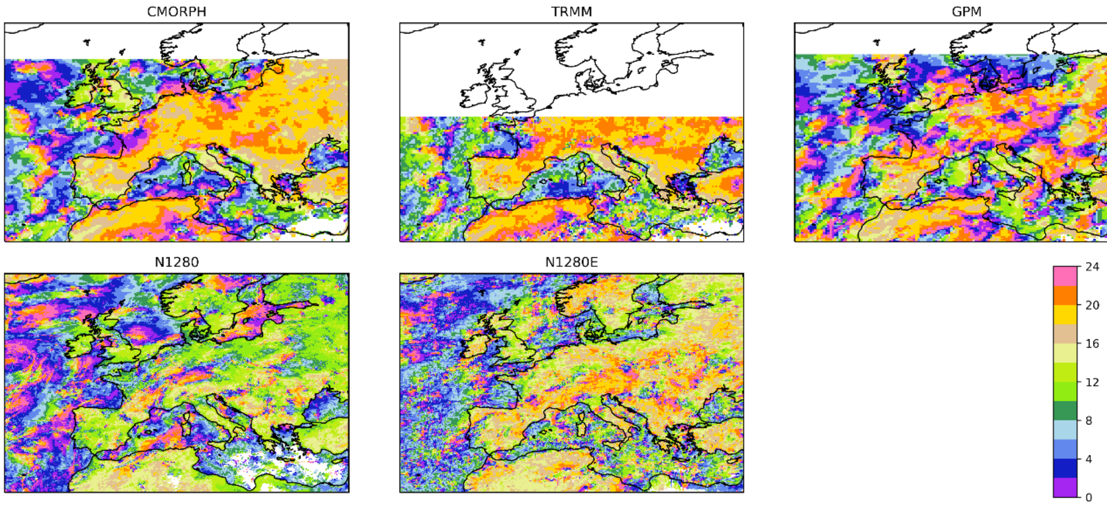

Figure 9: The timing of peak precipitation (local time) for July-August-September for three observational datasets (CMORPH, TRMM, GPM) and models with standard parameterised convection (N1280) and explicit convective (N1280E).

3.2 Stochastics parameterizations

Three configurations were implemented and tested in the EC-Earth3P model: stochasticity in the atmosphere only, using the standard SPPT scheme; one with stochasticity in ocean and sea-ice components only; finally, one with stochasticity in atmosphere, ocean, sea-ice and the land-component of the model. Further details of each configuration can be found in Table 2, including how each configuration is labelled in discussion/figures. In each case, 3 ensemble members were generated following the hist-1950 and highres-future scenarios, thereby covering the period 1950-2050 with historical and projected forcings. An equivalent ensemble was generated using the default EC-Earth3P deterministic configuration. This control ensemble, along with the single member of EC-Earth3P-HR (high-resolution deterministic) are what the stochastic schemes are compared against. In all cases, the resolution is that of the low-resolution EC-Earth3P model, i.e. spectral resolution T255, 91 levels in the vertical and a 1 degree ocean. Note that since EC-Earth3P is the only PRIMAVERA model which has implemented these schemes, we restrict comparison to the EC-Earth3P model alone, including when comparing against improvements due to resolution.

We first quantify resource usage, then examine and attempt to quantify improvements, before drawing our conclusions. Because the number of topics considered over the course of PRIMAVERA is large, the discussion is necessarily kept brief.

Configuration	SPPT	Stochastic ocean + sea-ice	Fully probabilistic Earth-System Model
Abbreviation (experiment IDs)	SPPT (stc1, stc2, stc3)	OCE (otc1, otc2, otc3)	FESM (ftc1, ftc2, ftc3)
Description	The `stochastically perturbed parameterisation tendencies' scheme, used in the Integrated Forecast System (IFS) weather forecasts. See Palmer et al. (2009). Physics tendencies are multiplicatively perturbed by a random pattern with a decorrelation time-scale of 6hrs and length-scale of 500km. The standard deviation of the perturbation is 0.52 and clipped to always be between 0 and 2.	The schemes from Juricke et al. (2013), (2014), (2017). These add multiplicative perturbations to both tendencies and parameters of certain components of NEMO and LIM: The vertical diffusion, Gent McWilliams and turbulent kinetic energy mixing schemes (NEMO), the P* (sea-ice strength) scheme (LIM)	SPPT and stochastic ocean+sea-ice are turned on with default parameters. Land-component H-TESSEL has a stochastic parameter perturbation scheme added: the uncertain soil moisture parameters are multiplicatively perturbed, following the methodology of MacLeod et al. (2016).

TABLE 2: The stochastic schemes considered. The deterministic experiments are collectively referred to as CTRL (control).

3.2.1 Quantification of Resource Usage

Table 3 quantifies the resources associated with these schemes, as well as for the high-resolution EC-Earth3P-HR configuration. It can be seen that the most comprehensive set of stochastic schemes leads to an increase in runtime of less than 4 %, while the high-resolution model is associated with an almost 500% increase in cost, a difference in two orders of magnitude. It is important to note that while the low-resolution experiments all used the same number of cores, the high-resolution simulation used significantly more, implying that 500% is a large underestimate of the true difference. Because the choice of number of cores is user optional, one cannot give a definitive number. The choice made for the EC-Earth3P-HR simulation was based on optimising the runtime/core ratio: with this choice, the percentage increase in cost associated with increased resolution is around 1200%. As a result, even an improvement due to stochasticity which is an order of magnitude smaller than a comparable improvement due to resolution, is still highly valuable when factoring in this disparity in resource usage. Improvements that are directly comparable or exceeding that of higher resolution must be judged as extremely valuable and represent strong evidence for the usage of these schemes as a default in climate model simulations.

