Benchmark calculations of radiative forcing by greenhouse gases

Robert Pincus[1,1], Stefan Alexander Buehler[2,2], Manfred Brath[3,3], Omar Jamil[4,4], Frank Evans[5,5], James Manners[4,4], Raymond Menzel[6,6], Eli J. Mlawer[7,7], David Paynter[8,8], Rick Pernak[9,9], Cyril Crevoisier[10], and Yoann Tellier[10]

[1]University of Colorado/NOAA Earth System Research Laboratory
[2]Universität Hamburg
[3]University of Hamburg
[4]Met Office
[5]University of Colorado Boulder
[6]UCAR
[7]Atmospheric & Environmental Research
[8]GFDL/NOAA
[9]AER
[10]Laboratoire Météorologique Dynamique

November 30, 2022

Abstract

Changes in the concentration of greenhouse gases within the atmosphere lead to changes in radiative fluxes throughout the atmosphere. The value of this change, called the instantaneous radiative forcing, varies across climate models, due partly to differences in the distribution of clouds, humidity, and temperature across models, and partly due to errors introduced by approximate treatments of radiative transfer. 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 the representation of absorption and emission by greenhouse gases. The clear-sky instantaneous forcing by greenhouse gases 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 \si{\watt\per\square\meter}. Six contributing line-by-line models agree in their estimate of this forcing to within 0.025 \si{\watt\per\square\meter} while even recently-developed parameterizations have typical errors four or more times larger, suggesting both that the samples reveal true differences among line-by-line models and that parameterization error will be readily resolved. Agreement among line-by-line models is better in the longwave than in the shortwave where differing treatments of the water vapor vapor continuum affect estimates of forcing by carbon dioxide and methane. The impacts of clouds on instantaneous radiative forcing are roughly estimated, as are adjustments due to stratospheric temperature change. Adjustments are large only for ozone and for carbon dioxide, for which stratospheric cooling introduces modest non-linearity.

Benchmark calculations of radiative forcing by greenhouse gases

Robert Pincus[1,2], Stefan A. Buehler[3], Manfred Brath[3], Cyril Crevoisier[4], Omar Jamil[5], K. Franklin Evans[6], James Manners[5,7], Raymond L. Menzel[8,9], Eli J. Mlawer[10], David Paynter[8], Rick L. Pernak[10], Yoann Tellier[4]

[1]Cooperative Institute for Environmental Studies, University of Colorado, Boulder, Colorado, USA
[2]NOAA Physical Sciences Lab, Boulder, Colorado, USA
[3]Meteorological Institute, Department of Earth Sciences, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany
[4]Laboratoire de Météorologie Dynamique, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France
[5]Met Office, Exeter, UK
[6]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado, USA
[7]Global Systems Institute, Exeter University, Exeter, UK
[8]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA
[9]University Corporation for Atmospheric Research, Princeton, New Jersey, USA
[10]Atmospheric and Environmental Research, Lexington, Massachusetts, USA

Key Points:

- Mean clear-sky instantaneous radiative forcing by greenhouse gases is computed with six benchmark models using 100 atmospheric profiles.
- Sampling error is several times smaller than the level of disagreement among models, which is itself smaller than parameterization error.
- The impacts of clouds and stratospheric adjustment are roughly estimated; adjustments are large only for carbon dioxide and ozone.

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 throughout the atmosphere. The value of this change, called the instantaneous radiative forcing, varies across climate models, due partly to differences in the distribution of clouds, humidity, and temperature across models, and partly due to errors introduced by approximate treatments of radiative transfer. 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 the representation of absorption and emission by greenhouse gases. The clear-sky instantaneous forcing by greenhouse gases 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 $\mathrm{W\,m^{-2}}$. Six contributing line-by-line models agree in their estimate of this forcing to within 0.025 $\mathrm{W\,m^{-2}}$ while even recently-developed parameterizations have typical errors four or more times larger, suggesting both that the samples reveal true differences among line-by-line models and that parameterization error will be readily resolved. Agreement among line-by-line models is better in the longwave than in the shortwave where differing treatments of the water vapor vapor continuum affect estimates of forcing by carbon dioxide and methane. The impacts of clouds on instantaneous radiative forcing are estimated from climate model simulations. The adjustment due to stratospheric temperature change by assuming fixed dynamical heating. Adjustments are large only for ozone and for carbon dioxide, for which stratospheric cooling introduces modest non-linearity.

