Recent progress and review of issues related to Physics Dynamics Coupling in geophysical models

Markus Gross[1], Hui Wan[2], Philip J. Rasch[2], Peter M. Caldwell [3], David L. Williamson[4], Daniel Klocke[5], Christiane Jablonowski[6], Diana R. Thatcher[6], Nigel Wood[7], Mike Cullen[7], Bob Beare[8], Martin Willett[7], Florian Lemarié[9], Eric Blayo[9], Sylvie Malardel[10], Piet Termonia[11], Almut Gassmann[12], Peter H. Lauritzen[4], Hans Johansen[13], Colin M. Zarzycki[4], Koichi Sakaguchi[2], and Ruby Leung[2]

[1]Departamento de Oceanografía Física, Carretera Ensenada-Tijuana 3918, Ensenada BC 22860, México.
[2]Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99354, USA.
[3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
[4]National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA.
[5]Hans Ertel Center for Weather Research, Deutscher Wetterdienst (DWD), Germany
[6]University of Michigan, Department of Climate and Space Sciences and Engineering, 2455 Hayward St., Ann Arbor, MI 48109, USA.
[7]Met Office, FitzRoy Road, Exeter, EX1 3PB, United Kingdom.
[8]CEMPS, Exeter University, Prince of Wales Road, Exeter, EX4 4SB, UK.
[9]INRIA, Univ. Grenoble-Alpes, LJK, CNRS, Grenoble F-38000, France.
[10]ECMWF, Shinfield Park, Reading, RG2 9AX, UK.
[11]Royal Meteorological Institute of Belgium, Ringlaan 3, Avenue Circulaire, B-1180 Brussels, Belgium.
[12]IAP Kühlungsborn, Leibniz-Institut für Atmosphärenphysik e.V. an der Universität Rostock, Schlossstraße 6, 18225 Kühlungsborn, Germany
[13]Applied Numerical Algorithms Group, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, USA

Corresponding author: M Gross, `mgross@cicese.mx`

Abstract

Geophysical models of the atmosphere and ocean invariably involve parameterizations. These represent two distinct areas: Subgrid processes that the model cannot resolve, and diabatic sources in the equations, due to radiation for example. Hence, coupling between these physics parameterizations and the resolved fluid dynamics and also between the dynamics of the air and water, is necessary. In this paper weather and climate models are used to illustrate the problems. Nevertheless the same applies to other geophysical models. This coupling is an important aspect of geophysical models. However, often model development is strictly segregated into either physics or dynamics. As a consequence, this area has many unanswered questions. Recent developments in the design of dynamical cores, extended process physics and predicted future changes of the computational infrastructure are increasing complexity. This paper reviews the state-of-the-art of the physics-dynamics coupling in geophysical models, surveys the analysis techniques, and illustrates open questions in this field. This paper focuses on two objectives: To illustrate the phenomenology of the coupling problem with references to examples in the literature and to show how the problem can be analysed. Proposals are made on how to advance the understanding and upcoming challenges with emerging modeling strategies. This paper is of interest to model developers who aim to improve the models and have to make choices on and test new implementations, to users who have to understand choices presented to them and finally users of outputs, who have to distinguish physical features from numerical problems in the model data.

1 Introduction

In the context of this publication geophysical models are weather, climate, and Earth system models that describe fluid dynamics of the oceans and the atmosphere as well as their interactions with various physical, chemical, and biogeochemical processes that occur within those fluids or at the Earth's surface. The aim of such a geophysical model is to predict a spatially and temporally discrete representation of the true solution. This true solution is defined by a set of equations describing the physics (e.g., balances of momentum, energy and mass) and chemistry (and possibly even the biogeochemistry) of the geophysical system. Discrete approximations, in space and time, to these equations are necessary in order to numerically solve these equations using a computer, to produce simulations approximating the original physical system in the form of a space-time average of the governing equations. Spatial and temporal discretizations are two distinct yet related aspects, and are symbolically shown in Figure 1a by the curved surfaces and the arrows perpendicular to the surfaces, respectively. Considering the finite resolutions that are practically affordable in terms of computational cost, some component models are further divided into sub components representing processes (phenomena) that are resolved or unresolved (under-resolved). In the atmosphere and ocean models, the sub components that describe the resolved fluid dynamics are commonly known as the dynamical cores or simply "dynamics", while the representation of unresolved or under-resolved processes is referred to as the subgrid-scale parameterization or simply "physics", which operate on time and space scales much below the model resolution.

In this context Physics Dynamics Coupling (PDC) is defined as the formulation and implementation of the coupling between any two (or more) physical components of the modelling system under consideration. The term physical component is used to represent any of: an individual physical parameterization; a collection of such parametrizations; the dynamical core (for example, of the atmosphere or the ocean); or a modelling subsystem (such as the atmosphere and ocean models in an Earth System Model). The formulation and implementation of the coupling should address the following issues:

- the compatibility of the thermodynamic formulation between components;
- the discrete representation of the interaction between components that represent a possibly vast (and vastly different) range of time and space scales;

- the possible use of different resolutions between components (including variable versus fixed resolutions);
- and different spatial and temporal discretizations of the governing equations (for example spectral versus grid point versus finite element).

Thererfore, as Figure 1a aims to illustrate, PDC is not limited to only the one dimensional interaction between physics and dynamics. A key challange in the above is the design of space-time integration schemes for the different components that, when combined, reproduce the space-time averaged behaviour of *the whole system* being modelled.

In part due to the very high level of complexity of the real-world system, those models are typically developed, evaluated, and applied in units called component models that correspond to the different target systems, for example the atmosphere, ocean, land, glacier, and sea ice. The schematic shown as Figure 1a includes only two component models for simplicity: the atmosphere and the ocean. Those components are inherently coupled to each other through the momentum, mass and energy exchanges at their interfaces.

The parameterizations are typically organized by processes, for example cumulus convection and cloud microphysics in the atmosphere, and lateral and vertical mixing in the ocean. Some of these processes are symbolized by clip art icons in Figure 1a. Different processes can, and do in real models, reside at different locations in the space-time domain. For example the characteristic time scales associated with cloud microphysics and planetary-scale advection are vastly different.

The wide ranges of spatial and temporal scales that are associated with the different elements in Figure 1a have naturally resulted in different foci in research and the compartmentalization of the model codes. This compartmentalization and separation is necessary in order to understand and gain insights into the complex system and to render the model development manageable and traceable.

This, however, leads to what is known as splitting, i.e., evaluating in isolation, in time and or space, the response and feedback of a process to the evolution of model state, assuming the other processes stay unchanged during a certain time interval or that processes are evaluated in a pre-determined hierarchy, sequentially progressing the model state from one process (or family of processes) to the next. While splitting is useful and unavoidable, it can also lead to undesirable features in the numerical solutions since many processes are linked (coupled) to each other. In fact, many interactions are known to exist between the dynamical cores and the parameterizations, between different parameterized processes, and between component models such as the atmosphere and ocean. The multiplicity of timescales and their broad span from microseconds to months means that the time averaging required by the numerical solution is highly non-trivial.

The modeling errors inevitably introduced by the splitting procedure as outlined above are the core theme of the present paper. This is arguably a key, yet under investigated topic (however the community is starting to embrace the topic [*Gross et al.*, 2016]). In the past, the much lower spatial resolution and much simpler model formulation have been the dominant sources of model error. In recent years, however, rapid enhancement of computing capabilities has allowed for substantial increase in model resolution as well as the incorporation of much more comprehensive description of subgrid-scale phenomena. Examples of the latter include the life cycles of atmospheric aerosols and cloud droplets, which involve processes at spatial scales of nanometers to microns. If the coupling between the components does not "transport" sufficient information back and forth from the dynamical core to the physics, then the most accurate scheme in the dynamics alone will not improve the quality of the model simulations as may be otherwise expected. Sufficient information is to be understood as each component being able to utilise the input from other components in a physical meaningful and compatible way. For example, high order dynamics will not be able to reveal its full potential if coupled with low order physics. The low order physics cannot react physically to

the high order information from the dynamics. Nevertheless the physics will create forcings which in turn will influence the dynamics significantly. Thus numerical issues in the parameterized physics and in the process coupling can be bottlenecks in the reduction of overall model error. Also, in the prediction of the discrete representation (which may be point wise or space-time average over the grid box) of the true solution a subgrid model is required, but the subgrid model can only be formulated by assuming a scale separation, between the resolved and unresolved scale. This scale seperation depends on reolution and becomes more difficult and or questionable as resolution increases (c.f., sections 5 and 9).

The present paper presents examples and reviews the ongoing efforts in addressing those issues, and discuss an overarching topic of thermodynamic consistency across model components. Also discussed are a hierarchy of analysis methods aiming at furthering the understanding and provide a means of analysis of the intricate interactions, using mathematical analysis, reduced equations, model with simplified physics, and full model analysis.

The remainder of the paper is organized as follows:

Section 2 reviews the historical development of the field, indicating key publications and advances, analysis and tests performed so far and the evidence available. *Sections 3* focuses on issues related to process splitting in the time stepping algorithm, which can strongly affect model behavior in various ways. Process splitting is commonly used to make the approximation of each process more practical and to obtain an easy-to-maintain modular structure of the model source code. It may be argued that this is a pragmatic and/or convenient choice. Since different processes can produce compensating effects or can compete for resources, large numerical errors can result if processes are allowed to operate in isolation for periods of time longer than the typical time scales of their interactions. Solution convergence with respect to time step size is discussed in section 4 in the context of the spatial resolution being kept constant. While increasing only the temporal resolution does not guarantee that the solution will converge to the observed atmospheric motions in the real world, it is argued that the convergence analysis can be informative and help to improve the understanding of the process interactions in a model.

The following section (*section 5*) then proceeds to discuss that while analytically a convergence study (in time and space) will eventually converge towards the solution of the Navier Stokes equations, this would require resolution of the Kolmogorov scale, which is far from current resolutions and computational capability. Therefore, alternatively, an artificial scale separation has to be imposed if convergence is required or desired.

Section 5 illustrates methods of evaluating the coupling strategies. It is first necessary to distinguish the parts of the physics that force the system (e.g. radiation) from those (the subgrid model) that represent subgrid dynamics and thermodynamics and vanish at infinite resolution. Coupling of the subgrid model can be evaluated using a (reduced) continuous equation set. This defines a scale separation and allows a precise definition of the required solution in which the subgrid model plays an important part. In particular, the accurate reproduction of asymptotic limits where subgrid transports play a key role can be checked. Different coupling strategies can thus be analyzed in a less abstract way than the analysis by *Dubal et al.* [2006]. The importance of validating subgrid models and coupling strategies against the averaged data from much higher resolution models is illustrated using a convection example.

Section 6 emphasizes that ideally there would be a standard test procedure and established benchmark results across a whole range of models, with tests that isolate the components while still reflecting the model complexity and hence maintaining relevance. However, so far the design and testing strategy for simplified tests which stress the coupling and provide useful results has proven very difficult indeed. A proposal for idealized testing of the coupling is made and some results from already implemented tests are provided.

(a)

(b)
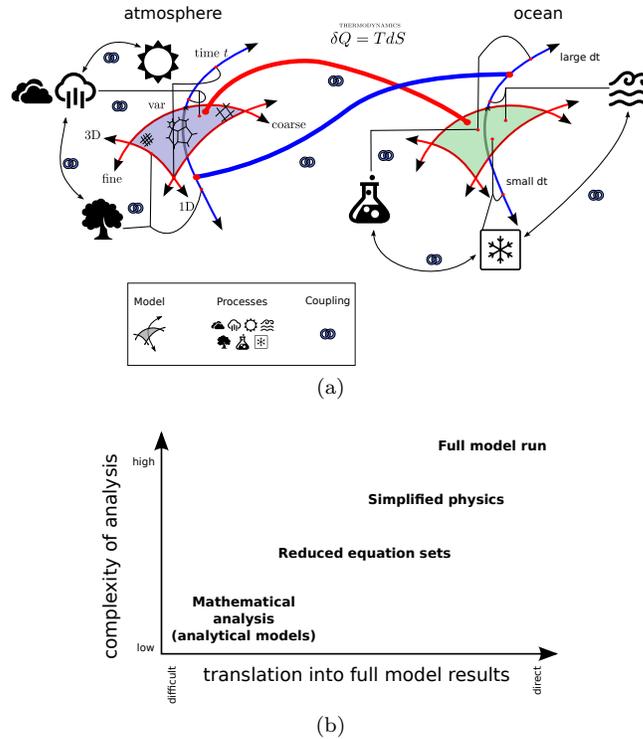
Figure 1. Schematic representation of PDC. For simplicity panel a) shows only two models, an ocean model and an atmosphere model. Both of these have spatial scales (here indicated by the plane with red lines) and temporal scales (indicated by the blue axis). These are coupled (thick lines), i.e. one domain in the spatial plane maps into the spatial plane of the other model (thick red line) and similarly in the temporal axis (thick blue line). In the spatial plane aspects such as grid type, fixed versus variable resolution, one dimensional vs three dimensional and fine versus coarse are shown as some of the aspects of the spatial resolution that can vary in between models and does not necessarily have a straight forward mapping. Then each of these models has its own ecosystem of parameterizations (an arbitrary set of processes was chosen here for illustration only), which interact with the model and themselves - via coupling. These parameterizations as well occupy potentially - or almost certainly - different areas on the spatial plane and temporal axis. All of this exists in front of a background problem of thermodynamics, which ultimately governs them all (or is ought to, anyhow). Panel b) shows the four tier scheme of investigation, ranging from (by necessity) abstract analysis, via reduced equation sets (with less necessity for abstraction) to simplified physics tests and finally full model runs. The complexity of the analysis increases from one to the other. The manner in which the results and conclusions from the experimentation can inform the production runs ranges from "difficult", i.e. results are expected in the form of guidance or informing a choice that needs to be made in the design phase, to "direct", meaning that a benefit can be demonstrated straight away by producing an improved forecast.

Section 7 focuses on the coupling in between different models, as depicted in Figure 1. An exchange of information has to occur on both the space plane and the time axis. Numerical results suggest a correlation between coupling errors and model uncertainties. Ocean-Atmosphere (OA) coupling requires the careful study of the mathematical and numerical formulation of physical parameterizations. In analogy to section 6, a lack of simplified models and reference test cases is observed.

Section 8 summarizes the state of the current scientific discussion around the thermodynamic compatibility. Even with ideal coupling in the time and space domains, different models are coupled. This has to be done in accordance with the laws of thermodynamics. The dynamics of a model describes exclusively reversible phenomena, whereas the physics of a model describes irreversible phenomena.

Sections 9 and 10 discuss the complexity of the interaction of parameterizations and resolution change. As resolution increases more detail is included in the solution to the governing equations, leaving less for the parameterizations. An example, amongst others, are mesoscale eddies in the oceans. Also some assumptions cease to be valid, such as a single continuity equation in the presence of sea ice. In this paper, first issues relating to uncertainties in the parameterizations themselves as horizontal resolutions approach and pass convection permitting regimes, the gray zone, are defined and illustrated with examples. Limitations of the current models in the gray zones alongside a short review of the progress to date are presented and illustrated using the example of the *Ala*din A*ro*me (ALARO) configuration of the *A*ire *L*imitée *A*daptation dynamique *D*éveloppement *I*nter*N*ational (ALADIN) model [*De Troch et al.*, 2013].

Section 10 discuss new and emerging modeling strategies of separating physics and dynamics grids (10.1), and how time-stepping/process-splitting (sections 2 and 3) and scale-awareness of deep convection (section 9) can interact and pose a challenge to models using spatialy varying horizontal resolution (10.2)

Section 11 summarizes the contributions from the different areas in the context of the whole picture elaborated on in the individual sections, aims to strengthen the awareness of this aspect in the community and invites the community to resolve and understand the issues jointly.

2 Historical review

The history of PDC probably starts with the first General Circulation Model (GCM) simulations. In the late 1960s, *Manabe and Bryan* [1969], at the Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey, presented one of the first simulations of the global atmospheric circulation coupled to ocean processes.

There were two models: one for the atmosphere, developed by Manabe, a meteorologist, who developed the atmosphere model; and one for the Ocean, developed by Bryan, an oceanographer with meteorological training. This separation of model domains remains today, nearly 50 years later. Manabe and Bryan joined forces to create a computational system that coupled their models. The winds and rain would contribute to the ocean currents, and the sea-surface temperatures and evaporation contribute to the circulation of the atmosphere. They soon realized that the coupling required much more than just passing on the forcing from one model to the other. Simply computing the fluxes in one model and applying them to the other did not yield a satisfactory outcome.

In the early days the coupling problem was a practical one: how to combine the different time scales and fit the model to the available computing capacity. The challenge was to get a working model.

Over the next twenty years computational resources were increasing at an accelerating rate. This had two implications. First, there was the capacity to run more and more coupled models and to increase the complexity. And second, with more data available it became possible to study the interaction, the coupling in more detail. The emphasis shifted from getting it running in the first place to investigating and contrasting different strategies and formulations. The literature details a long series of investigations and reports of problems faced when coupling the dynamical systems to their forcings.

Lander and Hoskins [1997] investigated and discussed the scales generated by an atmospheric model and argued that not all of them should be utilized in the physical parameterizations, because they are artifacts of the solution procedure, and not part of the solution itself. If the scales are close to the truncation limit and the model is not strongly damped there is significant noise present in the solution. If, partly in response to the noise, the model is damped then it is most strongly so near the truncation limit. Also the discretization error, which is always present, is most significant here. They therefore suggest that these "unbelievable" scales should not be used in the parameterizations and recommend the use of a coarser grid to evaluate the forcing from the parameterizations. The non-linear character of some physical parameterizations and their sensitivity to small perturbations can otherwise quickly lead to the growth of noise, rather than a correct approximation of the physics of the system (cf. sections 9 and 10.1). Interestingly recent work carried out at the European Centre for Medium-Range Weather Forecasts (ECMWF) goes in the opposite direction, i.e. using higher spatial resolution for the parameterizations. Their results clearly show benefit from doing this, as further elaborated in section 9.

Caya et al. [1998] investigated temporal aspects of the coupling. They discuss the effect of "splitting", i.e. applying the parameterizations subsequently to time stepping the dynamics, in combination with long time steps (15 min), such as admissible by Semi Lagrangian (SL) models. They argue that the splitting error can become unacceptably large.

Probably one of the first studies utilizing the full model and varying coupling parameters systematically is *Williamson* [1999]. Here the grid of the physical parameterizations and scale of the external surface forcing are held fixed while the horizontal resolution of the dynamical core is increased. This is shown to aid the convergence of tropical Hadley circulation, with increasing dynamical resolution, however it does not converge if the physics grid is not held constant. This is attributed to the forcing of smaller scales from the physics, indicating that the parameterizations at the coarser grid do not include their own forcings from the finer scale, i.e. missing processes.

Focusing on the vertical and in particular the resolution of the physical parameterizations, *Molod* [2009] analyzes the full model response to a refinement of the vertical physics grid. It was found that this benefits fields which are computed directly in the physical parameterizations, and in the vertical structure of the relative humidity and mass stream function, in line with the *Williamson* [1999] result, i.e. resolving some of the processes that are not captured by the parameterizations at coarse resolutions.

Wedi [1999] shows that the model performance can be improved by grouping certain parameterizations together and using predictors to improve the input from the dynamics into the parameterizations.

The suite of parameterizations is split into two groups. One to be evaluated at the arrival point and the other at the departure point. When compared with a simpler fractional stepping (or sequential or time split scheme) the following benefits are observed: second order accuracy, increase in stability, reduction of the time step dependence and numerical noise, improved mass conservation, more accurate forecasts with respect to the root mean square error and anomaly correlations and improved tropical cyclone tracks.

It becomes apparent that there are several options and clearly, some options are better than others. It is not always feasible to construct a new coupling scheme from scratch, imple-

ment it and test it in fully operational forecast mode in order to determine if it is better or not. Some form of analysis would be desirable, to test ideas and evaluate potential performance improvements.

Extending the framework presented by *Caya et al.* [1998], both in complexity of the sample problems as well as the coupling mechanisms, *Staniforth et al.* [2002a] and *Staniforth et al.* [2002b] analyze the explicit, implicit, split-implicit and symmetrized split-implicit coupling. They highlight that the stability of the explicit coupling is very restrictive for fast damping processes, such as vertical diffusion in the boundary layer at high resolution, thus rendering the explicit coupling computationally inefficient for practical applications. The authors show that this can be addressed by using implicit coupling, however it leads to "a highly nonlinear and computationally difficult and expensive problem to solve". The split implicit coupling addresses this but reduces the accuracy.

Back to full model analysis, *Williamson* [2002] reports statistically relevant differences when comparing time-split (sequential) and process-split (parallel) couplings to a simulation with the original version of the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM) Version 3 (CCM3). However, owing partly to the small time step used, these differences were small, highlighting the difficulty in clearly differentiating better from worse coupling mechanisms using the full model output alone. See also section 3.

Cullen and Salmond [2003] present a predictor corrector scheme that can give some of the advantages of a fully-implicit scheme and show that the use of more than one physics evaluation per time step significantly improves the accuracy in a model problem. An attempt is made to classify slow and fast processes. Using the predictor scheme short-time variability is reduced and a transfer from convective to dynamic precipitation observed in consequence. In what is possibly so far the most convincing demonstration on what difference the temporal coupling can make on a forecast, *Beljaars et al.* [2004] argue that, for the ECMWF Integrated Forecasting System (IFS), sequential splitting (tendencies of the explicit processes are computed first and are used as input to the subsequent implicit fast process) is preferable over parallel splitting (tendencies of all the parameterized processes are computed independently of each other) for problems with multiple time scales, because a balance between processes is obtained during the time integration. See also section 3.

In an analytically tractable framework - as mentioned above - this practical demonstration of the benefits of the sequential splitting is followed up by *Dubal et al.* [2004, 2005, 2006], using mathematical analysis. They conclude that while some advantages exist for parallel splitting over sequential splitting (e.g., parallel computation and not requiring an ordering of physical processes), the sequential-split methods are more flexible when it comes to eliminating splitting errors.

The issue of ongoing non-convergence of model results is highlighted in *Williamson* [2008]. Analyzing convergence runs with resolution varying from T42 to T340 truncation and 40 to 5 minutes time step, convergence is observed in larger scales of the zonal average equatorial precipitation and equatorial wave propagation. However, a non-convergent mass shift from polar to equatorial regions and a zonal average cloud fraction decrease was observed. In general, the simulations show a sensitivity to the parameterizations time step as well as to the horizontal resolution. Even when the time step is fixed, global averages do not converge with increasing resolution for all fields. For example, there is no indication that either precipitable water or precipitation converges with increasing resolution. This renders the analysis of the coupling more difficult as it is not immediately obvious how to generate a reference solution that can be used to test for coupling errors, using full model runs. The problem of attribution of errors has been more recently investigated in *Wan et al.* [2013]. Only relatively recently has the importance of the coupling in its own right been recognized and efforts to address these issues on a multi disciplinary level are underway [*Gross et al.*, 2016].

This brief - and by no means comprehensive - review is meant to illustrate the vast array of considerations made in the context of coupling models and their parameterizations. Furthermore the difficulties faced when attempting to analyze the impacts and consequences and when designing new coupling algorithms and, indeed, parameterizations, is illustrated. The remainder of the present publication will illuminate the problem from different directions and highlight current progress in the respective areas, starting with splitting of processes in the discrete model.

3 Time stepping errors introduced by splitting

Weather and climate models rely on discretizing time and space dimensions in order to make calculations computationally affordable. Numerical errors from spatial and temporal discretization can be closely related in some situations but more straightforward to separate in other cases. In this section, the focus is exclusively on time discretization by discussing model behaviors with fixed spatial resolution and varied time steps.

3.1 Impact of time stepping errors

Time Step size can have a large impact on the behavior of weather and climate models. For example, in one version of the European Centre - Hamburg model (ECHAM) Version 5 (ECHAM5) climate model [*Roeckner et al.*, 2003, 2006], the equilibrium climate sensitivity (i.e., the global-average equilibrium surface temperature change in response to doubling carbon di-oxide (CO_2)) was found to vary by a factor of two when the model's time step size was varied between 5 min and 40 min (Figure 2). While solution sensitivity to time step size is not at all surprising from a mathematical perspective, such large discrepancies are undesirable numerical artifacts for model users who assume the models reflect the state-of-the-art understanding of the workings of the real-world system. In practice, it might be possible to "tuned away" the time step sensitivity by using different parameter values for different step sizes; however, there exists the danger that such tuning might result in error compensation that cannot be guaranteed for other applications (e.g., simulations under different forcing scenarios). To improve the credibility of future climate projections, it would be useful to revise the model and reduce the sensitivity to time step so as to provide the confidence that results from the numerical models are reasonably accurate solutions of the underlying continuous physics equations.

Strong sensitivities to model time step have been seen in other models as well. For instance, *Wan et al.* [2014] showed that clouds and precipitation simulated by the Community Atmosphere Model (CAM) version 5 changes substantially when the model time step is reduced from 30 min (the default value) to 4 min. *Zhang et al.* [2012] found that the impact of swapping aerosol nucleation parameterizations on sulfuric acid gas and aerosol concentrations was overwhelmed by the effect of changing the time stepping scheme used for solving the sulfuric acid gas equation in the aerosol-climate model (ECHAM-HAM). For ECMWF's weather forecast model IFS, *Beljaars et al.* [2004] showed that revising the numerical coupling between the dynamical core and turbulent momentum diffusion can substantially improve the 24 hour forecast of 10 m wind speed when using a 60 min time step (which was the operational value at the time). *Williamson* [2002] mentioned that when the splitting method within the parameterization suite was modified, CCM3 produced a climate equilibrium that was substantially different from the default model in some small contiguous areas. In other areas, the climates were similar, but the balances producing them were very different. Most of the studies cited above and the additional examples mentioned below indicate that *it is often the combination of coupling between processes and long time steps which cause time stepping problems in contemporary models*. The remainder of this section is focused on coupling issues, though it is acknowledged that long time steps can cause problems within individual processes as well.
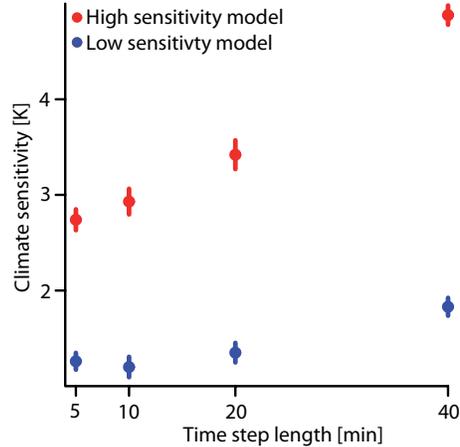
Figure 2. Equilibrium climate sensitivity in ECHAM5 [*Roeckner et al.*, 2003, 2006] slab-ocean simulations at T31L19 resolution. Red and blue markers indicate high- and low-sensitivity models which differ only in a few uncertain parameters in the physics parameterizations [*Klocke et al.*, 2011]. For each time step size listed on the x-axis, the model's climate sensitivity is computed as the difference between 10 year present-day simulations and the last 10 yrs of a 50 yr doubled CO_2 experiment. Error bars indicate inter-annual variability of global- and annual-mean surface temperature.

