Benchmark calculations of radiative forcing by greenhouse gases

Robert Pincus[1,2], Stefan A. Buehler[3], Manfred Brath[3], Omar Jamil[4], K. Franklin Evans[5], James Manners[4,6], Raymond L. Menzel[7,8], Eli J. Mlawer[9], David Paynter[8], Rick L. Pernak[9]

[1]Cooperative Institute for Environmental Studies, University of Colorado, Boulder, Colorado, USA
[2]NOAA/Earth System Research Lab, Physical Sciences Division, Boulder, Colorado, USA
[3]Informatics and Natural Sciences Department of Earth Sciences, Meteorological Institute, Faculty of Mathematics, Universität Hamburg, Hamburg, Germany
[4]Met Office, Exeter, UK
[5]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado, USA
[6]Global Systems Institute, Exeter University, Exeter, UK
[7]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA
[8]University Corporation for Atmospheric Research, Princeton, New Jersey, USA
[9]Atmospheric and Environmental Research, Lexington, Massachusetts, USA

Key Points:

- Clear-sky instantaneous radiative forcing by greenhouse gases is computed with six benchmark models using 100 profiles.
- Sampling error in the global, annual mean is much smaller than the level of disagreement among models, which is typically about 0.01 $\mathrm{W\,m^{-2}}$.
- The impacts of clouds and stratospheric cooling are assessed using simple assumptions.

Corresponding author: Robert Pincus, `Robert.Pincus@colorado.edu`

Abstract

Changes in the concentration of greenhouse gases within the atmosphere lead to changes in radiative fluxes within the atmosphere and at its boundaries. This paper describes an experiment within the Radiative Forcing Model Intercomparision Project that uses benchmark calculations made with line-by-line models to identify parameterization error in this quantity. The instantaneous forcing to which the world has been subject is computed using a set of 100 profiles, selected from a re-analysis of present-day conditions, that represent the global annual mean forcing with sampling errors of less than 0.01 $\mathrm{W\,m^{-2}}$. Agreement in estimates of forcing among six contributing line-by-line models is excellent, with standard deviations typically less than 0.025 $\mathrm{W\,m^{-2}}$, suggesting that parameterization error will be readily resolved. The impact of clouds on this forcing is estimated using diagnostic calculations across a range of climate models, while adjustments due to stratospheric temperature re-equilibration are estimated assuming fixed dynamical heating.

1 Providing global-scale benchmarks for radiation parameterizations

One of the three questions motivating the sixth phase of the Coupled Model Intercomparison Project (CMIP6, see Eyring et al., 2016) is "How does the Earth system respond to forcing?" The degree to which this question can be addressed depends partly on how well the forcing can be characterized. The measure most useful in explaining the long-term response of surface temperature is the *effective radiative forcing*, defined as change in radiative flux at the top of the atmosphere after accounting for adjustments (changes in the opacity of the atmosphere not associated with mean surface warming, see Sherwood et al., 2015). In support of CMIP6 the Radiative Forcing Model Intercomparison Project (RFMIP, see Pincus, Forster, & Stevens, 2016) will characterize the forcing to which models are subject using fixed-SST experiments (Hansen, 2005; Rotstayn & Penner, 2001) in which atmospheric composition and land use are varied but surface temperature response is suppressed (Forster et al., 2016).

The models participating in the previous phase of CMIP translated prescribed changes in atmospheric composition into a relatively wide range of effective radiative forcing (e.g. Chung & Soden, 2015) even in the absence of adjustments. Some of this variability is due to how the model-specific distributions of clouds and water vapor mask the radiative impact of changes in greenhouse gas concentrations (e.g. Huang, Tan, & Xia, 2016). Additional variability, however, is due to model error in the *instantaneous radiative forcing*, i.e. the change in flux in the absence of adjustments, as illustrated by comparisons that use prescribed atmospheric conditions to (Collins et al., 2006; Ellingson, Ellis, & Fels, 1991; Oreopoulos et al., 2012; Pincus et al., 2015) to eliminate other causes of disagreement.

RFMIP will complement the characterization of effective radiative forcing with an assessment of errors in computations of clear-sky instantaneous radiative forcing due to greenhouse gases and aerosols. This assessment, identified within CMIP6 as experiment *rad-irf*, is possible because there is little fundamental uncertainty. Using reference "line-by-line" models, atmospheric conditions and gas concentrations can be mapped to extinction with high fidelity at the very fine spectral resolution needed to resolve each of the millions of absorption lines. Fluxes computed with high spectral and angular resolution are then limited in precision primarily by uncertainty in inputs. These benchmark models are known to be in very good agreement with observations (e.g. Alvarado et al., 2013; Kiel et al., 2016), especially in the absence of difficult-to-characterize clouds, given current knowledge of spectroscopy.

Previous assessments of radiative transfer parameterizations, focused on understanding the causes of error, have examined the response to perturbations around a small numbers of atmospheric profiles. RFMIP builds on this long history by focusing

on the global scale relevant for climate modeling. As we explain below, we make this link by carefully choosing a small number of atmospheric states that nonetheless sample the conditions needed to determine global-mean clear-sky instantaneous radiative forcing by greenhouse gases. A number of reference modeling groups have provided fluxes for these sets of conditions, providing both a benchmark and information as to how reasonable choices might affect those benchmarks given current understanding.

Here we exploit the calculations requested by RFMIP to move towards benchmark estimates of the true radiative forcing to which the earth has been subject due to increases well-mixed greenhouse gases. We describe the construction of a small set of atmospheric columns that can be used to accurately reproduce global-mean, annual-mean instantaneous radiative forcing by greenhouse gases, summarize the reference calculations being supplied and demonstrate the very small level of disagreement among them. We highlight the values of clear-sky instantaneous radiative forcing for a range of changes in atmospheric composition relative to pre-industrial conditions and cautiously extend these benchmark estimates towards more useful estimates that include the impact of clouds and adjustments.

