nextGEMS: entering the era of kilometer-scale Earth system modeling

Hans Segura[1,*], Xabier Pedruzo-Bagazgoitia[2,*], Philipp Weiss[3,*], Sebastian K. Müller[4,*], Thomas Rackow[2,*], Junhong Lee[1,*], Edgar Dolores-Tesillos[5,*], Imme Benedict[6,*], Matthias Aengenheyster[2], Razvan Aguridan[2,7], Gabriele Arduini[2], Alexander J. Baker[8], Jiawei Bao[9], Swantje Bastin[1], Eulàlia Baulenas[7], Tobias Becker[2], Sebastian Beyer[10], Hendryk Bockelmann[11], Nils Brüggemann[1], Lukas Brunner[12], Suvarchal K. Cheedela[10], Sushant Das[13], Jasper Denissen[2], Ian Dragaud[12], Piotr Dziekan[14], Madeleine Ekblom[15], Jan Frederik Engels[11], Monika Esch[1], Richard Forbes[2], Claudia Frauen[11], Lilli Freischem[3], Diego García-Maroto[16], Philipp Geier[2], Paul Gierz[10], Álvaro González-Cervera[16], Katherine Grayson[7], Matthew Griffith[2], Oliver Gutjahr[1], Helmuth Haak[1], Ioan Hadade[2], Kerstin Haslehner[17], Shabeh ul Hasson[18], Jan Hegewald[19], Lukas Kluft[1], Aleksei Koldunov[10], Nikolay Koldunov[10], Tobias Kölling[1], Shunya Koseki[20], Sergey Kosukhin[1], Josh Kousal[2], Peter Kuma[21], Arjun U. Kumar[1], Rumeng Li[22], Nicolas Maury[23], Maximilian Meindl[17], Sebastian Milinski[2], Kristian Mogensen[2], Bimochan Niraula[11], Jakub Nowak[14], Divya Sri Praturi[1], Ulrike Proske[24], Dian Putrasahan[1], René Redler[1], David Santuy[16], Domokos Sármány[2], Reiner Schnur[1], Patrick Scholz[10], Dmitry Sidorenko[10], Dorian Spät[17], Birgit Sützl[2], Daisuke Takasuka[25], Adrian Tompkins[26], Alejandro Uribe[21], Mirco Valentini[2], Menno Veerman[6], Aiko Voigt[17], Sarah Warnau[6,27], Fabian Wachsmann[11], Marta Wacławczyk[14], Nils Wedi[2], Karl-Hermann Wieners[1], Jonathan Wille[28], Marius Winkler[1], Yuting Wu[1], Florian Ziemen[11], Janos Zimmermann[11], Frida A.-M. Bender[21], Dragana Bojovic[7], Sandrine Bony[23], Simona Bordoni[4], Patrice Brehmer[29], Marcus Dengler[30], Emanuel Dutra[31], Saliou Faye[32], Erich Fischer[28], Chiel van Heerwaarden[6], Cathy Hohenegger[1], Heikki Järvinen[15], Markus Jochum[33], Thomas Jung[10], Johann H. Jungclaus[1], Noel S. Keenlyside[20], Daniel Klocke[1], Heike Konow[1], Martina Klose[22], Szymon Malinowski[14], Olivia Martius[5], Thorsten Mauritsen[21], Juan Pedro Mellado[12], Theresa Mieslinger[1], Elsa Mohino[16], Hanna Pawłowska[14], Karsten Peters-von Gehlen[11], Abdoulaye Sarré[32], Pajam Sobhani[34], Philip Stier[3], Lauri Tuppi[15], Pier Luigi Vidale[8], Irina Sandu[2], and Bjorn Stevens[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]European Centre for Medium-Range Weather Forecasts, Bonn, Germany
[3]Department of Physics, University of Oxford, Oxford, United Kingdom
[4]University of Trento, Trento, Italy
[5]Institute of Geography, University of Bern, Bern, Switzerland
[6]Meteorology and Air Quality Group, Wageningen University, Wageningen, the Netherlands
[7]Barcelona Supercomputing Center, Barcelona, Spain
[8]National Centre for Atmospheric Science and Department of Meteorology, University of Reading, Reading, United Kingdom
[9]Institute of Science and Technology Austria, Klosterneuburg, Austria
[10]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[11]German Climate Computing Center, Hamburg, Germany
[12]Department of Earth System Sciences, University of Hamburg, Hamburg, Germany
[13]Department of Earth and Atmospheric Sciences, National Institute of Technology Rourkela, Rourkela, India
[14]Institute of Geophysics, Faculty of Physics, University of Warsaw, Warsaw, Poland

[15]Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, Helsinki, Finland
[16]Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid, Madrid, Spain
[17]Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria
[18]HAREME Lab, Earth and Society Research Hub (ESRAH), University of Hamburg, Hamburg, Germany
[19]Gauß-IT-Zentrum, Braunschweig University of Technology (GITZ), Braunschweig, Germany
[20]Geophysical Institute, University of Bergen, Bergen, Norway
[21]Department of Meteorology, Stockholm University, Stockholm, Sweden
[22]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
[23]LMD/IPSL, CNRS, Université Pierre et Marie Curie, Paris, France
[24]Hydrology and Environmental Hydraulics Group, Wageningen University, Wageningen, the Netherlands
[25]Department of Geophysics, Tohoku University, Sendai, Japan
[26]Abdus Salam International Centre for Theoretical Physics, Trieste, Italy
[27]Wetsus, European Centre of Excellence for Sustainable Water Technology, Leeuwarden, the Netherlands
[28]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[29]Institut de Recherche pour le Développement, Dakar, Senegal
[30]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[31]Instituto Português do Mar e da Atmosfera, Lisbon, Portugal
[32]Institut Sénégalais de Recherches Agricoles, Dakar, Senegal
[33]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[34]Latest Thinking GmbH, Hamburg, Germany
[*]These authors contributed equally to this work.

Correspondence: Hans Segura (hans.segura@mpimet.mpg.de), Xabier Pedruzo-Bagazgoitia (xabier.pedruzo@ecmwf.int),
Philipp Weiss (philipp.weiss@physics.ox.ac.uk), Sebastian K. Müller (sebastian.mueller@unitn.it), Thomas Rackow
(thomas.rackow@ecmwf.int)

Abstract. The Next Generation of Earth Modeling Systems (nextGEMS) project aimed to produce multi-decadal climate simulations, for the first time, with resolved kilometer-scale (km-scale) processes in the ocean, land, and atmosphere. In only three years, nextGEMS achieved this milestone with the two km-scale Earth system models, ICOsahedral Non-hydrostatic model (ICON) and Integrated Forecasting System coupled to the Finite-volumE Sea ice-Ocean Model (IFS-FESOM). nextGEMS was

5   based on three cornerstones: 1) developing km-scale Earth system models with small errors in the energy and water balance, 2) performing km-scale climate simulations with a throughput greater than one simulated year per day, and 3) facilitating new workflows for an efficient analysis of the large simulations with common data structures and output variables. These cornerstones shaped the timeline of nextGEMS, divided into four cycles. Each cycle marked the release of a new configuration of ICON and IFS-FESOM, which were evaluated at hackathons. The participants in hackathons included experts from climate

10   science, software engineering, and high-performance computing, as well as users from the energy and agricultural sectors. The continuous efforts over the four cycles allowed us to produce 30-year simulations of ICON and IFS-FESOM, spanning the period 2020–2049 under the SSP3-7.0 scenario. The throughput was about 500 simulated days per day on the Levante supercomputer of the German Climate Computing Center (DKRZ). The simulations employed a horizontal grid of about 5 km resolution in the ocean and 10 km resolution in the atmosphere and land. Aside from this technical achievement, the simula-

15   tions allowed us to gain new insights into the realism of ICON and IFS-FESOM. Beyond its timeframe, nextGEMS builds the

foundation of the Climate Change Adaptation Digital Twin developed in the Destination Earth initiative and paves the way for future European research on climate change.

1 Introduction

The advent of exascale supercomputers and progress in numerical modeling opens the door to a new way of simulating our
20   Earth system (Schär et al., 2020; Slingo et al., 2022). Several international initiatives aim to resolve kilometer-scale (km-scale) processes explicitly using horizontal grid spacings equal to or less than 10 km globally in the atmosphere (e.g., Satoh et al., 2008), land (e.g., Kollet and Maxwell, 2006), and ocean (e.g., Maltrud and McClean, 2005). Resolving km-scale processes explicitly makes it possible to represent more accurately the horizontal and vertical transfer of mass and energy and the circulation that it entails. In practice, it brings climate simulations to a level of granularity that has long been proven necessary
25   to understand regional climate impact and support climate adaptation and mitigation.

Deep convective motions in the atmosphere, for example, redistribute moisture and energy, influencing the tropical vertical temperature profile and the global hydrological cycle (Kuang, 2010; Prein et al., 2017; Tian and Kuang, 2019; Bao et al., 2024). Km-scale atmospheric simulations, also referred to as storm-resolving simulations, resolve deep convection, capturing mesoscale convective systems (Peters et al., 2019; Prein et al., 2020; Becker et al., 2021), convectively coupled equatorial
30   waves (Miura et al., 2007; Holloway et al., 2013; Senf et al., 2018), and tropical cyclones (Gentry and Lackmann, 2010; Judt et al., 2021; Baker et al., 2024). Km-scale simulations also provide a more detailed representation of the land in terms of surface topography and heterogeneity and its impact on local-, regional-, and large-scale circulations (Sandu et al., 2019). Previous studies showed that such simulations improve the representation of atmospheric blockings (Woollings et al., 2018; Schiemann et al., 2020), weaken the soil moisture–precipitation feedback (Lee and Hohenegger, 2024), and impact the circulations gener-
35   ated by surface-radiation interactions such as land-sea breezes (Birch et al., 2015). Km-scale oceanic simulations, also referred to as eddy-resolving simulations, resolve mesoscale eddies (Hewitt et al., 2022), increasing the global kinetic energy (Chassignet et al., 2020) and vertical transport of heat (Griffies et al., 2015). Previous studies showed that the higher resolution also impacts the internal variability of the ocean (Penduff et al., 2010), the timing of Antarctic sea ice decrease, and the magnitude of the projected sea level rise (van Westen and Dijkstra, 2021; Rackow et al., 2022).
40   The early insights gained from storm- and eddy-resolving simulations were a strong motivation for the Next Generation of Earth Modelling Systems (nextGEMS) project funded by the European Horizon 2020 programme. nextGEMS aimed to build the next generation of km-scale Earth system models, namely the ICOsahedral Non-hydrostatic model (ICON; Hohenegger et al., 2023) and the Integrated Forecasting System (IFS; Rackow et al., 2025b). The latter can be run together with either the Finite-VolumE Sea ice–Ocean Model (FESOM; Scholz et al., 2019), used in this project, or the Nucleus for European
45   Modelling of the Ocean model (NEMO; Madec and the NEMO System Team, 2023).

While ICON and IFS-FESOM both resolve km-scale processes, their history and philosophy differ. ICON relies on a simple framework and aims to minimize the use of parameterization schemes. In the atmosphere, it parameterizes only small-scale processes related to turbulence, cloud microphysics, and radiation and does not employ any scheme for convection (Hoheneg-

ger et al., 2023). IFS-FESOM, on the other hand, operates as a weather model and incorporates years of model tuning to obtain
50   accurate forecasts. In the atmosphere, it employs sophisticated parameterization schemes for various processes including con-
vection (ECMWF, 2023b; Rackow et al., 2025b). The two approaches of ICON and IFS-FESOM are complementary and allow
us to examine whether explicitly resolving km-scale processes is enough to capture the main features of our climate and to
what extent the simulation quality depends on the ability to fine-tune the remaining small scales.

nextGEMS had the visionary goal to produce and analyze, for the first time, multi-decadal km-scale climate simulations
55   with coupled atmosphere, land, and ocean under a Shared Socioeconomic Pathway (SSP) scenario. To facilitate the model
development, nextGEMS explored new ways to foster the collaboration of project participants with expertise in software
engineering, model development, and climate physics. The timeline of nextGEMS was divided into four cycles. Each cycle
marked the release of new simulations with ICON and IFS-FESOM, which were prepared for and analyzed at large hackathons
with over 100 participants. Two main problems had to be solved to produce multi-decadal km-scale simulations. First, we had
60   to improve the computational throughput and simplify the analysis of simulation data. Second, we had to achieve simulations
with an energetically consistent climate, i.e., a near-stationary climate with a top-of-atmosphere (TOA) energy balance close
to zero with global conservation of mass and energy and no near-surface temperature drift.

nextGEMS achieved this visionary goal in only three years. In Cycle 4, we produced km-scale simulations of our climate
system over multiple decades with a competitive throughput of about 500 simulated days per day (SDPD). Here, we document
65   these simulations and review the evolution of both models, ICON and IFS-FESOM. Throughout the project, we gained knowl-
edge of technical and scientific aspects and encountered positive and negative surprises when analyzing the simulation data.
To create a storyline, we translated that learning process into four questions:

1. Radiation balance (Section 4.1): Can we adjust the radiative properties and formation mechanisms of clouds in order to
    correct the global radiation balance and limit the drifts in the global surface temperature?

70 2. Key features of mean climate (Section 4.2): Are key features of the mean climate correctly represented in a km-scale
    simulation with an energetically consistent climate?

3. Local- to synoptic-scale phenomena (Section 4.3): Does an energetically consistent climate constrain the patterns of
    local-, meso- and synoptic-scale phenomena?

4. Time scales of regional patterns (Section 4.4): Over what simulation times do regional patterns emerge?

75   This paper is structured as follows. In Section 2, we provide an overview of the project concept and models, ICON and
IFS-FESOM. In Section 3, we report the main changes made in both models over the four cycles. In Section 4, we discuss
the four questions stated above and examine the realism of our km-scale simulations. In Sections 5 and 6, we summarize the
development of the past three years and provide an outlook for the future of nextGEMS.

2 nextGEMS concept and models

2.1 Concept

nextGEMS followed an innovative approach to create an integrated community of experts from different domains, including climate science, model development, and high-performance computing (HPC). The model development was structured into four cycles from October 2021 to March 2024. The cycles had an average length of about 8 months and facilitated the fast-paced evolution of the two models, ICON and IFS-FESOM. Each cycle was marked by a coordinated set of simulations with both models and by a hackathon, where more than 100 project participants came together for a week of interactive coding, debugging, plotting, and technical and scientific discussions. The hackathons turned the model development into a collaborative endeavor, with close interaction between the modeling centers and their partner institutions in 14 countries. The hackathons were also used as opportunities to engage with industry stakeholders and show the potential of such km-scale climate simulations for applications such as wind and solar energy or agriculture and fisheries. The interaction patterns between participants were analysed across hackathons, revealing an increase in inter-institutional cooperation and cross-disciplinary exchange over the three years. The outcome and feedback from the hackathons were also critical in the preparation of the next cycle, always aiming to produce more scientifically sound and computationally efficient simulations over longer periods. Figure 1 shows the timeline of the whole project from Cycle 1 to 4. The simulation times evolved from a few weeks with throughputs of about 10 SDPD to three decades with throughputs of about 500 SDPD.

The simulations were produced and stored on the HPC-system Levante (HLRE-4 Levante, 2024). The participants of hackathons worked directly on Levante instead of copying large amounts of simulation data to individual systems. We took several steps to make the data analysis user-friendly and model-agnostic. Most simulation data was provided via Intake catalogs, a Python package for searching and loading data (https://github.com/intake/intake). From Cycle 2 onwards, the simulations were published on the World Data Center for Climate (WDCC). And from Cycle 3 onwards, the simulations were unified into a single Intake catalog (https://data.nextgems-h2020.eu/catalog.yaml). The output of ICON was stored as Zarr datasets (https://github.com/zarr-developers/zarr-python), and the output of IFS-FESOM was indexed using gribscan (Kölling et al., 2024), a tool to scan gridded binary (GRIB) files and create Zarr-compatible indices such that users can access the underlying GRIB files as Zarr datasets. In addition, example notebooks for an initial analysis were provided via the easygems website (https://easy.gems.dkrz.de and https://github.com/nextGEMS).