Stochastic parameterisation scheme/Hi-res	SPPT	Stochastic ocean + sea-ice	Fully stochastic (atmosphere, ocean, land and sea-ice)	Increased horizontal resolution (EC-Earth3P-HR)
Percentage increase in runtime relative to low-resolution deterministic	2.7%	0.8%	3.8%	465%

TABLE 3: The computational cost of different stochastic schemes and increased resolution. The runtime here is measured as the time to finish a single model year. For the low-resolution simulations, the number of cores used was identical for all configurations. For the high-res model, a significantly larger number of cores were used.

3.2.2 Quantification of benefits

We now examine the impact of adding the various stochastic schemes to EC-Earth3P in the hist-1950 coupled simulations; impacts on forced-SST simulations was done prior to PRIMAVERA project, so only fully coupled simulations are considered here. In some cases, a clear quantification of impact can be made (e.g. by considering reduction in biases against a fixed source of observations), while in others this is less clear (e.g. where observations are poorly constrained). We therefore present results and then draw conservative conclusions at the end.

3.2.2.1 Pre-PRIMAVERA Results

Two main pieces of work relevant for WP4 were carried out early on in PRIMAVERA with the use of pre-PRIMAVERA datasets. The first, Strommen et al. (2019), considered the impact of SPPT, the stochastic land-scheme and the Independent SPPT scheme (ISPPT) on EC-Earth3P, using forced-SST simulations. These results were found to be broadly positive on the mean state and variance, with associated reductions in RMS errors for some key variables compared to observations. However, the impact was found to be too large for some large-scale quantities such as global mean atmospheric water vapor. A key conclusion of this paper was that models should be re-tuned once a stochastic scheme has been added: the experimental protocol of PRIMAVERA was such that a tuning of this sort would have made for an unfair comparison against the HR model, so was not carried out. As comparable HR simulations were not available during the work of ibid, a comparison was not made to HR. However, many of the mean state improvements are replicated also in the coupled simulations considered in PRIMAVERA proper, so it is plausible that these stochastic schemes would perform competitively with many aspects of HR also in the forced-SST configuration.

The second paper, Strommen et al. (2019c), showed that SPPT can change the climate sensitivity of EC-Earth3P. In that case, changes to low-cloud cover and cloud feedbacks reduced the overall sensitivity of the model when SPPT was activated. However, changes were not appreciable until after 2050, when the RCP8.5 scenario forcing becomes increasingly strong, so we did not a priori expect to see significant differences in the hist-1950 scenario. As we will see in the next section, this expectation was met after the PRIMAVERA experiments were analysed. However, it is important to remember that in longer term projections, changes in sensitivity may emerge that are non-negligible (of the order >10%).

3.2.2.2 Energy budget changes

A key conclusion of both Strommen et al. (2019) and Strommen et al. (2019c) was the fact that stochastic schemes can notably change a model's energy budget. We therefore computed the full energy budget across the three configurations, decomposed into the individual components. The result can be seen in Figure 10. The changes for SPPT and FESM are consistent with those in ibid, with the dominant impact being a reduction in total solar radiation. This is in turn explained by a rapid increase in cloud liquid water in these experiments, which make the clouds more reflective; this increase is a result of the interaction between the mean-zero stochastic perturbations and the highly non-linear process of condensation. The net impact is around -0.2 W/m2 in both cases, which while statistically significant, is relatively modest in magnitude. The OCE configuration has no significant impact on the overall energy budget. The impact of HR (not shown) is notably larger and required extensive tuning to bring into closer line with observed values. It is an important take-home message that stochastic schemes, while necessitating further tuning, are generally likely to be significantly less severe in mean state impacts and therefore require a considerably less strenuous tuning effort.

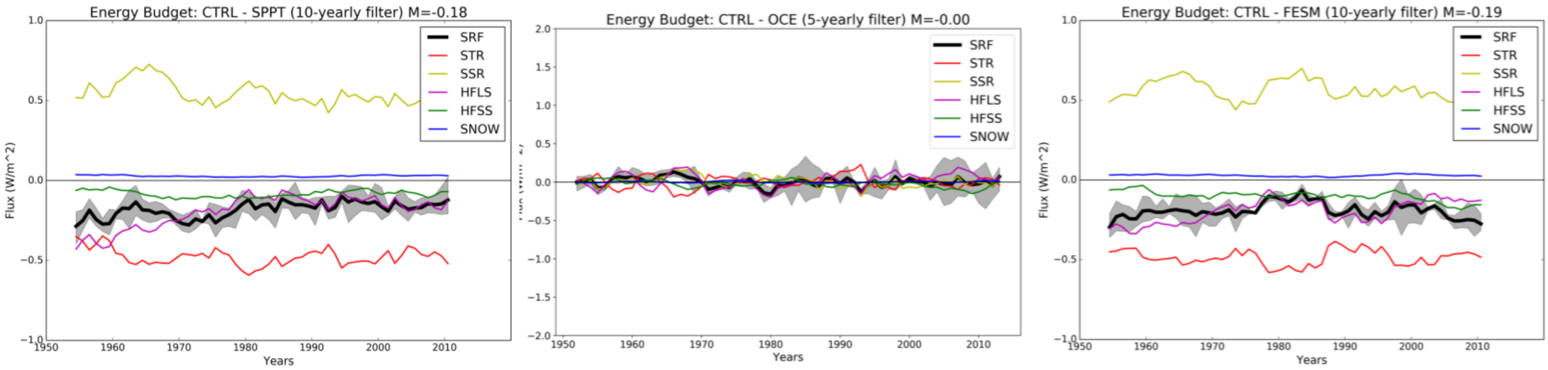

Figure 10: Difference between surface fluxes (CTRL minus SPPT on left panel, CTRL minus OCE on the middle panel, CTRL minus FESMon the right panel). Thermal radiation (STR, red), solar radiation (SSR, yellow), sensible heat flux (HFSS, green), latent heat flux (HFLS, mauve) and latent heat release of snow (SNOW, blue). Net surface energy, the sum of these terms, is shown in black (SRF).