2 Making global-mean benchmarks practical

Large-scale line-by-line calculations have become increasingly practical, and the RFMIP effort to diagnose errors in instantaneous radiative forcing by aerosols will apply line-by-line modeling at relatively low spectral resolution (Jones et al., 2017) to eight global snapshots for each participating model. Errors in global mean, annual mean clear-sky instantaneous radiative forcing by greenhouse gases, however, can be assessed with a much more parsimonious set of atmospheric conditions. This is because temporal variations of temperature and water vapor are relatively slow and have a modest impact on the sensitivity of flux to changes in greenhouse gas concentrations. Many previous calculations (see Etminan, Myhre, Highwood, & Shine, 2016, for a recent example), in fact, estimate global mean, annual mean values using just two or three profiles, based on work in the 1990s showing that even such simple representations of latitudinal variability are sufficient to constrain flux changes at the tropopause to within about a percent (Freckleton et al., 1998; Myhre, Highwood, Shine, & Stordal, 1998).

Here we describe the construction of a set of atmospheric profiles designed to determine *error* in global-mean in instantaneous radiative forcing, obtained using a reference model on a very large number of atmospheric and surface conditions to determine the (present-day) radiative forcing, and choosing a subset of these conditions that minimizes the sampling error across a range of measures in radiative forcing. As we demonstrate below, the same set of profiles also provides an accurate sample of the parameterization or approximation error in radiative forcing.

2.1 Computing global-mean, annual mean radiative fluxes and flux perturbations

We characterize the range of conditions in the present-day atmosphere using a single year (2014) of the ERA-Interim reanalysis (Dee et al., 2011). We sample temperature, pressure, specific humidity, ozone mixing ratios, and surface temperature and albedo on a 1.5°grid every 10.25 days. Sampling at high latitudes is reduced to maintain roughly equal area weighting. Concentrations of other greenhouse gases (CO_2, CH_4, N_2O, HCFCs 22 and 134a, CFCs 11, 12, and 113, and CCl_4) use 2014 values from NOAA greenhouse gas inventories and are assumed to be spatially uniform. We assume that these 823,680 profiles adequately represent global-mean, annual-mean clear-sky conditions.

We apply a series of 17 perturbations (detailed in the Supplemental Information) to these conditions, including varying concentrations of greenhouse gases (especially CO_2), temperature, and humidity. Some temperature perturbations include spatial patterns obtained from climate change simulations made for CMIP5. The perturbations are intended to sample error across a wide range of conditions. The perturbations are similar to, but not quite the same as, those used by the final RFMIP experiments in Section 3, because the RFMIP protocol was not fully established when we performed these calculations.

Reference fluxes for present-day conditions and each perturbation are computed using the UK Met Office SOCRATES (Suite Of Community RAdiative Transfer codes based on Edwards & Slingo, 1996) using a very high-resolution k-distribution with 300 bands in the longwave and 260 bands in the shortwave (Walters et al., 2019), that agrees quite well with line-by-line models (e.g Pincus et al., 2015). The spectral overlap of gases is treated with equivalent extinction with corrected scaling. Clouds and aerosols are not considered, consistent with the RFMIP protocol.

We also compute fluxes for these sets of atmospheric conditions with an approximate model: RRTMG (Iacono, Mlawer, Clough, & Morcrette, 2000; Mlawer, Taubman, Brown, Iacono, & Clough, 1997), which is based on somewhat older spectroscopic information and so is expected to have errors with a potential dependence on atmospheric state.

2.2 Choosing a set of globally-representative profiles

We seek a small subset of atmospheric profiles that best reproduces the global, annual mean obtained from the full calculation. To identify such a set we must quantify what we mean by "best" by defining a cost or objective function with which to measure sampling error. Because the goal of RFMIP is to establish accuracy in calculations of radiative forcing, our objective function O is defined in terms of the change in flux between each of the 17 perturbations and present-day conditions. (For perturbations in which the only change is to greenhouse gas concentrations this quantity is precisely the instantaneous radiative forcing.) The objective function includes errors in changes of upward flux at the top of the atmosphere and downward flux at the surface as well as changes in flux divergence in above and below the tropopause (as determined by Wilcox, Hoskins, & Shine, 2011); each quantity is computed for both longwave and shortwave fluxes. We guard against compensating errors related to temperature or humidity by further considering 9 roughly equal-area latitude bands centered on the equator. We choose an l^2 norm so that

$$O = \left[\frac{1}{N_{\mathrm{pert}} N_{\mathrm{flux}} N_{\mathrm{lat}}} \sum_{l=1}^{N_{\mathrm{lat}}} \sum_{p=1}^{N_{\mathrm{pert}}} \sum_{q=1}^{N_{\mathrm{quant}}} \left(\Delta F_{l,p,q}^{(\mathrm{samp})} - \Delta F_{l,p,q}^{(\mathrm{true})} \right)^2 \right]^{1/2} \tag{1}$$

where $\Delta F_{l,p,q}$ describes the average change in flux or flux divergence, as computed with the reference model, between perturbation p and present-day conditions in latitude band l for quantity q. The objective function includes the four flux quantities for both longwave and shortwave fluxes ($N_{quant} = 8$).

