A missing source of uncertainty: forcing-dependent model parameter sensitivity

Xiuhua Zhu
CEN, University of Hamburg, Bundesstraße 53, 20146 Hamburg, Germany
E-mail: xiuhua.zhu@uni-hamburg.de

Keywords: model uncertainty, parameter sensitivity, climate forcing, CMIP, surface temperature

Abstract

Climate modelling groups usually work with a handful of model versions (parameter combinations) that reproduce certain targeted aspects of observed climate within a certain validity range and apply them for studying future climate change. What is of concern is whether these retained model versions, with respect to their de-selected counterparts, continue being optimal for future climate that is supposed to distinctly differ from the present one. Extrapolating model performances beyond their validity range requires model parameter sensitivity (i.e., changes in model output due to changes in model parameters) remains more or less stationary despite different forcing conditions. This requirement, however, is shown to be ill-grounded by exemplified analyses of resolution sensitivity in an Earth System Model under different forcing conditions, whereby model resolution is handled as a model parameter in a wider sense: (i) Model resolution sensitivity depends on the forcing conditions applied; moreover, (ii) The further the forcing deviates from a reference state, the earlier one can detect a systematic change in model resolution sensitivity, in particular, in its spatial details. Therefore, model parameter sensitivity and forcing conditions should be evaluated as a compound; failure to account for this relation leads to a systematic underestimation of uncertainty in forced responses of climate models, thereby imposing hazardous impacts on practical applications of CMIP outputs.

Introduction

Uncertainty assessment is crucial to model-based climate change research, because it determines the confidence one may assign to simulated climate signals and consequently how these signals should be distributed under practical circumstances. There are three primary sources of uncertainty: internal climate variability (ICV), forcing uncertainty, and model uncertainty (Stainforth et al 2007, Hawkins and Sutton 2009, Yip et al 2011, Knutti and Sedláček 2013, Wang et al 2015, Woldemeskel et al 2016). The first two describe uncertainty in initial and boundary conditions that serve as inputs for climate models, with the former stemming from the chaotic nature of the system (Smith et al 1999, Palmer et al 2005), and the latter dealing with unknowns in forecasting anthropogenic emissions, aerosol emissions, land-use changes, and solar variability for the future. These two sources of uncertainty lie outside of the climate model development scope and cannot be significantly reduced. In contrast, model uncertainty can be better constrained by improving our understanding of climate processes, based on which models are constructed and validated.

Model uncertainty comprises structural and parametric uncertainty and is often studied in a dichotomous manner. The former is addressed by means of multi-model ensemble, and the latter by means of perturbed physics ensemble using a single model (manifested by changing model parameters). Since all models are merely abstraction, simplification and interpretation of reality, no matter how detailed and physically reasonable a model’s structure is, structural uncertainty will not be significantly reduced and thus is also referred to as model inadequacy (Stainforth et al 2007). Nevertheless, assuming that the collective response of multi-models will
largely cancel out the structural imperfection of each individual model and thereby achieve a better representation of the actual system (Hagedorn *et al.* 2005), the Coupled Model Intercomparison Project (CMIP) employs multi-model ensemble to sample model structural uncertainty while leaving the parametric uncertainty almost unrepresented, because all models are provided with model parameters fixed and parameterization schemes predetermined.

A great example of applying perturbed physics ensemble for studying model uncertainty is a multi-thousand-member grand ensemble of simulations that allow combinations of perturbations in six model parameters in a general circulation model. The resulting extreme climate sensitivity ranges from 1.9 K to more than 11 K (Stainforth *et al.* 2005) and is almost three times the range projected by climate models that participate in the CMIP-phase 5 (2.1 K–4.7 K, Andrews *et al.* 2012). This large discrepancy unambiguously points out the drastic underestimation of model uncertainty already on the level of individual models. Unsurprisingly, such high sensitivity to model physics (parameterizations) has also been revealed in the recent CMIP6 (e.g., Zelinka *et al.* 2020).

