AGU100 ADVANCING EARTH AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2017MS001233

This article is a companion to Giorgetta et al. (2018) https://doi.org/10.1029/2017MS001242.

Key Points:
- Article presents evaluation of atmosphere component of new ICON Earth system model
- The new MPI atmospheric ICON-A model partly outperforms ECHAM6.3
- ICON-A is flexible to be run at grid spacings from a few tens of meters to hundreds of kilometers

Correspondence to:
T. Crueger,
traute.crueger@mpimet.mpg.de

Citation:
Crueger, T., Giorgetta, M. A., Brokopf, R., Esch, M., Fiedler, S., Hohenegger, C., et al. (2018). ICON-A, the atmosphere component of the ICON Earth system model: II. Model evaluation. *Journal of Advances in Modeling Earth Systems*, *10*, 1638–1662. https://doi.org/10.1029/2017MS001233

Received 15 NOV 2017
Accepted 23 MAY 2018
Accepted article online 9 JUN 2018
Published online 21 JUL 2018

ICON-A, The Atmosphere Component of the ICON Earth System Model: II. Model Evaluation

T. Crueger[1], M. A. Giorgetta[1], R. Brokopf[1], M. Esch[1], S. Fiedler[1], C. Hohenegger[1], L. Kornblueh[1], T. Mauritsen[1], C. Nam[1,2], A. K. Naumann[1], K. Peters[1,3], S. Rast[1], E. Roeckner[1], M. Sakradzija[1], H. Schmidt[1], J. Vial[1], R. Vogel[1,4], and B. Stevens[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Institute for Meteorology, University of Leipzig, Leipzig, Germany, [3]Now at Deutsches Klimarechenzentrum GmbH (DKRZ), Hamburg, Germany, [4]Now at LMD/IPSL, CNRS, Sorbonne Université, Paris, France

Abstract We evaluate the new icosahedral nonhydrostatic atmospheric (ICON-A) general circulation model of the Max Planck Institute for Meteorology that is flexible to be run at grid spacings from a few tens of meters to hundreds of kilometers. A simulation with ICON-A at a low resolution (160 km) is compared to a not-tuned fourfold higher-resolution simulation (40 km). Simulations using the last release of the ECHAM climate model (ECHAM6.3) are also presented at two different resolutions. The ICON-A simulations provide a compelling representation of the climate and its variability. The climate of the low-resolution ICON-A is even slightly better than that of ECHAM6.3. Improvements are obtained in aspects that are sensitive to the representation of orography, including the representation of cloud fields over eastern-boundary currents, the latitudinal distribution of cloud top heights, and the spatial distribution of convection over the Indian Ocean and the Maritime Continent. Precipitation over land is enhanced, in particular at high-resolution ICON-A. The response of precipitation to El Niño sea surface temperature variability is close to observations, particularly over the eastern Indian Ocean. Some parameterization changes lead to improvements, for example, with respect to rain intensities and the representation of equatorial waves, but also imply a warmer troposphere, which we suggest leads to an unrealistic poleward mass shift. Many biases familiar to ECHAM6.3 are also evident in ICON-A, namely, a too zonal SPCZ, an inadequate representation of north hemispheric blocking, and a relatively poor representation of tropical intraseasonal variability.

1. Introduction

ICON-A denotes the atmospheric component of the new icosahedral nonhydrostatic Earth system model (ICON-ESM), developed by the Max Planck Institute (MPI) for Meteorology. ICON-A has been developed to replace ECHAM, which over the past three decades had served as the atmospheric component of what was initially called the coupled-climate model of the Max Planck Institute and which, with the introduction of the carbon cycle, more recently became known as the Max Planck Institute Earth system model (MPI-ESM). ECHAM is comprised of a dynamical core taken from the European Center for Medium Range Weather Forecasts (ECMWF) and a suite of physical parameterizations that have been developed, implemented, and fine-tuned throughout its history (Roeckner et al., 1996, 2006; Stevens et al., 2013). Simulations based on ECHAM have been performed in support of every phase of Coupled Model Intercomparison Project (CMIP), and in the informal way in which these are measured, it has generally been recognized as one of the leading climate models worldwide. Its last release, ECHAM6.3 was made in preparation for participation in Coupled Model Intercomparison Project Phase 6 (CMIP6), as the atmospheric component of the MPI-ESM (version 1.2; see section 2).

Replacing such a well-tuned model is not an easy task, but it offers an opportunity to learn things, such as what factors influence the quality of a climate model. In addition, it allows us to document basic features of the climate of ICON-A for eventual users as a matter of good scientific practice. A technical description of ICON-A is provided in the companion article by Giorgetta et al. (2018).

The use of the ICON-A development to shed light on factors that influence the quality of simple measures of a model's representation of Earth's climate is aided by the fact that ECHAM6.3 and ICON-A share almost the same

physics. The main differences among the models are as follows: (i) ECHAM6.3's development, spanning over three decades, has allowed for a greater opportunity to identify bugs and fine-tune the physics to get the most out of the models; and (ii) ICON-A shares the same dynamical core (Zängl et al., 2015) as the numerical weather prediction (NWP) version of the same model, whereas ECHAM's dynamics were modeled by the Simmons et al. (1989) spectral-space dynamical core developed for the ECMWF. To provide context in assessing what aspects of a climate model are important for its representation of the climate system, for both ECHAM6.3 and ICON-A, we present simulations at different resolutions. This allows us to compare, for example, differences between the two models to differences that would emerge from a modest (factor of 2 to 4) change in the resolution of one of the models.

This article is the companion paper of Giorgetta et al. (2018), which describes the construction and tuning of the ICON-A model. Here, in this paper, we endeavor to evaluate the salient climatological model features of ICON-A. This second part is also the first and only comprehensive evaluation of ICON-A and is organized as follows: In section 2 the salient features of the different models analyzed in this paper, the type of simulations that form the basis of this analysis and the data sources for the evaluation of the model's output are summarized. Subsequently, section 3 presents an overview of the mean state—including indicators of the circulation, the energy budget and factors influencing it, and the precipitation—of ICON-A in comparison to observations and ECHAM6.3. The variability of the model is evaluated, particularly in light of variability for climate and the growing appreciation that models, which adequately capture both the mean climatological state and the patterns and strength of the variability within that state, prove more difficult to be constructed than ones that do one or the other.

2. Models, Experiments, and Data

We evaluate the ICON-A at a resolution over which it has been tuned, and a not-tuned version of the same model at fourfold higher spatial resolution. Two tuned and one not-tuned ECHAM6.3 GCM versions, as well as observations, serve as a reference. In the companion article, Giorgetta et al. (2018) present the technical description of ICON-A. ECHAM6.3 differs from its predecessor ECHAM6.1 (Stevens et al., 2013) with respect to the formulation of a few physical processes. Subgrid-scale cloudiness is represented using an assumed humidity distribution function scheme developed by Sundqvist et al. (1989). In ECHAM6.3, to reduce marine stratocumulus biases (Stevens et al., 2013), a modification of the treatment of subgrid-scale cloudiness has been introduced. In addition, the radiation scheme PSrad (Pincus & Stevens, 2013) is implemented to ECHAM6.3. A major change in ECHAM6.3 compared to previous versions represents the treatment of radiative effects of anthropogenic aerosol in the shortwave radiation calculation. Here the climatology of anthropogenic aerosols used in previous ECHAM versions has been replaced by the MACv2-SP parameterization for the anthropogenic aerosol optical properties, which induces an associated Twomey effect (Fiedler et al., 2017; Stevens et al., 2017).

Minor updates in ECHAM6.3 concern the convection scheme, which is the scheme of Tiedtke (1989) and with extensions proposed by Nordeng (1994). Furthermore, modifications have been performed to the stratiform cloud scheme and the turbulent transfer scheme, partly to eliminate bugs. The land surface is coupled to the atmosphere by the land component of MPI-ESM1.2, JSBACH (Reick et al., 2013). New in ECHAM6.3 is the transfer and storage of soil water, now treated in a 5-layer scheme (Hagemann & Stacke, 2015). These changes necessitated some retuning of the model, as will be documented in a forthcoming model description paper.

The main difference between ICON-A and ECHAM6.3 is that ICON-A has a completely new dynamical core, which solves the compressible equations of motion in physical (rather than spectral) space using a vertical discretization in geometric rather than in pressure levels. For reference, ECHAM6.3 has a spectral-transform dynamical core, which solves the primitive equations with a vertical discretization in mass space (pressure levels). The dynamical core in ICON-A is identical to that used in the NWP version of ICON-A and arises from the joint development of ICON-A by the German Weather Service (DWD) and the MPI, described by Zängl et al. (2015). Differences in the treatment of dynamics also imply differences in the treatment of boundary conditions, particularly at the upper boundary, but also in the treatment of dissipation on the grid scale so as to ensure numerical stability and damp computational modes. Along with a new dynamical core, ICON-A has an entirely new computational infrastructure, encompassing its handling of input and output data, housekeeping, and event handling as well as the architectural structure of the model from a software engineering point of view.

Table 1
Naming and Salient Features of the Experiments Investigated in This Article

Name	Model version	Discretization	Grid spacing	Tuning	Time step
ICON-R2B4	ICON-AES-1.3.00	R2B4, 47 *z* levels	160 km	Yes	600 s
ICON-R2B6 untuned	ICON-AES-1.3.00	R2B6, 47 *z* levels	40 km	R2B4 settings	150 s
ECHAM-LR-(R1/R2)	ECHAM6.3.04	T63, 47 *p* levels	200 km	Yes	450 s
ECHAM-HR untuned	ECHAM6.3.04	T127, 95 *p* levels	100 km	LR settings	200 s
ECHAM-HR	ECHAM6.3.04	T127, 95 *p* levels	100 km	Yes	200 s

Note. ECHAM-LR, ECHAM-LR-R1, and ECHAM-LR-R2 represent an ensemble of ECHAM-LR experiments. For the spectral core the grid spacing of the transform grid at the equator is specified.

The physics package of ICON-A has been ported to ICON-A from ECHAM6.3, and is sometimes referred to as the ECHAM physics, or the ICON-A component of the ESM physics, to differentiate them from the NWP physics. They include a suite of physical parameterizations for the representation of diabatic processes (Stevens et al., 2013). Structural changes to the model's physics were only performed for the turbulent fluxes of the boundary layer. Here the turbulent kinetic energy scheme of Brinkop and Roeckner (1995) that has been used in ECHAM for many years has been replaced by a scheme for the total turbulence energy (Mauritsen et al., 2007; Pithan et al., 2015).

Additional changes of the ECHAM6.3 physics have been made in ICON-A because of its different grids and discretization. Simply porting the physics line by line from ECHAM6.3 to ICON-A led to a substantial degradation of the climate, which required tuning to ameliorate model performances. This tuning process is described in detail in Giorgetta et al. (2018) and only briefly summarized here. In the convection scheme of ICON-A, the entrainment rates have been doubled for middle and deep convective updraft and downdraft and the relaxation time for convective available potential energy has been reduced by 50%. In addition, in the ICON-A implementation of the ECHAM6.3 physics, precipitation has been made independent of the thickness of a convective cloud. In ECHAM6.3 and all earlier versions of ECHAM, these thicknesses were prescribed to be 150 hPa over the ocean and 300 hPa over land. In the experiments with ECHAM6.3, one high resolution (HR) and the low resolution (LR) experiment (see below and Table 1) were individually tuned to optimize their climate. For reference, we also consider an ECHAM-HR experiment that utilizes the ECHAM-LR parameter settings. In addition to changes in the convection scheme, ICON-A differs from ECHAM6.3 physics in its tuning of the gravity wave drag. As described in Part 1, there are separate parameterizations for orographic and nonorographic gravity waves. These are strongly influenced by surface orography, which differs substantially between ICON-A's physical space representation and the spectral space representation in ECHAM6.3. Hence, very different parameter settings in ECHAM6.3 and ICON-A have been selected. In the case of ICON-A, the model tuning has focused on a realistic shallow tropospheric circulation, and not on the circulation of the stratosphere, something that is reflected also in the quality of the simulated climate in these two regions of the atmosphere. Work is ongoing to improve the representation of the middle atmosphere.

The ICON-A experiments were performed with ICON-AES-1.3.00, the latest release of the model, in two different horizontal resolutions. The configuration, denoted as R2B4, has an approximately isotropic horizontal grid mesh of around 160 km; the other (R2B6) is a fourfold refinement, roughly a 40-km horizontal mesh. Both configurations adopt the same vertical discretization, with 47 vertical levels (see Table 1), extending from the surface up to 0.01 hPa. The layers have thicknesses ranging about 300 m at 1,000 m height, and about 1,000 m at 10,000 m height. Detailed information concerning the vertical levels are given in Giorgetta et al. (2018). The physics of the ICON-R2B4 configuration has been tuned, while in the R2B6 configuration, all parameter settings are held constant. Hence, ICON-R2B6 represents an untuned configuration.

The ECHAM6.3 simulations are performed with the ECHAM6.3.04 version. Two configurations have been developed for participation in CMIP6, differing in their horizontal and vertical grid spacings. Both were individually tuned to optimize their climate. This tuning mainly refers to parameters of the convection scheme. The most important are the convection conversion rate for cloud water to rain that has been reduced from 2.5×10^{-4} to 1.5×10^{-4} and the timescale for removal of convective available potential energy that has been decreased from 2 to 1 hr. One configuration uses a T63 truncation (with a latitude-longitude transform grid containing 192 degrees of freedom at each latitude, and 96 at each longitude) and 47 vertical levels (LR). T63 denotes a triangular truncation of the spherical harmonics to 63 wave numbers, representing grid distances

Table 2
Observational Data Used for Evaluation in This Article

Quantity	Name	Period	Reference
TOA radiative properties	CERES-EBAF-Ed4.0	2000/03–2016/02	Loeb et al. (2009)
Surface radiative properties	CERES-EBAF-Surface-Ed4.0	2000/03–2016/02	Kato et al. (2013)
Total cloud fraction	CALIPSO-GOCCP-v2.9	2006/06–2015/12	Chepfer et al. (2010)
Energy budget	SS12	2000–2012	Stevens and Schwartz (2012)
Barbados cloud fraction	CloudNet (v0.10.2)	2011/01–2015/09	Stevens et al. (2016)
Leipzig cloud fraction	CloudNet (v0.10.2)	2011/08–2015/12	n/a
Total precipitation daily	GPCP-V1.2	1997–2014	Huffman et al. (2001)
Total precipitation monthly	GPCP-V2.2	1978–2013	Adler et al. (2003)
Zonal mean temperature	ERA interim	1979–2008	Dee et al. (2011)
Zonal mean zonal wind	ERA interim	1979–2008	Dee et al. (2011)
Sea level pressure	ERA interim	1979–2008	Dee et al. (2011)
Sea surface temperature	HadISST1	1979–2008	Rayner et al. (2003)

Note. The observations of cloud fraction were processed using the CloudNet algorithms following Illingworth et al. (2007).

of 1.87° by 1.87°, corresponding to about 200-km grid spacing near the equator. The second configuration (ECHAM-HR) has a T127 truncation (394 points around a latitude circle) and 95 vertical levels. The lowest 12 levels of the 95 vertical level model are distributed similarly to the lower 12 levels in the 47 vertical level version; hence, the HR version essentially yields a better vertical resolution of the upper troposphere and the stratosphere. Both versions extend from the surface up to 0.01 hPa, corresponding to a height of about 80 km (Stevens et al., 2013). For a better assessment of the changes going from ICON-R2B4 to ICON-R2B6, we performed a second ECHAM-HR simulation, which utilizes the parameter settings of ECHAM-LR, and is called ECHAM-HR untuned. Table 1 provides an overview of the experiments and their notations.