We identify optimal subsets of profiles from within the complete set using simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983). Because the optimization is stochastic we perform 25 independent optimizations for each of a range of subset sizes. We save the realization with the lowest value of O although this choice has little impact as the standard deviation across realizations is small (roughly 6% of the mean sampling error), so that the sampling error in the best realization is only about 10% smaller than the mean (Figure 1). Simulated annealing produces sampling errors substantially lower than purely random sampling (by a factor of 19 for 100 columns, not

shown). The choice of columns is reasonably robust: sampling error in the independent estimate of mean radiative forcing with RRTMG is only modestly larger (15% for 100 columns) than for the reference calculations used in the optimization.

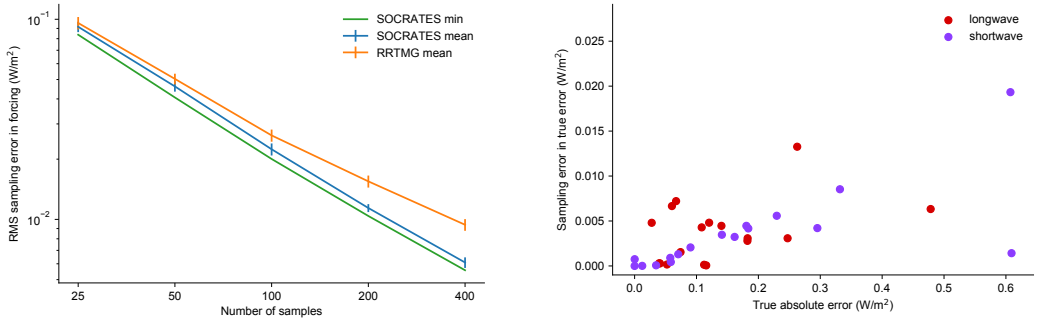

Figure 1. Left: values of the cost function O, an aggregate measure of error across regions, changes in atmospheric conditions, and measures of flux (Eq. 1) as a function of the number of optimal columns. The simulated annealing method used to chose the columns is stochastic; the mean and standard deviation across realizations is shown along with the value of sample error from the best-fit realization used in further calculations. The choice of columns based on reference radiative transfer calculations ("SOCRATES high-res") is robust, producing only modestly larger sampling errors for approximate calculations ("RRTMG"). Right: Absolute value of the sampling error $\mathcal{E}^{(\mathrm{samp})} - \mathcal{E}^{(\mathrm{true})}$ in estimates of the approximation error $\mathcal{E} = \Delta F^{(\mathrm{approx})} - \Delta F^{(\mathrm{ref})}$ sought by RFMIP. Errors shown are for the mean of100 samples representing the global, annual mean, for changes in upwelling longwave flux at the top of the atmosphere (red) and downwelling shortwave flux at the surface (purple) from 17 perturbations. Parameterization errors range from 0 to about 0 to 0.6 $\mathrm{W\,m^{-2}}$ in the global, annual mean; sampling error almost always less than 0.01 $\mathrm{W\,m^{-2}}$.

Columns chosen to minimize sampling error in mean radiative forcing also provide accurate estimates of parameterization error $\mathcal{E} = \Delta F^{(\mathrm{approx})} - \Delta F^{(\mathrm{ref})}$ in that forcing. Fig. 1 shows the sampling error $\mathcal{E}^{(\mathrm{samp})}_{p,q} - \mathcal{E}^{(\mathrm{true})}_{p,q}$ in estimates of the global, annual mean parameterization error for RRTMG compared to high-resolution SOCRATES calculations for the 17 perturbations used to develop the column samples. True absolute errors from RRTMG range from near 0 to 0.6 $\mathrm{W\,m^{-2}}$ in the global, annual mean; sampling error in these estimates is almost always less than 0.01 $\mathrm{W\,m^{-2}}$.

The RFMIP protocol uses the set of 100 columns with the lowest value of the objective function O. As a consequence of optimizing the sampling for radiative forcing, fluxes for any individual state including the present-day baseline are themselves subject to sampling errors: global mean insolation in our sample, for example, is 335.1 $\mathrm{W\,m^{-2}}$ (c.f. the true mean of $\sim 1361/4 = 340.25$ $\mathrm{W\,m^{-2}}$). In addition, using a single set of columns for both longwave and shortwave calculations means that the sun is below the horizon for roughly half the set of columns.

3 Radiation calculations with reference models

3.1 Contributions and variants

To date results from five benchmark models are available for experiment *rad-irf*: ARTS 2.3 (Buehler et al., 2018), provided by the University of Hamburg; LBLRTM v12.8 (Clough et al., 2005), provided by Atmospheric and Environmental Research;

the SOCRATES model described in Sec. 2.1, provided by the UK Met Office; the Reference Forward Model (Dudhia, 2017), provided by the NOAA Geophysical Fluid Dynamics Lab; and GRTCODE, a new line-by-line code developed at GFDL. Most models used the spectroscopic information from HITRAN 2012 (Rothman et al., 2013) although GRTCODE results use HITRAN 2016 (Gordon et al., 2017) and LBLRTM used the aer_v_3.6 line file, which is based on HITRAN 2012 but includes small changes to improve comparisons with select observations. With one exception noted below the models use variants of the MT_CKD continuum (Mlawer et al., 2012).

These five models provided thirteen sets of longwave fluxes and seven sets of shortwave fluxes. This multiplicity arises because some models provided calculations for slightly different sets of greenhouse gases, called "forcing variants" within CMIP and RFMIP, and/or slightly different model configurations ("physics variants").

Climate models participating in CMIP6 may specify well-mixed greenhouse concentrations using one of three forcing variants described by Meinshausen et al. (2017): using some or all of the 43 greenhouse gases provided in the forcing data set; by prescribing CO_2, CH_4, N_2O, CFC−12, and an "equivalent" concentration of CFC−11 to represent all other gases; or using CO_2, CH_4, N_2O, and equivalent concentrations of CFC−11 and HFC−134 a. (Concentrations of water vapor and ozone are drawn from reanalysis, as described in Sec. 2.1.) Some models provided results for more than one of these forcing variants.

