Assessment of Climate Biases in OpenIFS Version 43R3 across Model Horizontal Resolutions and Time Steps

Abhishek Savita[1], Joakim Kjellsson[1,2], Robin Pilch Kedzierski[3,1], Mojib Latif[1,2], Tabea Rahm[1,2], Sebastian Wahl[1] and Wonsun Park[4,5]

Referee #1

Review of "Assessment of Climate Biases in OpenIFS Version 43R3 across Model Horizontal Resolutions and Time Steps"

This study evaluates the sensitivity of an atmospheric general circulation model (OpenIFS cycle 43R3) to different combinations of time-step and horizontal resolution. The authors evaluate several aspects of the mean climate and variability in simulations with prescribed sea surface temperatures (SSTs) and sea-ice over the the period 1979-2019. The authors identify several regions where reducing model time-step from 60 minutes to 15 minutes can have positive impacts on systematic biases that are comparable to the impact of increasing resolution. The manuscript is clear and concise and the topic is within the scope of GMD. The results of this study will likely be of interest to the many users of the IFS model in weather and climate sciences. In particular, these results raise interesting questions about model development strategy when it is often necessary to work with cheaper and/or reduced resolution surrogates of more expensive operational/production configurations. I believe this manuscript can be suitable for publication in GMD but I have several comments that I think would improve and clarify the current manuscript.

Thank you very much, and happy to hear that reviewer finds this manuscript interesting and relevant to IFS modelling community.

Major comments:

(1) In section 3.1 the authors focus on biases of near-surface fields, and how these are alleviated with reduced time-step and/or increased resolution.

I encourage the authors to extend their analysis to other levels in the atmosphere (e.g. zonal means of temperature/wind against model/pressure levels). Given the changes in convection and vertical mixing identified later in the paper, I think it is possible that the authors will find similar sensitivities in the troposphere. I also think it is possible that changes at other levels may result in increased rather than reduced biases. This is fine, as the most interesting aspect is the sensitivity to time-step and how this varies with region (e.g. is it limited to near-surface/troposphere). I think it is unlikely that reducing time-step will improve biases in all regions/levels, so it would also be interesting to discuss and interpret any regions of increased bias (e.g. whether they might indicate a role for compensating errors).

We have added an additional section addressing the zonal mean wind and temperature biases (lines 271-296):

"We examined the zonal mean u wind bias at different model levels, and it is shown in Figures 4. We find that zonal mean u wind bias over the tropical region (40ºS and 40ºN) is positive and independent of model horizontal resolution and model time step (Fig. 4b-f). The OIFS-HRA-15m configuration has a relatively large negative bias in the Northern Hemisphere compared to the other configurations. The OIFS-LRA-15m and OIFS-HRA-15m zonal mean u wind bias is similar to that in Robert et al. (2018). However, the zonal mean u wind bias in the Southern Hemisphere is not consistent across horizontal resolution or model time step. The zonal mean u wind bias in midlatitudes (i.e., 70ºS to 50ºS) is positive and large in the OIFS-LRA-1h configuration and reduces throughout the pressure levels by shortening the model time step in the coarse resolution OpenIFS configuration (i.e., OIFS-LRA-30m and OIFS-LRA-15m). Whereas, the negative zonal wind bias south 70ºS in the coarse resolution configuration is consistent across the different time steps (Fig. 3b-d). It is also interesting to note that both OIFS-MRA-15m and OIFS-HRA-15m configurations exhibit a negative bias over the Southern Ocean (SO) at most of the pressure levels, which is not seen in the standard OIFS-LRA-1h, nor in the OIFS-LRA-30m or OIFS-LRA-15m configurations. Overall, we conclude that by reducing the model time step in the coarse resolution configuration, we improve winds not only at the surface but also at higher model levels mostly over the SO. A similar conclusion does not hold for the OIFS-MRA-15m and OIFS-HRA-15m configurations, as both suffer from large negative bias over the SO.

We also examined the zonal mean temperature bias at different pressure levels. We find a cold bias (1.5 to 6 °C) in the troposphere and lower stratosphere, and a warm bias (1.5 to 6 °C) above the stratosphere across the configurations (Figure not shown). This indicates that OpenIFS simulations (independent of model time step and horizontal resolution) are colder than observations in the lower stratosphere and warmer above. The cold bias in the lower stratosphere is larger in the high resolution (i.e., OIFS-HRA-15m), and the warm bias above the stratosphere is smaller compared to the other configurations. Robert et al. (2018) noticed a similar zonal mean temperature bias and speculated that the zonal mean temperature bias is linked with the sensitivity of spurious mixing due to convection and diffusion."

