Global Simulations of the Atmosphere at 1.45 km Grid-Spacing with the Integrated Forecasting System

Journal Article**Author(s):**

Dueben, Peter D.; Wedi, Nils; Saarinen, Sarni; Zeman, Christian

Publication date:

2020-06-20

Permanent link:

<https://doi.org/10.3929/ethz-b-000425651>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Journal of the Meteorological Society of Japan 98(3), <https://doi.org/10.2151/jmsj.2020-016>

Funding acknowledgement:

823988 - INFRAEDI-02-2018 | Excellence in Simulation of Weather and Climate in Europe, Phase 2 (EC)

Global Simulations of the Atmosphere at 1.45 km Grid-Spacing with the Integrated Forecasting System

Peter D. DUEBEN, Nils WEDI, Sami SAARINEN

European Centre for Medium-Range Weather Forecasts, UK

and

Christian ZEMAN

Eidgenössische Technische Hochschule Zürich, Switzerland

(Manuscript received 11 July 2019, in final form 11 February 2020)

Abstract

Global simulations with 1.45 km grid spacing are presented that were performed using 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 is presented, including recent date initial conditions, real-world topography, and state-of-the-art physical parametrizations, as well as diabatic forcing including shallow convection, turbulent diffusion, radiation and five categories for the water substance (vapor, liquid, ice, rain, and snow). Simulations are evaluated with regard to computational efficiency and model fidelity. Scaling results are presented, which were performed on the fastest supercomputer in Europe, Piz Daint (Top 500, November 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 discretization 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

Citation Dueben, P. D., N. Wedi, S. Saarinen, and C. Zeman, 2020: Global simulations of the atmosphere at 1.45 km grid-spacing with the Integrated Forecasting System. *J. Meteor. Soc. Japan*, **98**, 551–572, doi:10.2151/jmsj.2020-016.

1. Introduction

The complexity and quality of weather and climate models have 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 the performance of individual processors

Corresponding author: Peter D. Dueben, European Centre for Medium-Range Weather Forecasts, Shinfield Rd, Reading RG2 9AX, UK
E-mail: peter.dueben@ecmwf.int
J-stage Advance Published Date: 17 March 2020

is now inducing challenges for the domain of weather and climate modeling. It is getting more complicated to efficiently use 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 modeling 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 explicitly resolve all features of the Earth System, several sub grid features need to be parametrized. This involves a description of the statistical contributions of subgrid-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, parametrizations describe diabatic effects, such as radiation and water phase changes, as well as processes for which equations that describe the underlying physical behavior are unknown, such as soil processes. As resolution is increased, features of the Earth System such as deep convection can be explicitly represented on the computational grid of the model simulation.

Deep convection plays a significant 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 nine 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 kilometers, they are entering the so-called grey-zone of convection, where certain limiting assumptions underlying deep convection parametrization—that deep convection can be represented as a bulk parametrization scheme based on an ensemble of independent convective updrafts within a grid cell—cannot be justified anymore – if grid cells

partially or fully represent a single updraft. If the deep convection parametrization scheme is switched off, convection is explicitly simulated by the governing equations. However, convective cells will be significantly bigger when compared with 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 several of the chosen grid sizes, and unrealistically sized convective cells may cause a degradation of forecast skill in comparison with coarser simulations, where convection parametrization is used. Global simulations in which deep convection parametrization is 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 $O(1)$ km that would potentially allow sufficient resolution of deep convection 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 or 2.5 km. As discussed by 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 mesoscale eddies when compared to today’s models and can simulate ocean tides explicitly. Meso-scale eddies are essential to represent ocean variability accurately.

Until 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 h long simulation at 870 m grid spacing with the Nonhydrostatic ICosahedral Atmospheric Model (NICAM) (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 idealized 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; Müller et al. 2018). An overview of the history of global storm-resolving models can be found in the paper by 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° 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 (RCMIP) that is comparing global model simulations for an idealization of the climate system to have sufficient understanding of 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 with 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 Orlanski (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 ranging 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 the sea for grid spacings ranging from 16 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 presented by 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 exhibit the characteristic hydrostatic (non-trapped) behavior.