In addition, some models provided calculations with slightly reconfigured models. ARTS 2.3 does not normally include CO_2 line mixing but provided a second physics variant that did so. High spectral resolution calculations with SOCRATES are themselves considered a second physics variant of the lower-resolution calculations made during simulations with the host model HadGEM; a third variant uses the MT_CKD 3.2 treatment of the water vapor continuum in lieu of the CAVIAR continuum used in the development of the parameterization.

3.2 Instantaneous clear-sky forcing at present day

Figure 2 shows an example calculation of instantaneous radiative forcing, i.e. the change in net downward flux at TOA and surface and the change in net absorption across the atmosphere (net flux at TOA minus net at surface), here for the change between present-day and pre-industrial conditions. Increased greenhouse gas concentrations in the present day increase the opacity of the atmosphere. In the longwave this acts to decrease outgoing longwave at the TOA and increase downward longwave at the surface. The increase in downwelling surface radiation is smaller than the decrease in outgoing longwave, resulting in decreased radiative cooling across the atmosphere. In the shortwave there a near-zero increase in scattering back to space but an increase in atmospheric absorption, resulting in diminished solar radiation at the surface.

Agreement among the line-by-line models is excellent: the standard deviation across all six quantities is less than 0.025 $\mathrm{W\,m^{-2}}$ with the exception of LW absorption, where the standard deviation is 0.033 $\mathrm{W\,m^{-2}}$. There is no systematic variation across forcing variants, indicating that the equivalent concentrations accurately summarize the radiative impact of the neglected gases in the transition from pre-industrial to present-day conditions.

Changes in shortwave flux between pre-industrial and present-day are substantially smaller than in the longwave. The standard deviations are commensurate with those in the longwave, but diversity in atmospheric absorption and surface forcing is dominated by physics variant 2 of the SOCRATES code, which is unique among the models in using the CAVIAR treatment for continuum absorption by water vapor (Ptashnik, McPheat, Shine, Smith, & Williams, 2011; Ptashnik et al., 2013). Absorp-

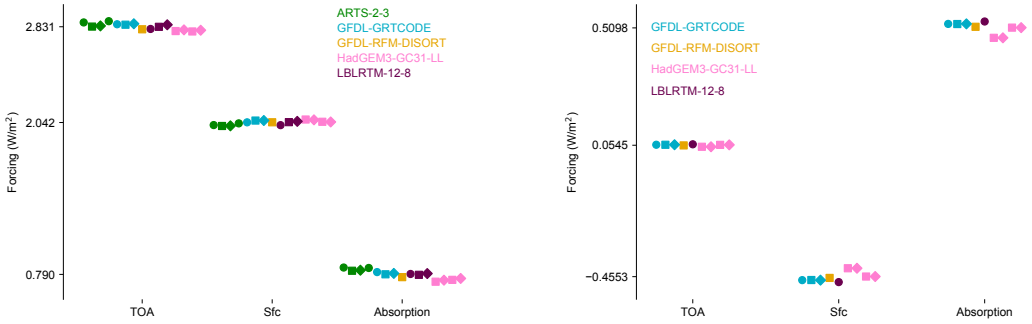

Figure 2. Global, annual mean instantaneous clear-sky radiative forcing by greenhouse gases at present-day, relative to pre-industrial conditions, as computed by benchmark radiative transfer models. Longwave results are on the left, shortwave results on the right, with the reference model denoted by the color. Model names follow the RFMIP convention with contributions from SOCRATES labeled as HadGEM3 to link the results to the host climate model. Results include multiple representations of greenhouse gas changes (circles, squares, and diamonds corresponding to forcing variants 1, 2, and 3) and small variants in the treatment of some physical processes as explained in the text. All variants of the reference models agree well in longwave calculations, while SOCRATES results in the shortwave show the small but noticeable impact of different treatments of the H_2O continuum, which overlaps with absorption by other gases in the near-infrared and so affects forcing by those gases.

tion in the near infrared in the CAVIAR continuum is substantially larger than in the MT_CKD continuum on which all other models rely, especially where water vapor absorption coincides with absorption lines of CO_2, CH_4 and N_2O. This makes the effect of these gases less significant and so reduces their forcing between pre-industrial and present-day concentrations.

Global-mean values of clear-sky instantaneous radiative forcing for a range of well-mixed greenhouse gases, averaged across all available reference models, are provided in Table 3.2.

3.3 Establishing a benchmark for parameterization error

Experiment *rad-irf* is intended to assess error in the parameterization of clear-sky radiation in the climate models participating in CMIP6. Resolving this error is only possible if the disagreement among benchmark models is small relative to the typical difference between a parameterization and the reference models themselves. Figure 3, which compares error from three modern parameterizations to the variability across the reference models, suggests that the benchmark calculation is likely to meet this goal. Results are shown for forcing across all 17 perturbations in experiment *rad-irf*. Errors relative to LBLRTM v12.8 are shown the for low-resolution version of SOCRATES, as used in the HadGEM model; for the parameterization used in the GFDL's AM4 model (Zhao et al., 2018); and for the newly-developed RTE+RRTMGP code (Pincus, Mlawer, & Delamere, 2019) which is trained on calculations with LBLRTM v12.8. These three parameterizations all use recent spectroscopic information and so are likely to be among the parameterizations with the smallest error. Nonetheless the error in each parameterization is almost always larger than the standard deviation across reference models, indicating differences between parameterizations and all reference models are dominated by parameterization error.

