AGU PUBLICATIONS

Geophysical Research Letters

FRONTIER ARTICLE
10.1002/2017GL076101

Key Points:
- Earth system models (ESMs) and their parameterization schemes can be radically improved by data assimilation and machine learning
- ESMs can integrate and learn from global observations from space and from local high-resolution simulations
- Ensemble Kalman inversion and Markov chain Monte Carlo methods show promise as learning algorithms for ESMs

Correspondence to:
T. Schneider,
tapio@caltech.edu

Citation:
Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*, 12,396–12,417. https://doi.org/10.1002/2017GL076101

Received 30 AUG 2017
Accepted 23 NOV 2017
Accepted article online 30 NOV 2017
Published online 29 DEC 2017

©2017. American Geophysical Union. All Rights Reserved.

Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations

Tapio Schneider[1,2], Shiwei Lan[1], Andrew Stuart[1], and João Teixeira[2]

[1]California Institute of Technology, Pasadena, CA, USA, [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Abstract Climate projections continue to be marred by large uncertainties, which originate in processes that need to be parameterized, such as clouds, convection, and ecosystems. But rapid progress is now within reach. New computational tools and methods from data assimilation and machine learning make it possible to integrate global observations and local high-resolution simulations in an Earth system model (ESM) that systematically learns from both and quantifies uncertainties. Here we propose a blueprint for such an ESM. We outline how parameterization schemes can learn from global observations and targeted high-resolution simulations, for example, of clouds and convection, through matching low-order statistics between ESMs, observations, and high-resolution simulations. We illustrate learning algorithms for ESMs with a simple dynamical system that shares characteristics of the climate system; and we discuss the opportunities the proposed framework presents and the challenges that remain to realize it.

1. Introduction

Climate models are built around models of the atmosphere, which are based on the laws of thermodynamics and on Newton's laws of motion for air as a fluid. Since they were first developed in the 1960s (Kasahara & Washington, 1967, Manabe et al., 1965; Mintz, 1965; Smagorinsky, 1963, 1965), they have evolved from atmosphere-only models, via coupled atmosphere-ocean models with dynamic oceans, to Earth system models (ESMs) with dynamic cryospheres and biogeochemical cycles (Bretherton et al., 2012; Intergovernmental Panel on Climate Change, 2013). Atmosphere and ocean models compute approximate numerical solutions to the laws of fluid dynamics and thermodynamics on a computational grid. For the atmosphere, the computational grid currently consists of $O(10^7)$ cells, spaced $O(10\text{ km})$–$O(100\text{ km})$ apart in the horizontal; for the oceans, the grid consists of $O(10^8)$ cells, spaced $O(10\text{ km})$ apart in the horizontal. But scales smaller than the mesh size of a climate model cannot be resolved yet are essential for its predictive capabilities. The unresolved scales are modeled by a variety of semiempirical parameterization schemes, which represent the dynamics on subgrid scales as parametric functions of the resolved dynamics on the computational grid (Stensrud, 2007). For example, the dynamical scales of stratocumulus clouds, the most common type of boundary layer clouds, are $O(10\text{ m})$ and smaller, which will remain unresolvable on the computational grid of global atmosphere models for the foreseeable future (Wood, 2012; Schneider et al., 2017). Similarly, the submesoscale dynamics of oceans that may be important for biological processes near the surface have length scales of $O(100\text{ m})$, which will also remain unresolvable for the foreseeable future (Fox-Kemper et al., 2014). Such smaller-scale dynamics in the atmosphere and oceans must be represented in climate models through parameterization schemes. Additionally, ESMs contain parameterization schemes for many processes for which the governing equations are not known or are only poorly known, for example, ecological or biogeochemical processes.

All of these parameterization schemes contain parameters that are uncertain, and the structure of the equations underlying them is uncertain itself. That is, there is parametric and structural uncertainty (Draper, 1995). For example, entrainment and detrainment rates are parameters or parametric functions of state variables such as the vertical velocity of updrafts. They control the interaction of convective clouds with their environment and affect cloud properties and climate. But how they depend on state variables is uncertain, as is the structure of the closure equations in which they appear (e.g., de Rooy et al., 2013; Holloway & Nee, 2009; Neelin et al., 2009; Nie & Kuang, 2012; Romps & Kuang, 2010; Stainforth et al., 2005). Or, as another example, the residence times of carbon in different reservoirs (e.g., soil, litter, and plants) control how rapidly and where

in the biosphere carbon accumulates. They affect the climate response of the biosphere. But they are likewise uncertain, differing by $O(1)$ factors among models (Bloom et al., 2016, Friend et al., 2014; Friedlingstein et al., 2006, 2014). Typically, parameterization schemes are developed and parameters in them are estimated independently of the model into which they are eventually incorporated. They are tested with observations from field studies at a relatively small number of locations. For processes such as boundary layer turbulence that are computable if sufficiently high resolution is available, parameterization schemes are increasingly also tested with data generated computationally in local process studies with high-resolution models (e.g., Jakob, 2003, 2010). After the parameterization schemes are developed and incorporated in a climate model or ESM, modelers adjust ("tune") parameters to satisfy large-scale physical constraints, such as a closed energy balance at the top of the atmosphere (TOA), or selected observational constraints, such as reproduction of the twentieth century global-mean surface temperature record. This model tuning process currently relies on knowledge and intuition of the modelers about plausible ranges of the tunable parameters and about the effect of parameter changes on the simulated climate of a model (Flato et al., 2013, Golaz et al., 2013; Hourdin et al., 2013, 2017; Mauritsen et al., 2012; Randall & Wielicki, 1997). But because of the nonlinear and interacting multiscale nature of the climate system, the simulated climate can depend sensitively and in unexpected ways on settings of tunable parameters (e.g., Suzuki et al., 2013; Zhao et al., 2016). It also remains unclear to what extent the resulting parameter choice is optimal, or how uncertain it is. Moreover, typically only a minute fraction of the available observations is used in the tuning process, usually only highly aggregated data such as global or large-scale mean values accumulated over periods of years or more. In part, this may be done to avoid overfitting, but more importantly, it is done because the tuning process usually involves parameter adjustments by hand, which each must be evaluated by a forward integration of the model. This makes the tuning process tedious and precludes adjustments of a larger set of parameters to fit more complex observational data sets or a wider range of high-resolution process simulations. It also precludes quantification of uncertainties (Hourdin et al., 2017; Schirber et al., 2013).

Climate models have improved over the past decades, leading, for example, to better simulations of El Niño, storm tracks, and tropical waves (Flato et al., 2013; Guilyardi et al., 2009; Hung et al., 2013). Weather prediction models, the higher-resolution siblings of the climate models' atmospheric component, have undergone a parallel evolution. Along with data assimilation techniques for the initialization of weather forecasts, this has led to great strides in the accuracy of weather forecasts (Bauer et al., 2015). But the accuracy of climate projections has not improved as much, and unacceptably large uncertainties remain. For example, if one asks how high CO_2 concentrations can rise before Earth's surface will have warmed 2°C above preindustrial temperatures—the warming target of the 2015 Paris Agreement, of which about 1°C remains because about 1°C has already been realized—the answers range from 480 to 600 ppm across current climate models (Schneider et al., 2017). A CO_2 concentration of 480 ppm will be reached in the late 2030s or early 2040s; 600 ppm may not be reached before 2060 even if CO_2 emissions continue to increase rapidly. Between these extremes lie vastly different optimal policy responses and socioeconomic costs of climate change (Hope, 2015).

These large and long-standing uncertainties in climate projections have their root in uncertainties in parameterization schemes. Parameterizations of clouds dominate the uncertainties in physical processes (Brient & Schneider, 2016, Bony et al., 2006; Cess et al., 1989, 1990; Soden & Held, 2006; Stephens, 2005; Vial et al., 2013; Webb et al., 2013). There are uncertainties both in the representation of the turbulent dynamics of clouds and in the representation of their microphysics, which control, for example, the distribution of droplet sizes in a cloud, the fraction of cloud condensate that precipitates out, and the phase partitioning of cloud condensate into liquid and ice (e.g., Bodas-Salcedo et al., 2014; Golaz et al., 2013; Jiang et al., 2012; Kay et al., 2016; Stainforth et al., 2005; Suzuki et al., 2013; Zhao et al., 2016). Additionally, there are numerous other parameterized processes that contribute to uncertainties in climate projections. For example, it is not precisely known what fraction of the CO_2 that is emitted by human activities will remain in the atmosphere, and so it is uncertain which emission pathways will lead to a given atmospheric CO_2 concentration target (Friedlingstein, 2015; Knutti et al., 2008; Meinshausen et al., 2009). Currently, only about half the emitted CO_2 accumulates in the atmosphere. The other half is taken up by oceans and on land. It is unclear in particular what fraction of the emitted CO_2 terrestrial ecosystems will take up in the future (Canadell et al., 2007, Friend et al., 2014; Friedlingstein et al., 2006, 2014; Knorr, 2009; Le Queéré et al., 2013; Todd-Brown et al., 2013). Reducing such uncertainties through the traditional approach to developing and improving parameterization schemes—attempting to develop one "correct" global parameterization scheme for each process

in isolation, on the basis of observational or computational process studies that are usually focused on specific regions—has met only limited success (Jakob, 2003, 2010; Randall, 2013).

Here we propose a new approach to improving parameterization schemes. The new approach invests considerable computational effort up front to exploit global observations and targeted high-resolution simulations through the use of data assimilation and machine learning within physical, biological, and chemical process models. We first outline in broad terms how we envision ESMs to learn from global observations and targeted high-resolution simulations (section 2). Then we discuss in more concrete terms the framework underlying such learning ESMs (section 3). We illustrate the approach by learning parameters in a relatively simple dynamical system that mimics characteristics of the atmosphere and oceans (section 4). We conclude with an outlook of the opportunities the framework we outline presents and of the research program that needs to be pursued to realize it (section 5).

2. Learning From Observations and Targeted High-Resolution Simulations

2.1. Information Sources for Parameterization Schemes

Parameterization schemes can learn from two sources of information:

1. *Global observations*. We live in the golden age of Earth observations from space (L'Ecuyer et al., 2015). A suite of satellites flying in the formation known as the A-train has been streaming coordinated measurements of the composition of the atmosphere and of physical variables in the Earth system. We have nearly simultaneous measurements of variables such as temperature, humidity, and cloud and sea ice cover, with global coverage for more than a decade (Jiang et al., 2012, Simmons et al., 2016; Stephens et al., 2002, 2017). Space-based measurements of biogeochemical tracers and processes, such as measurements of column average CO_2 concentrations and of photosynthesis in terrestrial ecosystems, are also beginning to become available (e.g., Bloom et al., 2016; Crisp et al., 2004; Eldering et al., 2017; Frankenberg et al., 2011; Frankenberg et al., 2014; Joiner et al., 2011; Liu et al., 2017; Sun et al., 2017; Yokota et al., 2009), and so are more detailed observations of the cryosphere (e.g., Gardner et al., 2013; Shepherd et al., 2012; Vaughan et al., 2013). Parameterization schemes can learn from such space-based global data, which can be augmented and validated with more detailed local observations from the ground and from field studies.
2. *Local high-resolution simulations*. Some processes parameterized in ESMs are in principle computable, only the globally achievable resolution precludes their explicit computation. For example, the turbulent dynamics (though currently not the microphysics) of clouds can be computed with high fidelity in limited domains in large-eddy simulations (LES) with grid spacings of O(10 m) (Khairoutdinov et al., 2009, Matheou & Chung, 2014; Pressel et al., 2015, 2017; Siebesma et al., 2003; Stevens et al., 2005; Schalkwijk et al., 2015). Increased computational performance has made LES domain widths of O(10 km)–O(100 km) feasible in recent years, while the horizontal mesh size in atmosphere models has shrunk, to the point that the two scales have converged. Thus, while global LES that reliably resolve low clouds such as cumulus or stratocumulus will not be feasible for decades, it is possible to nest LES in selected grid columns of atmosphere models and conduct high-fidelity local simulations of cloud dynamics in them (Schneider et al., 2017). Local high-resolution simulations of ocean mesoscale turbulence or sea ice dynamics can be conducted similarly. Parameterization schemes can learn from such nested high-resolution simulations.

