Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

This is a PDF of a manuscript that has been peer-reviewed and accepted for publication. As the article has not yet been formatted, copy edited or proofread, the final published version may be different from the early online release.

This pre-publication manuscript may be downloaded, distributed and used under the provisions of the Creative Commons Attribution 4.0 International (CC BY 4.0) license. It may be cited using the DOI below.

The DOI for this manuscript is
DOI:10.2151/jmsj.2020-016
J-STAGE Advance published date: March 17th, 2020
The final manuscript after publication will replace the preliminary version at the above DOI once it is available.

Global simulations of the atmosphere at 1.45 km grid-spacing with the Integrated Forecasting System

Peter D. Dueben

European Centre for Medium-Range Weather Forecasts, Reading, UK

Nils Wedi

European Centre for Medium-Range Weather Forecasts, Reading, UK

Sami Saarinen

European Centre for Medium-Range Weather Forecasts, Reading, UK

Christian Zeman

Eidgenössische Technische Hochschule Zürich

February 19, 2020

Corresponding author: Peter D. Dueben, European Centre for Medium-Range Weather Forecasts, Shinfield Rd, Reading RG2 9AX, UK.
E-mail: peter.dueben@ecmwf.int

Abstract

Global simulations with 1.45 km grid-spacing are presented that were performed with the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). Simulations are uncoupled (without ocean, sea-ice or wave model), using 62 or 137 vertical levels and the full complexity of weather forecast simulations including recent date initial conditions, real-world topography, and state-of-the-art physical parametrizations and diabatic forcing including shallow convection, turbulent diffusion, radiation and five categories for the water substance (vapour, liquid, ice, rain, snow). Simulations are evaluated with regard to computational efficiency and model fidelity. Scaling results are presented that were performed on the fastest supercomputer in Europe - Piz Daint (Top 500, Nov 2018). Important choices for the model configuration at this unprecedented resolution for the IFS are discussed such as the use of hydrostatic and non-hydrostatic equations or the time resolution of physical phenomena which is defined by the length of the time step.

Our simulations indicate that the IFS model — based on spectral transforms with a semi-implicit, semi-Lagrangian time-stepping scheme in contrast to more local discretisation techniques — can provide a meaningful baseline reference for O(1) km global simulations.

Keywords global cloud-resolving modelling; global storm-resolving modelling; hydrostatic equations; high-performance computing; scalability

1. Introduction

The complexity and quality of weather and climate models has improved at a remarkable speed during the last decades (Bauer et al. (2015)) and the steady increase in computing power has allowed for a steady increase in model resolution and complexity of forecast models. However, the recent slow-down of the increase in performance of individual processors is now generating challenges for the domain of weather and climate modelling. It is getting more complicated to make efficient use of modern and future supercomputers that require applications to use massive parallelism of up to $O(10^6)$ processing units and heterogeneous hardware including Central Processing Units (CPUs), Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), Field-Programmable Gate Arrays (FPGAs) and more. This is difficult for weather and climate models that are comprised of O(1 million) lines of model code, require diverse mathematical algorithms within a single modelling framework, and are often written in different styles of coding for the different model components.

As the model resolution of global atmospheric simulations is always insufficient to resolve all features of the Earth System explicitly, several sub-grid features need to be parametrised. This involves a description of the statistical contributions of sub-grid scale processes on the mean flow, expressed in terms of the mean flow parameters. This closure thus relies on the averaged equations and explicit expressions for the higher-order terms arising from the perturbations of the mean flow. In addition, parametrisations describe diabatic effects such as radiation and water phase changes as well as processes for which equations that describe the underlying physical behaviour are unknown, such as soil processes. As resolution is increased, features of the Earth System such as deep convection can be represented explicitly on the computational grid of the model simulation.

Deep convection plays a fundamental role for the vertical transport of energy in the tropics which is driving the global circulation of the atmosphere through well known circulation patterns such as the Hadley Cell and the meandering inter-tropical convergence zone (ITCZ). Today, only few global weather and climate models run routinely at a grid-spacing of less than 10 km. Several of these models have contributed to the *DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains* (DYAMOND) model inter-comparison for which 9 models performed 40 day global simulations at a grid-spacing finer than 5 km (Stevens et al. (2019)). However, as global weather and climate models are approaching grid-spacings of a few kilometres, they are entering the so-called "grey-zone" of convection, where certain limiting assumptions underlying deep convection parametrisation – that deep convection can be represented as a bulk parametrisation scheme based on an ensemble of independent convective updrafts within a grid-cell – cannot be justified any more – if grid-cells partially or fully represent a single updraft. If the deep convection parametrisation scheme is switched off, convection is explicitly simulated by the governing equations. However, convective cells will be significantly bigger when compared to convective cells in the real-world if the resolution of the model is insufficient. As a result, explicitly simulated convective cells assume the size of one or multiples of the chosen grid-size, and unrealistically sized convective cells may cause a degradation of forecast skill in comparison to coarser simulations, where convection parametrisation is used. Global simulations with deep convection parametrisation switched off are often called "cloud-resolving". We will, however, refer to our simulations as "storm-resolving" following the convention of the DYAMOND project (Stevens et al. (2019)) since a grid-spacing of 1.45 km will still not be sufficient to resolve individual clouds.

It has been suggested to move from O(10 km) grid-spacing to a grid-spacing of O(1km) that would potentially allow to resolve deep convection sufficiently for global weather and climate models and to skip the resolution range in between. There is substantial experience in the limited area community in Europe (Termonia et al (2018)), where forecast models operate routinely at intermediate grid-spacings of 5 km or 2.5 km. As discussed in Neumann et al. (2018), a grid-spacing of O(1 km) for global atmosphere models would potentially show a number of improvements, including the representation of topographic gravity waves and surface drag that are induced by explicitly represented small-scale topography, and the ability to assimilate satellite data at its native resolution. Ocean models at O(1 km) grid-spacing can resolve a larger fraction of meso-scale eddies that are essential to represent ocean variability accurately, and the ability to explicitly simulate ocean tides.

Till today there are only a small number of simulations of the (near-)global atmosphere with a grid-spacing close to or beyond 1 km. These include a seminal 12 hour long simulation at 870 m grid-spacing with the Nonhydrostatic ICosahedral Atmospheric Model (NICAM) model (Miyamoto et al. (2013)). There was also a simulation of the atmosphere between the latitudes -80° and 80° for 10 days with the *Consortium for Small-scale Modeling* (COSMO) model at a grid-spacing of 930 m that was performed for the idealised test case of a baroclinic instability (Fuhrer et al. (2018)). Model simulations at slightly lower resolution have been presented in various papers (Miura et al. (2007), Satoh et al. (2008), Fudeyasu et al. (2008), Skamarock et al. (2014), Michalakes et al. (2015), and Müller et al. (2015)).

An overview on the history of global storm-resolving models can be found in Satoh et al. (2019).

Many simulations have also been performed at high resolution but for limited domains. Bretherton and Khairoutdinov (2015) simulated a $20,480 \times 10,240$ km equatorial channel for 30 days at 4 km grid-spacing and Leutwyler et al. (2017) show 3-month-long simulation with 2.2 km grid-spacing on a European-scale computational domain using the COSMO model. Yang et al. (2016) performed simulations with a moist baroclinic instability test in a β-plane three-dimensional channel resembling the latitude range between 18 and 72 degree north with a horizontal grid-spacing of 488 m. Heinze et al. (2017) present large eddy simulations with ICON over almost the entire area of Germany with 156 m grid-spacing for weather-type timescales. It is also worth mentioning the radiative-convective equilibrium model inter-comparison project (RCEMIP) that is comparing global model simulations for an idealization of the climate system to understand more about clouds, convection and climate sensitivity, and to quantify differences between models (Wing et al. (2018)).

A number of global weather and climate models are using the hydrostatic approximation within the so called set of primitive (shallow atmosphere) equations for operational forecasts. This approach assumes vertical accelerations to be small compared to the balancing forces of gravity and the vertical pressure gradient. This is typically valid when the ratio of vertical to horizontal length scales of motion is small. As a result, vertical velocity becomes a diagnostic variable that can be derived from the continuity equation, and for energetic consistency additional acceleration terms in the horizontal momentum equations on the sphere are dropped. More recent work allows to relax this traditional approximation despite continuing to use the hydrostatic assumption (Tort and Dubos (2014)).

When the aspect ratio of vertical to horizontal motions becomes approximately one (Jeevanjee (2017)), the hydrostatic approximation will become invalid. However, the precise grid-spacing when this is happening seems to depend on the particular model, the model configuration, and the significance for the features of interest. Daley (1988) suggests that for global models with a spectral truncation numbers greater than 400 ($>$ 25 km grid-spacing) the non-hydrostatic set of equations should be used. However, Ross and Orlanksi (1978) found only little difference between hydrostatic and non-hydrostatic two-dimensional simulations of an idealized cold front at a resolution of 20 km. For a similar case, Orlanski (1981) found significant differences at a resolution of 8 km. Dudhia (1993) simulated a cold front with a hydrostatic and non-hydrostatic model configuration and both versions produced similar results for grid-spacings of 6.67 km. Kato (1997) found that a hydrostatic model with idealized moist convection overestimated precipitation at 5 km grid-spacing. Weisman et al. (1997) performed simulations of a squall line with different grid-spacings reaching from 20 km to 1 km and found that the hydrostatic model overestimated the maximum vertical velocity at grid-spacings of 4 km and lower. Jeevanjee (2017) ran idealized radiative-convective-equilibrium simulations over sea for grid-spacings ranging from 16 km to 0.0625 km and found that the hydrostatic model started to overestimate the vertical velocities for grid-spacings smaller than 2 km. Often quoted are also situations with vertical wind shear, where vertically propagating gravity waves are trapped in the lee of the mountain and energy propagates horizontally rather than vertically (Keller (1994)). Hydrostatic models do not "see" the shear, and gravity waves propagate vertically upwards. However, as shown in Wedi and Smolarkiewicz (2009), if the mountain is not resolved with a sufficient number of grid points relative to the mountain width (and for the given flow regime), also non-hydrostatic models will show the characteristic hydrostatic (non-trapped) behaviour.

There is no consensus in the literature which spatial discretisation scheme would be most appropriate for global storm-resolving simulations. Indeed, all of the common approaches for the development of dynamical cores, including finite difference, finite volume, finite element or spectral methods, seem to be capable of running global model simulations at a grid-spacing of only a few kilometres on state-of-the-art supercomputers (Michalakes et al.

(2015) and Stevens et al. (2019)). There is also no consensus whether explicit, semi-implicit or fully implicit time stepping schemes are most promising for use in global storm-resolving simulations. Implicit schemes allow to use a larger time-step in comparison to explicit schemes since they are not bound by Courant-Friedrichs-Lewy-type constraints for fast wave motions. However, Fuhrer et al. (2018) argue that even fully implicit, global, convection-resolving climate simulations at $1-2$ km grid spacing cannot be considered a viable option when using a time step larger than $40-60$ s since sound wave propagation and important diabatic processes are not resolved in time, potentially leading to a change in the history of the flow evolution. Instead, they use a split-explicit time stepping scheme with a time step of 6 s when running with a grid-spacing of less than a kilometre. In contrast, Yang et al. (2016) propose to work with implicit schemes and show results using a large time step of 240 s when running at a grid-spacing of less than a kilometre to achieve the best time-to-solution for simulations. As pointed out in Wedi et al. (2015) and evident in Mengaldo et al. (2018), different time-stepping approaches may incur low-order time truncation errors compared to the nominal spatial truncation error of a given model, especially if time and space are handled independently. Thus a careful analysis of time truncation error at 1 km global grid-spacing is pending.

This paper presents global simulations of the atmosphere with the *In-*

tegrated Forecasting System (IFS) with up to 1.45 km grid-spacing. The performance of the IFS is discussed and scalability tests on the Piz Daint supercomputer and the two supercomputers of the *European Centre for Medium-Range Weather Forecasts* (ECMWF) are presented. These scaling results provide a good benchmark for the improvements in efficiency that would be required to allow for global, operational weather forecasts or climate projections at storm-resolving resolution. A first scientific evaluation of the IFS model fidelity for simulations at storm-resolving 1.45 km grid-spacing is presented. This includes a discussion of the effective resolution of atmospheric dynamics from energy spectra (Abdalla et al. (2013)) and a limited assessment how choices for the model configuration, for example regarding the use of non-hydrostatic equations, the parametrisation of convection or the time step length influence model simulations.