The analysis and documentation of ICON-A's climate is based on AMIP experiments. These were performed according to the AMIP II protocol (Taylor et al., 2000), prescribing monthly sea surface temperatures (SSTs), sea ice concentrations, observed solar irradiance and greenhouse gas concentrations up to the year 2000, and the RCP4.5 concentrations thereafter (Taylor et al., 2012). While all ICON-A and ECHAM6.3 experiments follow this AMIP protocol, they were however initialized differently. ICON-A starts from an initialized state derived from ECMWF analysis data, whereas ECHAM6.3 is initialized from an idealized state with January 1976 boundary and forcing data. From January 1979 on, a physically consistent state is assumed. All experiments span a 36-year (1979 to 2014) period.

Table 3
Quantities and Validation Data Used to Establish the Climatological Model Biases Shown in Figures 1 and 2

Quantity	Validation data	Period	Reference
Air pressure at sea level	ERA interim	1979–2008	Dee et al. (2011)
TOA outgoing longwave radiation	CERES	2001–2013	Loeb et al. (2012)
TOA outgoing shortwave radiation	CERES	2001–2013	Loeb et al. (2012)
Surface eastward wind stress	ERA interim	1979–2008	Dee et al. (2011)
Surface northward wind stress	ERA interim	1979–2008	Dee et al. (2011)
Column water vapor content	NVAP	1988–1999	Randel et al. (1996)
Total precipitation over ocean	HOAPS	1988–2005	Andersson et al. (2010)
Total precipitation over land	GPCP-V2.2	1979–2008	Adler et al. (2003)
Surface land temperature	HadCRU4	1979–2008	Jones et al. (2012)
Temperature (850 hPa)	ERA interim	1979–2008	Dee et al. (2011)
Stationary waves (500 hPa)	ERA interim (geopotential)	1979–2008	Dee et al. (2011)
Zonal mean temperature (up to 10 hPa)	ERA interim	1979–2008	Dee et al. (2011)
Zonal mean zonal wind (up to 10 hPa)	ERA interim	1979–2008	Dee et al. (2011)

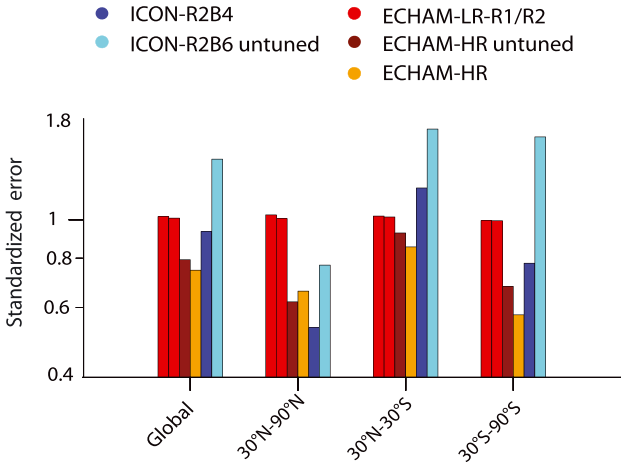

Figure 1. Standardized annual mean climatological errors of the experiments related to the reference experiment ECHAM-LR. A value smaller/larger than 1 indicates a smaller/larger bias compared to ECHAM-LR (period: 1979–2008).

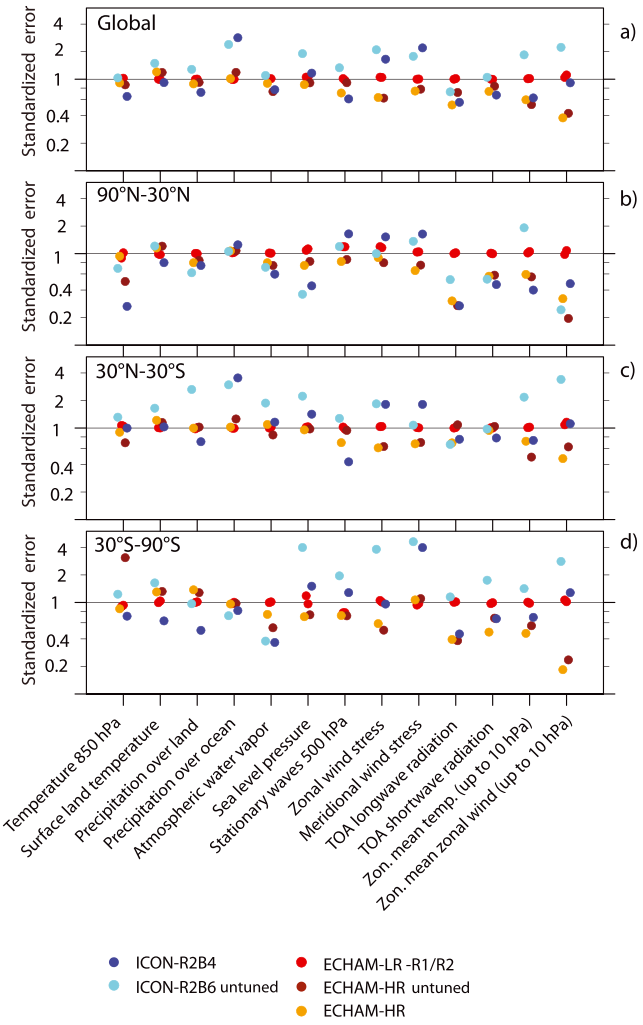

Figure 2. Same as Figure 1 for the single variables (see Table 3).

Observational data used in this article are mostly taken from satellite climatologies as indicated in Table 2. In addition, some evaluations are based on comparisons to reanalyses of meteorological data using the ERA interim data set (Dee et al., 2011), and cloud fields are further analyzed using measurements performed at the Barbados Cloud Observatory (BCO; Stevens et al., 2016) and at Leipzig. Table 3 refers to the observational data used to calculate the skill scores discussed in section 3.

3. Mean State

For a rough initial assessment of the mean state, we utilize a performance index motivated by Reichler and Kim (2008) and Watterson et al. (1999) and utilized by Stevens et al. (2013). This is based on a comparison of the annual climatology of single variables with satellite or reanalysis based estimates of the true climatology (see Table 3). The relative weight of each field is given by the observed interannual variance field. Finally, the relative error, compared to a reference experiment (in this case ECHAM-LR), is calculated. For the purpose of comparison, simulated fields are interpolated from their native grids to the native grids of the observational data sets. A performance index of unity denotes the performance as good as ECHAM-LR. Values larger than unity represent larger departures from the observations and are indicative of worse performance relative to ECHAM-LR, and smaller values imply smaller departures from the observations, hence, better performance. These scores give a first hint at model biases keeping in mind that for individual fields a large bias may reflect pattern changes, for example, shifts of rain bands, as much as a mean bias. The scores, averaged over all variables listed in Table 3, are shown in Figure 1, and for single atmospheric quantities in Figure 2 on the global scale, for the tropics (30°S to 30°N) and for the northern (30°N to 90°N) and southern extra tropics (90°S to 30°S) covering the period from 1979 to 2008. As a measure of model-internal variability, two additional ECHAM6.3-LR experiments (ECHAM-LR-R1 and ECHAM-LR-R2) are included to form a small ECHAM6.3-LR ensemble. These two ECHAM6.3 experiments reveal similar biases as the reference experiment ECHAM-LR. Small deviations are only found in the extra tropics for the stationary waves in 500 hPa. This implies that the spread of the scores within the ensemble remains marginal and that larger deviations between the scores of the experiments can be attributed to differences between the models and/or their configurations.

At a glance, it is evident that by this metric, the ICON-R2B4 model performs similarly, or slightly better, than the well-established ECHAM-LR model (Figure 1). The performance of ICON-R2B4 is especially good in the northern extra tropics, where it even outperforms ECHAM-HR. ICON-R2B4's performance is less satisfactory in the tropics, but not much worse than ECHAM-LR. Comparing ECHAM-LR with ECHAM-HR untuned and ICON-R2B4 with ICON-R2B6 (untuned) reveals a strong discrepancy: While the higher-resolution ICON-A experiment is clearly worse than its low-resolution counterpart, the opposite is the case for the corresponding ECHAM6.3 configurations. We speculate that for ECHAM6.3, where the tuning is more of a fine-tuning of a model whose physics and parameter choices have coevolved over decades, the model is less sensitive to its tuning parameters. For ICON-A the tuning has been more of a first attempt at identifying large inconsistencies and may be more grid dependent at this stage of development. Another explanation for relatively modest influence of tuning, as compared to resolution, for ECHAM6.3 is that (i) the changes in horizontal resolution are more modest than for ICON-A

and (ii) the improved vertical resolution may bring improvements that offset degradations associated with the lack of fine tuning. For ICON-A, this long-lasting tuning has not been done so far. The improvement of ECHAM-HR untuned compared to ECHAM-LR is thus due to the increased resolution, the further improvement going from the untuned to the tuned ECHAM-HR to the tuning procedure. For some details, we also find improvements going from ICON-R2B4 to ICON-R2B6 untuned (see below). However, the degradation due to the missing tuning dominates these improvements. On the other hand, the two ICON-A configurations differ by a factor of four in horizontal resolution and share the same vertical grid, whereas the ECHAM6.3 configurations differ by a factor of two in horizontal resolution and also differ in the vertical grid. Therefore, also these different resolution changes may have an impact on the models' performance besides the parameter settings.

The analysis of the scores for the single quantities in Figure 2 suggests that improvements in ICON-R2B4 relative to ECHAM-LR are most evident in the top-of-atmosphere (or TOA) radiative balance, and these improvements are evident in both the tropics and extra tropics individually. On a global scale, the main degradation in the simulations is in the precipitation over the ocean, and the surface-wind stress; this mostly comes from the tropics where heating and winds are closely related (Figure 2a). This metric leaves it unclear to what extent precipitation biases (discussed further in section 3.4) arise from changes in the convective parameterization design versus the effects of the dynamical core. It might also be that the stronger increase of resolution in ICON-A (from 160 to 40 km) compared to ECHAM6.3 (from 200 to 100 km) lead to improper parameters in the parameterizations in ICON-R2B6.

The single-field analysis also highlights a somewhat ambiguous impact of resolution. Whereas ECHAM-HR and even ECHAM-HR-untuned biases are reduced or unchanged relative to those of ECHAM-LR in essentially every field and in each of the regions, ICON-R2B6 has larger or unchanged biases relative to its low-resolution counterpart. A higher-resolution model with parameterized convection is, by mere virtue of its longer time-to-solution, more difficult to tune. ICON-A appears more sensitive to the tuning, but it remains unclear if this is because the tuning of this model is still in an early stage, or if it reflects the effect of larger changes in the horizontal mesh than what was considered for ECHAM6.3, or the lack of compensating improvements that might come with improved vertical resolution. For ICON-A, tuning is probably still more important. However, we expect this situation to change as one begins to replace parameterized processes with resolved ones, as for instance in storm resolving or large-eddy simulations (Heinold et al., 2013; Heinze et al., 2017; Klocke et al., 2017; Miyamoto et al., 2013; Tomita et al., 2005).

3.1. Stationary Waves and Zonal Mean Structure of Temperature and Winds

ICON-A's mean climate, and the influence of the dynamical core is further explored by examining the sea-level pressure (PSL), as a measure of the stationary wave structure. At the summary level presented in Figure 2a, global annual biases in PSL are broadly similar, with the exception of ICON-R2B6, across the different models and model configurations. This similarity belies larger changes in the seasonal patterns (Figure 3). Figure 3 presents the differences (b–f and h–l) between the simulated PSL and the observed values as represented by the ERA-interim (a and g) for the solstice seasons. In these seasons there is a consistent tendency of ICON-A to underestimate the PSL relative to ERA-interim over large regions in the tropics, primarily over oceans. Only in boreal summer in the northern hemisphere, ICON-R2B4 reveals evidence of an improvement especially compared to the ECHAM-LR representation of the observed PSL patterns (Figures 3h and 3k). Generally, in both versions of ICON-A, the low-pressure regions of the extra tropics are underrepresented in the mean of June to August (JJA), leading to weaker midlatitude meridional pressure gradients, and in the case of the southern hemisphere, a pressure pattern bias that is consistent with a poleward bias in the position of the storm tracks.

The stationary wave pattern in northern hemisphere winter (Figures 3a–3f) is worse in ICON-A than for ECHAM6.3 (compared to ERA interim), independent of the resolution chosen. Interestingly, ICON-A simulates both a weaker Azores High and Icelandic Low resulting in an underestimation of the climatologically averaged horizontal pressure gradient near the surface in the mean of December to February (DJF). These differences affect the baroclinic conditions over the North Atlantic consistent with an underestimation of the North Atlantic storm track in ICON-A from December to February.