Of course, both observations and high-resolution simulations have been exploited in the development of parameterization schemes for some time. For example, data assimilation techniques have been used to estimate parameters in parameterization schemes from observations. Parameters especially in cloud, convection, and precipitation parameterizations have been estimated by minimizing errors in short-term weather forecasts over time scales of hours or days (e.g., Aksoy et al., 2006; Emanuel & Živković Rothman, 1999; Grell & Dévényi, 2002; Ruiz et al., 2013, 2015; Schirber et al., 2013) or by minimizing deviations between simulated and observed longer-term aggregates of climate statistics, such as global-mean TOA radiative fluxes accumulated over seasons or years (e.g., Jackson et al., 2008; Jarvinen et al., 2010; Neelin et al., 2010; Solonen et al., 2012; Tett et al., 2013). High-resolution simulations have been used to provide detailed dynamical information such as vertical velocity and turbulence kinetic energy profiles in convective clouds, which are not easily available from observations. They have often been employed to augment observations from local field studies, and parameterization schemes have been fit to and evaluated with the observations and the high-resolution simulations used in tandem (e.g., de Rooy et al., 2013; Hohenegger & Bretherton, 2011; Liu et al., 2001; Siebesma et al., 2003, 2007; Stevens et al., 2005; Romps, 2016). High-resolution deep convection-resolving simulations with O(1 km) horizontal grid spacing and, most recently, LES with O(100 m) horizontal grid spacing have also

been nested in small, usually two-dimensional subdomains of atmospheric grid columns, as a parameterization surrogate that explicitly resolves some aspects of cloud dynamics (e.g., Grabowski & Smolarkiewicz, 1999; Grabowski, 2001; Grabowski, 2016; Khairoutdinov & Randall, 2001; Khairoutdinov et al., 2005; Randall et al., 2003; Randall, 2013; Parishani et al., 2017). Such multiscale modeling approaches, often called superparameterization, have led to markedly improved simulations, for example, of the Asian monsoon, of tropical surface temperatures, and of precipitation and its diurnal cycle, albeit at great computational expense (e.g., Benedict & Randall, 2009; Pritchard & Somerville, 2009a, 2009b; Stan et al., 2010; DeMott et al., 2013). However, multiscale modeling relies on a scale separation between the global-model mesh size and the domain size of the nested high-resolution simulation (Weinan et al., 2007). Multiscale modeling is computationally advantageous relative to global high-resolution simulations as long as it suffices for the nested high-resolution simulation to subsample only a small fraction of the footprint of a global-model grid column, and to extrapolate the information so obtained to the entire footprint on the basis of statistical homogeneity assumptions. As the mesh size of global atmosphere models shrinks to horizontal scales of kilometers—resolutions that are already feasible in short integrations or limited areas and that will become routine in the next decade (Ban et al., 2015; Ohno et al., 2016; Palmer, 2014; Schneider et al., 2017)—the scale separation to the minimum necessary domain size of nested high-resolution simulations will disappear, and with it the computational advantage of multiscale modeling.

What we propose here combines elements of these existing approaches in a novel way. At its core are still parameterization schemes that are based on physical, biological, or chemical process models, whose mathematical structure is developed on the basis of theory, local observations, and, where possible, high-resolution simulations. But we propose that these parameterization schemes, when they are embedded in ESMs, learn directly from observations and high-resolution simulations that both sample the globe. High-resolution simulations are employed in a targeted way—akin to targeted or adaptive observations in weather forecasting (Bishop et al., 2001; Lorenz & Emanuel, 1998; Palmer et al., 1998)—to reduce uncertainties where observations are insufficient to obtain tight parameter estimates. Instead of incorporating high-resolution simulations globally in a small fraction of the footprint of each grid column like in multiscale modeling approaches, the ESM we envision deploys them locally, in entire grid columns, albeit only in a small subset of them. High-resolution simulations can be targeted to grid columns selected based on measures of uncertainty about model parameters. If the nested high-resolution simulations feed back onto the ESM, this corresponds to a locally extreme mesh refinement; however, two-way nesting may not always be necessary (e.g., Moeng et al., 2007; Zhu et al., 2010). The model learns parameters from observations and from nested high-resolution simulations in a computationally intensive learning phase, after which it can be used in a computationally more efficient manner, like models in use today. Nonetheless, even in simulations of climates beyond what has been observed, bursts of targeted high-resolution simulations can continue to be deployed to refine parameters and estimate their uncertainties.

2.2. Computable and Noncomputable Parameters

Learning from high-resolution simulations and observations is aimed at determining two different kinds of parameters in parameterization schemes: *computable* and *noncomputable* parameters. (Since parameters and parametric functions of state variables play essentially the same role in our discussion, we simply use the term parameter, with the understanding that this can include parametric functions and even nonparametric functions.) Computable parameters are those that can, in principle, be inferred from high-resolution simulations alone. They include parameters in radiative transfer schemes, which can be inferred from detailed line-by-line calculations; dynamical parameters in cloud turbulence parameterizations, such as entrainment rates, which can be inferred from LES; or parameters in ocean mixing parameterizations, which can be inferred from high-resolution simulations. Noncomputable parameters are parameters that, currently, cannot be inferred from high-resolution simulations, either because computational limitations make it necessary for them to also appear in parameterization schemes in high-resolution simulations, or because the microscopic equations governing the processes in question are unknown. They include parameters in cloud microphysics parameterizations, which are still necessary to include in LES, and many parameters characterizing ecological and biogeochemical processes, whose governing equations are unknown. Cloud microphysics parameters will increasingly become computable through direct numerical simulation (Devenish et al., 2012; Grabowski & Wang, 2013), but ecological and biogeochemical parameters will remain noncomputable for the foreseeable future. Both computable and noncomputable parameters can, in principle, be learned from observations; the only restrictions to their identifiability come from the well-posedness of the learning

problem and its computational tractability. But only computable parameters can be learned from targeted high-resolution simulations. To be able to learn computable parameters, it is essential to represent non-computable aspects of a parameterization scheme consistently in the high-resolution simulation and in the parameterization scheme that is to learn from the high-resolution simulation. For example, radiative transfer and microphysical processes need to be represented consistently in a high-resolution LES and in a parameterization scheme if the parameterization scheme is to learn computable dynamical parameters such as entrainment rates from the LES.

This approach presents challenges for parameter learning, since it implies the need to use observational data and high-resolution simulations in tandem to improve model parameterizations. But it also presents an opportunity: in doing so, the reliability and predictive power of ESMs can be improved, and uncertainties in parameters and predictions can be quantified.

2.3. Objectives: Bias Reduction and Exploitation of Emergent Constraints

Computational tractability is paramount for the success of any parameter learning algorithm for ESMs (e.g., Annan & Hargreaves, 2007; Jackson et al., 2008; Neelin et al., 2010; Solonen et al., 2012). The central issue is the number of times the objective function needs to be evaluated, and hence an ESM needs to be run, in the process of parameter learning. Standard parameter estimation and inverse problem approaches may require $O(10^5)$ function or derivative evaluations to learn $O(100)$ parameters, especially if uncertainty in the estimates is also required (Cotter et al., 2013). This many forward integrations and/or derivative evaluations of ESMs are not feasible if each involves accumulation of longer-term climate statistics. Fast parameterized processes in climate models often exhibit errors within a few hours or days of integration that are similar to errors in the mean state of the model (Klocke & Rodwell, 2014; Ma et al., 2013; Phillips et al., 2004; Rodwell & Palmer, 2007; Xie et al., 2012). This has given rise to hopes that it may suffice to evaluate objective functions by weather hindcasts over time scales of only hours, making many evaluations of an objective function feasible (Aksoy et al., 2006; Ruiz et al., 2013; Wan et al., 2014). But experience has shown that such short-term optimization may not always lead to the desired improvements in climate simulations (Schirber et al., 2013). Additionally, slower parameterized processes, for example, involving biogeochemical cycles or the cryosphere, require longer integration times to accumulate statistics entering any meaningful objective function. Therefore, we focus on objective functions involving climate statistics accumulated over windows that we anticipate to be wide compared with the O(10 days) time scale over which the atmosphere forgets its initial condition. Then the accumulated statistics do not depend sensitively on atmospheric initial conditions. This reduces the onus of correctly assimilating atmospheric initial conditions in parameter learning, which would be required if one were to match simulated and observed trajectories, as in approaches that assimilate model parameters jointly with the state of the system by augmenting state vectors with parameters (e.g., Aksoy et al., 2006; Anderson et al., 2009; Dee, 2005). The minimum window over which climate statistics will need to be accumulated will vary from process to process, generally being longer for slower processes (e.g., the cryosphere) than faster processes (e.g., the atmosphere). For slower processes whose initial condition is not forgotten over the accumulation window, it will remain necessary to correctly assimilate initial conditions.

The objective functions to be minimized in the learning phase can be chosen to directly minimize biases in climate simulations, for example, precipitation biases such as the longstanding double-ITCZ bias in the tropics (Adam et al., 2016, 2017; Lin, 2007; Li & Xie, 2014; Zhang et al., 2015), or cloud cover biases such as the "too few–too bright" bias in the subtropics (Karlsson et al., 2008; Nam et al., 2012; Webb et al., 2001; Zhang et al., 2005). Because the sensitivity with which an ESM responds to increases in greenhouse gas concentrations correlates with the spatial structure of some of these biases in the models (e.g., Tian, 2015; Siler et al., 2017), minimizing regional biases will likely reduce uncertainties in climate projections, in addition to leading to more reliable simulations of the present climate. To minimize biases, the objective function needs to include mean-field terms penalizing mismatch between spatially and at least seasonally resolved simulated and observed mean fields, for example, of precipitation, ecosystem primary productivity, and TOA radiative energy fluxes.

Additionally, there is a growing literature on "emergent constraints," which typically are fluctuation-dissipation relationships that relate measurable fluctuations in the present climate to the response of the climate system to perturbations (Collins et al., 2012; Hall & Qu, 2006; Klein & Hall, 2015). For example, how strongly tropical low-cloud cover covaries with surface temperature from year to year or even seasonally in the present climate correlates in climate models with the amplitude of the cloud response to global warming

(Qu et al., 2014, 2015; Brient & Schneider, 2016). Therefore, the observable low-cloud cover covariation with surface temperature in the present climate can be used to constrain the cloud response to global warming. Or as another example, how strongly atmospheric CO_2 concentrations covary with surface temperature in the present climate correlates in climate models with the amplitude of the terrestrial ecosystem response to global warming (e.g., the balance between CO_2 fertilization of plants and enhanced soil and plant respiration under warming) (Cox et al., 2013; Wenzel et al., 2014). Therefore, the observable CO_2 concentration covariation with surface temperature can be used to constrain the terrestrial ecosystem response to global warming. Such emergent constraints are usually used post facto, in the evaluation of ESMs. They lead to inferences about the likelihood of a model given the measured natural variations, and they therefore can be used to assess how likely it is that its climate change projections are correct (e.g., Brient & Schneider, 2016). But emergent constraints usually are not used directly to improve models. In what we propose, they are used directly to learn parameters in ESMs and to reduce uncertainties in the climate response. To do so, covariance terms (e.g., between surface temperature and cloud cover or TOA radiative fluxes, or between surface temperature and CO_2 concentrations) need to be included in the objective function.

The choice of objective functions to be employed is key to the success of what we propose. The use of time-averaged statistics such as mean-field and covariance terms will make the objective functions smoother and hence reduce the computational cost of minimization, compared with minimizing objective functions that directly penalize mismatch between simulated and observed trajectories of the Earth system. From the point of view of statistical theory, the objective functions should contain the sufficient statistics for the parameters of interest, but what these are is not usually known a priori. In practice, the choice of objective functions will be guided by expertise specific to the relevant subdomains of Earth system science, as well as computational cost. Given that current ESM components such as clouds and the carbon cycle exhibit large seasonal biases (e.g., Lin et al., 2014; Keppel-Aleks et al., 2012; Karlsson & Svensson, 2013), and their response to long-term warming in some respects resembles their response to seasonal variations (e.g., Brient & Schneider, 2016; Wenzel et al., 2016), accumulating seasonal statistics in the objective functions suggests itself as a starting point.

3. Machine Learning Framework for Earth System Models

3.1. Models and Data

To outline how we envision parameterization schemes in ESMs to learn from diverse data, we first set up notation. Let $\theta = (\theta_c, \theta_n)$ denote the vector of model parameters to be learned, consisting of computable parameters θ_c that can be learned from high-resolution simulations, and noncomputable parameters θ_n that can only be learned from observations (for example, because high-resolution simulations themselves depend on θ_n). The parameters θ appear in parameterization schemes in a model, which may be viewed as a map $\mathcal{G}$, parameterized by time t, that takes the parameters θ to the state variables $\boldsymbol{x}$,

$$\boldsymbol{x}(t) = \mathcal{G}(\theta, t). \tag{1}$$

The state variables $\boldsymbol{x}$ can include temperatures, humidity variables, and cloud, cryosphere, and biogeochemical variables, and the map $\mathcal{G}$ may depend on initial conditions and time-evolving boundary or forcing conditions. The map $\mathcal{G}$ typically represents a global ESM. The state variables $\boldsymbol{x}$ are linked to observables $\boldsymbol{y}$ through a map $\mathcal{H}$ representing an observing system, so that

$$\boldsymbol{y}(t) = \mathcal{H}(\boldsymbol{x}(t)). \tag{2}$$

The observables $\boldsymbol{y}$ might represent surface temperatures, CO_2 concentrations, or spectral radiances emanating from the TOA. The map $\mathcal{H}$ in practice will be realized through an observing system simulator, which simulates how observables $\boldsymbol{y}$ are impacted by a multitude of state variables $\boldsymbol{x}$. The actual observations (e.g., space-based measurements) are denoted by $\tilde{\boldsymbol{y}}$, so $\boldsymbol{y}(t) - \tilde{\boldsymbol{y}}(t)$ is the mismatch between simulations and observations. Since $\boldsymbol{y}$ is parameterized by θ, while $\tilde{\boldsymbol{y}}$ is independent of θ, mismatches between $\boldsymbol{y}$ and $\tilde{\boldsymbol{y}}$ can be used to learn about θ.