The two models developed in nextGEMS, ICON and IFS-FESOM, can simulate the atmosphere, ocean and sea ice, and land and their interactions at km-scales. Moreover, they can include additional components of the Earth system, such as the carbon cycle (and aerosols). For this reason, we refer to ICON and IFS-FESOM as km-scale Earth system models, even though the simulations presented here were conducted without any carbon cycle module. Figure 2 shows an overview of how these components and their interactions are represented in each model. The two models are complementary. While both models simulate the Earth system, ICON can be characterized as a research model with a minimal set of parameterization schemes, whereas IFS-FESOM can be characterized as an operational model with elaborate parameterization schemes and
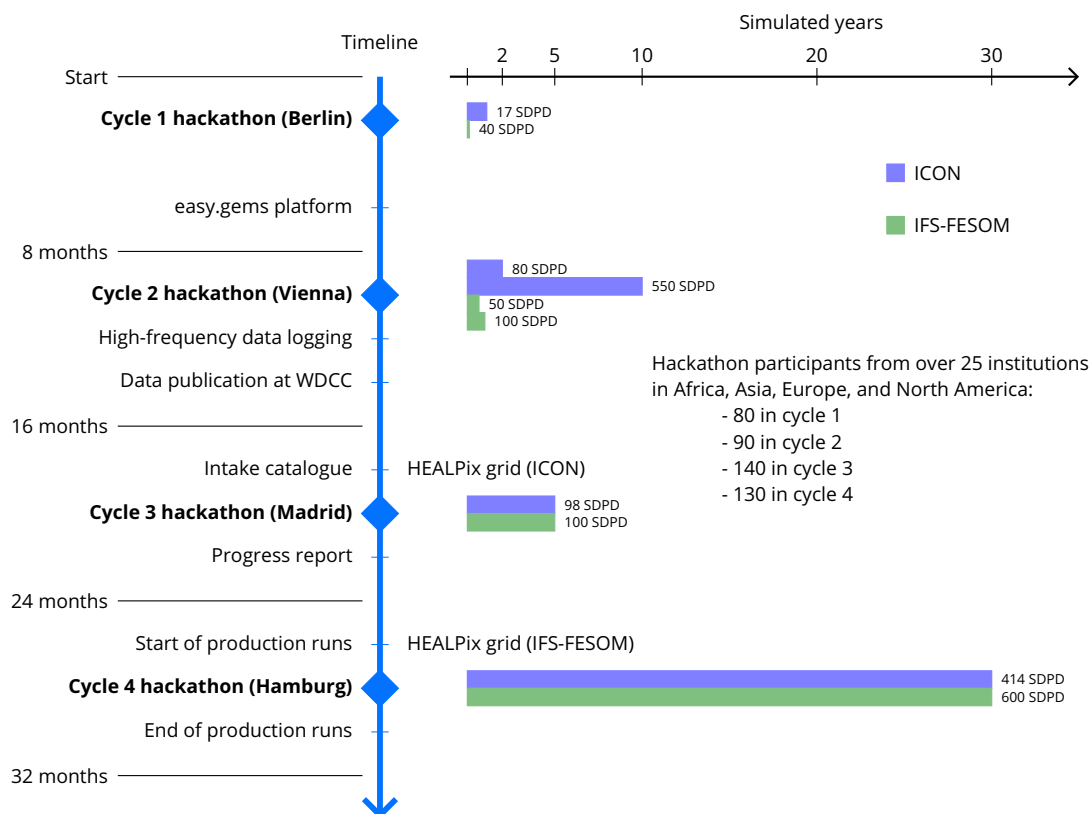
Figure 1. Timeline of nextGEMS from Cycle 1 to 4. The vertical axis shows the development cycles, including hackathons and milestones, and the horizontal axis shows the simulated years, including throughputs in simulated days per day (SDPD).

tuning methods. The key features of both models are described in the next two sections and the developments over the four cycles are summarized in Section 3 (and Appendix A).

2.2 Models

2.2.1 ICON

115

The ICOsahedral Non-hydrostatic model (ICON) is developed by the ICON Partnership. A detailed first description of the km-scale version was presented by Hohenegger et al. (2023). In ICON, the Earth system is divided into three coupled components: atmosphere, land, and ocean. All three components are discretized with an icosahedral-triangular C grid (Zängl et al., 2015).

The atmosphere is modeled with non-hydrostatic equations (Zängl et al., 2015). It includes tendencies from parameterized

120 processes: turbulence (Dipankar et al., 2015; Lee et al., 2022), radiation (Pincus et al., 2019), and cloud microphysics (Baldauf et al., 2011). It is discretized with terrain-following levels in the vertical (Leuenberger et al., 2010). In addition, it includes the one-moment aerosol module HAM-lite developed in nextGEMS (Weiss et al., 2024). HAM-lite represents aerosols as
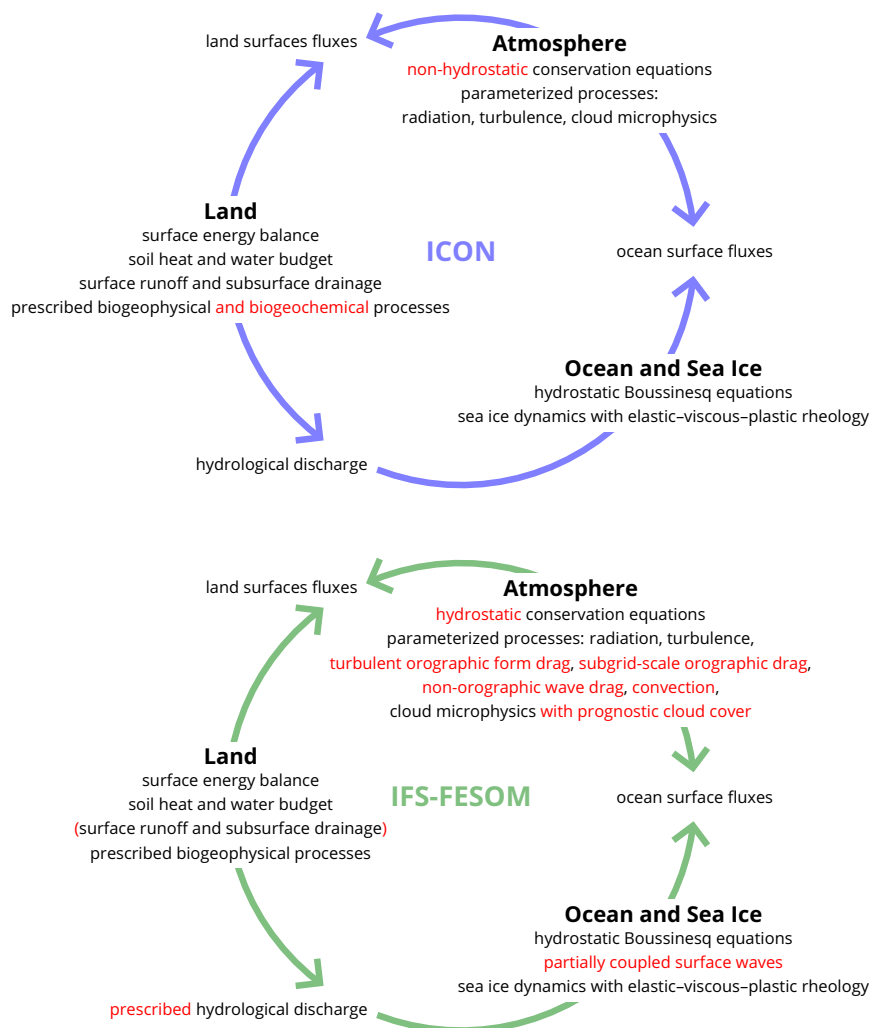
6

Figure 2. Overview of the Earth system models ICON and IFS-FESOM. Differences between the two models are highlighted in red.

an ensemble of log-normal modes with prescribed sizes and compositions to reduce the computational costs. The aerosol modes are transported as prognostic tracers and are coupled with the parameterized processes mentioned above. The land is
125 represented with the land surface model Jena Scheme for Biosphere-Atmopshere Coupling in Hamburg (JSBACH; Reick et al., 2021) and interacts with the atmosphere via surface fluxes and with the ocean via hydrological discharge. It is discretized with multiple soil layers (Reick et al., 2021). The ocean is modeled with hydrostatic Boussinesq equations (Korn et al., 2022) and interacts with the atmosphere via the Yet Another Coupler (YAC; Hanke et al., 2016). It is discretized with variable levels in the vertical that follow the movement of the ocean surface (Korn et al., 2022). The ocean dynamics are spun up as described in
130 Hohenegger et al. (2023) for Cycle 1 and as described in Appendix A for later cycles. Lastly, the sea ice dynamics are modeled based on elastic-viscous-plastic rheology (Danilov et al., 2015).

2.2.2 IFS-FESOM

The Integrated Forecasting System (IFS) is developed and maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF) in collaboration with institutions in its member and cooperating states. A detailed description of the
135 model can be found in the official documentation (ECMWF, 2023a, b). The IFS is centered around the atmosphere coupled with other components of the Earth system such as land, ocean, and sea ice. The atmosphere and land are discretized with an octahedral-reduced Gaussian (TCo) grid (Malardel and Wedi, 2016). The ocean is discretized for nextGEMS with either a triangulated 0.25-degree grid (tORCA025) or an eddy-resolving (NG5) grid (Rackow et al., 2025b).

The atmosphere is modeled with hydrostatic equations and includes parameterization schemes for radiation, turbulence,
140 turbulent orographic form drag, subgrid-scale orographic drag, non-orographic wave drag, convection, and cloud microphysics including a prognostic cloud cover (ECMWF, 2023b). It is discretized with hybrid levels in the vertical, transitioning from terrain-following coordinates at lower levels to pressure-level coordinates at higher levels (Simmons and Strüfing, 1983). The land is represented with the land surface model ECLand (Boussetta et al., 2021) and discretized with four soil layers. In its current configuration, ECLand computes local runoff, but it does not route the runoff nor provide it as freshwater input into
145 the ocean. Instead, the hydrological discharge is prescribed for all coastal points. In Cycle 1, the ocean is represented with the FESOM2.1 model (Wang et al., 2014; Scholz et al., 2019), developed and maintained by the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI). It is discretized with the tORCA025 grid using a parameterization for mesoscale eddies (Gent and Mcwilliams, 1990; Gent et al., 1995). In the later cycles, the ocean is updated to FESOM2.5 and is discretized with the NG5 grid using an eddy-resolving resolution of 5 kilometers (Rackow et al., 2025b). In all cycles,
150 the vertical dimension is discretized with Arbitrary-Lagrangian-Eulerian coordinates. The ocean dynamics are spun up in stand-alone mode for 5 years with atmospheric forcing from the ERA5 reanalysis (Hersbach et al., 2020) before the coupled IFS-FESOM simulations are started. Lastly, waves are represented with the ecWAM model (Janssen, 2004), and sea ice is represented with the FESIM model (Danilov et al., 2015, 2017).

2.2.3 Output convergence

155 A large effort was made to harmonize the model output and to facilitate an easy comparison between the two models. As an important step towards this end, ICON and IFS-FESOM both provided their output on a HEALPix grid in Cycle 4 (Górski et al., 2005). HEALPix stands for Hierarchical Equal Area isoLatitude Pixelation of a sphere. The pixels of such a grid all have the same area and are located on lines of constant latitude. In addition, the hierarchical tessellation of such a grid makes it possible to visualize and process the data on different resolutions (https://healpix.sourceforge.io). In ICON, the HEALPix
160 grid was incorporated with an improved version of the YAC coupler (Hanke et al., 2016) and the newly developed HIOPY module (https://gitlab.gwdg.de/ican/hiopy/). In IFS-FESOM, the HEALPix grid was incorporated with the multIO framework (Sarmany et al., 2024) and the MIR package (Maciel et al., 2017). This achievement was made possible thanks to synergistic work with the Climate Adaptation Digital Twin of Destination Earth. Apart from the HEALPix grid, developers of ICON and IFS-FESOM worked closely together to provide a minimum set of common variables, to output atmospheric variables at a

165   common set of pressure levels, and to output variables on a common, height-dependent frequency. From Cycle 2 onwards, IFS-FESOM provided some post-processed output of general interest, such as monthly means and coarser output on a regular 0.25-degree grid (Rackow et al., 2025b). From Cycle 3 onwards, the output was computed on the fly by multIO (Sarmany et al., 2024) and MIR (Maciel et al., 2017) before writing it to disk.

3   Simulations

170   In this section, we describe the important characteristics of each cycle, focusing on the simulation period, computational throughput, and model biases in terms of mass and energy conservation. To refer to the simulations, we use the following conventions: ICON-C*X* and IFS_F-C*X*, in which *X* indicates the cycle and IFS_F stands for IFS-FESOM. In the case of two simulations from the same cycle, we added the letters A and B, e.g., ICON-C2-A and ICON-C2-B or IFS_F-C2-A and IFS_F-C2-B. Table 1 provides an overview of the simulations from Cycles 1 to 4. All simulations started on 20 January 2020. The last

175   column shows whether the simulation was energetically consistent, as defined in Section 1. The simulation data was published on the World Data Center for Climate (Wieners et al., 2023; Koldunov et al., 2023; Wieners et al., 2024).

    While there is a clear increase in throughput over all cycles, the input/output (I/O) operations in IFS_F-C2-A and IFS_F-C2-B were extremely expensive as there were no dedicated I/O servers. The computational throughput was, therefore, less than expected for 864 nodes. This has been solved with the introduction of MultIO (Sarmany et al., 2024) in IFS and FESOM in

180   Cycle 3 and Cycle 4, as explained in Section 3.3 and Section 3.4, respectively.

3.1   Cycle one

The first cycle started with the ICON and IFS-FESOM versions presented by Hohenegger et al. (2023) and Rackow et al. (2025b), respectively. The simulation with ICON, referred to as ICON-C1, used a horizontal grid of 5 km in the atmosphere and ocean. In addition, there were 90 vertical levels in the atmosphere, 128 vertical levels in the ocean, and 5 soil layers in the

185   land. The time step was 40 s for the atmosphere (and land) and 80 s for the ocean. Note that the same simulation was presented by Hohenegger et al. (2023) under the name G_AO_5km. ICON-C1 was integrated over one year on the Mistral supercomputer of DKRZ. The simulation with IFS-FESOM, referred to as IFS_F-C1, was based on IFS CY46R1 (ECMWF, 2019) and it used a horizontal grid spacing of 4.4 km in the atmosphere and 25 km in the ocean (a triangulated version of NEMO's ORCA025 grid for use with FESOM, coined 'tORCA025'). IFS_F-C1 was integrated over 40 days on the Cray supercomputer of ECMWF.

190   Both Mistral and Cray supercomputers were decommissioned in 2021 and replaced with Levante (HLRE-4 Levante, 2024) and Atos (ECMWF, 2024), respectively.

    ICON-C1 and IFS_F-C1 both showed large imbalances in the energy budget. In ICON-C1, the imbalance in the atmospheric energy budget of about $-4\,\mathrm{W\,m^{-2}}$ caused a drift in the global mean surface temperature of about $-2\,\mathrm{K}$ (Mauritsen et al., 2022; Hohenegger et al., 2023). In IFS_F-C1, the imbalance in the energy budget of $6.4\,\mathrm{W\,m^{-2}}$ (and smaller at coarser resolutions)

195   was related to a leak in the water budget in the atmosphere. This leak caused a spurious increase in the total precipitation of about $4.6\,\%$ (Rackow et al., 2025b). On the technical side, the throughput in ICON-C1 was 17 SDPD on 400 nodes of the

Table 1. Overview of all simulations from nextGEMS Cycle 1 to 4 performed with ICON and IFS-FESOM (IFS_F). Δx_A and Δx_O indicate the horizontal resolution of the atmosphere and ocean, respectively. Period indicates the integration time. Nodes refers to the number of CPU nodes on Levante (128 cores in total, 256 GB main memory) or an equivalent number of nodes (indicated with *). Throughput refers to the simulated days per day (SDPD), normalized by the number of nodes (SDPD node^{-1}). The last two columns show the atmospheric radiative forcing and whether the simulation is energetically consistent. The simulation data of Cycles 2 to 4 was published on the World Data Center for Climate (Wieners et al., 2023; Koldunov et al., 2023; Wieners et al., 2024).

	Δx_A (km)	Δx_O (km)	Period	Nodes	Throughput (SDPD, SDPD node^{-1})	Forcing	Energetically consistent
Cycle 1							
ICON-C1	5	5	406 days	100*	17, 0.17	2020	No
IFS_F-C1	4.4	25	40 days	20*	40, 2	2020	No
Cycle 2							
ICON-C2-A	5	5	2 years	400	80, 0.2	2020	No
ICON-C2-B	10	10	10 years	400	550, 1.375	2020	No
IFS_F-C2-A	2.8	5	8 months	864*	50, 0.057	2020	No
IFS_F-C2-B	4.4	5	1 year	864*	100, 0.116	2020	No
Cycle 3							
ICON-C3	5	5	5 years	530	98, 0.185	2020	No
IFS_F-C3	4.4	5	5 years	269	100, 0.372	2020	Yes
Cycle 4							
ICON-C4	10	5	30 years	464	414, 0.892	SSP3-7.0	Yes
IFS_F-C4	9	5	30 years	269	600, 2.23	SSP3-7.0	Yes

Mistral supercomputer, roughly corresponding to 100 nodes of Levante, whereas the throughput in IFS_F-C1 was 40 SDPD using 60 nodes of the Cray supercomputer, roughly corresponding to 20 nodes of Levante. With these numbers, ICON could simulate 0.17 SDPD per node, 10-fold less than IFS (2 SDPD per node). As a rule of thumb, a possible explanation is the coarse horizontal grid spacing in the ocean module in IFS. The different machines used to simulate ICON-C1 and IFS_F-C1 could also be an additional reason for higher SDPD per node in IFS. Thus, aside from the large energy imbalance, the small computational throughputs were also the main obstacles to the upcoming decadal simulations.