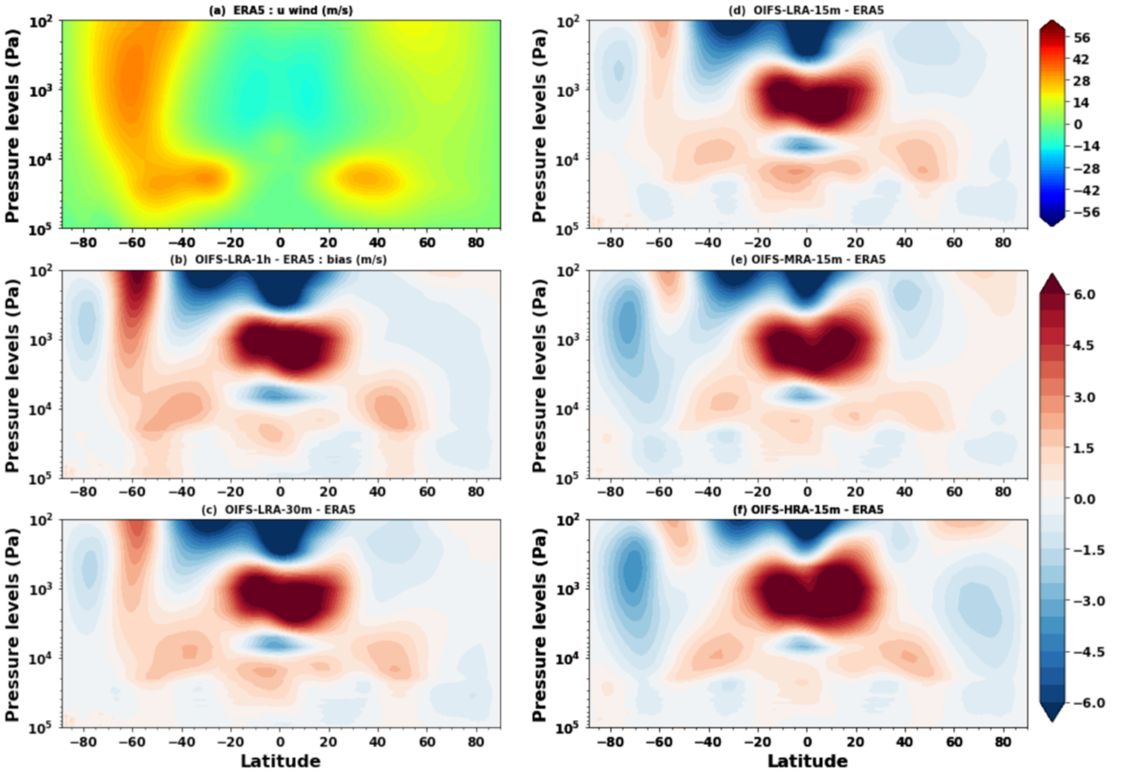


Figure 4. (a) Annual zonal mean ERA5 zonal wind [ms^{-1}]. (b-d) Annual zonal mean zonal wind [ms^{-1}] bias for different model time steps (1h (b), 30m (c), and 15m (d)) using ~100 km resolution, and (e-f) with different horizontal resolutions, ~50 (e) and ~25 km (f), respectively. Biases are computed with respect to ERA5 over the period 1979–2019.

(2) The abstract concludes with the general statement that "reducing the time step in the OpenIFS model, one can alleviate some climate biases at a lower cost than by increasing the horizontal resolution." I would like the authors to to add some discussion of whether they expect their results to generalise to resolutions and/or time-steps not tested in this manuscript. For example, how far is the LR configuration from converging? Would reducing time-step in a much higher resolution model (e.g. 9km) bring similar benefits? Depending on these additions, the authors may wish to qualify the concluding line of the abstract.
As suggested by the reviewer, we have modified the general conclusion to more specific as (lines 28-29): "Reducing the time step in the coarse resolution (~100 km) OpenIFS model, one can alleviate some climate biases at a lower cost than by increasing the horizontal resolution.