Section 2 provides details of the model configuration that was used. Section 3 will discuss the performance and scaling behaviour of the model. Section 4 will present the scientific evaluation of model runs. Section 5 and 6 will provide a discussion and conclusions.

2. A description of the IFS

We perform model simulations with the un-coupled IFS atmosphere model cycle 45r2 (no ocean, sea-ice or wave model, since these currently limit the scalability of the coupled system at 1.45 km grid-spacing). The IFS is a spectral transform model where prognostic variables have a dual representation in grid-point space and global spectral space represented via spherical harmonic basis functions. The latter facilitates easy computations of horizontal gradients and the Laplacian operator relevant for horizontal wave propagation. The special property of the horizontal Laplacian operator in spectral space on the sphere conveniently transforms the three-dimensional Helmholtz problem, arising from the semi-implicit discretisation, into an array (for each zonal wavenumber) of two-dimensional matrix operator inversions. Importantly, products of terms, (semi-Lagrangian) advection and all (columnar) physical parametrizations are computed in grid-point space. Water substances have only a representation in grid-point space. A cubic octahedral (reduced) Gaussian grid is used for this purpose (Wedi (2014) and Malardel et al. (2016)).

To transform between grid-point and spectral space requires the subsequent use of a Legendre Transformation and a Fast Fourier Transformation (called "transforms" in the rest of the paper). To improve performance for the calculation of the Legendre transformation, that shows a computational complexity proportional to N^3 with the truncation wave number N, a so-called "Fast Legendre Transformation" was introduced that is trading performance against accuracy and achieving a scaling behaviour of $N^2 log^3(N)$

(Wedi et al. (2013)). To avoid the transforms for high-resolution simulations in the future, ECMWF is also developing an alternative dynamical core based on a finite volume discretisation with the same collocation of prognostic variables as in the current IFS (*Integrated Forecasting System-Finite Volume Model* (IFS-FVM); Küehnlein et al. (2019)). However, in this paper we use IFS to refer to the spectral transform model.

The IFS is based on a semi-implicit semi-Lagrangian time-stepping scheme with no decentering that allows for the use of long time steps. We are using the same time-step for both dynamics and all physics at a grid-spacing of 1.45 km. There are two exceptions with turbulent vertical diffusion using two sub-steps and an hourly call frequency for radiative transfer calculations. Model simulations are initialised from the 9 km operational analysis of ECMWF at 13th October 2016 0h UTC, suitably interpolated using the integrated interpolation and post-processing software of Arpege/IFS ("https://www.umr-cnrm.fr/gmapdoc") to the target grid that is used for storm-resolving simulations. Next to the transforms, the calculation of the physical parametrisation schemes ("physics") and the semi-Lagrangian advection scheme are the largest contributors to computational cost of simulations that are both calculated in grid-point space. Only a comparably small fraction of the cost is generated by calculations in spectral space, mostly related to the semi-implicit timestepping scheme.

Most of the model simulations were performed with the single precision version of the IFS using 32 bits to represent real numbers. This version is using single precision for almost the entire model integration (Dueben and Palmer (2014) and Vana et al. (2017)). The quality of forecast simulations is equivalent between double and single precision simulations. However, the use of single precision is causing a small error in mass conservation and a global mass fixer is used in these simulations. The global mass fixer is cheap and easy to apply within a spectral model. The use of single precision is reducing runtime by approximately 40% (dependent on the Message Passing Interface (MPI) / Open Multi-Processing (OpenMP) configuration of the runs) and memory requirements are reduced significantly which makes it possible to run simulations also on a much smaller number of nodes for testing.

We perform both hydrostatic and non-hydrostatic simulations. Non-hydrostatic simulations are formulating the non-hydrostatic system in a mass-based vertical coordinate and adding prognostic variables for the vertical velocity and a deviation from the hydrostatic pressure. The resulting semi-implicit system is more complicated when compared to hydrostatic simulations but similarly solved in spectral space (see for example Voitus et al. (2019) and references therein).

We will compare model simulations that are using a different number of iterations for the optional predictor-corrector (PC) time-stepping scheme required for stability in the non-hydrostatic model. The advection (horizontal and vertical) and the entire spectral semi-implicit solve, including the spectral transforms of several prognostic variables, are required in each iteration, which is causing a significant increase of the computational cost of the non-hydrostatic model (see Wedi et al. (2009), (2013) for details). A second difference is the use of a finite element discretisation scheme in the vertical direction for standard hydrostatic simulations and a finite difference scheme that is currently used when running in non-hydrostatic mode. As detailed in Bubnova et al. (1995) the vertical discretisation has to be bespoke to ensure that the discrete and continuous system of equations are consistent. Improved consistency together with better treatment of vertical boundary conditions for the non-hydrostatic configuration may be achieved through changes to the equations and corresponding changes to the solution algorithm as detailed in Voitus et al. (2019). A vertical finite element scheme for the non-hydrostatic equations is also under active development (see also https://www.ecmwf.int/en/newsletter/161/news/ecmwf-tests-new-numerical-scheme-vertical-grid) but neither of these developments are available for experimentation at a grid-spacing of 1.45 km.

Most of the model simulations that are presented in this paper are based on the cubic octahedral grid with an average 1.45 km (TCo7999) grid-spacing (1.25 km near the equator). Operational weather forecasts at ECMWF use a cubic octahedral Gaussian grid with 9 km grid-spacing (TCo1279) for deterministic forecasts and 18 km grid-spacing (TCo639) for ensemble predictions with 50 ensemble members.

We use the standard procedures for ECMWF to generate high-resolution topography fields (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation). The 30" orography data derived from different sources is spectrally fitted to T15999, slightly filtered in spectral space, and truncated to the 7999 truncation. The resulting field is used as input to the IFS simulations. There is no representation of sub-grid scale topography within the high-resolution simulations at 1.45 km grid-spacing.

We use the standard set of physical parametrisation schemes of the IFS for all forecasts presented in this paper, cloud microphysics with 5 categories for water substance (vapour, liquid, ice, rain, snow), radiation, shallow convection, turbulent vertical diffusion, and the ECMWF land-surface model (HTESSEL; https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation). The parametrisation of deep convection is switched off for simulations at 1.45 km grid-spacing. However, the parametrisation of shallow convection remains active in all simulations. Fig. 1

Figure 1 shows the topography as it is used in deterministic operational forecasts at ECMWF and in the 1.45 km grid-spacing. The detailed representation of topography is one compelling reason to increase resolution of atmospheric models in complex terrain. Indeed, the figures show a remarkable level of detail with a significant improvement of the representation of valleys for mountain ranges such as the Alps or the Himalayas if grid-spacing is reduced to 1.45 km.

3. Scalability

ECMWF is spending significant resources to optimise simulations with the IFS for present and future high performance computers as part of its scalability programme (see https://www.ecmwf.int/en/about/what-we-do/scalability or Müller et al. (2019) for the Energy-efficient Scalable Algorithms for Weather Prediction at Exascale (ESCAPE) project as examples).

We have performed model simulations with the IFS on the two supercomputers of ECMWF and the Piz Daint supercomputer at the Swiss National Supercomputing Centre (CSCS). ECMWF has two identical CRAY compute clusters. Each of them has 3610 Cray XC40 nodes and a peak performance of 4.25 petaflop. Every node has two Intel E5-2695v4 Broadwell CPUs. Each CPU has 18 compute cores. Piz Daint is the fastest supercomputer in Europe and #6 on the June 2019 TOP500 list (www.top500.org/lists/2019/06/) with a peak performance of 27.15 petaflop. The Cray XC50 has a total of

5704 nodes that are equipped with one 12-core Intel E5-2690 v3 Haswell CPU with 64 Gigabytes of memory and one NVIDIA Tesla P100 GPU per node, interconnected with the Cray Aries network. The simulations of this study did not use the GPUs for computations.

The simulations on Piz Daint for this paper were performed using a hybrid MPI/OpenMP configuration with either 4880 tasks with 12 threads per task or 9776 tasks and 6 threads per task, utilizing 4880 nodes or 4888 nodes, respectively. The two configurations produced similar performance (see Table 1). The performance results that are presented in the following do not consider model initialisation and focus solely on the resources used during model timesteps. Fig. 2

Figure 2 is showing the cost distribution of the different model components for simulations with the IFS at the grid-spacing that is used for routine weather forecasts at ECMWF (9 km), as well as simulations with 1.45 km grid-spacing in hydrostatic and non-hydrostatic mode. The use of single precision will not change the cost fraction significantly as long as I/O and the *Nucleus for European Modelling of the Ocean* (NEMO) ocean model are switched off. The relative cost for spectral transforms are higher for the non-hydrostatic configuration since additional transformations between spectral and grid-point space are required (Wedi et al. (2013)). The hydrostatic simulation is using a finite-element discretisation for the vertical and no predictor-corrector scheme (similar to H-FE-120DT in the next section and the operational setting at ECMWF) while the non-hydrostatic simulation is using a finite-difference discretisation for the vertical and one iteration of the predictor-corrector scheme. Fig. 3

Figure 3 shows the scaling behaviour of the IFS on Piz Daint for simulations with 1.45 km grid-spacing and Table 1 provides information about the simulations. The hydrostatic configurations are significantly less expensive in comparison to non-hydrostatic simulations. Both model configurations show reasonable (strong) scaling behaviour when using most of the available nodes on the supercomputer, in particular for the non-hydrostatic case. However, the data is limited, and the comparison of hydrostatic and non-hydrostatic configurations indicate that the run with significantly shorter elapsed time per time-step appears to be effected by latency within the global communications of the transforms (not shown). Nevertheless, the efficiency of simulations is reaching 0.19 simulated years per day (SYPD) of computation for the hydrostatic model. To allow operational weather and climate simulations would require a throughput of approximately one forecast year per day of computation (obviously also with I/O switched on and with ocean and wave model coupled).

While the performance results of this paper are promising, it should be noted that simulations with the IFS with shorter time-step size, two or three predictor-corrector iterations, and 137 vertical levels, as compared and presented in section 4, will naturally increase computational cost significantly. These simulations scale in the same way, but at higher overall time-to-solution.

The model configuration that is closest for a comparison of performance are the COSMO simulations from Fuhrer et al. (2018) which have documented 0.043 (0.23) SYPD for 930 m (1.9 km) with near-global simulations of the COSMO model scaling to nearly 4888 GPU-accelerated nodes on Piz Daint. Schulthess et al. (2019) is coming to the conclusion that there is a shortfall factor of 101x for the COSMO model and a shortfall factor of 247x for the non-hydrostatic IFS model with a projected 30 s timestep to reach global simulations at 1 km resolution with a throughout of 1 SYPD when running both models on Piz Daint. This does not necessarily indicate that the COSMO model is more efficient since this comparison penalises IFS for using a larger time-step and not using the GPU resources on each node. Under this caveat, we conclude that the IFS simulations presented in this paper are competitive compared to other models. We also list energy consumption figures in Table 1 for our IFS simulations as reported by Piz Daint, which will be useful for future reference since energy-to-solution is an emerging measure of efficiency for Earth-System models. Here, we measure in units of actually consumed MegaWatt hours (MWh) per simulated year

(SY). Table 1

4. Scientific evaluation of selected simulations

In this section, we compare model fidelity between different model simulations at 1.45 km grid-spacing. To identify the impact of different options for the model configuration, we mainly compare six different model runs that are described in Table 2. Table 2

Unfortunately, we could not run the non-hydrostatic simulation with 30 second timesteps and real-world topography as it became unstable. However, this instability could be removed using a more strongly filtered version of the orography (not shown here), or no orography (see notopo-NH-FD-DT30). Furthermore, we anticipate that changes to the non-hydrostatic configuration that are currently implemented, will help to remove these instabilities (Voitus et al. (2019)).

The section will show results for global spectra of horizontal kinetic energy (Section 4.1), probability density functions (PDFs), spectra and snapshots of vertical velocity (Section 4.2), PDFs of precipitation (Section 4.3), plots for satellite simulations in comparison to real satellite data (Section 4.4) and preliminary results for forecast errors of high-resolution simulations (Section 4.5).

4.1 Energy spectra

Fig. 4

To get a first impression of model fidelity for the different model simulations, we have plotted the spectra for horizontal kinetic energy in Figure 4. It should be noted that the spectra presented here are only snapshots and that the model is still not spun-up completely after 12 hours of simulations. The data to average over a longer time period is not available for these simulations. However, we do not expect the qualitative differences between our simulations to change significantly.