Local high-resolution simulations nested in a grid column of an ESM may be viewed as a time-dependent map $\mathcal{L}$ from the state variables $\boldsymbol{x}$ of the ESM to simulated state variables $\tilde{\boldsymbol{z}}$,

$$\tilde{\boldsymbol{z}}(t) = \mathcal{L}(\theta_n, t; \boldsymbol{x}). \tag{3}$$

The map $\mathcal{L}$ is parameterized by noncomputable parameters θ_n and time t, and it can involve the time history of the state variables $\boldsymbol{x}$ up to time t. The vector $\tilde{\boldsymbol{z}}$ contains statistics of high-resolution variables whose counterparts in the ESM are computed by parameterization schemes, such as the mean cloud cover or liquid water content in a grid box. The corresponding variables $\boldsymbol{z}$ in the ESM are obtained by a time-dependent map $\mathcal{L}$ that takes state variables $\boldsymbol{x}$ and parameters θ to $\boldsymbol{z}$,

$$\boldsymbol{z}(t) = \mathcal{S}(\theta, t; \boldsymbol{x}). \tag{4}$$

The map $\mathcal{S}$ typically represents a single grid column of the ESM with its parameterization schemes, taking as input $\boldsymbol{x}$ from the ESM. It is structurally similar to $\mathcal{L}$. Crucially, however, $\mathcal{L}$ generally depends on all parameters $\theta = (\theta_c, \theta_n)$, while $\mathcal{L}$ only depends on noncomputable parameters θ_n. Thus, the mismatch $\boldsymbol{z}(t) - \tilde{\boldsymbol{z}}(t)$ can be used to learn about the computable parameters θ_c.

The same framework also covers other ways of learning about parameterizations schemes from data. For example, the map $\mathcal{G}$ may represent a single grid column of an ESM, driven by time-evolving boundary conditions from reanalysis data at selected sites. Observations at the sites can then be used to learn about the parameterization schemes in the column (Neggers et al., 2012). Or, similarly, the map $\mathcal{G}$ may represent a local high-resolution simulation driven by reanalysis data, with parameterization schemes, for example, for cloud microphysics, about which one wants to learn from observations.

3.2. Objective Functions

Objective functions are defined through mismatch between the simulated data $\boldsymbol{y}$ and observations $\tilde{\boldsymbol{y}}$, on the one hand, and simulated data $\boldsymbol{z}$ and high-resolution simulations $\tilde{\boldsymbol{z}}$, on the other hand. We define mismatches using time-averaged statistics, because they do not suffer from sensitivity to atmospheric initial conditions; indeed, matching trajectories directly requires assimilating atmospheric initial conditions, which would make it difficult to disentangle mismatches due to errors in climatically unimportant atmospheric initial conditions from those due to parameterization errors. However, the time averages can still depend on initial conditions for slowly evolving components of the Earth system, such as ocean circulations or ice sheets.

We denote the time average of a function $\phi(t)$ over the time interval $[t_0, t_0 + T]$ by

$$\langle \phi \rangle_T = \frac{1}{T} \int_{t_0}^{t_0+T} \phi(t)\, \mathrm{d}t. \tag{5}$$

The observational objective function can then be written in the generic form

$$J_o(\theta) = \frac{1}{2} \| \langle \boldsymbol{f}(\boldsymbol{y}) \rangle_T - \langle \boldsymbol{f}(\tilde{\boldsymbol{y}}) \rangle_T \|^2_{\Sigma_y} \tag{6}$$

with the two-norm

$$\| \cdot \|_{\Sigma_y} = \| \Sigma_y^{-\frac{1}{2}} \cdot \| \tag{7}$$

normalized by error standard deviations and covariance information captured in Σ_y. The function $\boldsymbol{f}$ of the observables typically involves first- and second-order quantities, for example,

$$\boldsymbol{f}(\boldsymbol{y}) = \begin{pmatrix} \boldsymbol{y} \\ y_i' y_j' \end{pmatrix}, \tag{8}$$

where, for any observable ϕ, $\phi'(t) = \phi(t) - \langle \phi \rangle_T$ denotes the fluctuation of ϕ about its mean $\langle \phi \rangle_T$. With $\boldsymbol{f}$ given by (8), the objective function penalizes mismatch between the vectors of mean values $\langle \boldsymbol{y} \rangle_T$ and $\langle \tilde{\boldsymbol{y}} \rangle_T$ and between the covariance components $\langle y_i' y_j' \rangle_T$ and $\langle \tilde{y}_i' \tilde{y}_j' \rangle_T$ for some indices i and j. The least squares form of the objective function (6) follows from assuming an error model

$$\boldsymbol{f}(\boldsymbol{y}) = \boldsymbol{f}(\tilde{\boldsymbol{y}}) + \boldsymbol{\eta}, \tag{9}$$

with the matrix Σ_y encoding an assumed covariance structure of the noise vector $\boldsymbol{\eta}$. The relevant components of Σ_y may be chosen very small for quantities that are used as constraints on the ESM (e.g., the requirement of a closed global energy balance at TOA).

For the mismatch to high-resolution simulations, we accumulate statistics over an ensemble of high-resolution simulations in different grid columns of the ESM and at different times, possibly, but not necessarily, also accumulating in time. We denote the corresponding ensemble and time average by $\langle \phi \rangle_E$, and define an objective function analogously to that for the observations through

$$J_s(\theta_c) = \frac{1}{2} \| \langle \boldsymbol{g}(\boldsymbol{z}) \rangle_E - \langle \boldsymbol{g}(\tilde{\boldsymbol{z}}) \rangle_E \|^2_{\Sigma_z}. \tag{10}$$

Like the function $\boldsymbol{f}$ above, the function $\boldsymbol{g}$ typically involves first- and second-order quantities, and the least squares form of the objective functions follows from the assumed covariance structure Σ_z of the noise.

3.3. Learning Algorithms

Learning algorithms attempt to choose parameters θ that minimize J_o and J_s. However, minimization of J_o and J_s does not always determine the parameters uniquely, for example, if there are strongly correlated parameters or if the number of parameters to be learned exceeds the number of available observational degrees of freedom. In such cases, regularization is necessary to choose a good solution for the parameters among the multitude of possible solutions. This may be achieved in various ways: by adding to the least-squares objective functions (6) and (10), regularizing penalty terms that incorporate prior knowledge about the parameters (Engl et al., 1996), by Bayesian probabilistic regularization (Kaipio & Somersalo, 2005), or by restriction of the parameters to a subset, as in ensemble Kalman inversion (Iglesias et al., 2013).

All of these regularization approaches may be useful in ESMs. They involve different trade-offs between computational expense and the amount of information about the parameters they provide.

1. Classical regularized least squares leads to an optimization problem that is typically tackled by gradient descent or Gauss-Newton methods, in which derivatives of the parameter-to-data map are employed (Nocedal & Wright, 2006). Such methods usually require $O(10^2)$ integrations of the forward model or evaluations of its derivatives with respect to parameters.
2. Bayesian inversions usually employ Markov chain Monte Carlo (MCMC) methods (Brooks et al., 2011) and variants such as sequential Monte Carlo (Del Moral et al., 2006) to approximate the posterior probability density function (PDF) of parameters, given data and a prior PDF. A PDF of parameters provides much more information than a point estimate and consequently MCMC methods typically require many more forward model integrations, sometimes on the order of $O(10^5)$. The computational demands can be decreased by an order of magnitude by judicious use of derivative information where available (see Beskos et al., 2017, and references therein) or by improved sampling strategies (e.g., Jackson et al., 2008; Jarvinen et al., 2010; Solonen et al., 2012). Nonetheless, the cost remains orders of magnitude higher than for optimization techniques.
3. Ensemble Kalman methods are easily parallelizable, derivative-free alternatives to the classical optimization and Bayesian approaches (Houtekamer & Zhang, 2016). Although theory for them is less well developed, empirical evidence demonstrates behavior similar to derivative-based algorithms in complex inversion problems, with a comparable number of forward model integrations (Iglesias, 2016). Ensemble methods for joint state and parameter estimation have recently been systematically developed (Bocquet & Sakov, 2013, 2014; Carrassi et al., 2017), and they are emerging as a promising way to solve inverse problems and to obtain qualitative estimates of uncertainty. However, numerical experiments have indicated that such uncertainty information is qualitative at best: the Kalman methods invoke Gaussian assumptions, which may not be justified, and even if the Gaussian approximation holds, the ensemble sizes needed for uncertainty quantification may not be practical (Iglesias et al., 2013; Law & Stuart, 2012).

An important consideration is how to blend the information about parameters contained in the high-resolution simulations and in the observations. One approach is as follows, although others may turn out to be preferable. Minimizing the high-resolution objective function J_s in principle gives the computable parameters θ_c as an implicit function of the noncomputable parameters θ_n. This implicit function may then be used as prior information to minimize the observational objective function J_o over θ. Bayesian MCMC approaches may be feasible for fitting J_s, since the single-column model $\mathcal{L}$ is relatively cheap to evaluate, and the ensemble of high-resolution simulations $\mathcal{L}$ needed may not be large. Although Bayesian approaches may not be feasible for fitting J_o, for which accumulation of statistics of the model $\mathcal{G}$ is required, this hierarchical approach does have the potential to incorporate detailed uncertainty estimates coming from the high-resolution simulations.

The choice of normalization (i.e., Σ_y and Σ_z) in the objective functions plays a significant role in parameter learning, and learning about it has been demonstrated to have considerable impact on data assimilation for weather forecasts (Dee, 1995; Stewart et al., 2014). We will not discuss this issue in any detail, but note it may be addressed by the use of hierarchical Bayesian methodology and ensemble Kalman analogs. Nor will we dwell on the important issue of structural uncertainty—model error—other than to note that this can, in principle, be addressed through the inverse problem approach advocated here: additional unknown parameters, placed judiciously within the model to account for model error, can be learned from data

(Dee, 2005; Kennedy & O'Hagan, 2001). The choice of normalization is especially important in this context as it relates to disentangling learning about model error from learning about the other parameters of interest.

Learning algorithms for ESMs can be developed further in several ways:

1. Minimization of the objective functions J_o and J_s may be performed by online filtering algorithms, akin to those used in the initialization of weather forecasts, which sequentially update parameters as information becomes available (Law et al., 2015). This can reduce the number of forward model integrations required for parameter estimation, and it can allow parameterization schemes to learn adaptively from high-resolution simulations during the course of a global simulation.
2. Where to employ targeted high-resolution simulations can be chosen to optimize aspects of the learning process. The simplest approach would be to deploy them randomly, for example, by selecting regions with a probability proportional to their climatological cloud fraction for high-resolution simulations of clouds. More efficient would be techniques of optimal experimental design (see Alexanderian et al., 2016, and references therein), within online filtering algorithms. With such techniques, high-resolution simulations could be generated to order, to update aspects of parameterization schemes that have the most influence on the global system with which they interact.

Progress along these lines will require innovation. For example, filtering algorithms need to be adapted to deal with strong serial correlations such as those that arise when averages $\langle\phi\rangle_{T_i}$ are accumulated over increasing spans $T_i < T_{i+1}$ and parameters are updated from one average $\langle\phi\rangle_{T_i}$ to a longer average $\langle\phi\rangle_{T_{i+1}}$. And optimal experimental design techniques require the development of cheap computational methods to evaluate sensitivities of the ESM to individual aspects of parameterization schemes.

4. Illustration With Dynamical System

We envision ESMs eventually to learn parameters online, with targeted high-resolution simulations triggering parameter updates on the fly. Here we want to illustrate in off-line mode some of the opportunities and challenges of learning parameters in a relatively simple dynamical system. We use the Lorenz-96 model (Lorenz, 1996), which has nonlinearities resembling the advective nonlinearities of fluid dynamics and a multiscale coupling of slow and fast variables similar to what is seen in ESMs. The model has been used extensively in the development and testing of data assimilation methods (e.g., Anderson, 2001; Lorenz & Emanuel, 1998; Ott et al., 2004).

4.1. Lorenz-96 Model

The Lorenz-96 model consists of K slow variables X_k ($k = 1, \dots, K$), each of which is coupled to J fast variables $Y_{j,k}$ ($j = 1, \dots, J$) (Lorenz, 1996):

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - hc\bar{Y}_k, \tag{11}$$

$$\frac{1}{c}\frac{\mathrm{d}Y_{j,k}}{\mathrm{d}t} = -bY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + \frac{h}{J}X_k. \tag{12}$$

The overbar denotes the mean value over j,

$$\bar{Y}_k = \frac{1}{J}\sum_{j=1}^{J} Y_{j,k}. \tag{13}$$

Both the slow and fast variables are taken to be periodic in k and j, forming a cyclic chain with $X_{k+K} = X_k$, $Y_{j,k+K} = Y_{j,k}$, and $Y_{j+J,k} = Y_{j,k+1}$. The slow variables X may be viewed as resolved-scale variables and the fast variables Y as unresolved variables in an ESM. Each of the K slow variables X_k may represent a property such as surface air temperature in a cyclic chain of grid cells spanning a latitude circle. Each slow variable X_k affects the J fast variables $Y_{j,k}$ in the grid cell, which might represent cloud-scale variables such as liquid water path in each of J cumulus clouds. In turn, the mean value of the fast variables over the cell, $\bar{Y}_k$, feeds back onto the slow variables X_k. The strength of the coupling between fast and slow variables is controlled by the parameter h, which represents an interaction coefficient, for example, an entrainment rate that couples cloud-scale variables to their large-scale environment. Time is nondimensionalized by the linear-damping time scale of the slow variables, which we nominally take to be 1 day, a typical thermal relaxation time of surface temperatures (Swanson & Pierrehumbert, 1997). The parameter c controls how rapidly the fast variables are damped relative

to the slow; it may be interpreted as a microphysical parameter controlling relaxation of cloud variables, such as a precipitation efficiency. The parameter F controls the strength of the external large-scale forcing and b the amplitude of the nonlinear interactions among the fast variables. Following Lorenz (1996), albeit relabeling parameters, we choose $K = 36$, $J = 10$, $h = 1$, and $F = c = b = 10$, which ensures chaotic dynamics of the system.