(3) What is the impact time-step/resolution on the representation of extremes? It is plausible that changes in time-step that improve the mean state have a limited impact on extremes that are more sensitive to horizontal resolution (e.g. orographic precipitation or tropical cyclones). As cited by the authors, the mean climate of the 25km and 50 km HighResMIP configurations of IFS are very similar. However, the differences in horizontal resolution are evident in the representation of extremes (examples below):
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019JD032184
Bador et al. (2020)
https://journals.ametsoc.org/view/journals/clim/33/7/jcli-d-19-0639.1.xml
Roberts J. et al (2020)
This is an excellent point. However, the main focus in this study is the mean state biases. The OpenIFS' model time-step sensitivity to extremes will be discussed in detail in a separate manuscript.

Minor comments:

Introduction: This section would benefit from an overview of the "expected" impacts of reducing time-step in simple models in terms of truncation error and how this might not always hold true in a more complex system. For example, in simple finite difference models, solutions converge as grid-spacing and time-step are decreased due to reduced truncation errors. The choice of time-step and grid-spacing may also be constrained by stability criteria. However, this intuition does not always hold in complex models due to the coupling between many different elements. For instance, it is plausible that the unconditionally stable semi-implicit semi-Lagrangian scheme used in the IFS allows a user to configure the model with a long time step to reduce the cost. Later developments on top of this configuration introduce compensating errors in other aspects of the physics that reduce biases. Reducing the time step at a later stage may then leads to increased biases as the model configuration has been implicitly tuned for a particular combination of time-step and resolution.

We added some text to the Introduction section (see below) (lines 75-82):

"While the semi-implicit semi-Lagrangian scheme, as used in OpenIFS, is unconditionally stable and the time step can be chosen to be very long, a shorter time step generally leads to a decrease in truncation error in the finite differences and thus a more accurate representation of the model dynamics. The physics parameterisations, which are computed independently of each other in OpenIFS, also benefit from a shorter time step as it will allow the various parameterisations to be coupled at a higher frequency (Beljaars et al. 2018). However, model parameters for e.g., convection or diffusion may be tuned for a specific time step and shortening the time step can therefore, in some cases, increase model error. Hence, a shorter model time step is expected to reduce biases in model dynamics, e.g., winds, while the results for parameterised processes, e.g., precipitation, may be mixed".

Lines 110-119. The authors state that their "detailed analysis of phase speed in such a manner is novel in literature". This may be the case, but I would like the authors to provide a more detailed summary of the method used to diagnose the amplitude and phase speed of extratropical Rossby waves (e.g. a bullet point list of the main processing steps). The current description is insufficient for reproduction of the analysis. In particular, it is not clear from the text how wave packets and associated phase speeds are diagnosed.

We revised the texts so that we can reproduce the analysis by following the steps (lines 132-137)

"We then applied the Fourier decomposition analysis to determine amplitude and position for each Rossby wave number at each latitude as a function of time. Phase speed is computed as the difference in the daily position of each wave, and stored at the midpoints in the time dimension. For consistency, wave amplitudes are interpolated to the midpoints in time as well. Lastly, seasonal averages are computed from the daily data for the boreal and austral winter seasons over the time period 1979–2019. In the case of phase speed, it is weighed by the corresponding daily (midpoint) amplitude squared when computing the seasonal averages in order to account for the impact of higher-amplitude events."

Line 169. Typo? Should be "Roberts et al. (2018)" as in intro?

This has been fixed.