In the current CMIP framework, decisions on model structure, such as what subsystems to include (e.g., the carbon cycle, stratospheric dynamics), are often reached first. Subsequently, model parameters and parameterization schemes are tuned and calibrated to reproduce the preindustrial or present-day climate with an acceptable accuracy (the tuning climate, Mauritsen *et al.* 2012, Hourdin *et al.* 2017). The model setup that generates the ‘optimal’ fitting is then used for future climate projections, which in fact extrapolates the model’s ‘optimal’ performance beyond its validity condition into a new climate that is supposed to greatly differ from the tuning climate. Such an extrapolation requires model parameter sensitivity that determines relative contributions of each parameter to changes in model outputs to be independent of the climates or the forcing conditions so that the ‘optimal’ remains optimal and the ‘poor’ remains poor. However, whether or not the above-assumed independence holds has so far seldom been examined.

This analysis attempts to elicit whether model parameter sensitivity varies when exposed to a different forcing by examining model resolution sensitivity (RS) in surface temperature (ST) simulated by the Max Planck Institute Earth System Model (MPI-ESM). By doing so, model resolution is treated as a parameter in a more general sense, and resolution sensitivity Rs is used to elicit and showcase the dependence of model parameter sensitivity on external forcing. Three experiments (PiControl, 1pctCO2, abrupt4xCO2, see table 1) that participated in CMIP5, each with two realizations that differ in resolution (LR—low resolution and MR—Mixed Resolution), are analyzed. Model RS, defined by ST differences between LR and MR (MR minus LR), is measured by global mean ST (ST) and spatial features of the first two moments of simulated ST (temporal mean μ and standard deviation σ). Forcing-induced changes are measured by changes in RS between PiControl and transient simulations (1pctCO2 and abrupt4xCO2). If RS turns to vary with forcing conditions, it then suggests a dependence of model parameter sensitivity on the forcing applied and will have far-reaching implications for all uncertainty analyses based on CMIP-alike frameworks.

### Resolution sensitivity in PiControl

Global mean ST (ST) shows a clear response to the forcing applied (figure 1(a)): it stays steady at around 13.5 °C in PiControl. In response to the 1% CO2 increase per year, it increases at a steady rate to about 19 °C at the end of the simulation (after 150 years), whereas in abrupt4xCO2, it warms immediately upon the forcing and later at a slower rate reaching approximately 19.5 °C after 150 years of simulation. In contrast, resolution sensitivity in global mean ST (Rs(ST)) are relatively stable in all three integrations despite the different forcing (figure 1(b)): MR is slightly warmer than LR by 0.25 °C in PiControl, 0.2 °C in 1pctCO2 and 0.1 °C in abrupt4xCO2.

The warmer mean climate in MR can be observed on a global scale, with the exception in the North Atlantic subpolar basin, the Okhotsk Sea, and in the Southern Ocean on both sides of the Antarctic circumpolar current (figure 2(a)). Cold anomalies in the first two regions are mainly contributed by reduced heat transport owing to weaker oceanic gyres, whereas cooling near the Antarctic circumpolar current may relate to the fact that in comparison to LR, MR simulates a less well-connected south Atlantic Ocean and Indian ocean (thus reduced transport of warm and salty water from Indian Ocean to South Atlantic) (Jungclaus *et al.* 2013).

This global warming is accompanied by enhanced variability of ST over the two poles and significantly weaker variability in regions with high eddy kinetic energy, which corresponds to energetic ocean currents.
Lumpkin and Johnson 2013), including the Kuroshio current extension eastward from Japan, near Newfoundland where the North Atlantic current makes its way eastward between the subtropical and subpolar gyre, at the equatorial Pacific where the Equatorial current is located and in southeast Indian ocean corresponding spatially to the south Indian Ocean countercurrent (figure 2(b)). In subsequent analyses, figure 2 serves as reference RS maps.

Resolution sensitivity in transient simulations
RS in the transient simulations (estimated for the last 30 years of both 1pctCO2 and abrupt4xCO2) displays distinctly different spatial features from those in PiControl. RS in mean ST (RS_μ) shows a colder North — warmer south pattern in both transient simulations: strong cooling occurs over the Arctic and the North Atlantic basin, in the subtropical part of Africa, and as well over Australia and Amazonia, whereas wide-spread warming is observed in the Southern Ocean and over the Antarctic; there is also a tendency of cooling in the tropical continent in the Northern Hemisphere, e.g., over north America and tropical Eurasia extending from Middle-East to China (left column, figure 3). RS in variability (RS_σ) is characterized by prevalent weakening over the Arctic and a large area of the tropical oceans and by strengthening in the mid-latitude oceans and over the eastern equatorial Pacific (right column, figure 3).