The zonal pattern of PSL biases are, as would be expected, consistent with temperature biases in the troposphere. Figure 4 presents the annual and zonal mean temperature in the ERA interim reanalysis from

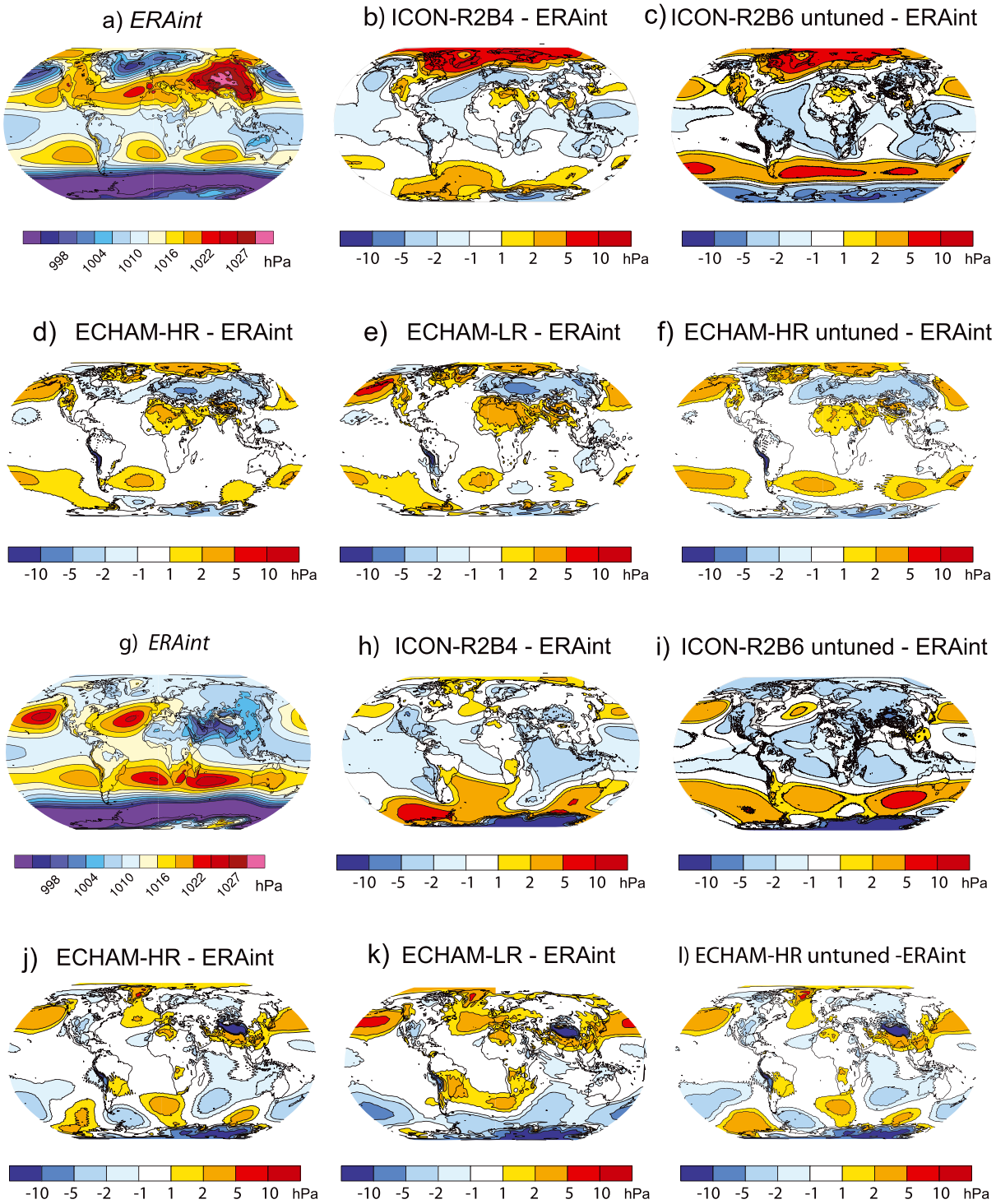

Figure 3. Mean PSL climatology; DJF (a–f), and JJA (g–l). ERA interim (a, g), and differences between the simulations and ERA interim: ICON-R2B4 (b, h), ICON-R2B6 (c, i), ECHAM-HR (d, j), ECHAM-LR (e, k), and ECHAM-HR untuned (f, l).

the surface to the lower mesosphere (0.1 hPa), and the differences between the five model simulations and the reanalysis. Throughout the tropics, the models are up to 2 K too warm in the lower and middle troposphere and this feature is somewhat more pronounced in the ICON-A simulations (between 2 and 4 K). Near the tropopause, all the models are too cold, especially in the ICON-A simulations, which is more evident in the untuned high-resolution simulation ICON-R2B6.

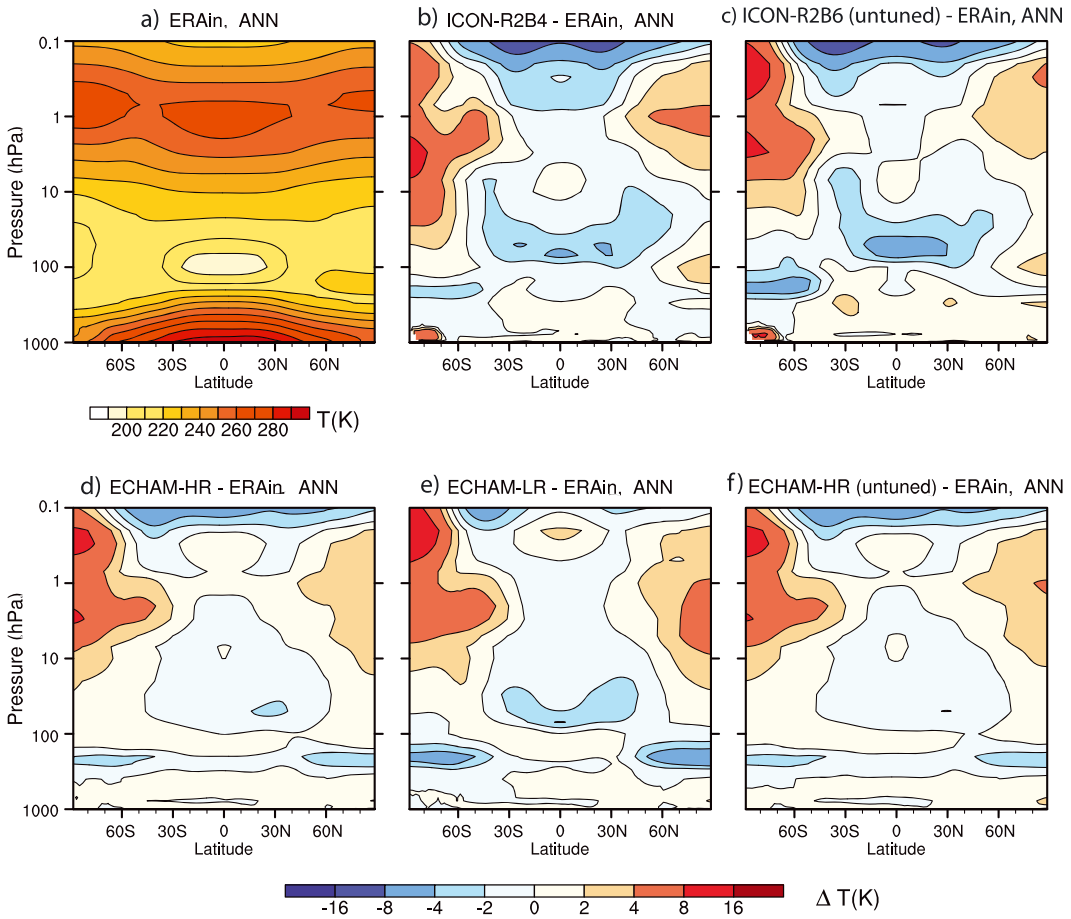

Figure 4. Multiyear annual and zonal mean temperature (K) for ERA interim (a), and anomalies w.r.t. ERA interim: ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e), and ECHAM-HR untuned (f).

The thermal structure of the tropical atmosphere is strongly influenced by convection. For a given SST, more entrainment tends to favor a warmer troposphere and more equatorial convection (Becker et al., 2017; Möbis & Stevens, 2012). Hence, the warmer troposphere throughout the tropics and below 200 hPa in the ICON-A simulations is consistent with the changes to the parameter settings of the convection scheme in the course of its tuning (see section 2). The warmer tropical troposphere, and cooler temperatures at high latitudes are reflected in the zonally averaged zonal winds (Figure 5). The westerlies are strengthened and shifted poleward, with less separation between the eddy-driven and subtropical jets in the southern hemisphere as compared to observations.

In the stratosphere, ICON-A shows a slight increase of the biases compared to ECHAM6.3. The stratosphere is too cold in the tropics (up to 8 K in the lower and even more than 8 K in the high stratosphere), and too warm in the extra tropics (around 8 K in the northern extra tropics; Figure 4). The tropical cold bias is bottom heavy in both ICON-A and ECHAM6.3, and the warm bias is top heavy, which as a consequence of thermal wind balance is reflected in the prograde (easterly) zonally averaged zonal wind bias through the middle stratosphere, which is even more pronounced in the ICON-A simulations (Figure 5). The differences among the models and their configurations are small, and perhaps only the ECHAM-HR simulations (tuned and untuned) stand out as somewhat different, particularly in their tropical on-equator wind bias, and seemingly smaller lower-stratosphere temperature biases. Such difference may be indicative of the better ability of this model (with much higher vertical resolution in the stratosphere) to capture the interaction among the waves and the mean flow, as reflected also in its ability to capture (albeit with parameterized wave sources) the quasi-biennial oscillation. It may also reflect the rewards of having spent more attention to the tuning of this model's representation of the stratosphere (pretuned versions of ECHAM-HR had biases that were more similar to ECHAM-LR), precisely because its representation of important processes is improved by resolution.

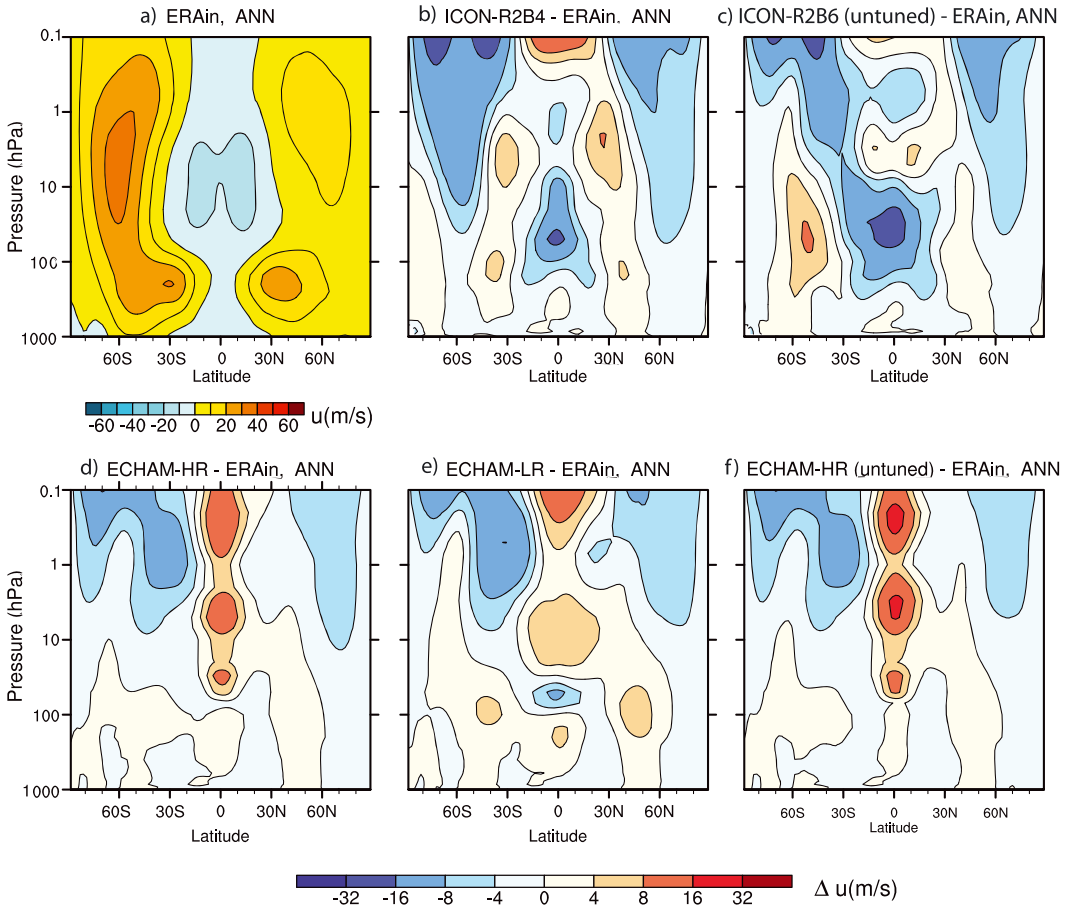

Figure 5. Multiyear annual and zonal mean zonal wind (m/s) for ERA interim (a), and anomalies w.r.t. ERA interim: ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e), and ECHAM-HR untuned (f).

3.2. Energy Budget

From the point of view of the global mean energy budget, it is difficult to clearly favor one model or one model configuration over the other. Energy budget terms for the different simulations are summarized in Table 4. The differences across the ECHAM-LR ensemble are not shown as they agree to within ±0.1 W/m^2 for all variables shown. Comparisons are made directly to CERES and to the synthesis by Stevens and Schwartz (2012) whereby the latter was also influenced strongly by the CERES measurements. Some consistent patterns that do emerge are the tendency for the magnitude of cloud radiative effects to decrease with increased resolution. As a whole the TOA longwave and shortwave cloud radiative effects in ICON-A are smaller in magnitude than ECHAM6.3 and further from what is observed, but their net effect is more similar to what is observed. The substantially colder surface (0.4 K, despite identical SSTs) explains differences in the surface energy budget in ICON-A relative to ECHAM6.3—for instance, larger sensible heat fluxes and less upward longwave radiation. Both ICON-A and ECHAM6.3 are similar in the extent to which the atmospheric energy budget is closed, with differences between the net flux at the top of the atmosphere and the surface generally less than or equal to 0.5 W/m^2 for ECHAM-LR and ICON-R2B4, and below 2.0 W/m^2 for the high-resolution simulations.

The pattern of the reflected shortwave radiation in both ICON-A and ECHAM6.3 is similarly well captured in comparison to CERES. This is illustrated using Taylor diagrams (Taylor, 2001) for the solstice seasons and both, the all and clear sky fields (Figure 6). Clear-sky values are extremely well correlated with the observations, but the variability is smaller. This could arise from a poor characterization of variability in surface albedo, variability in aerosol burdens (in association with mineral dust transport), but may also be a residual cloud effect in the CERES data. The simulated variability in the all-sky shortwave fluxes is better represented in ICON-R2B6 as compared to ICON-R2B4 and the tuned ECHAM6.3 simulations (Figure 6), which is indicative of a better representation of cloud variability as this dominates the all-sky signal.