The quadratic nonlinearities in this dynamical system resemble advective nonlinearities, for example, in the sense that they conserve the quadratic invariants ("energies") $\sum_k X_k^2$ and $\sum_j Y_{j,k}^2$ (Lorenz & Emanuel, 1998). The interaction between the slow and fast variables conserves the "total energy" $\sum_k \left(X_k^2 + \sum_j Y_{j,k}^2 \right)$. Energies are damped by the linear terms; they are prevented from decaying to zero by the external forcing F. Eventually, the system approaches a statistically steady state in which driving by the external forcing F balances the linear damping.

Let $\langle \cdot \rangle_\infty$ denote a long-term time mean in the statistically steady state, and note that all slow variables X_k are statistically identical, as are all fast variables $Y_{j,k}$, so we can use the generic symbols X and Y in statistics of the variables. Multiplication of (11) by X_k, using that all variables X_k are statistically identical, and averaging shows that, in the statistically steady state, second moments of the slow variables satisfy

$$\langle X^2 \rangle_\infty = F \langle X \rangle_\infty - hc \langle X \bar{Y} \rangle_\infty . \tag{14}$$

Similarly, second moments of the fast variables satisfy

$$\langle \overline{Y^2} \rangle_\infty = \frac{h}{J} \langle X \bar{Y} \rangle_\infty , \tag{15}$$

where the overbar again denotes a mean value over the fast-variable index j. That is, the interaction coefficient h can be determined from estimates of the one-point statistics $\langle \overline{Y^2} \rangle_\infty$ and $\langle X \bar{Y} \rangle_\infty$. Its inverse is proportional to the regression coefficient of the fast variables onto the slow: $h^{-1} \propto \langle X \bar{Y} \rangle_\infty / \langle \overline{Y^2} \rangle_\infty$. So the regression of the fast variables onto the slow can be viewed as providing an "emergent constraint" on the system, insofar as the interaction coefficient h affects the response of the system to perturbations (e.g., in F). Estimates of $\langle X^2 \rangle_\infty$ and $\langle X \rangle_\infty$ provide an additional constraint (14) on the parameters F and c. Taking mean values of the dynamical equations (11) and (12) would provide further constraints on these parameters, as well as on b, in terms of two-point statistics involving shifts in k and j, for example, covariances of X_k and X_{k-1}.

In what follows, we demonstrate the performance of learning algorithms in a perfect-model setting, first focusing on one-point statistics to show how to learn about parameters in the full dynamical system from them. Subsequently, we use two-point statistics to learn about parameters in a single "grid column" of fast variables only.

4.2. Parameter Learning in Perfect-Model Setting

We generate data from the dynamical system (11) and (12) with the parameters $\boldsymbol{\theta} = (F, h, c, b)$ set to $\tilde{\boldsymbol{\theta}} = (10, 1, 10, 10)$. The role of "observations" $\tilde{\boldsymbol{y}} = (\tilde{X}, \tilde{Y})$ in the perfect-model setting is played by data $\tilde{X}$ and $\tilde{Y}$ generated by the dynamical system with parameters $\boldsymbol{\theta}$ set to their "true" values $\tilde{\boldsymbol{\theta}}$. That is, the dynamical system (11) and (12) with parameters $\boldsymbol{\theta}$ stands for the global model $\mathcal{G}$, the observing system map $\mathcal{H}$ is the identity, and the data $\tilde{X}$ and $\tilde{Y}$ generated by the dynamical system with parameters $\tilde{\boldsymbol{\theta}}$ is a surrogate for observations. The parameters $\boldsymbol{\theta}$ of the dynamical system are then learned by matching statistics $\langle \phi \rangle_T$ accumulated over $T = 100$ days (with 1 day denoting the unit time of the system), using discrete sums in place of the time integral in the average (5) and minimizing the "observational" objective function

$$J_o(\boldsymbol{\theta}) = \frac{1}{2} \| \langle \boldsymbol{f}(X, Y) \rangle_T - \langle \boldsymbol{f}(\tilde{X}, \tilde{Y}) \rangle_\infty \|_\Sigma^2 . \tag{16}$$

The moment function to be matched,

$$\boldsymbol{f}(X, Y) = \begin{pmatrix} X \\ \bar{Y} \\ X^2 \\ X\bar{Y} \\ \overline{Y^2} \end{pmatrix} , \tag{17}$$

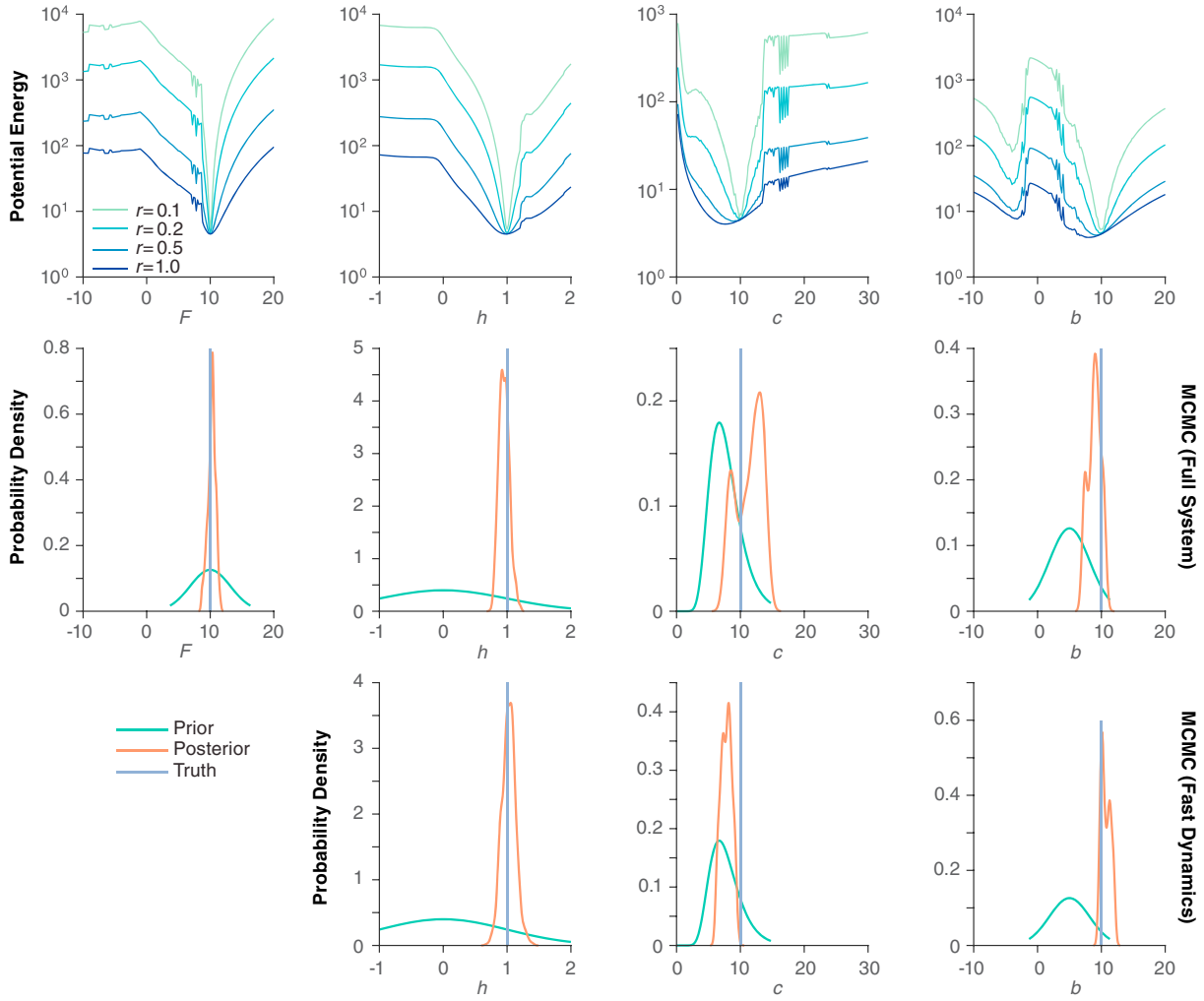

Figure 1. Probability density functions for parameter learning in Lorenz-96 model. (top row) Marginal potential energies (negative logarithm of posterior PDF) at different noise levels r. Each marginal potential energy is obtained by varying the parameter on the abscissa while holding other parameters fixed at their true values. (middle row) Marginal prior and posterior PDFs estimated by MCMC from the full dynamics (11) and (12), with noise level $r = 0.5$ and accumulating over $T = 100$ days for each sample. (bottom row) Marginal prior and posterior PDFs estimated by MCMC from the fast dynamics (12) only, likewise with noise level $r' = 0.5$ but accumulating over only $T' = 20$ days for each sample.

has an entry for each of the $K = 36$ indices k, giving a vector of length $5K = 180$. The noise covariance matrix Σ is chosen to be diagonal, with entries that are proportional to the sample variances of the moments contained in the vector $\boldsymbol{f}$,

$$\mathrm{diag}\Sigma = r^2 \left[\mathrm{var}(X), \mathrm{var}(\bar{Y}), \mathrm{var}(X^2), \mathrm{var}(X\bar{Y}), \mathrm{var}(\overline{Y^2})\right]. \tag{18}$$

Here $\mathrm{var}(\phi)$ denotes the variance of ϕ, and r is an empirical parameter indicating the noise level. The variances $\mathrm{var}(\phi)$ and the "true moments" $\langle \boldsymbol{f}(\tilde{X}, \tilde{Y})\rangle_\infty$ are estimated from a long (46,416 days) control simulation of the dynamical system with the true parameters $\tilde{\boldsymbol{\theta}}$.

As an illustrative example, we use normal priors for $(\theta_1, \theta_2, \theta_4) = (F, h, b)$, with mean values $(\mu_1, \mu_2, \mu_4) = (10, 0, 5)$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_4^2) = (10, 1, 10)$. Enforcing positivity of c, we use a log-normal prior for $\theta_3 = c$, with a mean value $\mu_3 = 2$ and variance $\sigma_3^2 = 0.1$ for $\log c$ (i.e., a mean value of 7.4 for c). We take the parameters a priori to be uncorrelated, so that the prior covariance matrix is diagonal.

To illustrate the landscape learning algorithms have to navigate, Figure 1 (top row) shows sections through the potential energy, defined as the negative logarithm of the posterior PDF,

$$U(\boldsymbol{\theta}) = J_o(\boldsymbol{\theta}) - \sum_{i=1}^{4} \log p_i(\theta_i), \tag{19}$$

where $p_i(\theta_i)$ is the prior PDF of parameter θ_i. The figure shows the marginal potential energies obtained as one parameter at a time is varied and the objective function $J_o(\boldsymbol{\theta})$ is accumulated by forward integration, while the other parameters are held fixed at their true values. As the noise level r increases, the contribution of the log-likelihood of the data ($\propto J_o$) is downweighted relative to the prior, the posterior modes shift toward the prior modes, and the posterior is smoothed. Here the objective function $J_o(\boldsymbol{\theta})$ for each parameter setting is accumulated over a long period (10^4 days) to minimize sampling variability. However, even with this wide accumulation window, sampling variability remains in some parameter regimes and there noticeably affects $J_o(\boldsymbol{\theta})$. An example is the roughness around $c = 17$, which appears to be caused by metastability on time scales longer than the accumulation window. The roughness could be smoothed by accumulating over periods that are yet longer, or by averaging over an ensemble of initial conditions, but analogous smoothing might be impractical for ESMs. Time-averaged ESM statistics may exhibit similarly rough dependencies on some parameters (e.g., Suzuki et al., 2013; Zhao et al., 2016), although the dependence on other parameters appears to be relatively smooth (e.g., Neelin et al., 2010), perhaps because ESM parameters targeted for tuning are chosen for the smooth dependence of the climate state on them. Roughness of the potential energy landscape can present challenges for learning algorithms, which may get stuck in local minima. Note also the bimodality in b, which arises because the one-point statistics we fit cannot easily distinguish prograde wave modes of the system (which propagate toward increasing k) from retrograde modes (cf. Lorenz & Emanuel, 1998).

4.2.1. Bayesian Inversion

We use the random-walk Metropolis (RWM) MCMC algorithm (Brooks et al., 2011) for a full Bayesian inversion of parameters in the dynamical system (11) and (12), thereby sampling from the posterior PDF. To reduce burn-in (MCMC spin-up) time, we initialize the algorithm close to the true parameter values with the result of an ensemble Kalman inversion (see below). The RWM algorithm is then run over 2,200 iterations, the first 200 iterations are discarded as burn-in, and the posterior PDF is estimated by binning every other of the remaining 2,000 samples. The objective function for each sample is accumulated over $T = 100$ days, using the end state of the previous forward integration as initial condition for the next one, without discarding any spin-up after a parameter update.

The resulting marginal posterior PDFs do not all peak exactly at the true parameter values, but the true parameter values lie in a region that contains most of the posterior probability mass (Figure 1, middle row). The posterior PDFs indicate the uncertainties inherent in estimating the parameters. The posterior PDF of c has the largest spread, in terms of standard deviation normalized by mean, indicating relatively large uncertainty in this parameter. The uncertainty appears to arise from the roughness of the potential energy (Figure 1, top row), which reflects inherent sensitivity of the system response to parameter variability; additional roughness of the posterior PDFs may be caused by sampling variability from finite-time averages (Wang et al., 2014). For all four parameters, the posterior PDFs differ significantly from the priors, demonstrating the information content provided by the synthetic data. Finally, although these results have been obtained with $O(10^3)$ forward model integrations and objective function evaluations, more objective function evaluations may be required for more complex forward models, such as ESMs.