3.2.2.3 Mean state and variance biases

In order to keep the discussion more contained, we restrict to four key variables deemed of particular importance for the global scale climate, namely sea-surface temperature (tos), two-metre temperature (tas), total cloud cover (clt) and precipitation (pr).

Figure 11 shows the impact due to each stochastic configuration on the long-term monthly mean and variance of sea-surface temperature (tos). EC-Earth3P shows common mean state biases, including a too-warm North Pacific, a too-cool North Atlantic and a slightly too-warm Indian Ocean basin. There is also evidence of the split ITCZ bias. All four of these biases are alleviated in both the SPPT and FESM configurations. In OCE, the main bias reduction is a substantial warming of the North Atlantic Ocean, of particular importance for realistic representations of European climate. While OCE by itself amplifies to some extent biases in the tropical Pacific, these are cancelled out by the impact of SPPT in the FESM configuration. The effect is also positive on model variance, with a decrease in unrealistically high variances over the North Pacific/Atlantic (not shown). Overall, while HR configuration is overall doing better than FESM in terms of RMSE and variance, FESM is reducing the biases by about a third of what HR does.

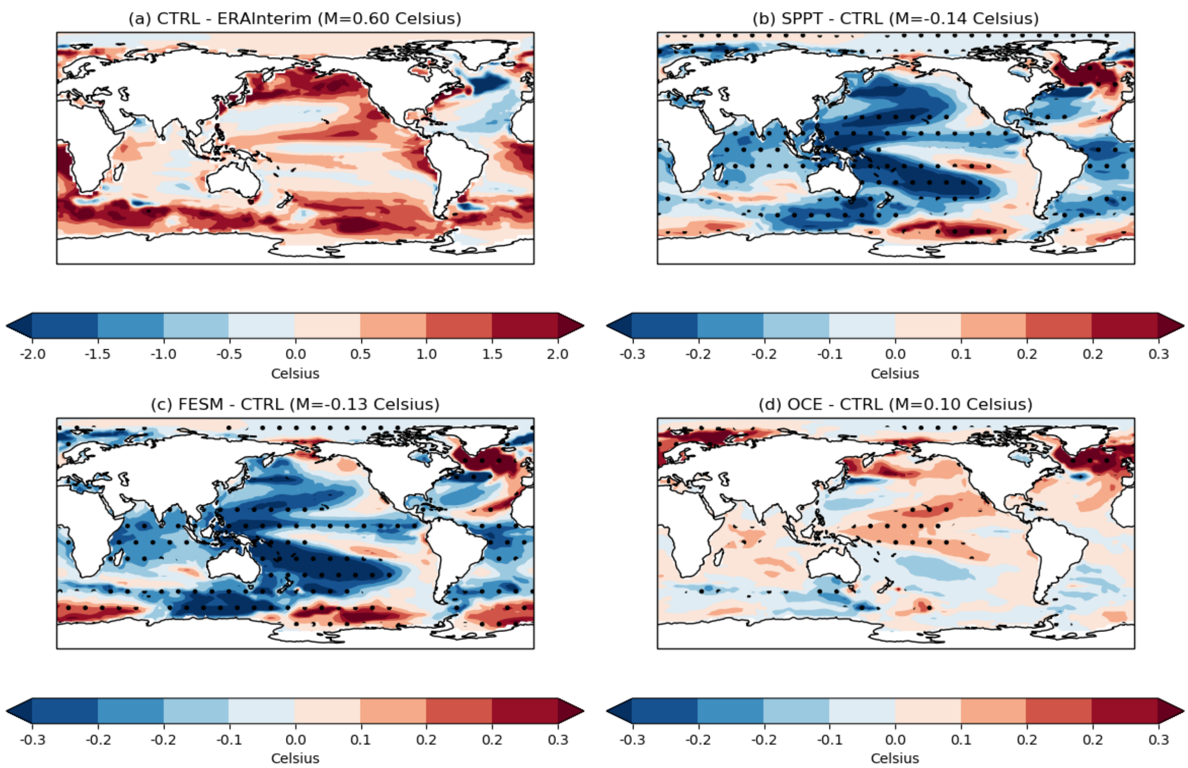

Figure 11: Impact of stochastic schemes on sea-surface temperature (tos) means, 1980-2015. In (a), CTRL minus ERA-Interim, (b) SPPT minus CTRL, (c) FESM minus CTRL and (d) OCE minus CTRL. Dots indicate regions where the change is significant to a 95% confidence interval. In (b), (c) and (d), locations where the colour is opposite to that of (a) indicate a reduction in bias.

Interestingly, comparable improvements in surface temperature (tas) are not easy to identify, implying that the improved tos state is not translating up to the atmosphere. This suggests changes to air-sea interactions in the SPPT and FESM configurations, something we revisit later. Overall, while HR makes the spatial pattern bias worse, it does appear to improve the timeseries evolution in a manner which none of the stochastic configurations achieve.