To track how RS has evolved with time in the transient simulations until it becomes statistically different from that in the PiControl in spatial patterns (as shown in figure 3), spatial features of RS_μ (RS_σ), which are estimated over a 30-year window that slides through the entire 150 years of transient simulations for both LR and MR, are correlated to the respective reference map from PiControl (figures 2(a), (b)); the coefficients vary with time in color from black to light brown (dashed line in figure 4). Impacts of internal climate variability are sampled by performing the Monte-Carlo simulation for PiControl (see Method, gray dots in figure 4, dash-dotted lines define the 95% CI along each axis).

PiControl
RS in mean ST (RS_μ) shows high spatial consistency with correlation coefficients exceeding 0.91 for a 30-year window length at the 95% CI (x-axis, figures 4(a), (b)). However, the strong robustness revealed above does not extend to the spatial pattern of temporal standard deviation, which has much lower correlation coefficients (lying between [0.3, 0.65], horizontal dash-dotted lines in figure 4).

Transient simulations (1pctCO2 and abrupt4xCO2)
Both transient simulations show substantial changes in RS_μ but not in a similar manner in RS_σ. For instance, systematic changes in RS_μ can be detected 21 years after the forcing is applied in 1pctCO2 (year 21 denotes the

![Figure 1](attachment:image.png)

Figure 1. (a) Global mean surface temperature (ST) in LR and (b) its difference with respect to MR (MR minus LR) that defines resolution sensitivity (RS) in (ST) RS_μ<sub>ST</sub>. Only 250 years from PiControl are displayed. Black: PiControl, blue: 1pctCO2, and red: abrupt4xCO2. In (b) dashed lines denote temporal means.
starting year of a window of 30 years, thus it refers to the period of 1870–1899 and already immediately in the first simulation year in abrupt4xCO2 (the period of 1850–1879) (figures 4(a), (b), see the location of the very first black star). It means that beyond these time points RS_μ exhibits systematic changes so that it is statistically detectable (with correlation coefficients constantly below the lower bound of the 95% CI, 0.91). In contrast, changes in RS_σ do not project clear deviations from the ICV range.

**Summary and discussion**

Many uncertainty studies often assume that the three sources of uncertainty, namely, ICV-related uncertainty, model uncertainty, and forcing uncertainty, are independent of each other and accordingly can be conveniently decomposed (e.g., Hawkins and Sutton 2009) (Note that in CMIP-alike frameworks the term 'model uncertainty' does not consider the uncertainty in the model parameter space but only refers to model structural errors). Despite the increasing inter-model dependence that has evolved along model development (Collins et al 2011, Masson and Knutti 2011), the above assumption of independence of uncertainty sources may hold to a
large extent. For instance, a study based on various climate models has demonstrated that each model’s large-scale deviation from the multi-model-mean remains stationary despite strong climate change (from PiControl to abrupt4xCO2, Krinner and Flanner 2018), which can be interpreted as the distance between any two models (two different model structures) being stationary despite different forcing conditions. In other words, changes in forcing conditions and model structural errors could be regarded as being independent of each other. Does this conclusion also apply to model parameter sensitivity, namely, changes in model climate induced by changes in model parameters?

To answer the above question, this study relaxes the definition of model parameter that usually refers exclusively to physics-related parameters (such as those in the convection scheme) and regards model resolution as a model parameter. Resolution sensitivity (RS) that denotes differences between model simulations with different resolution (LR and MR) is then examined in terms of its dependence on forcing conditions to infer parameter sensitivity-forcing interactions. This relaxation is done because parametric uncertainty is not being addressed in the current CMIP framework and thus there is no way to directly address parameter sensitivity-forcing interactions. Therefore, it shall be viewed as a practicable and economic bypass that enables us to elicit and showcase the dependence of model parameter sensitivity on external forcing by examining a model’s

![Figure 3. RS_μ and RS_σ in 1pctCO2 and abrupt 4xCO2 (°C), estimated for the last 30 years of transient simulations. Only significant values are shown (5%, Monte Carlo test, see Appendix for details).](image)