4.2.2. Ensemble Kalman Inversion

Ensemble Kalman inversion may be an attractive learning algorithm for ESMs when Bayesian inversion with MCMC is computationally too demanding. To illustrate its performance, we use the algorithm of Iglesias et al. (2013), initializing ensembles of size M with parameters drawn from the prior PDFs. In the analysis step of the Kalman inversion, we perturb the target data by addition of noise with zero mean and variance given by (18), that is, replacing $\langle \boldsymbol{f}(\tilde{X}, \tilde{Y}) \rangle_\infty$ by $\langle \boldsymbol{f}(\tilde{X}, \tilde{Y}) \rangle_\infty + \boldsymbol{\eta}^{(j)}$ with $\boldsymbol{\eta}^{(j)} \sim \mathcal{N}(0, \Sigma)$ for each ensemble member j. As in the MCMC algorithm, the objective function for each parameter setting is accumulated over $T = 100$ days, without discarding any spin-up after each parameter update. As initial state for the integration of the ensemble, we use a state drawn from the statistically steady state of a simulation with the true parameters.

Table 1 summarizes the solutions obtained by this ensemble Kalman inversion after $N_{\max} = 25$ iterations, for different ensemble sizes M and noise levels r. The ensemble mean of the Kalman inversion provides reasonable parameter estimates. But the ensemble standard deviation does not always provide quantitatively accurate uncertainty information. For example, for low noise levels, the true parameter values often lie more than 2 standard deviations away from the ensemble mean. The ensemble spread also differs quantitatively from the posterior spread in the MCMC simulations. In experiments in which we did not perturb the target data, the smaller ensembles ($M = 10$) occasionally collapsed, with each ensemble member giving the same point estimate of the parameters. In such cases, the ensemble contains no uncertainty information, illustrating potential

Table 1
Ensemble Means and Standard Deviations for the Parameters $\theta = (F, h, c, b)$ Obtained by Ensemble Kalman Inversions for Different Ensemble Sizes M and Different Noise Levels r

Noise	Mean ($M = 10$)	Mean ($M = 100$)	Std ($M = 100$)
$r = 0.1$	(9.62, 0.579, 9.37, 2.63)	(9.71, 0.992, 8.70, 9.95)	(0.023, 0.001, 0.104, 0.022)
$r = 0.2$	(9.57, 0.516, 7.90, 3.15)	(9.77, 0.994, 9.07, 10.04)	(0.107, 0.005, 0.524, 0.103)
$r = 0.5$	(9.77, 0.522, 9.29, 5.31)	(9.63, 0.982, 8.34, 9.93)	(0.295, 0.017, 1.477, 0.350)
$r = 1.0$	(9.70, 0.633, 7.68, 6.13)	(9.53, 0.952, 7.97, 9.37)	(0.385, 0.039, 1.964, 0.701)

pitfalls of using ensemble Kalman inversion for uncertainty quantification. However, with the perturbed data and for larger ensembles, the ensemble standard deviation is qualitatively consistent with the posterior PDF estimated by MCMC (Figure 1, middle row). It provides some uncertainty information, especially for higher noise levels, for example, in the sense that the parameter c is demonstrably the most uncertain (Table 1 and Figure 2b). Methods such as localization and variance inflation can help with issues related to ensemble collapse and can also be used to improve ensemble statistics more generally (see Law et al., 2015, and references therein). However, systematic principles for their application with the aim of correctly reproducing Bayesian posterior statistics have not been found, and so we have not adopted this approach.

The ensemble Kalman inversion typically converges within a few iterations (Figure 2 indicates $\lesssim 5$ iterations when $M = 100$). Larger ensembles lead to solutions closer to the truth (Figure 2a). Convergence within five iterations for ensembles of size 10 or 100 implies 50 or 500 objective function evaluations, representing substantial computational savings over the MCMC algorithm with 2,000 objective function evaluations. These computational savings come at the expense of detailed uncertainty information. Where the optimal trade-off lies between computational efficiency, on the one hand, and precision of parameter estimates and uncertainty quantification, on the other hand, remains to be investigated.

4.3. Parameter Learning From Fast Dynamics

Finally, we investigate learning about parameters from the fast dynamics (12) alone. This is similar to learning about computable parameters from local high-resolution simulations, for example, of clouds. That is, the fast dynamics (12) with the true parameters $\tilde{\theta}$ stand for the high-resolution model $\mathcal{L}$, which generates data $\tilde{\boldsymbol{z}} = \tilde{Y}$, and the fast dynamics with parameters θ play the role of the single-column model $\mathcal{L}$, which generates data $\boldsymbol{z} = Y$. We choose $k = 1$ and fix $X_1 = 2.556$, a value taken from the statistically steady state of the full dynamics. There are three parameters to learn from the fast dynamics: $(\theta_2, \theta_3, \theta_4) = (h, c, b)$. The one-point statistics $(\bar{Y}_1, \overline{Y_1^2})$ of the fast variables are not enough to recover all three. Therefore, we consider the moment function

$$\boldsymbol{g}(Y) = \begin{pmatrix} Y_{j,1} \\ Y_{j,1} Y_{j',1} \end{pmatrix}, \qquad j, j' \in \{1, \ldots, J\}, \tag{20}$$

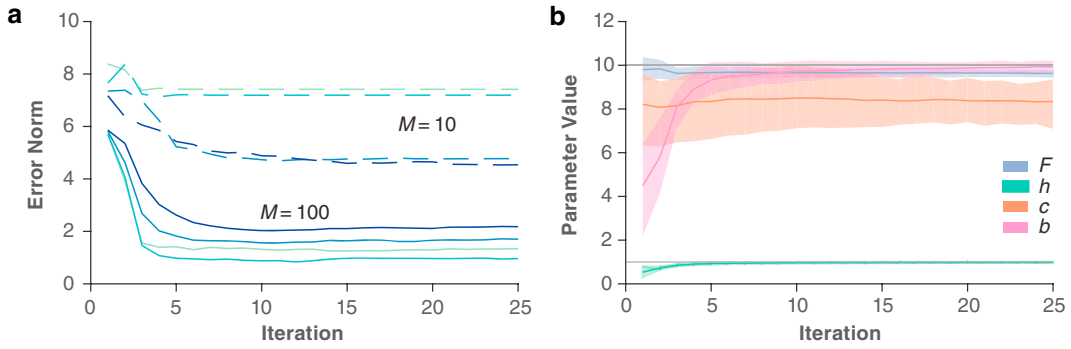

Figure 2. Convergence of ensemble Kalman inversion. (a) Error norm $\|\hat{\theta} - \tilde{\theta}\|$ of ensemble-mean parameter estimate $\hat{\theta}$ as a function of iteration, for different ensemble sizes M and noise levels r. Dashed lines for $M = 10$, and solid lines for $M = 100$. Color coding of noise levels $r = 0.1$, 0.2, 0.5, and 1.0 as in Figure 1 (top row). (b) Ensemble-mean value $\hat{\theta}_i$ of the four parameters as a function of iteration (solid lines), with interquartile range of ensemble (shading from 25th to 75th percentile), for $M = 100$ and $r = 0.5$. The objective function for each parameter setting is accumulated over $T = 100$ days.

containing all first moments and second moments (including cross-moments), giving a vector of length $J + J(J+1)/2 = 65$. We minimize the "high-resolution" objective function

$$J_s(\theta_2, \theta_3, \theta_4) = \frac{1}{2} \| \langle \boldsymbol{g}(Y) \rangle_{T'} - \langle \boldsymbol{g}(\tilde{Y}) \rangle_{\infty} \|^2_{\Sigma'}, \tag{21}$$

using a diagonal noise covariance matrix Σ' with diagonal elements proportional to the variances of the statistics in $\boldsymbol{g}$, with a noise level r' analogous to the noise covariance matrix (18). The variances of the statistics are estimated from a long control integration of the fast dynamics with fixed $X_1 = 2.556$. Because the fast variables Y evolve more rapidly than the slow variables X, we accumulate statistics over only $T' = 20$ days.

Bayesian inversion with RWM, with the same priors and algorithmic settings as before and with noise level $r' = 0.5$, again gives marginal posterior PDFs with modes close to the truth (Figure 1, bottom row). The posterior PDFs exhibit similar multimodality and reflect similar uncertainties and biases of posterior modes as those obtained from the full dynamics, especially with respect to the relatively large uncertainties in c (cf. Figure 1, middle row).

These examples illustrate the potential of learning about parameters from observations and from local high-resolution simulations under selected conditions (here for just one value of the slow variable X_1). An important question for future investigations is to what extent such results generalize to imperfect parameterization schemes, whose dynamics is usually not identical to the data-generating dynamics, so that structural in addition to parametric uncertainties arise. This issue can be studied for the Lorenz-96 system, for example, by using approximate models as parameterizations of the fast dynamics (e.g., Crommelin & Vanden-Eijnden, 2008; Fatkullin & Vanden-Eijnden, 2004; Wilks, 2005).

5. Outlook

Just as weather forecasts have made great strides over the past decades, thanks to improvements in the assimilation of observations (Bauer et al., 2015), climate projections can advance similarly by harnessing observations and modern computational capabilities more systematically. New methods from data assimilation, inverse problems, and machine learning make it possible to integrate observations and targeted high-resolution simulations in an ESM that learns from both and uses both to quantify uncertainties. As an objective of such parameter learning we propose the reduction of biases and exploitation of emergent constraints through the matching of mean values and covariance components between ESMs, observations, and targeted high-resolution simulations.

Coordinated space-based observations of crucial processes in the climate system are now available. For example, more than a decade's worth of coordinated observations of clouds, precipitation, temperature, and humidity with global coverage is available; parameterizations of clouds, convection, and turbulence can learn from them. Or simultaneous measurements of CO_2 concentrations and photosynthesis are becoming available; parameterizations of terrestrial ecosystems can learn from them. So far, such observations have been primarily used to evaluate models and identify their deficiencies. Their potential to improve models has not yet been harnessed. Additionally, it is feasible to conduct faithful local high-resolution simulations of processes such as the dynamics of clouds or sea ice, which are in principle computable but are too costly to compute globally. Parameterizations can also learn from such high-resolution simulations, either online by nesting them in an ESM or off-line by creating libraries of high-resolution simulations representing different regions and climates to learn from. Such a systematic approach to learning parameterizations from data allows the quantification of uncertainties in parameterizations, which in turn can be used to produce ensembles of climate simulations to quantify the uncertainty in predictions.

The machine learning of parameterizations in our view should be informed by the governing equations of subgrid-scale processes whenever they are known. The governing equations can be systematically coarse-grained, for example, by modeling the joint PDF of the relevant variables as a mixture of Gaussian kernels and generating moment equations for the modeled PDF from the governing equations (cf. Firl & Randall, 2015; Golaz et al., 2002; Guo et al., 2015; Lappen & Randall, 2001a). The closure parameters that necessarily arise in any such coarse graining of nonlinear governing equations can then be learned from a broad range of observations and high-resolution simulations, as parametric or nonparametric functions of ESM state variables (cf. Parish & Duraisamy, 2016). The fineness of the coarse graining (measured by the number of Gaussian kernels in the above example) can adapt to the information available to learn closure parameters.

Such equation-informed machine learning will provide a more versatile means of modeling subgrid-scale processes than the traditional approach of fixing closure parameters ad hoc or on the basis of a small sample of observations or high-resolution simulations. Because parameterizations learned within the structure of the known governing equations respect the relevant symmetries and conservation laws to within the closure approximations, they likely have greater out-of-sample predictive power than unstructured parameterization schemes, such as neural networks that are fit to subgrid-scale processes without explicit regard for symmetries and conservation laws (e.g., Krasnopolsky et al., 2013). Out-of-sample predictive power will be crucial if high-resolution simulations performed in selected locations and under selected conditions are to provide information globally and in changed climates. However, for noncomputable processes whose governing equations are unknown, like many ecological or biogeochemical processes, more empirical, data-driven parameterization approaches may well be called for.

An ESM that is designed from the outset to learn systematically from observations and high-resolution simulations represents an opportunity to achieve a leap in fidelity of parameterization schemes and thus of climate projections. Such an ESM can be expected to have attendant benefits for weather forecasting, because weather forecasting models and the atmospheric component of ESMs are essentially the same. However, challenges lie along the path toward realization of such an ESM:

1. We need innovation in learning algorithms. Our relatively simple example showed that parameters in a perfect-model setting can be learned effectively and efficiently by ensemble Kalman inversion. It remains to investigate questions such as the optimal ensemble size in Kalman inversions, how to adapt inversion algorithms to imperfect models, and how to quantify uncertainties. To increase computational efficiency, online filtering algorithms need to be developed that update parameters on the fly as Earth system statistics are being accumulated.
2. We need investigations of the best metrics to use when learning parameterization schemes from observations or high-resolution simulations. For example, are least-squares objective functions the best ones to use? Which covariance components or other statistics should be included in the objective functions? There are trade-offs between the number of covariance components that can be estimated from data and the information they can provide about parameterization schemes.
3. We need innovation in how learning from observations should interact with learning from targeted high-resolution simulations. How should high-resolution simulations be targeted? Where is the optimum trade-off between the added computational cost of conducting high-resolution simulations and the marginal information about parameterization schemes they provide?
4. We need innovation in parameterization schemes themselves, to design them such that they can learn effectively from diverse data sources and can be systematically refined when more information becomes available. It will be important to develop parameterizations that treat subgrid-scale motions (e.g., boundary layer turbulence, shallow convection, and deep convection) in a unified manner, to eliminate artificial spectral gaps that do not exist in nature and to reduce the number of correlated parameters in the schemes (e.g., Guo et al., 2015; Lappen & Randall, 2001a, 2001b; Köhler et al., 2011; Suselj et al., 2013; Park, 2014a, 2014b). Novel approaches that exploit ideas ranging from stochastic parameterization to systematic coarse-graining likely have roles to play here (e.g., Berner et al., 2017; Lucarini et al., 2014; Klein & Majda, 2006; Majda et al., 2003; Majda et al., 2008; Majda, 2012; Palmer & Williams, 2010; Palmer et al., 2005; Wouters et al., 2016; Wouters & Lucarini, 2013). Furthermore, as the resolution of ESMs increases, it will also be necessary to revisit the common practice of modeling subgrid-scale dynamics in grid columns, because the lateral exchange of subgrid-scale information across grid columns will play increasingly important roles.