Table 4
Energy Budget Terms and CRE from CERES Observations, SS12 Best Estimates, and Simulation of ICON-A and ECHAM6.3 in W/m^2

	CERES	SS12	ICON-A R2B4	ICON-A R2B6 Untuned	ECHAM6.3 LR	ECHAM6.3 HR Untuned	ECHAM6.3 HR
TOA							
SW down	340	340	340.3	340.3	340.4	340.4	340.4
SW up	99	100	99.0	96.6	102.3	97.4	98.1
LW up	240	239	240.8	242.1	238.6	242.4	240.9
CRE SW	−46		−43.0	−40.7	−47.8	−42.9	−43.8
CRE LW	28		22.7	21.7	24.5	21.8	22.9
Surface							
SW down	187	187	192.2	194.9	185.4	191.6	190.5
SW up	23	25	26.7	27.3	24.3	25.0	24.4
LW down	345	342	344.1	343.3	347.1	344.1	344.4
LW up	399	397	397.3	397.7	399.5	399.3	399.5
CRE SW	−50		−45.7	−43.3	−51.0	−45.7	−46.2
CRE LW	30		27.4	24.6	27.4	25.1	24.7
LHF		86	85.7	87.4	84.1	85.9	84.5
SHF		20	26.6	24.1	24.8	24.5	24.6
T			288.1	288.1	288.5	288.5	288.5
Net							
TOA	1	1	0.5	1.6	−0.5	0.6	1.4
TOA CRE	−18		−20.3	−19.0	−23.3	−21.2	−20.9
Surface		1	0.0	1.8	−0.2	1.1	1.9
Surface CRE	−20		−18.4	−18.7	−23.6	−20.7	−21.5
LHF + SHF		106	112.3	111.5	108.9	110.4	109.2

Note. CERES observations originate from "CERES_ EBAF_ Ed4.0" and "CERES EBAF-Surface_ Ed4.0" for the period March 2000 to February 2016 (Loeb et al. 2013, Kato et al. 2013). SS12 estimates are from Stevens and Schwartz (2012), who do not separate the upward and downward contributions to the net surface shortwave irradiance, which they estimate at 162 W/m^2, with a range of 23–30 W/m^2 for the upward surface SW irradiance. They also indicate an uncertainty of more than ±5 W/m^2 for the surface energy budget terms, and less than ±2 W/m^2 at the top of the atmosphere. ICON-A and ECHAM6.3 model output are averaged for the time period 1979 to 2008. TOA = top of the atmosphere; SW = shortwave irradiance; LW = longwave irradiance; CRE = cloud radiative effect; LHF = the heat flux from evaporation and sublimation; SHF = the sensible heat transfer from the surface; T = temperature.

3.3. Cloudiness

In terms of the spatial variability of cloud radiative properties and cloudiness, ICON-A and ECHAM6.3 show a broadly similar ability to capture the main features of the observations (Figures 7 and 8). In Figure 7, showing the seasonal climatology of the solstice seasons, the physical space dynamical core in ICON-A removes the spectral ringing that is evident, particularly near regions of great topographic contrast, at the grid-scale in the ECHAM6.3 simulations. Perhaps partly related to this issue, ICON-A also better represents the spatial distribution of the cloud radiative effect (CRE) in the subtropical stratocumulus regions, and there is some evidence that ICON-A better colocates CRE features relative to the land-masses of the maritime continent. And whereas the CRE features in the storm-tracks of the northern hemisphere during JJA are better represented by ICON-A, the generally poorer representation of the southern hemispheric storm tracks is also reflected in ICON-A's representation of CRE features in this region (Figures 3 and 7).

In Figure 8, the annual means of total cloud cover of the models are compared with the space-borne lidar cloud-amount measurements of CALIPSO-GOCCP (Chepfer et al., 2010), which were developed for GCM comparisons from the CALIPSO Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) Level 1 (Winker et al., 2007) data product. The comparison of the simulated cloudiness, taken directly from the models rather than from a lidar simulator, with the observations supports inferences drawn from the CRE analysis above. For instance, the ICON-A shows improved stratocumulus cloud cover in both the low- and high-resolution

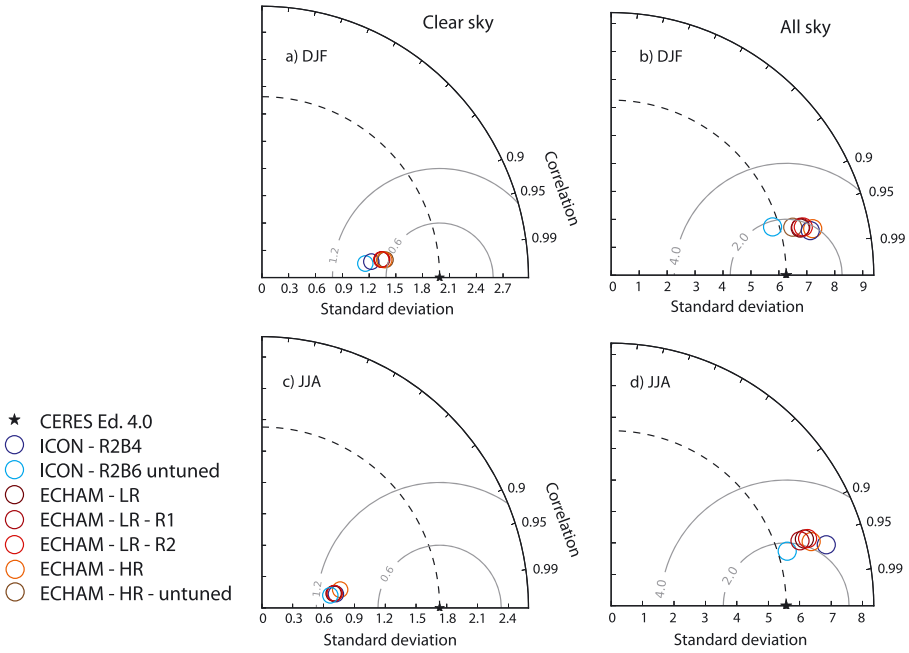

Figure 6. Taylor diagrams of reflected shortwave radiation at the top of atmosphere. Shown are the correlation coefficient, year-to-year standard deviation and the root mean square error following Taylor (2001) for Northern hemisphere (top) winter and (bottom) summer for (left) clear-sky and (right) all-sky conditions using CERES Ed. 4 for 2001–2014 as benchmark.

simulations compared to ECHAM6.3, although they are still slightly underestimated and too close to the equator compared to the observations. From the distributions of cloud cover and cloud radiative effects in Figures 7 and 8, a persistent problem for many models can be seen: The low-level clouds are insufficiently pervasive, yet their cloud-radiative effects are, if anything, too strong (Nam et al., 2012). This is especially evident in trade-wind regions. The comparison with the CALIPSO-GOCCP product also highlights deficiencies that were less apparent in the CRE fields, for instance, insufficient zonal contrast in cloudiness over the Indian Ocean in ICON-A, which is less apparent in ECHAM6.3. In addition, there is a too zonal and overly extensive SPCZ, especially for ICON-A. Over northern Eurasia the simulations tend to be cloudier than observed, though the extra tropical oceans tend to be less cloudy.

The zonally averaged cloud amount as a function of height and latitude shows that the vertical distribution of clouds in ICON-A and ECHAM6.3 are broadly similar across resolutions (Figure 9). One exception are clouds in ICON-A around 60°N, which are clearly closer to the observations in ICON-R2B6 than in ICON-R2B4 and the ECHAM6.3 simulations. Even without a lidar simulator, several features are evident from the comparison of the simulations with CALIPSO-GOCCP observations. These include too few low clouds in the tropics, a bias that would be even more pronounced if a simulator had been used and thus accounted for the inability of the lidar to see low clouds beneath optically deep high clouds. In contrast, there is some hint, but this is more difficult to establish without the simulator, that at higher latitudes, the cloudiness is more surface concentrated than observed. Focusing on the cloud-top distribution, ICON-A has a more hemispherically symmetric distribution of high clouds than is observed, or as is simulated by ECHAM6.3, but the latitudinal distribution of the highest clouds is better represented by ICON-A as compared to ECHAM6.3, in both hemispheres, and relatively insensitive to resolution.

Ground-based measurements of cloud distributions support the inferences drawn above regarding biases in profiles of cloudiness in the tropics and in the middle to high latitudes. For this purpose we compare the simulations to long-term measurements from two ground stations, one tropical maritime, and one central European station. The latter is the Leipzig site and the former the BCO (Stevens et al., 2016), that has been shown to be representative of the broader trade wind region (Medeiros & Nuijens, 2016). Profiles at these stations are compared to statistics drawn from a broader region within the simulations: from 10°N to 20°N and from 50°W to 60°W for the BCO and from 50°N to 53°N and from 9°W to 15°W for Leipzig. The homogeneity of the trade winds upstream of Barbados motivates the use of a larger sampling area in that case.

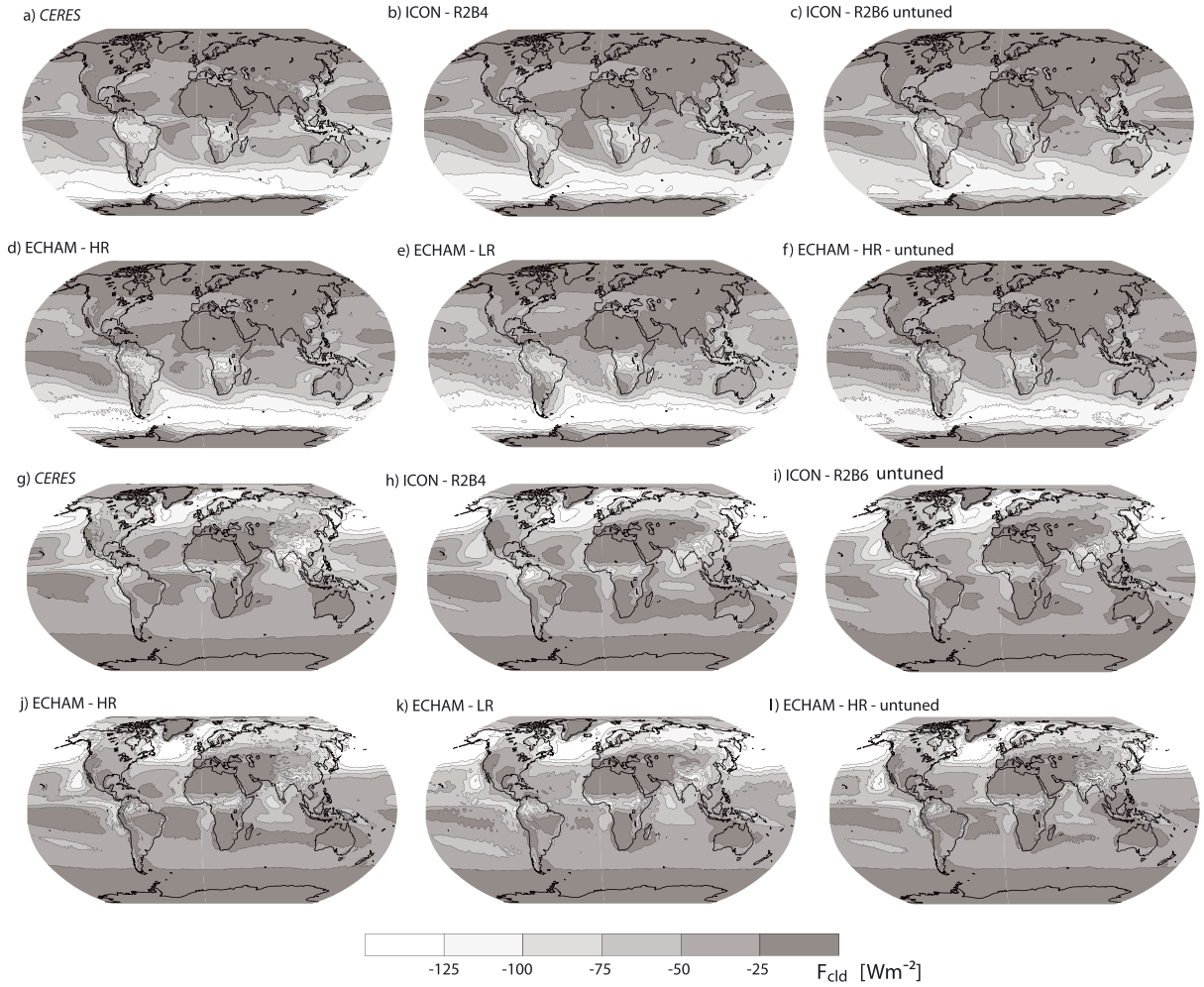

Figure 7. Cloud radiative effect averaged over DJF (a–f) and JJA (g–l). Shown are shortwave radiative effects of clouds from CERES Ed. 4 (a, g), ICON-R2B4 (b, h), ICON-R2B6 (c, i), ECHAM-HR (d, j), the ensemble mean of the low-resolution ECHAM-LR (e, k), and ECHAM-HR untuned (f, l); (period: 2001–2014).

Both ground sites make use of advanced remote sensing and the CloudNet (Illingworth et al., 2007) algorithms to characterize cloud distributions in height and time. Figure 10 demonstrates that ICON-A and ECHAM6.3 are similarly skillful in capturing the differences in cloudiness between the two sites, and with season at each site, individually. For instance, over the BCO both ICON-A and ECHAM6.3 show the tendency toward more top-heavy shallow cloud profiles in the winter season, a signature of the observations. However, the simulations generally underestimate cloud amount over the subtropical oceans, and overestimate cloud amount over midlatitude land. The tropical cloud bias is most pronounced in the lower troposphere, in association with fair-weather clouds. The midlatitude bias appears to mostly be due to too many high clouds, although low clouds are also too prominent during JJA in the comparison to the Leipzig measurements. All the simulations also fail to adequately differentiate the structure of low clouds, particularly the relatively higher cloud base over land, that is evident when comparing the observations over Leipzig in summer with the fair-weather profiles over Barbados.