Table 1. Mean instantaneous radiative forcing across all available benchmark models, forcing variants, and physics variants, in $\mathrm{W\,m^{-2}}$. Forcing is the difference between net downward radiation under perturbed conditions minus those under pre-industrial (PI) conditions; because the profiles provided for experiment *rad-irf* are perturbed around present-day (PD) conditions the difference required may be indirect, as explained in the table. Values are provided for the top of the atmosphere (TOA) and surface (Sfc). RFMIP experiment *rad-irf* contains further perturbations meant to assess errors in temperature and humidity dependence.

Experiment	LW TOA	LW Sfc	SW TOA	SW Sfc
Computed as difference from perturbation "PI"				
Present-day	2.831	2.042	0.055	-0.455
Future	7.417	5.567	0.355	-1.393
Last Glacial Maximum	-2.387	-1.417	-0.065	0.316
Computed as negative difference from perturbation "PD"				
Present-day CO_2	1.311	0.930	0.029	-0.165
Present-day CH_4	0.612	0.274	0.055	-0.242
Present-day N_2O	0.205	0.088	0.002	-0.011
Present-day O_3	0.129	0.326	-0.032	-0.033
Present-day halocarbons	0.533	0.394	0.000	-0.001
Computed as difference from perturbation "PI CO2"				
$\frac{1}{2}\times CO_2$	-2.702	-1.792	-0.050	0.274
$2\times CO_2$	2.714	1.978	0.064	-0.367
$3\times CO_2$	4.308	3.260	0.110	-0.629
$4\times CO_2$	5.443	4.253	0.146	-0.840
$8\times CO_2$	8.207	7.042	0.252	-1.442

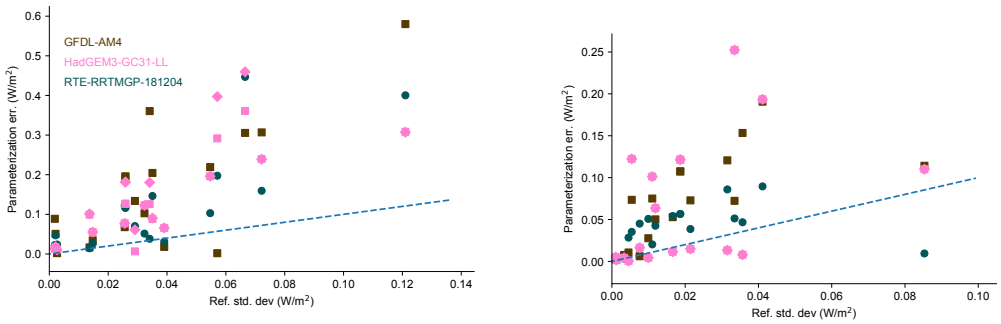

Figure 3. Absolute error in instantaneous radiative forcing (longwave at the top of atmosphere on the left, shortwave at the surface on the right) as computed by three modern parameterizations as a function of amount of disagreement across the reference models. Results are shown for all available forcing and physics variants for each of the 17 perturbations in experiment *rad-irf*. Error is assessed relative to LBLRTM v12.8 on which the RTE+RRTMGP parameterization is trained, minimizing the error for this parameterization. Regardless of which model is used as the benchmark, however, the error in each of parameterization exceeds the standard deviation of results from the reference models for a large majority of perturbations, indicating that the reference calculations reported here are accurate enough to resolve parameterization error.

4 Moving towards effective radiative forcing

4.1 Accounting for clouds

Though RFMIP experiment *rad-irf* was designed to assess parameterization error it offers an opportunity to refine benchmark calculations of the radiative forcing experienced by Earth due to various composition changes. Estimates of the instantaneous radiative forcing must be modified to account for clouds which screen the impact of changes in greenhouse gas concentrations. Previous efforts to establish benchmarks (e.g. Etminan et al., 2016; Myhre, Stordal, Gausemel, Nielsen, & Mahieu, 2006) have made this problem tractable by using two atmospheric profiles (see Sec. 2) each combined with three sets of representative cloud properties as observed by passive satellite instruments. Sampling errors in the global, annual mean are of order 1% although the cloud data are not well-suited to estimates of masking at the surface. A complete calculation would sample the co-variability of clouds, temperature, humidity, and ozone (assuming concentrations of other greenhouse gases vary primarily in the vertical) requiring vastly more computation than is required for clear skies.

As an alternative we have examined the ratio of all-sky to clear-sky instantaneous radiative forcing by greenhouse gases in the few available simulations from CMIP6. The Cloud Feedbacks Model Intercomparison Project (Webb et al., 2017) requests, at low priority, calculations with CO_2 concentrations quadrupled from pre-industrial concentrations; two models have made such calculations available at this writing (HadGEM3 (Walters et al., 2019) for experiment *amip* and IPSL-CM6A for experiment *historical*). We have also made diagnostics radiation calculations in GFDL's AM4 model using pre-industrial greenhouse gas concentrations during RFMIP "fixed-SST" experiments in which these concentrations are normally held constant at present-day values.

Results are provided in Table S2 in the supplemental material. A decade ago Andrews and Forster (2008) found that the presence of clouds reduced longwave instantaneous radiative forcing from quadrupled CO_2 concentrations by amounts ranging

from 9 to 20%, depending on the model (see their Table S2). As the distribution of clouds simulated by climate models has continued to move closer to observations (Klein et al., 2013) the estimated impact on top-of-atmosphere forcing has grown while the range across models and experiments has decreased (in Table S2 it is 23.6% to 26.5%). Clouds have a similar impact on shortwave forcing at the surface and an even larger impact on longwave forcing at the surface, though weaker observational constraints on the vertical structure of clouds allow for greater diversity across models.