3.2 Cycle two

In the second cycle of IFS-FESOM, the IFS version was upgraded to CY47R3 (ECMWF, 2021). This included, among other changes, the inclusion of a 5-layer snow scheme, compared to the previously used bulk scheme. In addition, the atmospheric imbalance of water and, therefore, energy was significantly reduced to less than $1\,\mathrm{W\,m^{-2}}$. This was done by activating global tracer mass fixers (Rackow et al., 2025b). Such modifications also improved the NWP configuration of IFS-NEMO at ECMWF
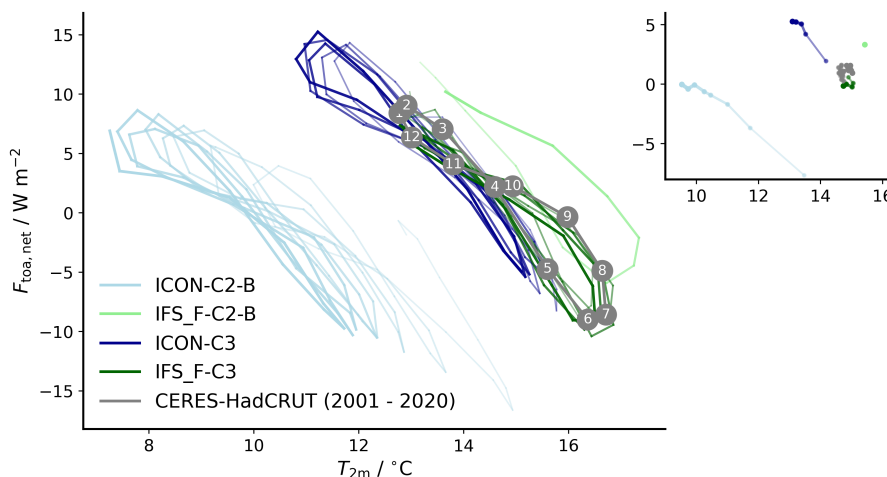
Figure 3. Annual cycles of the monthly mean 2-meter temperature and top-of-atmosphere radiation balance from ICON-C2-B, ICON-C3, IFS_F-C2-B, and IFS_F-C3. Note that the radiation balance is computed as the difference of the downwelling and upwelling radiation, i.e., $F_{\text{toa,net}} = F^{\downarrow}_{\text{toa,sw}} - F^{\uparrow}_{\text{toa,sw}} - F^{\uparrow}_{\text{toa,lw}}$ (Mauritsen et al., 2022). The widths and opacities of the lines increase with time. Gray line shows the observational reference averaged over 2001 to 2020, i.e., 2-meter temperature from HadCRUT (Morice et al., 2021) and radiation balance from CERES-EBAF (Loeb et al., 2018). The months are labeled with numbers. The corresponding annual means are shown in the top-right inset. Like for the annual cycles, the widths and opacities of the lines increase with time.

and were, therefore, incorporated into its operational version CY48r1 (Rackow et al., 2025b). Two IFS-FESOM simulations, IFS_F-C2-A and IFS_F-C2-B, were provided. Both simulations shared a common horizontal grid spacing in the ocean ($\Delta x = 5\,\text{km}$) but used different grid spacings in the atmosphere. IFS_F-C2-A used a horizontal grid spacing of 2.8 km and was integrated for 8 months, whereas IFS_F-C2-B used a horizontal grid spacing of 4.4 km and was integrated for 1 year. Both simulations were subject to a warming trend on the order of $1\,\text{K}\,\text{y}^{-1}$ as shown in Figure 3.

In the second cycle of ICON, the energy leak in Cycle 1 was traced back to bugs in the dynamical core, cloud microphysics, and surface fluxes within the turbulence scheme. Bugs in the microphysics and turbulence were resolved, and the energy imbalance was reduced. Besides these fixes, the new radiation scheme RTE-RRTMGP (Pincus et al., 2019) was implemented, and a cloud inhomogeneity factor was introduced to tune the radiation balance at the top of the atmosphere, as discussed further in Section 4.1. In the ocean module, a new vertical coordinate system with thinner surface levels was used. Further details on the ICON development in this cycle are summarized in Appendix A. After 8 test runs, two simulations were provided. ICON-C2-A used a horizontal grid spacing of 5 km and was integrated for 2 years, whereas ICON-C2-B used a horizontal grid spacing of 10 km and was integrated for 10 years. Both simulations were subject to a cold drift, as shown in Figure 3.

IFS_F-C2-A and IFS_F-C2-B were both performed on the Atos supercomputer of ECMWF, where one node is roughly equivalent to one node in Levante. The throughput in IFS_F-C2-A (2.8 km) was 50 SDPD on 864 nodes, giving a normalized throughput of 0.057 SDPD per node. Coarsening the resolution by roughly a factor of 2 in IFS_F-C2-B (4.4 km) increased the

throughput by a factor of 2 on the same number of nodes, resulting in a normalized throughput of 0.116 SDPD per node. The
17-fold reduction in the normalized throughput in IFS_F-C2-B compared to IFS_F-C1 is explained by several factors. In Cycle
2, IFS-FESOM introduced an eddy-resolving ocean. This means an 8-fold increase in grid points and a 5-fold smaller timestep,
resulting in 40-fold higher costs for the ocean. More output variables were written in Cycle 2 with a higher spatial and temporal
resolution. Moreover, the IO configuration for IFS-FESOM was suboptimal as it did not make use of any IO servers yet. In
addition, full ocean support for hybrid MPI-OpenMP parallelization was only supported with the release of FESOM version
230 2.5 (used from Cycle 3 onwards). We also decided to make use of a large number of available nodes on the Atos supercomputer
at the time, despite the lack of an optimized IO configuration. Thus, even with a considerably greater number of resources in
Cycle 2 compared to Cycle 1, which sped up the computing part of the model, the total throughput of IFS-FESOM, including
IO, did not scale accordingly, and a large number of used nodes, therefore, reduces the normalized throughput considerably.

ICON-C2-A and ICON-C2-B were both performed on the Levante supercomputer. The throughput in ICON-C2-A (5 km)
235 was 80 SDPD on 400 nodes, meaning 0.2 SDPD per node. Coarsening the resolution by a factor of 2 in ICON-C2-B (10 km)
increased the throughput by a factor of 7 on the same number of nodes, increasing the normalized throughout to 1.375 SDPD
per node. With the same resolution in ICON, simulations in Cycle 1 and 2 have the same throughput per node, meaning that
the increase in throughput was due to the use of more nodes.

3.3 Cycle three

240 In the third cycle of IFS-FESOM, the base version of IFS was upgraded to CY48R1 (ECMWF, 2023a, b; Rackow et al., 2025b).
On top of that, the representation of uncertain cloud and microphysical processes was optimised to yield improved shortwave
and longwave TOA fluxes and, as a result, a top-of-atmosphere radiation budget within observational uncertainty. The reduced
cloud base mass-flux approach was introduced to increase the convective organization in simulations at 4.4 km resolution.
In the land module, land use/land cover information was revised to describe km-scales better (Boussetta et al., 2021), and
245 a parametrization for urban processes was added (McNorton et al., 2021, 2023). FESOM was upgraded to version 2.5 with
significant changes in the atmosphere-ocean coupling, such as the coupling of ocean surface velocities and accounting for the
enthalpy of snow falling into the ocean (Rackow et al., 2025b). With all these modifications, IFS-FESOM was integrated in
time for 5 years using a horizontal grid spacing of 4.4 km in the atmosphere and 5 km on average in the ocean (IFS_F-C3). In
contrast to previous cycles, IFS_F-C3 was not subject to drift in the surface temperature compared to observations, indicating
250 an energetically consistent climate (Figure 3). Moreover, the radiation fluxes at the top of the atmosphere were also close to
observations (Rackow et al., 2025b).

In the third cycle of ICON, the energy imbalance was further reduced by fixing bugs in the cloud microphysics and turbulence
scheme. In the calculation of surface fluxes, the heat capacity at constant pressure was replaced by the heat capacity at constant
volume. ICON computes those fluxes under the assumption of constant volume (not pressure). This change reduced the amount
255 of energy transferred from the surface to the atmosphere by about 29 %. Like in the previous cycle, the cloud inhomogeneity
factor was used to tune the radiation balance. Further changes in the ocean and land modules are summarized in Appendix A.
After 27 test runs, ICON was integrated for 5 years using a horizontal grid spacing of 5 km in the ocean and atmosphere. ICON-

C3 exhibited a cooling trend of about $0.33\,\mathrm{K\,y^{-1}}$, as shown in Figure 3. This cold drift could be traced back to remaining energy leaks in the dynamical core.

All simulations with IFS-FESOM were performed on the Levante supercomputer from Cycle 3 onwards. The throughput per node showed a 3-fold increase in Cycle 3 (0.372 SDPD per node) compared to Cycle 2 (0.116 SDPD per node), explaining the similar throughput (100 SDPD) even if IFS_F-C3 used a third of the resources of IFS_F-C2-B. The resource efficiency was increased substantially by optimizing the I/O operations, and from Cycle 3 onwards, I/O servers via multIO were introduced in IFS. Moreover, less 3D ocean output was written in Cycle 3 compared to Cycle 2 (only the upper 300 m at 3-hourly frequency), which dramatically reduced the time spent for I/O in IFS-FESOM. In ICON, the throughput per node in Cycle 3 (0.185 SDPD per node) was roughly similar to Cycle 2 (0.2 SDPD per node). Thus, the use of slightly more resources in Cycle 3 (530 nodes) compared to Cycle 2 (400 nodes) increased the throughput from 80 to 98 SDPD.

3.4 Cycle four

In the fourth cycle of IFS-FESOM, given the success of providing energetically consistent simulations in Cycle 3, the last changes needed were mostly related to preparing input for climate projections. In addition, only daily output was written for ocean fields, in line with IFS-FESOM simulations for the Climate Change Adaptation Digital Twin in Destination Earth. IFS_F-C4 used a horizontal grid spacing of 9 km in the atmosphere and 5 km in the ocean and was integrated for 30 years.

In Cycle 4, the big challenge in ICON was to simulate an energetically consistent climate. To this end, the following changes in the atmosphere module were made. The diffusivity in the momentum equation of the dynamical core was modified to account for the persistent but small leak in internal energy in the dynamical core. The old turbulent mixing scheme VDIFF was upgraded to enable both explicit and implicit numerical solvers, to replace the implicit atmosphere-surface coupling with an explicit coupling, and to shift the diffused thermal quantity from dry static energy to internal energy. This new upgrade is called TMX, and it contributed to a further reduction of the energy leak at the surface. Lastly, the cloud inhomogeneity factor was linked with the lower tropospheric stability to better account for the cloud type, as discussed in Section 4.1. Further details on the development of ICON-C4 are summarized in Appendix A. Furthermore, the implementations to reduce the energy leak in ICON are now taken as a base for future simulations. After numerous short test runs, ICON-C4 was integrated for 30 years using a horizontal grid spacing of 10 km in the atmosphere and 5 km in the ocean.

ICON-C4 and IFS_F-C4 both used greenhouse gas concentrations, including ozone, following the Coupled Model Intercomparison Project Phase 6 (CMIP6) for the forcing time series of a SSP3-7.0 scenario (O'Neill et al., 2016). In addition, ICON-C4 used the time-varying Max Planck Institute Aerosol Climatology (MACv2-SP; Stevens et al., 2017) for anthropogenic aerosols together with the Stenchikov climatology from 1999 for volcanic aerosols (Stenchikov et al., 1998). IFS_F-C4 used the time-varying tropospheric climatology provided by the CONsistent representation of temporal variations of boundary Forcings in reanalysES and Seasonal forecasts (CONFESS) project (Stockdale et al., 2024), in which volcanic aerosols from 1850 with no large eruptions were applied to all years after 2014.

In view of the project timeline and less than desired access to computational resources, a compromise had to be made between the model resolution and simulation length for the production simulations. This led to 30-year simulations at resolutions of

about 10 km in the atmosphere and 5 km in the ocean. In ICON-C4, the 10 km grid spacing of the atmosphere allowed us to reach a throughput per node of 0.892 SDPD per node, and using 464 nodes (18% of Levante's nodes), the total throughput was 414 SDPD. In IFS_F-C4, the 9 km grid spacing in the atmosphere, together with further improvements in I/O operations by using multIO for both IFS and FESOM, gave a throughput per node of 2.23 SDPD per node. Using 269 nodes (9% of Levante's nodes), IFS_F-C4 simulated 600 SDPD.

Thus, in less than three years of model development, nextGEMS developed two models with an energetically consistent climate, which can be a persistent issue in established climate models (e.g. Sanderson and Rugenstein, 2022), and with a competitive throughput. Taken together, multidecadal and global km-scale climate simulations, in which land, atmosphere, and ocean are coupled, have now become a reality. This opens new ways to analyze regional atmospheric, oceanic, and land processes on global scales and their changes in a global warming context.

4 Insights into the realism of km-scale simulations

In this part, we examine the realism of the Cycle 4 simulations outlined in Section 3.4. To create a storyline, we discuss four central questions on the radiation balance, key features of mean climate, local- to synoptic-scale phenomena, and time scales of regional patterns, as introduced in Section 1. The reader should note that robust and definite answers to these questions will form gradually as separate studies from the different scientific groups of nextGEMS emerge.

4.1 Radiation balance

As outlined in Section 3, it was possible to improve the conservation of mass and energy and eliminate the large drifts in the near-surface temperature observed in the early simulations. This does not imply necessarily, however, that the energy balance at the top-of-atmosphere agrees with observations. In both ICON and IFS-FESOM, cloud radiative properties and formation mechanisms were adjusted to reduce the remaining biases. However, the tuning strategies in ICON and IFS-FESOM differed. The low cloud cover, for example, was too high in ICON but too low in IFS-FESOM.

In ICON, the cloud cover was adjusted in two steps targeting the turbulent mixing and cloud brightness. First, we adjusted the mixing in the Smagorinsky scheme, which depends on the Richardson number, following the formulation of Louis (1979). This adjustment allows some mixing or entrainment in situations where the traditional Smagorinsky scheme would yield no mixing. Mixing at large Richardson numbers controls stratiform boundary layer clouds but has only small effects on trade wind clouds. Second, we accounted for the lack of a cloud fraction scheme through a cloud inhomogeneity factor (ζ), which depends on the lower-tropospheric stability (LTS). LTS is defined as the difference in potential temperature between the free troposphere and the surface and is strongly correlated with the stratiform low cloud cover (Wood and Bretherton, 2006). In theory, the inhomogeneity factor is equal to one for fully-resolved, homogeneous clouds and is less than one for partially-resolved, inhomogeneous clouds, as discussed by (Cahalan et al., 1994; Mauritsen et al., 2012). It scales down the cloud water and ice before the shortwave fluxes are calculated in the radiation scheme. In ICON, the inhomogeneity factor acts only on the liquid clouds, whereas the ice clouds remain unchanged. It increases non-linearly from about 0.4 ($\zeta_{\min}$) at a lower-tropospheric

stability of $0\,\mathrm{K}$ to about 0.8 ($\zeta_{\max}$) at a lower-tropospheric stability of $30\,\mathrm{K}$, i.e.,

$$\zeta = \zeta_{\min} + (\zeta_{\max} - \zeta_{\min})\left(1 - \frac{\arctan 2\,(c_1, \mathrm{LTS} - c_2)}{\pi}\right), \tag{1}$$

where $\zeta_{\min} = 0.4$, $\zeta_{\max} = 0.8$, $c_1 = 2\,\mathrm{K}$ and $c_2 = 20\,\mathrm{K}$ are tuning parameters.

In IFS-FESOM, the cloud cover is a prognostic variable of the cloud microphysics scheme (ECMWF, 2023b). It shows a resolution and time step dependence. Overall, it is smaller at higher resolution with shorter time steps than at coarser resolutions with longer time steps. To find a configuration that is close to observations of both longwave and shortwave radiative fluxes from CERES-EBAF (Loeb et al., 2018), the cloud cover was modified as documented in Rackow et al. (2025b) by

- reducing the inhomogeneity enhancement factor for accretion from 3 to 2,

- reducing the cloud edge erosion from $6 \cdot 10^{-6}$ to $4 \cdot 10^{-6}$, and

- assuming a constant effective cloud spacing, following recommendations by Fielding et al. (2020).

Overall, these changes led to an increase in the low cloud cover. In addition, the high cloud cover was increased in areas with strong deep convective activity by

- decreasing a threshold that limits the minimum size of the ice effective radius from 60 to 40, in agreement with observational evidence

- and changing from cubic to linear departure point interpolation in the semi-Lagrangian advection scheme for all water species except vapour.

In combination, these changes in the cloud cover in IFS-FESOM increased the outgoing shortwave radiation by about $5\,\mathrm{W\,m^{-2}}$, while decreasing outgoing longwave radiation by about $3\,\mathrm{W\,m^{-2}}$. This led to a well-balanced radiation budget at the TOA with both shortwave and longwave fluxes in realistic ranges.