3.4. Precipitation

The climatological spatial distribution of precipitation is also broadly similar in ICON-A as compared to ECHAM6.3, and both capture the main features present in the observations. This is illustrated with the help of Figure 11, which compares the annual climatologies of the simulations with that from GPCP. Differences of detail, which can have a profound effect on the structure of the tropics and tropical circulation, are nevertheless evident. Generally, ICON-A has a stronger tendency of the double ITCZ than ECHAM6.3. Apart from that, the ICON-R2B6 precipitation distribution is more similar to the observations over the western Indian Ocean, and into the maritime continent, where more precipitation locates over the major islands and

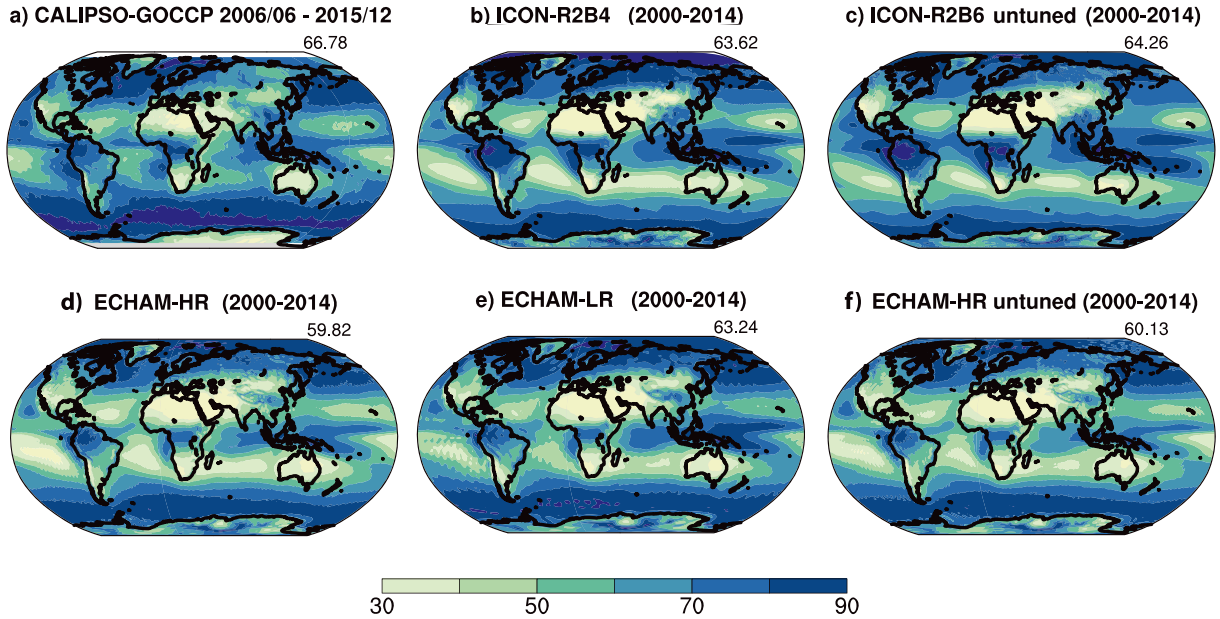

Figure 8. Distribution of annual mean total cloud cover and their global means (number atop the single figures) for CALIPSO-GOCCP (a), ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e), and ECHAM-HR untuned (f) are presented for the stated time periods.

topographical features. In ICON-A, precipitation extends further northward over the Bay of Bengal, and into southeast Asia. This may reflect the advantage of a physical-space dynamical core in representing sharp topographic gradients, or changes performed to the convection scheme, in particular allowing shallow clouds to rain. This latter effect would also be consistent with the generally better representation of precipitation over land over the west of northern South America in ICON-A, where ECHAM6.3 produces systematically too less rainfall. Precipitation over land is amplified with higher resolution, as expected (Demory et al., 2014), but it is suggested this can be tuned away as in ECHAM-HR (see ECHAM-HR untuned versus ECHAM-HR in Table 5).

The increase in precipitation over land (Figure 11 and Table 5), as well as further differences in the simulation of precipitation in ICON-A compared to ECHAM6.3 (see below), may be related to the removal of a differential (land-sea) minimum cloud depth for precipitation formation in the ICON-A implementation of the ECHAM6.3 physics, or the better ability of the ICON-A dynamical core to represent orographic land-sea contrasts. Not only is the transport of moisture from ocean to land enhanced, but the precipitation couples more strongly to the circulation in ICON-A than in ECHAM6.3. Regime sorted analysis of precipitation as function of omega at 500 hPa (not shown) shows a stronger relationship between the two in ICON-R2B4 as compared to ECHAM-LR. This stronger relationship is a common feature seen in observations. Thus, ICON-A better represents the observed physical connection between precipitation and omega (Davies et al., 2013; Peters et al., 2013). At the same time, regime-sorted analysis of precipitation as function of SSTs reveals weaker precipitation in ICON-A than in ECHAM6.3 for warm SSTs. For instance, precipitation peaks at a SST of 30 °C, both in ICON-A and ECHAM6.3, but the amounts are 6.8 versus 9.3 mm/day.

As do most climate models with parameterized convection, both ECHAM6.3 and ICON-A overestimate the fractional contribution of precipitation from low intensities (2 mm/day) to the total precipitation as compared to GPCP-observations in the Tropics (Figure 12). This bias is evident over both land and sea alike (Figures 12b and 12c) and if anything is more pronounced, in the ICON-A simulations, especially over the ocean. The ECHAM6.3 simulations have an unrealistically pronounced preferred rainfall rate near 10 mm/day over the entire tropics, especially over land (Figures 12a and 12c), a feature that is less evident in ICON-A. Thus, in particular, ICON-R2B6 is closer to the observations over land.

The El Niño-Southern Oscillation (ENSO) precipitation teleconnection gives an indicator of how well precipitation couples to SST anomalies, which in AMIP experiments are imposed. In Figure 13 the precipitation regressed against SST anomalies averaged over the Niño 3.4 region is presented for the December–February period, when ENSO variability is largest. As is the case for some precipitation features itself, the higher-resolution simulation of ICON-A (R2B6) has some characteristics, which more accurately

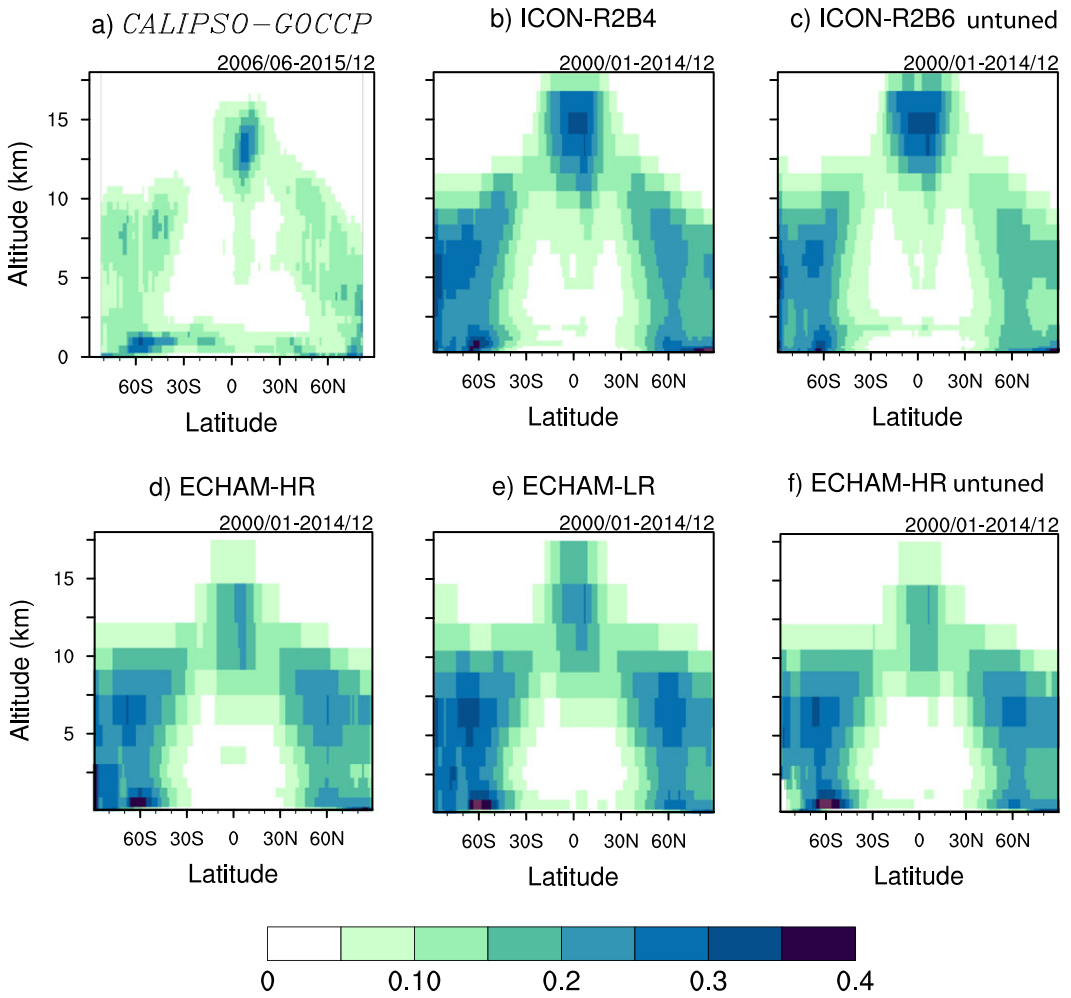

Figure 9. Annual and zonal mean of the vertical distribution of cloud fraction for CALIPSO-GOCCP (a), ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e) , and ECHAM-HR untuned (f) are presented for the stated time periods.

reflect what is observed, for instance, more rain over the western Indian Ocean, a more pronounced drying east of India, and a local maximum over equatorial South America. On the other hand, the rainfall response over the SPCZ is more realistic in ICON-R2B4.

4. Variability

To explore how this first release of ICON-A represents tropical variability relative to ECHAM6.3, we present an analysis of the interannual variability of the simulations in the extra tropics, in the form of the annular modes and blocking, and in the tropics, in the form of variability ranging from around weekly to intraseasonal timescales, as measured by equatorial waves.

4.1. Extra Tropical Variability

We calculate the North Atlantic Oscillation (NAO) following Hurrell (1995) as the leading Empirical Orthogonal Function (EOF) of PSL over the area 20°N to 80°N and 90°W to 40°W. The NAO is analyzed in DJF, when it is most pronounced. The Southern Annual Mode (SAM) is calculated as the leading EOF of the monthly mean PSL fields over the global extra tropical southern hemisphere (20°S to 90°S). For the analysis of the SAM we focus on the September to November (SON) season, when the interaction between the zonal flow and the planetary wave activity is strongest, thereby revealing a considerable impact, for example, on the lower stratosphere circulation and the southern hemisphere trade winds (Thompson & Wallace, 2000).

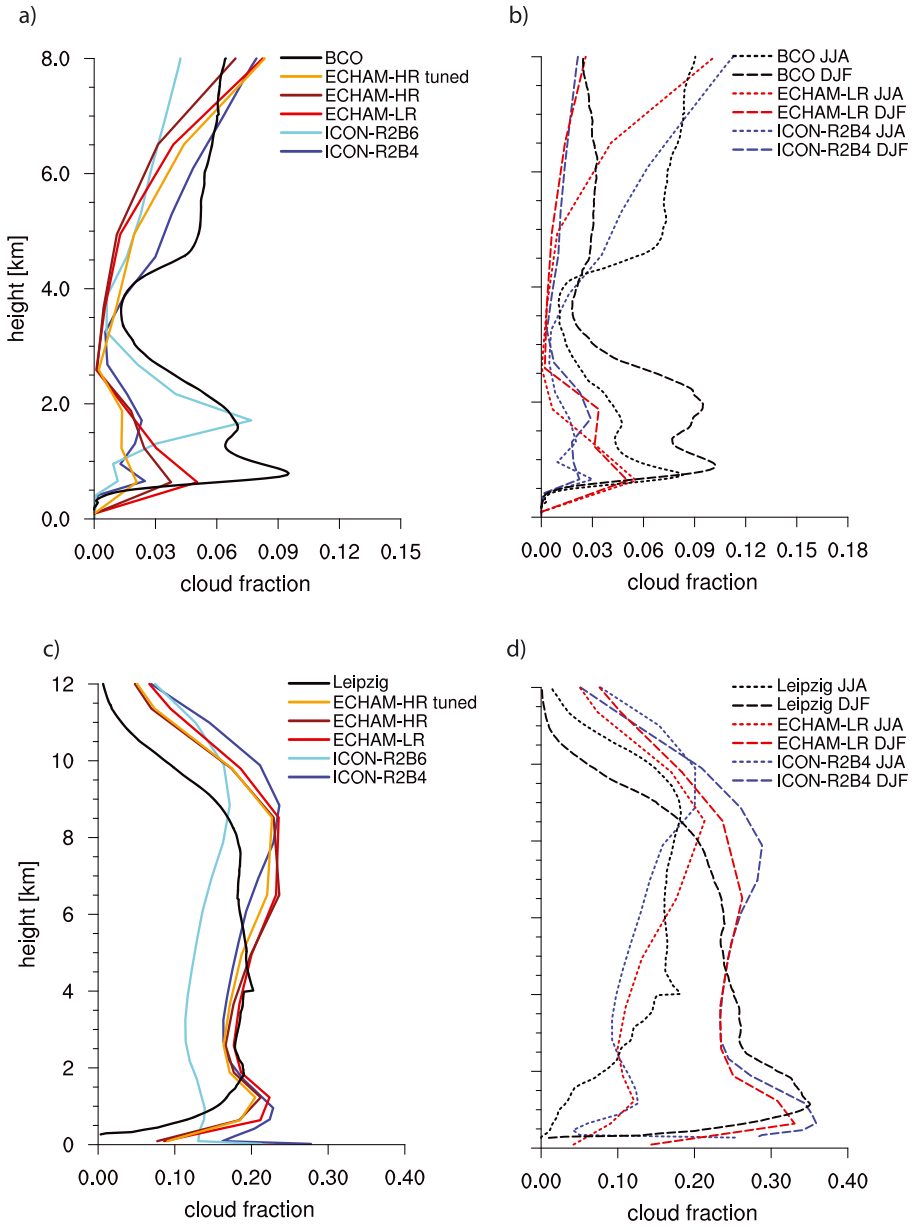

Figure 10. Annual (left) and seasonal (right) mean vertical profiles of cloud fraction for the Barbados (top) and Leipzig (bottom) observation sites. For the observations, the following CloudNet target categories are included in the cloud fraction estimate: cloud droplets only, ice, melting ice, and the mixed categories "drizzle/rain and cloud droplets," "ice and supercooled droplets," and "melting ice and cloud droplets."

All simulations capture the observed structure and strength of the NAO well (Figures 14a–14f), differing mostly in detail. ICON-A tends to underestimate the amplitude of the NAO dipole compared to the observations, consistent with its too weak stationary wave structure in the extra tropics (see section 3), and its center is displaced equatorward relative to what is observed. ECHAM-LR and ECHAM-HR (tuned) do a slightly better job in capturing the dipole position and strength. Its representation of the NAO explains less of the interannual variance in PSL (45.8% for ERA interim, 46.5% for ICON-R2B4, and 38.5% for ECHAM-LR), but it extends this mode of variability further to the west—particularly the high-pressure part of the pattern—than is observed. An increase of resolution without tuning leads to a slight degradation of the NAO in ICON-A and ECHAM6.3 with respect to both the strength and the amount of explained interannual variability. However, tuning ECHAM-HR reimproves the NAO.