4.2 Accounting for one adjustment

As noted earlier, the measure of forcing most closely related to temperature response is effective radiative forcing: the sum of the instantaneous radiative forcing, computable with robust radiative transfer models, and adjustments made by the physical climate system in the absence of surface temperature change (Sherwood et al., 2015). Adjustments, like forcing, result from a difference in two states and so are not directly observable. Many adjustments involve changes to circulations and clouds across a range of scales (e.g. Bretherton, Blossey, & Jones, 2013; Merlis, 2015) and can only be assessed with dynamical models for which establishing benchmarks is impractical.

In the climate models used to assess the global magnitude and distributions of adjustments, the dominant adjustment to greenhouse gas forcing is consistently the cooling of the stratosphere, partly because various tropospheric adjustments counteract each other (Smith et al., 2018). This cooling was first noted by Manabe and Wetherald (1967) and identified as an adjustment to longwave forcing by Hansen, Sato, and Ruedy (1997). The magnitude of this adjustment can be computed to a good approximation by assuming that dynamical heating in the stratosphere is fixed (Fels, Mahlman, Schwarzkopf, & Sinclair, 1980): computing the radiative cooling rate of the stratosphere under baseline (present-day) conditions, assuming that this cooling is balanced by longwave heating, and then finding the temperature profile necessary to obtain the same cooling profile under changed greenhouse gas concentrations. We follow Myhre et al. (2006) and Etminan et al. (2016) in supplying this first-order estimate of adjustments, which we compute by iterating with GRTCODE model at somewhat reduced spectral resolution until radiative heating rates reach their values in the present-day atmosphere.

The impact of stratospheric adjustment on forcing estimates is provided in Table S3 in the supplemental material. Stratospheric temperature changes affect shortwave forcing only through the temperature dependence of spectroscopy and thus has impacts smaller than 10^{-3} W m^{-2} across all experiments. Longwave forcing at the surface is only modestly affected by stratospheric adjustments because the unperturbed troposphere is nearly opaque in the spectral regions in which greenhouse gases cool. At the top of the atmosphere, longwave forcing amplifies clear-sky radiative forcing by ozone by a factor of almost 1.9, and amplifies clear-sky radiative forcing by carbon dioxide by a factor which increases with CO_2 concentration (from about 53% for a halving of CO_2 to 70% for octupling). The amplification of nitrous oxide forcing is modest, while forcing by methane and halocarbons is modestly damped.

5 Constraints on radiative forcing

Previous work (e.g. Chung & Soden, 2015; Soden, Collins, & Feldman, 2018) has established that the instantaneous radiative forcing for a given change in atmospheric composition can vary widely among climate models. This diversity has two distinct sources: parameterization error and variety in the distributions of temperature, humidity, and clouds between models. By using accurate models across a representative set of observed conditions we have shown that the true value of clear-sky instantaneous

radiative forcing can be determined quite precisely, with all-sky estimates limited primarily by challenges in representing the co-variability of clouds and atmospheric state. This highlights the distinction between climate model diversity and true uncertainty in estimates of instantaneous radiative radiative forcing. Adjustments arising from greenhouse gas forcing, however, remain a currently-irreducible source of uncertainty in attempts to estimate the true effective radiative forcing to which our planet has been subject and a source of poorly-constrained diversity among model estimates of effective radiative forcing.

Two caveats apply to our estimates of clear-sky instantaneous radiative forcing. First, RFMIP explores parameterization in perturbations around present-day conditions, so that our estimates of instantaneous radiative forcing are based on present-day distributions of temperature and humidity. Since forcing depends modestly on both quantities (Huang et al., 2016) our estimates of forcing are slightly enhanced relative to calculations that use pre-industrial conditions. Second, in the interests of highlighting model error in the representation of absorption by gases, the *rad-irf* protocol specifies spectrally-constant surface albedo and emissivity as obtained from ERA-Interim. Shortwave forcing at the top of the atmosphere, which arises from the sensitivity to radiation reflected at the surface to greenhouse gases, can be dramatically overestimated if the surface albedo is overestimated in the spectral range affected by a given gas (Oreopoulos et al., 2012). The small values of shortwave forcing in Table 3.2 suggest that the simple treatment of surface albedo is not likely to cause a large error, but accounting for spectral variations in surface albedo would be a useful exercise.

The agreement in global-mean instantaneous radiative among reference models, though encouraging, is consistent with almost 30 years of experience: Ellingson et al. (1991), for example, report that most of their line-by-line results for flux agree to within 1%. The agreement arises partly because radiative forcing, as the difference between two calculations, is also less sensitive to assumptions or subtle differences between models because many variations cancel out (Mlynczak et al., 2016). In our data set, however, the level of agreement in fluxes across models at the atmosphere's boundaries is commensurate with the variability in forcing estimates. The agreement in both fluxes and forcing arises because the models rely on the same underlying physics applied to small variants around the same spectroscopic data, so that the accuracy is limited by current spectroscopic knowledge more than by the ability to calculate fluxes from that knowledge. So while spectroscopic knowledge is now demonstrable more complete than it was 30 years ago (Mlawer & Turner, 2016), small variations in forcing estimates – high precision – should be understood as being being conditioned on this knowledge rather than evidence of true accuracy.