Figure 4 shows the time series of the monthly mean 2-meter temperature (Figure 4a) and annual cycles of the monthly mean 2-meter temperature and top-of-atmosphere radiation balance (Figures 4b,c) over the whole simulation period. The annual means of the near-surface temperature and radiation balance are shown in Figures 4b,c. The corresponding observations from CERES (Loeb et al., 2018) and HadCRUT (Morice et al., 2021) are shown as well. As discussed by Mauritsen et al. (2012, 2022), the annual cycles are shaped like a figure eight. In the first half of the year, the radiation balance decreases while the near-surface temperature increases. In the second half, the radiation balance increases while the near-surface temperature decreases. The near-surface temperature lags behind the radiation balance, as would be expected from the large heat capacity of the Earth system. Overall, the values of IFS-FESOM agree better with observations than the values of ICON. The near-surface temperatures are $13.83\,^\circ\mathrm{C}$ for ICON, $14.94\,^\circ\mathrm{C}$ for IFS-FESOM, and $14.91\,^\circ\mathrm{C}$ for HadCRUT, averaged over the first four years from 2020 to 2023. And the top-of-atmosphere radiation balances are $0.25\,\mathrm{W\,m^{-2}}$ for ICON, $0.51\,\mathrm{W\,m^{-2}}$ for IFS-FESOM, and $1.46\,\mathrm{W\,m^{-2}}$ for CERES. Moreover, the near-surface temperature of ICON cools in contrast to IFS-FESOM. We assume
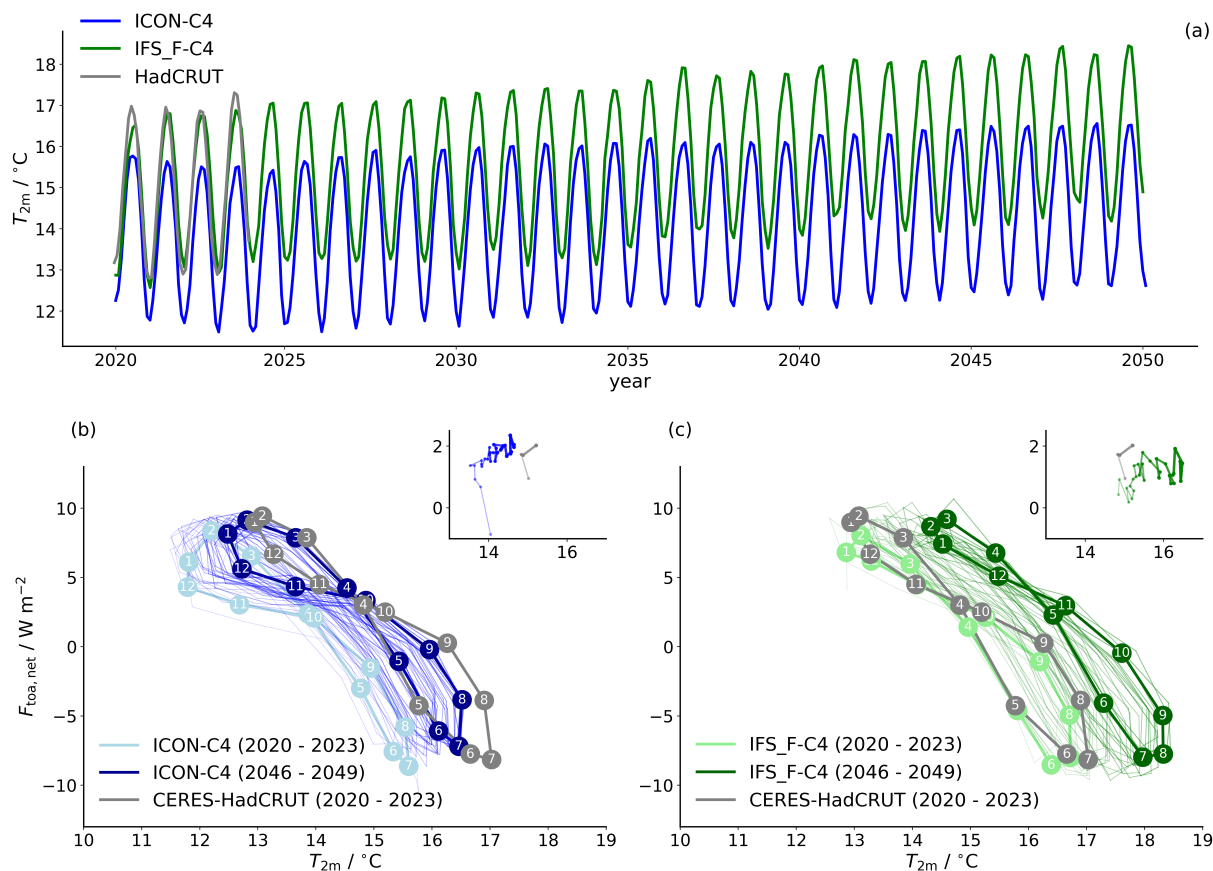
Figure 4. Time series of the monthly mean 2-meter temperature from ICON-C4 and IFS_F-C4 (a). Annual cycles of the monthly mean 2-meter temperature and top-of-atmosphere radiation balance from ICON-C4 (b) and IFS_F-C4 (c). Thick lines indicate averages over 2020 to 2023 (light colors) and 2046 to 2049 (dark colors), where months are labeled with numbers. Thin lines indicate individual cycles from 2020 to 2049, where widths and opacities increase with time. Gray lines show the observational reference, i.e., 2-meter temperature from HadCRUT (Morice et al., 2021) and radiation balance from CERES-EBAF (Loeb et al., 2018). The corresponding annual means are shown in the top-right insets. As for the annual cycles, the widths and opacities increase with time.

that this initial cooling is related to the adjustment of the atmosphere to the ocean, which is spun up with ERA5 forcing. In
355 the subsequent years, the radiation balance is positive, and the near-surface temperature increases in line with the SSP3-7.0
scenario in both ICON and IFS-FESOM. A detailed analysis of the shortwave reflectivity of stratocumulus clouds, which are
only partially resolved in km-scale Earth System models, is presented in Section 4.3.2.

4.2 Key features of mean climate

Our next question is whether a simulation with an energetically consistent climate, together with solving the oceanic and at-
360 mospheric flows on the km-scale, can have an adequate representation of the Earth's mean climatic features. To address this,
we select two key large-scale features in the energetics of the climate system: (i) the tropical rainbelt, and (ii) the pattern of sea
surface temperature (SST). The role played by km-scale processes in convection gives an additional reason to analyze the rep-
resentation of the tropical rainbelt, which has been notoriously difficult to reproduce in coarse-resolution Earth system models
(Mechoso et al., 1995; Lin, 2007). The analysis of the SST patterns, that coarse-resolution Earth system models fall short in
365 reproducing, is important because being able to capture it has emerged as an important factor in determining climate sensitivity.
The tropical rainbelt is here defined as the quantile 80 of the yearly precipitation mean in the tropics, i.e., it corresponds to
the 20% wettest region in the tropics (30°S-30°N; see Segura et al., 2022). Using ICON-C4 and IFS_F-C4, respectively, the
yearly precipitation mean is computed from the first five complete years of integration, 2021-2025. This pragmatic choice is
made here to allow nearly one year of spinup, from 20th January to 31st December 2020, and because computing the yearly
370 mean for the entire 30-year period would include the impact of the SSP3-7.0 forcings on the climate. In these five years, the El
Niño index 3.4 is in near-neutral conditions in both simulations, with the exception of one year in IFS_F-C4. This means that
the biases and the correct representation of large-scale features are not affected by the ENSO variability.

The structure of the terrestrial tropical rainbelt is encouragingly similar between both simulations and observations from
the precipitation satellite product Integrated Multi-satellitE Retrievals for GPM (IMERG; Huffman et al., 2019) (Figure 5).
375 The location is well reproduced, albeit with a reduced area in both simulations over South America. Regarding the tropical
ocean, ICON-C4 and IFS_F-C4 show a similar structure of the tropical rainbelt in the Eastern Pacific that matches IMERG.
However, a less consistent picture between both simulations appears over other tropical oceans. While the oceanic tropical
rainbelt is relatively well reproduced in IFS_F-C4 in terms of pattern, the oceanic tropical rainbelt is less well reproduced in
ICON-C4. The westward extension of the tropical rainbelt is too small in ICON in the Atlantic Ocean, while IFS does extend
380 westward similarly to observations. The meridional extent is, however, underestimated. In the Indo-Pacific region, ICON-C4
underestimates precipitation in the equatorial region, causing the well-known double ITCZ bias in this region, observed in
traditional climate models. This bias is evident in the zonal mean precipitation over the western Pacific (Figure 5c). In the
equatorial western Pacific, ICON-C4 simulates 7 mm d^{-1} less precipitation than what is expected from observations. In the
case of IFS_F-C4, the dry bias at the equator over the western Pacific is smaller than in ICON-C4. Precipitation is 3 mm d^{-1}
385 lower than IMERG in the equatorial western Pacific. The bias in IFS_F-C4 vanishes in the equatorial Pacific when IFS is
coupled with the NEMO ocean model (Rackow et al., 2025b), pointing to different ocean surface representations and coupling
choices as potential drivers of the development of this bias in IFS. IFS_F-C4 also presents a reduced area of tropical rainbelt
in the Bay of Bengal and over Southeast Asia (Figure 5b).

Overall, precipitation over tropical oceans appears to be less constrained than over land. The remaining biases in precipitation
390 over the tropical Pacific are linked to deviations in the tropical SST patterns, as observed in traditional climate models (Lin,
2007). Figure 6 shows the yearly-mean pattern of SST for the 2021-2025 period of simulation in IFS_F-C4 and ICON-C4, and
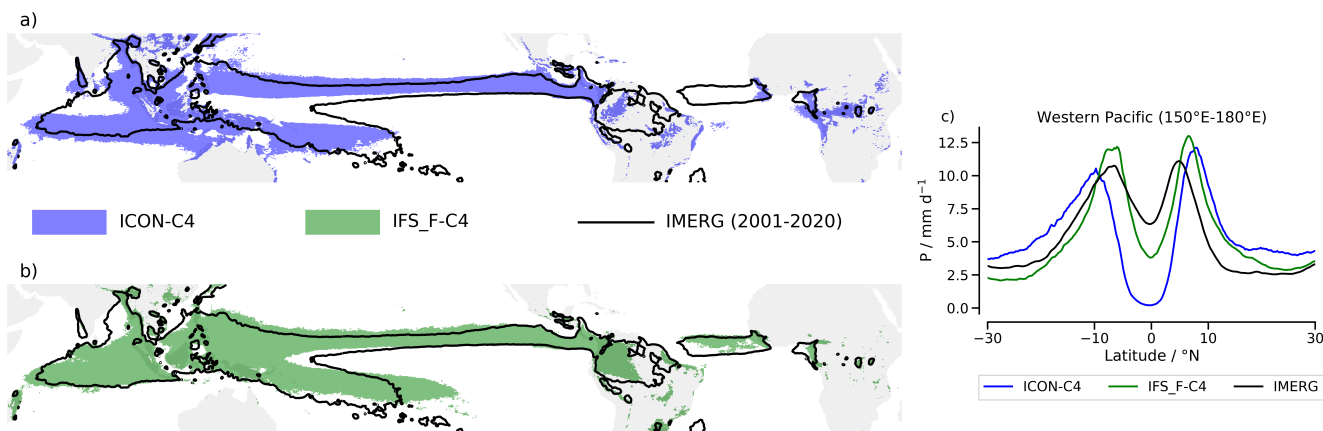
Figure 5. On the left, tropical rainbelt in ICON-C4 (a) and IFS_F-C4 (b) averaged over 2021 to 2025. The tropical rainbelt from IMERG (Huffman et al., 2019) averaged over 2001 to 2020 is outlined by black contour lines. On the right, zonal mean precipitation corresponding to the rainbelts on the left averaged over the western Pacific (c).

the pattern of SST for the HadISST climatology (2001-2020). In the Pacific, the ICON-C4 and IFS_F-C4 represent the cold-warm gradient between the Eastern and western Pacific similar to observations. However, in ICON-C4, there is a westward extension of cold waters reaching the western Pacific, constraining the development of the West Pacific warm pool. In IFS_F-
395 C4, this bias is not as prominent, with less westward extension of the cold tongue. In Figure 6, the warmest regions in the western Pacific and the western Atlantic are weaker than in observations. The West Pacific warm pool in IFS_F-C4 is 3-4 K warmer than the tropical SST mean, while in observations, the warm pool can be 4-5 K warmer than the tropical mean. The 1 K difference is also observed in the western Atlantic. In ICON-C4, the western Pacific warm pool is only 2 K warmer than the tropical mean. A similar value is observed in the western Atlantic.

400 The comparison between ICON-C4 and IFS_F-C4 raises some preliminary conclusions pointing to the positive and negative surprises in developing the two models. The fact that ICON-C4 can represent the tropical rainbelt over land and the Eastern Pacific indicates that a horizontal grid spacing of the order of 10 km is sufficient to reproduce the structure of precipitation in those regions, and that is possible with a minimum set of parameterizations. A better representation of the tropical rainbelt in IFS-_F-C4 than in ICON-C4 reflects the extensive experience of IFS as a weather forecast model with a long history in
405 model tuning to match the precipitation pattern. The difference between the representation of the SST pattern between ICON-C4 and IFS-_F-C4 suggests that a better representation of the equatorial SST pattern and the tropical rainbelt is linked. In ICON, simulations using prescribed SST with a similar configuration to nextGEMS show the importance of surface heat fluxes in light near-surface winds to get a single tropical rainbelt in the Indo-Pacific region (Segura et al., 2024). The experiments conducted by Segura et al. (2024) and Takasuka et al. (2024) indicate that fine-tuning subgrid-scale processes can produce a
410 correct representation of the tropical rainbelt without the use of convective parameterization.
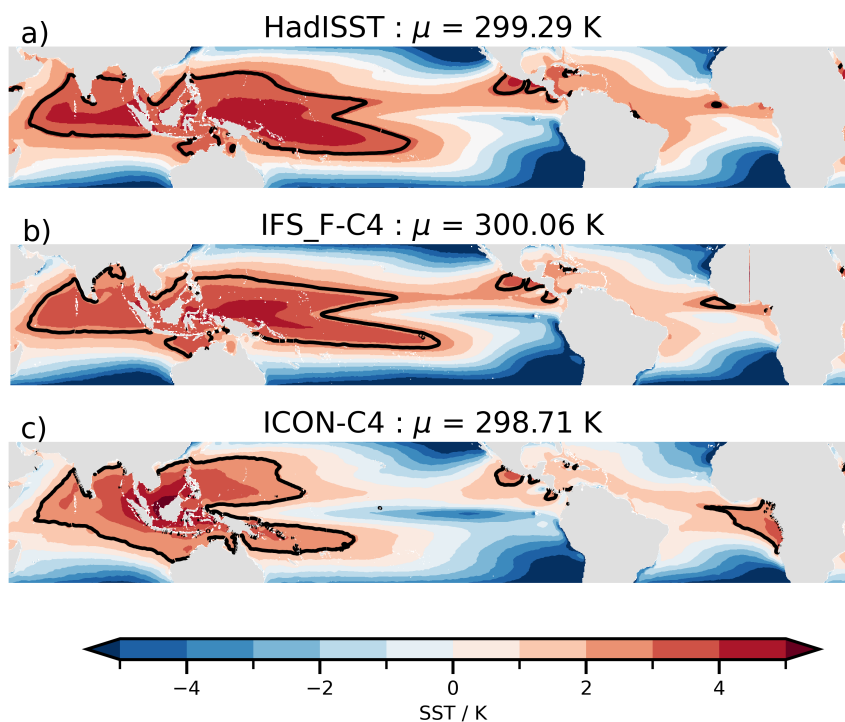
Figure 6. Annual mean sea surface temperature (SST) from HadISST (Rayner et al., 2003) averaged over 2001 to 2020 (a) and from IFS_F-C4 (b) and ICON-C4 (c) averaged over 2021 to 2025. The spatial mean, denoted as μ in the subplot titles, is subtracted from each product. And the 80th percentile is outlined by black contour lines.

4.3 Local- to synoptic-scale phenomena

In Section 4.1 we demonstrated that the nextGEMS simulations result in an energetically consistent global climate. As shown in Section 4.2, some of the typical spatial precipitation and SST patterns in the Pacific are acceptably reproduced, but some long-standing issues remain. In the following subsections, we investigate to what extent an energetically consistent climate translates into a constraint for local, meso-scale and synoptic-scale phenomena.

4.3.1 Local-scale phenomena

We here take the soil moisture–precipitation feedback as an exemplifying process of how the two models represent a local-scale phenomenon. The reasoning behind this is that local convection, on scales below 100 km, plays a dominant role in this coupling. The soil moisture-precipitation feedback is already apparent in the first year (Figure 7). Both ICON-C4 and IFS_F-C4 simulations generally reproduce similar patterns of where strong and weak soil moisture–precipitation feedback occurs within the first five years. This is quantified by the correlation coefficient between total-column soil moisture and precipitation during the boreal summer. The two models agree with each other on the regions with relatively strong feedback, such as Mexico,
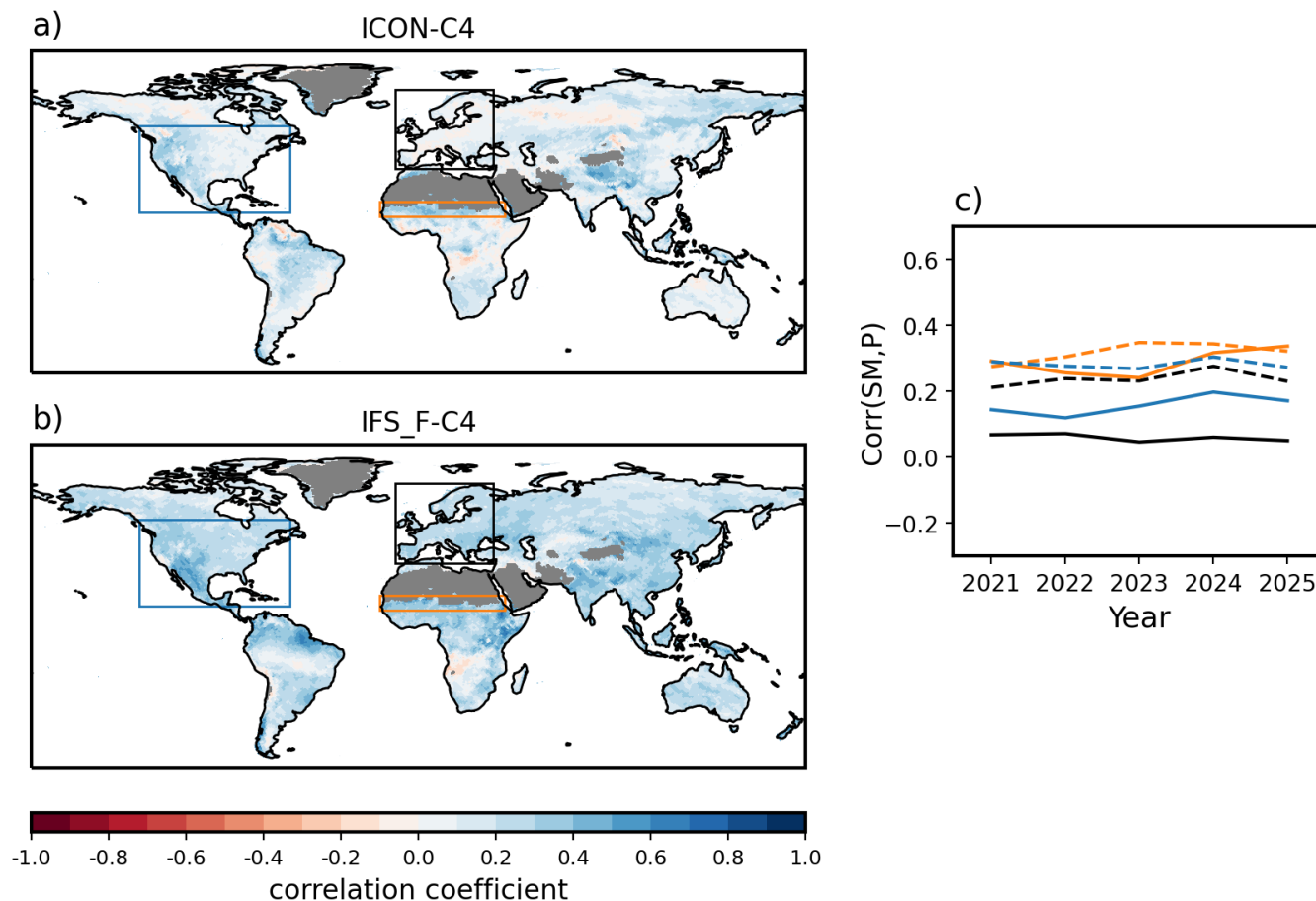
19

Figure 7. Soil moisture–precipitation feedback quantified with the correlation between daily mean total-column soil moisture (SM) and precipitation (P) during JJA of 2021 to 2025 for ICON-C4 (a) and IFS_F-C4 (b). Time series of the correlation coefficient (c) in ICON-C4 (full lines) and IFS_F-C4 (dashed lines) averaged over Europe [black box in (a) and (b)], Sahel [orange box in (a) and (b)], and the United States and Mexico [blue box in (a) and (b)]. Areas where the precipitation is smaller than $0.1\,\mathrm{mm\,d^{-1}}$ in both simulations are grayed out.

the southern United States and the Sahel. These strong feedback regions align well with the hot spots of land-atmosphere coupling where evapotranspiration is strongly controlled by soil moisture (Koster et al., 2004). The models do not share all the

425   features. ICON-C4 generally shows weaker feedback strength globally than IFS_F-C4. The reason for the weaker feedback can be attributed, in part, to how each model represents convection (Lee and Hohenegger, 2024). Convective parameterization causes precipitation to be more sensitive to surface evapotranspiration, leading to stronger feedback compared to the results with explicit convection (Hohenegger et al., 2009). The time series of land-atmosphere feedback over Europe, the Sahel, the United States, and Mexico show that the correlation coefficients established in the first year remain relatively stable over time,

430   with only minor fluctuations due to interannual variability.