Turning to total cloud cover (clt, Figure 12), we found no significant impact from the OCE configuration, while SPPT and FESM have broadly similar patterns, unsurprisingly perhaps as the atmospheric component is likely having the dominant effect on cloud formation. Both FESM and HR are associated with notable reductions in biases of both the mean and variance clt. FESM mostly improves the tropical Pacific, a change likely interlinked with improvements to the split ITCZ (discussed below). While FESM is able to improve variability of the clouds that the CTRL model produces, it cannot trigger a step change in regional formation associated with specific cloud types. Nevertheless, the improved mean state due to FESM (measured in terms of reduction in RMS) is around half of what HR achieves, a major success.

Finally, we consider precipitation. The CTRL model shows the classic `split ITCZ' bias, common to most climate models. FESM acts partially to alleviate this, though not as much as HR. Indeed, inspection of RMS changes (not shown) suggest that the net effect of FESM is close to 0, as the rain-band is broadened too far southwards, leading to an effective cancellation between a reduction in bias close to the equator and an increased bias further south (as well as over the Indian Ocean). We note that the PESM configuration (with Independent SPPT as opposed to regular SPPT) performed notably well here (not shown), performing competitively with HR in terms of bias reduction. This suggests that short-term variability introduced by FESM/SPPT is probably too large and too autocorrelated in time/space, leading to overly large changes on monthly timescales.

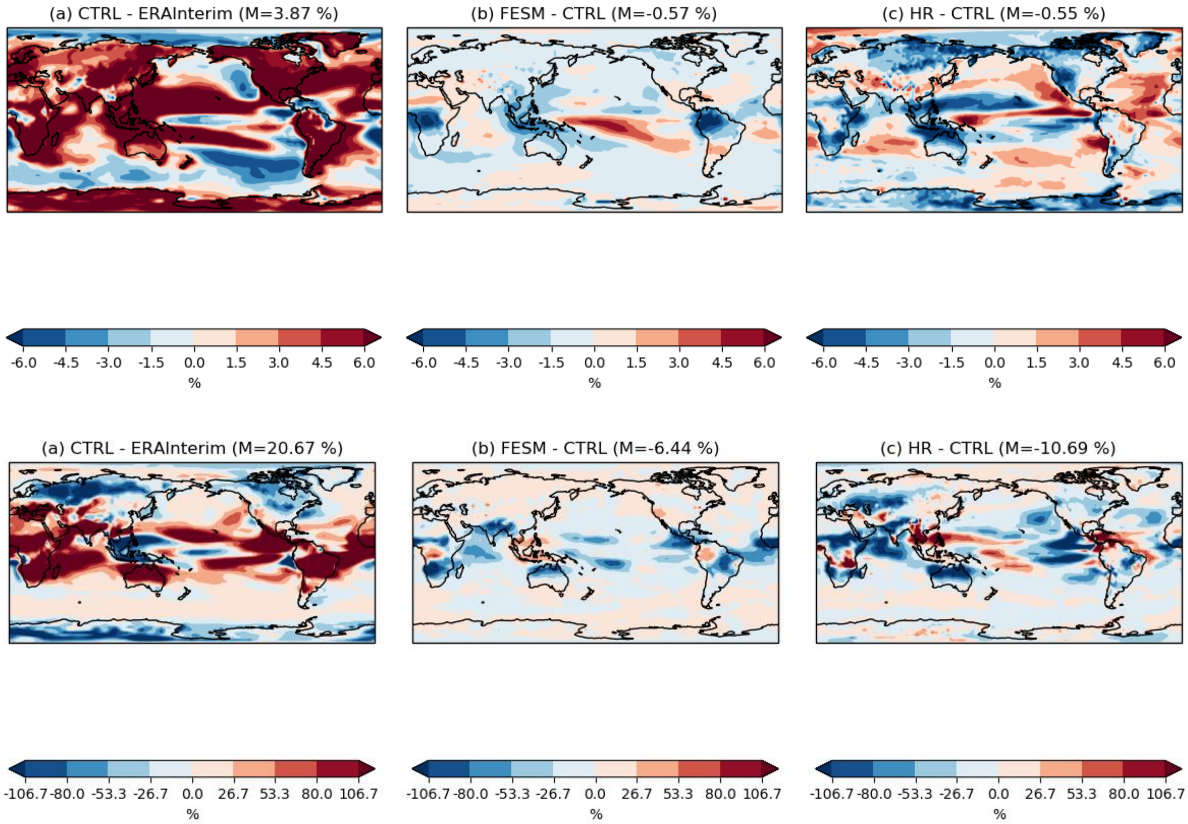

Figure 12: The total cloud cover (clt) monthly mean field (1980-2015) (top panel) and monthly variance field (1980-2015) (bottom panel). In (a), the bias CTRL minus ERA-Interim; in (b) FESM minus CTRL; in (c) HR minus CTRL. Regions in (b) and (c) where the colours are opposite of that in (a) indicate a reduction in the bias.

3.2.2.3 Representation of Euro-Atlantic Weather Regimes

Recall here that Euro-Atlantic weather regimes are thought of as distinct and quasi-persistent regions of phase space of geopotential height at 500hPa (Z500) in the Euro-Atlantic region: see Strommen et al. (2019b) and references therein. Clustering algorithms typically produce 4 regimes: a positive and negative NAO pattern (NAO+ and NAO-), an Atlantic Ridge (Atl. Ridge) pattern and a Scandinavian Blocking (Sc. Blocking) pattern. Their role in modulating European weather and climate is well known and a topic of frequent study.

It has been shown in earlier studies (e.g. Strommen et al. 2019b) that increased horizontal resolution tends to improve some aspects of regime structure but not others. In particular, the sharpness of the regimes (i.e. the extent to which the phase space of Z500 splits up into distinct and well-separated regimes) appears to go up robustly with resolution, but the location of the regimes in phase space (i.e. the appearance of the spatial patterns themselves) can still suffer from consistent biases. In Dawson et al. (2015) some evidence was found that stochastic physics can improve the regime structure as well, on par with the impact of increased resolution. In ibid though only a particular model was used with only one ensemble member. We strengthen the conclusion by utilising the multi-ensemble PRIMAVERA experiments, and find both an increase in sharpness (or significance) of regime structure as well as a consistent improvement in the regime patterns with SPPT (as well as FESM).