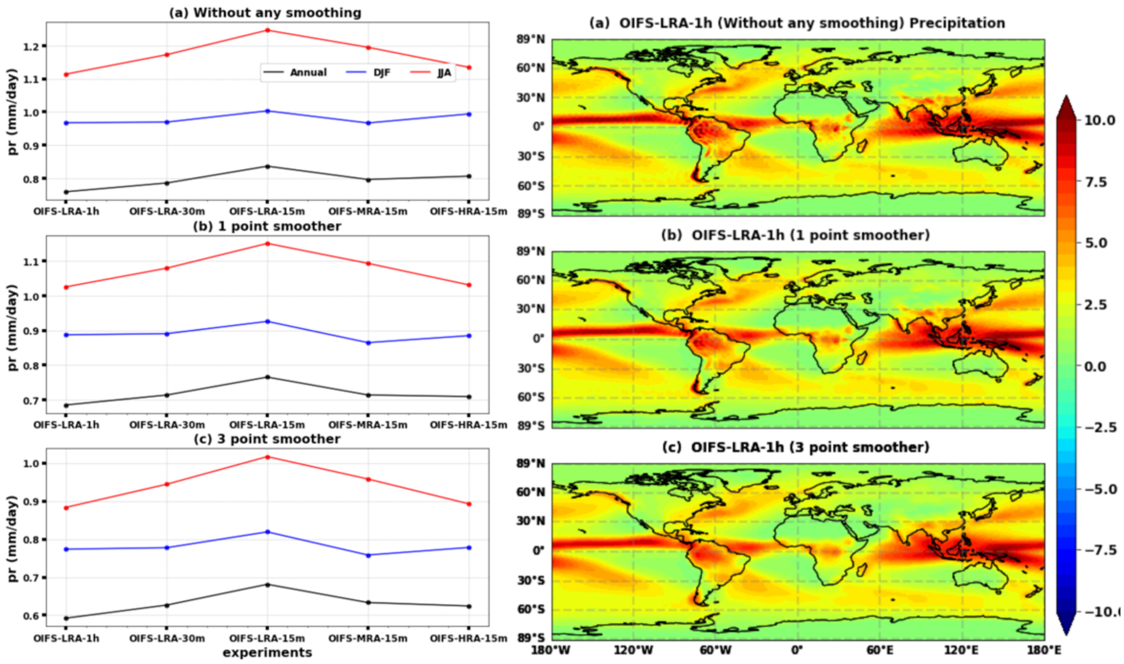
Line 198. Is it correct to include Coriolis? Work done by Coriolis term should be zero since it acts perpendicular to motion of air parcels.

Yes, the Coriolis term is included in the DYN part because we are analyzing zonal wind tendency. However, we have not quantified the individual contribution of the Coriolis term to the DYN term.

Lines 226-228: It is possible that the lower precipitation RMSE in OIFS-LRA-1h is due to a "double penalty" effect that penalises higher resolution models, which have more structure in the precipitation fields. Is the precipitation in the LR-1h experiment notably smoother? Other metrics (e.g. fractions skill score) may provide a different ranking of models. More details on double-penalty effects and fraction skill score here: https://www.ecmwf.int/en/about/media-centre/science-blog/2023/verifying-high-resolution-forecasts

To verify our result, we computed the RMSE by smoothing the data and found a similar conclusion. However, the RMSE values differ in magnitude if we compare them with without smoothing. We have not computed the fraction skill score as our conclusions are insensitive to the double penalty effect (see the figure below; (left) RMSE and (right) time mean Precipitation). We now added this information in the main text as well (lines 259-263).

"We have computed the SAT and precipitation biases with a 3-point smoothing, i.e., approximately 3x3 degree spatial smoothing, which eliminates the wiggles near steep topography arising from the Gibbs' phenomenon in the model spectral fields. We find that smoothing the fields does not change the main result that precipitation biases increase with shorter time step in Tco95 and then decreases somewhat with higher horizontal resolution. Hence, the wiggles are not the main source of precipitation biases and their presence does not impact the findings in this study.”



Lines 237-238 and figure 4: Why do the authors standardise the wave amplitude biases in figure 4 instead of showing the absolute values? This standardisation emphasises errors in regions the authors argue are unimportant, which complicates interpretation of the plots. Specifically, the authors focus their analysis of Rossby waves on the "region where the wave amplitude is larger than 5 ms-1 is termed core region, which mostly covers the area that is occupied by the thick black contours in Fig. 4". However, biases are presented "relative to ERA5 (model – ERA5), normalized by the ERA5 detrended variability expressed by the standard deviation", which highlights errors in the high-latitude high-wavenumber waves that are dismissed by the authors as "unimportant as these waves have a small amplitude and little effect on variability".
We updated the figures with absolute error and modified the texts accordingly (pages 9-10 and also provided texts below).