There is no consensus in the literature which spatial discretization 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 kilometers on state-of-the-art supercomputers (Michalakes et al. 2015; Stevens et al. 2019). There is also no consensus on whether explicit, semi-implicit or fully implicit timestepping schemes are most promising for use in global storm-resolving simulations. Implicit schemes allow to use a larger timestep in comparison with 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 because 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 timestep of 6 s when running with a grid spacing of less than a kilometer. In contrast, Yang et al. (2016) propose to work with implicit schemes and show results using a large timestep of 240 s when running at a grid spacing of less than a kilometer to achieve the best time-to-solution for simulations. As pointed out by Wedi et al. (2015) and evident in the paper by Mengaldo et al. (2018), different timestepping approaches may incur low-order time truncation errors compared with 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 Integrated Forecasting System (IFS) with up to 1.45 km grid spacing. The performance of the IFS is discussed, and the 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 parametrization of convection, or the timestep length influence model simulations.

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

2. A description of the IFS

We perform model simulations using the uncoupled IFS atmosphere model cycle 45r2 (no ocean, sea ice, or wave model, because 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 discretization, 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; Malardel et al. 2016).

To transform between grid-point and spectral spaces 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, which shows a computational complexity proportional to N^3 with the truncation wavenumber N , a so-called “fast Legendre transformation” was introduced that is trading performance against accuracy and achieving a scaling behavior 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 non-hydrostatic dynamical core based on a finite-volume discretization with the same collocation of prognostic variables as in the current IFS (Integrated Forecasting System–Finite-Volume Module (IFS-FVM); Kühnlein 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 timestepping scheme with no decentering that allows for the use of long timesteps. We are using the same timestep 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 initialized from the 9 km operational analysis of ECMWF on October 13, 2016, 0 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, physical parametrization schemes (“physics”) and the semi-Lagrangian advection scheme are the largest contributors to the 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

using 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 (Düben and Palmer 2014; 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 making 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 with hydrostatic simulations but similarly solved in spectral space (see e.g., 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) timestepping 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 in the computational cost of the non-hydrostatic model (for details, see Wedi et al. 2009, 2013). The second difference is the use of a finite-element discretization 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 by Bubnova et al. (1995), the vertical discretization 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 by 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 subgrid-scale topography within the high-resolution simulations at 1.45 km grid spacing.

We use the standard set of physical parametrization schemes of the IFS for all forecasts presented in this paper, cloud microphysics with five categories for water substance (vapor, liquid, ice, rain, and 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 parametrization of deep convection is switched off for simulations at 1.45 km grid spacing. However, the parametrization of shallow convection remains active in all simulations.

Figure 1 presents 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 present 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 optimize simulations using the IFS for present and future high performance computers as part of its scalability program (see Bauer et al. 2020, or the paper by Müller et al. (2019) for the Energy-Efficient Scalable Algorithms for Weather Prediction at Exascale (ESCAPE) project as examples).

We have performed model simulations using IFS on the two supercomputers of ECMWF and the Piz Daint

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 indicated by black contour.

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 GB 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.

Figure 2 is presents the cost distribution of the different model components for simulations using 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 Modeling of the Ocean (NEMO) model are switched off. The relative cost for spectral transforms are higher for the non-hydrostatic configuration, because additional transformations between spectral and grid-point spaces are required (Wedi et al. 2013). The hydrostatic simulation is using a finite-element discretization for the vertical and no predictor–corrector scheme (similar to H-FE-120DT in the next section and the operational setting at ECMWF), whereas the non-hydrostatic simulation is using a finite-difference discretization for the vertical direction and one iteration of the predictor–corrector scheme.

Figure 3 presents the scaling behavior 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 with non-hydrostatic simulations. Both model configurations exhibited a reasonable (strong) scaling behavior when using most of the available nodes on the supercomputer, in

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.

particular, for the non-hydrostatic case. However, the data is limited, and the comparison of hydrostatic and non-hydrostatic configurations indicates that the run with significantly shorter elapsed time per timestep 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. Allowing 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 models coupled).

While the performance results of this paper are promising, it should be noted that simulations using

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.

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 discretization method, the length of the timestep, and the number of predictor–corrector (PC) iterations.

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

the IFS with a shorter timestep 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) conclude that there is a shortfall factor of 101× for the COSMO model and a shortfall factor of 247× for the non-hydrostatic IFS model with a projected 30 s timestep to reach global simulations at 1 km resolution with a throughput 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 penalizes IFS for using a larger timestep 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 with 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).

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 described in Table 2.

Unfortunately, we could not run the non-hydrostatic simulation with 30 s 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 remove these instabilities (Voitus et al. 2019).