3.2 Splitting in the solution procedure

In order to facilitate the development, maintenance and practicality of numerical algorithms and model source code, parameterizations in weather and climate models are typically organized as separate modules for different processes. Here processes refer to individual physical phenomena such as cloud droplet formation or turbulent transport of chemical tracers. These may or may not be implemented in individual parameterizations. The process coupling discussed in this section includes the connection between different parameterizations, the connection between a parameterization and the host GCM, or the connection between different physical phenomena within an individual parameterization. Splitting is employed to evaluate the tendency terms for each process and to combine their effects to advance the discrete solution in time. The two most popular methods of splitting in operational models are parallel splitting (computing all process tendencies from the same model state, then using the sum of tendencies to march forward) and sequential splitting (computing a tendency, then either passing it together with the original model state to the next process or updating the model state and passing the new state to the next process). *Beljaars et al.* [2004] advocate sequential splitting with processes ordered from slowest to fastest in order to allow processes to feed and balance each other within each model step. It is worth noting that the benefits of sequential splitting depend on what information from already-calculated processes is used in subsequent process calculations. IFS uses both state information *and* tendencies from previous processes in some subsequent process calculations (hereafter referred to as sequential-tendency splitting), meaning that the processes see the tendencies of some of the previous process, but the model state is updated at the end of the time step. CAM physics simply updates the model state whenever a new tendency is available (hereafter sequential-update splitting). Since sequential-tendency splitting shares more information than sequential-update splitting or parallel splitting, it is unsurprising that it performs better. More sophisticated coupling has also been shown to be beneficial for very specific processes. For example, in the Semi-Lagrangian Averaging of Physical Parameterizations (SLAVEPP) algorithm of *Wedi* [1999], the tendencies are evaluated at both the departure and arrival points of the semi Lagrangian trajectory and then averaged. A predictor-corrector scheme is used to connect convective and stratiform clouds with the rest of the model. Below

only examples of coupling problems related to long time steps involving sequentially-split models are considered, because this is the prevailing configuration of global models.

3.3 Issues with splitting

One way that splitting causes errors is by skewing the competition for resources (e.g., cloud water, energy, or Convective Available Potential Energy (CAPE)) between processes. For example, convective instability can be removed by shallow convection, deep convection, or resolved-scale heating-induced motions. *Williamson* [2013] provides an example of competition for resources in a sequential-update split model. There it is noted that resolved-scale heating in CAM4 is applied as a hard adjustment which removes all supersaturation in a single time step, while CAM4 deep convection has a fixed timescale (30 min) for CAPE removal. As the time step decreases the fixed time-scale process does less to remove CAPE, while the hard adjustment does more. Since parameterized and resolved-scale deep convection have very different effects in CAM4, changing the model time step alters the ability of these processes to compete for convective instability and manifests as strong time step sensitivity. *Williamson* [2013] presents a simple model problem to illustrate the ramifications of this time step/time-scale interaction. The example provided results in extreme model behavior due to the interaction between the dynamics and the parameterizations. While this might be described as a time step sensitivity, it is actually a sensitivity to the ratio of parameterization time scales which changes with time step. Less drastic sensitivities have been seen by other investigators which appear to be related to the time-scale ratio issue. *Mishra and Sahany* [2013] found sensitivity to time step in the average tropical rainfall amount in CAM3 multi-year simulations, noting it was associated with the change in partitioning between convective and large-scale precipitation. *Reed et al.* [2012] showed a strong sensitivity in the strength of idealized tropical cyclones in high resolution CAM5 to time step, relating it to the accompanying change to the partitioning between convective and large-scale precipitation. In both studies the time scale of the convection was not changed and thus the ratio of time scales changed. This issue of partitioning is a typical symptom observed in models that use spatial resolutions in the gray zone of cumulus convection. More discussions on the gray zone can be found in section 9. It is worth noting that although the examples cited above are all from models that use sequential splitting, competition for resources is also a large problem for parallel splitting because it can result in unrealistically strong removal of resources. The most egregious cases of this (e.g., negative values) are typically resolved by simply rescaling tendencies to prevent over-consumption. This approach may leave more subtle cases untreated and, where applied, results in solving a different set of equations than originally intended. Another example of the partitioning problem was shown in the work of *Wan et al.* [2013], in which case the sulfuric acid condensation and aerosol nucleation acted as two sink processes in the sulfuric acid gas budget in the ECHAM-HAM model. The authors of that paper argued that more accurate simulations of the process rates, and consequently, more accurate near-surface concentrations of aerosol particles and cloud condensation nuclei, can be obtained when a solver handles the competing processes simultaneously without splitting.

A second scenario causing coupling problems is when one process is a source for something the other process consumes. If these processes are coupled by sequential-update splitting, the first process might push the quantity of interest to unreasonably high levels while the second process might pull it to unreasonably low levels. With parallel splitting the consuming process does not see a state immediately influenced by the source process until the following time step by which time the excess may have been modified by some other process. An example of such a push/pull problem with sequential-update splitting in CAM5 was presented in *Gettelman et al.* [2015], who note that macrophysics (condensation/evaporation + cloud fraction) is the main source of cloud water which is subsequently depleted by microphysical processes. By sub-stepping macro- and microphysics together they were able to obtain more realistic model behavior. *Wan et al.* [2013] describes another push/pull problem related to sulfuric acid gas budget in ECHAM-HAM. The study compared multiple time

stepping schemes for the coupling of sulfuric acid gas production (source) and condensation (sink). Results show that when the discrete time step is long compared to the characteristic condensation time scale, sequential splitting between production and condensation leads to a substantial overestimate of the condensation rate even when the individual processes are represented with accurate solutions of the split equations. It is argued that when practical to do so, the strongly interacting sources and sinks should be solved simultaneously. A third example is presented in *Beljaars et al.* [2004] for IFS. The near-surface wind speed is mainly affected by the pressure gradient force, Coriolis force, and the turbulent friction. Sensitivity tests showed that if the turbulent diffusion coefficients are computed after the model state variables have been updated by the dynamics-induced tendencies, positive biases in the intermediate wind speeds will lead to overestimation of turbulent friction thus negative bias in the 24 hour wind forecast.

Process coupling issues can lead to large time stepping errors and strong dependence on process ordering when splitting allows processes to operate in isolation for too long. This is not uncommon in operational models where the time step is often chosen to minimize computational cost without explicitly considering accuracy. An example is given by *Gettelman et al.* [2015], who note that sequential-update splitting with forward-Euler time stepping in CAM5 microphysics creates negative cloud water when computed tendencies are multiplied by inappropriately long time steps. Another example was provided in *Williamson and Olson* [2003], who found that aqua-planet simulations conducted with the NCAR CCM3 model had a single narrow peak of zonal mean precipitation at the equator when the Eulerian dynamical core was used, while simulations using the semi-Lagrangian dynamical core had a double-Inter-Tropical Convergence Zone (ITCZ) (i.e. a precipitation minimum at the equator, and two maxima straddling the equator). This sensitivity was attributed to the different time step sizes used for the physics parameterizations in the two model configurations (20 min for Eulerian, 60 min for semi-Lagrangian) rather than the dynamical cores themselves. The explanation the authors provided was that with sequential splitting, longer time steps lead to the accumulation of more CAPE, allowing convection to initiate further from the equator. The resulting condensational heating and secondary circulation further reinforce convection away from the equator. Similar changes to ITCZ shape in aqua-planet simulations with the CAM3 model have also been reported by *Li et al.* [2011].

3.4 Addressing the splitting problem

Tighter coupling between processes is necessary to alleviate the time stepping problems noted in section 3.1. A crude way to do this is by simply using shorter time steps, perhaps by sub-stepping clusters of tightly-coupled processes [*Gettelman et al.*, 2015]. Sequential-tendency splitting can also be used to allow faster processes to better react to the effects of slower processes. Passing specific information from one process to another can also be useful. For example, entrainment at the top of the cloudy boundary layer in the turbulence schemes by *Lock et al.* [2000] and *Bretherton and Park* [2009] is strongly affected by thermal instability diagnosed directly from radiative heating profiles. A benefit was also demonstrated in the aforementioned work on including dynamics information in the computation of turbulent surface drag [*Beljaars et al.*, 2004]. More recently, several parameterizations have been developed which handle multiple atmospheric processes in a unified way. Two such schemes which combine turbulence and shallow convection calculations are the Eddy Diffusivity Mass Flux (EDMF) approach of *Siebesma et al.* [2007] and the Cloud Layers Unified by Binormals (CLUBB) approach of *Golaz et al.* [2002a]. A scheme that unifies shallow and deep convection has been developed by *Park* [2014].

4 Time step convergence

Due to constraints on computational resources, global weather and climate models typically use time step sizes in the range from a few minutes to an hour. Ideally, the time

stepping and process-splitting methods should have sufficient accuracy at these step sizes to make the corresponding numerical errors small compared to the uncertainties associated with physically-based simplifications in the model equation system. One possible way to determine whether the time stepping accuracy is adequate is to check if the change in numerical solution caused by varying time step size stays below a practical tolerance defined through physical reasoning. Such an exercise can be interpreted as an assessment of time step convergence. In this section, convergence in the time dimension is discussed under the assumption of unchanged spatial resolution and model formulation. In this case, the asymptote of the discrete solutions – if they converge at all – is unlikely the best possible approximation of the real world, because inaccuracies associated with the analytic simplifications in the parameterizations and errors resulting from the spatial discretization are not alleviated by simply reducing the model time step. In other words, convergence in time alone describes the behavior of the numerical solutions of an analytically simplified, semi-discrete equation set, thus addressing only one aspect of the modeling problem.

In numerical analysis, convergence refers to the property of a numerical method that the discrete solution approaches the exact solution as the step size approaches zero. A scheme can be further characterized by its order of accuracy which describes an analytic relationship between the time step size and the local truncation error. While convergence tests (in this mathematical sense) are a standard part of dynamical core development, they have rarely been performed with model configurations combining both fluid dynamics and parameterized physics. Performing convergence tests in full-complexity models is not straightforward, for two reasons:

1. In the absence of analytic solutions, a "proxy ground truth" is needed in a convergence analysis. Conventional convergence studies in computational fluid dynamics involve the grids for all (spatial and temporal) coordinates going to zero; applying the same test strategy to weather and climate models can cause great difficulties in the interpretation of the results, because the parameterization schemes are likely to have undesirable sensitivities to spatial resolution (which is an issue somewhat separate from time stepping error) especially when the gray zone (section 9) is approached. Recent studies of *Teixeira et al.* [2007] and *Wan et al.* [2015] obtained reference solutions by running their models (Navy Operational Global Atmospheric Prediction System (NOGAPS) and CAM5, respectively) with very small time step sizes. *Wan et al.* [2015] argued that "convergence toward this proxy is a necessary but insufficient condition for the convergence toward the true solution". A caveat is that the model variables might converge to an unintended and/or unphysical state when time step alone goes to zero. Additionally, physical parameterizations are often designed to work within a particular range of time steps and using them outside of that range may violate physical assumptions. For example, most climate models assume that supersaturation with respect to liquid water is removed instantaneously, which is not true on timescales less than about 10 sec (cf *Squires* [1952]). Thus it is probably better to interpret the difference with respect to the reference solution as a metric of time step sensitivity rather than as error relative to the true solution of the chosen equation system.

2. Another difficulty in time step convergence analysis is that the expected behavior of operational models is not yet well established. Note that the concept of "order of accuracy" was originally developed for deterministic differential equations under the assumption that the solution is time-differentiable at least up to a certain order, but in full-fledged weather and climate models the condition is not always fulfilled. *Hodyss et al.* [2013] used numerical simulations of the diffusion-advection equation to demonstrate that when the time stepping scheme does not resolve the parameterized physical processes, the numerical solutions will behave as predicted by stochastic theory, resulting in a substantially reduced convergence rate. The implication of their results is profound: since weather and climate models include impactful processes (e.g., microphysics, turbulence) with time scales of seconds or smaller while models are

typically run with time steps several minutes in length, the originally expected first-(or higher-) order convergence might not be realizable. In deterministic numerical analysis, convergence is defined in the limit of small time steps, when model response to time step change can be approximated by a Taylor series truncated at the order of convergence. The large time steps used operationally in weather and climate models may fall outside the domain of validity of the truncated Taylor approximation, so the practical impact of reducing the model time step may be very different than predicted by deterministic numerical analysis. The slow convergence of the CAM5 simulations described by *Wan et al.* [2015] is likely an indication that the default model time step is far too long to resolve all the intrinsic time scales associated with the equation system. How to assess and improve solution accuracy in such a situation is a topic that requires further investigation

In addition, the traditional truncation error analysis often quantifies the numerical error of a discretization scheme in a single time step, i.e. the local truncation error, while in practice the global error accumulated in all the steps leading to a fixed simulation time is perhaps a more relevant metric. *Teixeira et al.* [2007] conducted a number of simulations with different time step sizes using NOGAPS, a quasigeostrophic (QG) model, and the Lorenz equations. They found that NOGAPS converged at first order near the start of their simulations, but the chaotic nature of nonlinear dynamical systems eventually caused simulations with different time steps to diverge into uncorrelated sequences of weather events. When one attempts to examine convergence rates beyond the first few steps of a simulation, uncertainties associated with the nonlinear nature of the equation system ("internal variability") need to be taken into account.

In operational weather and climate models, the magnitude of error obtained at a given resolution or given cost is the most important and practical measure of the quality of the time stepping method. Nevertheless, despite the abovementioned complication, a convergence analysis might still provide useful information about the numerical properties of the discrete model system, especially when the results deviate from the expected behavior. For example, *Wan et al.* [2015] found that in CAM5 the parameterizations that converge slower also have stronger time step sensitivity. In their case, the convergence rate provides a clear hint on which components of the model have inadequate numerical treatment thus require more attention in future development. Using the single-column version of a model to isolate the impact of physics parameterizations from fluid dynamics problems may help identify whether time stepping problems are related to physics, dynamics, or the interaction between the two, however the single-column approach requires a large number of cases to capture the variety of physical processes approximated in the models. Sub-stepping an individual process or a cluster of processes is a widely used strategy for improving stability and accuracy for faster components in a system. For example, the spectral-dynamical core in CAM5 uses multiple levels of sub-stepping for the adiabatic fluid dynamics, resolved-scale tracer transport, and numerical diffusion [cf. Table 1 in *Wan et al.*, 2015]. *Gettelman et al.* [2015] noted that using smaller time steps for the stratiform cloud parameterization had a positive impact on the model behavior. From the perspective of convergence analysis, sub-stepping clusters of problematic processes while keeping all aspects of other processes untouched can be useful for finding problems with certain schemes or process coupling. Replacing the time-integration method for certain schemes may play a similar role to sub-stepping. However, it should be kept in mind that while sub-stepping can improve the stability and accuracy of individual processes, it cannot address the splitting problem discussed in section 3 unless the strongly interacting processes are sub-stepped as a cluster.

As mentioned above, time step convergence tests of full-complexity models is a rarely-conducted exercise in the weather and climate modeling community. It will be interesting to see the outcome of the ongoing efforts in this direction.

With these real world issues and examples in mind the paper now proceeds into a more theoretical area, a mathematical analysis approach to the coupling.

5 Insights from models with simplified equation sets

Two examples of PDC are discussed below where the resolved scale behavior is strongly dependent on the subgridscale dynamics. This analysis highlights situations where the combination of resolved and subgrid terms is critical, e.g. in representing the total transport as the sum of resolved and subgrid transport. As the averaging scales are reduced, the subgrid contribution will reduce and be taken over by the resolved contribution.

5.1 Interaction of convection with balanced dynamics

In this case the spatial averaging scale is relatively large, and so the semi-geostrophic model, which is an accurate approximation to the governing equations on large scales, *Cullen* [2006], can be used as a proxy for the evolution of the spatially averaged equations. The behavior of this model can then be compared with solutions of the true governing equations with a much finer averaging scale which resolves convection explicitly. This then has implications for the design of models with parameterized convection.

The semi-geostrophic model includes the effect of large static stability variations, which are essential in considering interactions with convection. For illustration the in-compressible Boussinesq form of the equations in Cartesian geometry is used. Following *Cullen and Salmond* [2003], the equation for the ageostrophic wind is written as

$$\mathbf{Q}\mathbf{u}_{ag} + \frac{\partial}{\partial t}\nabla p = \mathbf{H}, \tag{1}$$

where

$$\mathbf{Q} = \begin{pmatrix} fv_{gx} + f^2 & fv_{gy} & fv_{gz} \\ -fu_{gx} & f^2 - fu_{gy} & -fu_{gz} \\ g\theta_x/\theta_0 & g\theta_y/\theta_0 & g\theta_z/\theta_0 \end{pmatrix}, \tag{2}$$

$$\mathbf{H} = \begin{pmatrix} -f\mathbf{u}_g \cdot \nabla v_g + F_1 \\ f\mathbf{u}_g \cdot \nabla u_g + F_2 \\ -g\mathbf{u}_g \cdot \nabla\theta/\theta_0 + S \end{pmatrix}.$$

Here $\mathbf{u} = (u, v, w)$ is the velocity, with suffix g indicating geostrophic and suffix ag indicating ageostrophic values. Suffices x, y and z indicate spatial derivatives. f is the Coriolis parameter, g is the acceleration due to gravity and θ is the potential temperature with reference value θ_0. F_1, F_2 and S are momentum and thermodynamic forcing terms respectively.

Under semi-geostrophic dynamics, the ageostrophic flow is determined diagnostically, and thus represents a response to the dynamical and physical forcing represented in equation (1). The strength of the response is determined by the eigenvalues of $\mathbf{Q}$, which represent the inertial and static stability of the atmospheric state. The geostrophic state would be expected to be described by the resolved flow in numerical models. However, the ageostrophic circulation required to maintain geostrophic balance would include subgrid-scale transports as well as resolved ageostrophic transport.

In the presence of moisture, the static stability is reduced by latent heating. This could be expressed, neglecting precipitation, by replacing θ by the equivalent potential temperature θ_e in saturated regions. In the presence of moist instability, $\mathbf{Q}$ would then have a negative eigenvalue. As illustrated in *Holt* [1990], this will result in convective transport rather than smooth vertical motion. The effect is that convective mass transport would replace the ascending branch of the ageostrophic circulation, while the downward ageostrophic circulation would be a smooth transport.
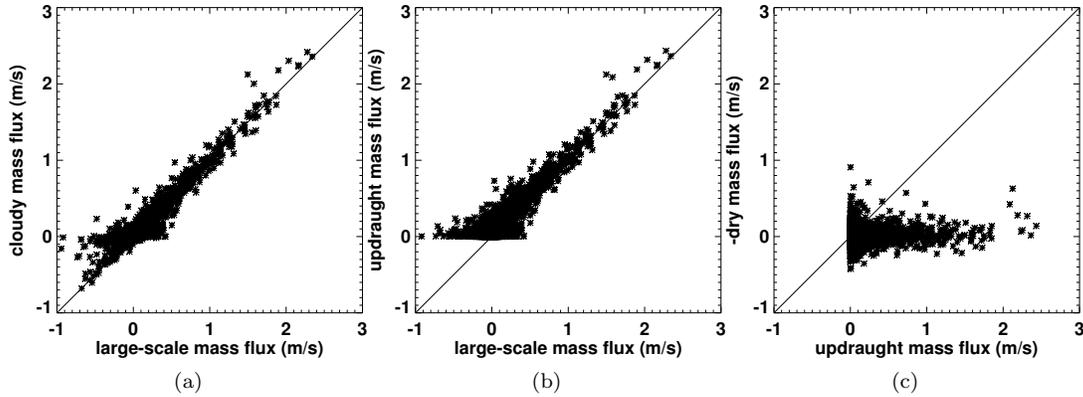
Figure 3. Scatter plots of cloudy mass flux against large-scale mass flux (a), cloudy updraft mass flux against large-scale mass flux (b), and minus dry mass flux against cloudy updraft mass flux (c). All in units of m s^{-1}. The data is taken from a height of 3195 m and is averaged in the horizontal to scale of 24 km. Met Office Unified Model.

This prediction is illustrated using a convection-permitting simulation performed as part of the Earth system model bias reduction and assessing abrupt climate change (EMBRACE) `http://cordis.europa.eu/project/rcn/99891_en.html` project. The simulation uses a configuration similar to that used operationally at the Met Office for United Kingdom (UK)-area short-range weather prediction [see *Holloway et al.*, 2012, for details] but with changes made to improve the representation of tropical convection and gravity waves. It has a horizontal resolution of 2.2 km with a large 8800 km by 5700 km domain centered on the tropical Indian ocean and 118 vertical levels with a 78 km lid. Within its domain the convection-permitting simulation was run freely after being initialized from the operational Met Office global model analysis valid at 0000 (UTC) on 18 August 2011. The lateral boundary conditions were provided every time step by a global model that was reinitialized from Met Office operational analyses every 6 hours. The data presented here was taken from 0000 UTC on 30 August 2011 and hence the convection permitting simulation was fully spun up.

The high resolution gridpoints are classified as cloudy or dry depending on the presence or not of cloud condensate: the cloudy areas are further subdivided into ascending and descending. The high resolution gridpoints are then aggregated onto a 24 km grid (a typical resolution at which convective parameterization is used) so that for each 24 km gridpoint a cloudy and dry mass flux is obtained, cloudy updrafts and downdrafts, and also the total large-scale mass flux.

Figure 3 shows that, for 24 km gridpoints that have some cloud, there is a close match between the total large-scale mass flux and the cloudy mass flux and hence most of the vertical motion happens within the cloudy areas (cf. section 9.2). The values of the dry mass flux are unrelated to the cloudy updraft mass flux. This means that there is no local compensating subsidence within the 24 km gridbox to match the cloudy updraft mass flux as is usually assumed in convective parameterization. The subsidence is instead spread over the whole domain. This is in agreement with the idea that the large-scale ascent is represented by convective plumes, while the subsidence is spread over a much wider region. This suggests that a radical rethink of (convective) parameterization strategy is required.

5.2 Interaction of the boundary layer with balanced dynamics

In this case the characterization of a simple model as the asymptotic limit of the full equations is exploited, and used to compare the effectiveness of different coupling strategies. A large scale balance is defined, which should be represented in the resolved numerical solutions, while the circulation required to maintain it will be described by both resolved and subgrid-scale transports. The inclusion of the boundary layer makes a fundamental change to the large scale balance because of the need to satisfy the no-slip boundary condition. Thus the balance is defined by the Ekman relations

$$\frac{\partial p}{\partial x} - f v_e = F_1(\mathbf{u}_e),$$
$$\frac{\partial p}{\partial y} + f u_e = F_2(\mathbf{u}_e).$$

(3)

(u_e, v_e) are the components of the Ekman velocity, and F_1 and F_2 represent the parameterized friction terms, which will depend on the horizontal momentum as indicated, as well as the thermodynamic structure. These equations can be solved for $\mathbf{u}_e$ given that $\mathbf{u}_e = \mathbf{u}_g$ at the top of the boundary layer and is zero at the ground.

Beare and Cullen [2013] derive equations analogous to equation 1 for the circulation required to maintain Ekman balance in time in the presence of dynamical and physical forcing. The ageostrophic circulation in semi-geostrophic theory is a second order accurate approximation in Rossby number to the velocity in the Euler equations. However, the equivalent circulation in the boundary layer is only first order accurate, as is the Ekman balance itself.

Now the effectiveness of schemes to couple the boundary layer with the balanced dynamics is demonstrated by following the method of *Cullen* [2007]. This experiment is described in detail by *Beare and Cullen* [2016]. A vertical slice model is used to construct a sequence of solutions of the boundary layer driven by a baroclinic wave where the Rossby number U/fL, with U and L denoting horizontal velocity and length scales, respectively, is progressively reduced. This is achieved by maintaining the same initial structure in the pressure and potential temperature while simultaneously increasing the Coriolis parameter and reducing the wind speed. The difference between the predicted circulation and the solution of the hydrostatic equations is then calculated. The expected result is second order convergence outside the boundary layer and first order inside. However, the boundary layer is found to become shallower as the Rossby number (Ro) is reduced, giving an overall convergence rate of $Ro^{1.7}$.

The results are compared using three numerical implementations. The control simulation uses a standard implicit time stepping, but the mixing coefficients F_1 and F_2 are evaluated only at the beginning of the time step. The *Wood et al.* [2007] scheme is a stable single step scheme which is unconditionally stable and second-order accurate. This is achieved by assuming a polynomial dependence of F_1, F_2 on wind speed. The K-update scheme includes the updated value of the boundary layer mixing coefficient at the new time level in each time step as described by *Cullen and Salmond* [2003] as well as the more accurate representation of the diffusion process in *Wood et al.* [2007]. This allows it to represent the balanced solution more accurately.