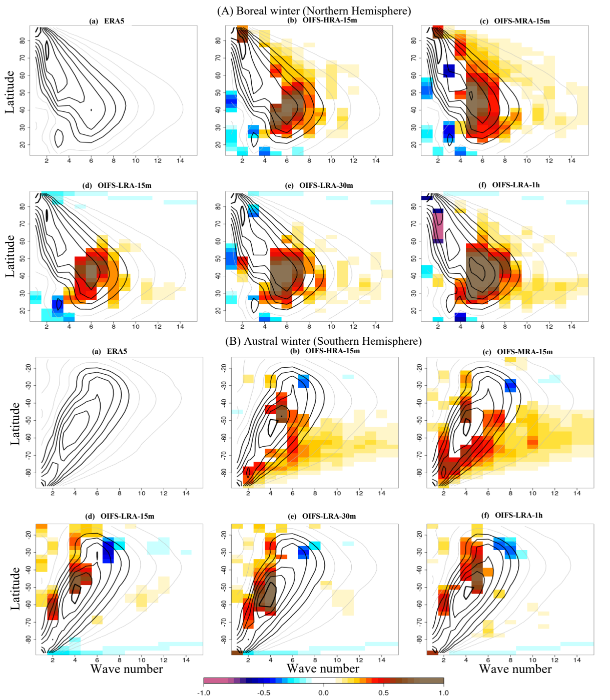


Figure 5. (A) The Rossby wave amplitude (contours) for different wave numbers in the Northern Hemisphere at 300 hPa (a) in ERA5 observation and (b-f) in the OIFS model simulations during 1979-2019 in DJF (i.e., boreal winter). The color shows difference of wave amplitude between the model and ERA5 where it is significant on the 95 % confidence level. The wave amplitude and contour interval are shown in ms^{-1}. The grey contours start from 2 ms^{-1} and the black contours from 5 ms^{-1} and the contour interval is 1 ms^{-1}. (B) is similar to (A), but for JJA (i.e., austral winter).

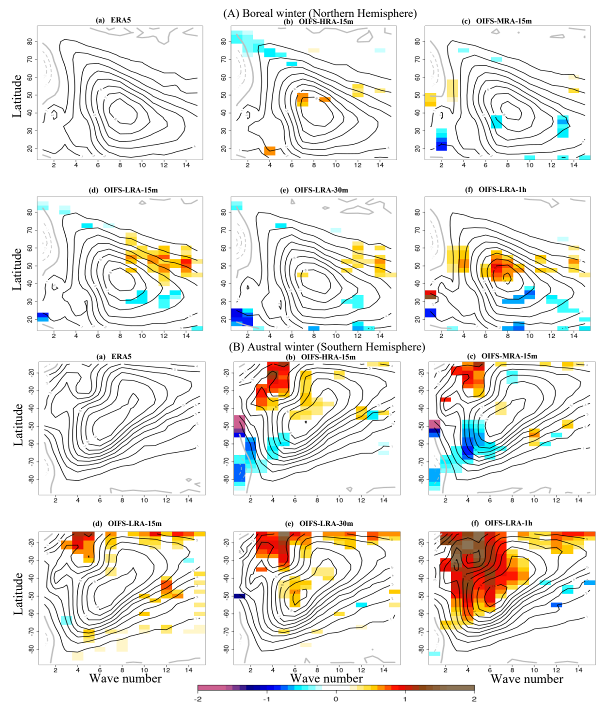


Figure 6. (A) The Rossby wave phase speed (contours) for different wave numbers at 300 hPa in the Northern Hemisphere in ERA5 (a) observation and (b-f) in the OIFS model simulations during 1979-2019 in DJF (i.e., boreal winter). The color shows the difference of wave phase speed between model and ERA5 where it is significant on the 95 % confidence level. The wave phase speed and contour interval are shown in ms^{-1}. The black

contours start from 1 $ms^{-1}$and the contour interval is 1 ms^{-1}. Panel (B) is similar to panel (A), but for JJA (i.e., austral winter). The dahs contours show a negative phase speed and a grey contour shows a zero-phase speed.