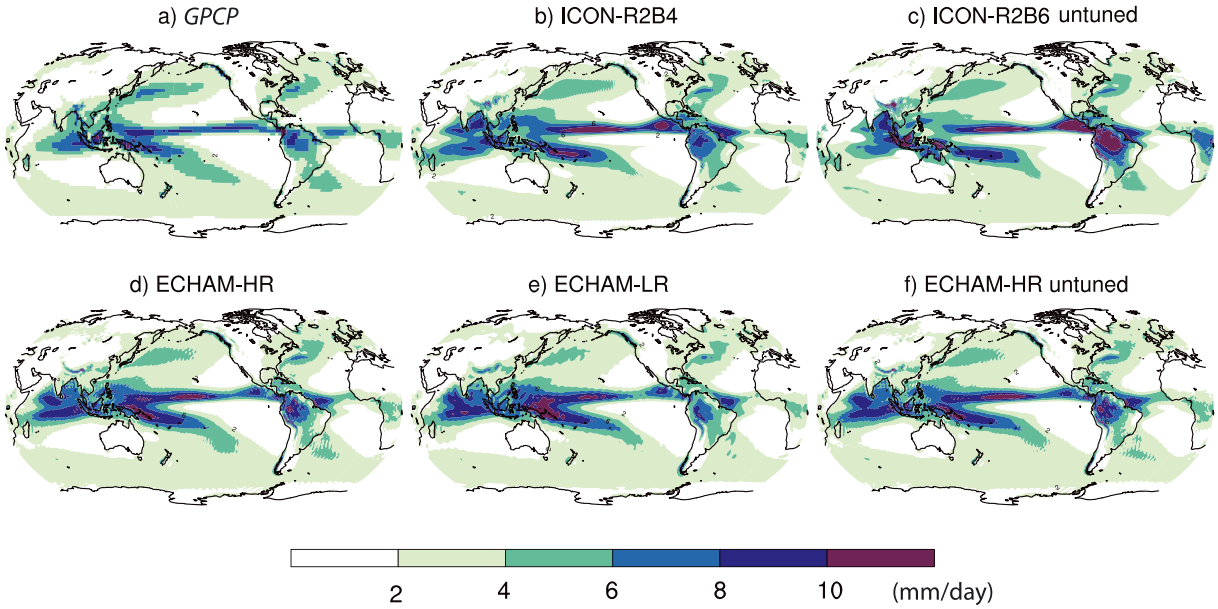

Figure 11. Mean precipitation climatology (1978–2013) from GPCP and in the simulations; GPCP (a), ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e), and ECHAM-HR untuned (f).

Table 5
Mean Precipitation and Moisture Convergence (Precipitation-Evaporation, Net) in the Simulations, Global, for the Tropics (30°N to 30°S), Land and Ocean and the Global Convective Rain Fraction (in Percent)

	GPCP	ICON-A R2B4	ICON-A R2B6 Untuned	ECHAM6.3 LR	ECHAM6.3 HR Untuned	ECHAM6.3 HR
Global						
Precipitation	3.60	3.40	3.48	3.37	3.45	3.39
Evaporation		−3.43	−3.50	−3.36	−3.43	−3.38
Net		−0.03	−0.02	0.01	0.01	0.01
Convective rain fraction		64.2	63.2	64.2	63.8	62.2
Global land						
Precipitation	2.95	2.61	3.04	2.45	2.55	2.48
Evaporation		−1.84	−1.92	−1.79	−1.84	−1.77
Net		0.78	1.11	0.66	0.72	0.71
Global ocean						
Precipitation	3.88	3.74	3.67	3.76	3.83	3.78
Evaporation		−4.11	−4.17	−4.02	−4.11	−4.06
Net		−0.37	−0.50	−0.27	−0.28	−0.28
Tropics						
Precipitation	4.08	4.17	4.31	4.10	4.19	4.08
Evaporation		−4.73	−4.81	−4.70	−4.75	−4.67
Net		−0.56	−0.50	−0.60	−0.57	−0.60
Tropics land						
Precipitation	4.16	3.66	4.65	3.26	3.45	3.31
Evaporation		−2.75	−2.93	−2.61	−2.66	−2.56
Net		0.91	1.71	0.64	0.79	0.75
Tropics ocean						
Precipitation	4.06	4.35	4.19	4.40	4.45	4.35
Evaporation		−5.44	−5.48	−5.44	−5.50	−5.43
Net		−1.09	−1.29	−1.04	−1.06	−1.08

Note. Precipitation and moisture convergence values are given in 10×10^{-5} kg·m^{-2}·s^{-1} for the time period 1979 to 2008. Differences among members of the ECHAM-LR small ensemble are less than the precision of the stated values in the table.

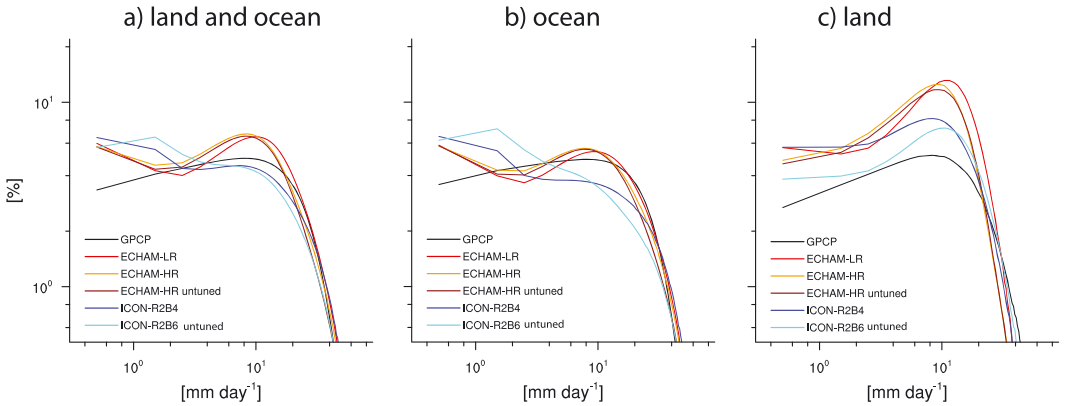

Figure 12. Fractional contribution in percent per rain rate to daily mean total precipitation in the tropics (30°S–30°N) for GPCP and the simulations. (a) All grid points, (b) ocean only, and (c) land only. All data are evaluated at T63 grid spacing.

For the SAM, ICON-A much better represents the observed amplitude and zonal asymmetry of the pattern (Figures 14h and 14i). This is evident in the equatorward extension of low pressure near 120°W. The centers of action in the high-pressure field, which has a tri-pole structure are much better represented in the higher-resolution models for both, ICON-A and ECHAM6.3. Taken together one might be tempted to say that the improvement of the SAM in ICON-A relative to ECHAM6.3 more than offsets the weak deterioration of the representation of the NAO. At the very least ICON-A is not clearly worse than ECHAM6.3 in its ability to represent interannual variability in patterns of PSL in the extra tropics.

4.1.1. Blocking

Another important mode of variability in the northern hemisphere extra tropics is the occurrence of quasi-stationary features, so-called blocking. Large-scale atmospheric blocking is important for regional climate variability and extreme weather events in the extra tropics (Carrera et al., 2004; Rex, 1950a, 1950b; Sillmann & Croci-Maspoli, 2009; Trigo et al., 2004). Here we perform a blocking analysis based on daily mean 500-hPa geopotential height using the blocking detection method described in Vial and Osborn (2012).

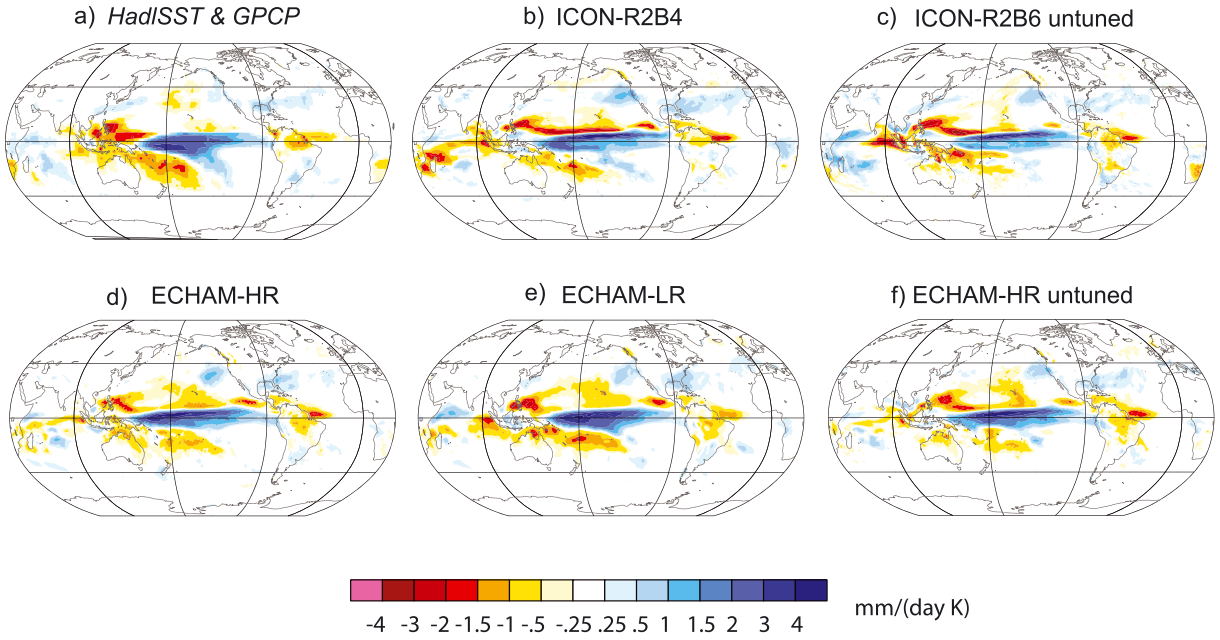

Figure 13. Precipitation response to the Niño 3.4 index, derived by regressing the DJF mean precipitation anomalies onto the DJF mean Niño-3.4 index anomalies. (a) HadISST1 and GPCP-V2.2 precipitation (see Table 2), (b) ICON-R2B4, (c) ICON-R2B6, (d) ECHAM-HR, (e) ECHAM-LR, and (f) ECHAM-HR untuned (period: 1979–2008).

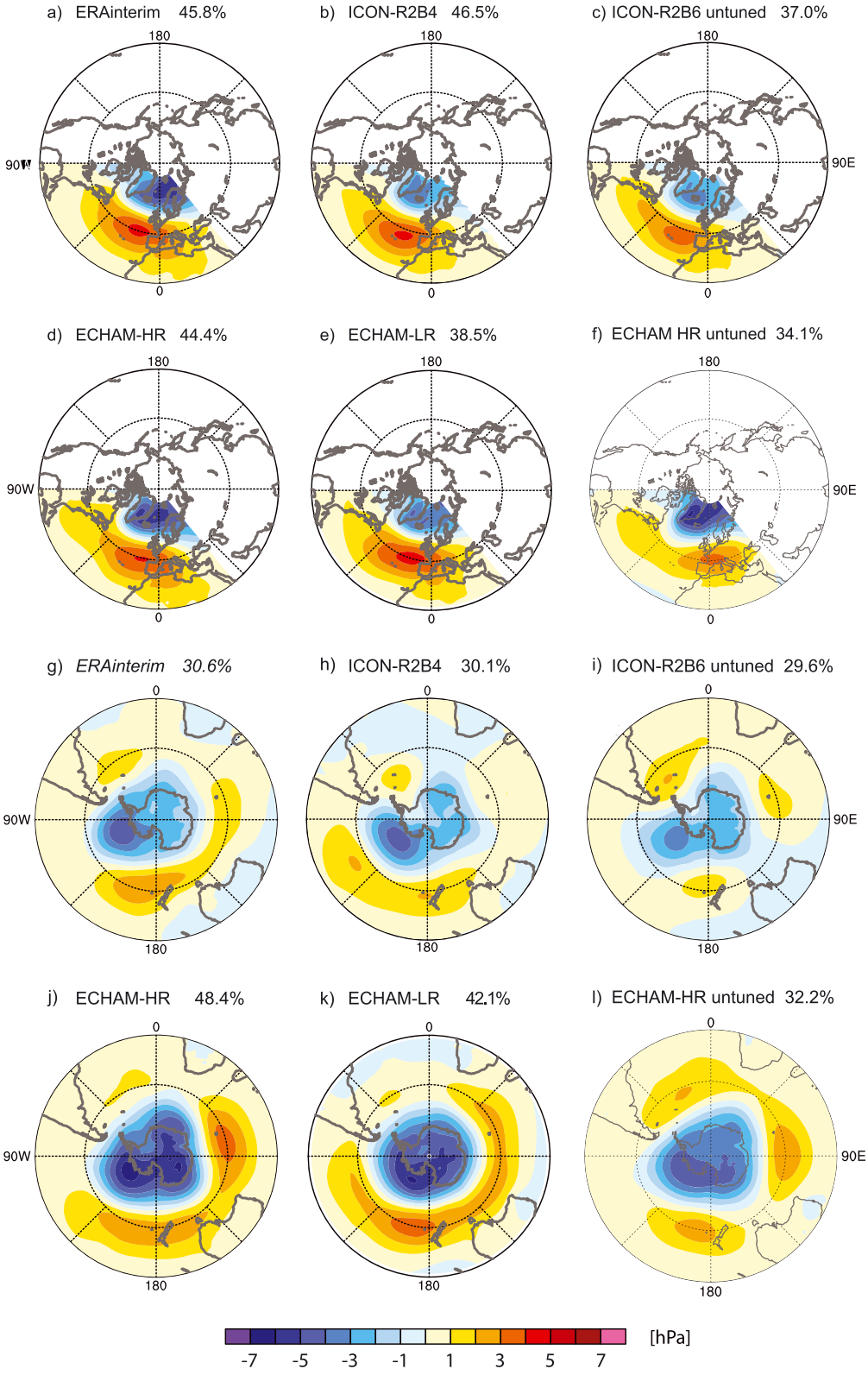

Figure 14. Winter (DJF) North Atlantic Oscillation (NAO) pattern (a–f) and Austral spring (September to November [SON]) Southern Annular Mode (SAM) pattern (g–l) for ERA-Interim (a, g), ICON-R2B4 (b, h), ICON-R2B6 (c, i), ECHAM-HR (d, j), ECHAM-LR (e, k), and ECHAM-HR untuned (f, l). The NAO/SAM is calculated as the first mode of variability of an Empirical Orthogonal Function-analysis of sea level pressure of the shown area (Hurrell, 1995; Thompson & Wallace, 2000); (period: 1980–2008/1979–2008).