Figure 4 shows the difference between primitive equation simulations using different boundary-layer time stepping schemes and the balanced model. At smaller Rossby numbers, all primitive equation models follow the ideal $Ro^{1.7}$ line. However, above Ro = 0.08, the primitive equation model using the implicit scheme starts to deviate significantly above the ideal line, and no longer converges at the required rate. The primitive equation model using the K-update scheme deviates slightly above the ideal line at Ro = 0.1. The hydrostatic primitive equation (HPE) model using the *Wood et al.* [2007] scheme follows the ideal $Ro^{1.7}$ line for the range of Ro shown. Both the K-update and *Wood et al.* [2007] schemes account
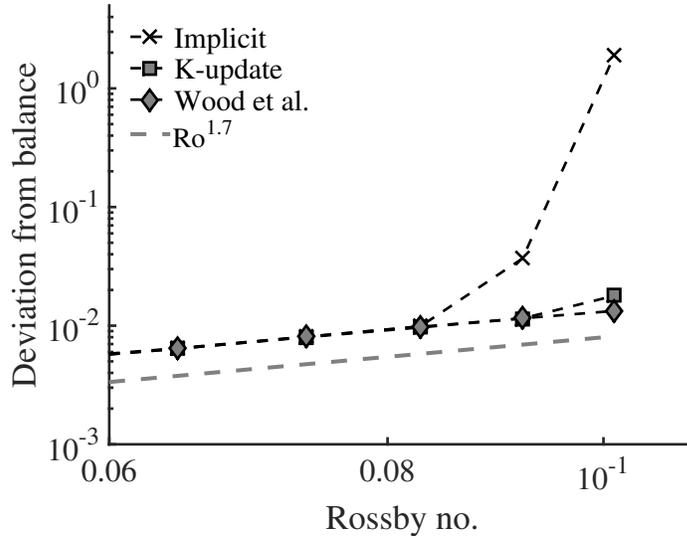
Figure 4. Convergence to balanced solution of the vertical slice primitive equation simulations [*Beare and Cullen*, 2016] for different time stepping schemes: Implicit, K-update and *Wood et al.* [2007]. $Ro^{1.7}$ is shown in gray for reference. .

for the variation of the boundary-layer diffusion across the time step, giving the improved convergence properties compared to the Implicit scheme. The deviation from the Ekman-balanced models thus exposes differences in the numerical methods employed.

5.3 Summary

Two approaches of validating methods of dynamics-physics coupling were demonstrated, given that the required averaged solution of the full equations cannot be described exactly as the solution of a set of partial differential equations. Section 5.1 illustrated that the validation of subgrid models should be against the averaged data from much higher resolution models. How the suggested changes translate into the full model also needs to be explored. Section 5.2 showed how methods of coupling of subgrid models can be validated by the accurate reproduction of asymptotic limits where subgrid transports are a key part of the limit solution. A future aim is to create protocols and setups for more complex models.

6 Analyzing the coupling of dynamical cores with a hierarchy of GCM test cases

One of the recurring questions is: Which PDC strategy is better? The answer depends crucially on the objective of the model run. Is it a climate run or a weather forecast? Is the model already severely time step restricted, such as Eulerian formulations, or are long time steps permitted, as in semi-implicit semi-Lagrangian models? The former may be less susceptible to coupling errors, assuming the physics and dynamics time steps are not too disparate, due to the higher temporal resolution and less scope for non-linear evolution (or splitting error). But even when these questions have been answered, in the full model context it is far from trivial to say which is better since the large number of factors involved quickly blur the answers. Therefore, testing is essential, and it is proposed that a hierarchy of idealized GCM test cases gives easier access to an improved understanding of the coupling mechanisms.

6.1 Idealized testing of GCMs

Full model testing has been discussed above and, for example, *Wan et al.* [2015] proposed various analysis techniques to better understand the impact of the physics time step on the model behavior. In an idealized framework, however, the parameterizations and lower boundary conditions are more constrained, which exposes the impact of the physics-dynamics coupling strategy on the simulation in a clearer way.

6.1.1 The different nature and sources of error

If models did not require tuning [*Hourdin et al.*, 2017], the answers would perhaps be more obvious. However, if a novel coupling method is implemented in an already tuned model, the solution is likely to be worse for the new coupling method if the model is then not re-tuned, even if the new coupling strategy would lead to a superior solution in the absence of tuning. Model tuning inevitably tunes against errors that are independent of the parameters tweaked in the tuning process (i.e., compensating errors). In this case, multiple errors may exist, but the superposition of errors introduced to minimize other errors may result in "shadowing of errors" if only the final solution is taken into account during tuning processes. Remove one of these errors and the result will be worse, despite having eliminated an error. For example, removing (or reducing) errors in the coupling of a mature model may result in a degraded final solution for these reasons. This highlights a key challenge in PDC: Not one single experiment will yield all the answers. The different techniques presented here have to be taken as a cohort of interrogation. Each has to be interpreted under there individual limitations. Combined it should be possible to derive clearer guidelines and understanding of the complex interactions. Not one analysis method is valid or one limitation invalidates the other. In a slightly modified version of Abraham Kaplans "The Conduct of Inquiry": The models are undeniably beautiful, and a man may justly be proud to be seen in their company. But they **have** their hidden vices. The question is, after all, not only whether they are good to look at, but whether we are able to interpret their results in the context of their limitations.

6.1.2 The GCM test case hierarchy

Due to the interconnected sources of error illustrated above it seems reasonable to implement and standardize an idealized testing protocol. It should be idealized in such way that the complexity of physical parameterizations is present in the forcing, but not in the implementation and that the implementation is generalized, allowing for direct comparisons between models. For the dynamical core, tests with idealized forcing exist, such as the Held-Suarez test case [*Held and Suarez*, 1994]. The Held-Suarez forcing was formulated for a dry and flat planet and includes a thermal relaxation mechanism and low-level Rayleigh friction. These mimic the effects of radiation and boundary layer mixing, respectively. However, the adjustment processes in the Held-Suarez test case are rather slow and do not challenge the physics-dynamics coupling sufficiently. A missing key ingredient is moisture. The latent heat exchanges due to water phase transitions are desirable in order to challenge the coupling mechanisms.

The "simple-physics" package by *Reed and Jablonowski* [2012] makes progress in this aspect. It incorporates bulk aerodynamic surface fluxes and diffusive boundary layer mixing processes of heat, moisture and momentum, a large-scale condensation scheme without a cloud phase, and utilizes an ocean-covered surface with prescribed sea surface temperatures (SSTs) as a lower-boundary condition. The Fortran source code is publicly available (`https://earthsystemcog.org/projects/dcmip-2012/`, click on 'Fortran Routines' in left navigation bar and download the attached 'simple physics suite' at the bottom of the page), removing the uncertainty of the implementation, and the suite is simplistic enough to be easily reproduced within varying model frameworks. However, the simple-physics package lacks radiation and is therefore only suitable for short-term simulations. This was remedied by *Thatcher and Jablonowski* [2016] who combined the ideas of the *Reed and*

Jablonowski [2012] simple-physics package and the Held-Suarez forcing to create a moist version of the Held-Suarez test. The resulting Moist Idealized Test Case (MITC) with Newtonian thermal relaxation mimicking "radiation" is suitable for long-term simulations and has been shown to reveal the intricacies of the physics-dynamics coupling as further highlighted in section 6.2. MITC can be considered a moist idealized test of intermediate complexity. The MITC Fortran routine is available as a supplement to the journal article from `http://www.geosci-model-dev.net/9/1263/2016/`.

The next step in the test case hierarchy points to simplified physics formulations with a radiation scheme and unconstrained SSTs that are e.g. determined by a slab ocean model (also called "mixed-layer" model). *Frierson et al.* [2006] presented a gray-radiation GCM, which possesses desirable ingredients such as radiation, an interactive slab ocean, large-scale precipitation, and surface/boundary layer schemes. However, the physics suite is not sufficiently documented to be easily reproducible and comparable to other models. If more realistic ocean temperatures are desired, a slab ocean scheme can also be augmented with a set of specified surface flux adjustments (commonly called "q-flux adjustments"). These can be added to the slab model's temperature tendency equation at each time step in order to maintain a seasonal cycle of realistic ocean temperatures.

A final step in the idealized model hierarchy are long-term "aqua-planet" simulations on a flat and ocean-covered Earth that utilize the complex physical parameterization package of a GCM. The lower boundary condition can either be based on prescribed SSTs as in *Neale and Hoskins* [2000] or a slab ocean approach with predicted SSTs as in *Lee et al.* [2008]. Aqua-planet simulations are popular for idealized climate studies. Here, we demonstrate that they can also provide insight into the delicate interplay between the physical parameterizations and the numerical schemes of dynamical cores with their associated diffusion (section 6.3). Ideally, in between the two well observed and understood boundary conditions, SST and incoming shortwave radiation at the top of the atmosphere, as much as possible should be left to the model, ie. variables should be allowed to propagate freely and not be prescribed or constrained to reference profiles or background states. If a certain coupling scheme or model formulation has an impact on the Hadley circulation, for example, then the test case should be able to show a trend towards this.

6.2 Simplified physics assessments

Figure 5 displays an example of how the MITC approach by *Thatcher and Jablonowski* [2016] can provide information about the physics-dynamics coupling strategy. The figure shows instantaneous, randomly selected snapshots of the 850 hPa vertical pressure velocities and precipitation rates in MITC simulations with the CAM5 model [*Neale et al.*, 2010a]. The depicted CAM5 dynamical cores are the Finite-Volume (FV) model [*Lin*, 2004], the spectral transform Eulerian (EUL) dynamical core, and the Spectral Element (SE) model [*Taylor and Fournier*, 2010; *Dennis et al.*, 2012a]. These are run at the horizontal resolutions $1° \times 1°$ (FV, $\approx$111 km), the triangular truncation T85 with a quadratic Gaussian grid (EUL, $\approx$156 km), and in the "*ne*30*np*4" (SE) configuration which corresponds to a grid spacing of about 111 km. All dynamical cores use the same 30 vertical levels. Their positions are documented in the Appendix of *Reed and Jablonowski* [2012].

The three dynamical cores are coupled to the identical MITC physics package [*Thatcher and Jablonowski*, 2016] and run for multiple years. Within the MITC physics package, the coupling strategy of the various physical processes follows the sequential-update approach which is also detailed in *Thatcher and Jablonowski* [2016]. However, the physics-dynamics coupling strategies differ. The FV dynamical core (Figures 5a,e) with a dynamics time step of 180 s is coupled to the physics package in a time-split (sequential) way and applies the physical forcings every 1800 s (physics time step). The EUL dynamical core (Figures 5b,f) is coupled to the physics in a process-split (parallel) way. EUL applies the physical forcings every 600 s which is identical to EUL's dynamics time step. The SE dynamical core (Fig-
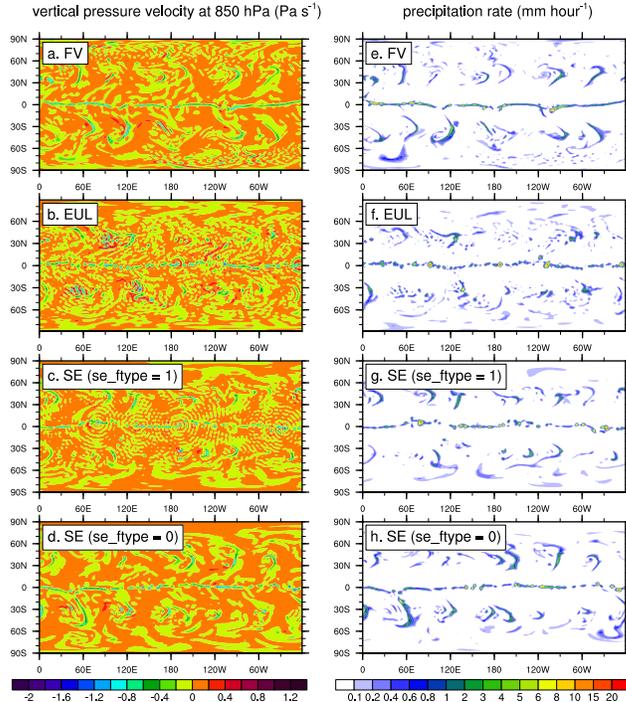
Figure 5. Snapshots of instantaneous (left) 850 hPa vertical pressure velocities and (right) precipitation rates in MITC simulations with the (a,e) CAM-FV, (b,f) CAM-EUL and (c,d,g,h) CAM-SE dynamical cores. se_ftype = 1 (c,g) denotes a physics-dynamics coupling with the long physics time step, se_ftype = 0 (d,h) couples with a sub-cycled, short dynamics time step. The physics time steps are 1800 s (FV, SE) and 600 s (EUL), the dynamics time steps are 180 s (FV), 600 s (EUL) and 300 s (SE). In the case of SE with se_ftype=0 the forcing was gradually applied every 300 s. The EUL dynamical core is coupled to the physics in a process-split (parallel) way, the SE and FV physics-dynamics coupling is time-split.

ures 5c,d,g,h) with a dynamics time step of 300 s is coupled to the physical parameterizations in a time-split way with a physics time step of 1800 s as FV. However, two coupling options exist in SE which either apply the physical forcings as a sudden adjustment after the long 1800 s physics time step (se_ftype = 1) or gradually within the sub-cycled dynamical core (se_ftype = 0) every 300 s.

Figures 5(c,d,g,h) document that the choice of the coupling strategy in CAM5-SE has significant impact on the simulation. The intense gridscale (or gridpoint) storms [*Williamson*, 2013], that develop along the equator in all models (seen in the precipitation rates in the right column), lead to a circular gravity wave ringing patterns in the 850 hPa vertical pressure velocity ω in CAM5-SE when coupled with the long 1800 s physics time step (se_ftype = 1, Figure 5c). The centers of the circular ω patterns coincide with the positions of the strongest precipitation rates in Figure 5g, which suggests that the intense latent heat release at these locations initiates the gravity wave noise. The gravity wave response to the impulsive physical forcing is large-scale, so that the explicitly-applied diffusion in CAM5-SE does not filter out its propagation. *Thatcher and Jablonowski* [2016] found that the gravity wave noise can be remedied when changing the coupling strategy in CAM-SE. In case of se_ftype = 0 (Figures 5d,h) the physical forcing tendencies are gradually applied within the CAM-SE dynamical core every 300 s. The strong grid-scale storms are still present in the precipitation field (Figure 5h). However, the more gradual forcing reduces the latent heat impulses and leads to a smooth vertical pressure velocity (Figure 5d). Similar sensitivities to the se_ftype setting were also found in full-complexity CAM-SE climate simulations (Peter Lauritzen (NCAR)

personal communication, 2015). Therefore, the CAM-SE se_ftype default was switched from 1 to 0. This shows that simpler modeling frameworks help expose the causes and effects of the physics-dynamics coupling choices.

It is also informative to compare these CAM-SE characteristics to the alternative FV and EUL dynamical cores. As SE (se_ftype = 1), the FV model (Figures 5a,e) also adjusts the state variables with the long 1800 s physics time step and experiences equatorial grid-point storms of similar magnitude (Figure 5e). However, the damping characteristics of the two dynamical cores differ [*Jablonowski and Williamson*, 2011] and FV can more effectively damp grid-scale noise due to its built-in, local monotonicity constraints. Therefore, FV distributes the large latent heating impulses more smoothly which leads to a smooth distribution of its vertical pressure velocity (Figure 5a). In contrast, the EUL model is built upon a global spectral numerical method which is known for its difficulty representing sharp contrasts locally. Here, the large latent heating impulses near the peak precipitation rates (Figure 5f), lead to the so-called Gibbs ringing effect (see also *Jablonowski and Williamson* [2011]). The Gibbs ringing is visible in EUL's vertical pressure velocity field (Figure 5b) and manifests itself as a noisy pattern (broken contours). The noise is even present in the midlatitudinal regions where organized precipitation bands should dominate. EUL's shorter 600 s physics time step (in comparison to the 1800 s used in FV and SE) is not able to prevent these numerical Gibbs oscillations.

6.3 Aqua-planet assessments

Another example of how full-physics aqua-planet simulations can give insight into the physics-dynamics interplay is shown in Figures 6 and 7. The figures provide information about the shape of the ITCZ in CAM5 aqua-planet simulations with prescribed SSTs (CONTROL case in *Neale and Hoskins* [2000]). As in section 6.2 the CAM5 dynamical cores EUL, FV and SE are assessed at the resolutions T85 (EUL) and 111 km (SE, FV) with 30 levels. In addition, the figures include the CAM5 spectral transform semi-Lagrangian (SLD) T85 dynamical core. All model simulations are run for 2.5 years, and the first six months are disregarded (spin-up period). The models use the dynamics time steps 300 s (SE), 180 s (FV), 600 s (EUL) and 1800 s (SLD), which are paired with the physics time steps 1800 s (SE, FV, SLD) and 600 s (EUL).

The shape of the ITCZ in aqua-planet simulations has been a topic of debate for over a decade. Some models show a single equatorial peak of the ITCZ precipitation rate whereas other models are characterized by a double ITCZ in the subtropics. *Blackburn et al.* [2013] even called the double ITCZ one of the "modern modeling mysteries". The suggested mechanisms that govern the shape and strength of the ITCZ vary widely and are ambiguous. *Williamson and Olson* [2003] found a dependence on the physics time step, time stepping scheme, the dynamical core and the strength of the horizontal diffusion. *Mishra et al.* [2008] discussed the ITCZ time step dependencies and physics changes, *Rajendran et al.* [2013] discussed the SSTs impact on the ITCZ, *Lee et al.* [2003] and *Möbis and Stevens* [2012] investigated the role of the convection scheme, *Williamson* [2008] reported on the sensitivities to horizontal resolution, and *Landu et al.* [2014] discussed the ITCZ sensitivity to two dynamical cores, their resolutions and strengths of the low-level moisture transports. More recently, *Medeiros et al.* [2015, 2016] compared the ITCZs in the aqua-planet configurations of models that participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) and provided an aqua-planet reference solution for NCAR's CAM5.3 (version 5.3) model.

Figure 6 depicts the time-mean (averaged over the last 2 years) zonal-mean precipitation rate in all four CAM5 dynamical cores in aqua-planet mode which are driven by the identical CAM5 physical parameterization package. The physics package is described in *Neale et al.* [2010a]. Here, we highlight that it contains the University of Washington (UW) moist turbulent planetary boundary layer (PBL) scheme by *Bretherton and Park* [2009] which is based on an assessment of the turbulent kinetic energy. This PBL scheme is tightly
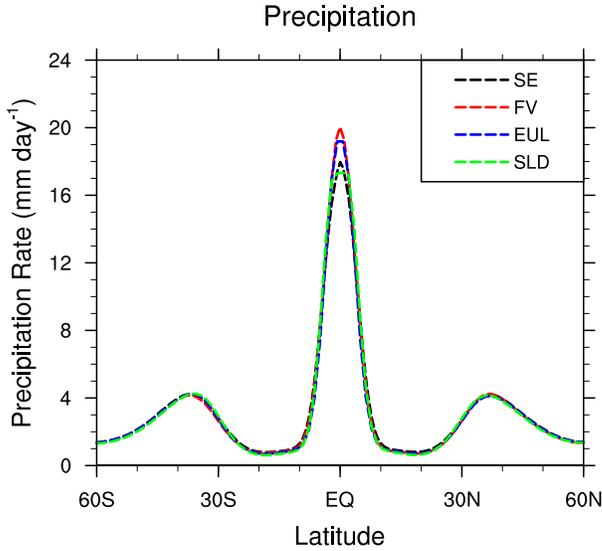
Figure 6. 2-year-mean zonal-mean precipitation rate in four aqua-planet simulations with the CAM5 dynamical cores SE (111 km), FV (111 km), EUL (T85), SLD (T85) and the default CAM5 physics package.

linked to the UW shallow convection parameterization, developed by *Park and Bretherton* [2009], through a cloud-base mass flux. In addition, the CAM5 deep convection scheme is based on the mass-flux parameterization by *Zhang and McFarlane* [1995], which has been enhanced by *Richter and Rasch* [2008] and *Neale et al.* [2008] to account for convective momentum transport and dilute plumes.

Figure 6 shows that the precipitation rates in all four aqua-planet simulations are remarkably similar. They all show a single ITCZ and equatorial peaks that range between 17.5 – 20 mm day^{-1}. This is in sharp contrast to the assessments by *Blackburn et al.* [2013] who intercompared 16 different model simulations that participated in the Aqua-Planet Experiment (APE) [*Blackburn and Hoskins*, 2013]. The peaks in the APE models ranged from 10-34 mm day^{-1} with an almost even split between single versus double ITCZ models. Since the APE models are characterized by vastly different dynamical cores, resolutions, physical parameterizations and coupling strategies this makes it very difficult to distinguish between causes and effects. When taking Figure 6 into account though, it seems feasible that most differences in APE models are likely triggered by different physical parameterizations.

Here a single aqua-planet framework is promoted as a "control environment" for idealized assessments of the physics-dynamics interplay. An example is given in Figure 7 which intercompares the CAM5 SE (111 km) and SLD T85 dynamical cores with 30 levels when coupled to the alternative physical parameterization scheme CLUBB [*Golaz et al.*, 2002a,b; *Bogenschutz et al.*, 2012, 2013]. CLUBB replaces CAM5's default PBL, macrophysics and shallow convection scheme. The Zhang-McFarlane deep convection scheme [*Zhang and McFarlane*, 1995] is still used. CAM5-SE/CLUBB is shown with two different settings of the fourth-order horizontal diffusion coefficient. Figures 7(a,d) depict the default diffusion coefficient 1×10^{15} m^{4}s^{-1} for the 111 km grid spacing (labeled *ne30np4*). The Figures 7(b,e) show the SE results with an increased diffusion coefficient of 5×10^{15} m^{4}s^{-1}. The SLD T85 dynamical core (Figures 7(c,f)) does not apply any explicitly-added diffusion since its numerical scheme already provides sufficient implicit numerical diffusion. The simulations shown in Figure 7 are 1.5 years long and the first six months are discarded (model spin-up period). The physics and dynamics time steps for SE and SLD are the same as quoted before.

The top row of Figure 7 shows the latitude-pressure cross section of a one-year-mean zonal-mean vertical pressure velocity in the tropics for SE (a,b) and SLD (c). The bottom
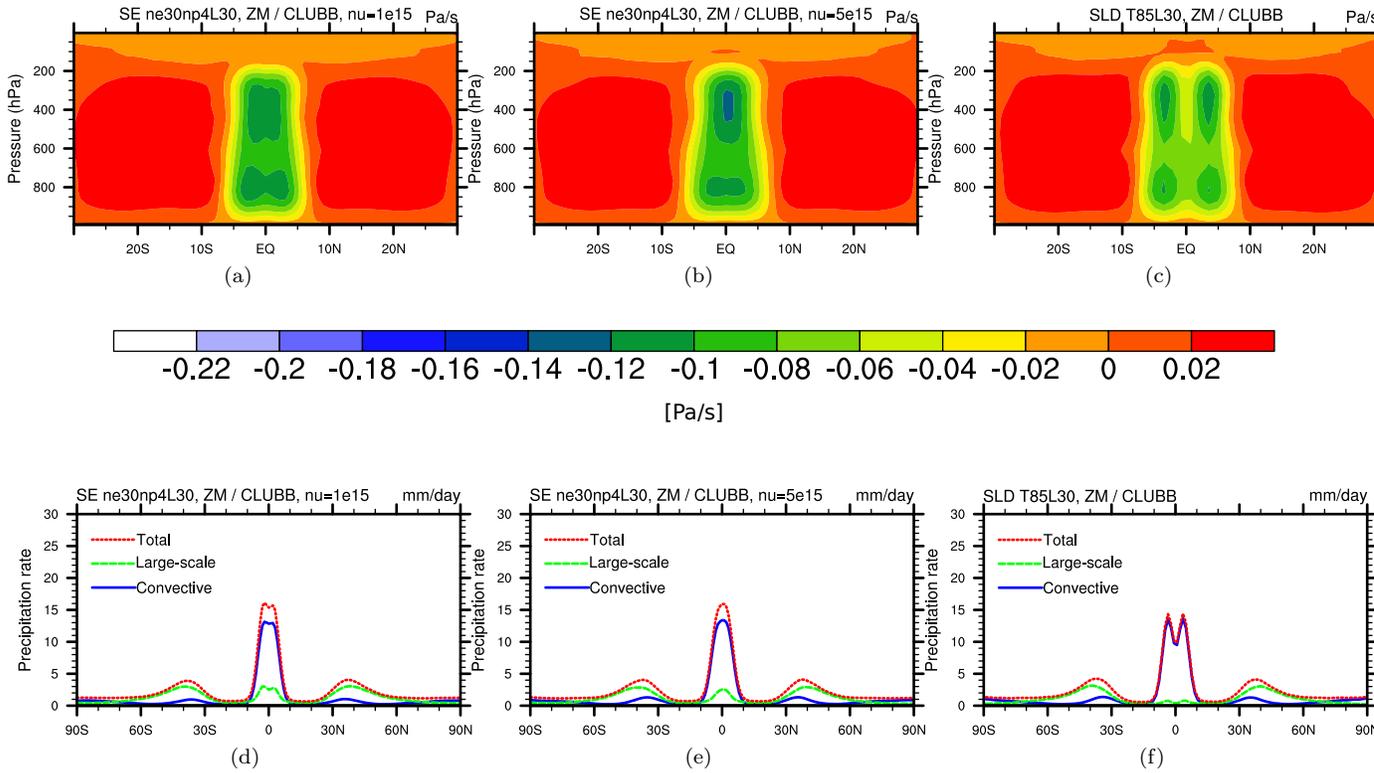
Figure 7. Aqua-planet simulations with the alternative CLUBB PBL, macrophysics and shallow convection scheme in CAM5. Top: Latitude-pressure cross section of the one-year-mean zonal-mean vertical pressure velocity in the tropics for the dynamical cores (a) SE with diffusion coefficient 1×10^{15} m^{4}s^{-1}, (b) SE with diffusion coefficient 5×10^{15} m^{4}s^{-1} and (c) SLD without explicit horizontal diffusion. Bottom: One-year-mean zonal-mean precipitation rates of the three runs, split into total (red), large-scale (green) and convective (blue) precipitation.

row presents the one-year-mean zonal-mean precipitation rates of the three runs, split into total, large-scale and convective precipitation. The total precipitation rate can be directly compared to Figure 6. Two observations are striking. First, the switch to the CLUBB scheme causes the SE and SLD dynamical cores with default diffusion settings (Figures 7(a,c,d,f)) to switch from the single ITCZ shown in Figure 6 to a double ITCZ structure. Second, the appearance of a weak double ITCZ structure in SE (Figures 7(a,d)) is highly dependent on the choice of the horizontal diffusion coefficient. The increased diffusion coefficient in Figures 7(b,e) impacts the moisture processes in a way that convert the weak double ITCZ in the default SE run to a single ITCZ peak. This brief assessment highlights the strength of an idealized testing framework in order to shed light on the physics-dynamics interactions. It is suggested that this approach can also be used to analyze the effects of different physics-dynamics coupling strategies.