The section will present results for the 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 with real satellite data (Section 4.4) and preliminary results for forecast errors of high-resolution simulations (Section 4.5).

4.1 Energy spectra

To get a first impression of model fidelity for the different model simulations, we have plotted the spectra for horizontal kinetic energy in Fig. 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 h 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 s timestep for the same model configuration. The energy level is slightly higher for notopo-NH-FD-DT30 in comparison with 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 a comparison that

Fig. 4. Spectra for horizontal kinetic energy for simulations using 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, $\Delta t = 450$ s, 0 PC). The vertical black lines indicate the grid spacing of 5 km and 10 km, respectively.

clearly fails to transition between the -3 and $-5/3$ scaling behaviors.

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 with the expected scaling, is between 5 km 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 seven to eight times or six times the grid spacing, respectively.

To make differences between the simulations more visible, we plot the horizontal kinetic energy spectra with a compensation for the $-5/3$ scaling in Fig. 5. IFS shows more deviations from the theoretical $-5/3$ curve at 200 hPa when compared with the results at 500 hPa. More recent comparisons with 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 Fig. 4, the NH-FD-DT60 exhibits spurious behavior 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 with the rotational part even for high wavenumbers. Notably, the vertical finite-element discretization is of higher order than the first- to second- order finite-difference discretization.

4.2 Vertical velocity

For further insight how the different model configurations represent vertical motions (and potentially convection), we have also plotted variance spectra of vertical velocity in Fig. 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 Fig. 5, the two non-hydrostatic simulations (NH-FD-DT60 and notopo-NH-DT30) show a spurious increase in variance for small scales. There are clear differences visible between the simulations H-FE-DT120 and H-FE-DT60 indicating that the dynamics are not yet converged with the timestep. However, differences when changing the vertical resolution and the vertical discretization from finite element (H-FE-DT60) to finite difference (H-FD-DT60) are even larger.

Figure 7 presents two-dimensional plots of vertical velocity in the tropics. H-FE-DT120 and H-FE-DT60 are showing larger-scale structures when compared with the other simulations but stronger convective regions. The simulations without topography are less active (notopo-H-FD-DT30 and notopo-NH-FD-

Fig. 5. Spectra of horizontal kinetic energy for simulations using the IFS at 1.45 km grid spacing 12 h into the forecast at 200 hPa (left) and 500 hPa (right) for the six model configurations. The plots show the total and 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 hPa (left) and 500 hPa (right) 12 h into the forecast.

Fig. 7. Vertical velocity [m s^{-1}] at 250 hPa 12 h into the forecast for a tropical area around Indonesia (7S/120W/3N/140W). Please note the nonlinear color scale.

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

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 h into the forecast are available.

Run	Height	$> 10 \text{ m s}^{-1}$	$> 20 \text{ m s}^{-1}$	$> 30 \text{ m s}^{-1}$
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

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 Fig. 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 indicating that the difference between hydrostatic and non-hydrostatic simulations is smaller when compared with other changes in the model configuration.

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 with regional simulations.

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^{-1} 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 discretization and higher resolution in the vertical direction (H-FE-DT60 and H-FE-DT120) show the highest up-ward velocities whereas 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 discretization and hydrostatic equations (H-FD-DT60) is showing the smallest vertical velocities. The signal is qualitatively consistent if considered after 12 h and 24 h (see Table 3 for numbers for a subset of runs).

4.3 Precipitation

The shape and distribution of precipitation should change significantly as grid spacing is reduced from 9 to 1.45 km. Figure 9 presents the PDFs of total precipitation for different simulations. For the simulation with 9 km grid spacing and parametrized convection, the number of grid points with heavy precipitation is significantly reduced indicating an ability to improve the representation of local precipitation when resolution is increased. The simulation with 120 s 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 with the hydrostatic simulations. However, differences are of the same order of magnitude as the other changes of the model configuration, such as the timestep or vertical discretization.