4.3.2 Meso-scale phenomena

As an example of a meso-scale phenomenon, we look into maritime stratocumulus cloud fields, typically occurring over cool currents off the western side of continents.

Of relevance for the representation of stratocumulus clouds in nextGEMS models were the following specifications. First of all, we note that ICON-C4 refuses any parameterization of convective clouds, while IFS_F-C4 employs a shallow convection scheme as well as a modified deep convection scheme. It was already discussed in Section 4.1 that for both ICON-C4 and IFS_F-C4 the *cloud inhomogeneity factor* (ζ), which addresses the subgrid-scale inhomogeneities of clouds, was adjusted under the assumption that on the finer grids, greater inhomogeneity is explicitly represented. For IFS_F-C4 also a parameter controlling cloud edge erosion was reduced in order to increase cloud cover of low clouds. Both of these parameters affect cloud-radiation interaction and, as such, impact the radiation balance at the top-of-atmosphere, showing stratocumulus clouds as a good example of how microscale processes impact global scales.

In Figure 8a, we show the annual cycle of cloud water path for IFS_F-C4 and ICON-C4 and cloud area fraction for IFS_F-C4 and CERES satellite observations, averaged over the ocean near the coast of California, one of the main stratocumulus regions (Klein and Hartmann, 1993). Figure 8b shows albedo, given as the fraction of top-of-atmosphere outgoing shortwave radiative flux (RSUT) and incoming shortwave radiative flux, as well as RSUT individually. Note that for the simulations, we average over the years between 2021 and 2025, and for observations between 2000 and 2023, meaning that the periods hardly overlap. We recognize from the cloud properties and albedo two maxima in CERES observations, one in summer and one in winter. The annual cycle in albedo is well represented in ICON-C4, showing agreement in both summer and winter peaks. Disregarding the systematic differences of cloud water path, we find that particularly the summer peak is well represented in ICON-C4. In turn, IFS_F-C4 simulates a too-early peak in May and one in winter. Looking at RSUT, we find that the summer peak dominates the radiative effects of the Californian stratocumulus clouds. ICON-C4 simulates the annual cycle well, while IFS_F-C4 does not reach the summer peak amplitude in RSUT, which is in line with the absent summer peak in the cloud water path.

From the annual cycles shown in Figure 8 we identify summer (JJA) as the radiatively most relevant season of marine stratocumulus off the coast of California. In Figure 9 we show the associated spatial maps of RSUT for JJA along with the cloud properties and RSUT biases, here averaged over the 2021-2025 period for both simulations and the 2016-2020 period for CERES. For ICON-C4, but particularly for IFS_F-C4, RSUT is underestimated. In CERES, the stratocumulus cloud field off the coast of California, as shown in Figure 9c, can be seen in cloud fraction and relates clearly to RSUT. For ICON-C4 and IFS_F-C4, we find that clouds have too small a liquid water path or are too far off the coast. As a consequence, associated biases in RSUT are substantial.

We summarize that the nextGEMS simulations' energetically consistent climate comes along with a stratocumulus cloud field in the Bay of California whose radiative effects (RSUT) are well in phase with that of CERES satellite observations. In ICON-C4, we find both the bi-modality of the annual cycle of cloudiness and albedo, as well as the peak amplitude of RSUT, matching well with observations. For IFS_F-C4, just like in CMIP5 and CMIP6 (Jian et al., 2021), the peak amplitude of RSUT is underestimated, particularly in summer, when stratocumulus clouds have their strongest radiative effects. The reasons
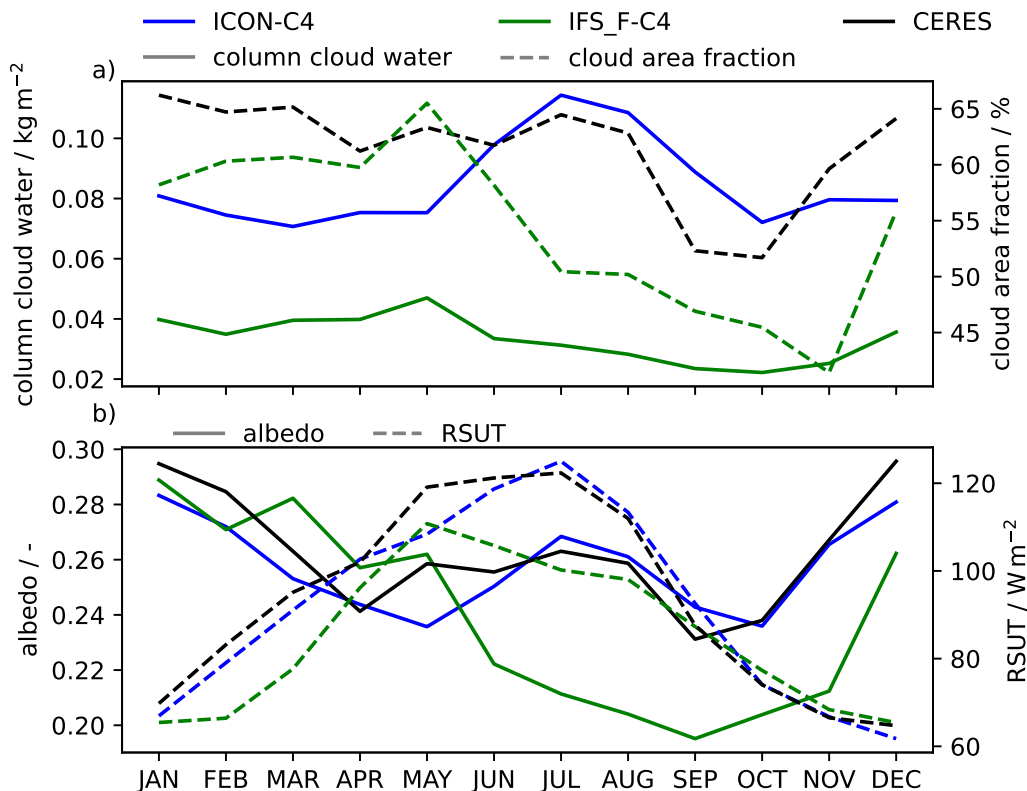
Figure 8. Annual cycle of stratocumulus cloud and radiative properties, averaged of 2021 to 2025 and over the ocean west of the Californian coast (longitude [115°W, 127.5°W] and latitude [27.5°N, 38°N], blue rectangle in Figure 9. Panel a) shows cloud (liquid) water path (kg m^{-2}) and cloud area fraction (%). Panel b) shows the the albedo and top-of-atmosphere upward radiative flux (RSUT / Wm^{-2}). Note that for ICON-C4 cloud fraction is not available, while for CERES column cloud water is not available.

465   for this are the absence or ocean-wards dislocation of stratocumulus, potentially inaccurate cloud-radiative properties, and under-resolved mixing processes of stratocumulus. Eventually, it would be of great interest to investigate the implementation of aerosol effects on cloud formation and brightening in km-scale Earth system models in regard to stratocumulus clouds and their radiative effects.

"Whether a realistic representation of global climate constrains patterns of regional phenomena?" can not be fully answered
470   because a proof of causal relationship between the global climate and regional circulation is not feasible. Still, the representation of the exemplary regional circulation, stratocumulus cloud fields, is found satisfying. In particular, for ICON-C4, despite not employing a parameterization of shallow convection, we find that not only the phase but also the amplitude of stratocumulus cloud radiative effects agree very well with observations (Nowak et al., 2021). This is an encouraging result that demonstrates the capabilities of km-scale Earth system models, as well as that instead of using convective parameterization the fine-tuning
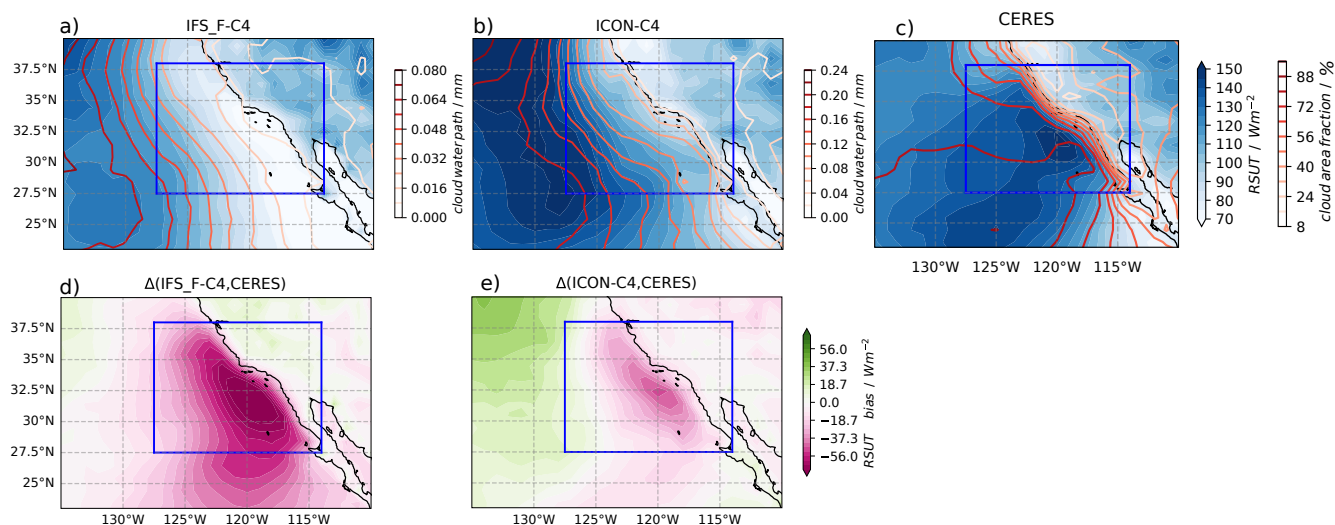475   of cloud microphysics is key to representing stratocumulus clouds.

Figure 9. Top-of-atmosphere upwelling shortwave flux (RSUT) over the ocean next to the Californian coast in JJA for ICON-C4 (a), IFS_F-C4 (b) and CERES (Loeb et al., 2018) (c), along with the cloud (liquid) water path ($kg\,m^{-2}$) for nextGEMS Cycle 4 models [2021; 2025] and cloud area fraction (%) for CERES [2016; 2020]. Panels (d) and (e) show biases in RSUT of ICON-C4 and IFS_F-C4 relative to CERES, respectively. The blue rectangle denotes the area considered for computing RSUT statistics in Figure 8.

4.3.3 Synoptic-scale phenomena

Atmospheric blocking is a key feature of mid-latitude synoptic-scale circulation, often linked to weather extremes such as heatwaves and cold spells. Here, we use atmospheric blocking as an example of a synoptic phenomenon. Figures 10a,b illustrates the climatology of blocking events (Figures 10a,b) and the 2021–2025 time series of blocking events for the "North Pacific" and "Central Europe" regions (Figure 10c). Overall, there is good agreement in the geographical distribution of atmospheric blocking between ICON-C4 and IFS_F-C4. However, IFS_F-C4 exhibits a higher frequency of blocking events compared to ICON-C4.

We hypothesize that the convergence in blocking location results between these models is influenced by their shared utilization of kilometer-scale orographic information. As demonstrated by Davini et al. (2022) using coarser-resolution simulations, higher-resolution orography significantly enhances the representation of atmospheric blocking. Specifically, the improved performance of IFS-FESOM in capturing blocking frequency can be attributed to its implementation of turbulent orographic form drag and subgrid-scale orography parametrizations, both of which are known to enhance the representation of atmospheric circulation features (e.g., Woollings et al., 2018).

The time series of annual blocking frequencies for the North Pacific and Central Europe regions reveal substantial interannual variability in both IFS_F-C4 and ICON-C4. Note that recent studies indicate that there is no significant trend in blocking frequency over the past decades (Wazneh et al., 2021). Traditional climate models tend to underestimate atmospheric block-
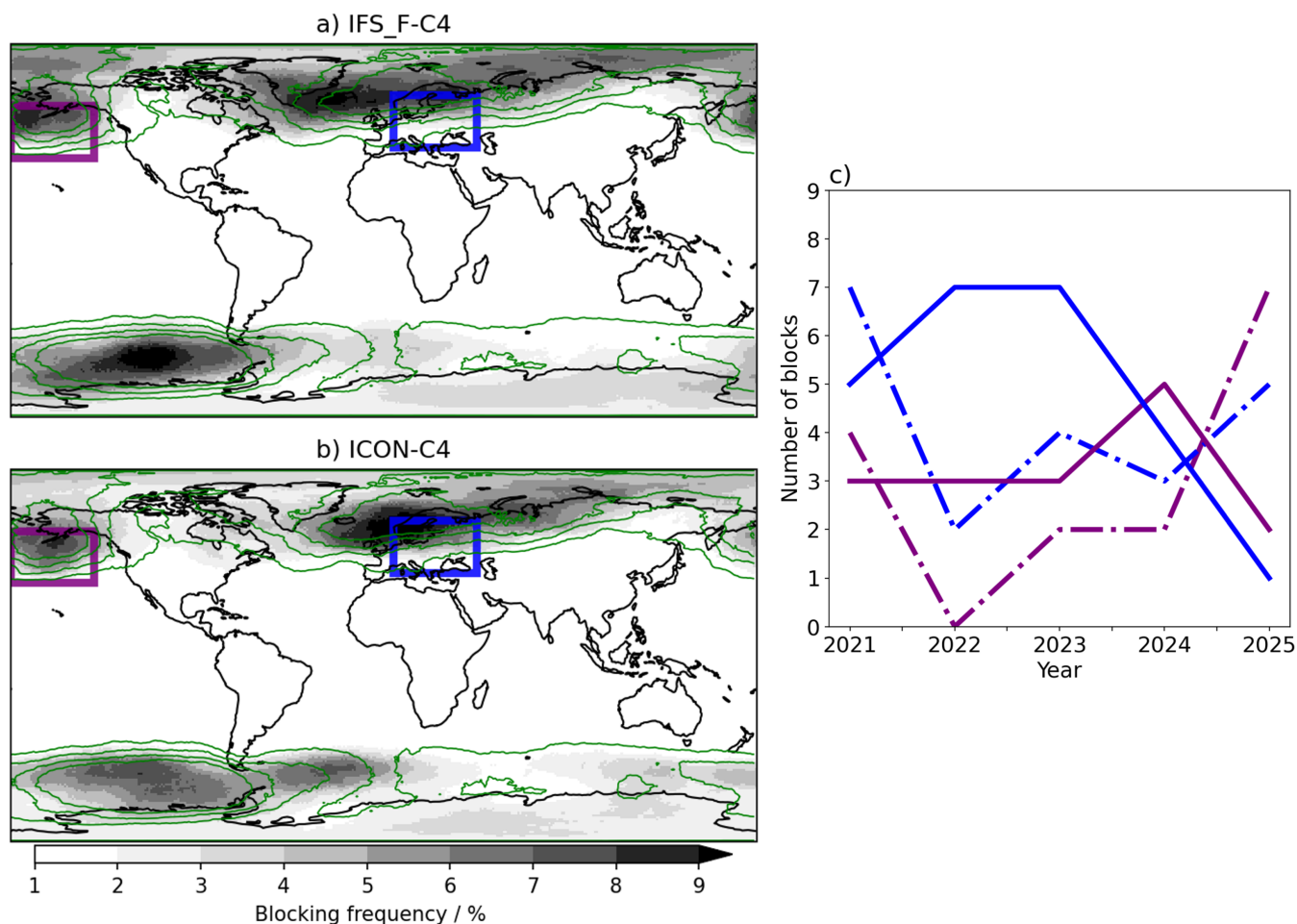
Figure 10. Annual blocking frequency in IFS_F-C4 (a) and ICON-C4 (d) during 2021–2025. Units are in percentage of blocked days relative to the total number of days per year. ERA5 climatology (2015–2019) is shown in green contours (2% intervals starting at 2%). Evolution of the number of blocks (c) in ICON-C4 (full lines) and IFS_F-C4 (dashed lines) passing over central Europe [blue box in (a) and (b)] and and North Pacific [purple box in (a) and (b)]. Blocking is identified as a persistent and quasi-stationary mid-level (500 hPa) geopotential height anomaly of the flow following the Schwierz index (Schwierz et al., 2004).

ing frequency (e.g., Dolores-Tesillos et al., 2024); thus, whether blocking will increase or decrease is still an open question (Berckmans et al., 2013; Davini and D'Andrea, 2016; Schiemann et al., 2017; Woollings et al., 2018).