“**Fig. 5** shows the Rossby wave amplitude (gray and black contours) for ERA5 and the individual OIFS simulations for the boreal winter (**Fig. 5A, DJF, Northern Hemisphere; NH)**) and austral winter (**Fig. 5B, JJA, Southern Hemisphere; SH**). The color in **Fig. 5** denotes the wave amplitude bias relative to ERA5 (model – ERA5). We focus only on those wave numbers and latitudes that have the highest wave amplitude, because these waves explain most of the variability. The region where the wave amplitude is larger than 5 ms^{-1} is termed “core region”, which mostly covers the area that is occupied by the thick black contours in **Fig. 5**. In DJF (NH), at north of 70° N, the Rossby wave numbers k=1 and k=2 have the largest amplitude in ERA5 whereas at the mid-latitudes (30° N to 60° N), the wave numbers between about k=3 and k=9 have large amplitude with the largest amplitude amounting to 8 ms^{-1} at about 40° N for the wave number k=6 (**Fig. 5Aa**). During JJA (SH), the wave amplitude is located in a similar core region (**Fig. 5Ba)** as that in DJF (NH). The amplitude is largest south of 70° S for the wave numbers k=1 and k=2 whereas at the mid-latitudes (45° S to 65° S), the wave numbers between about k=3 to 5 have large amplitude with the largest amplitude amounting to 9 ms^{-1} is found at 57.5° S for the wave number k=4 (**Fig. 5Ba**).

In DJF (NH) the OIFS-LRA-1h configuration exhibits a positive bias of ~1 ms^{-1} in Rossby wave amplitude (i.e., the waves amplitude bias in OIFS-LRA-1h is larger than the ERA-5) in the core region, in particular for wave numbers k=3-8 at latitudes between 25° N to 55° N and a negative bias at latitudes between 60° N to 80° N for waver number 2 (**Fig. 5Af**). The wave amplitude biases around the core region in OIFS-LRA-1h in the midlatitudes (20° N to 40° N) are small (~0.2) for the higher wave numbers and get better with a shorter time step configuration (OIFS-LRA-15m).

The Rossby wave amplitude biases in the OIFS-HRA-15m configuration are strongly reduced compared to the OIFS-LRA-1h configuration over the core region (**Fig. 5Ab and 5Af**). The Rossby wave amplitude bias reduction in the OIFS-MRA-15m configuration is mostly similar to that in the OIFS-HRA-15m configuration except for the wave number k=7 at 45° N, where the wave amplitude bias is larger in the OIFS-HRA-15m configuration (**Fig. 5Ab and 5Ac**). The OIFS-HRA-15 m and OIFS-MRA-15m configurations also exhibit a positive bias for wave number 2 at high-latitudes 60° N to 80° N. The OIFS-MRA-15m configuration also show a negative bias for the wave number 3 at latitudes between 60° N to 65° N in the core region, which is not present in the other configurations. The OIFS-HRA-15m and OIFS-MRA-15m configurations show similar bias around the core region as in the OIFS-LRA-1h configuration, i.e., high resolution and OIFS-LRA-1h configurations overestimate wave amplitudes for the higher wave numbers. The Rossby wave amplitude biases are progressively reduced from the OIFS-LRA-1h configuration to the OIFS-LRA-30m and OIFS-LRA-15m configurations (**Fig. 5Ad-Af**), indicating a sensitivity of model bias to the time step. The wave amplitude bias for wave number k=7 at 45° N exists in all the configurations, and it is smaller in the OIFS-LRA-15m and OIFS-MRA-15m configuration than in the other configurations. Overall, both OIFS-LRA-15m and OIFS-HRA-15m configurations are able to reproduce the observed Rossby-wave amplitudes in DJF (NH) better than OIFS-LRA-1h.

In JJA (SH), the Rossby wave amplitude bias in the core region is smaller than in DJF (NH) for all the configurations (**Fig. 5A and 5B**). OIFS-LRA-1h exhibits a positive bias of ~0.5 ms^{-1} in JJA (SH) for the wave number k=2 at latitude between ~50° S and ~62.5° S and for wave numbers k=4 to 5 between 30° S and 40° S (**Fig. 5Bf**). The OIFS-LRA-30m configuration shows a positive bias for the wave number k=2 to 5 at latitudes between 40° S and 70° S, which is larger than other configurations.
The OIFS-HRA-15m and OIFS-MRA-15m configurations exhibits a positive bias ~0.5 ms^{-1} around the core region and latitude 50° S to 70°S, which does not exist in the other coarse resolution configurations (**Fig. 5Bb-Bf).** The Rossby wave amplitude biases around the core region at the midlatitudes in the high-resolution simulations are consistent and large in the SH than the NH (Fig. 5Ab-c and 5Bb-c).