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 whereas convective precipitation is motivated

Fig. 9. PDFs of total precipitation integrated for the first 12 h for the simulations at 1.45 km grid spacing and a simulation with 9 km grid spacing (H, FE, 62 vertical levels, $\Delta t = 450$ s, 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 grid points being in the first bin of less than 0.1 mm precipitation.

by convective updrafts within the grid columns that are not represented explicitly if parametrization for convection is switched on. For storm-resolving simulations at 1.45 km grid spacing, the parametrization of deep convection is switched off while the parametrization 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

Table 4. Global average for large-scale precipitation, parametrized convective precipitation, and total precipitation integrated over the first and the second 12 h 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 h	1.3197	0.2779	1.5976
notopo-NH-FD-DT30 – 12 h	1.3107	0.2672	1.5780

within the first 12 h. The results in 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 in 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 with the simulation with 9 km grid spacing. Performing the same evaluation after 24 h does not change the conclusion (not all runs were simulated for the full 24 h). 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 parametrized deep convection should be performed for future publications.

4.4 Satellite simulators

Figures 10, 11 and 12 present 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 with the satellite data. It is evident that the higher resolution is beneficial for the representation of clouds with explicit cellular organization 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 with the simulations at higher resolution (see bottom left of Fig. 11). This indicates that the simulations at high resolution may require changes to the parametrization schemes, including changes to the parametrization 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 discretization in the vertical direction (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 are again rather small in comparison.

4.5 Forecast errors

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 the

results against the operational forecast configuration. The forecast error was calculated on a TCo639 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 whereas 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 behavior 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 timestep, number of iterations in the predictor–corrector scheme, and equations.

The non-hydrostatic simulations are showing some spurious behaviors for energy spectra (Fig. 5) and vertical velocity (Fig. 7). The hydrostatic simulation that was using a timestep of 120 s did not show a spurious behavior. However, the results are also different between the H-FE-DT120 and H-FE-DT60 simulation and this indicates that a timestep 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 Fig. 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 behavior of the spectra of the non-hydrostatic simulations at small scales (Fig. 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 the results in comparison with the other participating models—that were all non-hydrostatic—even at lead

Fig. 10. Simulated and observed top-of-the-atmosphere brightness temperatures derived from satellites and satellite simulators for October 16, 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 the results for the simulated satellite radiances of the operational weather forecast at ECMWF at 9 km grid spacing with parametrized deep convection and 137 vertical levels. The other plots show the results of the model simulations with 1.45 km grid spacing.

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

times of up to 40 days (Stevens et al. 2019). Since the hydrostatic simulations using the spectral IFS model are much cheaper when compared using non-hydrostatic simulations, we consider the hydrostatic

configuration to be a promising candidate for O(1 km) global model simulations at ECMWF. Moreover, an ensemble of H-FE-DT120 simulations with a much larger number of ensemble members may provide

Fig. 12. Same as Fig. 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, $\Delta t = 450$ s, 0 PC, coupled to NEMO and the wave model).

better forecast scores in comparison with 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 with the operational resolution for deterministic forecasts and that both parametrization 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 indicated 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 using the IFS may still require significant work before improvements in forecast scores can be realized 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 initialized from a lower-resolution analysis which leaves many degrees of freedom uninitialized 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 (Fig. 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 than 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 kilometers when comparing spectra at 9 km and 1.45 km grid spacing, and this structural difference is also observed in experiments that assess the impact of physical parametrization on energy spectra and on nonlinear spectral energy fluxes (Malardel et al. 2016).

6. Conclusion

In this paper, we document simulations using 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. The 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 skepticism of the community regarding the usefulness of spectral models for simulations at high resolution due to the bad scaling behavior 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 with other global models at slightly lower resolution (Michalakes et al. 2015), we argue that spectral discretization 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 is passed if 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 in the color scale in Fig. 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 modeling 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, e.g., Lawrence et al. 2018; Neumann et al. 2019; Schulthess et al. 2019; Schäer et al. 2019; Biercamp et al. 2019). While the results of this paper confirm that these simulations could be within reach soon, there is no question that it will require a large concerted European (or global) effort between modeling and supercomputing centers 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 modeling affordable and environmentally acceptable.

Acknowledgments

Many thanks to Cristina Lupu 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 Maria-Grazia Giuffreda 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 Union's Horizon 2020 research and innovation programme under grant agreement No 67162 and 800 987, respectively.