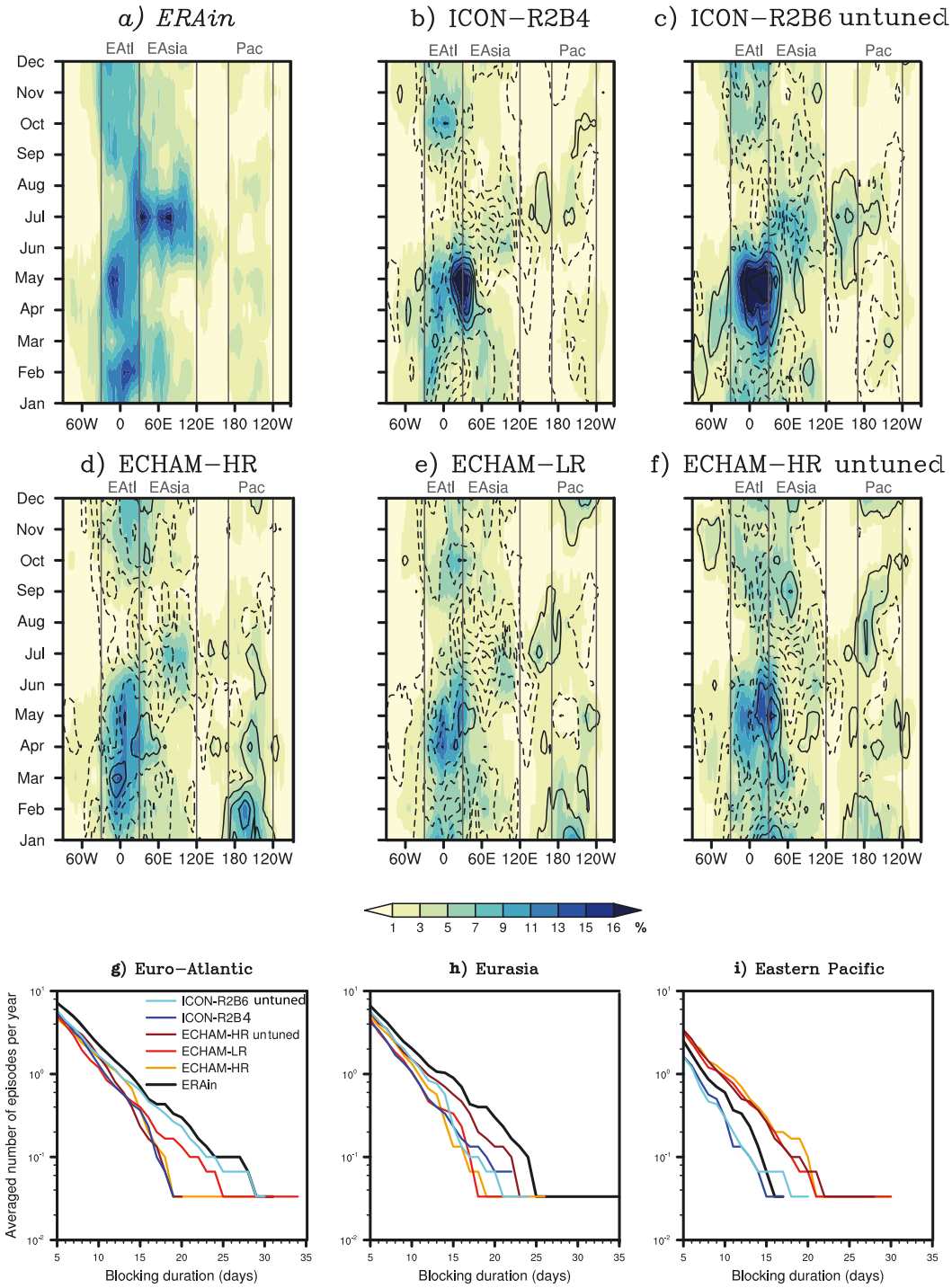

Figure 15. Upper and middle: Monthly climatology of large-scale blocking episode frequencies (in precent) as a function of longitude for ERA interim (a), ICON-R2B4 (b), ICON-R2B6 (c), ECHAM-HR (d), ECHAM-LR (e), and ECHAM-HR untuned (f). Contour lines in (b) to (f) represent the differences between the model simulations and the reanalysis. Underestimation of blocking frequencies is shown by the dashed lines, and overestimation by the solid lines. The thin vertical lines delineate the three regions of blocking occurrence: Euro-Atlantic (Eatl, 30°W–30°E), Eurasia (EAsia, 30°E–120°E) and Eastern Pacific (Pac, 170°E–240°E). Lower: Averaged number of large-scale blocking episodes per year as a function of duration (in days) for the Euro-Atlantic (g), Eurasian (h), and Eastern Pacific (i) sectors from ERA interim (black) and model simulations (colored lines); (period: 1979–2008).

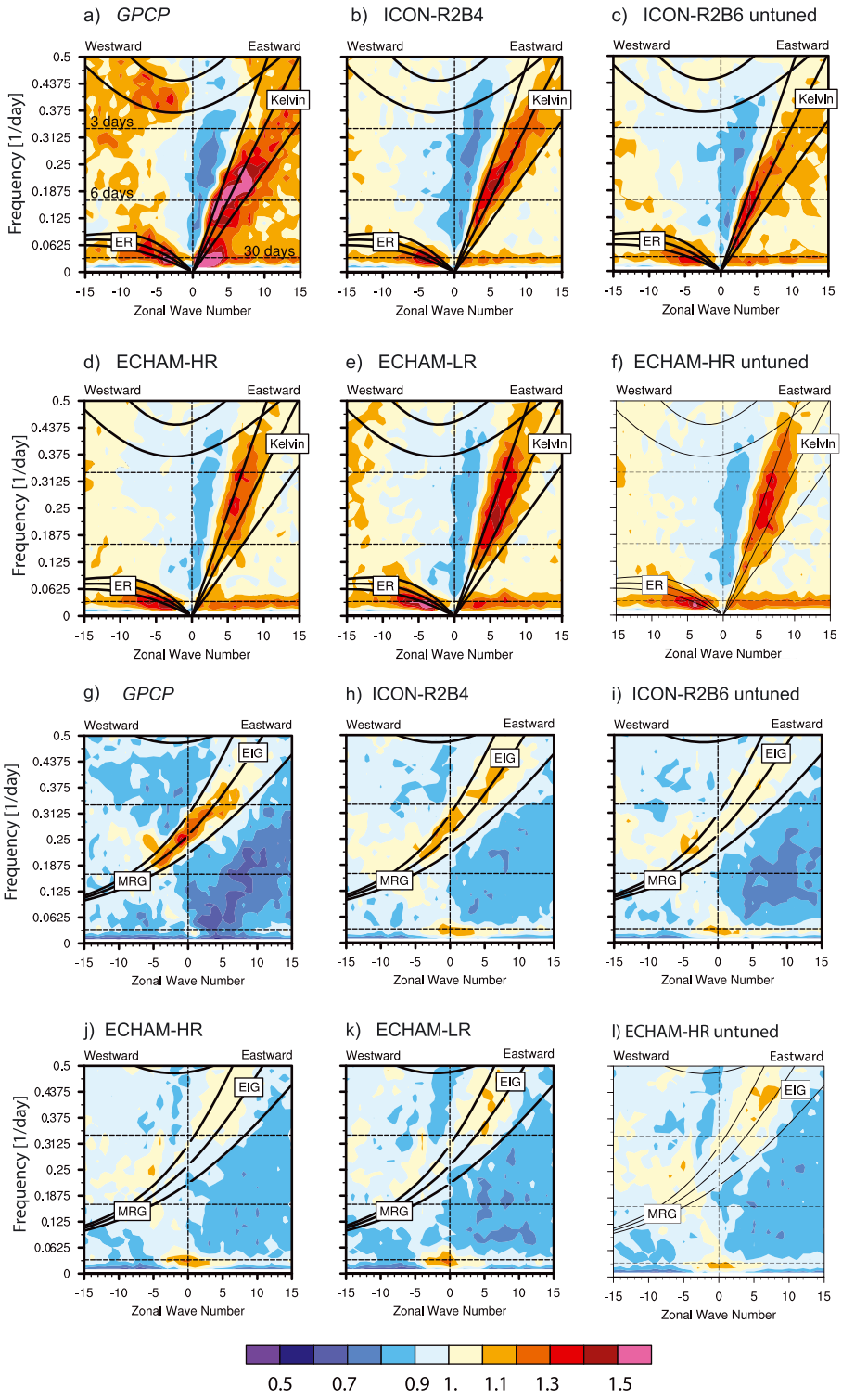

Figure 16. Wave number-frequency power spectra of the symmetric (upper) and antisymmetric (lower) component of precipitation, summed from 13°N to 13°S, plotted as ratios between raw power and the power of smoothed red noise background spectra. Positive/negative wave numbers represent eastward/westward propagating waves. Dispersion curves are shown for Kelvin and Eastward propagating inertio gravity waves (EIG) for equivalent depths of 12, 25, and 50 m (from right to left). Horizontal dashed lines represent periods of 3, 6, and 30 days (see (a)). In (a) the MJO signal appears with strong power on the eastward propagating part with wave numbers 1–3 and periods longer than 30 days. ERA interim (a, g), ICON-R2B4 (b, h), ICON-R2B6 (c, i), ECHAM-HR (d, j), ECHAM-LR (e, k), and ECHAM-HR untuned (f, l); (period: 1979–2008, for GPCP: 1997–2014; Kiladis et al., 2009).

There are considerable differences between the observed and simulated monthly climatological blocking frequency. This is illustrated with the help of Figures 15a–15f. Here we focus on the three well-known regions of blocking occurrence,: Euro-Atlantic (Eatl, −30°E to 30°E), Eurasia (EAsia, 30°E to 120°E) and Eastern Pacific (Pac, 170°E to 240°E). The cold seasons (winter and spring) account for a large part of the observed blocking activity in the Euro-Atlantic region, while the observed Eurasian block is mainly a summer phenomenon. Observed Pacific blocking events, while being less frequent, can occur most of the year (except in autumn). Figures 15g–15i presents the annual number of blocking events as a function of duration for the three sectors outlined above. Note that a logarithmic scale is used on the *y*-axis to illustrate the fact that the distributions of blocking lifetime follow an exponential decay (D'Andrea et al., 1998; Vial & Osborn, 2012).

Model biases common to all five simulations are (i) an underestimation of winter blocking frequencies over the Euro-Atlantic region (Figures 15b–15f), (ii) an underestimation of summer blocking frequencies and annual mean lifetimes over Eurasia (Figures 15b–15f and 15h), and (iii) a tendency for overestimating the frequency of Euro-Atlantic blocking episodes in spring, especially in ICON-A (Figures 15b–15f). The major differences between the two models are in the Pacific sector, where ICON-A simulates less frequent and less persistent blocking episodes and, as a result, is in better agreement with the reanalysis as compared to ECHAM6.3 (Figures 15b–15f and 15i). There is no systematic sensitivity in the model biases to resolution, but there is a tendency for an increased blocking frequency in the higher-resolution model versions (Figures 15b–15f), which in the Euro-Atlantic and Eurasian sectors tends to arise from an increase in the number of short-lasting events at the expense of longer episodes (cf. ECHAM6.3 in Figures 15g and 15h and ICON-A in Figure 15h). One exception stands out in the Euro-Atlantic, where increasing the resolution in ICON-A leads to a clear improvement of the blocking lifetime distribution (Figure 15g).

4.2. Tropical Variability

In the tropics, many important modes of variability depend on coupling with the SSTs. Some insight into tropical variability can, nonetheless, be gained by exploring the structure of convectively coupled waves. To assess the overall character of the convectively coupled equatorial waves, wave number-frequency analysis of daily precipitation, split into the symmetric and antisymmetric (with respect to the equator) components is commonly used (Crueger & Stevens, 2015; Jiang et al., 2015; Kiladis et al., 2009). The observed spectra, using GPCP input, reveal the strongest power for the waves indicated by the lines, representing the theoretical solutions of the shallow water theory (Wheeler & Kiladis, 1999). In Figure 16, positive wave numbers represent eastward propagating waves, and negative wave numbers westward propagation.

The simulations generally show less power associated with known equatorial waves than the observations. But if anything ICON-A yields a better representation of equatorial waves than ECHAM6.3. This is evident in both the presence and absence of power in the wave number-frequency domain. In the symmetric spectra ICON-R2B4 better represents low-frequency Kelvin waves and also has a hint of more correspondance with the observations in the low wave number but high-frequency westward propagating Equatorial Rossby waves (ER), often associated with tropical easterly wave disturbances and hurricanes. In the antisymmetric spectra, again ICON-R2B4 shows a more realistic power relative to ECHAM6.3 at frequencies and wave numbers associated with eastward propagating inertio gravity (EIG) or mixed Rossby gravity (MRG) waves. ICON-A also better represents the observed spectra by virtue of the lack of power associated with low wave number (equivalent depths larger than 50) Kelvin waves (Figure 16).

A prominent signal in the symmetric component of the observed spectrum at wave numbers between 1 and 3 and periods between 30 and 90 days is the Madden-Julian Oscillation (MJO; Figure 16a). Figures 16b–16f indicate that none of the simulations has a compelling MJO signal. From recent work with ECHAM6.3, we nevertheless know that the MJO can be tuned.

5. Summary and Outlook

This paper evaluates the first release of the new icosahedral nonhydrostatic atmospheric general circulation model (ICON-A). ICON-A, which will serve as the atmospheric component of the ICON Earth system model, presently uses (with only slight modifications) a suite of parameterizations for most sub-grid scale processes that was developed for ECHAM6.3. A tuned simulation with ICON-A at a relatively low resolution (160 km) is compared to a fourfold higher-resolution simulation with the same model, which has not been tuned.

Likewise simulations for the same period using the more established ECHAM6.3 climate model are also presented at low and high (twofold finer) resolution, the latter with a version utilizing the ECHAM-LR settings and, in addition, with a tuned version. The experiments presented are intended to document the climate of ICON-A for eventual users. They also shed some light on how aspects of the simulated climate depend on resolution, the dynamical core, and the extent to which they are influenced by tuning.

Overall, the simulations indicate that the present configuration of ICON-A, especially the tuned low-resolution ICON-R2B4 experiment, provides a compelling representation of the mean climate and its variability. Broadly speaking its climate is similar to that of the last release of the well-established ECHAM6.3 model. Improvements are noted in aspects of the simulations, which were known to be sensitive to the representation of orography, which is challenging in a spectral dynamical core. These include the representation of cloud fields over eastern-boundary currents, the latitudinal distribution of cloud top heights, and the spatial distribution of convection over the eastern Indian Ocean and the Maritime Continent. Overall, precipitation over land is enhanced, in particular in the high resolution ICON-A experiment and thus fits better with the observations. Tropical precipitation in ICON-A responds to El Niño SST variability broadly in better accord with the observations, particularly over the Indian Ocean. Some changes to the parameterizations that were introduced when porting the ECHAM6.3 physics to ICON-A also lead to improvements, for instance the removal of threshold convective depths required for rain production leads to the deletion of an unphysical dearth of rain intensities near 2 mm/day. An increase in the convective entrainment rate in the ICON-A implementation of the physics is thought to also explain slight improvements in the wave number-frequency spectra of equatorial waves, although an enhancement of the MJO could not be found. The changes to the convection also imply a warmer troposphere, which degrades some aspects of the climate, for instance, by shifting mass toward the poles.