The energy spectra show a spurious increase in energy for the NH-FD-DT60 configuration at small scales which is consistent with the instability that we experienced when using a 30 second timestep for the same model configuration. The energy level is slightly higher for notopo-NH-FD-DT30 in comparison to the other simulations at 200 hPa at small scales. The figure is also showing the spectra of a global simulation with 9 km grid-spacing for comparison that clearly fails to transition between the -3 and $-5/3$ scaling behaviour.

One way to assess the realism of horizontal kinetic energy spectra is to identify where the impact of dissipative mechanisms at the tail of the spectrum becomes evident via a departure from the theoretical $-5/3$ curve. The such defined effective resolution, for which the kinetic energy spectrum is reducing in comparison to the expected scaling, is between 5 and 10 km for the simulations at 1.45 km grid-spacing. This is consistent with other measurements of effective resolution from spectra (cf. Abdalla et al. 2013), for example Heinze et al. (2017) and Skamarock et al. (2014) who identify $7 - 8$ times or 6 times the grid-spacing, respectively. Fig. 5

To make differences between the simulations more visible, we plot the horizontal kinetic energy spectra with a compensation for the $-5/3$ scaling in Figure 5. IFS shows more deviations from the theoretical $-5/3$ curve at 200 hPa when compared to results at 500 hPa. More recent comparisons to other models in the DYAMOND project would suggest that this is both a spin-up feature but also specific to IFS (not shown). Overall, the different spectra are similar but differences are visible. Consistent with the total spectra in Figure 4, the NH-FD-DT60 show spurious behaviour at small scales. These features become less prominent if a less ambitious topography field with a coarser resolution is used (not shown here) and are small for the runs without topography (notopo-NH-FD-DT30). However, the non-hydrostatic simulations with topography may be repeated in future in light of ongoing model developments (Voitus et al. (2019)). For the equivalent hydrostatic simulation (H-FD-DT60), there is no increase in the spectra visible for the small scales but the divergent part shows a small bump close to wavenumber 4,000 at 200 hPa. In contrast to the other simulations, the two H-FE simulations have less energy in the divergent part of the spectrum at 500 hPa when compared to the rotational part even for high wavenumbers. Notably, the vertical finite element discretization is of higher order than the 1st-2nd order finite difference discretization.

4.2 Vertical velocity

Fig. 6

For further insight how the different model configurations represent vertical motions (and potentially convection), we have also plotted variance spectra of vertical velocity in Figure 6. As expected, the simulations without topography (notopo-H-FD-DT30 and notopo-NH-FD-DT30) show differences in the spectra of vertical velocity also for large scales. Consistent with the energy spectra in Figure 5, the two non-hydrostatic simulations (NH-FD-DT60 and notopo-NH-DT30) show a spurious increase of variance for small scales. There are clear differences visible between the simulations H-FE-DT120 and H-FE-DT60 which indicates that the dynamics are not yet converged with the timestep. However, differences when changing the vertical resolution and the vertical discretisation from finite element (H-FE-DT60) to finite difference (H-FD-DT60) are even larger.

Fig. 7

Figure 7 shows two-dimensional plots of vertical velocity in the tropics. H-FE-DT120 and H-FE-DT60 are showing larger-scale structures when compared to the other simulations but stronger convective regions. The simulations without topography are less active (notopo-H-FD-DT30 and notopo-NH-FD-DT30). NH-FD-DT60 is showing small-scale patterns of vertical velocities reminiscent of spectral ringing that may also be caused by spurious gravity waves. This signal is consistent with the spurious pattern in the energy spectra that were visible in Figure 5. Overall, differences between H-FD-DT60 and NH-FD-DT60 and between notopo-H-FD-DT30 and notopo-NH-FD-DT30 are rather small which indicates that the difference between hydrostatic and non-hydrostatic simulations is smaller when compared to other changes in the model configuration. Fig. 8

Figure 8 is comparing the probability distribution for vertical velocity for the four runs. Please note that it is not ideal to show only a single snapshot of the PDFs due to the short length of the simulation. While we do expect minor changes if results would be averaged over several independent timesteps, we do not expect qualitative differences in the results since the number of global sampling points is still substantial at least compared to regional simulations. Table 3

Differences in the distribution of vertical velocity are clearly visible for the different simulations and the two vertical levels. It is difficult to relate the measured values to observations but vertical velocities of more than 50 m/s may be unrealistically high. However, the actual number of cells with such large vertical velocities is very small (note the logarithmic scale with the total number of sampling points being 256 Million). Table 3 is listing the number of grid-points with large vertical velocities over the entire globe. The simulations with finite element discretisation and higher resolution in the vertical (H-FE-DT60 and H-FE-DT120) show the highest up-ward velocities while the two non-hydrostatic simulations (NH-FD-DT60 and notopo-NH-FD-DT30) are showing stronger negative velocities. In contrast, the simulation with finite difference discretisation and hydrostatic equations (H-FD-DT60) is showing the smallest vertical velocities. The signal is qualitatively consistent if considered after 12 and 24 hours (see Table 3 for numbers for a subset of runs).

4.3 Precipitation

Fig. 9

Table 4

The shape and distribution of precipitation should change significantly as grid-spacing is reduced from 9 km to 1.45 km. Figure 9 is showing the PDFs of total precipitation for different simulations. For the simulation with 9 km grid-spacing and parametrised convection, the number of grid-points with heavy precipitation is significantly reduced which indicates an ability to improve the representation of local precipitation when resolution is increased. The simulations with 120 seconds timestep (H-FE-DT120) is also showing a lower number of high-precipitation events. The non-hydrostatic simulations show a lower number of events with very large precipitation when compared to the hydrostatic simulations. However, differences are of the same order of magnitude to other changes of the model configuration such as the time step or vertical discretisation.

Within IFS, total precipitation per grid column can have two sources: large-scale precipitation and convective precipitation. Large scale precipitation represents precipitation from resolved atmospheric motions while convective precipitation is motivated by convective updrafts within the grid columns that are not represented explicitly if parametrisation for convection is switched on. For storm-resolving simulations at 1.45 km grid-spacing, the parametrisation of deep convection is switched off while the parametrisation of shallow convection is still enabled. We can therefore expect that convective precipitation will be reduced significantly for storm-resolving simulations and we would hope that large-scale precipitation would increase such that total precipitation is staying at the same level.

To test this hypothesis, Table 4 is presenting the averaged amount of precipitation over the entire globe within the first 12 hours. The results of the table are only based on a single model simulation and are therefore not well established in terms of statistics. However, as expected, all simulations at 1.45 km grid-spacing show a significant reduction of convective precipitation. The large-scale precipitation does indeed buffer the reduction and the amount of total precipitation is in fact increased by approximately 10-15% when compared to the simulation with 9 km grid-spacing. To perform the same evaluation after 24 hours does not change the conclusions (not all runs were simulated for the full 24 hours). However, a further evaluation how the transition between convective and large-scale precipitation is happening when resolution is steadily increased for simulations with and without parametrised deep convection should be performed for future publications.

4.4 Satellite simulators

Fig. 10

Fig. 11

Fig. 12

Figure 10, 11 and 12 show the results for the simulated satellite radiances for the different model runs. The plots were generated with the standard satellite simulator that is used at ECMWF which is based on RTTOV (Hocking et al. (2013)). All runs produce a cloud pattern that is realistic in comparison to the satellite data. It is evident that the higher resolution is beneficial for the representation of clouds with explicit cellular organisation absent in some of the convective areas for the simulation at 9 km grid-spacing. However, the representation of low level clouds seems to fit better to the satellite data for the 9 km simulation when compared to the simulations at higher resolution (see bottom left of Figure 11). This indicates that the simulations at high resolution may require changes to the parametrisation schemes, in particular of shallow convection and the cloud microphysics, but also their interaction with the boundary layer turbulent diffusion, e.g. Duran et al. (2018).

Consistent with the discussion of vertical velocity, the two simulations with finite element discretisation in the vertical (H-FE-DT120 and H-FE-DT60) appear to be too pop-corny with rather large convective cells. The differences between hydrostatic and non-hydrostatic equations is again rather small in comparison.

4.5 Forecast errors

Fig. 13

We have also calculated forecast errors for the headline scores of geopotential height at 500 hPa and temperature at 850 hPa. The two simulations without topography are not considered here. We compare results against the operational forecast configuration. The forecast error was calculated on a O639 octahedral reduced Gaussian grid with 18 km grid-spacing for all simulations. Please note that the forecast error for the operational forecast was calculated against the operational analysis while the other errors are calculated against the long-window analysis to allow for consistency with initial conditions. While these global forecast errors were calculated from a single forecast which does not provide a satisfying level of statistics, it is still evident that an increase in horizontal resolution does not necessarily lead to a reduction in forecast error for a single forecast. In contrast, the simulations with explicitly simulated deep convection show an increased forecast error. Interestingly, this behaviour is not observed in FV3 when comparing simulations at 3.25 km and 13 km grid-spacing (S.J. Lin, personal communication).

5. Discussion of model realism and design choices

The six model simulations with 1.45 km grid-spacing that were evaluated in the previous section provide some corner points with their choices for the length of the time-step, number of iterations in the predictor-corrector scheme, and equations.

The non-hydrostatic simulations are showing some spurious behaviour for energy spectra (Figure 5) and vertical velocity (Figure 7). The hydrostatic simulation that was using a timestep of 120 s did not show spurious behaviour. However, results are also different between the H-FE-DT120 and H-FE-DT60 simulation and this indicates that a time-step size of 120 s violates some time resolution aspects of either cloud/precipitation processes at vertical wind speeds typical for convective cells, or increased trajectory crossings within the semi-Lagrangian advection scheme itself. There is also an indication of too cold top-of-the-atmosphere brightness temperatures in the presence of deep convection (see Figure 11).

The simulations of this paper are entering the resolution range for which differences between hydrostatic and non-hydrostatic equations can be expected (Jeevanjee (2017)). For our simulations, H-FD-DT60 and NH-FD-

DT60 as well as notopo-H-FD-DT30 and notopo-NH-DT30 show similar results except for the spurious behaviour of the spectra of the non-hydrostatic simulations at small scales (Figure 6). Furthermore, to the authors best knowledge, the IFS simulations that were performed for the DYAMOND project at 4 km resolution showed no significant degradation in results in comparison to the other participating models – that were all non-hydrostatic – even at lead times up to 40 days (Stevens et al. (2019)). Since the hydrostatic simulations with the spectral IFS model are much cheaper when compared to non-hydrostatic simulations, we consider the hydrostatic configuration to be a promising candidate for O(1 km) global model simulations at ECMWF. There is also scope that an ensemble of H-FE-DT120 simulations with a much larger number of ensemble members may provide better forecast scores in comparison to an ensemble of H-FE-DT60 simulations at the same computational cost. In the same way as we propose reduced precision simulations, algorithmic choices that enhance the time- or energy-to-solution need to be fairly assessed.

The results of this paper show that forecast errors for Z500 and t850 are higher in comparison to the operational resolution for deterministic forecasts, and that both parametrisation schemes and dynamical core options will require further testing and adjustments to achieve optimal results. However, these results should not be over-interpreted since it is known from previous resolution upgrades at ECMWF that continuous efforts in improving the parametrizations for a given model resolution improve forecast scores. In any case, it is evident that storm-resolving simulations with the IFS may still require significant work before improvements in forecast scores can be realised as the relative weight of different parametrization schemes shifts. This is visible in the amount of total precipitation which is approximately 10-15% higher for storm-resolving simulations (see Table 4). Furthermore, the explicit representation of convective cells will increase variability in the tropics. This may help to improve ensemble spread but may also reduce skill for deterministic forecasts. The increased variability might require an increase in the number of ensemble members for ensemble predictions. This generates additional pressure for the development of highly efficient models to allow for global, operational ensemble simulations that run at storm-resolving resolution in the future. Notably, we have also initialised from a lower resolution analysis which leaves many degrees of freedom uninitialised and the problem of a global 1.45 km analysis is still formidable.