![Figure 4. Spatial correlations in RS_μ (x-axis) and RS_σ (y-axis) between reference maps from PiControl (figures 2(a), (b)) and RS from transient simulations (dash-starred) calculated over a window of 30 years that slides through the entire 150 years of transient simulations for LR and Mr. Coefficients vary with time in color from black to light brown, corresponding to start and end of the 150 years. Impacts of internal climate variability are sampled by correlating reference maps and RS statistics that are calculated for 30-yr segments randomly selected from PiControl (1000 repetitions, gray dots, 95% CI: dash-dotted lines). As long as the coefficient moves permanently outside of the 95% CI along either axis or in both directions, a systematic change in RS is detected.)](image)
resolution sensitivity RS. Model resolution, though being non-physical per definition, confines the scope within which parameterizations (model physics) are being tuned, compared and selected. For instance, parameterization schemes show resolution-dependent performances (e.g., Gao et al 2017, Yuval and O’Gorman 2020). And, after all, if the model resolution is high enough, subgrid processes would be resolved explicitly (non-parameterized). Viewed from this perspective, resolution is an indispensable part of the physics-based model parameter space.

This analysis has examined model resolution sensitivity (RS) of MPI-ESM under different forcing conditions. The following results are obtained:

(i) When measured by global mean ST, there is no visible relation between the choice of model resolution and the forcing, in spite of slightly different global mean values (figure 1(b)).

(ii) In contrast, taking into account the spatial patterns of mean ST and its variability reveals a completely different picture: depending on the forcing applied, spatial patterns of ST may distinctly differ (figures 2 versus 3), as has been manifested by the spatial correlation statistics;

(iii) furthermore, the further the forcing deviates from the reference forcing, the earlier a systematic change in spatial characteristics of mean ST (not of standard deviation) can be detected (1pctCO2 versus abrupt4xCO2 in figure 4).

The first result is reassuring, because coupled climate models are tuned in such a way that global mean ST as one of the most basic physical properties should at least be reasonably reproduced (e.g., Schmidt et al 2017). In particular, the final tuning of MPI-ESM has been carried out under preindustrial conditions (Mauritsen et al 2012). Since changes of global mean ST are determined by energy balance on the Earth surface in a zero-dimensional space, figure 1(b) shows that LR and MR represent this net balance in a way that is quantitatively comparable not only in PiControl but also in both simulations with transient CO2 forcing.

As soon as spatial information of simulated ST is taken into account, the coherence shown in figure 1(b) disappears. Distinguishable ST differences are observed on a global scale under the PiControl forcing (figure 2).

From an energy point of view, it suggests that the representation of the energy balance at each geographic grid-point is resolution dependent, even though LR and MR agrees on the zero-dimensional measure, i.e., global mean ST. As a matter of course, details of physical mechanisms and dynamic processes are represented differently in LR and MR, as has been observed in various climate variables on a global scale (Jungclaus et al 2013, Stevens et al 2013). It should be noted that it is not straightforward to conclude which of the two resolutions provides more informative results about the actual system (Kirtman et al 2012, Wehner et al 2013), and which configuration to use in CMIP may be a rather random selection in a scientific sense.

Of particular interest is the revealed dependence of RS on forcing conditions, which unfolds itself in space such that completely different spatial patterns are detected towards the end of the transient climate response simulations (figures 3 and 4). More importantly, the above-revealed dependence may extend to a larger pool of model parameters, as demonstrated in Dommenget (2015): global-wide changes in ST that result from perturbed model physics (parameters) in both the control climate and the warmer climate due to a doubling of CO2 are not only spatially different but meanwhile also parameter-dependent.

Such a dependence of parameter sensitivity on forcing conditions has so far seldom been discussed for its own sake for Earth system models. This largely results from the great mathematical and computational challenges imposed by the vast number of model parameters in a typical Earth system model. It accounts for the overall under-explored parametric uncertainty in Earth System models in comparison to some less complex subsystem models. The great endeavor by Stainforth et al (2005) serves as a good example to illustrate this point.