We also analyze the phase speed of Rossby waves for ERA5 and across the OIFS’ configurations for DJF (NH) and JJA (SH) seasons (**Fig. 6**). In the ERA5 dataset (**Fig. 6Aa**), the Rossby wave phase speed is positive (i.e., eastward moving, solid contour) for wave numbers greater than 2 (i.e., k>2) at most latitudes. The wave numbers k=1 to 2 have a positive wave phase speed from the equator to 55° N and a negative wave phase speed (i.e., westward moving, dashed contours) between 60° N and 80° N in DJF (NH) (**Fig. 6Aa**). The maximum phase speed is found at wave number k=8 at 40° N, while the minimum is found at wave number k=1 at 60° N (**Fig. 6Aa**). In JJA (SH) (**Fig. 6Ba**), the wave phase speeds are mostly positive and large for all the wave numbers and at each latitude, with the maximum phase speed is observed for the wave numbers between k=6 and k=8 and latitudes between 40° S and 60° S, and these waves are moving faster than that in DJF (NH).

The OIFS-LRA-1h configuration suffers from positive phase speed bias for wave numbers k=4 to 8 at latitudes between 42.5° N and 60° N, i.e., waves move faster eastward than in ERA5, and the bias is larger than 1 ms^{-1}. The bias of ~1 ms^{-1} for wave number k = 6 to 8 at 40° N and 60° N is of particular concern as it is near the maximum wave amplitudes in DJF (**Fig. 6Af)**. In general, phase speed biases in the OIFS-LRA-1h configuration are strongly reduced as either horizontal resolution is increased or time step is shortened (**Fig. 6Ab-5Af**). In JJA (SH), the OIFS-LRA-1h configuration exhibits a very large (between ~1.5-2 ms^{-1}) Rossby wave phase speed bias for most of the wave numbers, which is largest for the wave numbers k=2 to 8 between 15° S to 55° S (**Fig. 6Bf**). Large biases can be found between 15° S and 25° S (~1.5 ms^{-1}) for most of the wave numbers, but the wave activity is low there (**Fig. 6Bf**). The large phase speed biases are strongly reduced in the OIFS-LRA-30m and OIFS-LRA-15m configurations (**Fig. 6Bd-Bf**), indicating a strong sensitivity to the reduced biases in mean winds and wind speeds (**Fig. 1**). Overall, the Rossby wave speed bias in the OIFS-HRA-15m configuration is smaller than in the OIFS-LRA-1h configuration (**Fig. 6Bb and 6Bf**). However, we note that both the OIFS-MRA-15m and OIFS-HRA-15m configurations exhibit negative biases south of 55° S for wave numbers k= 1 to 5, that is, the eastward moving waves are slower than in the ERA5 (**Fig. 6Bb**).

The wave phase speed analysis reveals a clear improvement in the representation of the Rossby waves in the boreal winter (i.e., NH) when increasing the horizontal resolution and shortening the model time step compared to OIFS-LRA-1h configuration. In austral winter, however, the representation of Rossby wave amplitudes and phase speeds are the most realistic in OIFS-LRA-15m configuration, with longer time steps introducing too fast phase speeds and higher horizontal resolution introducing too slow phase speeds at wave number less than 6 (i.e., k<6)."

Section 3.2. How do the authors interpret the impact on Rossby wave amplitude/phase speed biases? For example, is it related to the representation of tropospheric jets and associated wave guides and their biases?
The representation of tropospheric jets and associated wave guides can be related to biases in Rossby wave packets (e.g., Giannakaki and Martius, (2016), Hakim, 2005 and Baumgart et al., 2018). We evaluate the amplitude and speed of each wavenumber individually that indicating at which scale the model biases occur. RWPs are then the combination (or sum) of the intermediate ones (e.g. wavenumbers 4-15), but diagnosing/tracking RWPs is a different analysis, and there's no consensus on the best method (Wolf and Wirth, 2017).

Section 3.3 and figure 7. What is the sampling uncertainty in these composites and estimates of pattern correlation (e.g. estimated using bootstrap resampling of available dates)? Are the differences between configurations significant?
We added sampling uncertainty using bootstrapping method with random 2000 iterations.