All simulations except one show vertical velocities that appear to be unrealistically large for a small number of grid-cells (Figure 8). This will require more detailed studies to disentangle the impact of microphysical processes and numerical choices. A more detailed evaluation of model fidelity for hydrostatic and non-hydrostatic simulations as well as different dynamical core choices and timesteps (including simulations with less then 30 s) will be performed in future studies. The large timesteps knowingly violate the time resolution required for some of the cloud related processes. However, given the logarithmic distribution of PDFs of precipitation and vertical velocity, it will be interesting to see in the future if for example the simulated climate is sensitive to this, or if this violation is acceptable if measured in climate or ensemble statistics.

Nevertheless, it is promising that all simulations are showing significant differences in the horizontal kinetic energy distribution even at scales of several hundred kilometres when comparing spectra at 9 km and 1.45 km grid-spacing, and this structural difference is also seen in experiments that assess the impact of physical parametrization on energy spectra and on non-linear spectral energy fluxes (Malardel et al. (2016)).

6. Conclusions

In this paper, we document simulations with the IFS that are running with a horizontal grid-spacing of 1.45 km from real-world initial conditions and with real-world topography on the fastest supercomputer in Europe. Results confirm that global storm-resolving simulations are possible today. A simulation that scales to almost the entire size of the fastest supercomputer in Europe can achieve 0.19 SYPD of computation (based on the H-FE-

DT120 configuration with 62 vertical levels). However, these simulations are generating only limited model output, are uncoupled, may require smaller timestep or non-hydrostatic adjustments, and would still be too slow to allow for operational weather and climate predictions that would require a throughput of at least 1 SYPD.

The IFS is performing reasonably well on the limited number of nodes on Piz Daint at 1.45 km and we expect a linear performance scaling if the number of CPUs per node would be increased. Given the scepticism of the community regarding the usefulness of spectral models for simulations at high resolution due to the bad scaling behaviour of the Legendre transformation, it is good news that the spectral IFS model is achieving throughput numbers that are competitive with grid-point models that are based on explicit timestepping schemes. Given the results of this paper and the high efficiency of the IFS in comparison to other global models at slightly lower resolution (Michalakes et al. (2015)), we argue that spectral discretisation combined with semi-implicit semi-Lagrangian time stepping schemes will remain highly competitive towards global storm-resolving simulations in the future. The use of half precision floating point arithmetic and hardware accelerators that were designed for deep learning may provide an additional speed-up for Legendre transformations (Hatfield et al. (2019)). This would, however, require further testing, in particular for simulations with high resolution.

We have presented figures of simulated satellite radiances, topography and model spectra that show improvements in realism and added value for global storm-resolving simulations. It is often argued that global storm-resolving model simulations are already able to pass the Turing test (suggested by Palmer (2016)). This test requires that it is not possible to distinguish between satellite observations and model simulations when looking at cloud fields. We claim that the simulations of this paper pass the Turing test since a simple change of the colour scale in Figure 10 would generate bigger differences than the differences that are visible between simulations and satellite observations. However, the results of this paper also show that differences between the real world and high-resolution simulations and differences between high-resolution simulations with different configurations are still significant and that it will still require significant work to find the optimal model configuration for storm-resolving models and to beat deterministic forecast scores of the current generation of weather models in operations.

The challenges that exascale supercomputing will bring to the domain of Earth System modelling and the likelihood that this will allow global storm-resolving simulations for operational weather and climate predictions have recently been outlined in several papers (see for example Lawrence et al. (2018), Neumann et al. (2019), Schulthess et al. (2019), Schäer et al. (2019), and Biercamp et al. (2019)). While the results of this paper confirm that these simulations could be within reach soon, there can be no question that it will require a large concerted European (or global) effort between modelling and supercomputing centres to face the significant challenges (adaptation to accelerators and heterogeneous hardware, the data avalanche (Balaji et al. (2018)), energy cost, etc.) to make global storm-resolving weather & climate modelling affordable and environmentally acceptable.

Acknowledgements

Many thanks to Cristina Lupus for a lot of help with the satellite simulator of ECMWF, to Pedro Maciel for significant support during the post-processing of model output data at high resolution and to Oliver Fuhrer, Christian Kühnlein, Masaki Satoh, and Bjorn Stevens as well as two anonymous reviewers for very valuable feedback. We gratefully acknowledge Thomas Schulthess and Giuffreda Maria Grazia for providing access to Piz Daint. Peter D. Düben gratefully acknowledges funding from the Royal Society for his University Research Fellowship and the ESIWACE and ESIWACE2 project. The ESIWACE and ESIWACE2 projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675191 and 823988. Nils Wedi acknowledges support from the ESCAPE / ESCAPE-2 projects under the European Unions Horizon 2020 research and innovation programme under grant agreement No 67162 and 800 987, respectively.

References

S. Abdalla, L. Isaksen, P.A.E.M. Janssen, and N. Wedi. Effective spectral resolution of ecmwf atmospheric forecast models. *ECMWF Newsletter*, 137, 2013.

V. Balaji, K. E. Taylor, M. Juckes, B. N. Lawrence, P. J. Durack, M. Lautenschlager, C. Blanton, L. Cinquini, S. Denvil, M. Elkington, F. Guglielmo, E. Guilyardi, D. Hassell, S. Kharin, S. Kindermann, S. Nikonov, A. Radhakrishnan, M. Stockhause, T. Weigel, and D. Williams. Requirements for a global data infrastructure in support of CMIP6. *Geoscientific Model Development*, 11(9), 2018.

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525, 2015. URL `https://doi.org/10.1038/nature14956`.

J. Biercamp, P. Bauer, P. Dueben, and B. Lawrence. A roadmap to the implementation of 1km earth system model ensembles. *ESiWACE Deliverable*, D1.2, 2019.

Christopher S. Bretherton and Marat F. Khairoutdinov. Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. *Journal of Advances in Modeling Earth Systems*, 7(4):1765–1787, 2015. doi: 10.1002/2015MS000499. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015MS000499`.

R. Bubnová, G. Hello, P. Bénard, and J.-F. Geleyn. Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Monthly Weather Review*, 123:515–535, 1995.

Roger Daley. The normal modes of the spherical non-hydrostatic equations with applications to the filtering of acoustic modes. *Tellus A: Dynamic Meteorology and Oceanography*, 40(2):96–106, 1988. doi: 10.3402/tellusa.v40i2.11785.

Peter D. Düben and T. N. Palmer. Benchmark tests for numerical forecasts on inexact hardware. *Monthly Weather Review*, 142:3809–3829, 2014.

J. Dudhia. A nonhydrostatic version of the Penn State-NCAR mesoscale model: validation tests and simulation of an Atlantic cyclone and cold front. *Monthly Weather Review*, 121 (5):1493–1513, 1993. ISSN 00270644. doi: 10.1175/1520-0493(1993)121¡1493:ANVOTP¿2.0.CO;2.

I. B. Duran, J.-F. Geleyn, F. Vana, J. Schmidli, and R. Brozkova. A turbulence scheme with two prognostic turbulence energies. *Journal of Atmospheric Sciences*, 75:3381–3401, 2018.

Hironori Fudeyasu, Yuqing Wang, Masaki Satoh, Tomoe Nasuno, Hiroaki Miura, and Wataru Yanase. Global cloud-system-resolving model nicam successfully simulated the lifecycles of two real tropical cyclones. *Geophysical Research Letters*, 35(22), 2008. doi: 10.1029/2008GL036003.

O. Fuhrer, T. Chadha, T. Hoefler, G. Kwasniewski, X. Lapillonne, D. Leutwyler, D. Lüthi, C. Osuna, C. Schär, T. C. Schulthess, and H. Vogt. Near-global climate simulation at 1 km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0. *Geoscientific Model Development*, 11 (4):1665–1681, 2018. doi: 10.5194/gmd-11-1665-2018. URL `https://www.geosci-model-dev.net/11/1665/2018/`.

S. Hatfield, M. Chantry, P. Dueben, and T. Palmer. Accelerating high-resolution weather models with deep-learning hardware. *PASC2019 Conference Proceedings*, 2019.

Rieke Heinze, Anurag Dipankar, Cintia Carbajal Henken, Christopher Moseley, Odran Sourdeval, Silke Trömel, Xinxin Xie, Panos Adamidis, Felix Ament, Holger Baars, Christian Barthlott, Andreas Behrendt, Ulrich Blahak, Sebastian Bley, Slavko Brdar, Matthias Brueck, Susanne Crewell, Hartwig Deneke, Paolo Di Girolamo, Raquel Evaristo, Jürgen Fischer, Christopher Frank, Petra Friederichs, Tobias Göcke, Ksenia Gorges, Luke Hande, Moritz Hanke, Akio Hansen, Hans-Christian Hege, Corinna Hoose, Thomas

Jahns, Norbert Kalthoff, Daniel Klocke, Stefan Kneifel, Peter Knippertz, Alexander Kuhn, Thriza van Laar, Andreas Macke, Vera Maurer, Bernhard Mayer, Catrin I. Meyer, Shravan K. Muppa, Roeland A. J. Neggers, Emiliano Orlandi, Florian Pantillon, Bernhard Pospichal, Niklas Röber, Leonhard Scheck, Axel Seifert, Patric Seifert, Fabian Senf, Pavan Siligam, Clemens Simmer, Sandra Steinke, Bjorn Stevens, Kathrin Wapler, Michael Weniger, Volker Wulfmeyer, Gnther Zngl, Dan Zhang, and Johannes Quaas. Large-eddy simulations over germany using icon: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, 143 (702):69–100, 2017. doi: 10.1002/qj.2947.

J. Hocking, P. Rayer, D. Rundle, R. Saunders, Matricardi, A. M., Geer, P. Brunel, and Vidot J. RTTOV v11 users guide. *NWP SAF report, Met Office*, page 107 pp., 2013.

Nadir Jeevanjee. Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, 9(6): 2304–2316, 2017. doi: 10.1002/2017MS001059. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001059`.

T. Kato. Hydrostatic and non-hydrostatic simulations of moist convection:

Review and further study. *Meteorology and Atmospheric Physics*, 63 (1-2):39–51, 1997. ISSN 01777971. doi: 10.1007/BF01025363.

Teddie L. Keller. Implications of the Hydrostatic Assumption on Atmospheric Gravity Waves. *Journal of the Atmospheric Sciences*, 51(13):1915–1929, 1994. ISSN 0022-4928. doi: 10.1175/1520-0469(1994)051¡1915:iothao¿2.0.co;2.

C. Kühnlein, W. Deconinck, R. Klein, S. Malardel, Z. P. Piotrowski, P. K. Smolarkiewicz, J. Szmelter, and N. P. Wedi. FVM 1.0: a nonhydrostatic finite-volume dynamical core for the IFS. *Geoscientific Model Development*, 12(2):651–676, 2019. doi: 10.5194/gmd-12-651-2019. URL `https://www.geosci-model-dev.net/12/651/2019/`.

B. N. Lawrence, M. Rezny, R. Budich, P. Bauer, J. Behrens, M. Carter, W. Deconinck, R. Ford, C. Maynard, S. Mullerworth, C. Osuna, A. Porter, K. Serradell, S. Valcke, N. Wedi, and S. Wilson. Crossing the chasm: how to develop weather and climate models for next generation computers? *Geoscientific Model Development*, 11(5):1799–1821, 2018. doi: 10.5194/gmd-11-1799-2018. URL `https://www.geosci-model-dev.net/11/1799/2018/`.

D. Leutwyler, D. Lüthi, N. Ban, O. Fuhrer, and C. Schär. Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research, Atmospheres*, 122:52375258, 2017. doi: 10.1002/qj.3502.

S. Malardel, Nils Wedi, Willem Deconinck, Michail Diamantakis, Christian Kühnlein, G. Mozdzynski, M. Hamrud, and Piotr Smolarkiewicz. A new grid for the IFS. *ECMWF Newsletter*, (146):23–28, 2016. doi: 10.21957/zwdu9u5i. URL `https://www.ecmwf.int/node/17262`.

G. Mengaldo, A. Wyszogrodski, M. Diamantakis, S-J Lock, F.X. Giraldo, and N. P. Wedi. Current and emerging time-integration strategies in global numerical weather and climate prediction. *Arch. Computat. Methods Eng.*, pages 1–22, 2018.

John Michalakes, Mark Govett, Rusty Benson, Tom Black, Hann-Ming Henry Juang, Alex Reinecke, and Bill Skamarock. AVEC report: NGGPS level-1 benchmarks and software evaluation. *Advanced Computing Evaluation Committee (AVEC)*, 2015. URL `https://repository.library.noaa.gov/view/noaa/18654`.