References

- Abdalla, S., L. Isaksen, P. A. E. M. Janssen, and N. Wedi, 2013: Effective spectral resolution of ECMWF atmospheric forecast models. *ECMWF Newsl.*, **137**, 19–22.
- Balaji, V., 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, 2018: Requirements for a global data infrastructure in support of CMIP6. *Geosci. Model Dev.*, **11**, 3659–3680.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55.
- Bauer, P., T. Quintino, N. Wedi, A. Bonanni, M. Chrust, W. Deconinck, M. Diamantakis, P. Düben, S. English, J. Flemming, P. Gillies, I. Hadade, J. Hawkes, M. Hawkins, O. Iffrig, C. Kuehnlein, M. A. Mueller, S. Saarinen, D. Sarmany, M. Sleigh, S. Smart, P. Smolarkiewicz, D. Thiemert, G. Tumolo, C. Wehrauch, and C. Zanna, 2020: *The ECMWF scalability programme: Progress and plan*. ECMWF Technical Memoranda, No.857, 110 pp, doi:10.21957/gdit22ulm.
- Biercamp, J., P. Bauer, P. Düben, and B. Lawrence, 2019: A roadmap to the implementation of 1km earth system model ensembles, D1.2. *Centre of Excellence in Simulation of Weather and Climate in Europe, ESiWACE*, doi:10.5281/zenodo.3361229.
- Bretherton, C. S., and M. F. Khairoutdinov, 2015: Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. *J. Adv. Model. Earth Syst.*, **7**, 1765–1787.
- Bubnová, R., G. Hello, P. Bénard, and J.-F. Geleyn, 1995: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Mon. Wea. Rev.*, **123**, 515–535.
- Daley, R., 1988: The normal modes of the spherical non-hydrostatic equations with applications to the filtering of acoustic modes. *Tellus A*, **40**, 96–106.
- Düben, P. D., and T. N. Palmer, 2014: Benchmark tests for numerical forecasts on inexact hardware. *Mon. Wea. Rev.*, **142**, 3809–3829.
- Dudhia, J., 1993: A nonhydrostatic version of the Penn State—NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.
- Duran, I. B., J.-F. Geleyn, F. Vana, J. Schmidli, and R. Brozkova, 2018: A turbulence scheme with two prognostic turbulence energies. *J. Atmos. Sci.*, **75**, 3381–3401.
- Fudeyasu, H., Y. Wang, M. Satoh, T. Nasuno, H. Miura, and W. Yanase, 2008: Global cloud-system-resolving model NICAM successfully simulated the lifecycles of two real tropical cyclones. *Geophys. Res. Lett.*, **35**, L22808, doi:10.1029/2008GL036003.
- Fuhrer, O., T. Chadha, T. Hoefler, G. Kwasniewski, X. Lapillonne, D. Leutwyler, D. Lüthi, C. Osuna, C. Schär, T. C. Schulthess, and H. Vogt, 2018: Near-global climate simulation at 1 km resolution: Establishing a performance baseline on 4888 GPUs with COSMO 5.0. *Geosci. Model Dev.*, **11**, 1665–1681.
- Hatfield, S., M. Chantry, P. Düben, and T. Palmer, 2019: Accelerating high-resolution weather models with deep-learning hardware. *Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '19)*, June 12–14, 2019, Zurich, Switzerland, Association for Computing Machinery, New York, NY, USA, 11 pp.
- Heinze, R., A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, C. Barthlott, A. Behrendt, U. Blahak, S. Bley, S. Brdar, M. Brueck, S. Crewell, H. Deneke, P. Di Girolamo, R. Evaristo, J. Fischer, C. Frank, P. Friederichs, T. Göcke, K. Gorges, L. Hande, M. Hanke, A. Hansen, H.-C. Hege, C. Hoose, T. Jahns, N. Kalthoff, D. Klocke, S. Kneifel, P. Knippertz, A. Kuhn, T. van Laar, A. Macke, V. Maurer, B. Mayer, C. I. Meyer, S. K. Muppa, R. A. J. Neggers, E. Orlandi, F. Pantillon, B. Pospichal, N. Röber, L. Scheck, A. Seifert, P. Seifert, F. Senf, P. Siligam, C. Simmer, S. Steinke, B. Stevens, K. Wapler, M. Weniger, V. Wulfmeyer, G. Zängl, D. Zhang, and J. Quaas, 2017: Large-eddy simulations over germany using ICON: A comprehensive evaluation. *Quart. J. Roy. Meteor. Soc.*, **143**, 69–100.
- Hocking, J., P. Rayer, D. Rundle, R. Saunders, M. Matricardi, A. Geer, P. Brunel, and J. Vidot, 2013: *RTTOV v11 users guide*. NWP SAF report, Met Office, 107 pp.
- Jeevanjee, N., 2017: Vertical velocity in the gray zone. *J. Adv. Model. Earth Syst.*, **9**, 2304–2316.
- Kato, T., 1997: Hydrostatic and non-hydrostatic simulations of moist convection: Review and further study. *Meteor. Atmos. Phys.*, **63**, 39–51.
- Keller, T. L., 1994: Implications of the hydrostatic assumption on atmospheric gravity waves. *J. Atmos. Sci.*, **51**, 1915–1929.
- Kuehnlein, C., W. Deconinck, R. Klein, S. Malardel, Z. P. Piotrowski, P. K. Smolarkiewicz, J. Szmelter, and N. P. Wedi, 2019: FVM 1.0: A nonhydrostatic finite-volume dynamical core for the IFS. *Geosci. Model Dev.*, **12**, 651–676.
- Lawrence, B. N., M. Rezný, 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, 2018: Crossing the chasm: How to develop weather and climate models for next generation computers? *Geosci. Model Dev.*, **11**, 1799–1821.
- Leutwyler, D., D. Lüthi, N. Ban, O. Fuhrer, and C. Schär, 2017: Evaluation of the convection-resolving climate modeling approach on continental scales. *J. Geophys. Res.: Atmos.*, **122**, 5237–5258.
- Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kuehnlein, G. Mozdzynski, M. Hamrud, and P. Smolarkiewicz, 2016: A new grid for the IFS. *ECMWF Newsl.*, **146**, 23–28.
- Mengaldo, G., A. Wyszogrodski, M. Diamantakis, S.-J. Lock, F. X. Giraldo, and N. P. Wedi, 2018: Current and emerging time-integration strategies in global numerical weather and climate prediction. *Arch. Com-*