6.4 Summary

The section demonstrated that simplified modeling frameworks and model hierarchies are useful tools to better understand the differences between formulations on a lower level of complexity. This will simplify experimentation and interpretation. It was shown how the tests are sensitive to the different physics-dynamics coupling strategies, physical parameterizations and diffusion settings and their interaction with the dynamical core. Here, simple prescribed SSTs (in aqua-planet mode) were used since the aim was to construct models that are as constrained as possible. Alternatively, slab ocean models with predicted SSTs and a closed surface energy budget represent the logical next step in the model hierarchy. In practice, models with significantly more complexity are utilized, with associated physical parameterizations, data assimilation, and other infrastructure. Coupling these components together (the same holds true for the land, ocean and ice models, chemistry, etc.) is non trivial, as the following section will describe.

7 Intra model coupling

In this section, the focus is on intra-models coupling problems within the climate modeling system, where the coupling occurs via an exchange of boundary conditions that transmit fluxes through a physical interface (e.g. the ocean/atmosphere, land/atmosphere, ice/atmosphere or ocean/sea-ice interface). A difficulty inherent to this type of application is that many distinct physical processes at different temporal and spatial scales, governed by different physical/conservation laws, must be simultaneously considered as a whole. This difficulty leads to intertwined physical, mathematical and computational delicacies. Algorithms to solve such coupled problems can be classified into two general categories

(i) *Monolithic method*: a single model representing all components to be coupled is defined. It requires each component to share the same space-time computational grid and computational framework. The advantage is that a tightly (strongly) coupled solution can be easily obtained. However, this approach is generally not tractable when trying to couple two individual models developed independently from each other with distinct numerical techniques, except for toy models [e.g. *Connors and Ganis*, 2011]. The monolithic approach has been used previously for land-atmosphere coupling when land surface processes were implemented as subroutines of **GCMs!** (**GCMs!**) but is currently abandoned to provide more modularity because of the increasing complexity of land surface models which are now treated as external modules [e.g. *Polcher et al.*, 1998; *Ryder et al.*, 2016].

(ii) *Partitioned/split method* : analogous to operator splitting, the full problem is split into smaller problems solved independently with boundary exchanges through their common interfaces [e.g. *Schulz et al.*, 2001; *Schmidt et al.*, 2004; *Large*, 2006]. This is the most frequently adopted and most natural option in coupled problems arising in Earth system modeling. However, the difficulty is that this type of approach can give

rise to various splitting errors and, thus, makes it difficult to recover a tightly coupled solution [*Keyes et al.*, 2013]. Analysis and attribution of these errors is not straightforward, as elaborated below. A comprehensive review about interface-coupled multiphysics systems in a broad sense can be found in *Keyes et al.* [2013].

Coupled problems arising in Earth system modeling cover a large range of aspects: parameterizations of turbulent boundary layers near interfaces, estimation of interfacial fluxes [*Schmidt et al.*, 2004; *Large*, 2006], space-time numerical schemes, matching of different grids at the interface [e.g. *Best et al.*, 2004; *Balaji et al.*, 2007], coupling algorithms [e.g. *Lemarié et al.*, 2015; *Ryder et al.*, 2016; *Beljaars et al.*, 2017], software implementation [e.g. *Valcke et al.*, 2012], etc, adding to the overall complexity of numerical models which are usually only considered on their own, neglecting connectivity.

In the present section the partitioned approach is considered and the example of the OA coupling is used to illustrate the delicacies in terms of physics/dynamics inconsistency inherent to intra-model coupling. Most of the issues presented here are not only relevant to OA coupling. Readers interested in more specific details on sea-ice/ocean or land surface/atmosphere coupling, for example, could refer to *Schmidt et al.* [2004] or *Ryder et al.* [2016] and the references therein.

7.1 Theoretical limitations of some of the current OA coupling methods

Most multiphysics coupling problems assume that all scales are resolved by the numerical models and that the boundary conditions at the interface are of Dirichlet or Neumann type (or a linear combination of both). In the case of the ocean-atmosphere problem the dynamical coupling is strongly influenced by physical parameterizations which makes the rigorous mathematical analysis not tractable. The numerical resolution of the OA coupling problem for practical applications is generally tackled in two different ways. Either by an exchange of instantaneous boundary data at the largest time step of the two models, this method is referred to as synchronous coupling, or by an exchange of averaged-in-time boundary data over a time interval $[t_i, t_{i+1}]$ (which is much larger than the largest time step). The latter method is referred to as asynchronous coupling. Those methods, described in Figure 8, are loose coupling schemes (in contrast to tight coupling schemes) in the sense that they correspond to only one iteration of an iterative process without reaching convergence [see *Lemarié et al.*, 2014, 2015]. Hence, they do not strictly provide the solution to the OA coupling problem, but an approximation of one since state variables of the two models are shifted by one time-step or a sequence of time-steps. The theoretical limitations of the synchronous and asynchronous methods are now explained further. In the synchronous coupling algorithm, the following errors are observed:

- *Aliasing errors*: significantly different time steps are used in each model (for the same horizontal resolution the oceanic model is integrated with a time step approximately ten times larger than the atmospheric model), as a consequence aliasing problems may arise and compromise stability [e.g. *Schluter and Pitsch*, 2005]
- *Synchronicity error* : air-sea fluxes are used as boundary conditions for the vertical turbulent diffusion terms which are treated implicitly in time, meaning that the fluxes at the interface are formally needed at time $t + \Delta t$ and not t (in Figure 8a). The explicit exchange of data in the synchronous coupling leads to an additional condition for the coupling to be stable even if unconditionally stable time stepping algorithms are used for vertical diffusion [*Lemarié et al.*, 2015; *Beljaars et al.*, 2017]. A way to circumvent those stability issues is to consider a synchronous coupling with an implicit exchange of date. In practice, this approach amounts to solve the local implicit problems in the ocean and the atmosphere in a monolithic way as one single implicit solver as often done for land-surface/atmosphere coupling [*Polcher et al.*, 1998; *Ryder et al.*, 2016]. Implicit flux coupling is so far seldom used in the context of OA coupled models.
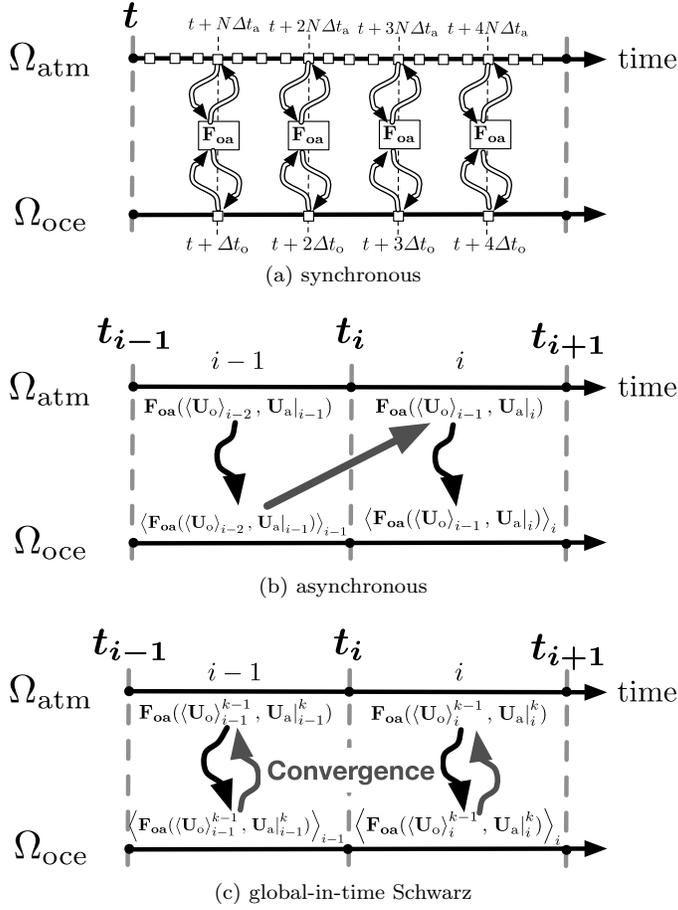
- *Physics-dynamics inconsistency error*: the uncertainties in the computation of air-sea fluxes at high-frequency through bulk formulations is extremely large (see discussion in section 2 in *Large* [2006], or *Foken* [2006]). The sources of those uncertainties are numerous, among them are the assumptions used to derive the continuous formulation of bulk formulae (constant-flux layer assumption, horizontal homogeneity, quasi stationarity, etc) and the fact that very few direct measurements exist to calibrate those semi-empirical formulations over the ocean. Moreover, the nonlinear problem associated to the estimation of bulk fluxes is often solved in an approximate way. In practice an averaging of the oceanic and atmospheric inputs to the bulk formulae should be required to minimize the uncertainty in air-sea fluxes [*Large*, 2006] meaning that an internally required time-scale $\Delta t_{\text{phys,req}}$ needs to be assumed for the parameterization scheme (a.k.a bulk formulation) to be valid, and $\Delta t_{\text{phys,req}}$ is generally larger than the model dynamical time step Δt_{dyn}. As a result, using a synchronous method can render the model solution very sensitive to the choice for the time step Δt_{dyn} since it is implicitly assumed that $\Delta t_{\text{phys}} = \min(\Delta t_{\text{dyn,req}}, \Delta t_{\text{dyn}})$, which can lead to large errors in the estimation of air-sea fluxes.

By construction the asynchronous coupling is expected to mitigate this latter issue since boundary data averaged-in-time are exchanged over a time interval $[t_i, t_{i+1}]$ generally much larger than the dynamical time-step. However, the asynchronous coupling algorithm also suffers from a synchronicity issue. Indeed, the oceanic state used on $[t_i, t_{i+1}]$ comes from the previous time window $[t_{i-1}, t_i]$ and not the current time window. Note that the lack of synchronicity is clearly visible in Figure 8b (oblique arrow). This error arises from the use of a non-iterative partitioned coupling approach. The asynchronous coupling does not permit an accurate representation of transient processes on short time scales, e.g., the diurnal SST cycle, which is undesirable especially when the space-time resolution is increased. This approach is however still used in numerous climate coupled models but an active research is currently in progress to minimize those synchronicity issues and allow correct phasing between the ocean and the atmosphere at a reasonable computational cost.

7.2 Reducing physics-dynamics inconsistency and splitting errors

In order to explore possible ways to reduce the above mentioned errors the theoretical framework of the Schwarz-like domain decomposition methods [e.g. *Lemarié et al.*, 2013] can be adopted. The idea behind those methods is to separate the original problem on the computational domain Ω, with $\Omega = \Omega_{\text{oce}} \cup \Omega_{\text{atm}}$ into subproblems on Ω_{oce} and Ω_{atm}, which can be solved separately. An iterative process is then applied to achieve convergence to the solution of the original problem. For the coupling of systems of partial differential equations (ignoring physical parameterizations) the converged solution obtained using the Schwarz algorithm is the same as the one obtained using a monolithic approach, within a given tolerance. It can be shown that the asynchronous coupling method corresponds to a single iteration of a global-in-time Schwarz Method, see Figure 8c. Respectively the synchronous coupling method corresponds to one single iteration of a local-in-time Schwarz method.

The usual methods (e.g. synchronous and asynchronous coupling) used in the context of ocean-atmosphere coupling are thus prone to splitting errors because they correspond to only one iteration of an iterative process without reaching convergence. There are so far very few studies aiming at quantifying the impact of these coupling errors on the coupled solutions. In *Connors and Ganis* [2011], using highly simplified models, it is shown that the use of a synchronous algorithm compared to a monolithic approach (or equivalently a Schwarz algorithm) leads to a larger model uncertainty in the sense that the resulting variance in model variables is larger. This result is based on an uncertainty quantification method using stochastic input parameters for the exchange coefficients involved in the air-sea flux computation. In *Lemarié et al.* [2014], numerical experiments using a mesoscale atmospheric model Weather Research and Forecasting (WRF) coupled with a regional oceanic model Regional Ocean Modeling System (ROMS) for a realistic simulation of a tropical cyclone have been

t

Ω_{atm}

$t + N\Delta t_a \quad t + 2N\Delta t_a \quad t + 3N\Delta t_a \quad t + 4N\Delta t_a$

time

$\mathbf{F_{oa}} \quad \mathbf{F_{oa}} \quad \mathbf{F_{oa}} \quad \mathbf{F_{oa}}$

Ω_{oce}

$t + \Delta t_o \quad t + 2\Delta t_o \quad t + 3\Delta t_o \quad t + 4\Delta t_o$

(a) synchronous

$t_{i-1} \qquad t_i \qquad t_{i+1}$

$i - 1 \qquad\qquad i$

Ω_{atm}

time

$\mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-2}, \mathbf{U_a}|_{i-1}) \quad \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-1}, \mathbf{U_a}|_i)$

Ω_{oce}

$\langle \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-2}, \mathbf{U_a}|_{i-1}) \rangle_{i-1} \quad \langle \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-1}, \mathbf{U_a}|_i) \rangle_i$

(b) asynchronous

$t_{i-1} \qquad t_i \qquad t_{i+1}$

$i - 1 \qquad\qquad i$

Ω_{atm}

time

$\mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-1}^{k-1}, \mathbf{U_a}|_{i-1}^{k}) \quad \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i}^{k-1}, \mathbf{U_a}|_{i}^{k})$

Convergence

Ω_{oce}

$\langle \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i-1}^{k-1}, \mathbf{U_a}|_{i-1}^{k}) \rangle_{i-1} \quad \langle \mathbf{F_{oa}}(\langle \mathbf{U_o} \rangle_{i}^{k-1}, \mathbf{U_a}|_{i}^{k}) \rangle_i$

(c) global-in-time Schwarz

Figure 8. Schematic view of the coupling between the computational domains of the atmosphere model, Ω_{atm}, and ocean model, Ω_{oce}, with time advancing to the right. The function $\mathbf{F_{oa}}(\mathbf{U_o}, \mathbf{U_a})$ represents the parameterization of air-sea fluxes with $\mathbf{U_o}$ (resp. $\mathbf{U_a}$) the oceanic (resp. atmospheric) state vector. $\langle \cdot \rangle$ is a given time averaging operator, and Δt_o, Δt_a the dynamical time step of the models such that $N = \Delta t_o / \Delta t_a$.

carried out. Ensemble simulations have been designed by perturbations of the coupling frequency and the initial conditions. One ensemble has been integrated using the global-in-time Schwarz Method and an other using the asynchronous method. The Schwarz iterative coupling methods leads to a significantly reduced spread in the ensemble results (in terms of cyclone trajectory and intensity), thus suggesting that a source of error is removed with respect to the asynchronous coupling case. The results of *Connors and Ganis* [2011] and *Lemarié et al.* [2014] emphasize empirically a correlation between the existence of splitting errors and model uncertainties.

Physics/dynamics inconsistencies in the context of coupled problems are hard to estimate since there is a lack of academic test-cases with reference solutions including physical parameterizations. It has been shown earlier that such inconsistencies can arise from coupling algorithms or nonconformities in the space-time computational grids but also from parameterization schemes for air-sea fluxes and turbulent boundary layers. However, the mathematical formulation of those schemes is often devised semi-empirically (e.g. by fitting independent measurements) and this can impair the regularity of the associated solutions [e.g. *Burchard et al.*, 2005; *Deleersnijder et al.*, 2008], giving rise to the development and persistence of "fibrillations" in model solutions. This complexity has to be taken into account when designing mathematically consistent and efficient intra-model coupling algorithms. Indeed the use of an iterative coupling method requires a certain degree of regularity (well-posedness) of the system of equations to be coupled, otherwise convergence is not guaranteed and/or the relevance of the converged solution could be questionable. For instance, the theoretical framework of the Schwarz methods could be used to derive intra-model compatibility/consistency constraints on the turbulent boundary layers parameterizations: a pair of parameterizations will be declared compatible if the associated iterative Schwarz algorithm converges. This task is challenging because parameterization schemes are often very complicated. To investigate those issues, working on simplified equation sets to focus on specific problems should be encouraged. For example, coupling in a Single Column Models (SCMs) is representative of the functioning of the three-dimensional coupled models (see *Schulz et al.* [2001] for an example in land surface/atmosphere coupling). Because one-dimensional coupled Large Eddy Simulation (LES) are increasingly used, they should provide reference solutions to coupled SCMs for various physical situations.

As emphasized above, the coupling of models can only succeed if the problems themselves are well posed, the coupling schemes sufficiently advanced to allow for convergence, and, crucially, the processes in different models are compatible. This can only be the case if their thermodynamic basis is compatible, otherwise no common state can exist (unless be coincidence). The aspect of the thermodynamic formulation will be explored in the following section on a example in the atmosphere.

8 Thermodynamics / inter process coupling

8.1 Inclusion of moisture related quantities

With increasing mesh resolution, the impacts of cloud physics processes become more and more variable. More extreme precipitation events become explicitly represented. Associated with such events there can be large, but transient, amounts of liquid water in a single model column (*Geleyn and Marquet* [2010]; *Bacmeister et al.* [2012]). From a physical viewpoint thermodynamic properties such as density, the gas constant or the specific heat of the air-cloud-precipitation mixture may no longer be approximated by the properties for dry air. Also the barycentric velocity $\mathbf{v} = \Sigma_i \mathbf{v_i} \varrho_i / \varrho$ (with i = dry air, water vapor, rain, snow, etc) may differ from the velocity of dry air when a significant amount of condensate is present. Those issues are especially important as the largest amount of available energy in the atmosphere is processed by condensation/evaporation processes and precipitation. It has also been found that cloud physics is responsible for most of the variability of regional extreme event forcasts [*Geleyn and Marquet*, 2010; *Bacmeister et al.*, 2012], e.g. flash floods.

Non-convergence in any model is largely driven by the different behavior of cloud physics as the time step is changed (see section 3). The single microphysical conversion rates, sedimentation rates, and turbulent or convective flux amounts are not exactly known. The only physical constraint available is correct energy transformation of all moisture related processes. The sole constraints of modeling are total mass conservation and total energy conservation. Those conservation properties on the global scale are generated by local phase changes and local latent heating rates. Therefore the expected changes from the current and often quite loose approximations to more physically consistent modeling will first appear on local scales and will have impact on the simulation of localized events such as flash floods due to extreme precipitation.

Another modeling constraint is interrelated tracer consistency (e.g. [*Lauritzen and Thuburn*, 2012]. If reactive gases are not transported coherently, spurious micro-physical or chemical reactions are the consequence. Such kind of inconsistencies will in turn also affect the radiation scheme if for instance ozone chemistry is explicitly modeled.

To some extent, the mass conservation constraint and the interrelated tracer consistency constraint do not go hand in hand. The modeling groups have two choices: either using specific quantities $q_i = \varrho_i / \sum_i \varrho_i$ or mixing ratios $r_i = \varrho_i / \varrho_d$, where ϱ_d is the density of dry air, as prognostic variables. The choice of specific quantities goes hand in hand with classical thermodynamic theory known from textbooks (c.f. [*de Groot and Mazur*, 1984]). This choice supports energy and mass conservation and leads easily to the derivation of a meaningful entropy budget equation. It is also consistent with the mass-weighted averages that are assumed in turbulence modeling. A drawback of specific quantities is that the dry air compartment is not a prognostic variable, but its evolution is determined by the continuity equation for total mass. The specific dry air content is then derived from $q_d = 1 - \sum_i q_i$, where i denotes all other constituents but dry air. Therefore, an interrelated tracer consistency between dry air and other constituents is numerically no longer possible. For interrelated tracer consistency, the choice of mixing ratios is more adequate. Then the first and second laws of thermodynamics have to be re-derived adequately, which may lead to other problems.

One of the fundamental assumptions for modeling multi-phase flows in the atmosphere is a constant temperature throughout the air parcel. Hence, falling precipitation is always assumed as having the same temperature as its environment. This assumption is very crude, but relaxing it would multiply the number of prognostic variables.

Only recently have the influences of correct thermodynamics been investigated. Although correct governing equation sets including all moisture effects have been known for some time (*Bannon* [2002]; *Zdunkowski and Bott* [2003]; *Wacker and Herbert* [2003]; *Wacker et al.* [2006]; *Catry et al.* [2007]; *Gassmann and Herzog* [2015]), a lot of models still approximate thermodynamics severely without the consequences being well understood. The PDC workshop series, *Gross et al.* [2016], intends to provide a platform for bringing attention to this issue.

Consistent thermodynamics are also needed to improve the understanding of the static stability of the atmosphere, which is important when diagnosing the intensity of turbulence. For this purpose, moist entropy, moist potential temperature and moist Brunt-Väisälä frequency have been derived by Marquet and Geleyn in a sequence of publications (e.g. *Marquet and Geleyn* [2013]).

8.2 Compatibility with the second law of thermodynamics

Surprisingly, modeling communities have not yet paid much attention to the compatibility of the model formulation with the second law of thermodynamics [*Gassmann and Herzog*, 2015]. The currently widely accepted argument supporting this disregards that turbulent fluxes are not considered structurally similar to molecular or viscous fluxes, which are indeed responsible for irreversible processes in nature [*de Groot and Mazur*, 1984]. On the
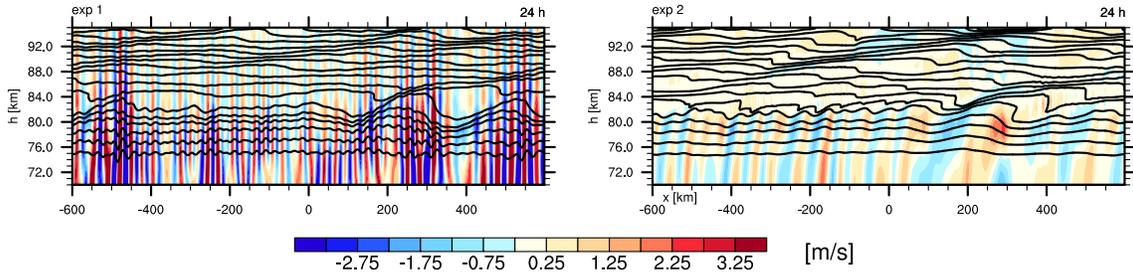
Figure 9.   Vertical velocity *w* (colors) and selected isentropes (contours) for the late phase of gravity waves breaking in the mesosphere for a classical turbulent heat flux parameterization (left) and a 2nd-law compliant variant (right). For this idealized 2-dimentional slice experiment, gravity waves are excited in the troposphere during the first 16 hours. They travel through an idealized summer stratosphere. The flow field after 24 hours is shown.

other hand, as already mentioned, the rôle of the physics parameterization is the irreversible conversion of some kind of available (kinetic or internal) energy into unavailable internal energy, whereas the dynamics is responsible for the two-way energy transfer. Since there is no other means than the dynamics for reversible energy conversions, the physics has to follow structurally the same rules as known from viscous or molecular processes. Therefore it is straightforward to raise the question of entropy production in the classical sense also for a numerically modeled atmosphere, as done for instance by *Goody* [2000]; *Bannon* [2015]; *Bannon and Lee* [2017]; *Pauluis and Held* [2002] and *Romps* [2008]. The last three authors recognize a serious problem: Turbulent heat fluxes proportional to the negative gradient of potential temperature give rise to negative entropy production and offend the second law if the atmosphere is stably stratified and turbulence is mechanically driven, as it is the case for gravity wave breaking. In this case, work has to be performed to push isentropes down in subgrid turbulent parameterizations. However, this work is not visible in the kinetic energy budget of the whole model. In order to solve the problem, *Akmaev* [2008] claims that the frictional heating – or shear production of turbulent kinetic energy – provides the necessary energy. However, from a process-oriented thermodynamic viewpoint, turbulent momentum fluxes and turbulent heat fluxes are unrelated processes according to Curie's principle, because the tensorial character of both fluxes is different [*de Groot and Mazur*, 1984]. Current and yet unpublished research (Gassmann, in preparation) aims at the resolution of the problem by introducing a turbulence-related vertical pressure gradient term in case of stable stratification. Then, the work done by buoyancy forces to push isentropes down becomes obvious as a loss of the vertical part of the resolved specific kinetic energy of vertical motion ($w^2/2$). The newly introduced turbulence-related vertical pressure gradient term acts like a Rayleigh damping term in the vertical velocity equation. Such kind of Rayleigh damping has already been used as a numerical remedy filtering gravity waves in a sponge layer of an atmospheric model by *Klemp et al.* [2008]. The vertical diffusion coefficient for potential temperature which is assigned to this apparent Rayleigh damping coefficient will have other properties than current diffusion coefficients. The diffusion coefficient depends on the kinetic energy of vertical motion. Therefore, isentropes are not pushed further down at their trough position and breaking waves can not amplify as would be the case in a conventional parameterization. Figure 9 demonstrates that a classical 2nd-law offending parameterization can lead to wave manifestation and amplification, but the 2nd-law compliant parameterization leads to wave attenuation.

If the turbulence is convectively driven, the vertical heat flux might be counter the local gradient of potential temperature in the upper part of the convective layer. The energy supply for upward mixing originates from the subgrid-scale kinetic energy (TKE). From the

perspective of the resolved model variables, the TKE and the internal energy are indistinguishable and are only representable together as internal energy. Unlike in the mechanically driven turbulence described above, an energy exchange between resolved kinetic energy and internal energy does not take place. Therefore the model equations must pretend a downgradient temperature (heat) flux $c_p \bar{\varrho} \overline{w''T''} = -c_p \bar{\varrho} K^T \partial_z \bar{T}$. Such a flux is always compliant with the 2nd law of thermodynamics for positive K^T. From the phenomenological side which knows about the distinction between TKE and internal energy, a flux formulation according to $c_p \bar{\varrho} \overline{w''T''} = -c_p \bar{\varrho} (\bar{T}/\bar{\theta}) K^\theta (\partial_z \bar{\theta} - \gamma)$ is adequate. The countergradient term γ does not lead to a violation of the second law as long as the total heat flux is down the true temperature gradient. In other words: for given K^θ and γ from an arbitrary parameterization, the 2nd law of thermodynamics requires from the relation $-c_p \bar{\varrho} K^T \partial_z \bar{T} = -c_p \bar{\varrho} (\bar{T}/\bar{\theta}) K^\theta (\partial_z \theta - \gamma)$ that K^T is a positive number. What is usually referred to as a counter-gradient flux is therefore not always violating the second law.