The time is right to seize the opportunities that the available global observations and our computational resources present. Fundamentally reengineering atmospheric parameterization schemes, such as cloud and boundary layer parameterizations, will become a necessity as atmosphere models, within the next decade, reach horizontal grid spacings of 1–10 km and begin to resolve deep convection (Schneider et al., 2017). At such resolutions, common assumptions made in existing parameterization schemes, such as that clouds and the planetary boundary layer adjust instantaneously to changes in resolved-scale dynamics, will become untenable. Additionally, advances in high-performance computing (e.g., many-core computational architectures based on graphical processing units) will soon require a redesign of the software infrastructure of ESMs (Bretherton et al., 2012; Schulthess, 2015; Schalkwijk et al., 2015). So it is timely now to reengineer ESMs and parameterization schemes and design them from the outset so that they can learn systematically from observations and targeted high-resolution simulations.

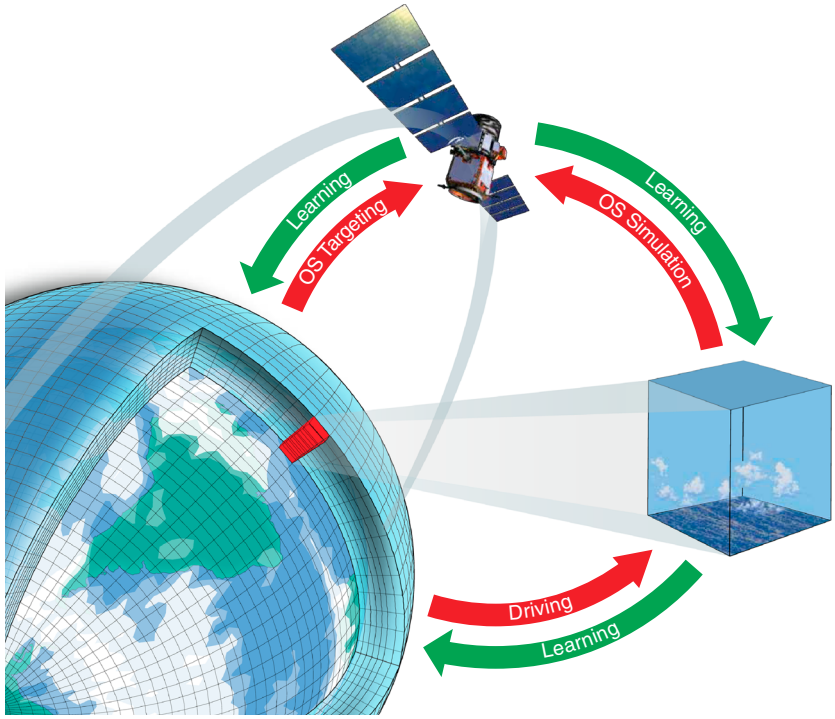

Figure 3. Schematic of Earth system modeling framework that integrates global observing systems (OS) and targeted high-resolution simulations. Not only can parameterizations in ESMs and high-resolution simulations learn from observations. The same framework and the observing system simulators that it needs to encompass can also be used for observing system simulation experiments to assess the value and benefits of new observing systems. (Image credit for satellite: NASA.)

Integrating observations and targeted high-resolution simulations in an Earth system modeling framework would have multiple attendant benefits. Solving the inverse problems of learning about parameterizations from observations requires observing system simulators that map model state variables to observables (Figure 3). The same observing system simulators, integrated in an Earth system modeling framework, can be used to answer questions about the value new observations would provide, for example, in terms of reduced uncertainties in ESMs. Addressing such questions in observing system simulation experiments (OSSEs) is increasingly required before the acquisition of new observing systems (e.g., as part of the U.S. Weather Research and Forecasting Innovation Act of 2017). They are naturally answered within the framework we propose.

Acknowledgments
We gratefully acknowledge financial support by Charles Trimble, by the Office of Naval Research (grant N00014-17-1-2079), and by the President's and Director's Fund of Caltech and the Jet Propulsion Laboratory. We also thank V. Balaji, Michael Keller, Dan McCleese, and John Worden for helpful discussions and comments on drafts, and Momme Hell for preparing Figure 3. The program code used in this paper is available at climate-dynamics.org/publications/. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Adam, O., Schneider, T., & Brient, F. (2017). Regional and seasonal variations of the double-ITCZ bias in CMIP5 models. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3909-1

Adam, O., Schneider, T., Brient, F., & Bischoff, T. (2016). Relation of the double-ITCZ bias to the atmospheric energy budget in climate models. *Geophysical Research Letters*, *43*, 7670–7677. https://doi.org/10.1002/2016GL069465

Aksoy, A., Zhang, F., & Nielsen-Gammon, J. W. (2006). Ensemble-based simultaneous state and parameter estimation with MM5. *Geophysical Research Letters*, *33*, L12801. https://doi.org/10.1029/2006GL026186

Alexanderian, A., Petra, N., Stadler, G., & Ghattas, O. (2016). A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. *SIAM Journal on Scientific Computing*, *38*, A243–A272.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*, 1283–1296. https://doi.org/10.1175/2009BAMS2618.1

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*, 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Annan, J. D., & Hargreaves, J. C. (2007). Efficient estimation and ensemble generation in climate modelling. *Philosophical Transactions of the Royal Society A*, *365*, 2077–2088. https://doi.org/10.1098/rsta.2007.2067

Ban, N., Schmidli, J., & Schär, C. (2015). Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, *42*, 1165–1172. https://doi.org/10.1002/2014GL062588

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*, 47–55. https://doi.org/10.1038/nature14956

Benedict, J. J., & Randall, D. A. (2009). Structure of the Madden-Julian oscillation in the superparameterized CAM. *Journal of Atmospheric Science*, *66*, 3277–3296. https://doi.org/10.1175/2009JAS3030.1

Berner, J., Achatz, U., Batte, L., Bengtsson, L., De La Camara, A., Christensen, H. M., … Yano, J.-I. (2017). Stochastic parameterization: Towards a new view of weather and climate models. *Bulletin of the American Meteorological Society*, *98*, 565–587. https://doi.org/10.1175/BAMS-D-15-00268.1

Beskos, A., Girolami, M., Lan, S., Farrell, P. E., & Stuart, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, *335*, 327–351. https://doi.org/10.1016/j.jcp.2016.12.041

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, *129*, 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 1285–1290. https://doi.org/10.5194/essd-5-165-2013

Bocquet, M., & Sakov, P. (2013). Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes in Geophysics*, *20*, 803–818. https://doi.org/10.5194/npg-20-803-2013

Bocquet, M., & Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, *140*, 1521–1535. https://doi.org/10.1002/qj.2236

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J.-L., … Yokohata, T. (2014). Origins of the solar radiation biases over the southern ocean in CFMIP2 models. *Journal of Climate*, *27*, 41–56. https://doi.org/10.1175/JCLI-D-13-00169.1

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., … Webb, M. J. (2006). How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, *19*, 3445–3482. https://doi.org/10.1175/JCLI3819.1

Bretherton, C., Balaji, V., Delworth, T., Dickinson, R. E., Edmonds, J. A., Famiglietti, J. S., … Smarr, L. L. (2012). *A national strategy for advancing climate modeling*. Washington, DC: The National Academies Press.

Brient, F., & Schneider, T. (2016). Constraints on climate sensitivity from space-based measurements of low-cloud reflection. *Journal of Climate*, *29*, 5821–5835. https://doi.org/10.1175/JCLI-D-15-0897.1

Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo* (619 pp.). Bora Raton, FL: Chapman and Hall/CRC.

Canadell, J. G., Quéré, C. L., Raupach, M. R., Field, C. B., Buitenhuis, E. T., & Ciais, P. (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 18,866–18,870. https://doi.org/10.1073 pnas.0702737104

Carrassi, A., Bocquet, M., Hannart, A., & Ghil, M. (2017). Estimating model evidence using data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*, 866–880. https://doi.org/10.1002/qj.2972

Cess, R. D., Potter, G., Blanchet, J., Boer, G., Ghan, S., Kiehl, J., … Yagai, I. (1989). Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. *Science*, *245*, 513–516.

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Del Genio, A. D., Déqué, M., … Zhang, M.-H. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, *95*, 16,601–16,615. https://doi.org/10.1029/JD095iD10p16601

Collins, M., Chandler, R. E., Cox, P. M., Huthnance, J. M., Rougier, J., & Stephenson, D. B. (2012). Quantifying future climate change. *Nature Climate Change*, *2*, 403–409. https://doi.org/10.1038/NCLIMATE1414

Cotter, S. L., Roberts, G. O., Stuart, A. M., & White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statistal Science*, *28*(3), 424–446.

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, *494*, 341–344. https://doi.org/10.1038/nature11882

Crisp, D., Atlas, R. M., Breon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., … Schroll, S. (2004). The orbiting carbon observatory (OCO) mission. *Advances in Space Research*, *34*, 700–709. https://doi.org/10.1016/j.asr.2003.08.062

Crommelin, D., & Vanden-Eijnden, E. (2008). Subgrid-scale parameterization with conditional Markov chains. *Journal of Atmospheric Science*, *65*, 2661–2675. https://doi.org/10.1175/2008JAS2566.1

de Rooy, W. C., Bechtold, P., Frohlich, K., Hohenegger, C., Jonker, H., Mironov, D., … Yano, J.-I. (2013). Entrainment and detrainment in cumulus convection: An overview. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1–19.

Dee, D. P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, *123*, 1128–1145. https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3323–3343. https://doi.org/10.1256/qj.05.137

Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B*, *68*, 411–436. https://doi.org/10.1111/j.1467-9868.2006.00553.x

DeMott, C. A., Stan, C., & Randall, D. A. (2013). Northward propagation mechanisms of the boreal summer intraseasonal oscillation in the ERA-Interim and SP-CCSM. *Journal of Climate*, *26*, 1973–1992. https://doi.org/10.1175/JCLI-D-12-00191.1

Devenish, B. J., Bartello, P., Brenguier, J.-L., Collins, L. R., Grabowski, W. W., IJzermans, R. H. A., … Warhaft, Z. (2012). Droplet growth in warm turbulent clouds. *Quarterly Journal of the Royal Meteorological Society*, *138*, 1401–1429. https://doi.org/10.1002/qj.1897

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B*, *57*, 45–97.

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., … Weir, B. (2017). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science*, *358*, eaam5745. https://doi.org/10.1126/science.aam5745

Emanuel, K. A., & Živković Rothman, M. (1999). Development and evaluation of a convection scheme for use in climate models. *Journal of Atmospheric Science*, *56*, 1766–1782. https://doi.org/10.1175/1520-0469(1999)056<1766:DAEOAC>2.0.CO;2

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems* (321 pp.). Dordrecht: Kluwer Academic.

Fatkullin, I., & Vanden-Eijnden, E. (2004). A computational strategy for multiscale systems with applications to Lorenz 96 model. *Journal of Computational Physics*, *200*, 605–638. https://doi.org/10.1016/j.jcp.2004.04.013

Firl, G. J., & Randall, D. A. (2015). Fitting and analyzing LES using multiple trivariate Gaussians. *Journal of Atmospheric Science*, *72*, 1094–1116. https://doi.org/10.1175/JAS-D-14-0192.1

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., … Rummukainen, M. (2013). Evaluation of climate models. In T. F. Stocker et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (chap. 9, pp. 741–853). Cambridge, UK, and New York: Cambridge University Press.