- put. Methods Eng.*, **26**, 663–684.
- Michalakes, J., M. Govett, R. Benson, T. Black, H. Juang, A. Reinecke, and B. Skamarock, 2015: *AVEC report: NGGPS level-1 benchmarks and software evaluation*. Advanced Computing Evaluation Committee (AVEC), 22 pp. [Available at <https://repository.library.noaa.gov/view/noaa/18654>.]
- Miura, H., M. Satoh, T. Nasuno, A. T. Noda, and K. Oouchi, 2007: A Madden-Julian oscillation event realistically simulated by a global cloud-resolving model. *Science*, **318**, 1763–1765.
- Miyamoto, Y., Y. Kajikawa, R. Yoshida, T. Yamaura, H. Yashiro, and H. Tomita, 2013: Deep moist atmospheric convection in a subkilometer global simulation. *Geophys. Res. Lett.*, **40**, 4922–4926.
- Müller, A., M. A. Kopera, S. Marras, L. C. Wilcox, T. Isaac, and F. X. Giraldo, 2018: Strong scaling for numerical weather prediction at petascale with the atmospheric model NUMA. *Int. J. High Perform. Comput. Appl.*, **33**, 411–426.
- Müller, A., 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. Ginton, 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. Sychala, 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, 2019: The ESCAPE project: Energy-efficient Scalable Algorithms for Weather Prediction at Exascale. *Geosci. Model Dev.*, **12**, 4425–4441.
- Neumann, P., P. D. Düben, P. Adamidis, P. Bauer, M. Brueck, L. Kornbluh, D. Klocke, B. Stevens, N. Wedi, and J. Biercamp, 2018: Assessing the scales in numerical weather and climate predictions: Will exascale be the rescue? *Philos. Trans. Roy. Soc. A*, **377**, 20180148, doi:10.1098/rsta.2018.0148.
- Orlanski, I., 1981: The quasi-hydrostatic approximation. *J. Atmos. Sci.*, **38**, 572–582.
- Palmer, T. N., 2016: A personal perspective on modelling the climate system. *Proc. Roy. Soc. A: Math. Phys. Eng. Sci.*, **472**, 20150772, doi:10.1098/rspa.2015.0772.
- Ross, B. B., and I. Orlanski, 1978: The circulation associated with a cold front. Part II: Moist case. *J. Atmos. Sci.*, **35**, 445–465.
- Satoh, M., T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga, 2008: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *J. Comput. Phys.*, **227**, 3486–3514.
- Satoh, M., B. Stevens, F. Judt, M. Khairoutdinov, S.-J. Lin, W. M. Putman, and P. Düben, 2019: Global cloud-resolving models. *Curr. Climate Change Rep.*, **5**, 172–184.
- Schär, C., O. Fuhrer, A. Arteaga, N. Ban, C. Charpiloz, S. Di Girolamo, L. Hentgen, T. Hoefler, X. Lapillonne, D. Leutwyler, K. Osterried, D. Panosetti, S. Rüdisühli, L. Schlemmer, T. Schulthess, M. Sprenger, S. Ubbiali, and H. Wernli, 2020: Kilometer-scale climate models: Prospects and challenges. *Bull. Amer. Meteor. Soc.*, **101**, E567–587.
- Schulthess, T. C., P. Bauer, N. Wedi, O. Fuhrer, T. Hoefler, and C. Schär, 2019: Reflecting on the goal and baseline for exascale computing: A roadmap based on weather and climate simulations. *Comput. Sci. Eng.*, **21**, 30–41.
- Skamarock, W. C., S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *J. Atmos. Sci.*, **71**, 4369–4381.
- Stevens, B., M. Satoh, L. Auger, J. Biercamp, C. Bretherton, X. Chen, P. Düben, F. Judt, M. Khairoutdinov, D. Klocke, C. Kodama, L. Kornbluh, S.-J. Lin, P. Neumann, W. M. Putman, N. Röber, R. Shibuya, B. Vannier, P. L. Vidale, N. Wedi, and L. Zhou, 2019: DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.*, **6**, 61, doi:10.1186/s40645-019-0304-z.
- Termonia, P., C. Fischer, E. Bazile, F. Bouyssel, R. Brožková, P. Bénard, B. Bochenek, D. Degrauwe, M. Derkova, 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, 2018: The ALADIN system and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geosci. Model Dev.*, **11**, 257–281.
- Tort, T., and T. Dubos, 2014: Dynamically consistent shallow-atmosphere equations with a complete Coriolis force. *Quart. J. Roy. Meteor. Soc.*, **140**, 2388–2392.
- Vana, F., P. Düben, S. Lang, T. Palmer, M. Leutbecher, D. Salmond, and G. Carver, 2017: Single precision in weather forecasting models: An evaluation with the IFS. *Mon. Wea. Rev.*, **145**, 495–502.
- Voitus, F., P. Bénard, C. Kühnlein, and N. P. Wedi, 2019: Semi-implicit integration of the unified equations in a mass-based coordinate: Model formulation and numerical testing. *Quart. J. Roy. Meteor. Soc.*, **145**, 3397–3408.
- Wedi, N. P., 2014: Increasing horizontal resolution in numerical weather prediction and climate simulations: Illusion or panacea? *Philos. Trans. Roy. Soc. London A: Math. Phys. Eng. Sci.*, **372**, 2018, doi:10.1098/rsta.2013.0289.
- Wedi, N. P., and P. K. Smolarkiewicz, 2009: A framework for testing global non-hydrostatic models. *Quart. J.*