Acknowledgments

All results for RFMIP experiment *rad-irf* are available on the Earth System Grid Federation (or at least will be by the time of publication - getting non-standard data published challenges the system). Preliminary data for Table S3 were provided by Tim Andrews and Alejandro Bodas-Salcedo of the UK Met Offce. Python scripts to produce the figures and tables (and the data, for the moment) are available at `https://owncloud.gwdg.de/index.php/s/5oX1PoqLcijzlSI` and will be placed in a repository before acceptance. ERA-Interim data were obtained from `https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim`. SOCRATES is available from `https://code.metoffice.gov.uk/trac/socrates` under an open source license but requires a free account from the UK Met Office to access the website. This work was financially supported by the Regional and Global Climate Modeling Program of the US Department of Energy Office of Environmental and Biological Sciences (grants DE-SC0012549 to RP and DE-SC0012399 to EJM). SB was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research

Foundation) under Germany's Excellence Strategy - EXC 2037 "Climate, Climatic Change, and Society" - Project Number: 390683824, contributing to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg.

References

Alvarado, M. J., Payne, V. H., Mlawer, E. J., Uymin, G., Shephard, M. W., Cady-Pereira, K. E., ... Moncet, J. L. (2013). Performance of the Line-By-Line Radiative Transfer Model (LBLRTM) for temperature, water vapor, and trace gas retrievals: recent updates evaluated with IASI case studies. *Atmos. Chem. Phys.*, *13*(14), 6687–6711. doi: 10.5194/acp-13-6687-2013

Andrews, T., & Forster, P. M. (2008). CO_2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.*, *35*(4), D14317. doi: 10.1029/2007GL032273

Bretherton, C. S., Blossey, P. N., & Jones, C. R. (2013). Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases. *J. Adv. Model. Earth Syst.*, *5*(2), 316–337. doi: 10.1002/jame.20019

Buehler, S. A., Mendrok, J., Eriksson, P., Perrin, A., Larsson, R., & Lemke, O. (2018). ARTS, the Atmospheric Radiative Transfer Simulator – version 2.2, the planetary toolbox edition. *Geosci. Model Dev.*, *11*(4), 1537–1556. doi: 10.5194/gmd-11-1537-2018

Chung, E.-S., & Soden, B. J. (2015). An assessment of methods for computing radiative forcing in climate models. *Environ. Res. Lett.*, *10*(7), 074004. doi: 10.1088/1748-9326/10/7/074004

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., ... Brown, P. D. (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, *91*(2), 233–244. doi: 10.1016/j.jqsrt.2004.05.058

Collins, W. D., Ramaswamy, V., Schwarzkopf, M. D., Sun, Y., Portmann, R. W., Fu, Q., ... Zhong, W. Y. (2006). Radiative forcing by well-mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). *J. Geophys. Res.*, *111*(D14), D14317. doi: 10.1029/2005JD006713

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Royal Met. Soc.*, *137*(656), 553–597. doi: 10.1002/qj.828

Dudhia, A. (2017). The Reference Forward Model (RFM). *J. Quant. Spectrosc. Radiat. Transfer*, *186*, 243–253. doi: 10.1016/j.jqsrt.2016.06.018

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quart. J. Royal Met. Soc.*, *122*(531), 689–719. doi: 10.1002/qj.49712253107

Ellingson, R. G., Ellis, J., & Fels, S. (1991). The intercomparison of radiation codes used in climate models: Long wave results. *J. Geophys. Res.*, *96*(D5), 8929–8953. doi: 10.1029/90JD01450

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.*, *43*(24), 12,614–12,623. doi: 10.1002/2016GL071930

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D., & Sinclair, R. W. (1980). Strato-

spheric Sensitivity to Perturbations in Ozone and Carbon Dioxide: Radiative and Dynamical Response. *J. Atmos. Sci.*, *37*(10), 2265–2297. doi: 10.1175/1520-0469(1980)037⟨2265:SSTPIO⟩2.0.CO;2

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., ... Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *J. Geophys. Res.*, *121*(20), 12,460–12,475. doi: 10.1002/2016JD025320

Freckleton, R. S., Highwood, E. J., Shine, K. P., Wild, O., Law, K. S., & Sanderson, M. G. (1998). Greenhouse gas radiative forcing: Effects of averaging and inhomogeneities in trace gas distribution. *Quart. J. Royal Met. Soc.*, *124*(550), 2099–2127. doi: 10.1002/qj.49712455014

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., ... Zak, E. J. (2017). The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*, *203*, 3–69. doi: 10.1016/j.jqsrt.2017.06.038

Hansen, J. (2005). Efficacy of climate forcings. *J. Geophys. Res.*, *110*(D18), 1042. doi: 10.1029/2005JD005776

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *J. Geophys. Res.*, *102*(D6), 6831–6864. doi: 10.1029/96JD03436

Huang, Y., Tan, X., & Xia, Y. (2016). Inhomogeneous radiative forcing of homogeneous greenhouse gases. *J. Geophys. Res.*, *121*(6), 2780–2789. doi: 10.1002/2015JD024569

Iacono, M. J., Mlawer, E. J., Clough, S. A., & Morcrette, J.-J. (2000). Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *J. Geophys. Res.*, *105*(D11), 14873–14890. doi: 10.1029/2000JD900091

Jones, A. L., Feldman, D. R., Freidenreich, S., Paynter, D., Ramaswamy, V., Collins, W. D., & Pincus, R. (2017). A New Paradigm for Diagnosing Contributions to Model Aerosol Forcing Error. *Geophys. Res. Lett.*, *44*(23), 12,004–12,012. doi: 10.1002/2017GL075933

Kiel, M., Wunch, D., Wennberg, P. O., Toon, G. C., Hase, F., & Blumenstock, T. (2016). Improved retrieval of gas abundances from near-infrared solar FTIR spectra measured at the Karlsruhe TCCON station. *Atmos. Meas. Tech.*, *9*(2), 669–682. doi: 10.5194/amt-9-669-2016

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680. doi: 10.1126/science.220.4598.671

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Gleckler, P. J. (2013). Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *J. Geophys. Res.*, *118*(3), 1329–1342. doi: 10.1002/jgrd.50141