Hiroaki Miura, Masaki Satoh, Tomoe Nasuno, Akira T. Noda, and Kazuyoshi Oouchi. A Madden-Julian oscillation event realistically simulated by a global cloud-resolving model. *Science*, 318(5857): 1763–1765, 2007. ISSN 0036-8075. doi: 10.1126/science.1148443. URL `http://science.sciencemag.org/content/318/5857/1763`.

Yoshiaki Miyamoto, Yoshiyuki Kajikawa, Ryuji Yoshida, Tsuyoshi Yamaura, Hisashi Yashiro, and Hirofumi Tomita. Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, 40 (18):4922–4926, 2013. doi: 10.1002/grl.50944. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50944`.

A. Müller, W. Deconinck, C. Kühnlein, G. Mengaldo, M. Lange, N. Wedi, P. Bauer, P. K. Smolarkiewicz, M. Diamantakis, S.-J. Lock, M. Hamrud, S. Saarinen, G. Mozdzynski, D. Thiemert, M. Glinton, P. Bénard, F. Voitus, C. Colavolpe, P. Marguinaud, Y. Zheng, J. Van Bever, D. Degrauwe, G. Smet, P. Termonia, K. P. Nielsen, B. H. Sass, J. W. Poulsen, P. Berg, C. Osuna, O. Fuhrer, V. Clement, M. Baldauf, M. Gillard, J. Szmelter, E. O'Brien, A. McKinstry, O. Robinson, P. Shukla, M. Lysaght, M. Kulczewski, M. Ciznicki, W. Piatek, S. Ciesielski, M. Błaewicz, K. Kurowski, M. Procyk, P. Spychala, B. Bosak, Z. Piotrowski, A. Wyszogrodzki, E. Raffin, C. Mazauric, D. Guibert, L. Douriez, X. Vigouroux, A. Gray, P. Messmer, A. J. Macfaden, and N. New. The ESCAPE project: Energy-efficient Scalable Algorithms for Weather Prediction at Exascale. *Geoscientific Model Development Discussions*, 2019:1–50, 2019. doi: 10.5194/gmd-2018-304. URL https://www.geosci-model-dev-discuss.net/gmd-2018-304/.

Andreas Müller, Michal A. Kopera, Simone Marras, Lucas C. Wilcox, Tobin Isaac, and Francis X. Giraldo. Strong scaling for numerical weather prediction at petascale with the atmospheric model NUMA. *CoRR*, abs/1511.01561, 2015. URL http://arxiv.org/abs/1511.01561.

Philipp Neumann, Peter D. Dueben, Panagiotis Adamidis, Peter Bauer, Matthias Brueck, Luis Kornblueh, Daniel Klocke, Bjorn Stevens, Nils Wedi, and Joachim Biercamp. Assessing the scales in numerical weather and climate predictions: Will exascale be the rescue? *Philosophical Transactions of the Royal Society A*, 377, 2018.

I. Orlanski. The quasi-hydrostatic approximation. *Journal of the Atmospheric Sciences*, 38(3):572–582, 1981. ISSN 00224928. doi: 10.1175/1520-0469(1981)038¡0572:TQHA¿2.0.CO;2.

T. N. Palmer. A personal perspective on modelling the climate system. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2188): 20150772, 2016. doi: 10.1098/rspa.2015.0772. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2015.0772.

B. B. Ross and I. Orlanski. The Circulation Associated with a Cold

Front. Part II: Moist Case. *Journal of the Atmospheric Sciences*, 35(3):445–465, 1978. ISSN 0022-4928. doi: 10.1175/1520-0469(1978)035¡0445:tcawac¿2.0.co;2.

M. Satoh, T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga. Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, 227(7):3486 – 3514, 2008. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2007.02.006. Predicting weather, climate and extreme events.

Masaki Satoh, Bjorn Stevens, Falko Judt, Marat Khairoutdinov, Shian-Jiann Lin, William M. Putman, and Peter Dueben. Global cloud-resolving models. *Current Climate Change Reports*, 5(3):172–184, Sep 2019. ISSN 2198-6061. doi: 10.1007/s40641-019-00131-0. URL `https://doi.org/10.1007/s40641-019-00131-0`.

C. Schär, A. Arteaga, N. Ban, C. Charpilloz, and et al. Kilometer-scale climate models: Prospects and challenges. *Submitted to BAMS*, 2019.

T. C. Schulthess, P. Bauer, N. Wedi, O. Fuhrer, T. Hoefler, and C. Schär. Reflecting on the goal and baseline for exascale computing: A roadmap based on weather and climate simulations. *Computing in Science Engineering*, 2019.

William C. Skamarock, Sang-Hun Park, Joseph B. Klemp, and Chris Snyder. Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *Journal of the Atmospheric Sciences*, 71 (11):4369–4381, 2014. doi: 10.1175/JAS-D-14-0114.1.

B. Stevens, M. Satoh, L. Auger, J. Biercamp, C. Bretherton, X. Chen, P. Dueben, F. Judt, M. Khairoutdinov, D. Klocke, C. Kodama, L. Kornblueh, S.-J. Lin, W. Putman, S. Ryosuke, P. Neumann, N. Roeber, B. Vannier, P.-L. Vidale, N. Wedi, and L. Zhou. DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Accepted in Progress in Earth and Planetary Science*, 2019.

P. Termonia, C. Fischer, E. Bazile, F. Bouyssel, R. Brozková, P. Bénard, B. Bochenek, D. Degrauwe, M. Derková, R. E. Khatib, R. Hamdi, J. Masek, P. Pottier, N. Pristov, Y. Seity, P. Smolkov, O. Spaniel, M. Tudor, Y. Wang, C. Wittmann, and A. Joly. The ALADIN system and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geoscientific Model Development*, 11:257–281, 2018. doi: 10.5194/gmd-11-257-2018. URL `https://www.geosci-model-dev.net/11/257/2018/gmd-11-257-2018.pdf`.

M. Tort and T. Dubos. Dynamically consistent shallow?atmosphere equations with a complete coriolis force. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2388–2392, 2014.

Filip Vana, Peter Dueben, Simon Lang, Tim Palmer, Martin Leutbecher, Deborah Salmond, and Glenn Carver. Single precision in weather forecasting models: An evaluation with the IFS. *Monthly Weather Review*, 145(2):495–502, 2017. doi: 10.1175/MWR-D-16-0228.1. URL `https://doi.org/10.1175/MWR-D-16-0228.1`.

F. Voitus, P. Bénard, C. Kühnlein, and N.P. Wedi. Semi-implicit integration of the unified equations in a mass-based coordinate: model formulation and numerical testing. *Quarterly Journal of the Royal Meteorological Society*, 2019. available online.

N. P. Wedi and P. K. Smolarkiewicz. A framework for testing global non-hydrostatic models. *Quarterly Journal of the Royal Meteorological Society*, 135:469–484, 2009.

N. P. Wedi, M. Hamrud, and G. Mozdzynski. A fast spherical harmonics transform for global NWP and climate models. *Monthly Weather Review*, 141:3450–3461, 2013.

N. P. Wedi, P. Bauer, W. Deconinck, M. Diamantakis, M. Hamrud, C. Kühnlein, S. Malardel, K. Mogensen, G. Mozdzynski, and P.K.

Smolarkiewicz. The modelling infrastructure of the integrated forecasting system: Recent advances and future challenges. Technical Report 760, Eur. Cent. For Medium-Range Weather Forecasts, Reading, UK, 2015.

Nils Wedi, K. Yessad, and A. Untch. The non-hydrostatic global IFS/ARPEGE model: model formulation and testing. (594):34, 10 2009.

Nils P. Wedi. Increasing horizontal resolution in numerical weather prediction and climate simulations: illusion or panacea? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 372(2018), 2014. ISSN 1364-503X. doi: 10.1098/rsta.2013.0289.

Morris L. Weisman, William C. Skamarock, and Joseph B. Klemp. The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125(4):527–548, 1997.

A. A. Wing, K. A. Reed, M. Satoh, B. Stevens, S. Bony, and T. Ohno. Radiative-convective equilibrium model intercomparison project. *Geoscientific Model Development*, 11 (2):793–813, 2018. doi: 10.5194/gmd-11-793-2018. URL https://www.geosci-model-dev.net/11/793/2018/.

C. Yang, W. Xue, H. Fu, H. You, X. Wang, Y. Ao, F. Liu, L. Gan, P. Xu, L. Wang, G. Yang, and W. Zheng. 10m-core scalable fully-implicit solver for nonhydrostatic atmospheric dynamics. In *SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 57–68, Nov 2016. doi: 10.1109/SC.2016.5.

List of Figures

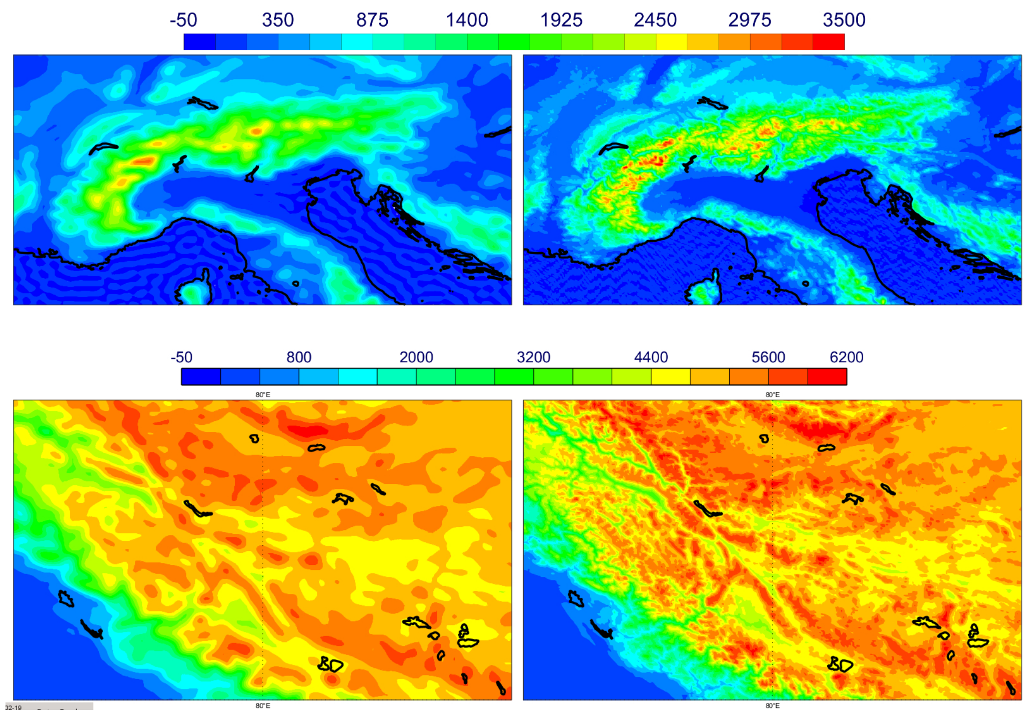

Fig. 1. Topography in [m] of the Alps (42N/4W/49N/18W; top) and parts of the Himalayas (25N/70W/N43/100W; bottom) for simulations at 9 km (left) and 1.45 km (right) grid-spacing. The land-water mask is drawn as black contour.