They allowed (only) 6 parameters in the atmosphere model that is coupled to a mixed layer ocean model to vary simultaneously. Even so, the experiment with a decent ensemble size was only made possible by thousands of volunteers worldwide who were willing to share their home computers to participate in the computation. It is needless to stress that their model has far less parameters than a typical Earth System model. On the other hand, it is to acknowledge that in comparison, parametric uncertainty and its dependence on forcing conditions have been relatively well explored in some climate subsystems. Examples include a carbon cycle model (Grieb et al 1999) and hydrological models (Gan et al 2014, Mockler et al 2016).

In practice, quite often, the multi-model ensemble spread of CMIP models is being analyzed as if it encompasses all model-related uncertainty and leaves the parametric uncertainty and its dependence on forcing entirely unaccounted (e.g., Beobide-Arsuaga et al 2021, Cai et al 2021). While such analyses may be received as normal practices for climate scientists, it may generate hazardous effects when CMIP outputs are used for practical purposes, for instance, to serve policy makers for future planning in the battle against adverse consequences of climate change (e.g., Ruane et al 2016, Chen et al 2020, UNEP 2020, Shiogama et al 2021).
these contexts, being unaware of the underestimation of model uncertainty, due to the unaccounted parametric uncertainty and its dependence on forcing conditions, will unavoidably lead to mis-planning of adaption measures that are designed to combat adverse consequences of climate change.

Specifically, the forcing-dependent parameter sensitivity exacerbates the under-representation of uncertainty in climate change studies, which takes place on two levels in CMIP-alike frameworks: firstly, model parametric uncertainty is far from being properly sampled on the level of individual models. Only one model version (i.e., parameter combination) is (or in a better case, a handful of model versions are) selected for each individual model, which is far from sufficient to describe the parametric uncertainty (Grieb et al 1999, Stainforth et al 2005). The resulting response of each individual model to a certain forcing is thus restrictively biased and further penetrates to multi-model ensemble statistics. Secondly, model parameter combinations that are predetermined under preindustrial or present-day conditions are employed to study model responses to different climate conditions (such as 1pctCO2 and abrupt4xCO2). Given the dependence of parameter sensitivity on forcing conditions, the parameter combination that is ‘optimal’ under, say, the present-day climate, is not likely to be ‘optimal’ any more for a different climate. In this sense, there is no reason to expect the multi-model ensemble spread across different CMIP generations to converge.

Bringing the above two aspects into the framework of uncertainty quantification of model parameters, the first point pertains to sensitivity analyses with a focused task of identifying the most sensitive parameters that have significant influences on simulated climate (or to screen out those exerting only marginal influences on model climate), whereby forcing conditions are usually held constant for different parameter combinations. Complementing this line of research, the second point puts the focus on interactions between parameters and external forcing, whereby forcing conditions are allowed to change, e.g., from preindustrial to historical forcing, or to scenario emission pathways. Specifically, the interaction between parameter sensitivity and forcing conditions becomes the subject of examination. Quite many efforts have been undertaken to address the first aspect (e.g., Grieb et al 1999, Gan et al 2014, Shi et al 2019). Model tuning could also be put in this category (Mauritsen et al 2012, Hourdin et al 2017). In contrast, the parameter-forcing interaction is receiving far less attention. It however affects all essential characteristics of model climate, such as global mean measures like climate sensitivity (e.g., Stainforth et al 2005) and spatial-temporal details of key climate variables (Dommenget 2015, Mockler et al 2016), as has been unambiguously demonstrated by this study.

It is worthwhile stressing that model resolution is a distinct parameter in the sense that it is mostly a subjective choice and reflects a compromise concerning model complexity, computational affordability, and model application priority, whereas physical parameters are allowed to take values from a certain numeric range or distribution and thus the choice of each individual parameter and their combination entails quantitative justification while ensuring physical principles, e.g., by defining and minimizing a cost function, and therefrom objective approaches are employed, as is the case in all sensitivity analysis methods (Hourdin et al 2017). This distinction, however, does not obviate the practice of inferring parameter sensitivity-forcing interactions by examining how resolution sensitivity varies under different forcing, which is reported in this work.

This study highlights two facts about CMIP models in addition to the un-represented parametric uncertainty: (i) parameter-forcing interactions are entirely missing, and (ii) the choice of climate metrics is of critical relevance not only for model tuning but also for performance assessment and model applications. It proposes that model parameter uncertainty and forcing conditions should be studied as a compound in a framework that incorporates perturbed physics ensemble and multi-model ensemble together. With the help of the initial condition ensemble method and stochastic parameterization, it enables a more complete representation of total uncertainty in forced model responses. The merit of such a combined approach has been successfully demonstrated (e.g., Yokohata et al 2010, Weisheimer et al 2011).