The discussed heat fluxes for stable and unstable stratification are the most intriguing examples where current understanding and practice has not yet matured. *Gassmann and Herzog* [2015] derive a slightly different approach for turbulent fluxes of water vapor and include a symmetric and trace-free tensor depending on shear and strain deformations in the turbulent momentum diffusion .

An issue which also becomes important with regard to the second law is the not well understood role of numerical diffusion. A lot of recent dynamical core developments have focused attention on geometrically compatible formulations, i.e. they require that some underlying geometrical law, such as the duality of divergence and gradient formulations or the curl-free gradient operator, hold in a discrete sense. Formulations that strictly respect this form of geometric compatibility lead to dynamical cores with truly reversible energy transformations, so that transformations between different kinds of energy are equally possible in both directions. Some diffusion in dynamical cores is still motivated numerically, even though this interferes with physically motivated turbulence parameterization schemes. Future work should consider the relation of these two effects (numerical diffusion / turbulence parameterization) and try to interpret numerical diffusion in a physical sense.

8.3 Future research

A question posed to the modeling communities is: Should all the parameterizations obey the second law? If the answer is yes then the consequence is that convection and gravity wave drag parameterization schemes have to be interpreted as special cases of ordinary subgrid-scale turbulence parameterizations. If the answer is no then are the efforts put into the developments of geometrically compatible dynamical cores worthwhile? Parameterization schemes of the EDMF type would then partly perform functions of the dynamical core, because they treat advection and diffusion at the same time.

9 The gray zone

As the resolution of geophysical models increases, the scale separation between the scales which are resolved by the model and the scales of the subgrid processes which are parameterized is vanishing. When a subgrid process, which is targeted by parameterizations at a lower resolution, is becoming partially resolved at a higher resolution, the model runs at a resolution in what is called the *gray zone* of this process.

9.1 Examples in current model configurations

The continuous resolution increase over the last 50 years has brought global models very close to the gray zone of convection. Limited area models for numerical weather prediction (NWP) have already jumped across the gray zone of convection, but they are now at

the verge of the gray zone of turbulence [*Wyngaard*, 2004; *Boutle et al.*, 2014; *Honnert and Masson*, 2014].

In the gray zone of a process, neither the subgrid representation of this process in a parameterization, nor the explicit representation by the prognostic parameters of the model are designed to accurately represent this process. Often, in practice, modelers either tune the existing parameterizations to extend their usage out of the range of validity of their fundamental hypotheses, or they switch off the parameterizations, even if the process is not yet well resolved but only "permitted" by the resolution, applying ad-hoc numerical filters, if necessary, to control the intensity of the process.

A recent example in the operational ECMWF model showed the difficulty of balancing the explicit and the parameterized representations of deep convection, even at a resolution of about 16 km which is not considered to be in a gray zone yet. When the convection scheme of the IFS was modified to improve the daily cycle of convection [*Bechtold et al.*, 2014], explicit convective clouds at isolated single grid points started to be diagnosed in calm conditions near mountainous and moist areas, leading locally to very unrealistic precipitation (Figure 10(a)). This new version of the convection scheme then delayed the onset of the parameterized convection toward the evening. The CAPE accumulates such that, in a region with weak orographic forcing of moist air in a no shear environment, an explicit convective cloud is permitted at a single grid point before the convection scheme is triggered. In the IFS, such single grid point structures are then pathologically amplified by the semi-Lagrangian advection scheme of the IFS dynamics [*Malardel and Ricard*, 2015]. The resulting unrealistic explicit deep convective clouds may last for several hours. This leads to spurious and very high precipitation rates at some grid points. Similar grid point storms have been reported in the literature for other global models, for example *Williamson* [2013] in CAM4, with T340 spectral truncation and a 5 min time step, (c.f. section 3) or in mesoscale limited area models at resolution in the 5-3 km range [*Malardel and Ricard*, 2015]. This example shows that, even with a hydrostatic model such as the IFS, explicit deep convective circulations are permitted at resolutions which are far too coarse to sample individual convective ascents in circumstances where the convection scheme is not triggered soon enough to release the CAPE. On the other hand, with the old version of the convection scheme, the parameterization was triggered earlier, and then the onset of the convection in the tropics was systematically too early. Finding the right balance between the explicit and the parameterized representation of convection everywhere around the globe becomes even more difficult in the gray zone of convection, at resolution around 5 km.

In the next sections, the main limitations of atmospheric models in the gray zone of convection and turbulence are listed. Recent attempts to design algorithms improving the seamless transition between resolved and parameterized treatments in these gray zones are presented.

9.2 Model limitations in the gray zones

In a gray zone, the explicit and the parameterized representations of a process are in "competition" in the numerical model. The result of this competition may be double counting or no counting at all, or, more often, the process is taken into account by both options but in a sub-optimal manner.

The prognostic equations which are solved by a numerical discrete model are the result of time and space filtering. This filtering creates an artificial cut-off in the continuous atmospheric spectrum between the processes which are represented by the mean prognostic variables of the models and the processes which are supposed to be "subgrid" and whose effect on the larger scales is parameterized. This cut-off scale is partly defined by the time/space resolution of the model and partly by the characteristics of the numerical schemes and physical parameterizations. In most numerical schemes, the largest errors are expected to happen at the cut-off scale, especially in regions of large gradients. Weaknesses of the numerical
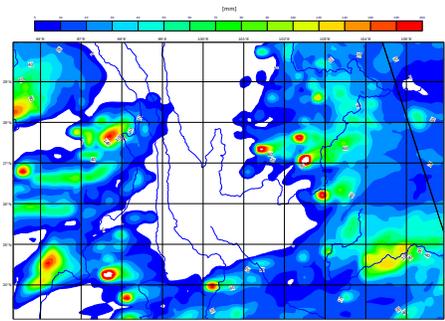
(a) Operational ECMWF forecast with a spectral truncation T1279 associated to a 16 km reduced Gaussian grid
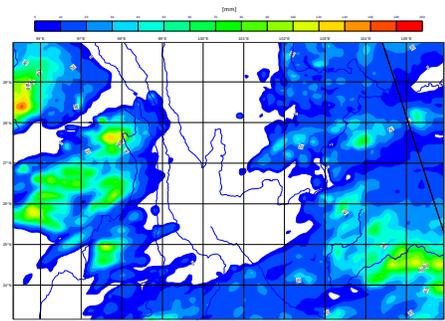
(b) IFS forecast with a spectral truncation T1279 associated to a 9 km reduced Gaussian grid.

Figure 10. 3 day accumulated surface large scale precipitation for forecasts starting on 2015-05-20 at 00 UTC VT:23 May 2015 00 UTC.

schemes such as large diffusion, large phase shift or non-conservation then directly affect the energy-containing circulation if they are permitted at this cut-off scale. Thus, as discussed by *Lander and Hoskins* [1997], physical parameterization should not force and should not be forced by the prognostic model variables containing variance at the grid scale. This statement is actually well illustrated with the case shown in Figure 10. With the horizontal resolution upgrade at the beginning of 2016, the IFS moved from a âĂIJlinearâĂİ grid to a âĂIJcubicâĂİ grid, keeping the same spectral truncation T1279. With the cubic grid, the smallest wavelength of the truncation is now represented by 4 points instead of 2 with the linear grid. Such a pairing between the spectral representation of the IFS and the model grid insure a better scale separation between the prognostic model variables and the subgrid effect computed in the physics package. In particular, single grid column resolved ascents are not allowed any more with a cubic grid. The development of grid point storms is then completely eliminated from the IFS forecast (Figure 10(b)).

Indeed, the formulation of most physical parameterizations is based on the clear scale separation, both in time and space, between a resolved environment and the parameterized processes which are treated as "perturbations" of the environment. The formulations are derived from a statistical evaluation of the impact of a large population of "perturbations" on the resolved flow, sometimes simplified by a "bulk" representation of the process, as, for example, in many convection schemes, a single convective cloud replaces a population of smaller cloud ascents. But, when the resolution increases, the grid becomes too small for a population of deep convective circulations to develop inside a grid box. In the gray zone of convection, the updrafts could cover a large fraction of the grid box area, and then the mean grid-box properties which are carried by the prognostic variables of the model should depart substantially from the updraft environment. The detrained material from the updrafts should also not be confined in the same grid column, but should be distributed over several grid columns (cf. section 5.1 and its Figure 3).

However, the Reynolds decomposition which is used to derive the eddy diffusivity formulation for non-local 1D turbulence or the mass flux formulation of most convection schemes does not allow any net mass transport by the "perturbations" in a grid box: updraft and downdraft have to cancel in the same grid box, rendering the problem of extending the detrainment to neighboring cells difficult to generalize (c.f. section 5.1).

The equilibrium hypothesis [*Arakawa and Schubert*, 1974] resulting from the hypothesis of the scale separation in time also starts to collapse when the time step of the model decreases. With an increase of resolution in time, the variability of the model variables becomes faster than the characteristic convective time scales, *Gerard and Geleyn* [2005]. If the closure of the convection scheme is too simple, spurious explicit convective storms are more likely to develop at high resolution [*Williamson*, 2013; *Gerard*, 2015] (cf. section 3.3).

The generalization from a 1D to a 3D treatment of a subgrid process often does not only involve the physical concepts behind the parameterization but also the code architecture of the model. Indeed, the scalability of the model benefits from the independence between the columns in the physics. Thus, the transition from 1D algorithms to 3D parameterizations involving horizontal exchanges with neighboring columns is also a challenge in term of code development and efficiency of the algorithms, especially for NWP and climate models.

9.3 Towards scale aware parameterizations

Efforts to develop scale-aware convective parameterization started in the Limited Area Model (LAM) community more than ten years ago [*Gerard and Geleyn*, 2005] and are now shared by a much larger community [*Arakawa and Wu*, 2013; *Gustafson et al.*, 2013; *Grell and Freitas*, 2014; *Siebesma*, 2015].

As already pointed out in the previous section, one of the main issues for the parameterization of deep convection in the gray zone is that, when the resolution of the model in-

creases, some of the condensates can be detrained across the gridbox and the ensuing compensating subsidence should take place within another grid box than the originating one. With the time step organization of NWP codes, this transport can only be handled by the advection of the dynamical core.

Piriou et al. [2007] observed that the advantage of Cloud-System-Resolving Models (CSRMs) with respect to parameterized budget equations is that the source terms for the convection can be separated into transport terms and microphysic terms and they argued that the two types can be treated independently. Moreover, if the condensation (and the cloudy evaporation) terms in cloud budgets models are computed by a microphysics scheme and provided as source terms to the environment, then the system can be closed, leading to CSRM-type equations that still do not contain detrainment terms. In that case, there is no need to directly rely on the budget equations to close the system.

However, to go from CSRMs to gridbox parameterizations, it is necessary to partition the grid box into a convective and a non convective part. *Gerard et al.* [2009] used the cloud scheme of *Xu and Randall* [1996] and *Smith* [1990] to introduce a protection of the cloud condensates in the convective part, to prevent their evaporation by the cloud microphysics scheme. Additionally they used a prognostic formulation of the convective mesh fraction of the updraft and a prognostic equation for the updraft-vertical velocity proposed in *Gerard and Geleyn* [2005]. The result is a CSRM-type set of equations without any explicit presence of detrainment terms. In other words, it interacts with the dynamics in the same manner as a CSRM-type of model does.

One can argue that bulk parameterizations should converge in their behavior to the behavior of CSRMs in the cloud-resolving limiting resolutions. If the prognostic equations of the mesh fraction and the updraft-vertical velocity scale properly, then the equations should converge to the equations of a CSRM. This yields a mechanism to control this convergence and to formulate a scale-aware parameterization of deep convection.

This approach was implemented in a scheme called the Modular Multiscale Microphysics and Transport scheme (3MT) and it formed the basis of the so-called ALARO-0 configuration of the ARPEGE-ALADIN system. *Gerard et al.* [2009] showed satisfactory results of this scheme with resolution ranging from the mesoscale up to 4 km, cf. Figure 11 of *Gerard et al.* [2009], where it can be seen that without 3MT the model did not resolve the organized convection satisfactorily, resolving only a small (in number and size) of intense gridpoint storms.

Recently, good results were found with an updated version of the scheme up to a resolution of about 1 km. *Meutter et al.* [2015] tested a version of the 3MT scheme which includes the parameterization of unsaturated downdrafts. They found downdraft mass fluxes that are sufficiently realistic so that they can be used operationally to forecast downbursts. *De Troch et al.* [2013] demonstrated that the ALARO model has an improved multiscale character than the former ALADIN configurations.

These efforts still need to be generalized for global NWP. There very different types of convective circulations, in the tropics and at higher latitudes, have to be well represented for medium range weather forecasting. It is also of importance for climate models in order to maintain a correct large scale balance. Recent results with variable resolution meshes [*Müller*, 2014] also show the need for scale-aware physics across the difficult range between $1 - 10$ km.

LAM models which are running at sub-kilometer resolutions do not use any parameterization of deep convection. But, the parameterization of eddies in the boundary layer is still needed. A blending between a 3D turbulence parameterization originally designed for LES and a 1D boundary layer parameterization suitable for coarser grid resolution has been shown to be very beneficial to the representation of clear or stratocumulus-topped boundary layers by *Boutle et al.* [2014]. In this case, the transition laws brought out by *Honnert and*

Masson [2014] are used to seamlessly drive the transition between unresolved to resolved turbulence. But more efforts are needed to generalize this approach to any regime in the boundary layer.

Malardel and Wedi [2016] identify the influence of subgrid scale parameterizations for the shape of the kinetic energy (KE) spectra as well as for the non-linear spectral fluxes at all scales. The artificial scale separation between resolved and subgrid processes modifies the natural turbulent energy cascade driven by advection. When the processes are parameterized, the circulation which is responsible for the average effect of the subgrid mixing is neither part of the "resolved" KE spectra, nor part of the non-linear spectral transfer, thus effectively disabling any energy cascade.

The temptation to enable the natural cascade by eliminating a particular parameterization too early in a gray zone is however risky as the model balances change at all scales as a result. Such practice may also have implications on the forecast error growth, as the predictability time of a k^{-3} system can be much longer than that of a $k^{-5/3}$ system [*Palmer et al.*, 2014]. However, it is unclear if a growing error is merely replaced by a much larger error injected at multiple scales when the process is parameterized.

The continuous increase of resolution is not the only change in the model environments that generate PDC challenges. Variable resolution and high order methods for example are two further areas that emerge on the horizon and are discussed in the following section.

10 Emerging challenges

The ecosystem of models is constantly evolving and new methods become available and feasible, replacing older, often somewhat simpler technologies. Currently the advent of high order finite element methods offers many more choices to the coupling than a grid point model would. Likewise spatially varying resolution and or adaptive refinement is used more and more often, partly due to the availability of mimetic methods which support this sort of models. This however is not without challenges to the coupling of the multiple (truncation) scales now present in the model.

10.1 Spatial physics-dynamics coupling with element-based high-order Galerkin methods

Numerical methods based on element-based high-order Galerkin discretization [see, e.g., *Durran*, 2010] have reached a level of maturity in which they are being considered for next generation weather and/or climate models. For example, the spectral-element dynamical core in NCAR's [CAM; *Neale et al.*, 2010a], referred to as CAM-SE [*Taylor et al.*, 2008; *Taylor and Fournier*, 2010; *Dennis et al.*, 2012b], is currently being used for high resolution climate modeling [e.g., *Small et al.*, 2014]. Other examples are *Giraldo and Restelli* [2008], *Nair et al.* [2009] and *Brdar et al.* [2013]. Below the focus is on CAM-SE, however, in principle the discussion applies to any element-based high-order Galerkin method.

To advance the solution to the equations of motion in time, element-based high-order Galerkin methods typically apply quadrature rules to numerically integrate basis functions over a reference element. The choice of quadrature rule is application dependent and can have consequences for the properties of the final algorithm; in particular algorithm efficiency. In CAM-SE Gauss-Lobatto-Legendre (GLL) quadrature is used which exactly integrates Lagrange polynomials up to degree $2p - 1$, where $p + 1$ is the number of quadrature points. For an introductory discussion on emerging Galerkin methods in the context of atmosphere modeling see, e.g., *Nair et al.* [2011].

Irrespective of the choice of quadrature rule, the quadrature points for higher-order methods are not equally spaced over the sphere and reference element. The higher the order, the more the quadrature points tend to cluster near the sides and, in particular, corners of

(a) ne4np4 quadrature points on cubed-sphere
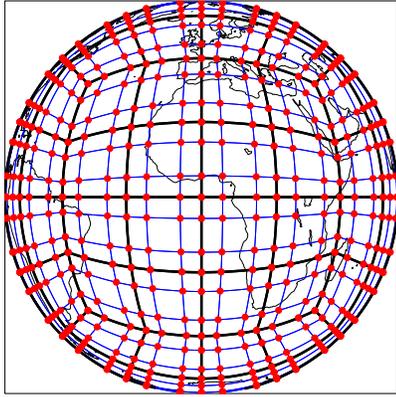
(b) ne2np8 quadrature points on cubed-sphere
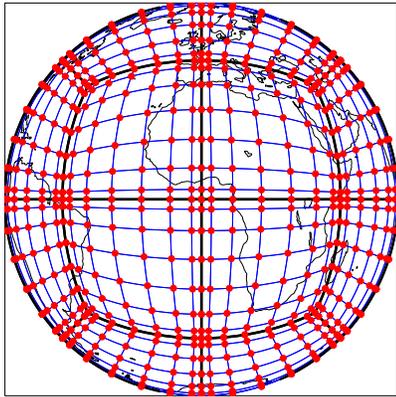
Figure 11. The CAM-SE GLL quadrature grid (red filled circles) on the cubed-sphere using (a) 4 × 4 elements per panel ($ne = 4$) and 4 × 4 quadrature points in each element ($np = 4$), referred to as *ne4np4* grid, and (b) 2 × 2 elements on each panel ($ne = 2$) and 8 × 8 quadrature points in each element ($np = 8$), referred to as the *ne2np8* grid. The average grid spacing at the Equator is approximately the same (7.5°) for both grids. The boundaries of the elements are marked with thick black lines.

the elements. As far as the authors are aware, current dynamical cores employing element-based high-order Galerkin methods use the quadrature point values for the state of the atmosphere passed to subgrid-scale parameterizations (physics). This approach follows the traditional model setup where physics and dynamics grids coincide. One may question if that is an appropriate choice for element-based high-order Galerkin methods. Physical parameterizations expect a state of the atmosphere representative of the area for which it should compute tendencies, a grid-cell averaged state of the atmosphere, for example. Obviously the quadrature point values are representative of the state of the atmosphere at the quadrature point and in the vicinity of the quadrature point but what area is associated with the quadrature point value? If one defines areas around the quadrature points so that the spherical area exactly matches the GLL quadrature point weight times the metric factor, a very irregular grid results. Hence the state of the atmosphere passed to physics is sampled anisotropically in space.
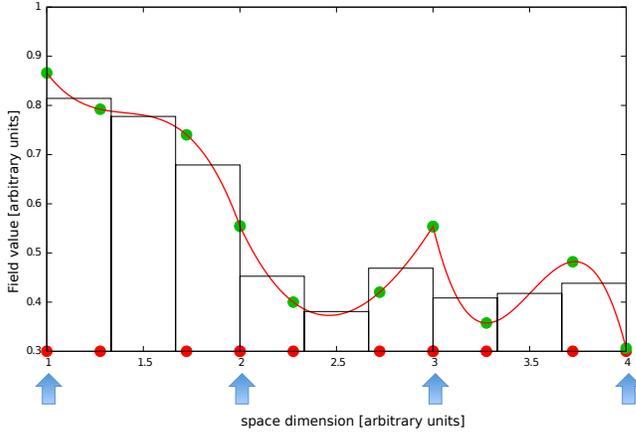
Figure 12. The figure shows three elements in one dimension. The edges of the elements are marked with blue arrows. The red curves are the degree 3 polynomials in each element and, following the CAM-SE algorithm, the polynomial values from each side of an element boundary are averaged. The filled green circles show the GLL quadrature point values and the red filled circles are the location of the GLL quadrature points in each element for np = 3. The histogram bar shows the cell averaged values on a nc = 3 physics grid (each element has been divided into 3 equal-sized control volumes) obtained by integrating the Lagrange basis functions over the control volumes.

Assuming that physics should be given a grid cell average value, it may be argued that it would be more consistent to integrate the basis functions within each element over quasi equal-area control volumes. From an implementation point of view it is convenient to have the control volumes sub-divide the element so that no control volume spans part of the neighboring elements. Note that the basis functions are C^∞ within each element but only C^0 at the element boundaries. If there is a strong grid-scale forcing at the element boundary, the physics grid value may be more representative than the extrema value (see Figure 12). This configuration, where physics and dynamics grids are separated, is referred to as *physgrid*. Care must be taken when mapping variables to and from dynamics and physics grids so that conservation properties are not violated. Here the arbitrary high-order, conservative and consistent remapping algorithm of *Ullrich and Taylor* [2015] is used. The algorithm consists of matrices that can be pre-computed: For mapping from the dynamics to the physics grid it consists of one matrix that performs a shape-preserving, but low-order, remap and another matrix that is not shape-preserving but high-order. The algorithm has been modified such that the two matrices in each element are optimally combined (linearly) so that the method is shape-preserving and, where possible, high-order. For mapping the tendencies back to the GLL quadrature grid a low-order conservative and shape-preserving method is used. The mapping algorithm accommodates any order of basis functions.

The next step is the choice of physics grid resolution: Same resolution as the dynamical core, coarser or finer? *Lander and Hoskins* [1997] argued, in the context of a spectral transform model, that the physical parameterizations should only be given believable scales. From linear theory it is well known that numerical methods used in the dynamical core do not represent the shortest wavelengths, e.g., the $2\Delta x$ wave, accurately. It may therefore be argued that the physical parameterizations should not be passed scales that, from linear theory, are not accurately represented. On the other hand, computing physics tendencies on a higher-resolution grid compared to the dynamical core may provide a better sampling of the atmospheric state, somewhat similar to the super-parameterization [*Grabowski*, 2001; *Khairoutdinov and Randall*, 2001] and sub-columns concepts [*Barker et al.*, 2002; *Pincus et al.*, 2003; *Thayer-Calder et al.*, 2015].

In this section the consequences of separating physics and dynamics grids in CAM-SE, as described above, are explored. The 1° version of CAM-SE is used, i.e. the $ne30np4$ configuration (30×30 elements on each panel, $ne = 30$, and 4×4 quadrature point, $np = 4$, in each element), and the physics tendencies are computed on the GLL grid (the grid of the dynamical core), a coarser (1.5°) physgrid, same resolution (1°) physgrid and finer (0.75°) resolution physgrid. The four configurations are referred to as $ne30np4$, $ne30np4nc2$, $ne30np4nc3$, $ne30np4nc4$, respectively, where $nc2$ refers to a 2×2 quasi equal-area physics grid in each element. Similarly for $nc3$ and $nc4$. Note that the GLL grid is the grid on which the dynamical core operates. Aqua-planet simulations [*Neale and Hoskins*, 2000] are performed with CAM4 physics [*Neale et al.*, 2010b] and the physics time step is the default 1800s. The reasoning behind choosing CAM4 physics instead of the newer CAM5 physics is that CAM4 physics is more resolution dependent [e.g. *Bacmeister et al.*, 2014; *Zarzycki et al.*, 2014a]. CAM4 physics is therefore expected to produce more physgrid resolution dependence than CAM5. Simulation length is 30 months and only the last 24 months are used for analysis. The code base used is revision 65448 of `https://svn-ccsm-models.cgd.ucar.edu/cam1/branches/physgrid`. Standard out-of-the-box namelist settings for the spectral element dynamical core were used.

Figure 13 shows the zonal-time average of surface pressure, total precipitation rate, total cloud fraction, and albedo as a function of latitude (from Equator to 80°N) for the different model configurations. The surface pressure field follows a slight decrease with increased physics grid resolution North of approximately 55°N. In the simulations presented in *Williamson* [2008] the surface pressure exhibits the same behavior when the model resolution was increased (see their Figure 4). Precipitation rates show little dependence on physics grid resolution except at the Equator. For total cloud fraction *Williamson* [2008] observed that the fraction decreased with increasing resolution. This is observed for the physgrid $nc = 2$ and $nc = 3$ simulations, whereas the $nc = 4$ cloud fractions are mostly bounded (in between) the $nc = 2$ and $nc = 4$ cloud fraction values. The same is observed for albedo. So for the zonal time-averaged fields there seems to be little dependence on physics grid resolution.

To investigate the effect of separating physics and dynamics grids on extreme events the fraction of time precipitation has been binned in 1 mm d^{-1} bins ranging from 0 to 120 mm d^{-1} (not shown). Only a very small dependence on physics grid resolution was observed for the high precipitation rate events.

In all, the physgrid configuration of CAM-SE has been demonstrated to produce aquaplanet results that are similar to the baseline (no physgrid) version. The dependence on physics grid resolution is different for different fields. The aquaplanet setup does not have stationary grid scale forcing and is only suitable for analyzing the free modes in the atmosphere. The next step is to investigate the effect of a physics grid on applications with stationary grid scale forcing (e.g., topography). CAM-SE had been found to produce some noise if topography is not sufficiently smoothed [*Lauritzen et al.*, 2015a]. The physgrid configuration has shown promise (not shown) in alleviating spurious grid-scale precipitation near steep orography due to the averaging over control volumes (especially near the edges of the elements). Similarly, the physgrid may improve simulations of other fields exposed to strong grid scale forcing such as photolysis driven tracers. An idealized test to investigate this has recently been developed [*Lauritzen et al.*, 2015b].