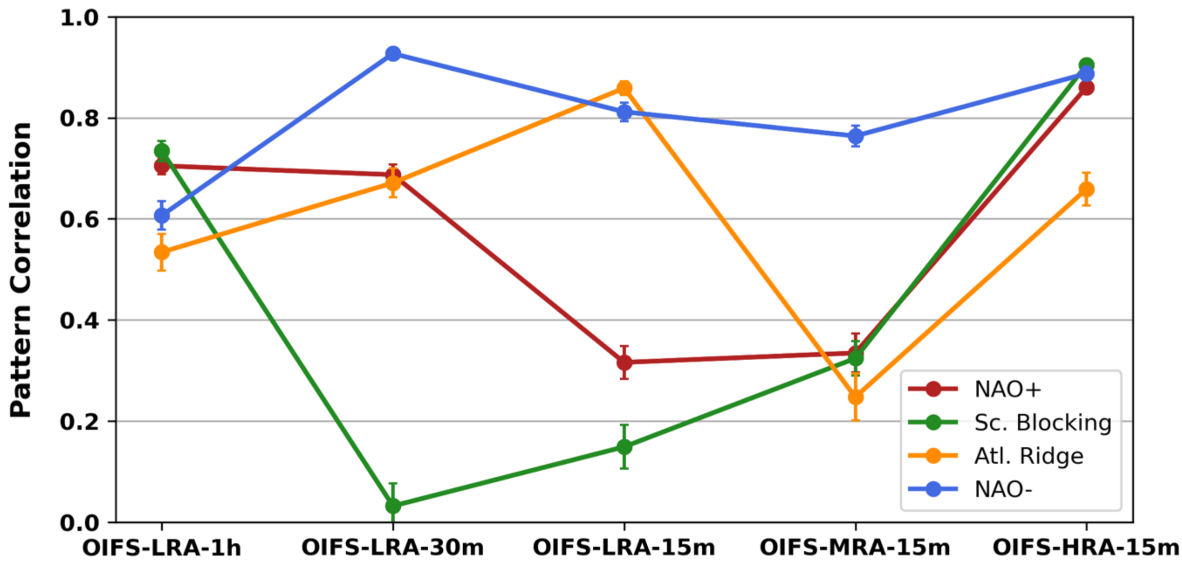


Figure. 8. Pattern correlation coefficient of the individual weather regime between OIFS model configurations and ERA5 for the period 1979-2019 for the DJF season. The error bars represent a 95% confidence interval.

Table 1. What is the HPC cost of the different configurations (e.g. core hours per model year)? Do they scale as expected from changes in time-step and number of grid points?
As the reviewer suggested, we added this information to Table 1 (lines 694-698).

Figure 2. What is the sampling uncertainty in these estimates of RMSE? Are the differences in RMSE between configurations significant?
We added the figure depicting RMSE with sampling uncertainty using bootstrapping method with random 2000 iterations (see below and page 22).

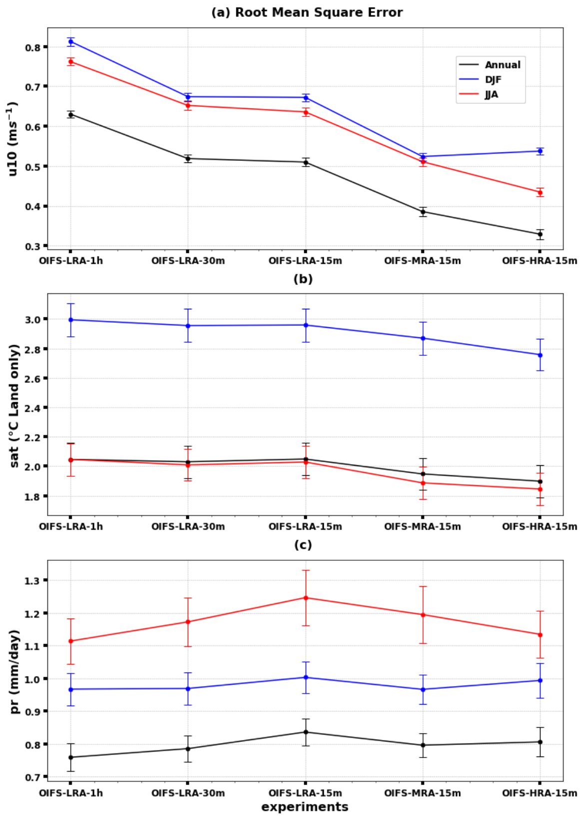


Figure 2. Root mean squared error of surface zonal wind (a), SAT (b), and precipitation (c) over the period 1979-2019 for all the configurations: annual (black) and seasonal mean (DJF: blue, JJA: red). The error bars represent a 95% confidence interval.