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 and/or temperature 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 et al., 2016) characterizes the forcing to which models are subject using "fixed-SST" experiments (Rotstayn & Penner, 2001; Hansen, 2005) in which atmospheric composition and land use are varied but the response of sea-surface temperature and sea ice concentrations 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, much of which remains even when model-specific adjustments are accounted for (e.g. Chung & Soden, 2015); intial results (Smith, Kramer, Myhre, et al., 2020) suggest that this diversity persists in CMIP6 models. Some of this variability is due a dependence on model state, especially how model-specific distributions of clouds and water vapor mask the radiative impact of changes in greenhouse gas concentrations (e.g. Huang et al., 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 (Ellingson et al., 1991; Collins et al., 2006; Oreopoulos et al., 2012; Pincus et al., 2015) to eliminate other causes of disagreement.

In an effort to untangle the contributions of state dependence and model error, RFMIP complements 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 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 reference 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 number 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 relatively 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 against which parameterizations can be evaluated and information as to how reasonable choices might affect those benchmarks given current understanding.

Here we describe the line-by-line calculations made for RFMIP and exploit them to move towards benchmark estimates of the true radiative forcing to which the earth has been subject due to increases in well-mixed greenhouse gases. We describe the construction of a small set of atmospheric profiles that can be used to accurately reproduce global-mean, annual-mean clear-sky instantaneous radiative forcing by greenhouse gases. We summarize the reference calculations supplied to date and highlight the values of clear-sky instantaneous radiative forcing for a range of changes in atmospheric composition relative to pre-industrial conditions. We show that sampling error from the small set of profiles is small enough that small differences among line-by-line calculations can be resolved, while variance among reference models is still less than even modern parameterized treatments, suggesting the the experiments can identify true variability across line-by-line models and parameterization error. We then cautiously extend these benchmark estimates towards more useful estimates that include the impact of clouds and adjustments.

2 Making global-mean benchmarks practical

Increasing computing power and more flexible software have made large-scale line-by-line calculations increasingly practical. Indeed RFMIP effort to diagnose errors in instantaneous radiative forcing by aerosols applies 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 et al., 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 et al., 1998).

Here we describe the construction of a set of atmospheric profiles designed to determine *error* in global-mean clear-sky instantaneous radiative forcing, obtained using a reference model on a very large number of atmospheric and surface conditions to determine this forcing and choosing a subset of these conditions that minimizes the sampling error across a range of measures. As we demonstrate below, the same set of pro-

Table 1. Perturbations around present-day (PD) conditions used to identify representative profiles. These are similar to, but not the same as, the perturbations used in RFMIP experiment *rad-irf* for reasons described in the text. Perturbations are applied to each profile drawn from ERA-Interim profile set. Carbon dioxide concentrations are relative to a pre-industrial (PI) volume mixing ratio of 278 ppmv. GHG refers to well-mixed greenhouse gases.Temperature T and relative humidity RH perturbations (12, 13) use the average of two models from the CMIP5 archive (GFDL-CM3 and GFDL-ESM2G) with relatively low and high climate sensitivities, respectively.

	Perturbation
1	PI 0.5×CO_2
2	PI 2×CO_2
3	PI 3×CO_2
4	PI 8×CO_2
5	PI CO_2 (278 ppmv)
6	PI CH_4 (0.722 ppmv)
7	PI N_2O (0.273 ppmv)
8	PI HFC (all HFC at zero)
9	PI O_3 (from CMIP6 PI ozone file)
10	PD +4K temperature, no H_2O change
11	PD +20% humidity
12	PI T, RH, O_3, GHG
13	2095 RCP8.5 T, RH, O_3, GHG
14	PI O_3, GHG
15	PI O_3, GHG, but PI 4×CO_2
16	2095 Avg Sens RCP4.5 O_3, GHG
17	2095 Avg Sens RCP8.5 O_3, GHG

files 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 Table 2.1) 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.

Our aim is to reproduce the mean of a set of reference fluxes, fully resolved in space and time and across the electromagnetic spectrum, computed for present-day conditions and each perturbation. The fluxes are computed using the UK Met Office SOCRATES (Suite Of Community RAdiative Transfer codes based on Edwards & Slingo, 1996) configured as a narrow-band model with a very high-resolution k-distribution with 300 bands in the longwave and 260 bands in the shortwave (Walters et al., 2019). This configuration agrees quite well with line-by-line models (e.g Pincus et al., 2015) and is one of the benchmark models described in Section 3.1. The spectral overlap of gases is treated with equivalent extinction with corrected scaling. Clouds and aerosols are not considered, consistent with the protocol for RFMIP experiment *rad-irf*.