This suggests that stochastic physics can lead to broad improvements in Euro-Atlantic regime structure. The impact here appears to be greater than HR (not shown), which improves sharpness but not the spatial patterns.

3.2.2.4 Arctic sea-ice

The introduction of a stochastic sea-ice scheme immediately invites for analysis of the impact on Arctic sea-ice concentration. Figure 13 shows the timeseries evolution of siconc across the hist-1950 simulations, compared to recent satellite estimates. It can be seen that HR follows a notably different trajectory, having started from a state of significantly lower sea-ice than the low-resolution configurations. Because observations are less reliable prior to the satellite era (1980 onwards), it is hard to conclude if this is positive or not. In the period 1980 onwards, it is generally the OCE configuration that tracks observations most closely. Much of this can be accounted for by the fact that the OCE experiments show a more realistic distribution of siconc during the peak month of September (not shown), where HR systematically overestimates the amount of sea-ice in general. It should be noted that looking at specific Arctic basins reveals that these impacts are not homogenous, with HR underestimating the total sea-ice for e.g. the Labrador Sea. The CTRL bias is broadly one of too much sea-ice, which SPPT and HR exacerbate. Only the OCE and FESM configurations serve to reduce the mean, albeit only minimally.

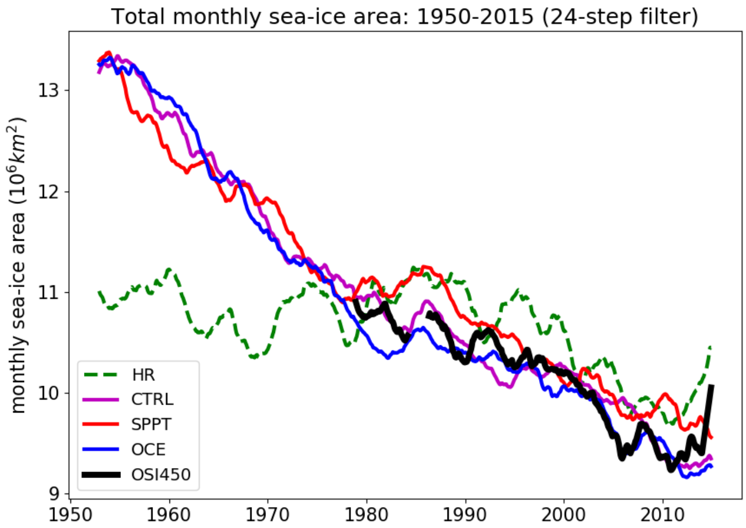

Figure 13: Time series evolution of Arctic sea-ice concentration (siconc). The observational dataset OSI450 is shown in black for comparison. A 24-month running mean has been applied to smooth the time series.

3.2.2.5 The Atlantic Meridional Overturning Circulation (AMOC)

Because of the influence of the mesoscale (eddies and boundary currents) on the AMOC, which are not resolved adequately until the resolution reaches around 1/12°, lower resolution models tend to have too weak an AMOC. It is therefore of natural interest to evaluate if stochasticity, particularly stochastic ocean-schemes, can impact on AMOC strength and variability.

OCE has only a minimal impact, increasing the mean circulation by around 0.5 Sv, while SPPT increases it by around 1 Sv. The impact of FESM was approximately the sum of the two. Much of the impact due to SPPT is taking place from around 1980 onwards. More generally, there is evidence of strong non-stationarity, which makes it hard to assess the significance of changes seen. However, we do find increased variance of surface winds in the upwelling region of the Labrador Sea with SPPT/FESM, implying increased mixing and an associated increase in the AMOC strength. The change due to SPPT is therefore likely to be robust and due to the increase in wind-speed variance.

We note that in both cases, the magnitude of the change is notably lower than that associated with switching to an eddy-resolving ocean. This is suggestive again of the notion that stochastic physics can only improve the variability of processes already being simulated, which may not always be sufficient to produce a step change in behaviour associated with increased resolution.

3.2.2.6 Tropical Cyclones

In ongoing work with UREAD, it has been found that the number of tropical cyclones goes up in the SPPT and FESM configurations. Figure 14 (taken from Vidale et al. 2020, submitted) shows an example of this using Pre-PRIMAVERA data from the SPHINX project, which had the benefit of including simulations with and without SPPT at different resolutions, as opposed to only one resolution for the PRIMAVERA simulations. It can be seen that adding SPPT increases the track density systematically. Quantitatively this increase is on the order of ~15%, and is roughly equal to the increase seen when passing from CTRL to HR. While this is also a noteworthy improvement due to SPPT, Figure 14 indicates that the difference is not resolution-independent and may be to some extent an example of an indiscriminate impact of the scheme. As a result, this may be a major improvement for some resolutions and less so for others. Further work is ongoing to shed more light on this, and better understand the respective roles of stochastic physics and model resolution on tropical cyclone genesis and development.

While Figure 14 shows this impact in pre-PRIMAVERA data, we verified that the same change occurs also in the PRIMAVERA version of EC-Earth3P. It is similar whether using SPPT or FESM, with no real change due to OCE. Vidale et al. (2020, submitted) also showed a similar increase in tropical cyclone frequency (~20%) in the HadGEM3-GC31 model when atmosphere stochastic schemes were included.