4.4 Time scales of regional patterns

495   The analysis in Section 4.1 brought another positive surprise: The TOA energy balance responded quite quickly to changes in the parameterization schemes. The tuning of the TOA energy balance shown in Section 4.1 was achieved in parallel in ICON

and IFS-FESOM by performing a number of short integrations. In the process, selected parameter values were modified within observational uncertainty to obtain a simulated TOA energy balance similar to the observed one. Indeed, the integration time was no longer than 12 days. This strategy was further used in IFS for testing across different resolutions. This means that with only a few days of km-scale simulations, one can build enough confidence on whether changes in parameter values converge to an improved response. Such a finding reinforces the possibilities of these km-scale models to be used more routinarily and in an operational context.

In the same line, robust regional circulation patterns in Section 4.2, Section 4.3.1 and Section 4.3.2 emerge within relatively short timespans, e.g. one-year periods. The double band of precipitation in the Western Pacific in ICON-C4 is present every year of the simulation and is already observed in the first 3 months of the simulation (not shown). The annual cycle of stratocumulus clouds is established already in the first year (Figure 8). Likewise, the soil moisture-precipitation feedback emerges already during the first year of simulations (Figure 7). The lesson that we want to communicate here is that it is unnecessary to run lengthy simulations to understand how parameterization changes will affect the TOA energy balance, the land-atmospheric coupling, the double ITCZ, and stratocumulus clouds. Thus, this lesson could be helpful for communities working on those problems. On the other hand, the large interannual variability on atmospheric blockings both in ICON and IFS-FESOM combined with the few occurrences of blocking in a given year (Fig. 10) suggests that multidecadal simulations are required to assess the performance and evolution of blocking frequencies properly.

5   Conclusions

In 2021, nextGEMS started with the goal of producing for the first time multidecadal climate simulations, in which the governing equations are solved at a horizontal resolution on the order of 10 kilometers or finer in the ocean, land, and atmosphere. With the resources available, nextGEMS provided 30-year climate simulations (from 2020 to 2050) under the SSP3-7.0 scenario with two different Earth system models, ICON and IFS-FESOM, with a horizontal grid spacing of 10 km in the atmosphere and 5 km in the ocean. The 30-year climate simulations were run on the supercomputer Levante, using 18% of its capacity in the case of ICON and 9% in the case of IFS-FESOM. While the limited computing resources forced the final horizontal grid spacing to be 10 km, nextGEMS prepared ICON and IFS-FESOM to be run on more powerful supercomputers than Levante, exploiting computer resources to produce climate simulations with a horizontal grid spacing of 5 km. The achievement of the 30-year climate simulations inspired this overview. Here, we presented the concept and progress of the project structured into four cycles and hackathons. In addition, we discussed the surprises we encountered and the lessons we learned when developing the models and analyzing the simulations. To create a storyline, we translated our learning process into four questions defined in Section 1.

Over the four cycles, the simulations with ICON and IFS-FESOM evolved from year-long simulations in Cycle 1 with significant mass and energy leaks to multi-decadal simulations in Cycle 4 with an energetically consistent climate, i.e., the response of the surface temperature to the radiative forcing was comparable to observations. The two 30-year simulations of Cycle 4, with ICON and IFS-FESOM, had a horizontal grid spacing of about $10\,\mathrm{km}$ in the atmosphere and land and $5\,\mathrm{km}$ in

530 the ocean. Notably, both simulations had a competitive throughput of 414 simulated days per day on 464 nodes for ICON and 600 simulated days per day on 269 nodes for IFS-FESOM.

In both models, ICON and IFS-FESOM, it was relatively easy to tune the cloud properties and bring the TOA radiation balance close to observations. The main reason was the quick response of the TOA radiation balance to changes in the cloud properties, allowing us to find optimal parameters with short simulations of about 12 days. The cloud properties were adjusted

535 in different ways in ICON and IFS-FESOM. While ICON targeted shallow and low-level stratocumulus clouds, IFS-FESOM targeted all types of clouds. Despite these different approaches, the bias in the TOA radiation balance could be reduced in both models, with a closer fit in IFS-FESOM.

The comparison between ICON and IFS-FESOM also gave insights into the representation of the mean climate. A km-scale model with no convection parameterization like ICON can reproduce the observed structure of the tropical rainbelt

540 over land in the Eastern Pacific, similar to IFS-FESOM, which uses a convection parameterization. However, ICON can not reproduce the pattern of the tropical rainbelt (still showing a double ITCZ) or the SST patterns in the western Pacific. Here, IFS-FESOM behaves more similarly to observations. Thus, solving explicitly km-scale processes with a grid spacing of 10 km is not sufficient to solve biases in the western Pacific, but fine-tuning in km-scale Earth system models can reduce the double ITCZ bias (Takasuka et al., 2024).

545 We also observed in ICON and IFS-FESOM that the seasonal cycle of stratocumulus clouds, and pattern of atmospheric blocking are well represented, i.e., similar to observations. This means that these regional phenomena are well constrained in km-scale climate simulations. However, there are some differences. ICON reproduces the seasonal cycle of the stratocumulus clouds better than IFS-FESOM. The different pathways for fine-tuning radiative properties of low-level clouds explain the discrepancy between ICON and IFS-FESOM. We also observe that the use of convective parameterization is not necessary

550 to represent this regional phenomenon correctly. On the other hand, the frequency of atmospheric blocking occurrence is better represented by IFS-FESOM, signaling the importance of sub-scale processes to represent the frequency of this synoptic circulation.

Our analysis also showed that short simulations of one year or even less are sufficient to develop and test many aspects of km-scale models such as biases in the tropical rainbelt, in the top-of-atmosphere radiation balance, the seasonal cycle of

555 stratocumulus clouds, or the soil moisture–precipitation feedback. This finding is particularly encouraging for other groups who work on resolving these biases or understanding the changes in regional circulation patterns under different climate scenarios. The resources for such short simulations are available on any of the EuroHPC systems, which brings us one step closer to the vision of using km-scale models in a research and operational setting.

6 Outlook

560 The development of ICON and IFS-FESOM continues within and beyond nextGEMS. To further increase the resolution and throughput, modeling centers are increasingly making use of exascale supercomputers such as the pan-European Large Unified Modern Infrastructure (LUMI, 2024). ICON, and to a lesser extent IFS-FESOM, have been adapted to perform well on such

machines primarily based on graphical processing units (Adamidis et al., 2024; Bishnoi et al., 2024). nextGEMS collaborates with other projects, such as WarmWorld (https://warmworld.de), on the optimization and modularization of the code base of

565   ICON, and with EERIE (https://eerie-project.eu), on the role of mesoscale processes in the ocean.

To improve the representation of the Earth system, ICON is currently including more components, such as the HAMOCC module for ocean biogeochemistry (Ilyina et al., 2013) and the HAM-lite module for interactive aerosols (Weiss et al., 2024). HAMOCC and HAM-lite have been evaluated with complementary simulations of up to one year. In preliminary simulations with ICON-HAMOCC at 10 kilometer resolution, biochemistry patterns at the regional scale appeared well reproduced in the

570   case of West African waters. On the other hand, sea surface chlorophyll-a begins to be reconstructed (Roussillon et al., 2023), which can be used to forecast marine productivity and its effect on shifts of exploited fish populations (Sarre et al., 2024) as well as trends in pelagic biomass (Diogoul et al., 2021). In simulations with ICON-HAM-lite at 5 kilometer resolution, key aerosol processes are captured including, for example, the formation of dust storms in the Sahara, wind-driven emissions of sea salt aerosols, or wilfire-driven emissions of carbonaceous aerosols. Figure 11 shows a scene of aerosol burdens on 5

575   September 2020 at 00:00 UTC. These examples show how the simulations produced so far, and those to come can be of interest for industry sectors such as solar and wind energy or agriculture and fisheries.

Related to this operational aspect, the two models developed in nextGEMS are also prototypes for the Climate Change Adaptation Digital Twin of the Destination Earth initiative (https://destination-earth.eu). The initiative aims to operationalize km-scale and multi-decadal climate simulations, assess the impacts of climate change, and evaluate adaptation strategies at

580   local and regional scales (Hoffmann et al., 2023; Sandu, 2024). For that initiative, the first projections from 2020 to 2040 were produced with ICON, IFS-NEMO, and IFS-FESOM at resolutions of $5\,\mathrm{km}$ and $4.4\,\mathrm{km}$ in the atmosphere and $5-10\,\mathrm{km}$ in the ocean, respectively. Moreover, the work on output harmonization in nextGEMS fed back into Destination Earth. The output of the models underpinning the Digital Twin is converted into a Generic State Vector, a set of selected variables in consistent units, frequencies, and on a common grid, which is then streamed to the different applications.

585   nextGEMS is the first step towards a European ecosystem of km-scale climate research and operational Earth system models, to which WarmWorld, EERIE, Destination Earth, and other projects contribute. These projects embody a cross-European effort to push forward the boundaries of climate science by reducing model uncertainties and providing climate information on local and regional scales where the impacts of climate change are felt (Prein et al., 2015; Gettelman et al., 2023). In this context, the Earth Virtualization Engine initiative (https://eve4climate.org) understood the necessity to provide fair free access to climate

590   information with local granularity, globally, using the best technology at hand (Stevens et al., 2024). Thus, the km-scale Earth system models developed in nextGEMS will not only serve scientific purposes but will also be tools to address critical questions for society and ecosystems regarding climate adaptation, mitigation, and risk management.

Code availability. The ICON model is available on the World Data Center for Climate, https://doi.org/10.35089/WDCC/IconRelease01, (ICON-partnership et al., 2024) under a permissive open source licence (https://opensource.org/license/BSD-3-clause). The FESOM2.5

595   model is a free software and available on Github (https://github.com/FESOM/fesom2). The latest version 2.5 including all developments used
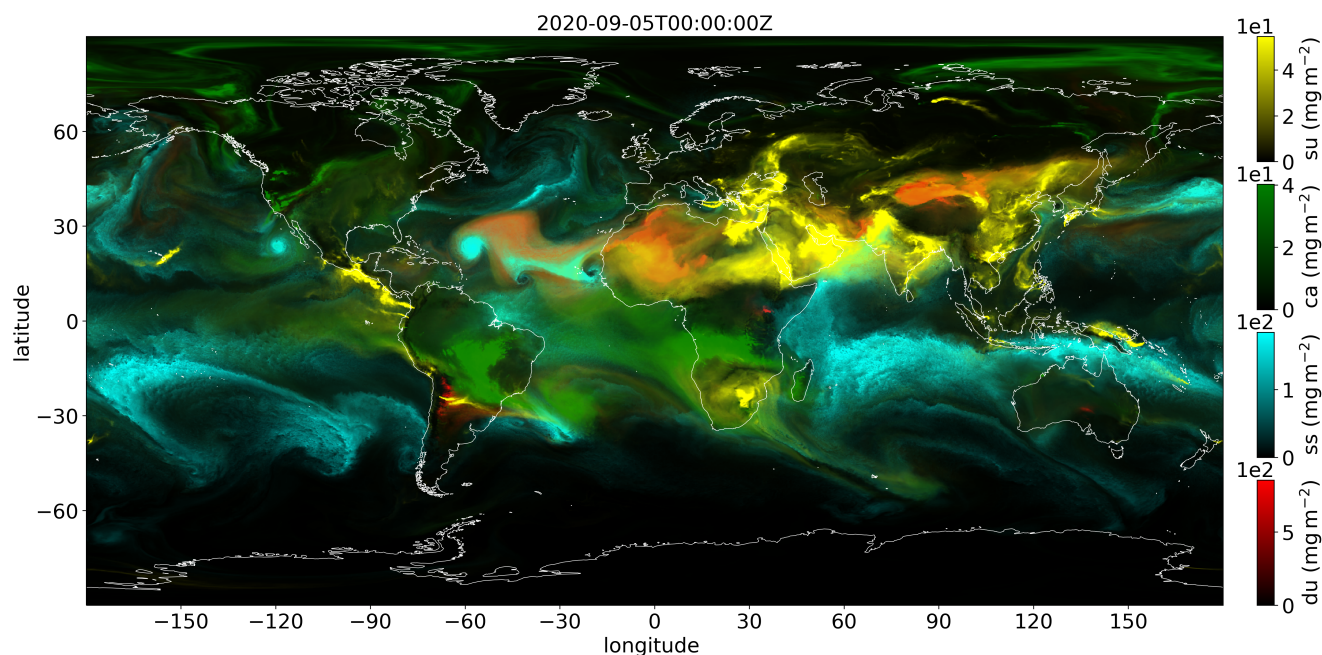
Figure 11. Scene of aerosols on 5 September 2020 at 00:00 UTC. Shown are column burdens of dust (du) in red, sea salt (ss) in blue, carbonaceous aerosol (ca) in green, and sulfuric aerosol (su) in yellow. The colormaps have a variable transparency which decreases from fully transparent at minima to fully opaque at maxima.

in nextGEMS Cycle 3 is archived on Zenodo, https://doi.org/10.5281/zenodo.10225420, (Rackow et al., 2023b). MultIO is free software and available on the GitHub of ECMWF (https://github.com/ecmwf). The IFS model is available subject to a licence agreement with ECMWF. ECMWF member-state weather services and approved partners will be granted access. The IFS code without modules for data assimilation is available for educational and academic purposes via an OpenIFS licence (http://www.ecmwf.int/en/research/projects/openifs). IFS code

600   modifications for nextGEMS have been separately archived on Zenodo, https://doi.org/10.5281/zenodo.10223576, (Rackow et al., 2023a). The data and scripts that we used to generate the figures are in https://doi.org/10.17617/3.QZHXMC (Segura, 2025).

Appendix A: Summary of model developments

Here, we summarize all developments in ICON over cycle 2 to 4.

610 **A1 Cycle 2**

In the atmosphere module, several bugs in the dynamical core, cloud microphysics, and surface fluxes were removed, which reduced the energy imbalance significantly. The old PSrad radiation scheme was replaced with the new RTE-RRTMGP scheme (Pincus et al., 2019). A cloud inhomogeneity factor was introduced to tune the radiation balance at the top of the atmosphere. In the land module, the GSWP3 input data of 2014 (Dirmeyer et al., 2006) was used to spin up the land reservoir for the hydrolog-
615 ical discharge. The equilibrium was reached after 5 years using a time step of $40\,\mathrm{s}$. The distribution area of freshwater entering the ocean was increased to a radius of $30\,\mathrm{km}$. In the ocean module, a new vertical coordinate system with thinner surface levels was introduced, and a new surface flux scheme was implemented. Moreover, the spin-up procedure was revised. The spin-up simulation was initialized with ORAS5 (Copernicus Climate Change Service, 2025) and forced with ERA5 (Hersbach et al., 2020) from 2010 to 2020 to establish a stationary eddy field. The sea surface temperature was nudged to ERA5 in the last year.
620 After 8 test runs, two simulations were provided. ICON-C2-A used a horizontal grid spacing of 5 km and was integrated for 2 years, whereas ICON-C2-B used a horizontal grid spacing of 10 km and was integrated for 10 years. ICON-C2-A and ICON-C2-B were both performed on on 400 nodes. The throughput was 80 SDPD for ICON-C2-A and 550 SDPD for ICON-C2-B. Coarsening the resolution by a factor of 2 increased the throughput by a factor of 7.

A2 Cycle 3

625 In the atmosphere module, the energy imbalance was further reduced by removing bugs in the cloud microphysics and turbulence scheme. The surface fluxes were calculated with the heat capacity at constant pressure instead of constant volume, which reduced the amount of energy transferred from the surface to the atmosphere by about $29\,\%$. In the land module, the heat capacity and conductivity maps were revised and the soil texture was accounted for in the hydrology scheme. In the ocean module, the vertical mixing was solved including Langmuir turbulence in the turbulent kinetic energy scheme. The turbulent
630 diffusion coefficient was decreased from 0.2 to 0.1. The tracer advection was treated with a new second-order scheme. The spin-up of the ocean was conducted in the same manner as in Cycle 2 except for the year 2019 being repeated. The dynamical core was solved with mixed precision, and the sea level pressure was coupled to the atmosphere. In contrast to the previous cycles, the output was compiled on the HEALpix grid (Górski et al., 2005). After 27 test runs, ICON-C3 was integrated over 5 years using a horizontal grid spacing of 5 km in the ocean and atmosphere. The throughput of ICON-C3 was 98 SDPD on 530
635 nodes compared to 80 SDPD on 400 nodes in ICON-C2-A.