Manabe, S., & Wetherald, R. T. (1967). Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity. *J. Atmos. Sci.*, *24*(3), 241–259. doi: 10.1175/1520-0469(1967)024⟨0241:TEOTAW⟩2.0.CO;2

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., ... Weiss, R. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.*, *10*(5), 2057–2116. doi: 10.5194/gmd-10-2057-2017

Merlis, T. M. (2015). Direct weakening of tropical circulations from masked CO_2; radiative forcing. *Proc Natl Acad Sci USA*, *112*(43), 13167. doi: 10.1073/pnas.1508268112

Mlawer, E. J., Payne, V. H., Moncet, J. L., Delamere, J. S., Alvarado, M. J., & Tobin, D. C. (2012). Development and recent evaluation of the MTCKD model of continuum absorption. *Phil. Trans. Royal Soc. A*, *370*(1968), 2520–2556. doi: 10.1021/jp710066f

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997).

Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, *102*(D14), 16663–16682. doi: 10.1029/97JD00237

Mlawer, E. J., & Turner, D. D. (2016). Spectral Radiation Measurements and Analysis in the ARM Program. *Meteorological Monographs*, *57*, 14.1–14.17. doi: 10.1175/AMSMONOGRAPHS-D-15-0027.1

Mlynczak, M. G., Daniels, T. S., Kratz, D. P., Feldman, D. R., Collins, W. D., Mlawer, E. J., ... Mast, J. C. (2016). The spectroscopic foundation of radiative forcing of climate by carbon dioxide. *Geophys. Res. Lett.*, *43*(10), 5318–5325. doi: 10.1002/2016GL068837

Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, *25*(14), 2715–2718. doi: 10.1029/98GL01908

Myhre, G., Stordal, F., Gausemel, I., Nielsen, C. J., & Mahieu, E. (2006). Line-by-line calculations of thermal infrared radiation representative for global condition: CFC-12 as an example. *J. Quant. Spectrosc. Radiat. Transfer*, *97*(3), 317–331. doi: 10.1016/j.jqsrt.2005.04.015

Oreopoulos, L., Mlawer, E., Delamere, J., Shippert, T., Cole, J., Fomin, B., ... Rossow, W. B. (2012). The Continual Intercomparison of Radiation Codes: Results from Phase I. *J. Geophys. Res.*, *117*, D06118. doi: 10.1029/2011JD016821

Pincus, R., Forster, P. M., & Stevens, B. (2016). The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6. *Geosci. Model Dev.*, *9*, 3447–3460. doi: 10.5194/gmd-9-3447-2016

Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing Accuracy, Efficiency, and Flexibility in Radiation Calculations for Dynamical Models. *J. Adv. Model. Earth Syst.*, *11*, 3074–3089. doi: 10.1029/2019MS001621

Pincus, R., Mlawer, E. J., Oreopoulos, L., Ackerman, A. S., Baek, S., Brath, M., ... Schwarzkopf, D. M. (2015). Radiative flux and forcing parameterization error in aerosol-free clear skies. *Geophys. Res. Lett.*, *42*(13), 5485–5492. doi: 10.1002/2015GL064291

Ptashnik, I. V., McPheat, R. A., Shine, K. P., Smith, K. M., & Williams, R. G. (2011, August). Water vapor self-continuum absorption in near-infrared windows derived from laboratory measurements. *J. Geophys. Res.*, *116*(D16), 488. doi: 10.1029/2011JD015603

Ptashnik, I. V., Petrova, T. M., Ponomarev, Y. N., Shine, K. P., Solodov, A. A., & Solodov, A. M. (2013). Near-infrared water vapour self-continuum at close to room temperature. *J. Quant. Spectrosc. Radiat. Transfer*, *120*, 23–35. doi: 10.1016/j.jqsrt.2013.02.016

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P. F., ... Wagner, G. (2013). The HITRAN2012 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*, *130*(0), 4–50. doi: 10.1016/j.jqsrt.2013.07.002

Rotstayn, L. D., & Penner, J. E. (2001). Indirect Aerosol Forcing, Quasi Forcing, and Climate Response. *J. Climate*, *14*(13), 2960–2975. doi: 10.1175/1520-0442(2001)014⟨2960:IAFQFA⟩2.0.CO;2

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., & Stevens, B. (2015). Adjustments in the forcing-feedback framework for understanding climate change. *Bull. Amer. Meteor. Soc.*, *96*, 217–228. doi: 10.1175/BAMS-D-13-00167.1

Smith, C. J., Kramer, R. J., Myhre, G., Forster, P. M., Soden, B. J., Andrews, T., ... Watson-Parris, D. (2018, November). Understanding Rapid Adjustments to Diverse Forcing Agents. *Geophys. Res. Lett.*, *45*(21), 12,023–12,031. doi: 10.1029/2018GL079826

Soden, B. J., Collins, W. D., & Feldman, D. R. (2018). Reducing uncertainties in

climate models. *Science*, *361*(6400), 326–327. doi: 10.1126/science.aau1864

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., ... Zerroukat, M. (2019). The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.*, *12*(5), 1909–1963. doi: 10.5194/gmd-12-1909-2019

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., ... Watanabe, M. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geosci. Model Dev.*, *10*(1), 359–384. doi: 10.5194/gmd-10-359-2017

Wilcox, L. J., Hoskins, B. J., & Shine, K. P. (2011). A global blended tropopause based on ERA data. Part I: Climatology. *Quart. J. Royal Met. Soc.*, *138*(664), 561–575. doi: 10.1002/qj.951

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B. (2018). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies. *J. Adv. Model. Earth Syst.*, *10*(3), 735–769. doi: 10.1002/2017MS001209