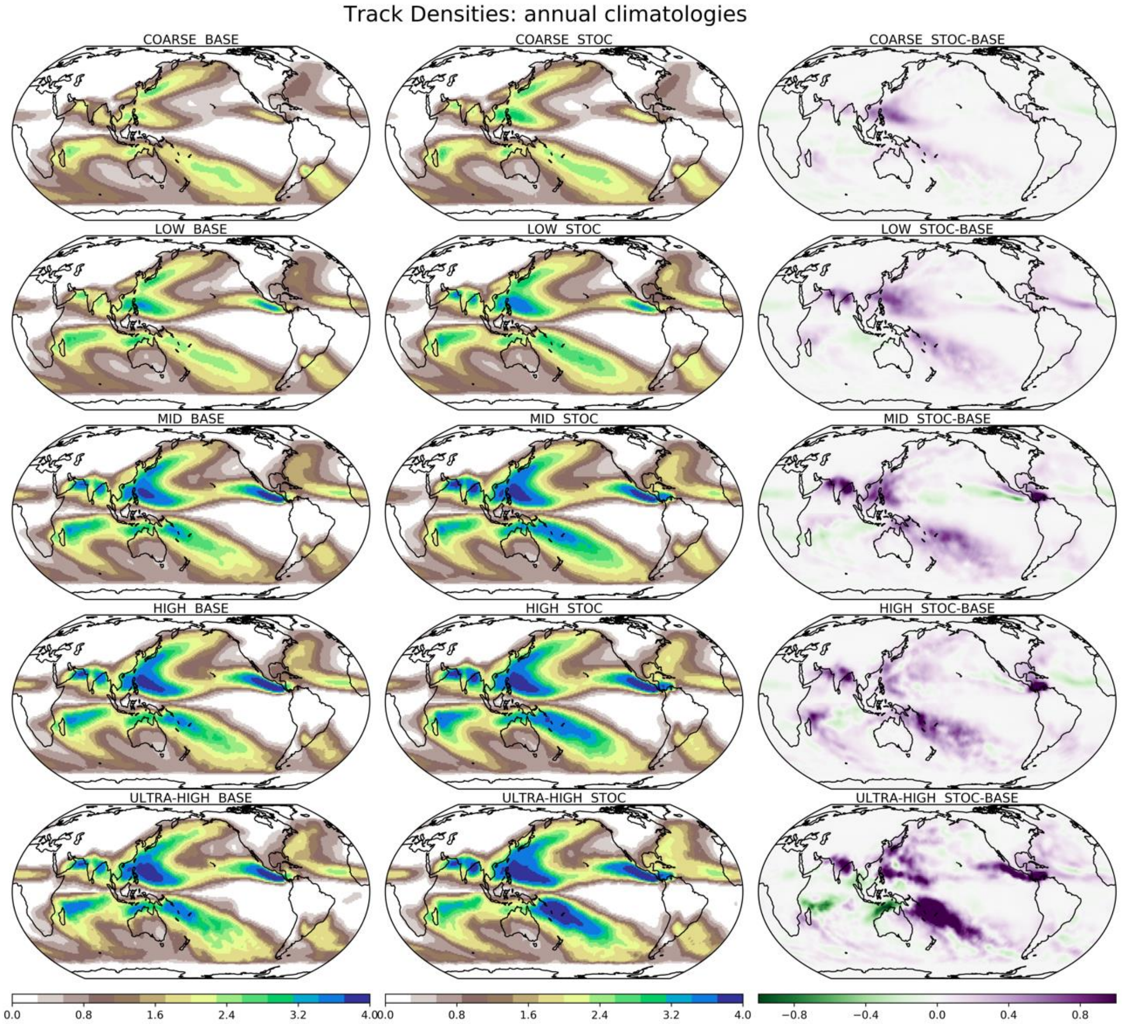

Figure 14: Tropical cyclone density for EC-Earth3P+SPPT at varying resolutions (lowest resolution at the top, highest resolution at the bottom). The left column shows CTRL simulations, the middle column shows with SPPT, and the right column shows the difference between the two.

3.2.2.7 Conclusions

Quantifying numerically the benefits of stochastic schemes is highly non-trivial, for the same reasons that numerically evaluating the quality of any climate model is non-trivial. Nevertheless, when evaluating changes to the model mean state and variance on climate timescales, we found that for several key variables (sea-surface temperature, cloud cover and precipitation), stochastic physics can achieve between 1/3 to 1/2 of the total bias reduction associated with HR. In some cases (surface temperature), HR increases the bias while stochastic physics has no significant impact. In other cases, stochastic physics can add a bias that HR does not; for example, SPPT tends to indiscriminately induce a big increase in cloud water which may, depending on the CTRL bias, be either an improvement or degradation of the mean state (Strommen et al. 2019c). It is therefore to be expected, in line with the conclusions of Strommen et al. (2019), that some tuning will be required when turning on stochastic schemes, but as shown in Section 3.2.1., this tuning is likely to be mild compared to the efforts associated with a step change in resolution. Because of the extremely low resource cost associated with stochastic schemes, a strong argument can therefore be made that these schemes are extremely cost-effective at reducing biases and improving variability.

Besides mean state improvements, we found improvements in the representation of precipitation extremes over Europe, Euro-Atlantic weather regimes and teleconnections to the North Atlantic Oscillation. Of particular relevance for European climate processes are these two latter improvements. The regime structure becomes more robust with SPPT/FESM, in a manner resembling the improvement due to HR. Unlike HR however, which often degrades the regime patterns, we also found an improvement in the spatial patterns of the regimes with SPPT/FESM. This is likely related to a more severe shift of the mean zonal wind fields associated with HR; the SPPT/FESM experiments appear to improve the underlying regime structure without perturbing the mean state as much, leading to a more straight-forward improvement. The fact that NAO teleconnections are better represented may also be related to this and has important implications for the representation of European climate variability in and of itself. The improvement due to stochastic physics appears in both cases to be comparable to (and possibly slightly better than) HR, though it is important to keep in mind that this conclusion might change after a more concerted tuning of the HR model.