A3 Cycle 4

In the atmosphere module, the diffusivity in the conservation of momentum was modified to account for the persistent leak of internal energy. The turbulent mixing scheme VDIFF was updated and renamed to TMX. This update came with several

key improvements: refactoring the code into an object-oriented structure, enabling both explicit and implicit numerical solvers,
640   replacing the implicit atmosphere-surface coupling with an explicit coupling, and replacing the dry static energy with the
internal energy in the diffusion scheme. Lastly, the cloud inhomogeneity factor was linked with the lower tropospheric stability
to better account for the cloud type. In the ocean module, the number of vertical levels was decreased from 128 to 72 levels,
while keeping thin layers in the upper ocean, and the turbulent mixing under sea ice was reduced. The ocean state spin-up was
conducted in the same manner as in Cycle 2 but without nudging of the sea surface temperature. ICON-C4 used ozone and
645   greenhouse gas concentrations following CMIP6 for the forcing time series of a SSP3-7.0 scenario (O'Neill et al., 2016). In
addition, ICON-C4 used the time-varying MACv2-SP climatology (Stevens et al., 2017) for anthropogenic aerosols together
with the Stenchikov climatology from 1999 for volcanic aerosols (Stenchikov et al., 1998). After numerous long test runs,
ICON-C4 was integrated over 30 years using a horizontal grid spacing of 10 km in the atmosphere and 5 km in the ocean. The
10 km grid spacing of the atmosphere, allowed us to reach a throughput of 414 SDPD on 464 nodes.

650   *Author contributions.*   HS, XP, PW, SM, TR, JL, ED, and IB analyzed the data, prepared the figures, and drafted the paper. MA, RA, GA,
AB, JB, SB, EB, TB, SB, HB, NB, LB, SC, SD, JD, ID, PD, ME, JE, ME, RF, CF, LF, DG, PG, PG, AG, KG, MG, OG, HH, IH, KH, SH,
JH, LK, AK, NK, TK, SK, SK, JK, PK, AK, RL, NM, MM, SM, KM, BN, JN, DP, UP, DP, RR, DS, DS, RS, PS, DS, DS, BS, DT, AT, AU,
MV, MV, AV, SW, FW, MW, NW, KW, JW, MW, YW, FZ, and JZ contributed to the project and revised the paper. FB, DB, SB, SB, SB,
PB, MD, ED, SF, EF, CH, CH, HJ, MJ, TJ, JJ, NK, DK, HK, MK, SM, OM, TM, JM, TM, EM, HP, KP, AS, PS, PS, LT, PV, IS, and BS
655   conceptualized and contributed to the project and revised the paper.

Competing interests.   Chiel van Heerwaarden is a member of the editorial board of the Geoscientific Model Development (GMD) journal.

References

Adamidis, P., Pfister, E., Bockelmann, H., Zobel, D., Beismann, J.-O., and Jacob, M.: The Real Challenges for Climate and Weather Mod-
elling on its Way to Sustained Exascale Performance: A Case Study using ICON (v2.6.6), Geoscientific Model Development Discussions,
2024, 1–23, https://doi.org/10.5194/gmd-2024-54, 2024.

665 Baker, A. J., Vannière, B., and Vidale, P. L.: On the Realism of Tropical Cyclone Intensification in Global Storm-Resolving Climate Models,
Geophysical Research Letters, 51, e2024GL109 841, https://doi.org/10.1029/2024GL109841, 2024.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numer-
ical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139, 3887 – 3905,
https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

670 Bao, J., Stevens, B., Kluft, L., and Muller, C.: Intensification of daily tropical precipitation extremes from more organized convection, Science
Advances, 10, eadj6801, https://doi.org/10.1126/sciadv.adj6801, 2024.

Becker, T., Bechtold, P., and Sandu, I.: Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations,
Quarterly Journal of the Royal Meteorological Society, 147, 4388–4407, https://doi.org/10.1002/qj.4185, 2021.

Berckmans, J., Woollings, T., Demory, M.-E., Vidale, P.-L., and Roberts, M.: Atmospheric blocking in a high resolution climate model:
675 influences of mean state, orography and eddy forcing, Atmospheric Science Letters, 14, 34–40, https://doi.org/10.1002/asl2.412, 2013.

Birch, C. E., Roberts, M. J., Garcia-Carreras, L., Ackerley, D., Reeder, M. J., Lock, A. P., and Schiemann, R.: Sea-Breeze Dynamics and
Convection Initiation: The Influence of Convective Parameterization in Weather and Climate Model Biases, Journal of Climate, 28, 8093
– 8108, https://doi.org/10.1175/JCLI-D-14-00850.1, 2015.

Bishnoi, A., Stein, O., Meyer, C. I., Redler, R., Eicker, N., Haak, H., Hoffmann, L., Klocke, D., Kornblueh, L., and Suarez, E.: Earth system
680 modeling on modular supercomputing architecture: coupled atmosphere–ocean simulations with ICON 2.6.6-rc, Geoscientific Model
Development, 17, 261–273, https://doi.org/10.5194/gmd-17-261-2024, 2024.

Boussetta, S., Balsamo, G., Arduini, G., Dutra, E., McNorton, J., Choulga, M., Agustí-Panareda, A., Beljaars, A., Wedi, N., Munõz-Sabater,
J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D., and Zsoter, E.: ECLand: The ECMWF
Land Surface Modelling System, Atmosphere, 12, https://doi.org/10.3390/atmos12060723, 2021.

685 Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Bell, T. L., and Snider, J. B.: The Albedo of Fractal Stratocumulus Clouds, Journal of
Atmospheric Sciences, 51, 2434 – 2455, https://doi.org/10.1175/1520-0469(1994)051<2434:TAOFSC>2.0.CO;2, 1994.

Chassignet, E. P., Yeager, S. G., Fox-Kemper, B., Bozec, A., Castruccio, F., Danabasoglu, G., Horvat, C., Kim, W. M., Koldunov, N., Li,
Y., Lin, P., Liu, H., Sein, D. V., Sidorenko, D., Wang, Q., and Xu, X.: Impact of horizontal resolution on global ocean–sea ice model
simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), Geoscientific Model
690 Development, 13, 4595–4637, https://doi.org/10.5194/gmd-13-4595-2020, 2020.

Copernicus Climate Change Service: ORAS5 global ocean reanalysis monthly data from 1958 to present, https://cds.climate.copernicus.eu/
datasets/reanalysis-oras5, 2025.

Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., Jung, T., and Schröter, J.: Finite-Element Sea Ice Model
(FESIM), version 2, Geoscientific Model Development, 8, 1747–1761, https://doi.org/10.5194/gmd-8-1747-2015, 2015.

695 Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volumE Sea ice–Ocean Model (FESOM2), Geoscientific Model Development,
10, 765–789, https://doi.org/10.5194/gmd-10-765-2017, 2017.

Davini, P. and D'Andrea, F.: Northern Hemisphere atmospheric blocking representation in global climate models: twenty years of improvements?, Journal of Climate, 29, 8823–8840, https://doi.org/10.1175/JCLI-D-16-0242.1, 2016.

Davini, P., Fabiano, F., and Sandu, I.: Orographic resolution driving the improvements associated with horizontal resolution increase in the Northern Hemisphere winter mid-latitudes, Weather and Climate Dynamics, 3, 535–553, https://doi.org/10.5194/wcd-3-535-2022, 2022.

Diogoul, N., Brehmer, P., Demarcq, H., El Ayoubi, S., Thiam, A., Sarre, A., Mouget, A., and Perrot, Y.: On the robustness of an eastern boundary upwelling ecosystem exposed to multiple stressors, Scientific Reports, 11, 1908, https://doi.org/10.1038/s41598-021-81549-1, 2021.

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., and Brdar, S.: Large eddy simulation using the general circulation model ICON, Journal of Advances in Modeling Earth Systems, 7, 963–986, https://doi.org/10.1002/2015MS000431, 2015.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface, Bulletin of the American Meteorological Society, 87, 1381 – 1398, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

Dolores-Tesillos, E., Martius, O., and Quinting, J.: On the role of moist and dry processes for atmospheric blocking biases in the Euro-Atlantic region in CMIP6, EGUsphere, 2024, 1–25, https://doi.org/10.5194/egusphere-2024-2878, 2024.

ECMWF: IFS Documentation CY46R1 - Part IV: Physical Processes, 4, ECMWF, https://doi.org/10.21957/xphfxep8c, 2019.

ECMWF: IFS Documentation CY47R3 - Part IV Physical processes, 4, ECMWF, https://doi.org/10.21957/eyrpir4vj, 2021.

ECMWF: IFS Documentation CY48R1 - Part III: Dynamics and Numerical Procedures, 3, ECMWF, https://doi.org/10.21957/26f0ad3473, 2023a.

ECMWF: IFS Documentation CY48R1 - Part IV: Physical Processes, 4, ECMWF, https://doi.org/10.21957/02054f0fbf, 2023b.

ECMWF: Atos Sequana XH2000, https://www.ecmwf.int/en/computing/our-facilities/supercomputer-facility, 2024.

Fielding, M., Schäfer, S., Hogan, R., and Forbes, R.: Parametrizing cloud geometry and its application in a subgrid cloud-edge erosion scheme, Q. J. R. Meteorol. Soc., 146, 1651–1667, https://doi.org/10.1002/qj.3758, 2020.

Gent, P. R. and Mcwilliams, J. C.: Isopycnal Mixing in Ocean Circulation Models, Journal of Physical Oceanography, 20, 150 – 155, https://doi.org/10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2, 1990.

Gent, P. R., Willebrand, J., McDougall, T. J., and McWilliams, J. C.: Parameterizing Eddy-Induced Tracer Transports in Ocean Circulation Models, Journal of Physical Oceanography, 25, 463 – 474, https://doi.org/10.1175/1520-0485(1995)025<0463:PEITTI>2.0.CO;2, 1995.

Gentry, M. S. and Lackmann, G. M.: Sensitivity of Simulated Tropical Cyclone Structure and Intensity to Horizontal Resolution, Monthly Weather Review, 138, 688 – 704, https://doi.org/10.1175/2009MWR2976.1, 2010.

Gettelman, A., Fox-Kemper, B., Flato, G., Klocke, D., Stamer, D., Stevens, B., and Vidale, P. L.: Kilometre-Scale Modelling of the Earth System: A New Paradigm for Climate Prediction, https://wmo.int/media/magazine-article/kilometre-scale-modelling-of-earth-system-new-paradigm-climate-prediction, 2023.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M.: HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, The Astrophysical Journal, 622, 759, https://doi.org/10.1086/427976, 2005.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., Wittenberg, A. T., Yin, J., and Zhang, R.: Impacts on Ocean Heat from Transient Mesoscale Eddies in a Hierarchy of Climate Models, Journal of Climate, 28, 952 – 977, https://doi.org/10.1175/JCLI-D-14-00353.1, 2015.

Hanke, M., Redler, R., Holfeld, T., and Yastremsky, M.: YAC 1.2.0: new aspects for coupling software in Earth system modelling, Geoscientific Model Development, 9, 2755–2769, https://doi.org/10.5194/gmd-9-2755-2016, 2016.

735   Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049,
740   https://doi.org/10.1002/qj.3803, 2020.

Hewitt, H., Fox-Kemper, B., Pearson, B., Roberts, M., and Klocke, D.: The small scales of the ocean may hold the key to surprises, Nature Climate Change, 12, 496–499, https://doi.org/10.1038/s41558-022-01386-6, 2022.

HLRE-4 Levante: Levante HPC System, https://www.dkrz.de/en/systems/hpc/hlre-4-levante, 2024.

Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., and Thiemert, D.: Destination Earth – A digital twin in support of climate services,
745   Climate Services, 30, 100 394, https://doi.org/10.1016/j.cliser.2023.100394, 2023.

Hohenegger, C., Brockhaus, P., Bretherton, C. S., and Schär, C.: The Soil Moisture–Precipitation Feedback in Simulations with Explicit and Parameterized Convection, Journal of Climate, 22, 5003 – 5020, https://doi.org/10.1175/2009JCLI2604.1, 2009.

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datseris, G., Esch, M., George, G., Giorgetta, M., Gutjahr,
750   O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemen, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geoscientific Model Development, 16,
755   779–811, https://doi.org/10.5194/gmd-16-779-2023, 2023.

Holloway, C. E., Woolnough, S. J., and Lister, G. M. S.: The Effects of Explicit versus Parameterized Convection on the MJO in a Large-Domain High-Resolution Tropical Case Study. Part I: Characterization of Large-Scale Organization and Propagation, Journal of the Atmospheric Sciences, 70, 1342 – 1369, https://doi.org/10.1175/JAS-D-12-0227.1, 2013.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., and Tan, J.: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06,
760   https://doi.org/10.5067/GPM/IMERG/3B-HH/06, 2019.

ICON-partnership, (DWD, MPI-M, DKRZ, KIT, and C2SM): ICON release 2024.01, https://doi.org/10.35089/WDCC/IconRelease01, 2024.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., and Núñez-Riboni, I.: Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, Journal of Advances in Modeling Earth Systems, 5, 287–315, https://doi.org/10.1029/2012MS000178, 2013.

765   Janssen, P.: The Interaction of Ocean Waves and Wind, Cambridge University Press, Cambridge, ISBN 9780521465403, https://doi.org/10.1017/CBO9780511525018, 2004.

Jian, B., Li, J., Wang, G., Zhao, Y., Li, Y., Wang, J., Zhang, M., and Huang, J.: Evaluation of the CMIP6 marine subtropical stratocumulus cloud albedo and its controlling factors, Atmospheric Chemistry and Physics, 21, 9809–9828, https://doi.org/10.5194/acp-21-9809-2021, 2021.

770   Judt, F., Klocke, D., Rios-Berrios, R., Vanniere, B., Ziemen, F., Auger, L., Biercamp, J., Bretherton, C., Chen, X., Düben, P., Hohenegger, C., Khairoutdinov, M., Kodama, C., Kornblueh, L., Lin, S.-J., Nakano, M., Neumann, P., Putman, W., Röber, N., Roberts, M., Satoh, M.,

Shibuya, R., Stevens, B., Vidale, P. L., Wedi, N., and Zhou, L.: Tropical Cyclones in Global Storm-Resolving Models, Journal of the Meteorological Society of Japan. Ser. II, 99, 579–602, https://doi.org/10.2151/jmsj.2021-029, 2021.

Klein, S. A. and Hartmann, D. L.: The seasonal cycle of low stratiform clouds, Journal of Climate, 6, 1587–1606, 1993.

775   Koldunov, N., Kölling, T., Pedruzo-Bagazgoitia, X., Rackow, T., Redler, R., Sidorenko, D., Wieners, K.-H., and Ziemen, F. A.: nextGEMS: output of the model development cycle 3 simulations for ICON and IFS, https://doi.org/10.26050/WDCC/nextGEMS_cyc3, 2023.

Kollet, S. J. and Maxwell, R. M.: Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, Advances in Water Resources, 29, 945–958, https://doi.org/10.1016/j.advwatres.2005.08.006, 2006.

Korn, P., Brüggemann, N., Jungclaus, J. H., Lorenz, S. J., Gutjahr, O., Haak, H., Linardakis, L., Mehlmann, C., Mikolajewicz, U., Notz,
780   D., Putrasahan, D. A., Singh, V., von Storch, J.-S., Zhu, X., and Marotzke, J.: ICON-O: The Ocean Component of the ICON Earth System Model—Global Simulation Characteristics and Local Telescoping Capability, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 952, https://doi.org/10.1029/2021MS002952, 2022.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D.,
785   Vasic, R., Xue, Y., and Yamada, T.: Regions of Strong Coupling Between Soil Moisture and Precipitation, Science, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

Kuang, Z.: Linear response functions of a cumulus ensemble to temperature and moisture perturbations and implications for the dynamics of convectively coupled waves, Journal of the atmospheric sciences, 67, 941–962, https://doi.org/10.1175/2009JAS3260.1, 2010.

Kölling, T., Kluft, L., and Rackow, T.: gribscan (v0.0.10), Zenodo, https://doi.org/10.5281/zenodo.10625189, 2024.

790   Lee, J. and Hohenegger, C.: Weaker land–atmosphere coupling in global storm-resolving simulation, Proceedings of the National Academy of Sciences, 121, e2314265 121, https://doi.org/10.1073/pnas.2314265121, 2024.

Lee, J., Hohenegger, C., Chlond, A., and Schnur, R.: The Climatic Role of Interactive Leaf Phenology in the Vegetation- Atmosphere System of Radiative-Convective Equilibrium Storm-Resolving Simulations, Tellus B: Chemical and Physical Meteorology, https://doi.org/10.16993/tellusb.26, 2022.

795   Leuenberger, D., Koller, M., Fuhrer, O., and Schär, C.: A Generalization of the SLEVE Vertical Coordinate, Monthly Weather Review, 138, 3683 – 3689, https://doi.org/10.1175/2010MWR3307.1, 2010.

Lin, J. L.: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean-atmosphere feedback analysis, Journal of Climate, 20, 4497–4525, https://doi.org/10.1175/JCLI4272.1, 2007.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and
800   the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, Journal of Climate, 31, 895 – 918, https://doi.org/10.1175/JCLI-D-17-0208.1, 2018.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Boundary-Layer Meteorology, 17, 187–202, https://doi.org/10.1007/BF00117978, 1979.

LUMI: Large Unified Modern Infrastructure, https://www.lumi-supercomputer.eu, 2024.

805   Maciel, P., Quintino, T., Modigliani, U., Dando, P., Raoult, B., Deconinck, W., Rathgeber, F., and Simarro, C.: The new ECMWF interpolation package MIR, https://doi.org/10.21957/h20rz8, 2017.

Madec, G. and the NEMO System Team: NEMO Ocean Engine Reference Manual, https://doi.org/10.5281/zenodo.8167700, 2023.

Malardel, S. and Wedi, N. P.: How does subgrid-scale parametrization influence nonlinear spectral energy fluxes in global NWP models?, Journal of Geophysical Research: Atmospheres, 121, 5395–5410, https://doi.org/10.1002/2015JD023970, 2016.

810 Maltrud, M. E. and McClean, J. L.: An eddy resolving global 1/10° ocean simulation, Ocean Modelling, 8, 31–54, https://doi.org/10.1016/j.ocemod.2003.12.001, 2005.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, https://doi.org/10.1029/2012MS000154, 2012.