10.2 Emerging challenges in PDC with multi-scale models

The last decade has seen a surge of atmospheric dynamical cores formulated on quasi-uniform grids, which do not suffer from the pole problem existing in latitude-longitude grids [*Williamson*, 2007]. Their development is also driven by the need to improve scalability on massively parallel computers and by diverse model applications from weather prediction to atmospheric chemistry and climate projections. For these applications, the dynamical cores must satisfy
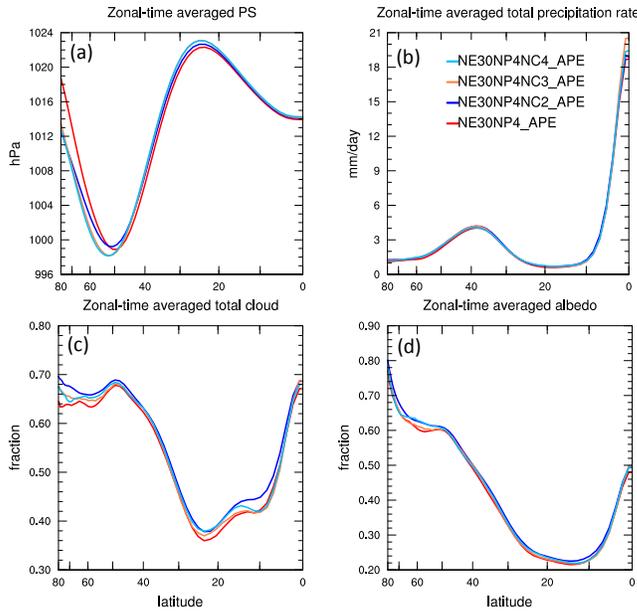
Figure 13. Zonal-time average (a) surface pressure, (b) total precipitation rate, (c) total cloud fraction, and (d) albedo as a function of latitude (from Equator to 80°N) for the different configurations of CAM-SE. The data has been averaged over a period of 24 months and mapped to a 1.5°×1.5° regular latitude-longitude grid for analysis.

several properties, such as conservation, compatible or mimetic properties, and accurate representation of global-to-meso-scale flows [*Taylor and Fournier*, 2010; *Ringler et al.*, 2010; *Skamarock*, 2011; *Staniforth and Thuburn*, 2012]. These numerical techniques, along with progress in grid generation [*Tomita et al.*, 2002; *Anderson et al.*, 2009; *Ju et al.*, 2011; *Walko and Avissar*, 2011], make it possible to increase grid resolution locally while maintaining a quasi-uniform resolution outside the refined domain. The associated grids are often described as unstructured because each cell is identified by a unique index and its connectivity to the neighboring cells, due to non-rectangular cell shapes and/or local coordinate system used in the numerical scheme. Unstructured grids are amenable to tiling complex geometry in irregular patterns. This is in contrast to traditional latitude-longitude grids in which each cell is regularly distributed and identified by global x, y indexing in both physical and computational space (structured grids). Examples of unstructured grids in quasi-uniform and variable resolution configurations are shown in Figure 14.

Local mesh refinement is also possible by stretched-grid methods on structured grids that are continuously and conformally transformed to achieve higher gridcell density over a specified region [*Schmidt*, 1977; *Staniforth and Mitchell*, 1978]; see *Fox-Rabinovitz et al.* [2006] and *McGregor* [2013] for general reviews. In this technique the number of the grid points remains the same after the transformation, so the increase in resolution over one region must be compensated by the decrease in resolution in the rest of the model domain. Recently, the stretched-grid method has been extended to unstructured grids by *Uchida et al.* [2016]. Here, all approaches that refine the horizontal resolutions over one or more regions on a global grid are referred to as the variable-resolution (VR) approach.

With global VR models, higher horizontal resolutions can be achieved in area(s) of interest while the computational burden is reduced relative to global high-resolution simulations due to coarser resolution over the remainder of the globe. The VR approach can avoid some of the known issues in limited-area models, such as the treatment of lateral bound-
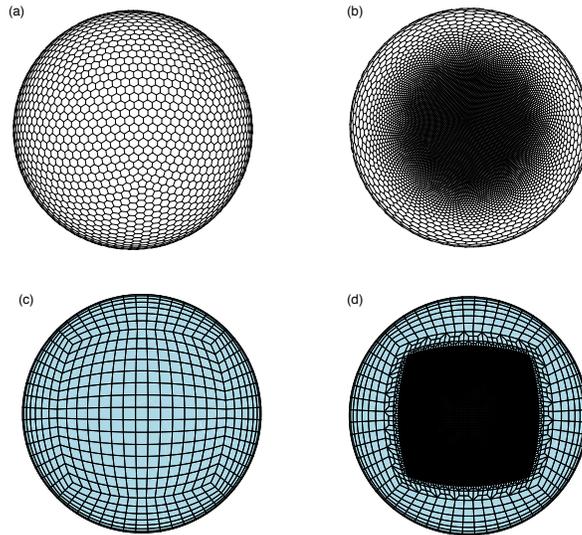
Figure 14. Examples of unstructured grids with quasi-uniform and variable resolutions: (a) quasi-uniform resolution centroidal Voronoi grid used for MPAS-A, (b) variable-resolution centroidal Voronoi grid, (c) quasi-uniform cubed sphere grid used for CAM-SE, and (d) variable-resolution cubed-sphere grid.

Table 1. List of references for variable-resolution modeling.

Model Name	Grid	References
CAM-SE	Cubed-sphere	*Zarzycki et al.* [2014a,b]; *Zarzycki and Jablonowski* [2014, 2015]; *Zarzycki et al.* [2015] *Rhoades et al.* [2016]; *Huang et al.* [2016]
HiRAM-FV3	Cubed-sphere	*Harris and Lin* [2013, 2014]; *Harris et al.* [2016]
MPAS-A	Centroidal Voronoi	*Ringler et al.* [2011]; *Skamarock et al.* [2012]; *Hagos et al.* [2013]; *Rauscher et al.* [2013] *Park et al.* [2013]; *Rauscher and Ringler* [2014]; *Park et al.* [2014]; *Sakaguchi et al.* [2015 *Fowler et al.* [2016]; *Zhao et al.* [2016]
NICAM	Icosahedral	*Tomita* [2008]; *Shibuya et al.* [2016]; *Uchida et al.* [2016]
OLAM	Icosahedral	*Medvigy et al.* [2008, 2010]; *Walko and Avissar* [2011]; *Medvigy et al.* [2011, 2013]

aries, consistency between the global and regional models, and lack of two-way interactions between the regional simulations and their driving global simulations [*Wang et al.*, 2004]. Idealized testing demonstrates that properly-designed numerical schemes on VR grids can provide additional fine-scale information at the regional scale without decreasing the accuracy of the global solution [*Ringler et al.*, 2011; *Ullrich and Jablonowski*, 2011; *Guba et al.*, 2014].

The advantages and challenges of VR weather and climate modeling have been actively studied. Table 1 provides a non-exhaustive list of recent studies of VR atmospheric model simulations for interested readers. The general consensus is that VR models can provide the benefits of high-resolution simulation inside or even outside the refined domain, such as improved orographic precipitation and snow cover [*Rhoades et al.*, 2016], tropical cyclones [*Zarzycki and Jablonowski*, 2014, 2015; *Zarzycki et al.*, 2014b], land cover representation [*Medvigy et al.*, 2011], remote influence from high-resolution regions [*Medvigy et al.*, 2013; *Sakaguchi et al.*, 2016], and overall regional climate metrics [*Medvigy et al.*,

2010; *Harris and Lin*, 2014; *Harris et al.*, 2016; *Sakaguchi et al.*, 2015; *Zarzycki et al.*, 2015; *Huang et al.*, 2016]. Boundary effects have also been evaluated, finding little artifacts in propagating waves throughout the variable-resolution domain [*Harris and Lin*, 2013; *Hagos et al.*, 2013; *Park et al.*, 2014; *Zarzycki et al.*, 2015]. So far the most challenging issue for VR models is related to (unphysical) sensitivity of physics parameterizations to spatial and temporal resolutions, although there are other potential challenges such as optimum topography smoothing on VR grids [e.g., *Zarzycki et al.* [2015]].

The previous sections (3, 4, and 9) illustrated several examples of undesirable sensitivities of weather and climate models to temporal and spatial resolutions. Specifically, section 3 discussed the mismatch between the predefined physical process time scales and the model time steps and how the mismatch affects the interaction between convection, cloud microphysics, and resolved dynamics in the sequential-update time splitting scheme. Similar sensitivities could negatively affect VR simulations that feature multiple resolutions within a single simulation. Indeed some studies in Table 1 reported non-negligible influence of resolution sensitivity on the overall quality of the VR simulations. For example, a striking difference in precipitation appears inside and outside the high-resolution domains in aquaplanet VR simulations using the CAM-SE dynamical core [*Zarzycki et al.*, 2014a] or the MPAS-A dynamical core [*Hagos et al.*, 2013; *Rauscher et al.*, 2013; *Zhao et al.*, 2016] with the CAM4 and CAM5 subgrid physics suites. In the following the effects of physics-dynamics coupling on the model sensitivity to spatial resolution in VR modeling are briefly explored.

Aquaplanet experiments were conducted using MPAS-A with the CAM4 physics as in *Williamson* [2013], who used the Eulerian spectral model with the same CAM4 physics. The version of MPAS-A used is the hydrostatic model described in *Park et al.* [2013] and *Rauscher et al.* [2013]. The model was configured with three different grids: quasi-uniform (QU) 240 km, QU 120 km, and a VR grid with 30 km grid spacing at the center of the refined domain over the equator, transitioning to 240 km grid spacing on the rest of the globe. There were no modifications to the CAM4 parameterization suite, and the same tuning and configuration were used in all simulations except for the numerical diffusion coefficient, which was adjusted based on gridcell size [*Rauscher et al.*, 2013]. The same dynamics time step of 100s was used in all simulations and for each grid cell in the VR simulations. The physics time step was defined independently of the dynamics time step. For each resolution, simulations with three different ratios (R) of the physics time step (Δt) to the convective relaxation time scale (τ) were run: $R = 1/6$ ($\Delta t = 600$ s and $\tau = 3600$ s), $R = 1/2$ ($\Delta t = 1800$ s and $\tau = 3600$ s), and $R = 1$ ($\Delta t = 600$ s and $\tau = 600$ s). Although different combinations of Δt and τ can produce the same R values, the analysis was limited to the aforementioned three combinations.

Comparing the two QU simulations, it is seen that as R approaches unity, the sensitivity to grid spacing of the total (convective + large-scale) precipitation (Figure 15a) as well as that of the fraction of convective to total precipitation (Figure 15b) become smaller.

This dependence of the resolution-sensitivity on R has a visible impact on the VR simulations. For R equals $1/6$, the zonal anomaly (relative to the zonal mean) of precipitation appears on the western or downwind side of the refinement (Figure 15c), and the attendant latent heat excites a Gill-type circulation apparent in the 200-hPa velocity potential (Figure 15e, also see *Hagos et al.* [2013], *Rauscher et al.* [2013], and *Zarzycki et al.* [2014a]). Theoretically the zonal anomaly should be nearly random spatially because there is no longitude-dependent forcings in the aquaplanet configuration. With $R = 1$, the model precipitation exhibits a substantially weaker sensitivity to the change of resolution (Figure 15d). Making τ comparable to Δt, convection is more active in removing the instability created by the resolved dynamics [*Williamson*, 2013]. However, the zonal anomaly is still visible and the undesirable Gill-type circulation is not entirely eliminated (Figure 15f).

More sophisticated modifications of convective schemes to address multiple resolutions have been proposed. From the above VR example, one may expect more physical
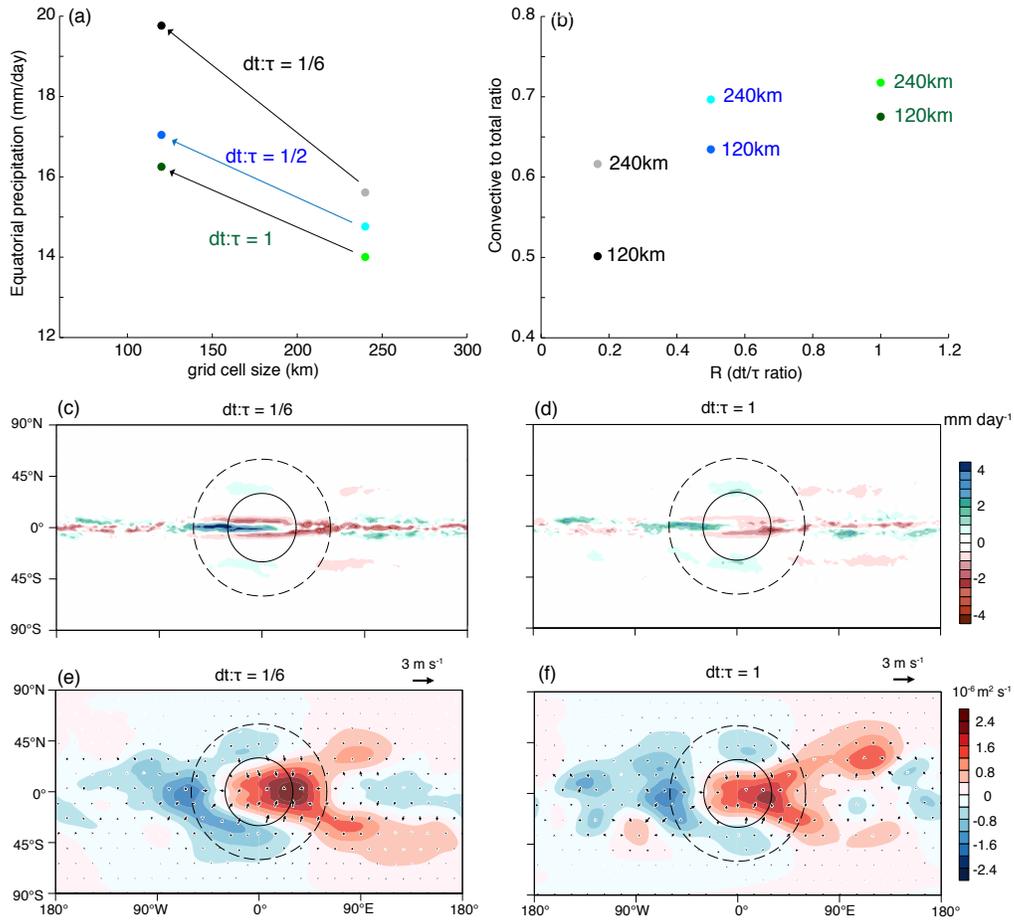
Figure 15. Influence of R ($\Delta t/\tau$) on the resolution sensitivity of the CAM4 physics (precipitation) to quasi-uniform and variable resolutions using MPAS-A. (a) Sensitivity of equatorial (+/- 2° latitude) precipitation to grid cell size (x-axis) in different values of R as represented by three arrows, (b) Fraction of convective precipitation as a function of R (x-axis) and grid cell size (240 km vs. 120 km), (c) Zonal anomaly of precipitation in a VR simulation with $R = 1/6$, (d) same as (c) but a VR simulation with $R = 1$, (e) zonal anomaly of velocity potential (shading) and divergent component of wind (arrows) with $R = 1/6$, and (f) Same as (e) but for $R = 1$. The solid and dashed circles in (c)-(f) represent the boundaries enclosing the domain with 30 km grid and the transition to 240 km grid domain, respectively.

behavior if τ is allowed to vary with the grid spacing as opposed to using a constant value across the globe. *Ma et al.* [2014] and *Gustafson et al.* [2014] suggested a simple formulation of τ as a linear function of grid spacing:

$$\tau = \max\left(\tau_{min}, \tau_{max}\frac{\Delta x}{\Delta x_{ref}}\right) \tag{4}$$

where $\tau_{min} = 600$ s, $\tau_{max} = 3600$ s, Δx is the grid spacing, and $\Delta x_{ref} = 120$ km (275 km in *Ma et al.* [2014]). This function is plotted in Figure 16a. *Fowler et al.* [2016] tested another simple method to achieve scale-aware representation of convection in VR MPAS-A simulations. They used the *Grell and Freitas* [2014] (hereafter GF) convection scheme, which follows the approach originally suggested by *Arakawa et al.* [2011] and *Arakawa and Wu* [2013]. In *Fowler et al.* [2016], the cloud-base mass flux is scaled by a quadratic function $(1 - \sigma)^2$, with σ being the fraction of convective cloud cover:

$$\sigma = \frac{0.04\pi}{A\varepsilon^2}, \tag{5}$$

where A is grid cell area and ε is the initial entrainment rate of $7 \times 10^{-5}\text{m}^{-1}$. With this grid-size dependent scaling, convective precipitation is mostly parameterized with $> \approx 40$ km grid spacing, but the fraction of parameterized convection rapidly decreases over the $5 - 30$ km range, and most convection is explicitly resolved with < 5 km grid spacing (Figure 16b). Repeating our aquaplanet simulations with the two modifications of *Ma et al.* [2014] and *Fowler et al.* [2016] is beyond the scope of this review. However, taking advantage of the simplicity of the two approaches, a heuristic, idealized analysis can be made to compare their resolution-sensitivity in the context of VR modeling and the time-scale mismatch described in *Williamson* [2013].
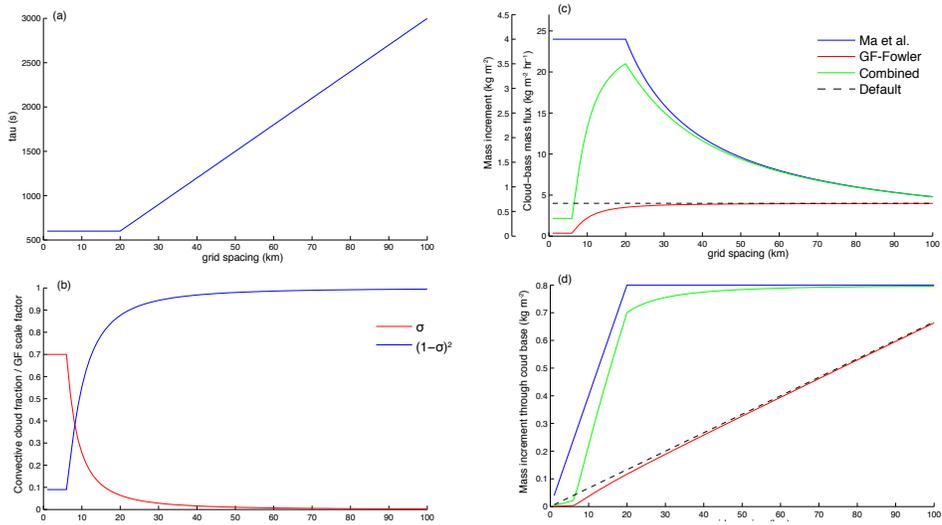
From the closure assumption of the Zhang and McFarlane scheme (*Zhang and McFarlane* [1995], hereafter ZM) used in CAM, the cloud base mass flux (M_b, kg m^{-2}hour^{-1}) can be diagnosed given certain values of CAPE and the CAPE consumption rate per unit cloud base mass flux, F:

$$M_b = \frac{\text{CAPE}}{\tau F}. \tag{6}$$

Here CAPE = 1000J kg^{-1} and $F = 250$J m^{2}kg^{-2}, respectively, and it is assumed that F does not change substantially with resolution for a given value of CAPE, which is reasonable based on CAM5 simulations on a one-degree and two-degree grid with the finite-volume dynamical core. The value chosen for F is higher than those typically found over tropical oceans ($F \approx 100$J m^{2}kg^{-2} for CAPE ≈ 1000J kg^{-1}), but convenient for showing different M_b in the same plot. In the default CAM, τ is 3600 s. In the *Ma et al.* [2014] approach, the constant τ was replaced with equation 4. For the GF approach as implemented by *Fowler et al.* [2016], M_b can be expressed as

$$M_b = (1 - \sigma)^2 \frac{\text{CAPE}}{\tau F} \tag{7}$$

In Figure 16c M_b is shown as a function of grid size (with the inside vertical axis). With constant τ, the default M_b is constant across resolutions. With the modification following GF- Fowler (red line), M_b gradually decreases with reduced grid spacing, reaching a minimum value that is specific to the implementation of *Fowler et al.* [2016]. This curve mimics the behavior of convection reported in their study. The *Ma et al.* [2014] approach exhibits a different behavior (blue line), with M_b increasing with decreasing grid spacing and reaching a maximum value at a grid size that depends on the tunable parameter τ_{min}. A simple combination of the two by substituting equation 4 in 7 is also shown (green line).

Figure 16. Illustration of the *Ma et al.* [2014] and *Fowler et al.* [2016] approaches for scale-aware convection using the Zhang and McFarlane closure. In (a) τ from *Ma et al.* [2014] is shown as a function of grid spacing. In (b) the fractional convective cloud cover (σ, red line) is shown and scaling factor for cloud-base mass flux used in *Fowler et al.* [2016]. In (c) the cloud-base mass flux (inside y-axis) is shown based on the Zhang and McFarlene closure (equation 4) with CAPE $=$ 1000J kg^{-1} and F $=$ 250 J m^2 kg^{-2} and different modifications. The four lines represent: Dashed: the default with τ $=$ 3600 s in equation 6 (Default); Blue: equation 6 with τ following equation 4 (*Ma et al.* [2014]); Red: equation 7 with τ $=$ 3600 s (GF-Fowler); and Green: equation 7 with τ as in equation 4 (Combined). The outside y-axis in (c) shows the mass increment through the cloud base for Δt $=$ 600 s (i.e., multiply each curve by 600). In (d), the mass increment through the cloud base is shown for the same cases in (c), using $\Delta t = 6 \times \Delta x$.

A different picture emerges when M_b is multiplied by Δt to obtain the mass increment over one time step, which is an important quantity as it represents the effectiveness of deep convection to remove instability generated over one time step. Figure 16d shows the mass increment assuming that Δt changes with Δx, for instance by considering the Courant–Friedrichs–Lewy (CFL) stability criteria. For simplicity, Δt is set to $6 \times \Delta x$ (e.g., recommendation in *Wang et al.* [2016]). The *Ma et al.* [2014] approach maintains a constant mass increment independent of grid spacing for gridcells larger than the gray zone. By doing so, this approach allows deep convection to be active even as Δx (and Δt) becomes smaller, thereby reducing the time step sensitivity elucidated by *Williamson* [2013]. The $R = 1$ case in Figure 15a,b is another demonstration of the same principle of allowing deep convection to be active by making τ and Δt comparable. On the other hand, the GF-Fowler approach produces a mass increment equal to or less than the default case. This suggests that the GF-Fowler approach will likely produce the same positive feedback and truncation-scale storms at high resolution as described in [*Williamson*, 2013], if it is implemented in CAM with the ZM scheme. This problem may be avoided by combining the GF-Fowler formulation of mass flux with equation 4 (green line). Note that the closure assumption in the GF scheme is different from that in the ZM scheme and does not include a predefined convection time scale (equation (19) in GF)), and both *Gustafson et al.* [2014] and *Fowler et al.* [2016] use the parallel-split, in which tendencies from the convection scheme do not directly affect the behavior of the cloud microphysics. Therefore the resolution-sensitivity described here is for illustrative purpose and is not directly applicable to their simulations.

In the VR models, it is common to use a constant dynamics time step that satisfies the stability criteria of the smallest grid cell for all the gridpoints in the global domain. Generally, the physics time step (Δt) is also fixed in relation to the dynamics time step. In this case the mass increment over Δt behaves in the same way as M_b (i.e., as in Figure 16c), because M_b is multiplied by the same constant Δt regardless of the grid spacing. The outside vertical axis of Figure 16c shows the mass increment for a constant Δt of 600 s. Based on this plot, the behavior of the GF-Fowler approach seems to be more desirable for VR models with a constant Δt and the CAM's sequential-update splitting. The modification by *Ma et al.* [2014] (and the result in Figure 15d,f with $R = 1$) would introduce larger mass increment from the parameterized convection for smaller grid spacing even beyond the gray zone, which is asymptotically erroneous. This artifact arises from the implicit assumption that Δt and Δx vary together, but this assumption breaks down in VR models. This simple analysis illustrates the dependence of scale-aware convection representation on the Physics-Dynamics coupling, such as the time splitting method or the co-variation between Δt and Δx, which has not been elucidated in the VR modeling framework.

It is worth noting that global simulations using the newer CAM5 parameterization suite shows improvement in some of these aspects, particularly with respect to cloud fraction and precipitation scaling in VR simulations [*Zarzycki et al.*, 2014a, 2015; *Zhao et al.*, 2016]. Both CAM4 and CAM5 use the same deep convective scheme; therefore aspects of the scale sensitivity seen with CAM4 may not be purely driven by the lack of scale-awareness such as related to the convective timescale. In aquaplanet simulations, *Zarzycki et al.* [2014a] found a smaller scale sensitive precipitation response with increasing resolution when using the CAM5 parameterization suite versus CAM4, a result that was replicated using more complex Atmospheric Model Intercomparison Project (AMIP)-style simulations with regional refinement [*Zarzycki et al.*, 2015]. In these cases, smaller anomalies in total precipitation with CAM5 are noted across grid spacings, which reduce the magnitude of spurious, physics-induced circulations as described above. Additionally, *O'Brien et al.* [2013] postulated that the large improvement in cloud fraction scaling is actually dominated by CAM5's new microphysical parameterizations. However, these results underscore the need to understand the complex relationships between the multitudes of components within parameterization suites that continue to grow in complexity.

In summary, examples illustrating process-coupling issues unique to VR models were presented. The example using MPAS-A-CAM4 demonstrated the relevance of time step/time-scale interaction [*Williamson*, 2013] to the VR framework. The analysis using the closure of the ZM scheme is admittedly idealistic, but illustrative of the subtle differences between two simple scale-aware approaches for representing convection (i.e., *Ma et al.* [2014] and *Fowler et al.* [2016]) and provided some testable hypotheses. In particular, a common practice in VR models to use the same physics time step across all gridpoints (regardless of their size) may not produce the desired effects of scale-aware convection parameterizations, depending on the type of process splitting. Previous studies also noted that scale-awareness in other parameterizations plays a role in determining the overall resolution-sensitivity and coupling to the dynamics. Based on the review and analysis, it is understood that novel approaches for understanding the interaction between resolved dynamics and subgrid parameterizations are required to fully exploit emerging themes in model development. For instance, a wide spectrum of scale-aware parameterizations (cf. section 9) as well as a variety of splitting methods (cf. section 3) should all be analyzed and evaluated in the VR framework. Having Δt dependent on the gridcell size may be one pathway to remove one dimension of the resolution-sensitivity specific to VR models so that scale-aware parameterizations tested for quasi-uniform resolution models would perform equally in VR models, although this could be technically challenging. An alternative pathway is to use a separate, uniform-resolution physics grid (cf. section 10.1) for calculation of tendencies from physics parameterizations. Such a separate physics grid was adopted by *Fox-Rabinovitz et al.* [2001] for their stretched-grid model. Addressing these physics-dynamics and process coupling issues, in addition to other fundamental challenges with respect to the current generation of numerical schemes and physics parameterizations, will be necessary to take full advantage of the capability of VR models.