Many biases familiar to ECHAM6.3 — and many other models for that matter – are also evident in the ICON-A simulations. These include a too zonal SPCZ, an inadequate representation of northern hemispheric blocking, a relatively poor representation of tropical intraseasonal variability, and a westerly jet that is too poleward in the southern hemisphere. The latter leads to the southern hemispheric eddy-driven jet being insufficiently differentiated from the subtropical jet, and large biases in cloud-radiative forcing in the southern hemisphere middle and high latitudes. ICON-A, as do many other models, simulates an insufficient coverage of subtropical clouds over the ocean, but so as to maintain about the right magnitude of cloud radiative effects, the simulated clouds tend to be too reflective. A novel analysis comparing the ICON-A and ECHAM6.3 simulations to ground based measurements for two stations, one over a midlatitude continental region, the other representative for the marine trade wind region, indicates that the models qualitatively capture many features of the differences between the cloud fields in the two locations, and their seasonal variability, but quantitatively they differ substantially from what is measured.

Compared to its lower resolution counter part, the higher (two-fold finer) resolution version of ECHAM6.3 has as good as, or even better, representation of almost every aspect of the climate that we looked at. The same is not true for the fourfold higher horizontal resolution (but same vertical resolution) version of ICON-A, which shows a strong degradation of the performance compared to its low-resolution counterpart. Two exceptions are in the representation of the interannual variability of PSL in the southern hemisphere and in the distribution of precipitation over tropical land relative to tropical ocean, and perhaps the teleconnection between tropical precipitation and ENSO like variability, both of which at least partly improve with resolution irrespective of which model is analyzed. A substantial difference between the ICON-A and ECHAM6.3 configurations used here is that the untuned high-resolution version of ECHAM6.3 improves its climate, which is not true for the high-resolution ICON-R2B6 simulation. This indicates that the scope for improving a model's climate by tuning and removing bugs is large for newer models that are still at a development stage, larger than simply increasing resolution without tuning the configuration. For mature models that had undergone extensive bug fixing and tuning in different configurations, like ECHAM6.3, the importance of tuning is much smaller, at least in the range of configurations considered here.

We believe that the potential to improve ICON-A by increasing the resolution is similar as compared to ECHAM6.3. This assumption is based on several reasons: Most recent experiments with ICON-R2B6 suggest changes of the sub-grid scale orography (SSO) representation, which could improve predominantly the southern hemisphere circulation. The latter actually revealed a strong bias increase in ICON-R2B6 compared to ICON-R2B4 (see Figure 2d). Therefore, we assume that a considerable amount of the overall ICON-R2B6 bias could be reduced by a properly represented SSO in the R2B6 configuration. In the tropics we

suppose that the bias increase—largely due to precipitation over land—is related to the convection scheme (see Figure 2c). The reason for this could be that the vertical resolution in ICON-R2B6 is not increased compared to ICON-R2B4. In ECHAM-HR the resolution is higher from about 1.5 km upward, improving the performance of the convection scheme. Experiments with ECHAM5 suggested that the horizontal and the vertical resolution should be chosen consistently (Retsch et al., 2017; Roeckner et al., 2006). Thus, it may help to increase not only the horizonal but also the vertical resolution in the ICON-R2B6 configuration. Apart from that, we expect the situation to change as one begins to replace parameterized processes with resolved ones.

A critical reader might ask what has been accomplished by investing so many resources in the development of a new model that has more or less the same properties as the old model. Being able to maintain ECHAM6.3's representation of the climate using a dynamical core that is mass conserving and is flexible enough to be run at grid spacings ranging from a few tens of meters to many hundreds of kilometers, which share the same software infrastructure as its host ocean and land models and which has been shown to scale on as many as 500,000 cores is, however, a quite satisfactory achievement. This is all the more satisfying given our experience in how much tuning and exception handling (identifying and fixing bugs) can improve the climate of a model.

Acknowledgments
The Max Planck Society for the Advancement of Science is thanked for their support of this research. Computational resources were made available by Deutsches Klimarechenzentrum (DKRZ) through support from the Bundesministerium für Bildung und Forschung (BMBF). T. Crueger has received funding by CliSAP, an excellence cluster at the University of Hamburg supported by the German Research Foundation. K. Peters, M. Sakradzija, and A. K. Naumann were supported by the Hans-Ertel Centre for Weather Research. This research network of universities, research institutes, and the Deutscher Wetterdienst is funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI). ERA-interim data sets are available from ECMWF (http://apps.ecmwf.int/datasets/data/interim-full-daily/). We acknowledge the CloudNet project (European Union contract EVK2-2000-00611) for providing the Leipzig and Barbados target classification data sets, which were produced by the Department of Meteorology, University of Reading, using measurements from the Leipzig and Barbados Cloud Observatory sites. The CloudNet data for Leipzig was downloaded from the ACTRIS Data portal. Helpful discussions with J. Buehl regarding the use of the CloudNet target classifications are acknowledged. CERES data were obtained from the NASA Langley Research Center CERES ordering tool at (http://ceres.larc.nasa.gov/). CALIPSO-GOCCP observations (orbit and statistics files) were downloaded from the CFMIP-Obs website (http://climserv.ipsl.polytechnique.fr/cfmip-obs/Calipso-goccp.html). GPCP Precipitation data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from http://www.esrl.noaa.gov/psd. The figures were produced using the NCAR Command Language (Version 6.2.1) [Software] (2014). Boulder, Colorado: UCAR/NCAR/CISL/VETS, https://doi.org/10.5065/D6WD3XH5. We thank K. Meraner for helpful comments on an earlier version of this manuscript. Primary data and scripts used in this paper are available via publications@mpimet.mpg.de.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., et al. (2003). The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-Present). *Journal of Hydrometeorology*, *4*, 1147–1167.

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Grassl, H., & Schulz, J. (2010). The Hamburg ocean atmosphere parameters and fluxes from satellite data—HOAPS-3. *Earth System Science Data*, *2*, 215–234.

Becker, T., Stevens, B., & Hohenegger, C. (2017). Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *Journal of Advances in Modeling Earth Systems*, *9*, 1488–1505. https://doi.org/10.1002/2016MS000865

Brinkop, S., & Roeckner, E. (1995). Sensitivity of a general circulation model to parameterizations of cloud-turbulence interactions in the atmospheric boundary layer. *Tellus A*, *47*(2), 197–220.

Carrera, M. L., Higgins, R. W., & Kousky, V. E. (2004). Downstream weather impacts associated with atmospheric blocking over the northeast Pacific. *Journal of Climate*, *17*(24), 4823–4839.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., et al. (2010). The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP). *Journal of Geophysical Research*, *115*, D00H16. https://doi.org/10.1029/2009JD012251

Crueger, T., & Stevens, B. (2015). The effect of atmospheric radiative heating by clouds on the Madden-Julian Oscillation. *Journal of Advances in Modeling Earth Systems*, *07*, 854–864. https://doi.org/10.1002/2015MS000434

D'Andrea, F., Tibaldi, S., Blackburn, M., Boer, G., Deéqué, M., Dix, M. R., et al. (1998). Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979–1988. *Climate Dynamics*, *14*(6), 385–407.

Davies, L., Jakob, C., May, P., Kumar, V. V., & Xie, S. (2013). Relationships between the large-scale atmosphere and the small-scale convective state for Darwin, Australia. *Journal of Geophysical Research: Atmospheres*, *118*, 11,534–11,545. https://doi.org/10.1002/jgrd.50645

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Demory, M.-E., Vidale, P. L., Roberts, M. J., Berrisford, P., Strachan, J., Schiemann, R., & Mizielinski, M. S. (2014). The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate Dynamics*, *42*, 2201–2225.

Fiedler, S., Stevens, B., & Mauritsen, T. (2017). On the sensitivity of anthropogenic aerosol forcing to model-internal variability and parameterizing a Twomey effect. *Journal of Advances in Modeling Earth Systems*, *9*, 1325–1341. https://doi.org/10.1002/2017MS000932

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., et al. (2018). ICON-A, the atmosphere component of the ICON Earth System Model: I. Model description. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1002/2017MS001242

Hagemann, S., & Stacke, T. (2015). Impact of the soil hydrology scheme on simulated soil moisture memory. *Climate Dynamics*, *44*, 1731–1750.

Heinold, B., Knippertz, P., Marsham, J. H., Fiedler, S., Dixon, N. S., Schepanski, K., et al. (2013). The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: Estimates from convection-permitting simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 4385–4400. https://doi.org/10.1002/jgrd.50402

Heinze, R., Dipankar, A., Henken, C., Moseley, C., Sourdeval, O., Trömel, S., et al. (2017). Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, *43*, 69–100.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., et al. (2001). Global precipitation at one-degree daily resolution from multisatellite observations. *Journal of Hydrometeorology*, *2*(1), 36–50.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*, 676–679.

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Delanoë, J., Pelon, J., et al. (2007). Cloudnet. *Bulletin of the American Meteorological Society*, *88*(6), 883–898.

Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., et al. (2015). Vertical structure and physical processes of the Madden-Julian Oscillation: Exploring key model physics in climate simulations. *Journal of Geophysical Research: Atmospheres*, *120*, 4718–4748. https://doi.org/10.1002/2014JD022375

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*, D05127. https://doi.org/10.1029/2011JD017139

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*, 2719–2740.

Kiladis, G. N., Wheeler, M. C., Haertel, P. T., Straub, K. H., & Roundy, P. E. (2009). Convectively coupled equatorial waves. *Reviews of Geophysics*, *47*, RG2003. https://doi.org/10.1029/2008RG000266

Klocke, D., Brueck, M., Hohenegger, C., & Stevens, B. (2017). Rediscovery of the doldrums in storm-resolving simulations over tropical Atlantic. *Nature Geoscience*, *10*, 891–896.

Loeb, N. G., Lyman, J. M., Johnson, G. C., Allan, R. P., Doelling, D. R., Wong, T., et al. (2012). Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, *5*, 110–113.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*, 748–766.

Mauritsen, T., Svensson, G., Zilitinkevich, S. S., Esau, I., Enger, L., & Grisogono, B. (2007). A total turbulent energy closure model for neutrally and stably stratified atmospheric boundary layers. *Journal of the Atmospheric Sciences*, *64*(11), 4113–4126.

Medeiros, B., & Nuijens, L. (2016). Clouds at Barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), E3062–E3070.

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., & Tomita, H. (2013). Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, *18*, 4922–4926. https://doi.org/10.1002/grl.50944

Möbis, B., & Stevens, B. (2012). Factors controlling the position of the Intertropical Convergence Zone on an aquaplanet. *Journal of Advances in Modeling Earth Systems*, *4*, M00A04. https://doi.org/10.1029/2012MS000199

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The "too few, too bright" tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*, L21801. https://doi.org/10.1029/2012GL053421

Nordeng, T. (1994). Extended versions of the convective parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. *ECMWF Techical Memorandum*, *206*, 41.

Peters, K., Jakob, C., Davies, L., Khouider, B., & Majda, A. J. (2013). Stochastic behavior of tropical convection in observations and a multicloud model. *Journal of the Atmospheric Sciences*, *70*(11), 3556–3575.

Pincus, R., & Stevens, B. (2013). Paths to accuracy for radiation parameterizations in atmospheric models. *Journal of Advances in Modeling Earth Systems*, *5*, 225–233. https://doi.org/10.1002/jame.20027

Pithan, F., Angevine, W., & Mauritsen, T. (2015). Improving a global model from the boundary layer: Total turbulent energy and the neutral limit Prandtl number. *Journal of Advances in Modeling Earth Systems*, *7*, 791–805. https://doi.org/10.1002/2014MS000382

Randel, D. L., Vonder Haar, T. H., Ringerud, M. A., Stephens, G. L., Greenwald, T. J., & Combs, C. L. (1996). A New Global Water Vapor Dataset. *Bulletin of the American Meteorological Society*, *77*, 1233–1246.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–311.

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*, 459–482. https://doi.org/10.1002/jame.20022

Retsch, M. H., Hohenegger, C., & Stevens, B. (2017). Vertical resolution refinement in an aqua-planet and its effect on the ITCZ. *Journal of Advances in Modeling Earth Systems*, *9*, 2425–2436. https://doi.org/10.1002/2017MS001010

Rex, D. F. (1950a). Blocking action in the middle troposphere and its effect upon regional climate: I. An aerological study of blocking action. *Tellus*, *2*(3), 196–211.

Rex, D. F. (1950b). Blocking action in the middle troposphere and its effect upon regional climate: II. The climatology of blocking action. *Tellus*, *2*(4), 275–301.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., et al. (1996). The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate (Report 218). Hamburg: Max Planck Institut für Meteorologie.

Roeckner, E. R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., et al. (2006). Brokopf Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *Journal of Climate*, *19*, 3771–3791.

Sillmann, J., & Croci-Maspoli, M. (2009). Present and future atmospheric blocking and its impact on European mean and extreme climate. *Geophysical Research Letters*, *36*, L10702. https://doi.org/10.1029/2009GL038259

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, C., & Wergen, W. (1989). The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. *Meteorology and Atmospheric Physics*, *40*(1), 28–60.

Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., et al. (2016). The Barbados Cloud Observatory—Anchoring investigations of clouds and circulation on the edge of the ITCZ. *Bulletin of the American Meteorological Society*, *97*, 787–801.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., et al. (2017). MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geoscientific Model Development*, *10*, 433–452.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). Atmospheric component of the MPI-M earth system model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*, 146–172. https://doi.org/10.1002/jame.20015

Stevens, B., & Schwartz, S. E. (2012). Observing and modeling Earth's energy flows. *Surveys in Geophysics*, *33*, 779–816.

Sundqvist, H., Berge, E., & Kristjánsson, J. E. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, *117*(8), 1641–1657.

Taylor, K. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Taylor, K., Williamson, D., & Zwiers, F. (2000). The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations. *PCMDI Rep*, *60*(28).

Thompson, D. W. J., & Wallace, J. M. (2000). Annular modes in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, *13*, 1000–1016.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800.

Tomita, H., Miura, H., Iga, S., Nasuno, T., & Satoh, M. (2005). A global cloud-resolving simulation: Preliminary results from an aqua planet experiment. *Geophysical Research Letters*, *32*, L8805. https://doi.org/10.1029/2005GL022459

Trigo, R. M., Trigo, I. F., DaCamara, C. C., & Osborn, T. J. (2004). Climate impact of the European winter blocking episodes from the NCEP/NCAR Reanalyses. *Climate Dynamics*, *23*(1), 17–28.

Vial, J., & Osborn, T. J. (2012). Assessment of atmosphere-ocean general circulation model simulations of winter northern hemisphere atmospheric blocking. *Climate Dynamics*, *39*(1-2), 95–112.

Watterson, I. G., Dix, M. R., & Colman, R. A. (1999). A comparison of present and doubled CO_2 climates and feedbacks simulated by three general circulation models. *Journal of Geophysical Research*, *104*(D2), 1943–1956.

Wheeler, M., & Kiladis, G. N. (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *Journal of the Atmospheric Sciences*, *56*, 374–399.

Winker, D., Hunt, W., & McGill, M. (2007). Initial performance assessment of CALIOP. *Geophysical Research Letters*, *34*, L19803. https://doi.org/10.1029/2007GL030135

Zängl, G., Reinert, D., Ripodas, P., & Baldauf, M. (2015). 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.