In terms of ocean variability, the OCE scheme does tend to improve variability, in accordance with previous studies. No meaningful change was found to the AMOC however. A better job seems to have been done by the sea-ice scheme, which improves some important aspects of Arctic sea-ice concentration, including mean state and variance. In particular, the behaviour during the formation period September-October matches observations more closely with OCE than for any other configuration (HR included), a result which may be closely linked with the improved teleconnection from the Arctic to the NAO. This is a major success of the stochastic sea-ice scheme which will be studied in more detail in upcoming work. The impact of HR on this teleconnection is hard to compare against, due to the disparity in ensemble size, but based on the 1 ensemble member available is at best equal in magnitude to OCE.

A final notable impact is the increase in tropical cyclone count due to SPPT/FESM. Both HR and these configurations lead to a comparable increase in their count, in both cases a clear improvement relative to CTRL. However, preliminary analysis suggests that the mechanism is different in both cases, with the impact of SPPT/FESM mainly due to the changed mean-state of atmospheric water vapour.

In conclusion, UOXF have carried out a broad and extensive amount of analysis of the impact of three different stochastic configurations on EC-Earth3P. We find improvements across most of these, often comparable to or even matching changes due to increased resolution. Overall, evidence suggests that HR has a greater positive impact. Indeed, the simple picture emerging is that the stochastic schemes considered here can improve the variability of the processes already simulated by CTRL, but cannot force CTRL to resolve smaller-scale processes. Some key processes are therefore improved by HR but unchanged with stochasticity (e.g. AMOC variability). However, because the resource cost associated with stochasticity is several orders of magnitude lower than the cost of HR, the benefit/cost ratio for stochasticity is almost certainly very high. Because the overall change to the models energy balance is small, it is likely that any additional tuning required to smooth away edges would be mild. It is therefore the strong recommendation from UOXF that EC-Earth3P can benefit significantly from adding any of the configurations tested to the default version of the model. In particular, the fully stochastic configuration FESM, may represent a permanent `upgrade’ to the model.

REFERENCES

Birch, C., M. Roberts, L. Garcia-Carreras, D. Ackerley, M. Reeder, A. Lock, 2015: Sea breeze dynamics and convection initiation: the influence of convective parameterisation in weather and climate model biases. J. Clim, 28, 8093-8108. doi: http://dx.doi.org/10.1175/JCLI-D-14-00850.1.

Christensen, H. M., Lock, S. J., Moroz, I. M., & Palmer, T. N. (2017). Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society*, *143*(706), 2168–2181. https://doi.org/10.1002/qj.3075

Dawson, A., & Palmer, T. N. (2015). Simulating weather regimes: impact of model resolution and stochastic parameterization. *Climate Dynamics*, *44*(7–8), 2177–2193. https://doi.org/10.1007/s00382-014-2238-x

Field, P. R., M. J. Roberts, J. M. Wilkinson, 2018: Simulated lightning in a convection permitting global model. Journal of Geophysical Research - Atmospheres, doi: https://doi.org/10.1029/2018JD029295.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), Geosci. Model Dev., 12, 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, 2019.

Haarsma, R, et al. (2020) HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model performance, data handling and validation. Geoscientific Model Development, https://doi.org/10.5194/gmd-2019-350.

Hewitt, H. T., and coauthors (2020): Resolving and Parameterising the Ocean Mesoscale in Earth System Models. Current Climate Change Reports, submitted.

Jackson, L.C., M. J. Roberts, H. T. Hewitt, D. Iovino, T. Koenigk, V. L. Meccia, C. D. Roberts, Y. Ruprich-Robert, R. A. Wood, 2019: Does ocean resolution affect the rate of AMOC weakening? Clim. Dyn., submitted.

Juricke, S., Palmer, T. N., Zanna, L., Juricke, S., Palmer, T. N., & Zanna, L. (2017). Stochastic sub-grid scale ocean mixing: Impacts on low frequency variability. *Journal of Climate*, JCLI-D-16-0539.1. https://doi.org/10.1175/JCLI-D-16-0539.1

Juricke, S., & Jung, T. (2014). Influence of stochastic sea ice parametrization on climate and the role of atmosphere-sea ice-ocean interaction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2018). https://doi.org/10.1098/rsta.2013.0283

Juricke, S., Lemke, P., Timmermann, R., & Rackow, T. (2013). Effects of stochastic ice strength perturbation on arctic finite element sea ice modeling. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-12-00388.1

Macleod, D. A., Cloke, H. L., Pappenberger, F., & Weisheimer, A. (2016). Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 79–90. https://doi.org/10.1002/qj.2631

Moreton, S., D. Ferreira, M. Roberts, H. Hewitt, 2020: Evaluating surface eddy properties in climate simulations with 'eddy-present' and 'eddy-rich' ocean resolution. Ocean Modelling, 147. https://doi.org/10.1016/j.ocemod.2020.101567

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., … Weisheimer, A. (2009). Stochastic Parametrization and Model Uncertainty. *{ECMWF Technical Memorandum}*, *598*, available at http://www.ecmwf.int/publications/.