11 Conclusions and outlook

The wide span of issues related to PDC in geophysical models has been presented. It is apparent that under the surface of all of these remain many unanswered questions.

The current efforts are disperse. The historical review in section 2 has highlighted that research in this area has been ongoing for some time by various groups focusing on individual aspects. However, this research does not always apply in a direct fashion to full models, partly due to the highly complex and cross disciplinary nature of the topic.

Probably the most sensitive parameter is the time step. Section 3 has shown that there is ample evidence in the commonly used models relating to a time step sensitivity. The cause and effect, however, is not as clearly identifiable. As section 4 has shown, convergence of the whole model run with a reduction of the time step is far from trivial to achieve and analyze. This is where the value of reduced models comes in. As has been shown above, there are different ways in which reduction can be achieved. One option is to reduce the equation set, as in section 5, which then renders the generation of a reference solution more straight forward and allows for more rigorous mathematical analysis. Another option, as discussed in section 6, is to reduce the complexity of the GCM. Obviously the balance has to be right. Oversimplification does not challenge the coupling as the real model would, overly complex setups make the analysis intractable.

The topic has then been taken to the next level of complexity when discussing the challenge of coupling two models to each other, as in the Atmosphere-Ocean case. In section 7 this has been described theoretically and reference been made to potential analysis methods.

In the background of all of the modeling and coupling activities naturally exists one fundamental common set of laws. Thermodynamics, the subject of section 8. Naturally one would think that all models should obey the laws of thermodynamics. However, due to different formulations and the at times more pragmatic attitudes towards the representation of

sub grid process and coupling to or coupling of models, the conformance with the laws of thermodynamics may not be as strict as would be desirable.

One could argue that this could be corrected readily. This however overlooks at least one crucial aspect: The wide range of parameters that exists and the models are subjected to. One of these aspects is resolution, for example. Models are run at different resolutions for different purposes or resolution passes through regions where one formulation ceases to be appropriate (parameterization) and the processes are resolved by the fundamental equations of motion (resolved convection). This problem has been discussed and illustrated in section 9.

As well as the sensitivity for the subject, the individual models and modeling approaches evolve with time. As pointed out by the last section, the emergence of variable resolution models and high-order Galerkin methods poses new challenges and yet another level of complexity for the coupling question. For example the gray zone now coexists with regions where the parameterization is well justified and regions where the model resolves processes, in the same model run. Not only is there the question of when in time to evaluate the parameterizations, or how to apply their forcings in time, there is also a choice of what the parameterizations should see, the element averaged values or node integration point values and how their forcings should then be applied.

Perhaps the hardest yet most important is how to analyze model performance in terms of the different schemes and hence determine which scheme is better. In order to achieve this, however, the analysis has to touch every one of the above aspects, in order to be relevant for the final model outcome. This is a challenge which the community has to tackle as a whole. The generation of standardized hierarchies of setups with their respective established reference solutions should aid in the translation from improved components into improved full model runs and enable insightful evaluation of changes made to the coupling.

Overall, decisions have to be made. Decisions demand guidance, objectively and systematically. This is an activity that impacts all of the modeling community, from developers to users. Due to its complexity it also has to be tackled by the community as a whole. The authors hope that this article will seed this development and provide a basis for this decision-making process and that the PDC workshop series (2014 in Ensenada, BC, Mexico, 2016 in Richland, WA, USA and 2018 in Reading, UK (tbc) - refer to `http://pdc.cicese.mx` for the latest information and material from the previous workshops) will provide a platform for this.

References

Akmaev, R. A. (2008), On the energetics of maximum-entropy temperature profiles, *Q. J. R. Meteorol. Soc.*, *134*, 187–197.

Anderson, B. D., S. E. Benzley, and S. J. Owen (2009), Automatic all quadrilateral mesh adaption through refinement and coarsening, *Proceedings of the 18th International Meshing Roundtable, IMR 2009*, pp. 557–574, doi:10.1007/978-3-642-04319-2_32.

Arakawa, A., and W. H. Schubert (1974), Interaction of a cumulus cloud ensemble with the large-scale environment, Part I, *Journal of the Atmospheric Sciences*, *31*(3), 674–701.

Arakawa, A., and C. Wu (2013), A unified representation of deep moist convection in numerical modeling of the atmosphere: Part I, *J. Atmos. Sci.*, *70*, 1977–1992.

Arakawa, A., J.-H. Jung, and C.-M. Wu (2011), Toward unification of the multiscale modeling of the atmosphere, *Atmospheric Chemistry and Physics*, *11*, 3731–3742, doi:10.5194/acp-11-3731-2011.

Bacmeister, J. T., P. H. Lauritzen, A. Dai, and J. E. Truesdale (2012), Assessing possible dynamical effects of condensate in high resolution climate models, *Geophys. Res. Lett.*, *39*, L04,806.

Bacmeister, J. T., M. F. Wehner, R. B. Neale, A. Gettelman, C. Hannay, P. H. Lauritzen, J. M. Caron, and J. E. Truesdale (2014), Exploratory high-resolution climate simulations using the community atmosphere model (cam), *J. Climate*, *27*(9), 3073–3099, doi:10.1175/JCLI-D-13-00387.1.

Balaji, V., J. Anderson, I. Held, M. Winton, S. Malyshev, and R. Stouffer (2007), The fms exchange grid: a mechanism for data exchange between earth system components on independent grids, *Proceedings of the 2005 International Conference on Parallel Computational Fluid Dynamics, College Park, MD, USA, Elsevier*.

Bannon, P. (2002), Theoretical foundation for models of moist convection, *J. Atmos. Sci.*, *59*, 1967–1982.

Bannon, P. R. (2015), Entropy production and climate efficiency, *Journal of the Atmospheric Sciences*, *72*(8), 3268–3280, doi:10.1175/JAS-D-14-0361.1.

Bannon, P. R., and S. Lee (2017), Toward quantifying the climate heat engine: Solar absorption and terrestrial emission temperatures and material entropy production, *Journal of the Atmospheric Sciences*, *74*(6), 1721–1734, doi:10.1175/JAS-D-16-0240.1.

Barker, H., R. Pincus, and J.-J. Morcrette (2002), The monte carlo independent column approximation: Application within large-scale models, in *Proceedings of the GCSS Workshop, 20-24 May 2002*, edited by Kananaskis, Alberta, Cananda, available at http://www.met.utah.edu/skrueger/gcss-2002/Extended-Abstracts.pdf.

Beare, R. J., and M. J. P. Cullen (2013), Diagnosis of boundary-layer circulations, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(1991), doi:10.1098/rsta.2011.0474.

Beare, R. J., and M. J. P. Cullen (2016), Validating weather and climate models at small Rossby numbers: including a boundary layer, *Quarterly Journal of the Royal Meteorological Society*, *142*(700), 2636–2645, doi:10.1002/qj.2852.

Bechtold, P., N. Semane, P. Lopez, J.-P., Chaboureau, A. Beljaars, and N. Bormann (2014), Representing equilibrium and non-equilibrium convection in large-scale models., *J. Atmos. Sci.*

Beljaars, A., P. Bechtold, M. Köhler, J. J. Morcrette, A. Tompkins, P. Viterbo, and N. Wedi (2004), The numerics of physical parameterization, in *Proc. ECMWF Workshop on Recent Developments in numerical methods for atmosphere and ocean modelling*, pp. 113–135, European Centre For Medium-Range Weather Forecasts, 6-10 September 2004, Shinfield Park, Reading, U.K.

Beljaars, A., E. Dutra, G. Balsamo, and F. Lemarié (2017), On the numerical stability of surface–atmosphere coupling in weather and climate models, *Geosci. Model Dev.*, *10*(2), 977–989, doi:10.5194/gmd-10-977-2017.

Best, M. J., A. Beljaars, J. Polcher, and P. Viterbo (2004), A proposed structure for coupling tiled surfaces with the planetary boundary layer, *Journal of Hydrometeorology*, *5*(6), 1271–1278, doi:10.1175/JHM-382.1.

Blackburn, M., and B. J. Hoskins (2013), Context and aims of the Aqua-Planet Experiment, *J. Meteor. Soc. Japan*, *91A*, 1–15.

Blackburn, M., D. L. Williamson, K. Nakajima, W. Ohfuchi, Y. O. Takahashi, Y.-Y. Hayashi, H. Nakamura, M. Ishiwatari, J. McGregor, H. Borth, V. Wirth, H. Frank, P. Bechtold, N. P. Wedi, H. Tomita, M. Satoh, M. Zhao, I. M. Held, M. J. Suarez, M.-I. Lee, M. Watanabe, M. Kimoto, Y. Liu, Z. Wang, A. Molod, K. Rajendran, A. Kitoh, and R. Stratton (2013), The Aqua-Planet Experiment (APE): Control SST simulation, *J. Meteor. Soc. Japan*, *91A*, 17–56.

Bogenschutz, P. A., A. Gettelman, H. Morrison, V. E. Larson, D. P. Schanen, N. R. Meyer, and C. Craig (2012), Unified parameterization of the planetary boundary layer and shallow convection with a higher-order turbulence closure in the Community Atmosphere Model: single-column experiments, *Geoscientific Model Development*, *5*, 1407–1423.

Bogenschutz, P. A., A. Gettelman, H. Morrison, V. E. Larson, C. Craig, and D. P. Schanen (2013), Higher-order turbulence closure and its impact on climate simulations in the Community Atmosphere Model, *J. Climate*, *26*, 9655–9676.

Boutle, I. A., J. E. J. Eyre, and A. P. Lock (2014), Seamless stratocumulus simulation across the turbulent gray zone, *Monthly Weather Review*, *142*(4), 1655–1668.

Brdar, S., M. Baldauf, A. Dedner, and R. Klöfkorn (2013), Comparison of dynamical cores for NWP models: comparison of COSMO and Dune, *Theoretical and Computational Fluid Dynamics*, *27*(3-4), 453–472, doi:10.1007/s00162-012-0264-z.

Bretherton, C. S., and S. Park (2009), A new moist turbulence parameterization in the Community Atmosphere Model, *J. Climate*, *22*, 3422–3448, doi:10.1175/2008JCLI2556.1.

Burchard, H., E. Deleersnijder, and G. Stoyan (2005), Some numerical aspects of turbulence-closure models, in *Marine Turbulence: Theories, Observations and Models*, edited by H. Baumert, J. Simpson, and J. Sündermann, pp. 197–206, Cambridge University Press.

Catry, B., J.-F. Geleyn, M. Tudor, P. Bénard, and A. Trojáková (2007), Flux-conservative thermodynamic equations in a massweighted framework, *Tellus A*, *59*, 71–79.

Caya, A., R. Laprise, and P. Zwack (1998), Consequences of using the splitting method for implementing physical forcings in a semi-implicit semi-Lagrangian model, *Monthly Weather Review*, *126*(6), 1707–1713.

Connors, J., and B. Ganis (2011), Stability of algorithms for a two domain natural convection problem and observed model uncertainty, *Computat. Geosci.*, *15*(3), 509–527, doi:10.1007/s10596-010-9219-x.

Cullen, M. (2007), Modelling atmospheric flows, *Acta Numerica*, *16*, 67–154, doi:10.1017/S0962492906290019.

Cullen, M. J. P. (2006), *A mathematical theory of large-scale atmospheric flow*, p. 259pp., Imperial College Press.

Cullen, M. J. P., and D. J. Salmond (2003), On the use of a predictor-corrector scheme to couple the dynamics with the physical parametrizations in the ECMWF model, *Quarterly Journal of the Royal Meteorological Society*, *129*(589), 1217–1236, doi:10.1256/qj.02.12.

de Groot, S. R., and P. Mazur (1984), *Nonequilibrium thermodynamics*, Dover.

De Troch, R., R. Hamdi, H. Van de Vyver, J.-F. Geleyn, and P. Termonia (2013), Multiscale performance of the ALARO-0 model for simulating extreme summer precipitation climatology in Belgium, *Journal of Climate*, *26*(22), 8895–8915, doi:10.1175/JCLI-D-12-00844.1.

Deleersnijder, E., E. Hanert, H. Burchard, and H. Dijkstra (2008), On the mathematical stability of stratified flow models with local turbulence closure schemes, *Ocean Dyn.*, *58*, 237–246, 10.1007/s10236-008-0145-6.

Dennis, J., J. Edwards, K. J. Evans, O. N. Guba, P. H. Lauritzen, A. A. Mirin, A. St-Cyr, M. A. Taylor, and P. H. Worley (2012a), CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, *Int. J. High Perf. Comput. Appl.*, *26*, 74–89, doi:10.1177/1094342011428142.

Dennis, J. M., J. Edwards, K. J. Evans, O. Guba, P. H. Lauritzen, A. A. Mirin, A. St-Cyr, M. A. Taylor, and P. H. Worley (2012b), CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, *Int. J. High. Perform. C.*, *26*(1), 74–89, doi: 10.1177/1094342011428142.

Dubal, M., N. Wood, and A. Staniforth (2004), Analysis of parallel versus sequential splittings for time-stepping physical parameterizations, *Monthly Weather Review*, *132*(1), 121–132, doi:10.1175/1520-0493(2004)131<0121:AOPVSS>2.0.CO;2.

Dubal, M., N. Wood, and A. Staniforth (2005), Mixed parallel-sequential-split schemes for time-stepping multiple physical parameterizations, *Monthly Weather Review*, *133*(4), 989–1002, doi:10.1175/MWR2893.1.

Dubal, M., N. Wood, and A. Staniforth (2006), Some numerical properties of approaches to physics-dynamics coupling for NWP, *Quarterly Journal of the Royal Meteorological Society*, *132*(614), 27–42, doi:10.1256/qj.05.49.

Durran, D. (2010), *Numerical Methods for Fluid Dynamics: With Applications to Geophysics*, *Texts in Applied Mathematics*, vol. 32, 2 ed., Springer, 516 p.

Foken, T. (2006), 50 years of the monin-obukhov similarity theory, *Boundary-Layer Meteorology*, *119*(3), 431–447, doi:10.1007/s10546-006-9048-6.

Fowler, L. D., W. C. Skamarock, G. A. Grell, S. R. Freitas, and M. G. Duda (2016), Analyzing the Grell-Freitas Convection Scheme from Hydrostatic to Nonhydrostatic Scales within a Global Model, *Monthly Weather Review*, *144*, 2285–2306, doi:10.1175/MWR-D-15-0311.1.

Fox-Rabinovitz, M. S., L. L. Takacs, R. C. Govindaraju, J. Suarez, Max, D. A. Office, G. Space, and G. S. Corporation (2001), A Variable-Resolution Stretched-Grid General Circulation Model : Regional Climate Simulation, *Monthly Weather Review*, *129*, 453–469.

Fox-Rabinovitz, M. S., J. Côté, B. Dugas, M. Déqué, and J. L. McGregor (2006), Variable resolution general circulation models: Stretched-grid model intercomparison project (SG-MIP), *Journal of Geophysical Research*, *111*, D16,104, doi:10.1029/2005JD006520.

Frierson, D. M. W., I. M. Held, and P. Zurita-Gotor (2006), A gray-radiation aquaplanet moist GCM. Part I: static stability and eddy scale, *J. Atmos. Sci.*, *63*, 2548–2566, doi: 10.1175/JAS3753.1.

Gassmann, A., and H.-J. Herzog (2015), How is local material entropy production represented in a numerical model?, *Q. J. R. Meteorol. Soc.*, *141*, 854–869.

Geleyn, J.-F., and P. Marquet (2010), Moist thermodynamics and moist turbulence for modelling at the non-hydrostatic scales, in *ECMWF Workshop on non-hydrostatic modelling*.

Gerard, L. (2015), Bulk mass-flux perturbation formulation for a unified approach of deep convection at high resolution, *Monthly Weather Review*, *143*(10), 4038–4063, doi:10.1175/MWR-D-15-0030.1.

Gerard, L., and J.-F. Geleyn (2005), Evolution of a subgrid deep convection parametrization in a limited-area model with increasing resolution, *Quarterly Journal of the Royal Meteorological Society*, *131*(610), 2293–2312, doi:10.1256/qj.04.72.

Gerard, L., J.-M. Piriou, R. Brozkova, J.-F. Geleyn, and D. Banciu (2009), Cloud and pre-cipitation parametrization in a meso-gamma-scale operational weather prediction model, *Mon. Wea. Rev.*, *137*, 3960–3976.

Gettelman, A., H. Morrison, S. Santos, P. Bogenschutz, and P. M. Caldwell (2015), Advanced two-moment bulk microphysics for global models. part II: Global model solutions and aerosol–cloud interactions, *Journal of Climate*, *28*(3), 1288–1307.

Giraldo, F., and M. Restelli (2008), A study of spectral element and discontinuous Galerkin methods for the Navier-Stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases, *J. Comput. Phys.*, *227*(8), 3849 – 3877, doi:http://dx.doi.org/10.1016/j.jcp.2007.12.009.

Golaz, J.-C., V. E. Larson, and W. R. Cotton (2002a), A PDF-based model for boundary layer clouds. Part I: Method and model description, *J. Atmos. Sci.*, *59*, 3540–3551.

Golaz, J.-C., V. E. Larson, and W. R. Cotton (2002b), A PDF-based model for boundary layer clouds. Part II: Model Results, *J. Atmos. Sci.*, *59*, 3552–3571.

Goody, R. (2000), Sourcea and sinks of climate entropy, *Q. J. R. Meteorol. Soc.*, *126*, 1953–1970.

Grabowski, W. W. (2001), Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP), *J. Atmos. Sci.*, *58*, 978–997.

Grell, G. A., and S. R. Freitas (2014), A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmos. Chem. Phys.*, *14*, 5233–5250.

Gross, M., S. Malardel, C. Jablonowski, and N. Wood (2016), Bridging the (knowledge) gap between physics and dynamics, *Bull. Amer. Soc.*, *97*, 137–142, doi:10.1175/BAMS-D-15-00103.1.

Guba, O., M. a. Taylor, P. a. Ullrich, J. R. Overfelt, and M. N. Levy (2014), The spectral element method (SEM) on variable-resolution grids: evaluating grid sensitivity and resolution-aware numerical viscosity, *Geoscientific Model Development*, *7*(2010), 2803–2816, doi:10.5194/gmd-7-2803-2014.

Gustafson, W. I., P.-L. Ma, H. Xiao, B. Singh, P. J. Rasch, and J. D. Fast (2013), The Separate Physics and Dynamics Experiment (SPADE) framework for determining resolution awareness: A case study of microphysics, *J. Geophys. Res. (Atmos.)*, *118*(16), 9258–9276.

Gustafson, W. I., P.-L. Ma, and B. Singh (2014), Precipitation characteristics of CAM5 physics at mesoscale resolution during MC3E and the impact of convective timescale choice, *Journal of Advances in Modeling Earth Systems*, *6*(4), 1271–1287, doi:10.1002/2014MS000334.

Hagos, S., R. Leung, S. A. Rauscher, and T. Ringler (2013), Errors characteristics of two grid refinement approaches in aqua-planet simulations: MPAS-A and WRF, *Monthly Weather Review*, *141*, 3022–3036, doi:10.1175/MWR-D-12-00338.1.

Harris, L. M., and S.-J. Lin (2013), A two-way nested global-regional dynamical core on the cubed-sphere grid, *Monthly Weather Review*, *141*, 283–306, doi:10.1175/MWR-D-11-00201.1.

Harris, L. M., and S.-J. Lin (2014), Global-to-regional nested grid climate simulations in the GFDL high resolution atmospheric model, *Journal of Climate*, *27*, 4890–4910, doi:10.1175/JCLI-D-13-00596.1.

Harris, L. M., S. J. Lin, and C. Y. Tu (2016), High-resolution climate simulations using GFDL HiRAM with a stretched global grid, *Journal of Climate*, *29*, 4293–4314, doi:10.1175/JCLI-D-15-0389.1.

Held, I. M., and M. J. Suarez (1994), A proposal for the intercomparison of dynamical cores of atmospheric general circulation models, *Bull. Amer. Meteor. Soc.*, *75*, 1825–1830.

Hodyss, D., K. C. Viner, A. Reinecke, and J. A. Hansen (2013), The impact of noisy physics on the stability and accuracy of physics–dynamics coupling, *Monthly Weather Review*, *141*(12), 4470–4486.

Holloway, C. E., S. J. Woolnough, and G. M. S. Lister (2012), Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study, *Q. J. R. Meteorol. Soc.*, *138*, 1692–1708, doi:10.1002/qj.1903.

Holt, M. W. (1990), Semi-geostrophic moist frontogenesis in a Lagrangian model, *Dyn. Atmos. Oceans*, *14*, 463–481.

Honnert, R., and V. Masson (2014), What is the smallest physically acceptable scale for 1D turbulence schemes?, *Frontiers in Earth Science*, *2*(27), doi:10.3389/feart.2014.00027.

Hourdin, F., T. Mauritsen, A. Gettelman, J.-C. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, F. Rauser, C. Rio, L. Tomassini, M. Watanabe, and D. Williamson (2017), The art and science of climate model tuning, *Bulletin of the American Meteorological Society*, *98*(3), 589–602, doi:10.1175/BAMS-D-15-00135.1.

Huang, X., A. M. Rhoades, P. A. Ullrich, and C. M. Zarzycki (2016), An evaluation of the variable-resolution CESM for modeling California's climate, *Journal of Advances in Modeling Earth Systems*, *8*, 345–369, doi:10.1002/2013MS000282.

Jablonowski, C., and D. L. Williamson (2011), The pros and cons of diffusion, filters and fixers in atmospheric general circulation models, in *Numerical Techniques for Global Atmospheric Models*, *Lecture Notes in Computational Science and Engineering*, vol. 80, edited by P. H. Lauritzen, C. Jablonowski, M. A. Taylor, and R. D. Nair, pp. 381–493, Springer.

Ju, L., T. Ringler, and M. Gunzburger (2011), Voronoi tessellations and their application to climate and global modeling, in *Numerical Techniques for Global Atmospheric Models*, *Lecture Notes in Computational Science and Engineering*, vol. 80, edited by P. Lauritzen, C. Jablonowski, M. Taylor, and R. Nair, chap. 10, pp. 313–342, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-11640-7.

Keyes, D. E., et al. (2013), Multiphysics simulations: Challenges and opportunities, *International Journal of High Performance Computing Applications*, *27*(1), 4–83, doi:10.1177/1094342012468181.

Khairoutdinov, M. F., and D. A. Randall (2001), A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results, *Geophysical Research Letters*, *28*(18), 3617–3620, doi:10.1029/2001GL013552.

Klemp, J. B., J. Dudhia, and A. D. Hassiotis (2008), An upper gravity-wave absorbing layer for NWP applications, *Mon. Wea. Rev.*, *136*, 3987–4004.

Klocke, D., R. Pincus, and J. Quaas (2011), On constraining estimates of climate sensitivity with present-day observations through model weighting, *Journal of Climate*, *24*(23), 6092–6099.

Lander, J., and B. J. Hoskins (1997), Believable scales and parameterizations in a spectral transform model, *Monthly Weather Review*, *125*(2), 292–303, doi:10.1175/1520-0493(1997)125<0292:BSAPIA>2.0.CO;2.

Landu, K., L. R. Leung, S. Hagos, V. Vinoj, S. A. Rauscher, T. Ringler, and M. Taylor (2014), The dependence of ITCZ structure on model resolution and dynamical core in aqua-planet simulations, *J. Climate*, *27*, 2375–2385.

Large, W. G. (2006), Surface fluxes for practitioners of global ocean data assimilation, in *An Integrated View of Oceanography: Ocean Weather Forecasting in the 21st Century*, edited by E. Chassignet and J. Verron, Kluwer.

Lauritzen, P. H., and J. Thuburn (2012), Evaluating advection/transport schemes using interrelated tracers, scatter plots and numerical mixing diagnostics, *Quarterly Journal of the Royal Meteorological Society*, *138*(665), 906–918, doi:10.1002/qj.986.

Lauritzen, P. H., J. T. Bacmeister, P. F. Callaghan, and M. A. Taylor (2015a), NCAR global model topography generation software for unstructured grids, *Geoscientific Model Development Discussions*, *8*(6), 4623–4651, doi:10.5194/gmdd-8-4623-2015.

Lauritzen, P. H., A. J. Conley, J.-F. Lamarque, F. Vitt, and M. A. Taylor (2015b), The terminator "toy" chemistry test: a simple tool to assess errors in transport schemes, *Geoscientific Model Development*, *8*(5), 1299–1313, doi:10.5194/gmd-8-1299-2015.

Lee, M.-I., I. Kang, and B. E. Mapes (2003), Impacts of cumulus convection parameterization on aqua-planet AGCM simulations of tropical intraseasonal variability, *J. Meteor. Soc. Jap.*, *81*(5), 963–992.

Lee, M.-I., M. J. Suarez, I.-S. Kang, I. M. Held, and D. Kim (2008), A moist benchmark calculation for atmospheric general circulation models, *J. Climate*, *21*, 4934–4954, doi:

10.1175/2008JCLI1891.1.

Lemarié, F., L. Debreu, and E. Blayo (2013), Toward an optimized global-in-time Schwarz algorithm for diffusion equations with discontinuous and spatially variable coefficients, part 2 : the variable coefficients case, *Electron. Trans. Numer. Anal.*, *40*, 170–186.

Lemarié, F., P. Marchesiello, L. Debreu, and E. Blayo (2014), Sensitivity of Ocean-Atmosphere Coupled Models to the Coupling Method : Example of Tropical Cyclone Erica, *Research Report RR-8651*, INRIA Grenoble.

Lemarié, F., E. Blayo, and L. Debreu (2015), Analysis of ocean-atmosphere coupling algorithms: Consistency and stability, *Procedia Computer Science*, *51*(0), 2066 – 2075, doi: http://dx.doi.org/10.1016/j.procs.2015.05.473, international Conference On Computational Science, {ICCS} 2015 Computational Science at the Gates of Nature.

Li, F., W. Collins, M. Wehner, D. Williamson, J. Olson, and C. Algieri (2011), Impact of horizontal resolution on simulation of precipitation extremes in an aqua-planet version of Community Atmospheric Model (CAM3), *Tellus A*, *63*(5), 884–892.

Lin, S.-J. (2004), A "vertically Lagrangian" finite-volume dynamical core for global models, *Mon. Wea. Rev.*, *132*, 2293–2307.