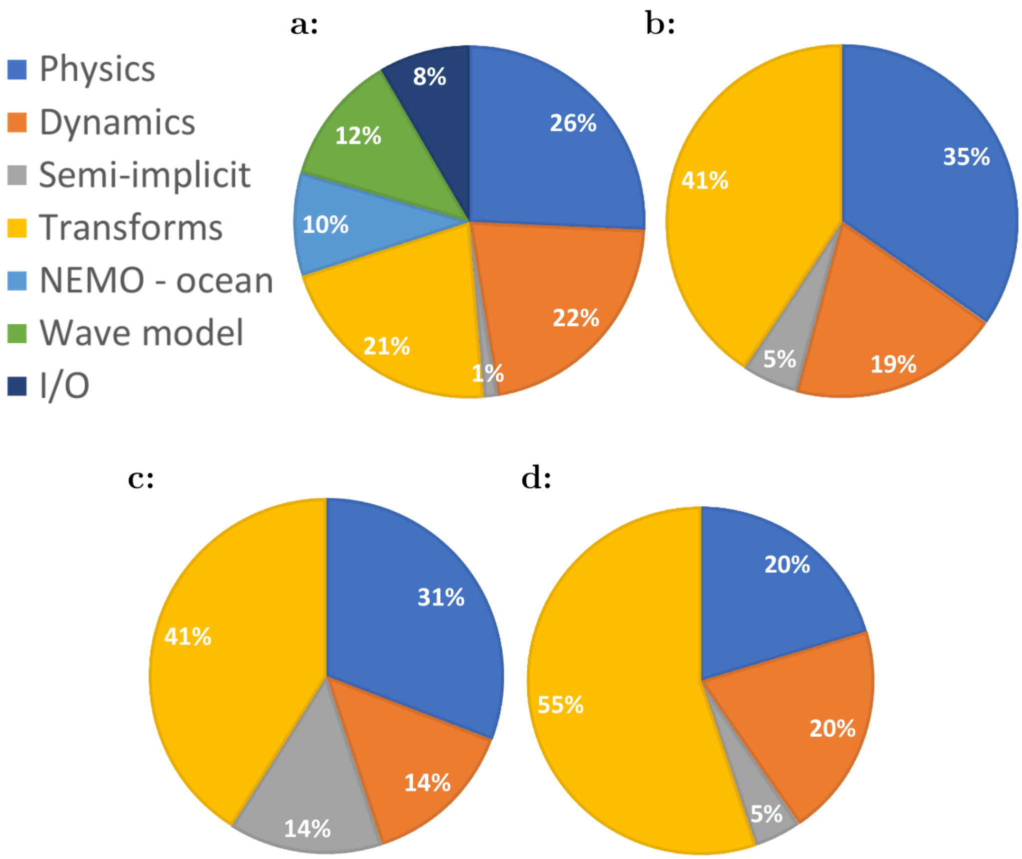

Fig. 2. Cost distribution for different simulations. **a:** Operational forecast at ECMWF at 9 km grid-spacing with I/O on the ECMWF computer, 137 vertical levels, hydrostatic equations and double precision. **b:** Forecast simulation at 9 km grid-spacing on the ECMWF computer but without I/O and with 62 vertical levels, hydrostatic equations and single precision. **c:** Forecast simulation at 1.45 km grid-spacing with 4880 nodes (12 threads per MPI task) of Piz Daint without I/O, 62 vertical levels, hydrostatic equations and single precision. **d:** same as **c** but for non-hydrostatic equations with one predictor-corrector iteration.

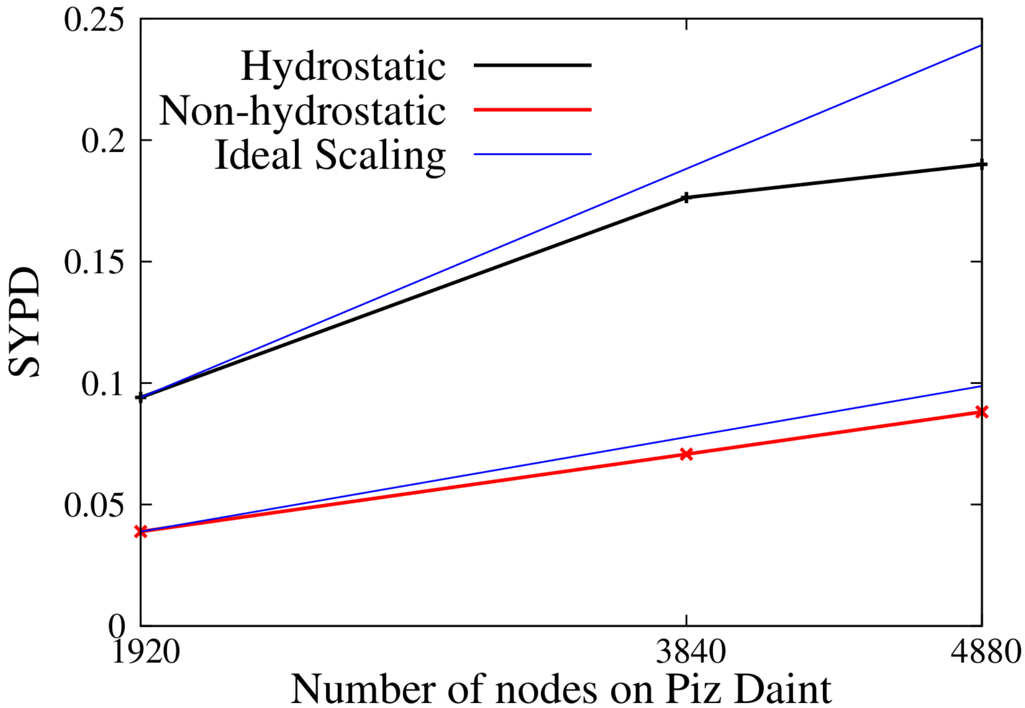

Fig. 3. (Strong) scaling of IFS simulations on Piz Daint at 1.45 km grid-spacing and with 62 vertical levels. Please note that this is not a logarithmic plot.

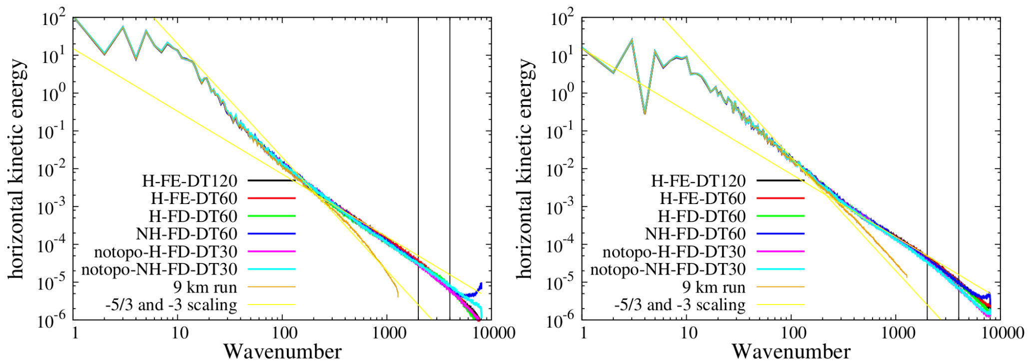

Fig. 4. Spectra for horizontal kinetic energy for simulations with the IFS at 200 hPa (left) and 500 hPa (right) for the six model configurations with 1.45 km grid-spacing and a simulation with 9 km grid-spacing (H, FE, 62 vertical levels, $\Delta t = 450s$, 0 PC). The vertical black lines mark the grid-spacing of 5 km and 10 km respectively.

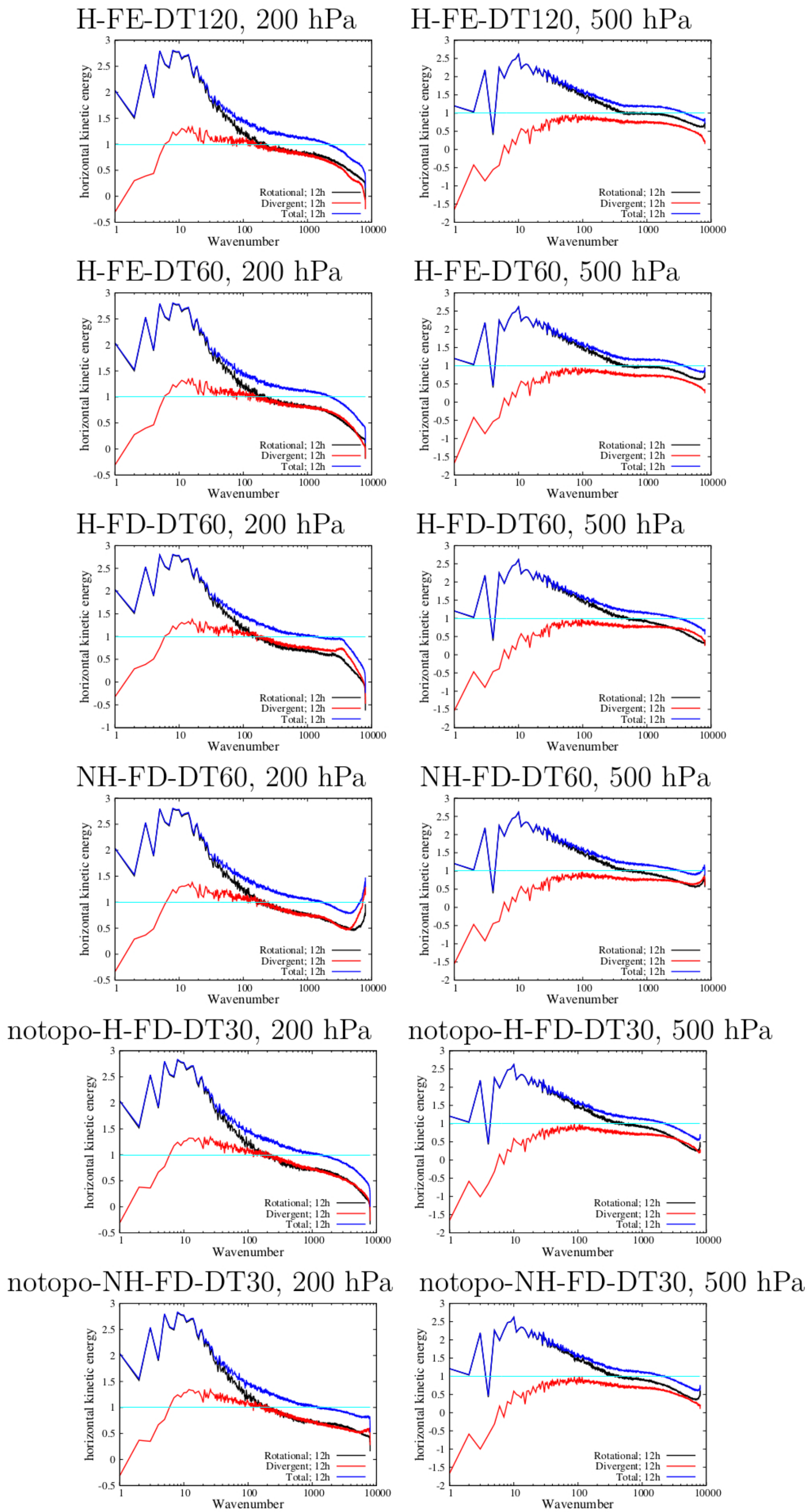

Fig. 5. Spectra of horizontal kinetic energy for simulations with the IFS at 1.45 km grid-spacing 12 hours into the forecast at 200 hPa (left) and 500 hPa (right) for the six model configurations. The plots show the total as well as the rotational and divergent components of the horizontal kinetic energy spectra. The coefficients were multiplied with $k^{5/3}$ to improve visibility. The light blue horizontal line indicates -5/3 scaling.

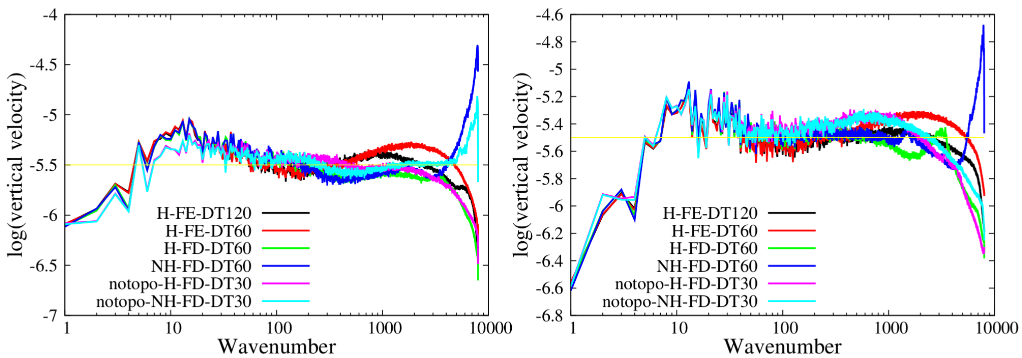

Fig. 6. Spectra of vertical velocity at 250 (left) and 500 (right) hPa 12 hours into the forecast.