Fox-Kemper, B., Bachman, S., Pearson, B., & Reckinger, S. (2014). Principles and advances in subgrid modelling for eddy-rich simulations. *CLIVAR Exchanges No.65*, *19*(2), 42–46.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., … Yokota, T. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*, L17706. https://doi.org/10.1029/2011GL048738

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., … Taylor, T. E. (2014). Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sensing of Environment*, *147*, 1–12. https://doi.org/10.1016/j.rse.2014.02.007

Friedlingstein, P. (2015). Carbon cycle feedbacks and future climate change. *Philosophical Transactions of the Royal Society A*, *373*, 20140421. https://doi.org/10.1098/rsta.2014.0421

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., … Zeng, N. (2006). Climate–carbon cycle feedback analysis: Results from the C_{4MIP} model intercomparison. *Journal of Climate*, *19*, 3337–3353. https://doi.org/10.1175/JCLI3800.1

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*, 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1

Friend, A. D., Lucht, W., Rademacher, T. T., Keribin, R., Betts, R., Cadule, P., … Woodward, F. I. (2014). Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO_2. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 3280–3285. https://doi.org/10.1073/pnas.1222477110

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., … Paul, F. (2013). A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. *Science*, *340*, 852–857. https://doi.org/10.1126/science.1234532

Golaz, J.-C., Horowitz, L. W., & Levy II, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, *40*, 2246–2251. https://doi.org/10.1002/grl.50232

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

Grabowski, W. W. (2001). Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP). *Journal of Atmospheric Science*, *58*, 978–997. https://doi.org/10.1175/1520-0469(2001)058<0978:CCPWTL>2.0.CO;2

Grabowski, W. W. (2016). Towards global large eddy simulation: Super-parameterization revisited. *Journal of the Meteorological Society of Japan*, *94*, 327–344. https://doi.org/10.2151/jmsj.2016-017

Grabowski, W. W., & Smolarkiewicz, P. K. (1999). CRCP: A cloud resolving convection parameterization for modeling the tropical convecting atmosphere. *Physica D*, *133*, 171–178. https://doi.org/10.1016/S0167-2789(99) 00104-9

Grabowski, W. W., & Wang, L.-P. (2013). Growth of cloud droplets in a turbulent environment. *Annual Review of Fluid Mechanics*, *45*, 293–324. https://doi.org/10.1146/annurev-fluid-011212-140750

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 1693. https://doi.org/10.1029/2002GL015311

Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., … Stockdale, T. (2009). Understanding El Niño in ocean-atmosphere general circulation models. *Bulletin of the American Meteorological Society*, *90*, 325–340. https://doi.org/10.1175/2008BAMS2387.1

Guo, H., Golaz, J.-C., Donner, L. J., Wyman, B., Zhao, M., & Ginoux, P. (2015). CLUBB as a unified cloud parameterization: Opportunities and challenges. *Geophysical Research Letters*, *42*, 4540–4547. https://doi.org/10.1002/2015GL063672

Hall, A., & Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, *33*, L03502. https://doi.org/10.1029/2005GL025127

Hohenegger, C., & Bretherton, C. S. (2011). Simulating deep convection with a shallow convection scheme. *Atmospheric Chemistry and Physics*, *11*, 10,389–10,406. https://doi.org/10.5194/acp-11-10389-2011

Holloway, C. E., & Neelin, J. D. (2009). Moisture vertical structure, column water vapor, and tropical deep convection. *Journal of Atmospheric Science*, *66*, 1665–1683.

Hope, C. (2015). The $10 trillion value of better information about the transient climate response. *Philosophical Transactions of the Royal Society A*, *373*, 20140429. https://doi.org/10.1098/rsta.2014.0429

Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F., … Roehrig, R. (2013). LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, *40*, 2193–2222. https://doi.org/10.1007/s00382-012-1343-y

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., … Williamson, D. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*, 589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*, 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1

Hung, M.-P., Lin, J.-L., Wang, W., Kim, D., Shinoda, T., & Weaver, S. J. (2013). MJO and convectively coupled equatorial waves simulated by CMIP5 climate models. *Journal of Climate*, *26*, 6185–6214. https://doi.org/10.1175/JCLI-D-12-00541.1

Iglesias, M. A. (2016). A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Problems*, *32*(025), 002. https://doi.org/10.1088/0266-5611/32/2/025002

Iglesias, M. A., Law, K. J. H., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse Problems*, *29*, 045001. https://doi.org/10.1088/0266-5611/29/4/045001

Intergovernmental Panel on Climate Change (2013). *Climate change 2013: The physical science basis*. New York: Cambridge University Press.

Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., & Bowman, K. P. (2008). Error reduction and convergence in climate prediction. *Journal of Climate*, *21*, 6698–6709. https://doi.org/10.1175/2008JCLI2112.1

Jakob, C. (2003). An improved strategy for the evaluation of cloud parameterizations in GCMs. *Bulletin of the American Meteorological Society*, *84*, 1387–1401. https://doi.org/10.1175/BAMS-84-10-1387

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bulletin of the American Meteorological Society*, *91*, 869–875. https://doi.org/10.1175/2009BAMS2898.1

Jarvinen, H., Raisanen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., … Haario, H. (2010). Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmospheric Chemistry and Physics*, *10*, 9993–10002. https://doi.org/10.5194/acp-10-9993-2010

Jiang, J. H., Su, H., Zhai, C., Perun, V. S., Genio, A. D., Nazarenko, L. S., … Stephens, G. L. (2012). Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. *Journal of Geophysical Research*, *117*, D14105. https://doi.org/10.1029/2011JD017237

Joiner, J., Yoshida, Y., Vasilkov, A., Corp, L., & Middleton, E. (2011). First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences*, *8*, 637–651. https://doi.org/10.5194/bg-8-637-2011

Kaipio, J., & Somersalo, E. (2005). *Statistical and computational inverse problems* (Vol. 160). New York: Springer.

Karlsson, J., & Svensson, G. (2013). Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble. *Geophysical Research Letters*, *40*, 4374–4379. https://doi.org/10.1002/grl.50768

Karlsson, J., Svensson, G., & Rodhe, H. (2008). Cloud radiative forcing of subtropical low level clouds in global models. *Climate Dynamics*, *30*, 779–788. https://doi.org/10.1007/s00382-007-0322-1

Kasahara, A., & Washington, W. M. (1967). NCAR global general circulation model of the atmosphere. *Monthly Weather Review*, *95*, 389–402.

Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., & Bitz, C. (2016). Global climate impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model (CESM). *Journal of Climate*, *29*, 4617–4636. https://doi.org/10.1175/JCLI-D-15-0358.1

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B*, *63*, 425–464. https://doi.org/10.1111/1467-9868.00294

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., … Wofsy, S. C. (2012). The imprint of surface fluxes and transport on variations in total column carbon dioxide. *Biogeosciences*, *9*, 875–891.

Khairoutdinov, M., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of Atmospheric Science*, *62*, 2136–2154. https://doi.org/10.1175/JAS3453.1

Khairoutdinov, M. F., Krueger, S. K., Moeng, C.-H., Bogenschutz, P. A., & Randall, D. A. (2009). Large-eddy simulation of maritime deep tropical convection. *Journal of Advances in Modeling Earth Systems*, *1*, 15. https://doi.org/10.3894/JAMES.2009.1.15

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

Klein, R., & Majda, A. J. (2006). Systematic multiscale models for deep convection on mesoscales. *Theoretical and Computational Fluid Dynamics*, *20*, 525–551. https://doi.org/10.1007/s00162-006-0027-9

Klein, S. A., & Hall, A. (2015). Emergent constraints for cloud feedbacks. *Current Climate Change Reports*, *1*, 276–287. https://doi.org/10.1007/s40641-015-0027-1

Klocke, D., & Rodwell, M. J. (2014). A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Quarterly Journal of the Royal Meteorological Society*, *140*, 517–524. https://doi.org/10.1002/qj.2172

Knorr, W. (2009). Is the airborne fraction of anthropogenic CO_2 emissions increasing? *Geophysical Research Letters*, *36*, L21710. https://doi.org/10.1029/2009GL040613

Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., … Wigley, T. M. L. (2008). A review of uncertainties in global temperature projections over the twenty-first century. *Journal of Climate*, *21*, 2651–2663. https://doi.org/10.1175/2007JCLI2119.1

Köhler, M., Ahlgrimm, M., & Beljaars, A. (2011). Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, *137*, 43–57. https://doi.org/10.1002/qj.713

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Belochitski, A. A. (2013). Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. *Advances in Artificial Neural Systems*, *2013*, 485913. https://doi.org/10.1155/2013/485913

Lappen, C.-L., & Randall, D. A. (2001a). Toward a unified parameterization of the boundary layer and moist convection. Part I: A new type of mass-flux model. *Journal of Atmospheric Sciences*, *58*, 2021–2036.

Lappen, C.-L., & Randall, D. A. (2001b). Toward a unified parameterization of the boundary layer and moist convection. Part II: Lateral mass exchanges and subplume-scale fluxes. *Journal of Atmospheric Sciences*, *58*, 2037–2051.

Law, K., Stuart, A., & Zygalakis, K. (2015). Data Assimilation: A Mathematical Introduction. In *Texts in Applied Mathematics* (Vol. 62). Cham: Springer.

Law, K. J. H., & Stuart, A. M. (2012). Evaluating data assimilation algorithms. *Monthly Weather Review*, *140*, 3757–3782. https://doi.org/10.1175/MWR-D-11-00257.1

Le Queéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., … Zeng, N. (2013). The global carbon budget 1959–2011. *Earth System Science Data*, *5*, 165–185. https://doi.org/10.5194/essd-5-165-2013

L'Ecuyer, T. S., Beaudoing, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., … Hilburn, K. (2015). The observed state of the energy budget in the early twenty-first century. *Journal of Climate*, *28*, 8319–8346. https://doi.org/10.1175/JCLI-D-14-00556.1

Li, G., & Xie, S.-P. (2014). Tropical biases in CMIP5 multi-model ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *Journal of Climate*, *27*, 1765–1780. https://doi.org/10.1175/JCLI-D-13-00337.1

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

Lin, J.-L., Qian, T., & Shinoda, T. (2014). Stratocumulus clouds in Southeastern Pacific simulated by eight CMIP5–CFMIP global climate models. *Journal of Climate*, *27*, 3000–3022. https://doi.org/10.1175/JCLI-D-13-00376.1

Liu, C., Moncrieff, M. W., & Grabowski, W. W. (2001). Hierarchical modelling of tropical convective systems using explicit and parametrized approaches. *Quarterly Journal of the Royal Meteorological Society*, *127*, 493–515.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., … Eldering, A. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science*, *358*, eaam5690. https://doi.org/10.1126/science.aam5690

Lorenz, E. N. (1996). Predictability—A problem partly solved. In Reprinted in T. N. Palmer & R. Hagedorn (Eds.), *Proceedings Seminar on Predictability, Predictability of Weather and Climate, Cambridge UP (2006)* (Vol. 1, pp. 1–18). Reading, Berkshire, UK: ECMWF.

Lorenz, E. N., & Emanuel, K. A. (1998). Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of Atmospheric Science*, *55*, 399–414. https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2

Lucarini, V., Blender, R., Herbert, C., Ragone, F., Pascale, S., & Wouters, J. (2014). Mathematical and physical ideas for climate science. *Reviews of Geophysics*, *52*, 809–859. https://doi.org/10.1002/2013RG000446

Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., & Zhang, Y. (2013). Metrics and diagnostics for precipitation-related processes in climate model short-range hindcasts. *Journal of Climate*, *26*, 1516–1534. https://doi.org/10.1175/JCLI-D-12-00235.1

Majda, A. J. (2012). Challenges in climate science and contemporary applied mathematics. *Communications on Pure and Applied Mathematics*, *65*, 920–948. https://doi.org/10.1002/cpa.21401

Majda, A. J., Franzke, C., & Khouider, B. (2008). An applied mathematics perspective on stochastic modelling for climate. *Philosophical Transactions of the Royal Society A*, *366*, 2429–2455. https://doi.org/10.1098/rsta.2008.0012

Majda, A. J., Timofeyev, I., & Vanden-Eijnden, E. (2003). Systematic strategies for stochastic mode reduction in climate. *Journal of Atmospheric Science*, *60*, 1705–1722.

Manabe, S., Smagorinsky, J., & Strickler, R. F. (1965). Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, *93*, 769–798.

Matheou, G., & Chung, D. (2014). Large-eddy simulation of stratified turbulence. Part II: Application of the stretched-vortex model to the atmospheric boundary layer. *Journal of Atmospheric Science*, *71*, 4439–4460. https://doi.org/10.1175/JAS-D-13-0306.1

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., … Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*, M00A01. https://doi.org/10.1029/2012MS000154

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., … Allen, M. R. (2009). Greenhouse-gas emission targets for limiting global warming to 2°C. *Nature*, *458*, 1158–1162. https://doi.org/10.1038/nature08017

Mintz, Y. (1965). Very long-term global integration of the primitive equations of atmospheric motion. In *WMO-IUGG Symposium on Research and Development Aspects of Long-Range Forecasting, Boulder, Colo.* (Vol. 1964, pp. 141–155). Geneva: World Meteorological Organization.

Moeng, C.-H., Dudhia, J., Klemp, J., & Sullivan, P. (2007). Examining two-way grid nesting for large eddy simulation of the PBL using the WRF Model. *Monthly Weather Review*, *135*, 2295–2311. https://doi.org/10.1175/MWR3406.1

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

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., & Meyerson, J. E. (2010). Considerations for parameter optimization and sensitivity in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21,349–21,354. https://doi.org/10.1073/pnas.1015473107

Neelin, J. D., Peters, O., & Hales, K. (2009). The transition to strong convection. *Journal of Atmospheric Science*, *66*, 2367–2384.

Neggers, R. A. J., Siebesma, A. P., & Heus, T. (2012). Continuous single-column model evaluation at a permanent meteorological supersite. *Bulletin of the American Meteorological Society*, *93*, 1389–1400. https://doi.org/10.1175/BAMS-D-11-00162.1

Nie, J., & Kuang, Z. (2012). Temperature and moisture perturbations: a comparison of large-eddy simulations and a convective parameterization based on stochastically entraining parcels. *Journal of Atmospheric Science*, *69*, 1936–1956.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). Springer Series in Operations Research. New York: Springer.