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith (2000), A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests, *Monthly Weather Review*, *128*, 3187.

Ma, P.-L., P. J. Rasch, J. D. Fast, R. C. Easter, W. I. Gustafson Jr., X. Liu, S. J. Ghan, and B. Singh (2014), Assessing the CAM5 physics suite in the WRF-Chem model: implementation, resolution sensitivity, and a first evaluation for a regional case study, *Geoscientific Model Development*, *7*(3), 755–778, doi:10.5194/gmd-7-755-2014.

Malardel, S., and D. Ricard (2015), An alternative cell-averaged departure point reconstruction for pointwise semi-Lagrangian transport schemes, *Quarterly Journal of the Royal Meteorological Society*, *141*, 2114–2126, doi:10.1002/qj.2509.

Malardel, S., and N. P. Wedi (2016), How does subgrid-scale parametrization influence nonlinear spectral energy fluxes in global NWP models?, *Journal of Geophysical Research: Atmospheres*, *121*(10), 5395–5410, doi:10.1002/2015JD023970, 2015JD023970.

Manabe, S., and K. Bryan (1969), Climate calculations with a combined ocean-atmosphere model, *Journal of the Atmospheric Sciences*, *26*(4), 786–789.

Marquet, P., and J.-F. Geleyn (2013), On a general definition of the squared Brunt–Väisälä frequency associated with the specific moist entropy potential temperature, *Q.J.R. Meteorol. Soc.*, *139*, 85Ű100.

McGregor, J. L. (2013), Recent developments in variable-resolution global climate modelling, *Climatic Change*, *129*(3), 369–380, doi:10.1007/s10584-013-0866-5.

Medeiros, B., B. Stevens, and S. Bony (2015), Using aquaplanets to understand the robust responses of comprehensive climate models to forcing, *Climate Dynamics*, *44*(7-8), 1957–1977.

Medeiros, B., D. L. Williamson, and J. G. Olson (2016), Reference aquaplanet climate in the Community Atmosphere Model, version 5, *Journal of Advances in Modeling Earth Systems*, *8*(1), 406–424.

Medvigy, D., R. L. Walko, and R. Avissar (2008), Modeling interannual variability of the Amazon hydroclimate, *Geophysical Research Letters*, *35*, L15,817, doi:10.1029/2008GL034941.

Medvigy, D., R. L. Walko, M. J. Otte, and R. Avissar (2010), The Ocean–Land–Atmosphere Model: Optimization and Evaluation of Simulated Radiative Fluxes and Precipitation, *Monthly Weather Review*, *138*, 1923–1939, doi:10.1175/2009MWR3131.1.

Medvigy, D., R. L. Walko, and R. Avissar (2011), Effects of deforestation on spatiotemporal distributions of precipitation in South America, *Journal of Climate*, *24*, 2147–2163, doi:10.1175/2010JCLI3882.1.

Medvigy, D., R. L. Walko, M. J. Otte, and R. Avissar (2013), Simulated changes in northwest U.S. climate in response to Amazon deforestation, *Journal of Climate*, *26*, 9115–9136, doi:10.1175/JCLI-D-12-00775.1.

Meutter, P. D., L. Gerard, G. Smet, K. Hamid, R. Hamdi, D. Degrauwe, and P. Termonia (2015), Predicting small-scale, short-lived downbursts: case study with the NWP limited-area ALARO model for the Pukkelpop thunderstorm, *Mon. Wea. Rev.*, *143*, 742–756, doi:10.1175/MWR-D-14-00290.1.

Mishra, S. K., and S. Sahany (2013), Effects of time step size on the simulation of tropical climate in NCAR-CAM3, *Climate Dynamics*, *37*, 689–704, doi:10.1007/s00382-011-0994-4.

Mishra, S. K., J. Srinivasan, and R. S. Nanjundiah (2008), The impact of the time step on the intensity of ITCZ in an aquaplanet GCM, *Mon. Wea. Rev.*, *136*, 4077–4091.

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

Molod, A. (2009), Running GCM physics and dynamics on different grids: algorithm and tests, *Tellus A*, *61*(3), 381–393, doi:10.1111/j.1600-0870.2009.00394.x.

Müller, A. (2014), Does high order and dynamic adaptative refinement improve the efficiency of atmospheric simulation?, PDC14, CICESE, Ensenada, Mexico.

Nair, R., H.-W. Choi, and H. Tufo (2009), Computational aspects of a scalable high-order discontinuous Galerkin atmospheric dynamical core, *Computers & Fluids*, *38*(2), 309 – 319, doi:http://dx.doi.org/10.1016/j.compfluid.2008.04.006.

Nair, R. D., M. N. Levy, and P. H. Lauritzen (2011), Emerging numerical methods for atmospheric modeling, in: P.H. Lauritzen, R.D. Nair, C. Jablonowski, M. Taylor (Eds.), Numerical techniques for global atmospheric models, *Lecture Notes in Computational Science and Engineering, Springer*, *80*, 251–311.

Neale, R. B., and B. J. Hoskins (2000), A standard test for AGCMs and their physical parameterizations. I: The proposal, *Atmos. Sci. Letters*, *1*, 101–107.

Neale, R. B., J. H. Richter, and M. Jochum (2008), The impact of convection on ENSO: From a delayed oscillator to a series of events, *J. Climate*, *21*, 5904–5924, doi:10.1175/2008JCLI2244.1.

Neale, R. B., C.-C. Chen, A. Gettelman, P. H. Lauritzen, S. Park, D. L. Williamson, A. J. Conley, R. Garcia, D. Kinnison, J.-F. Lamarque, D. Marsh, M. Mills, A. K. Smith, S. Tilmes, F. Vitt, H. Morrison, P. Cameron-Smith, W. D. Collins, M. J. Iacono, R. C. Easter, S. J. Ghan, X. Liu, P. J. Rasch, and M. A. Taylor (2010a), Description of the NCAR Community Atmosphere Model (CAM 5.0), *NCAR Technical Note NCAR/TN-486+STR*, National Center for Atmospheric Research, Boulder, Colorado, 268 pp., available from http://www.cesm.ucar.edu/models/cesm1.0/cam/.

Neale, R. B., J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, P. J. Rasch, S. J. Vavrus, M. A. Taylor, W. D. Collins, M. Zhang, and S.-J. Lin (2010b), Description of the NCAR Community Atmosphere Model (CAM 4.0), *NCAR Technical Note NCAR/TN-485+STR*, National Center for Atmospheric Research, Boulder, Colorado, 212 pp., available from http://www.cesm.ucar.edu/models/cesm1.0/cam/.

O'Brien, T. A., F. Li, W. D. Collins, S. A. Rauscher, T. D. Ringler, M. A. Taylor, S. M. Hagos, and R. L. Leung (2013), Observed scaling in clouds and precipitation and scale incognizance in regional to global atmospheric models, *Journal of Climate*, *26*, 9313–9333, doi:10.1175/JCLI-D-13-00005.1.

Palmer, T. N., A. Döring, and G. Seregin (2014), The real butterfly effect, *Nonlinearity*, *27*, R123, doi:10.1088/0951-7715/27/9/R123.

Park, S. (2014), A Unified Convection Scheme (UNICON). Part I: Formulation, *Journal of Atmospheric Sciences*, *71*, 3902–3930, doi:10.1175/JAS-D-13-0233.1.

Park, S., and C. S. Bretherton (2009), The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model, *J. Climate*, *22*, 3449–3469, doi:10.1175/2008JCLI2557.1.

Park, S.-H., W. C. Skamarock, J. B. Klemp, L. D. Fowler, and M. G. Duda (2013), Evaluation of global atmospheric solvers using extensions of the Jablonowski and Williamson baroclinic wave test case, *Monthly Weather Review*, *141*(9), 3116–3129, doi:10.1175/

MWR-D-12-00096.1.

Park, S.-H., J. B. Klemp, and W. C. Skamarock (2014), A Comparison of mesh refinement in the global MPAS-A and WRF models using an idealized normal-mode baroclinic wave simulation, *Monthly Weather Review*, *142*, 3614–3634, doi:10.1175/MWR-D-14-00004.1.

Pauluis, O., and I. M. Held (2002), Entropy budget of an atmosphere in radiative-convective equilibrium. Part I: Maximum work and frictional dissipation, *J. Atmos. Sci.*, *59*, 125–139.

Pincus, R., H. W. Barker, and J.-J. Morcrette (2003), A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, *Journal of Geophysical Research: Atmospheres*, *108*(D13), doi:10.1029/2002JD003322, 4376.

Piriou, J.-M., J.-L. Redelsperger, J.-F. Geleyn, J.-P. Lafore, and F. Guichard (2007), An approach for convective parameterization with memory, in separating microphysics and transport in grid-scale equations, *J. Atmos. Sci.*, *64*, 4127–4139.

Polcher, J., B. McAvaney, P. Viterbo, M.-A. Gaertner, A. Hahmann, J.-F. Mahfouf, J. Noilhan, T. Phillips, A. Pitman, C. Schlosser, J.-P. Schulz, B. Timbal, D. Verseghy, and Y. Xue (1998), A proposal for a general interface between land surface schemes and general circulation models, *Global and Planetary Change*, *19*(1âĂŞ4), 261 – 276, doi: https://doi.org/10.1016/S0921-8181(98)00052-6.

Rajendran, K., A. Kitoh, and J. Srinivasan (2013), Effect of SST variation on ITCZ in APE simulations, *J. Meteor. Soc. Japan*, *91A*, 195–215.

Rauscher, S. A., and T. D. Ringler (2014), Impact of variable-resolution meshes on midlatitude baroclinic eddies using CAM-MPAS-A, *Monthly Weather Review*, *142*, 4256–4268, doi:10.1175/MWR-D-13-00366.1.

Rauscher, S. A., T. D. Ringler, W. C. Skamarock, and A. A. Mirin (2013), Exploring a global multiresolution modeling approach using aquaplanet simulations, *Journal of Climate*, *26*(8), 2432–2452, doi:10.1175/JCLI-D-12-00154.1.

Reed, K. A., and C. Jablonowski (2012), Idealized tropical cyclone simulations of intermediate complexity: A test case for AGCMs, *Journal of Advances in Modeling Earth Systems*, *4*(2), doi:10.1029/2011MS000099, M04001.

Reed, K. A., C. Jablonowski, and M. A. Taylor (2012), Tropical cyclones in the spectral element configuration of the Community Atmosphere Model, *Atmospheric Science Letters*, *13*, 303–310, doi:10.1002/asl.399.

Rhoades, A., X. Huang, P. Ullrich, and C. M. Zarzycki (2016), Characterizing Sierra Nevada snowpack using variable-resolution CESM, *Journal of Applied Meteorology and Climatology*, *55*, 173–196, doi:10.1175/JAMC-D-15-0156.1.

Richter, J. H., and P. J. Rasch (2008), Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3, *J. Climate*, *21*, 1487–1499, doi:10.1175/2007JCLI1789.1.

Ringler, T. D., J. Thuburn, J. Klemp, and W. Skamarock (2010), A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids, *Journal of Computational Physics*, *229*, 3065–3090, doi:10.1016/j.jcp.2009.12.007.

Ringler, T. D., D. Jacobsen, M. Gunzburger, L. Ju, M. Duda, and W. Skamarock (2011), Exploring a multiresolution modeling approach within the shallow-water equations, *Monthly Weather Review*, *139*(11), 3348–3368, doi:10.1175/MWR-D-10-05049.1.

Roeckner, E., G. Bäuml, L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, I. Kirchner, L. Kornblueh, E. Manzini, A. Rhodin, U. Schlese, U. Schulzweida, and A. Tompkins (2003), The atmospheric general circulation model ECHAM5. PART I: model description., *Technical Report 349*, Max Planck Institute for Meteorology, Hamburg, Germany.

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

Romps, D. M. (2008), The dry entropy budget of a moist atmosphere, *J. Atmos. Sci.*, *65*, 3779–3799.

Ryder, J., J. Polcher, P. Peylin, C. Ottlé, Y. Chen, E. van Gorsel, V. Haverd, M. J. McGrath, K. Naudts, J. Otto, A. Valade, and S. Luyssaert (2016), A multi-layer land surface energy budget model for implicit coupling with global atmospheric simulations, *Geosci. Model Dev.*, *9*(1), 223–245, doi:10.5194/gmd-9-223-2016.

Sakaguchi, K., L. R. Leung, C. Zhao, Q. Yang, J. Lu, S. Hagos, S. A. Rauscher, L. Dong, T. D. Ringler, and P. H. Lauritzen (2015), Exploring a multi-resolution approach using AMIP simulations, *Journal of Climate*, doi:10.1175/JCLI-D-14-00729.1.

Sakaguchi, K., J. Lu, L. R. Leung, C. Zhao, Y. Li, and S. Hagos (2016), Sources and pathways of the upscale effects on the Southern Hemisphere jet in MPAS-CAM4 variable-Resolution simulations, *Journal of Advances in Modeling Earth Systems*, *accepted*.

Schluter, J. U., and H. Pitsch (2005), Antialiasing filters for coupled Reynolds-Averaged/Large-Eddy Simulations, *AIAA Journal*, *43*(3), 608–615.

Schmidt, F. (1977), Variable fine mesh in spectral global models, *Beiträge zur Physik der Atmosphäre*, *50*, 211–217.

Schmidt, G. A., C. M. Bitz, U. Mikolajewicz, and L.-B. Tremblay (2004), Ice-ocean boundary conditions for coupled models, *Ocean Model.*, *7*(1âĂŞ2), 59 – 74, doi:https://doi.org/10.1016/S1463-5003(03)00030-1.

Schulz, J.-P., L. Dümenil, and J. Polcher (2001), On the land surface-atmosphere coupling and its impact in a single-column atmospheric model, *Journal of Applied Meteorology*, *40*(3), 642–663, doi:10.1175/1520-0450(2001)040<0642:OTLSAC>2.0.CO;2.

Shibuya, R., H. Miura, and K. Sato (2016), A grid transformation method for a quasi-uniform, circular fine region using the spring dynamics, *Journal of the Meteorological Society of Japan*, *94*(5), 443–452, doi:http://doi.org/10.2151/jmsj.2016-022.

Siebesma, A. P. (2015), First workshop of the grey zone project, *GEWEX news*, *25*, 7–9.

Siebesma, A. P., P. M. M. Soares, and J. Teixeira (2007), A Combined Eddy-Diffusivity Mass-Flux Approach for the Convective Boundary Layer, *Journal of Atmospheric Sciences*, *64*, 1230, doi:10.1175/JAS3888.1.

Skamarock, W. C. (2011), Kinetic energy spectra and model filters, in *Numerical Techniques for Global Atmospheric Models*, *Lecture Notes in Computational Science and Engineering*, vol. 80, edited by P. Lauritzen, C. Jablonowski, M. Taylor, and R. Nair, chap. 14, pp. 495–512, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-11640-7.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler (2012), A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering, *Monthly Weather Review*, *140*(9), 3090–3105, doi:10.1175/MWR-D-11-00215.1.

Small, R. J., J. Bacmeister, D. Bailey, A. Baker, S. Bishop, F. Bryan, J. Caron, J. Dennis, P. Gent, H.-m. Hsu, M. Jochum, D. Lawrence, E. Muñoz, P. diNezio, T. Scheitlin, R. Tomas, J. Tribbia, Y.-h. Tseng, and M. Vertenstein (2014), A new synoptic scale resolving global climate simulation using the Community Earth System Model, *Journal of Advances in Modeling Earth Systems*, *6*(4), 1065–1094, doi:10.1002/2014MS000363.

Smith, R. N. B. (1990), A scheme for predicting layer clouds and their water content in a general circulation model, *Q. J. R. Meteorol. Soc.*, *116*, 435–46.

Squires, P. (1952), The growth of cloud drops by condensation, . *Aust. J. Sci. Res.*, *5*, 66–86.

Staniforth, A., and J. Thuburn (2012), Horizontal grids for global weather and climate prediction models: a review, *Quarterly Journal of the Royal Meteorological Society*, *138*(662), 1–26, doi:10.1002/qj.958.

Staniforth, A., N. Wood, and J. Côté (2002a), Analysis of the numerics of physics-dynamics coupling, *Quarterly Journal of the Royal Meteorological Society*, *128*(586), 2779–2799, doi:10.1256/qj.02.25.

Staniforth, A., N. Wood, and J. Côté (2002b), A simple comparison of four physics dynamics coupling schemes, *Monthly Weather Review*, *130*(12), 3129–3135, doi:10.1175/1520-0493(2002)130<3129:ASCOFP>2.0.CO;2.

Staniforth, A. N., and H. L. Mitchell (1978), A variable-resolution finite-element technique for regional forecasting with the primitive equations, *Monthly Weather Review*, *106*, 439–447.

Taylor, M., J. Edwards, and A. St-Cyr (2008), Petascale atmospheric models for the community climate system model: new developments and evaluation of scalable dynamical cores, *J. Phys.: Conf. Ser.*, *125*, doi:10.1088/1742-6596/125/1/012023.

Taylor, M. A., and A. Fournier (2010), A compatible and conservative spectral element method on unstructured grids, *J. Comp. Phys.*, *229*(17), 5879 – 5895, doi:10.1016/j.jcp. 2010.04.008.

Teixeira, J., C. A. Reynolds, and K. Judd (2007), Time step sensitivity of nonlinear atmospheric models: Numerical convergence, truncation error growth, and ensemble design, *J. Atmos. Sci.*, *64*, 175–189, doi:10.1175/JAS3824.1.

Thatcher, D. R., and C. Jablonowski (2016), A moist aquaplanet variant of the Held-Suarez test for atmospheric model dynamical cores, *Geoscientific Model Development*, *9*, 1263–1292, doi:10.5194/gmd-9-1263-2016.

Thayer-Calder, K., A. Gettelman, C. Craig, S. Goldhaber, P. A. Bogenschutz, C.-C. Chen, H. Morrison, J. Höft, E. Raut, B. M. Griffin, J. K. Weber, V. E. Larson, M. C. Wyant, M. Wang, Z. Guo, and S. J. Ghan (2015), A unified parameterization of clouds and turbulence using CLUBB and subcolumns in the Community Atmosphere Model, *Geoscientific Model Development*, *8*(12), 3801–3821, doi:10.5194/gmd-8-3801-2015.

Tomita, H. (2008), A stretched icosahedral grid by a new grid transformation, *Journal of the Meteorological Society of Japan*, *86A*, 107–119.

Tomita, H., M. Satoh, and K. Goto (2002), An Optimization of the Icosahedral Grid Modified by Spring Dynamics, *Journal of Computational Physics*, *183*(1), 307–331, doi: 10.1006/jcph.2002.7193.

Uchida, J., M. Mori, H. Nakamura, M. Satoh, K. Suzuki, and T. Nakajima (2016), Error and energy budget analysis of a nonhydrostatic stretched-grid global atmospheric model, *Monthly Weather Review*, *144*(4), 1423–1447, doi:10.1175/MWR-D-15-0271.1.

Ullrich, P. a., and C. Jablonowski (2011), An analysis of 1D finite-volume methods for geophysical problems on refined grids, *Journal of Computational Physics*, *230*(3), 706–725, doi:10.1016/j.jcp.2010.10.014.

Ullrich, P. A., and M. A. Taylor (2015), Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part I, *Mon. Wea. Rev.*, *143*, 2419–2440, doi:10.1175/ MWR-D-14-00343.1.

Valcke, S., V. Balaji, A. Craig, C. DeLuca, R. Dunlap, R. W. Ford, R. Jacob, J. Larson, R. O'Kuinghttons, G. D. Riley, and M. Vertenstein (2012), Coupling technologies for earth system modelling, *Geosci. Model Dev.*, *5*(6), 1589–1596, doi:10.5194/ gmd-5-1589-2012.

Wacker, U., and F. Herbert (2003), Continuity equations as expressions for local balances of masses in cloudy air, *Tellus A*, *55*, 247–254.

Wacker, U., T. Frisius, and F. Herbert (2006), Evaporation and precipitation surface effects in local mass continuity laws of moist air, *J. Atmos. Sci.*, *63*, 2642–2652.

Walko, R. L., and R. Avissar (2011), A direct method for constructing refined regions in unstructured conforming triangular–hexagonal computational grids: Application to OLAM, *Monthly Weather Review*, *139*(12), 3923–3937, doi:10.1175/MWR-D-11-00021.1.

Wan, H., P. J. Rasch, K. Zhang, J. Kazil, and L. R. Leung (2013), Numerical issues associated with compensating and competing processes in climate models: an example from ECHAM-HAM, *Geoscientific Model Development*, *6*(3), 861–874, doi:10.5194/ gmd-6-861-2013.

Wan, H., P. J. Rasch, K. Zhang, Y. Qian, H. Yan, and C. Zhao (2014), Short ensembles: an efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models, *Geosci. Model Dev.*, *7*(2), 1961–1977, doi:10.5194/gmd-7-1961-2014.

Wan, H., P. J. Rasch, M. A. Taylor, and C. Jablonowski (2015), Short-term time step convergence in a climate model, *Journal of Advances in Modeling Earth Systems*, *7*(1), 215–225,

doi:10.1002/2014MS000368.

Wang, W., C. Bruyère, M. Duda, J. Dudhia, D. Gill, M. Kavulich, K. Keene, H.-C. Lin, J. Michalakes, S. Rizvi, X. Zhang, J. Berner, and K. Fossell (2016), ARW version 3 modeling system user's guide, *Tech. Rep. January*, NCAR, Boulder, Colorado, doi: 10.5065/D68S4MVH.

Wang, Y., L. R. Leung, J. L. McGregor, D.-K. Lee, W.-C. Wang, Y. Ding, and F. Kimura (2004), Regional climate modeling: Progress, challenges, and prospects, *Journal of the Meteorological Society of Japan*, *82*(6), 1599–1628, doi:10.2151/jmsj.82.1599.

Wedi, N. (1999), *The numerical coupling of the physical parameterizations to the "dynamical" equations in a forecast model*, Technical Memorandum 274, ECMWF, Reading, U.K.

Williamson, D. L. (1999), Convergence of atmospheric simulations with increasing horizontal resolution and fixed forcing scales, *Tellus A*, *51*(5), 663–673, doi:10.1034/j.1600-0870.1999.00009.x.

Williamson, D. L. (2002), Time-split versus process-split coupling of parameterizations and dynamical core, *Monthly Weather Review*, *130*(8), 2024–2041.

Williamson, D. L. (2007), The evolution of dynamical cores for global atmospheric models, *Journal of the Meteorological Society of Japan. Ser. II*, *85B*, 241–269, doi:10.2151/jmsj.85B.241.

Williamson, D. L. (2008), Convergence of aqua-planet simulations with increasing resolution in the Community Atmospheric Model, Version 3, *Tellus A*, *60*(5), 848–862, doi:10.1111/j.1600-0870.2008.00339.x.

Williamson, D. L. (2013), The effect of time steps and time-scales on parametrization suites, *Quarterly Journal of the Royal Meteorological Society*, *139*(671), 548–560, doi:10.1002/qj.1992.

Williamson, D. L., and J. G. Olson (2003), Dependence of aqua-planet simulations on time step, *Quarterly Journal of the Royal Meteorological Society*, *129*(591), 2049–2064, doi:10.1256/qj.02.62.

Wood, N., M. Diamantakis, and A. Staniforth (2007), A monotonically damping, second-order accurate, unconditionally stable, numerical scheme for diffusion, *Q. J. R. Meteorol. Soc.*, *133*, 1559–1573.

Wyngaard, J. C. (2004), Toward numerical modeling in the "terra incognita", *Journal of the Atmospheric Sciences*, *61*(14), 1816–1826.

Xu, K.-M., and D. A. Randall (1996), A semi-empirical cloudiness parameterization for use in climate models, *J. Atmos. Sci.*, *53*, 3084–3102.

Zarzycki, C. M., and C. Jablonowski (2014), A multidecadal simulation of Atlantic tropical cyclones using a variable-resolution global atmospheric general circulation model, *Journal of Advances in Modeling Earth Systems*, *6*, doi:10.1002/2014MS000352.

Zarzycki, C. M., and C. Jablonowski (2015), Experimental tropical cyclone forecasts Using a variable-resolution global model, *Monthly Weather Review*, *143*, 4012–4037, doi:10.1175/MWR-D-15-0159.1.

Zarzycki, C. M., M. N. Levy, C. Jablonowski, J. R. Overfelt, M. A. Taylor, and P. A. Ullrich (2014a), Aquaplanet experiments using CAM's variable-resolution dynamical core, *J. Climate*, *27*(14), 5481–5503, doi:10.1175/JCLI-D-14-00004.1.

Zarzycki, C. M., C. Jablonowski, and M. a. Taylor (2014b), Using variable-resolution meshes to model tropical cyclones in the Community Atmosphere Model, *Monthly Weather Review*, *142*(3), 1221–1239, doi:10.1175/MWR-D-13-00179.1.

Zarzycki, C. M., C. Jablonowski, D. R. Thatcher, and M. A. Taylor (2015), Effects of localized grid refinement on the general circulation and climatology in the Community Atmosphere Model, *Journal of Climate*, *28*(7), 2777—2803, doi:10.1175/JCLI-D-14-00599.1.

Zdunkowski, W., and A. Bott (2003), *Dynamics of the Atmosphere: A course in Theoretical Meteorology*, Cambridge University Press.

Zhang, G. J., and N. A. McFarlane (1995), Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre General Circulation Model, *Atmos.–Ocean*, *33*(3), 407–446, doi:10.1080/07055900.1995.9649539.

Zhang, K., D. O'Donnell, J. Kazil, P. Stier, S. Kinne, U. Lohmann, S. Ferrachat, B. Croft, J. Quaas, H. Wan, S. Rast, and J. Feichter (2012), The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations, *"Atmos. Chem. Phys."*, *12*(3), 8911–8949, doi:10.5194/acp-12-8911-2012.

Zhao, C., L. R. Leung, S.-H. Park, S. Hagos, J. Lu, K. Sakaguchi, J.-H. Yoon, B. E. Harrop, W. C. Skamarock, and M. G. Duda (2016), Exploring the impacts of physics and resolution on aqua-planet simulations from a non-hydrostatic global variable-resolution modeling framework, *Journal of Advances in Modeling Earth Systems*, *8*, doi:doi:10.1002/2016MS000727.