Here it lays a potential of obtaining a too-broad range of uncertainty. While on the one side Bayesian inference may help identify succinct model setups and/or model combinations for a given observational dataset (e.g., Spiegelhalter et al 2002), on the other side, climate models should contain no inessential complexity relative to the problem at hand (Jeevanjee et al 2017), which essentially concerns the choice of climate metrics. This requirement of ‘fit for purpose’ should also be observed when determining the calibration criteria for model tuning, which is iteratively executed until the model climate fits the calibration criteria with an acceptable accuracy. Some current criteria only consider global measures or spatially aggregated model errors (Taylor 2001, Reichler and Kim 2008, Collins et al 2011, Qian et al 2016); spatial coherence that can be revealed by spatial correlation is not yet taken into account. If model results are meant to guide the design and planning of regional sustainable development, a reasonable degree of accuracy in spatial details is required. Naturally, spatial details of climate variables of concern should be assigned appropriate weighting in the calibration metric, although it should always be kept in mind that even a perfect knowledge of the past should not be extrapolated with the same confidence into the future.
Acknowledgments

This study is supported by National Natural Science Foundation of China (41776002) and Strategic Priority Research Program of Chinese Academy of Sciences (XDA20060501). The author thanks Stephan Kern, Frank Sielmann and Klaus Fraedrich for helpful comments. Thanks are extended to anonymous reviewers. The author likes to acknowledge the CMIP community for providing the climate model data, retained and globally distributed in the framework of the ESGF. The CMIP data of this study were replicated and made available for this study by the DKRZ.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Xiuhua Zhu @ https://orcid.org/0000-0001-8765-3976

References

Andrews T, Gregory J M, Webb M J and Taylor K E 2012 Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models Geophys. Res. Lett. 39 L09712

Bebobide-Assuaga G et al 2021 Uncertainty of ENSO-amplitude projections in CMIP5 and CMIP6 models Clim. Dyn. (https://doi.org/10.1007/s00382-021-05673-4)

Cai Z, You Q, Wu F, Chen H W, Chen D and Cohen J 2021 Arctic warming revealed by multiple CMIP6 models: evaluation of historical simulations and quantification of future projection uncertainties J. Climate 34 4871–92

Chen Y, Liu A and Cheng X 2020 Quantifying economic impacts of climate change under nine future emission scenarios within CMIP6 Total Environ. 703 134950

Collins M et al 2011 Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles Clim. Dyn. 36 1737–66

Dommeng D 2015 A simple model perturbed physics study of the simulated climate sensitivity uncertainty and its relation to control climate biases Clim. Dyn. 46 427–47

Gan Y, Duan Q, Gong W, Tong C, Sun Y, Chu W, Ye A, Miao C and Di Z 2014 A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model Environ. Modelling Safew. 51 269–85

Gao Y, Leung L R, Zhao C and Hagos S 2017 Sensitivity of US summer precipitation to model resolution and convective parameterizations across gray zone resolutions J. Geophys. Res. Atmos. 122 2714–33

Grieb T M et al 1999 Examination of model uncertainty and parameter interaction in a global carbon cycling model (GLOCCO) Environ. Int. 25 787–803

Hagedorn R, Doblas-Reyes F J and Palmer T N 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept Tellus A 57 219–33

Hawkins E and Sutton R 2009 The potential to narrow uncertainty in regional climate predictions Bull. Amer Meteor Soc 90 1095–107

Houdin F et al 2017 The art and science of climate model tuning Bull. Am. Meteorol. Soc. 98 589–602

Jeevanjee N, Hassanzadeh P, Hill S and Sheshadri A 2017 A perspective on climate model hierarchies Geosci. Model Dev. 10 3207–23

Knutti R and Sedlácek J 2013 Robustness and uncertainties in the new CMIP5 climate model projections Nat. Clim. Change 3 369–73

Knutti R and Flanner M G 2018 Striking stationarity of large-scale climate model bias patterns under strong climate change Proc Natl Acad Sci USA 115 9462–6