Renault L., Lemarié F., Arsouze T. (2019). On the implementation and consequences of the oceanic currents feedback in ocean–atmosphere coupled models. Ocean Modelling, Volume 141, 101423, ISSN 1463-5003, https://doi.org/10.1016/j.ocemod.2019.101423

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., Schiemann, R., Seddon, J., Vannière, B., and Vidale, P. L.: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments, Geosci. Model Dev., https://www.geosci-model-dev.net/12/4999/2019/, 2019.

Roberts, M.J. and 26 Coauthors, 2020: Sensitivity of the Atlantic Meridional Overturning Circulation to Model Resolution in CMIP6 HighResMIP Simulations and Implications for Future Changes. JAMES. https://doi.org/10.1029/2019MS002014

Strømmen, K., Christensen, H. M., Berner, J., & Palmer, T. N. (2018). The impact of stochastic parametrisations on the representation of the Asian summer monsoon. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3749-z

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

Strommen, K., Mavilia, I., Corti, S., Matsueda, M., Davini, P., von Hardenberg, J., … Mizuta, R. (2019b). The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL082843

Strommen, K., Watson, P. A. G., & Palmer, T. N. (2019c). The Impact of a Stochastic Parameterization Scheme on Climate Sensitivity in EC-Earth. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2019JD030732

Vidale, P.L., and coauthors, 2020: Impact of stochastic physics and model resolution on the simulation of Tropical Cyclones in climate GCMs. J. Clim., submitted.

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

4. Lessons Learnt

CMIP6 configuration climate models at standard resolution have been extensively tuned over years to reach good performance skills in terms of both the mean state of the climate as well as its variability. On the other hand, the configurations developed in the framework of PRIMAVERA WP4 at high or very high resolution (i.e. ~10km horizontal resolution) are new, with very limited knowledge of each component on climatic scale, and currently at the frontier of what can be done at present. Depending on the physical questions tackled, especially when focusing on large scale dynamics, there is therefore no obvious improvement in switching to very high resolution yet, but some processes do act rather differently compared to the current parameterisations. Finding strategies to tune these configurations with limited computational cost will be key to get their full potential and get the advantage of explicitly resolved oceanic eddies or atmospheric convection.

Running these configurations required some (sometimes extensive) adaptation of the tools currently used in each institute. Also, the huge extra cost in terms of technical issues (e.g. for pre-post treatment of files including cmorization, or regular crashes as more computational nodes are requested), computational resources (leading to submission of several projects to access supercomputers resources), storage (up to several hundreds of Tb can be needed for one simulation) or time to solution (e.g. VHR simulation at BSC has been running for 2 years continuously) associated with these configurations have proven to be quite frustrating. However, the pioneer work for this type of configuration will prove to be extremely valuable for coming projects and next generations of climate models.

As for the stochastic schemes considered here, they all lead to noteworthy improvements in various parts of the model. In particular, the fully stochastic configuration leads to broad improvements in the long-term climate mean and variability for key variables such as cloud cover, sea-surface temperature and precipitation: these bias reductions are between one third to half of what is seen when increasing resolution. Extremes over Europe are better represented, as are Euro-Atlantic circulation regimes. The number of tropical cyclones increases in a manner roughly equivalent to the change associated with increased resolution: in both cases this leads to a closer match with observations. While the stochastic ocean and sea-ice configuration leads to better Arctic sea-ice variability, including a more pronounced teleconnection to Europe, its impacts are to some extent overwhelmed when an atmospheric scheme is included. Therefore, depending on user priorities, a case can be made for either the fully stochastic configuration or the ocean and ice only configuration.

In all cases, the resources associated with a stochastic scheme are several orders of magnitude lower than the cost of increased resolution. As a result, while high resolution overall leads to greater improvements in the areas we considered, the benefits/cost ratio is likely to be very large. This suggests that adding a stochastic scheme to a climate model may be, in effect, a permanent and extremely cost-effective upgrade of the model.

5. Links Built

Overall, simulations at frontier resolution followed the CMIP6 protocol dedicated to high resolution simulations (i.e. HighResMIP) set in WP6, and all simulations produced in WP4 will be published on the ESGF node (link with WP9). Each institute produced a set of experiments with varying horizontal resolution either in the atmosphere or in the ocean, or both. The database that comes out of these simulations is an unprecedented tool to study the impact of numerical resolution (WP2). The eddy-resolving coupled GCMs complement Stream 1 simulations from WP6 that intend to parameterise eddies. This allows for comparative analysis that evaluates the role of eddies on oceanic and atmospheric processes, the climate response, consequences on European climate, etc., as evident in deliverables such as D2.5 and D4.5. Frontier simulations are also compared to WP3 simulations, specifically where different types of ocean mixing schemes are tested, and these analyses are reported in D3.4.

The MPI-ESM-ER (Max Planck Institute Earth System Model [eddy resolving]) configuration that was set up in PRIMAVERA has now been adopted for use in other projects such as FAFMIP and German BMBF-project HIPRED RACE. The demonstrator of EC-EarthP-VHR configuration used by BSC was developed in the ESIWACE project, and the production version developed and used in PRIMAVERA will be incorporated in EC-Earth workflow to be run more routinely in ESIWACE2 project.

As for the stochastic simulations, work in WP4 ended up building close links to WP2, particularly deliverables D2.2 and D2.5. This happened naturally, since we wished to not only analyse the impacts of stochastic physics in and of themselves, but also relative to impacts of increased resolution. We therefore ended up modifying our plans and analysis to be more closely aligned with analysis carried out by WP2 partners. Our work on the impact of stochasticity on the diurnal cycle of convection, and a comparison with increased resolution, ended up connecting with the DYAMOND Initiative. The analysis, documented in Deliverable D4.4, included a comparison with the ultra-high-resolution ECMWF-IFS model simulation, one of the models participating in DYAMOND.