- Roy. Meteor. Soc.*, **135**, 469–484.
- Wedi, N. P., K. Yessad, and A. Untch, 2009: *The nonhydrostatic global IFS/ARPEGE model: Model formulation and testing*. ECMWF Technical Memoranda, No.594, European Center for Medium-Range Weather Forecasts, Reading, UK, 36 pp.
- Wedi, N. P., M. Hamrud, and G. Mozdzynski, 2013: A fast spherical harmonics transform for global NWP and climate models. *Mon. Wea. Rev.*, **141**, 3450–3461.
- Wedi, N. P., P. Bauer, W. Deconinck, M. Diamantakis, M. Hamrud, C. Kühnlein, S. Malardel, K. Mogensen, G. Mozdzynski, and P. K. Smolarkiewicz, 2015: *The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges*. ECMWF Technical Memoranda, No.760, European Center for Medium-Range Weather Forecasts, Reading, UK, 18 pp.
- Weisman, M. L., W. C. Skamarock, and J. B. Klemm, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548.
- Wing, A. A., K. A. Reed, M. Satoh, B. Stevens, S. Bony, and T. Ohno, 2018: Radiative-convective equilibrium model intercomparison project. *Geosci. Model Dev.*, **11**, 793–813.
- Yang, C., W. Xue, H. Fu, H. You, X. Wang, Y. Ao, F. Liu, L. Gan, P. Xu, L. Wang, G. Yang, and W. Zheng, 2016: 10m-core scalable fully-implicit solver for nonhydrostatic atmospheric dynamics. *SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Salt Lake City, UT, USA, IEEE, 57–68.