Lumpkin R and Johnson G 2013 Global ocean surface velocities from drifters: mean, variance, Elnino-Southern oscillation response, and seasonal cycle J. Geophys. Res. Oceans 118 2992–3006

Masson D and Knutti R 2011 Climate model genealogy Geophys. Res. Lett. 38 L08703

Mauritsen T et al 2012 Tuning the climate of a global model J. Adv Model Earth Syst 4 M00A01

Mockler M, Chun K P, Saprina-Azuri G, Bruen M and Weaver H S 2016 Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations Adv. Water Resour. 97 299–313

Palmer T, Doblas-Reyes F, Hagedorn R and Weisheimer A 2003 Probabilistic prediction of climate using multi-model ensembles: from basics to applications Philos Trans R Soc Lond B Biol. Sci. 368 1991–9

Qian Y et al 2016 Uncertainty quantification in climate modeling and projection Bull. Am. Meteorol. Soc. 97 821–4

Reichler T and Kim J 2008 How well do coupled models simulate today’s climate? Bull. Am. Meteorol. Soc. 89 303–11

Ruane A C et al 2016 The vulnerability, impacts, adaptation and climate services advisory board (VIACS AB v1.0) contribution to CMIP6 Geosci. Model Dev. 9 3493–515

Schmidt G et al 2017 Practice and philosophy of climate model tuning across six US modeling centers Geosci. Model Dev. 10 3207–23

Shi Y et al 2019 How parameter specification of an earth system model of intermediate complexity influences its climate simulations Prog. Earth Planet Sci. 6 46

Shiogama H et al 2021 Selecting CMIP6-based future climate scenarios for impact and adaptation studies SOLA 17 57–62

Smith L A, Ziehmann C and Fraedrich K 1999 Uncertainty dynamics and predictability in chaotic systems Q J R Meteorol Soc 125 2855–86

Spiegelhalter D J, Best N G, Carlin B P and van der Linde A 2002 Bayesian measures of model complexity and fit J R Statist Soc B 64 583–639

Stainforth D A et al 2005 Uncertainty in predictions of the climate response to rising levels of greenhouse gases Nature 433 403–6

Xiuhua Zhu
Stainforth D A, Allen M R, Tredger E R and Smith L A 2007 Confidence, uncertainty and decision-support relevance in climate predictions Phil. Trans. R. Soc. A 365 2145–61
Stevens B et al 2013 Atmospheric component of the MPI-M earth system model: ECHAM6 J. Adv. Model Earth Syst. 5 146–72
Taylor K E 2001 Summarizing multiple aspects of model performance in a single diagram J. Geophys. Res. 106 7183–92
UNEP 2020 Projections of Future Coral Bleaching Conditions using IPCC CMIP6 models: climate Policy Implications, Management Applications, and Regional Seas summaries (Nairobi, Kenya: United Nations Environment Programme) http://hdl.handle.net/20.500.11822/34219
Wang X L, Feng Y and Swail V R 2015 Climate change signal and uncertainty in CMIP5-based projections of global ocean surface wave heights J. Geophys. Res. Oceans 120 3859–71
Wehner M F et al 2013 The effect of horizontal resolution on simulation quality in the community atmospheric model CAM5.1. J Adv Model Earth Syst 6 980–97
Weisheimer A, Palmer T N and Doblas-Reyes F J 2011 Assessment of representation of model uncertainty in monthly and seasonal forecast ensembles Geophys. Res. Lett. 38 L16703
Woldemeskel F M, Sharma A, Sivakumar B and Mehrotra R 2016 Quantification of precipitation and temperature uncertainties simulated by CMIP3 and CMIP5 models J. Geophys. Res. Atmos. 121 3–17
Yip S, Ferro C A T, Stephenson D and Hawkins E 2011 A simple coherent framework for partitioning uncertainty in climate predictions J. Climate 24 4634–43
Yokohata T et al 2010 Structural similarities and differences in climate responses to CO2 increase between two perturbed physics ensembles J. Climate 23 1392–410
Yuval J and O’Gorman P A 2020 Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions Nat. Commun. 11 3295
Zelinka M D et al 2020 Causes of higher climate sensitivity in CMIP6 models Geophys. Res. Lett. 47 e2019GL085782