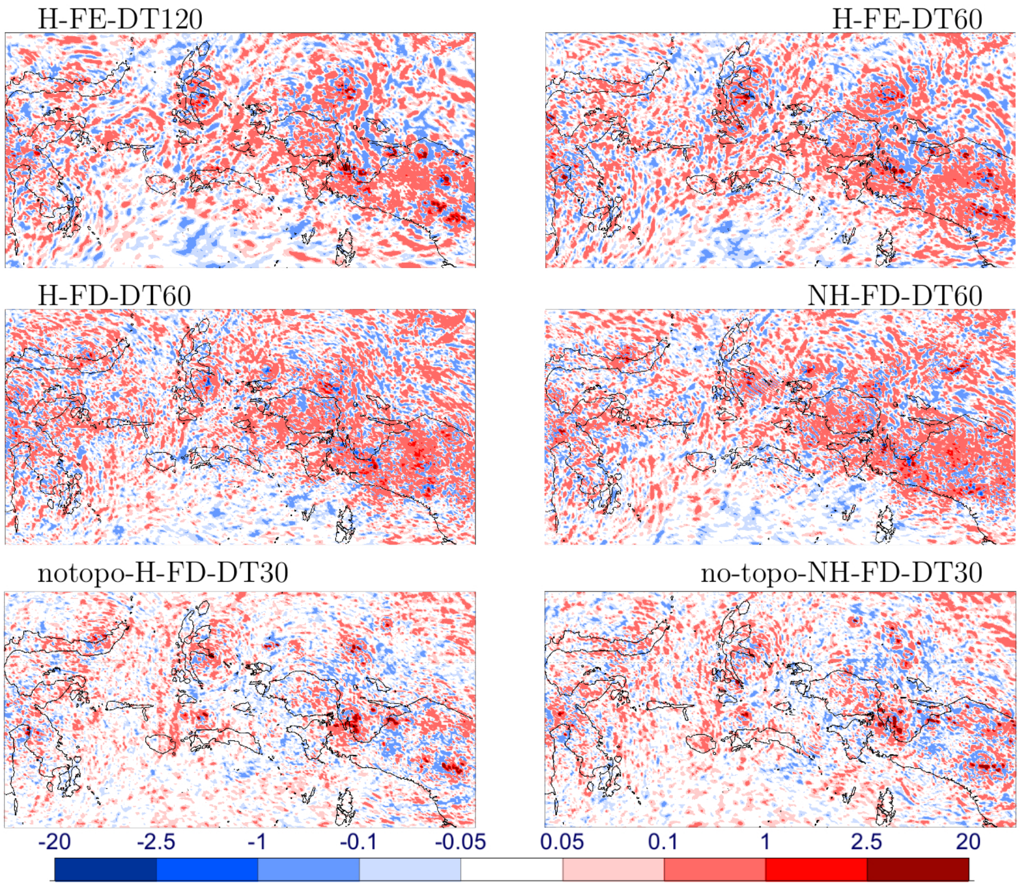

Fig. 7. Vertical velocity [m/s] at 250 hPa 12 hours into the forecast for a tropical area around Indonesia (7S/120W/3N/140W; top to bottom). Please note the non-linear colour scale.

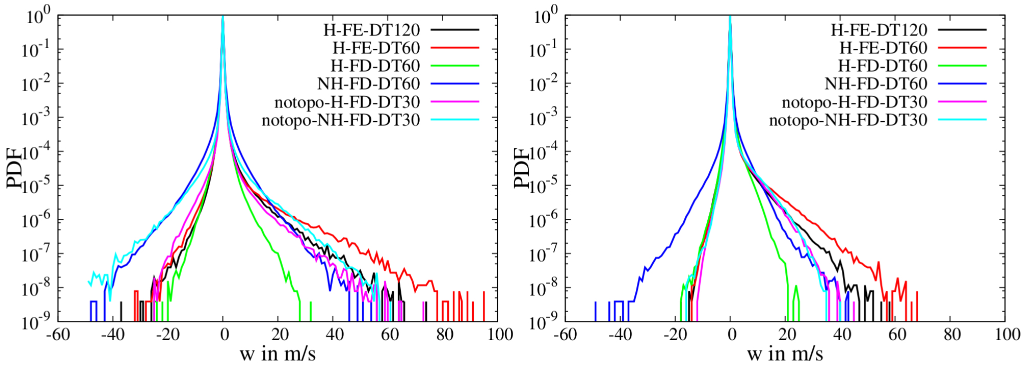

Fig. 8. PDFs of vertical velocity 12 hours into the forecast at 250 hPa (left) and 500 hPa (right). Please note the logarithmic scale on the y-axis.

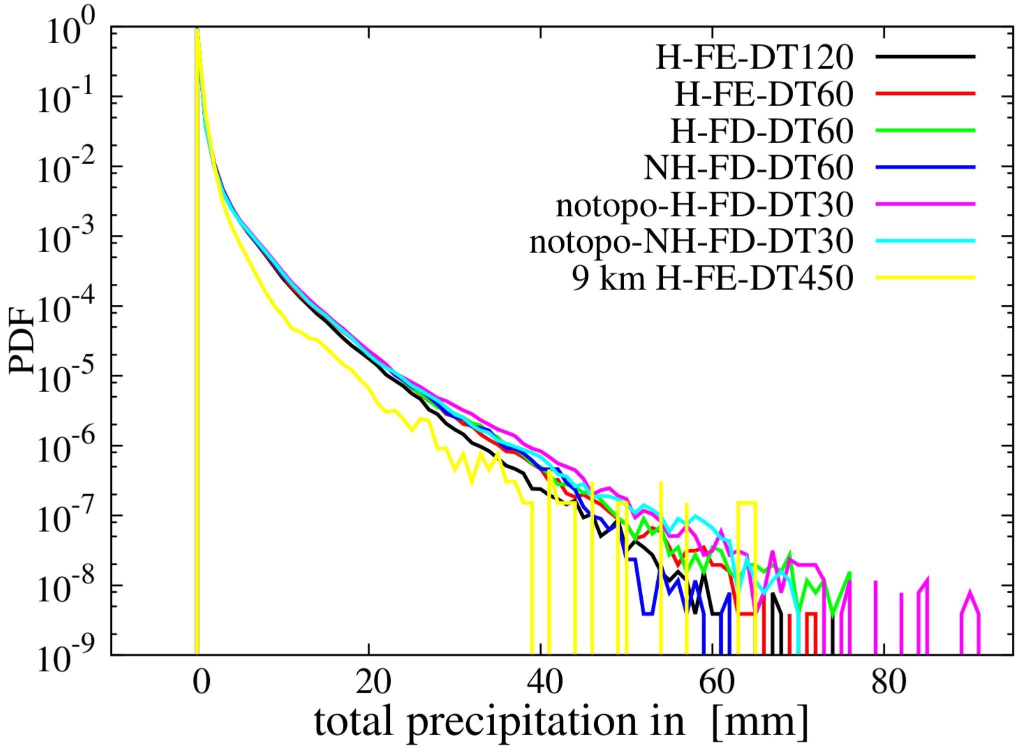

Fig. 9. PDFs of total precipitation integrated for the first 12 hours for the simulations at 1.45 km gridspacing and a simulation with 9 km gridspacing (H, FE, 62 vertical levels, $\Delta t = 450s$, 0 PC). Please note the logarithmic scale on the y-axis. The bin-size when calculating the PDF was 0.1 mm which results in the majority of gridpoints being in the first bin of less than 0.1 mm precipitation.

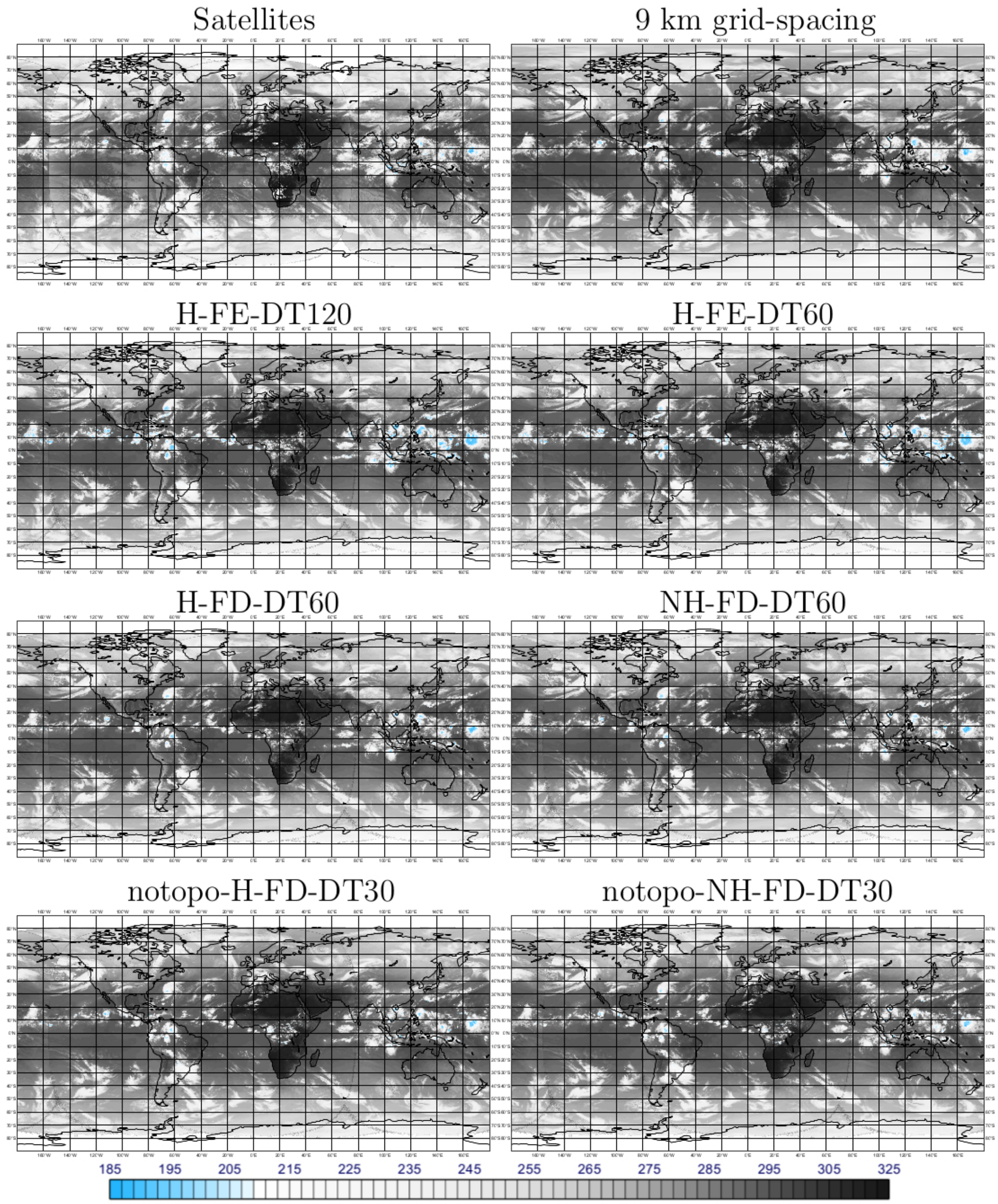

Fig. 10. Simulated and observed top-of-the-atmosphere brightness temperatures derived from satellites and satellite simulators for 16th October 2016, 12 UTC. We use data from different satellites to generate the panel on the top left (Meteosat-7 at 12 UTC and Meteosat-10 at 11:45 UTC from EUMETSAT, GOES-13 at 12 UTC and GOES-15 at 12 UTC from NOAA and Himawari-8 at 11 UTC from the Japan Meteorological Agency). The plot on the top right shows results for simulated satellite radiances of the operational weather forecast at ECMWF at 9 km grid-spacing with parametrised deep convection and 137 vertical levels. The other plots show results of the model simulations with 1.45 km gridspacing.