815 Mauritsen, T., Redler, R., Esch, M., Stevens, B., Hohenegger, C., Klocke, D., Brokopf, R., Haak, H., Linardakis, L., Röber, N., and Schnur, R.: Early Development and Tuning of a Global Coupled Cloud Resolving Model, and its Fast Response to Increasing CO2, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.16993/tellusa.54, 2022.

McNorton, J., Agustí-Panareda, A., Arduini, G., Balsamo, G., Bousserez, N., Boussetta, S., Chericoni, M., Choulga, M., Engelen, R., and Guevara, M.: An Urban Scheme for the ECMWF Integrated Forecasting System: Global Forecasts and Residential CO2 Emissions, Journal
820 of Advances in Modeling Earth Systems, 15, e2022MS003 286, https://doi.org/10.1029/2022MS003286, 2023.

McNorton, J. R., Arduini, G., Bousserez, N., Agustí-Panareda, A., Balsamo, G., Boussetta, S., Choulga, M., Hadade, I., and Hogan, R. J.: An Urban Scheme for the ECMWF Integrated Forecasting System: Single-Column and Global Offline Application, Journal of Advances in Modeling Earth Systems, 13, e2020MS002 375, https://doi.org/10.1029/2020MS002375, 2021.

Mechoso, C. R., Robertson, A. W., Barth, N., Davey, M. K., Delecluse, P., Gent, P. R., Ineson, S., Kirtman, B., Latif, M., Treut, H. L., Nagai,
825 T., Neelin, J. D., Philander, S. G. H., Polcher, J., Schopf, P. S., Stockdale, T., Suarez, M. J., Terray, L., Thual, O., and Tribbia, J. J.: The Seasonal Cycle over the Tropical Pacific in Coupled Ocean–Atmosphere General Circulation Models, Monthly Weather Review, 123, 2825–2838, https://doi.org/https://doi.org/10.1175/1520-0493(1995)123<2825:TSCOTT>2.0.CO;2, 1995.

Miura, H., Satoh, M., Nasuno, T., Noda, A. T., and Oouchi, K.: A Madden-Julian Oscillation Event Realistically Simulated by a Global Cloud-Resolving Model, Science, 318, 1763–1765, https://doi.org/10.1126/science.1148443, 2007.

830 Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/10.1029/2019JD032361, 2021.

Nowak, J. L., Siebert, H., Szodry, K.-E., and Malinowski, S. P.: Coupled and decoupled stratocumulus-topped boundary layers: turbulence properties, Atmospheric Chemistry and Physics, 21, 10 965–10 991, 2021.

835 O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Penduff, T., Juza, M., Brodeau, L., Smith, G. C., Barnier, B., Molines, J.-M., Treguier, A.-M., and Madec, G.: Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales, Ocean Science, 6, 269–284, https://doi.org/10.5194/os-6-269-
840 2010, 2010.

Peters, K., Hohenegger, C., and Klocke, D.: Different representation of mesoscale convective systems in convection-permitting and convection-parameterizing NWP models and its implications for large-scale forecast evolution, Atmosphere (Basel)., 10, 1–18, https://doi.org/10.3390/atmos10090503, 2019.

Pincus, R., Mlawer, E. J., and Delamere, J. S.: Balancing Accuracy, Efficiency, and Flexibility in Radiation Calculations for Dynamical
845 Models, Journal of Advances in Modeling Earth Systems, 11, 3074–3089, https://doi.org/10.1029/2019MS001621, 2019.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Reviews of Geophysics, 53, 323–361, https://doi.org/https://doi.org/10.1002/2014RG000475, 2015.

Prein, A. F., Rasmussen, R., and Stephens, G.: Challenges and advances in convection-permitting climate modeling, Bulletin of the American Meteorological Society, 98, 1027–1030, https://doi.org/10.1175/BAMS-D-16-0263.1, 2017.

Prein, A. F., Rasmussen, R., Castro, C. L., Dai, A., and Minder, J.: Special issue: Advances in convection-permitting climate modeling, Climate Dynamics, 55, 1–2, https://doi.org/10.1007/s00382-020-05240-3, 2020.

Rackow, T., Danilov, S., Goessling, H. F., Hellmer, H. H., Sein, D. V., Semmler, T., Sidorenko, D., and Jung, T.: Delayed Antarctic sea-ice decline in high-resolution climate change simulations, Nature Communications, 13, 637, https://doi.org/10.1038/s41467-022-28259-y, 2022.

Rackow, T., Becker, T., Forbes, R., and Fielding, M.: Source code changes to the Integrated Forecasting System (IFS) for nextGEMS simulations, https://doi.org/10.5281/zenodo.10223577, 2023a.

Rackow, T., Hegewald, J., Koldunov, N. V., Mogensen, K., Scholz, P., Sidorenko, D., and Streffing, J.: FESOM2.5 source code used in nextGEMS Cycle 3 simulations with IFS-FESOM, https://doi.org/10.5281/zenodo.10225420, 2023b.

Rackow, T., Becker, T., and Pedruzo-Bagazgoitia, X.: Namelist files and settings for 30-year km-scale nextGEMS Cycle 4 production simulations with IFS-FESOM, https://doi.org/10.5281/zenodo.14725225, 2025a.

Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., Bechtold, P., Beyer, S., Bidlot, J., Boussetta, S., Deconinck, W., Diamantakis, M., Dueben, P., Dutra, E., Forbes, R., Ghosh, R., Goessling, H. F., Hadade, I., Hegewald, J., Jung, T., Keeley, S., Kluft, L., Koldunov, N., Koldunov, A., Kölling, T., Kousal, J., Kühnlein, C., Maciel, P., Mogensen, K., Quintino, T., Polichtchouk, I., Reuter, B., Sármány, D., Scholz, P., Sidorenko, D., Streffing, J., Sützl, B., Takasuka, D., Tietsche, S., Valentini, M., Vannière, B., Wedi, N., Zampieri, L., and Ziemen, F.: Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2.5 and NEMOv3.4, Geoscientific Model Development, 18, 33–69, https://doi.org/10.5194/gmd-18-33-2025, 2025b.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/https://doi.org/10.1029/2002JD002670, 2003.

Reick, C. H., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J., Raddatz, T., Roeckner, E., Schnur, R., and Wilkenskjeld, S.: JSACH 3 - The land component of the MPI Earth System Model: Documentation of version 3.2, Tech. Rep. 240, Max-Planck-Institut für Meteorologie, https://doi.org/10.17617/2.3279802, 2021.

Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., and Martinez, E.: A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers, Frontiers in Marine Science, 10, https://doi.org/10.3389/fmars.2023.1077623, 2023.

Sanderson, B. M. and Rugenstein, M.: Potential for bias in effective climate sensitivity from state-dependent energetic imbalance, Earth System Dynamics, 13, 1715–1736, https://doi.org/10.5194/esd-13-1715-2022, 2022.

Sandu, I.: Destination Earth's digital twins and Digital Twin Engine – state of play, https://doi.org/10.21957/is1fc736jx, 2024.

Sandu, I., van Niekerk, A., Shepherd, T. G., Vosper, S. B., Zadra, A., Bacmeister, J., Beljaars, A., Brown, A. R., Dörnbrack, A., McFarlane, N., Pithan, F., and Svensson, G.: Impacts of orography on large-scale atmospheric circulation, npj Climate and Atmospheric Science, 2, 10, https://doi.org/10.1038/s41612-019-0065-9, 2019.

Sarmany, D., Valentini, M., Maciel, P., Geier, P., Smart, S., Aguridan, R., Hawkes, J., and Quintino, T.: MultIO: A Framework
for Message-Driven Data Routing For Weather and Climate Simulations, in: Proceedings of the Platform for Advanced Scien-
885   tific Computing Conference, PASC '24, Association for Computing Machinery, New York, NY, USA, ISBN 9798400706394,
https://doi.org/10.1145/3659914.3659938, 2024.

Sarre, A., Demarcq, H., Keenlyside, N., Krakstad, J.-O., El Ayoubi, S., Jeyid, A. M., Faye, S., Mbaye, A., Sidibeh, M., and Brehmer, P.:
Climate change impacts on small pelagic fish distribution in Northwest Africa: trends, shifts, and risk for food security, Scientific Reports,
14, 12 684, https://doi.org/10.1038/s41598-024-61734-8, 2024.

890 Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.: Nonhydrostatic icosahedral atmospheric model (NICAM) for global
cloud resolving simulations, Journal of Computational Physics, 227, 3486–3514, https://doi.org/10.1016/j.jcp.2007.02.006, 2008.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Girolamo, S. D., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried,
K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T. C., Sprenger, M., Ubbiali, S., and Wernli, H.: Kilometer-Scale Climate
Models: Prospects and Challenges, Bulletin of the American Meteorological Society, 101, E567 – E587, https://doi.org/10.1175/BAMS-
895   D-18-0167.1, 2020.

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F.,
and Jung, T.: The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models, Journal
of Climate, 30, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1, 2017.

Schiemann, R., Athanasiadis, P., Barriopedro, D., Doblas-Reyes, F., Lohmann, K., Roberts, M. J., Sein, D. V., Roberts, C. D., Terray, L., and
900   Vidale, P. L.: Northern Hemisphere blocking simulation in current climate models: evaluating progress from the Climate Model Intercom-
parison Project Phase 5 to 6 and sensitivity to resolution, Weather and Climate Dynamics, 1, 277–292, https://doi.org/10.5194/wcd-1-277-
2020, 2020.

Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., Sein, D., Smolentseva, M., Rakowsky, N., and Jung, T.: Assess-
ment of the Finite-volumE Sea ice-Ocean Model (FESOM2.0) – Part 1: Description of selected key model elements and comparison to its
905   predecessor version, Geoscientific Model Development, 12, 4875–4899, https://doi.org/10.5194/gmd-12-4875-2019, 2019.

Schwierz, C., Croci-Maspoli, M., and Davies, H. C.: Perspicacious indicators of atmospheric blocking, Geophysical Research Letters, 31,
https://doi.org/10.1029/2003GL019341, 2004.

Segura, H.: nextGEMS overview paper: scripts, https://doi.org/10.17617/3.QZHXMC, 2025.

Segura, H., Hohenegger, C., Wengel, C., and Stevens, B.: Learning by Doing: Seasonal and Diurnal Features of Tropical Precipitation in a
910   Global-Coupled Storm-Resolving Model, Geophys. Res. Lett., 49, 1–10, https://doi.org/10.1029/2022GL101796, 2022.

Segura, H., Bayley, C., Fiévet, R., Glöckner, H., Günther, M., Kluft, L., Naumann, A. K., Ortega, S., Praturi, D. S., Rixen, M., Schmidt, H.,
Winkler, M., Hohenegger, C., and Stevens, B.: A single tropical rainbelt in global storm-resolving models: the role of surface heat fluxes
over the warm pool, https://doi.org/10.22541/essoar.173482210.07670254/v1, 2024.

Senf, F., Klocke, D., and Brueck, M.: Size-Resolved Evaluation of Simulated Deep Tropical Convection, Monthly Weather Review, 146,
915   2161 – 2182, https://doi.org/10.1175/MWR-D-17-0378.1, 2018.

Simmons, A. J. and Strüfing, R.: Numerical forecasts of stratospheric warming events using a model with a hybrid vertical coordinate,
Quarterly Journal of the Royal Meteorological Society, 109, 81–111, https://doi.org/10.1002/qj.49710945905, 1983.

Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T., and Teutsch, G.: Ambitious partnership needed
for reliable climate prediction, Nature Climate Change, 12, 499–503, https://doi.org/10.1038/s41558-022-01384-8, 2022.

920 Stenchikov, G. L., Kirchner, I., Robock, A., Graf, H.-F., Antuña, J. C., Grainger, R. G., Lambert, A., and Thomason, L.: Radiative forcing from the 1991 Mount Pinatubo volcanic eruption, Journal of Geophysical Research: Atmospheres, 103, 13 837–13 857, https://doi.org/10.1029/98JD00693, 1998.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T.: MACv2-SP: a parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, Geoscientific Model Development, 10, 433–452,
925 https://doi.org/10.5194/gmd-10-433-2017, 2017.

Stevens, B., Adami, S., Ali, T., Anzt, H., Aslan, Z., Attinger, S., Bäck, J., Baehr, J., Bauer, P., Bernier, N., Bishop, B., Bockelmann, H., Bony, S., Brasseur, G., Bresch, D. N., Breyer, S., Brunet, G., Buttigieg, P. L., Cao, J., Castet, C., Cheng, Y., Choudhury, A. D., Coen, D., Crewell, S., Dabholkar, A., Dai, Q., Doblas-Reyes, F., Durran, D., Gaidi, A. E., Ewen, C., Exarchou, E., Eyring, V., Falkinhoff, F., Farrell, D., Forster, P. M., Frassoni, A., Frauen, C., Fuhrer, O., Gani, S., Gerber, E., Goldfarb, D., Grieger, J., Gruber, N., Hazeleger, W., Herken,
930 R., Hewitt, C., Hoefler, T., Hsu, H. H., Jacob, D., Jahn, A., Jakob, C., Jung, T., Kadow, C., Kang, I. S., Kang, S., Kashinath, K., Königslöw, K. K.-V., Klocke, D., Kloenne, U., Klöwer, M., Kodama, C., Kollet, S., Kölling, T., Kontkanen, J., Kopp, S., Koran, M., Kulmala, M., Lappalainen, H., Latifi, F., Lawrence, B., Lee, J. Y., Lejeun, Q., Lessig, C., Li, C., Lippert, T., Luterbacher, J., Manninen, P., Marotzke, J., Matsouoka, S., Merchant, C., Messmer, P., Michel, G., Michielsen, K., Miyakawa, T., Müller, J., Munir, R., Narayanasetti, S., Ndiaye, O., Nobre, C., Oberg, A., Oki, R., Özkan Haller, T., Palmer, T., Posey, S., Prein, A., Primus, O., Pritchard, M., Pullen, J., Putrasahan, D.,
935 Quaas, J., Raghavan, K., Ramaswamy, V., Rapp, M., Rauser, F., Reichstein, M., Revi, A., Saluja, S., Satoh, M., Schemann, V., Schemm, S., Poberaj, C. S., Schulthess, T., Senior, C., Shukla, J., Singh, M., Slingo, J., Sobel, A., Solman, S., Spitzer, J., Stier, P., Stocker, T., Strock, S., Su, H., Taalas, P., Taylor, J., Tegtmeier, S., Teutsch, G., Tompkins, A., Ulbrich, U., Vidale, P. L., Wu, C. M., Xu, H., Zaki, N., Zanna, L., Zhou, T., and Ziemen, F.: Earth Virtualization Engines (EVE), https://doi.org/10.5194/essd-16-2113-2024, 2024.

Stockdale, T., Senan, R., Hogan, R., Kipling, Z., and Flemming, J.: A new time-varying tropospheric aerosol climatology for the IFS,
940 https://doi.org/10.21957/ba371f56ts, 2024.

Takasuka, D., Kodama, C., Suematsu, T., Ohno, T., Yamada, Y., Seiki, T., Yashiro, H., Nakano, M., Miura, H., Noda, A. T., Nasuno, T., Miyakawa, T., and Masunaga, R.: How Can We Improve the Seamless Representation of Climatological Statistics and Weather Toward Reliable Global K-Scale Climate Simulations?, Journal of Advances in Modeling Earth Systems, 16, https://doi.org/10.1029/2023MS003701, 2024.

945 Tian, Y. and Kuang, Z.: Why does deep convection have different sensitivities to temperature perturbations in the lower versus upper troposphere?, Journal of the Atmospheric Sciences, 76, 27–41, https://doi.org/10.1175/JAS-D-18-0023.1, 2019.

van Westen, R. M. and Dijkstra, H. A.: Ocean eddies strongly affect global mean sea-level projections, Science Advances, 7, eabf1674, https://doi.org/10.1126/sciadv.abf1674, 2021.

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., and Schröter, J.: The Finite Element Sea Ice-
950 Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model, Geoscientific Model Development, 7, 663–693, https://doi.org/10.5194/gmd-7-663-2014, 2014.

Wazneh, H., Gachon, P., Laprise, R., de Vernal, A., and Tremblay, B.: Atmospheric blocking events in the North Atlantic: trends and links to climate anomalies and teleconnections, Climate Dynamics, 56, 2199–2221, https://doi.org/10.1007/s00382-020-05583-x, 2021.

Weiss, P., Herbert, R., and Stier, P.: ICON-HAM-lite: simulating the Earth system with interactive aerosols at kilometer scales, EGUsphere,
955 2024, 1–26, https://doi.org/10.5194/egusphere-2024-3325, 2024.

Wieners, K.-H., Ziemen, F. A., Koldunov, N., Pedruzo-Bagazgoitia, X., Rackow, T., Redler, R., Sidorenko, D., and Kölling, T.: nextGEMS: output of the model development cycle 2 simulations for ICON and IFS, https://doi.org/10.26050/WDCC/nextGEMS_cyc2, 2023.

Wieners, K.-H., Rackow, T., Aguridan, R., Becker, T., Beyer, S., Cheedela, S. K., Dreier, N.-A., Engels, J. F., Esch, M., Frauen, C., Klocke, D., Kölling, T., Pedruzo-Bagazgoitia, X., Putrasahan, D., Sidorenko, D., Schnur, R., Stevens, B., and Zimmermann, J.: nextGEMS: output

960 of the production simulations for ICON and IFS, https://doi.org/10.35095/WDCC/nextGEMS_prod_addinfov1, 2024.

Wood, R. and Bretherton, C. S.: On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability, Journal of Climate, 19, 6425 – 6432, https://doi.org/10.1175/JCLI3988.1, 2006.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., Sillmann, J., Lupo, A. R., and Seneviratne, S.: Blocking and its Response to Climate Change, Current Climate Change Reports, 4, 287–300, https://doi.org/10.1007/s40641-018-0108-z, 2018.

965 Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quarterly Journal of the Royal Meteorological Society, 141, 563–579, https://doi.org/10.1002/qj.2378, 2015.