Ohno, T., Satoh, M., & Yamada, Y. (2016). Warm cores, eyewall slopes, and intensities of tropical cyclones simulated by a 7-km-mesh global nonhydrostatic model. *Journal of Atmospheric Science*, *73*, 4289–4309. https://doi.org/10.1175/JAS-D-15-0318.1

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., … Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, *56*, 415–428. https://doi.org/10.1111/j.1600-0870.2004.00076.x

Palmer, T. (2014). Build high-resolution global climate models. *Nature*, *515*, 338–339. https://doi.org/10.1038/515338a

Palmer, T., & Williams, P. (2010). *Stochastic physics and climate modelling* (480 pp.). Cambridge, UK: Cambridge University Press.

Palmer, T. N., Gelaro, R., Barkmeijer, J., & Buizza, R. (1998). Singular vectors, metrics, and adaptive observations. *Journal of Atmospheric Science*, *55*, 633–653. https://doi.org/10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., & Leutbecher, M. (2005). Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, *33*, 163–193. https://doi.org/10.1146/annurev.earth.33.092203.122552

Parish, E. J., & Duraisamy, K. (2016). A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics*, *305*, 758–774. https://doi.org/10.1016/j.jcp.2015.11.012

Parishani, H., Pritchard, M. S., Bretherton, C. S., Wyant, M. C., & Khairoutdinov, M. (2017). Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *Journal of Advances in Modeling Earth Systems*, *9*, 1542–1571. https://doi.org/10.1002/2017MS000968

Park, S. (2014a). A unified convection scheme (UNICON). Part I: Formulation. *Journal of Atmospheric Science*, *71*, 3902–3930.

Park, S. (2014b). A unified convection scheme (UNICON). Part II: Simulation. *Journal of Atmospheric Science*, *71*, 3931–3973.

Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., … Yio, J. J. (2004). Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bulletin of the American Meteorological Society*, *85*, 1903–1915. https://doi.org/10.1175/BAMS-85-12-1903

Pressel, K. G., Kaul, C. M., Schneider, T., Tan, Z., & Mishra, S. (2015). Large-eddy simulation in an anelastic framework with closed water and entropy balances. *Journal of Advances in Modeling Earth Systems*, *7*, 1425–1456. https://doi.org/10.1002/2015MS000496

Pressel, K. G., Mishra, S., Schneider, T., Kaul, C. M., & Tan, Z. (2017). Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *Journal of Advances in Modeling Earth Systems*, *9*, 1342–1365. https://doi.org/10.1002/2016MS000778

Pritchard, M. S., & Somerville, R. C. J. (2009a). Empirical orthogonal function analysis of the diurnal cycle of precipitation in a multi-scale climate model. *Geophysical Research Letters*, *36*, L05812. https://doi.org/10.1029/2008GL036964

Pritchard, M. S., & Somerville, R. C. J. (2009b). Assessing the diurnal cycle of precipitation in a multi-scale climate model. *Journal of Advances in Modeling Earth Systems*, *1*, 12. https://doi.org/10.3894/JAMES.2009.1.12

Qu, X., Hall, A., Klein, S. A., & Caldwell, P. M. (2014). On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Climate Dynamics*, *42*, 2603–2626. https://doi.org/10.1007/s00382-013-1945-z

Qu, X., Hall, A., Klein, S. A., & DeAngelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*, 7767–7775. https://doi.org/10.1002/2015GL065627

Randall, D. A. (2013). Beyond deadlock. *Geophysical Research Letters*, *40*, 5970–5976. https://doi.org/10.1002/2013GL057998

Randall, D. A., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*, 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Randall, D. A., & Wielicki, B. A. (1997). Measurements, models, and hypotheses in the atmospheric sciences. *Bulletin of the American Meteorological Society*, *78*, 400–406.

Rodwell, M. J., & Palmer, T. N. (2007). Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, *133*, 129–146. https://doi.org/10.1002/qj.23

Romps, D. M. (2016). The stochastic parcel model: A deterministic parameterization of stochastically entraining convection. *Journal of Advances in Modeling Earth Systems*, *8*, 319–344. https://doi.org/10.1002/2015MS000537

Romps, D. M., & Kuang, Z. (2010). Nature versus nurture in shallow convection. *Journal of Atmospheric Science*, *67*, 1655–1666.

Ruiz, J., & Pulido, M. (2015). Parameter estimation using ensemble-based data assimilation in the presence of model error. *Monthly Weather Review*, *143*, 1568–1582. https://doi.org/10.1175/MWR-D-14-00017.1

Ruiz, J. J., Pulido, M., & Miyoshi, T. (2013). Estimating model parameters with ensemble-based data assimilation: A review. *Journal of the Meteorological Society of Japan*, *91*, 79–99. https://doi.org/10.2151/jmsj.2013-201

Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., & Van Meijgaard, E. (2015). Weather forecasting using GPU-based large-eddy simulations. *Bulletin of the American Meteorological Society*, *96*, 715–723. https://doi.org/10.1175/BAMS-D-14-00114.1

Schirber, S., Klocke, D., Pincus, R., Quaas, J., & Anderson, J. L. (2013). Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *Journal of Advances in Modeling Earth Systems*, *5*, 58–70. https://doi.org/10.1029/2012MS000167

Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*, 3–5. https://doi.org/10.1038/nclimate3190

Schulthess, T. C. (2015). Programming revisited. *Nature Physics*, *11*, 369–373.

Shepherd, A., Ivins, E. R., Geruo, A., Barletta, V. R., Bentley, M. J., Bettadpur, S., … Zwally, J. (2012). A reconciled estimate of ice-sheet mass balance. *Science*, *338*, 1183–1189. https://doi.org/10.1126/science.1228102

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., … Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of Atmospheric Science*, *60*, 1201–1219.

Siebesma, A. P., Soares, P. M. M., & Teixeira, J. (2007). A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of Atmospheric Science*, *64*, 1230–1248. https://doi.org/10.1175/JAS3888.1

Siler, N., Po-Chedley, S., & Bretherton, C. S. (2017). Variability in modeled cloud feedback tied to differences in the climatological spatial pattern of clouds. *Climate Dynamics*, *48*(305), 1–12. https://doi.org/10.1007/s00382-017-3673-2

Simmons, A., Fellous, J.-L., Ramaswamy, V., Trenberth, K., Asrar, G., Balmaseda, M., … Shepherd, T. (2016). Observation and integrated Earth-system science: A roadmap for 2016–2025. *Advances in Space Research*, *57*, 2037–2103. https://doi.org/10.1016/j.asr.2016.03.008

Smagorinsky, J. (1963). General circulation experiments with the primitive equations. I. The basic experiment. *Monthly Weather Review*, *91*, 99–164.

Smagorinsky, J., Manabe, S., & Holloway Jr., J. L. (1965). Numerical results from a nine-level general circulation model of the atmosphere. *Monthly Weather Review*, *93*, 727–768.

Soden, B. J., & Held, I. M. (2006). An assessment of climate feedbacks in coupled ocean-atmosphere models. *Journal of Climate*, *19*, 3354–3360. https://doi.org/10.1175/JCLI3799.1

Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J., & Järvinen, H. (2012). Efficient MCMC for climate model parameter estimation: Parallel adaptive chains and early rejection. *Bayesian Analysis*, *7*, 715–736. https://doi.org/10.1214/12-BA724

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., … Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, *433*, 403–406.

Stan, C., Khairoutdinov, M., DeMott, C. A., Krishnamurthy, V., Straus, D. M., Randall, D. A., … Shukla, J. (2010). An ocean-atmosphere climate simulation with an embedded cloud resolving model. *Geophysical Research Letters*, *37*, L01702. https://doi.org/10.1029/2009GL040822

Stensrud, D. J. (2007). *Parameterization schemes: Keys to understanding numerical weather prediction models* (477 pp.). Cambridge, UK: Cambridge University Press.

Stevens, B., Moeng, C.-H., Ackerman, A. S., Bretherton, C. S., Chlond, A., de Roode, S., … Zhu, P. (2005). Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Monthly Weather Review*, *133*, 1443–1462. https://doi.org/10.1175/MWR2930.1

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., … Lebsock, M. (2017). CloudSat and CALIPSO within the A-train: Ten years of actively observing the earth system. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-16-0324.1

Stephens, G. L. (2005). Cloud feedbacks in the climate system: A critical review. *Journal of Climate*, *18*, 237–273. https://doi.org/10.1175/JCLI-3243.1

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., … Cloudsat Science Team (2002). The CloudSat mission and the A-train. *Bulletin of the American Meteorological Society*, *83*, 1771–1790. https://doi.org/10.1175/BAMS-83-12-1771

Stewart, L., Dance, S. L., Nichols, N. K., Eyre, J., & Cameron, J. (2014). Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, *140*, 1236–1244. https://doi.org/10.1002/qj.2211

Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., … Yuen, K. (2017). OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science*, *358*, eaam5747. https://doi.org/10.1126/science.aam5747

Suselj, K., Teixeira, J., & Chung, D. (2013). A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *Journal of Atmospheric Science*, *70*, 1929–1953.

Suzuki, K., Golaz, J.-C., & Stephens, G. L. (2013). Evaluating cloud tuning in a climate model with satellite observations. *Geophysical Research Letters*, *40*, 4463–4468. https://doi.org/10.1002/grl.50874

Swanson, K. L., & Pierrehumbert, R. T. (1997). Lower-tropospheric heat transport in the Pacific storm track. *Journal of Atmospheric Science*, *54*, 1533–1543.

Tett, S. F. B., Mineter, M. J., Cartis, C., Rowlands, D. J., & Liu, P. (2013). Can top-of-atmosphere radiation measurements constrain climate predictions? Part I: Tuning. *Journal of Climate*, *26*, 9348–9366. https://doi.org/10.1175/JCLI-D-12-00595.1

Tian, B. (2015). Spread of model climate sensitivity linked to double-intertropical convergence zone bias. *Geophysical Research Letters*, *42*, 4133–4141. https://doi.org/10.1002/2015GL064119

Todd-Brown, K., Randerson, J., Post, W., Hoffman, F., Tarnocai, C., Schuur, E., & Allison, S. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, *10*, 1717–1736. https://doi.org/10.5194/bg-10-1717-2013

Vaughan, D. G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., … Zhang, T. (2013). Observations: Cryosphere, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 317–382).

Vial, J., Dufresne, J.-L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, *41*, 3339–3362. https://doi.org/10.1007/s00382-013-1725-9

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

Wang, Q., Hu, R., & Blonigan, P. (2014). Least squares shadowing sensitivity analysis of chaotic limit cycle oscillations. *Journal of Computational Physics*, *267*, 210–224. https://doi.org/10.1016/j.jcp.2014.03.002

Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models atmospheric climate models. *Climate Dynamics*, *17*, 905–922.

Webb, M. J., Lambert, F. H., & Gregory, J. M. (2013). Origins of differences in climate sensitivity, forcing and feedback in climate models. *Climate Dynamics*, *40*, 677–707. https://doi.org/10.1007/s00382-012-1336-x

Weinan, E., Engquist, B., Li, X., Ren, W., & Vanden-Eijnden, E. (2007). Heterogeneous multiscale methods: A review. *Communications in Computational Physics*, *3*, 367–450.

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2014). Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Biogeosciences*, *119*, 794–807.

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2016). Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO_2. *Nature*, *538*, 499–501. https://doi.org/10.1038/nature19772

Wilks, D. S. (2005). Effects of stochastic parametrizations in the Lorenz '96 system. *Quarterly Journal of the Royal Meteorological Society*, *131*, 389–407. https://doi.org/10.1256/qj.04.03

Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, *140*, 2373–2423. https://doi.org/10.1175/MWR-D-11-00121.1

Wouters, J., & Lucarini, V. (2013). Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach. *Journal of Statistical Physics*, *151*, 850–860. https://doi.org/10.1007/s10955-013-0726-8

Wouters, J., Dolaptchiev, S. I., & Lucarini, V. (2016). Parameterization of stochastic multiscale triads. *Nonlinear Processes in Geophysics*, *23*, 435–445. https://doi.org/10.5194/npg-23-435-2016

Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., & Zhang, Y. (2012). On the correspondence between short- and long-time-scale systematic errors in CAM4/CAM5 for the Year of Tropical Convection. *Journal of Climate*, *25*, 7937–7955. https://doi.org/10.1175/JCLI-D-12-00134.1

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., & Maksyutov, S. (2009). Global concentrations of CO_2 and CH_4 retrieved from GOSAT: First preliminary results. *SOLA*, *5*, 160–163.

Zhang, M. H., Lin, W. Y., Klein, S. A., Bacmeister, J. T., Bony, S., Cederwall, R. T., … Zhang, J. H. (2005). Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *Journal of Geophysical Research*, *110*, D15S02. https://doi.org/10.1029/2004JD005021

Zhang, X., Liu, H., & Zhang, M. (2015). Double ITCZ in coupled ocean-atmosphere models: From CMIP3 to CMIP5. *Geophysical Research Letters*, *42*, 8651–8659. https://doi.org/10.1002/2015GL065973

Zhao, M., Held, I. M., Ramaswamy, V., Lin, S.-J., Ming, Y., Ginoux, P., … Paynter, D. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, *29*, 543–560. https://doi.org/10.1175/JCLI-D-15-0191.1

Zhu, P., Albrecht, B. A., Ghate, V. P., & Zhu, Z. (2010). Multiple-scale simulations of stratocumulus clouds. *Journal of Geophysical Research*, *115*, D23201. https://doi.org/10.1029/2010JD014400