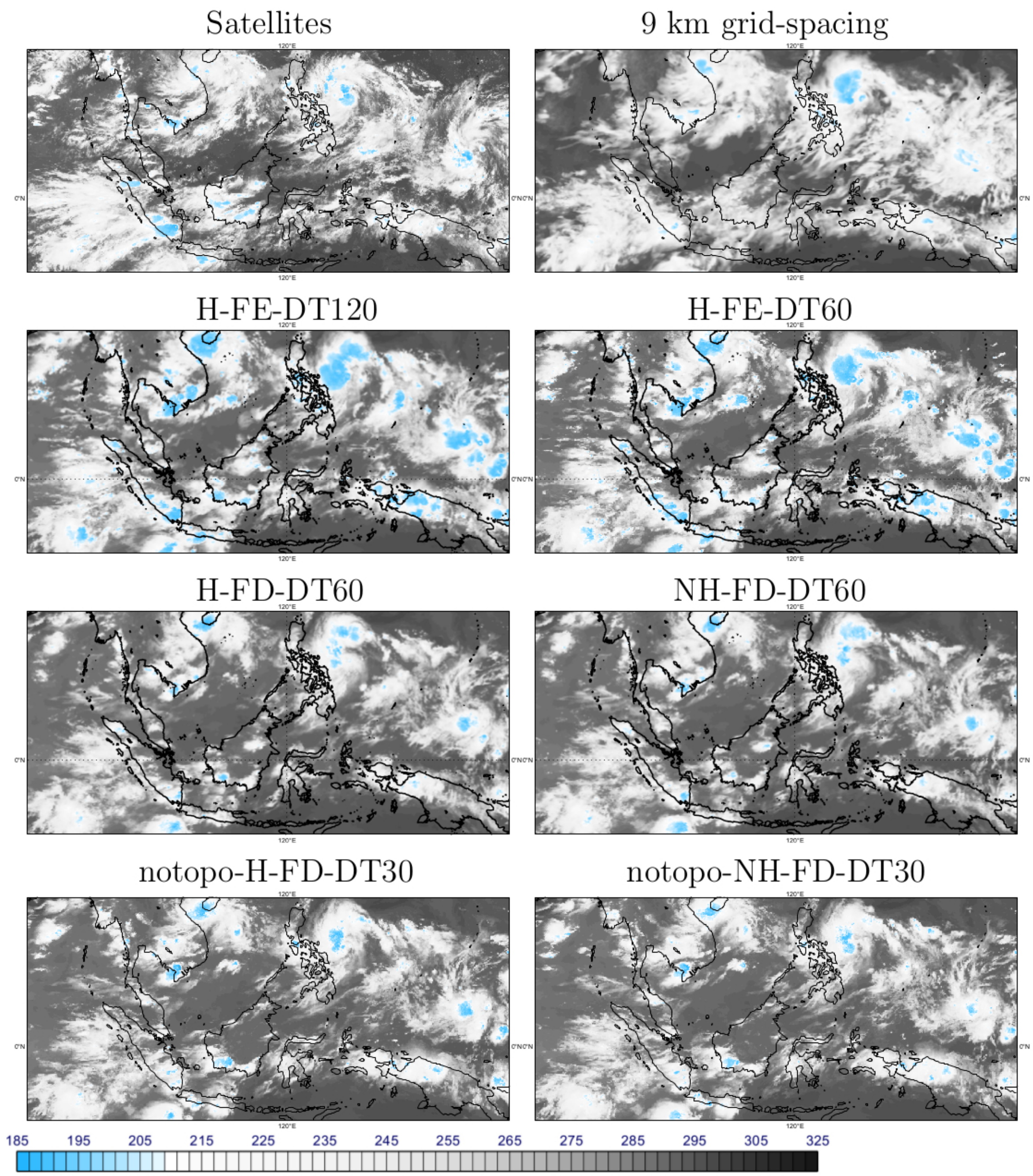

Fig. 11. Same as Figure 10 but for the area over Indonesia (10S/85W/20N/150W).

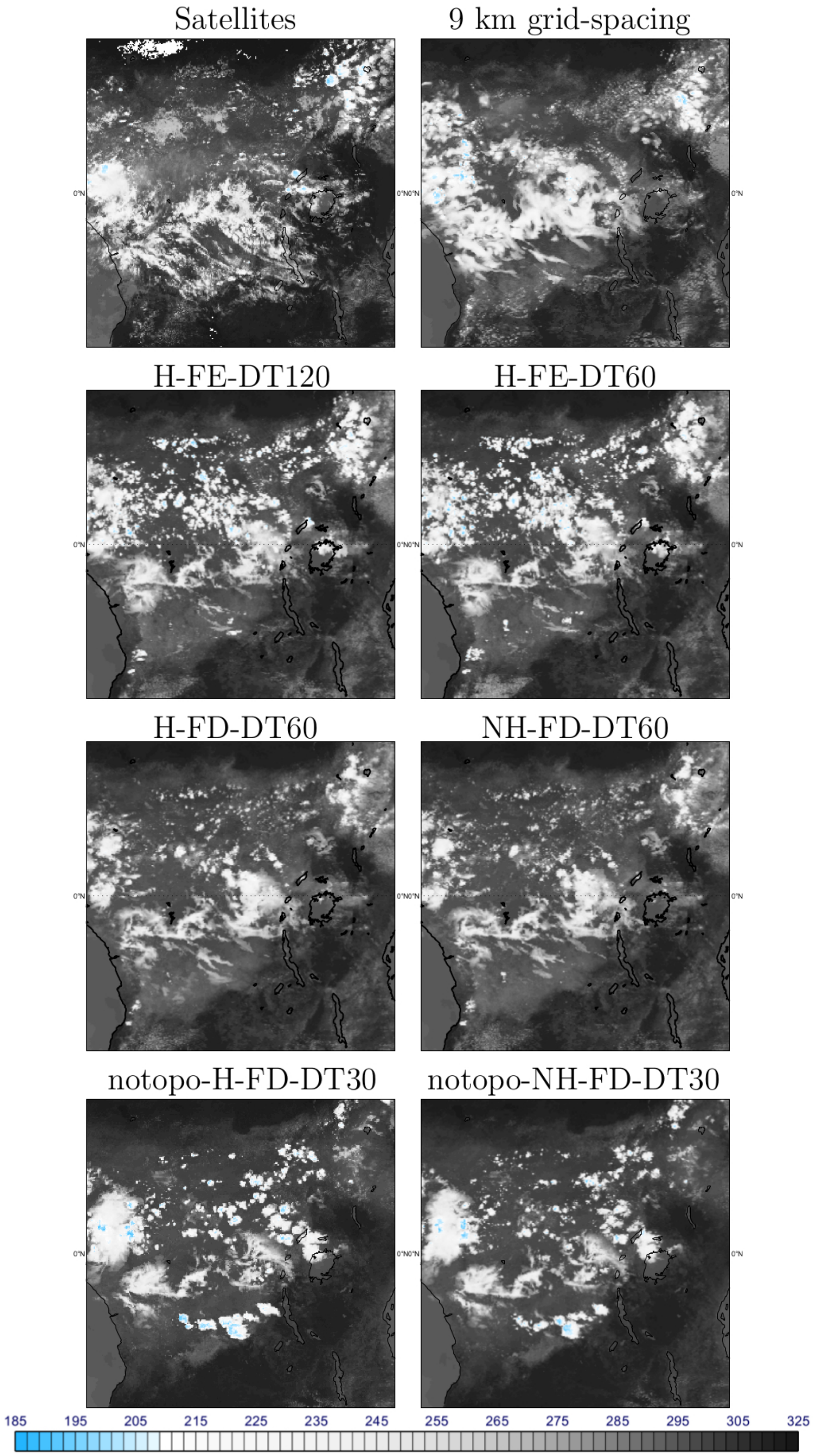

Fig. 12. Same as Figure 10 but for an area over Africa (15S/10W/15N/40W).

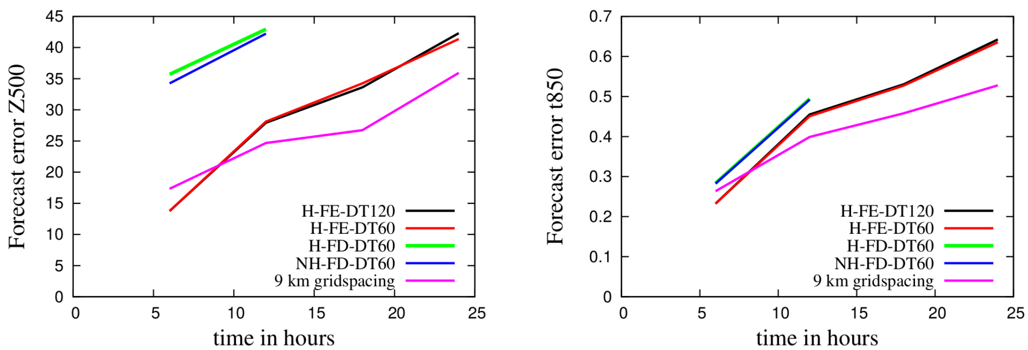

Fig. 13. Mean absolute error averaged over the globe plotted against forecast lead time that was calculated against analysis products for geopotential height at 500 hPa (Z500) and temperature at 850 hPa (t850) for a single forecast with different model configurations. The simulation with 9 km grid-spacing is the operational forecast at ECMWF (H, FE, 137 vertical levels, $\Delta t = 450s$, 0 PC, coupled to NEMO and the wave model).

List of Tables

Table 1. Scalability tests with the IFS on Piz Daint for simulations with 1.45 km horizontal grid-spacing and 62 vertical levels when running on 4880 or 4888 nodes. The GPUs of the compute nodes were not used.

Dycore option	#tasks and threads	Energy consumption per year	Throughput
Hydrostatic	4880 tasks; 12 threads per task	85.21 MWh/SY	0.190 SYPD
Non-hydrostatic	9776 tasks; 6 threads per task	191.74 MWh/SY	0.088 SYPD
Non-hydrostatic	4880 tasks; 12 threads per task	195.30 MWh/SY	0.085 SYPD

Table 2. Properties of the simulations that are evaluated in Section 4. The table provides the identifier that is used for each run in the rest of the paper, information whether the run was hydrostatic or non-hydrostatic and whether it was using real-world or flat topography, the number of vertical levels, the vertical discretisation method, the length of the timestep, as well as the number of predictor-corrector (PC) iterations.

Run Identifier	Hydrostatic?	Topo-graphy	Vertical levels	Vertical disc.	timestep and number of PC iterations
H-FE-DT120	Yes	Yes	137	Finite element	120s / 0 PC
H-FE-DT60	Yes	Yes	137	Finite element	60s / 0 PC
H-FD-DT60	Yes	Yes	62	Finite difference	60s / 3 PC
NH-FD-DT60	No	Yes	62	Finite difference	60s / 3 PC
notopo-H-FD-DT30	Yes	No	62	Finite difference	30s / 2 PC
notopo-NH-FD-DT30	No	No	62	Finite difference	30s / 2 PC

Table 3. Number of grid-points with large vertical velocities for the different runs. A simulation with 1.45 km grid-spacing has a total of $256,288,000$ grid-points per vertical level. For some of the runs the values for both 12 and 24 hours into the forecast are available.

Run	Height	>10 m/s	>20 m/s	>30 m/s
H-FE-DT120-12h	250 hPa	11,307	2,428	607
H-FE-DT120-24h	250 hPa	12,651	3,087	846
H-FE-DT60-12h	250 hPa	16,187	5,474	2,225
H-FE-DT60-24h	250 hPa	15,028	5,543	2,306
H-FD-DT60-12h	250 hPa	1,418	60	1
NH-FD-DT60-12h	250 hPa	21,914	1,870	219
notopo-H-FD-DT30	250 hPa	7,744	1,456	330
notopo-NH-FD-DT30	250 hPa	21,379	4,323	882
H-FE-DT120-12h	500 hPa	15,945	2,228	338
H-FE-DT120-24h	500 hPa	19,005	3,289	601
H-FE-DT60-12h	500 hPa	27,526	6,036	1,335
H-FE-DT60-24h	500 hPa	26,435	6,913	1,846
H-FD-DT60-12h	500 hPa	2,487	13	0
NH-FD-DT60-12h	500 hPa	11,992	450	63
notopo-H-FD-DT30	500 hPa	17,237	1,289	61
notopo-NH-FD-DT30	500 hPa	21,521	1,625	39

Table 4. Global average for large-scale precipitation, parametrised convective precipitation and total precipitation integrated over the first and the second 12 hours of the simulation (all in [mm]).

Run	Large-scale	Convective	Total
9 km grid-spacing – 12 h	0.7512	0.6521	1.4034
9 km grid-spacing – 24 h	0.6787	0.7325	1.4112
H-FE-DT120 – 12 h	1.4017	0.1658	1.5675
H-FE-DT120 – 24 h	1.4980	0.1582	1.6563
H-FE-DT60 – 12 h	1.3976	0.1805	1.5781
H-FE-DT60 – 24 h	1.4819	0.1695	1.6514
H-FD-DT60 – 12 h	1.3675	0.2325	1.6000
NH-FD-DT60 – 12 h	1.3695	0.2300	1.5992
notopo-H-FD-DT30 – 12	1.3197	0.2779	1.5976
notopo-NH-FD-DT30 – 12	1.3107	0.2672	1.5780