We also compute fluxes for these sets of atmospheric conditions with an approximate model: RRTMG (Mlawer et al., 1997; Iacono et al., 2000), 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 minimizes sampling error in the global, annual mean obtained from the full calculation. To identify such a subset 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 above and below the tropopause (the level of which is determined by Wilcox et al., 2011); each quantity is computed for both longwave and shortwave fluxes. We guard against compensating errors related to temperature, humidity, and surface albedo and emissivity by further considering 9 roughly equal-area latitude bands centered on the equator. We choose an l^2 norm so that

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

where $\Delta F_{l,p,q}$ describes the average change in flux or flux divergence, as computed with the reference model over the full set of profiles, between perturbation p and present-day conditions in latitude band l for quantity q, and $\widehat{\Delta F}_{l,p,q}$ the sampled estimate of the same quantity. 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 et al., 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 profiles, not shown). The choice of profiles is reasonably robust to the choice of model: sampling error in the independent estimate of mean radiative forcing with RRTMG is only modestly larger (15% for 100 profiles) than for calculations with the narrow-band configuration of SOCRATES.

Profiles chosen to minimize sampling error in mean radiative forcing also provide accurate estimates of parameterization error $\mathcal{E} = \Delta\tilde{F} - \Delta F$ in that forcing, where $\Delta\tilde{F}$

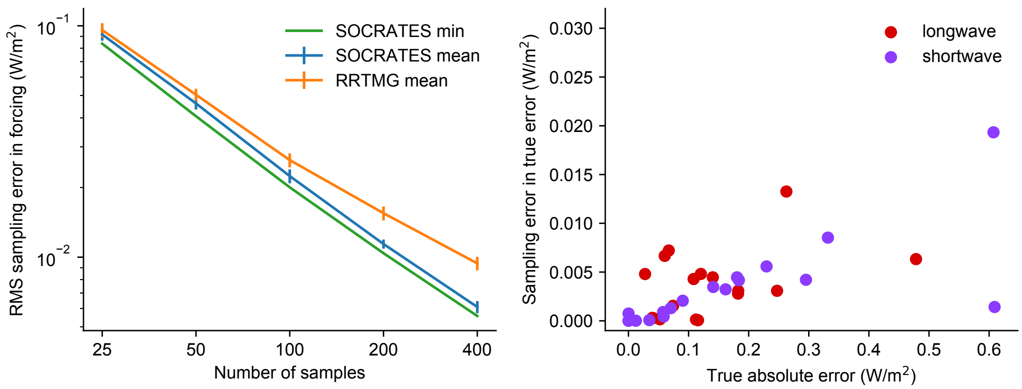

Figure 1. Left: values of the cost function O, an aggregate measure of error across regions, changes in atmospheric conditions, and measures of flux (Eq. 1) as a function of the number of optimal profiles. The simulated annealing method used to chose the profiles is stochastic; the mean and standard deviation across realizations is shown along with the value of sample error from the best-fit realization used in further calculations. The choice of profiles based on reference radiative transfer calculations ("SOCRATES") is robust, producing only modestly larger sampling errors for approximate calculations ("RRTMG"). Right: Absolute value of the sampling error $\hat{\mathcal{E}} - \mathcal{E}$ in estimates of the approximation error $\mathcal{E} = \Delta\tilde{F} - \Delta F$ sought by RFMIP. Errors shown are for the mean of 100 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 is almost always less than 0.01 $\mathrm{W\,m^{-2}}$.

is a computation made with an approximate model. Fig. 1 shows the sampling error $\hat{\mathcal{E}} - \mathcal{E}$ 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 profile 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 profiles 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 profiles for both longwave and shortwave calculations means that the sun is below the horizon for roughly half the set of profiles.

3 Radiation calculations with reference models

Experiment *rad-irf* requests fluxes for these 100 profiles and for 17 perturbations around present-day conditions, including changes in greenhouse gas concentrations, temperature, and humidity (see tables 3 and 4 in Pincus et al., 2016). Below we focus on the thirteen experiments in which gas concentrations alone are changed.

3.1 Contributions and variants

To date six benchmark models have contributed results: 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 narrow-band configuration 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; GRTCODE, a new line-by-line code developed at GFDL; and 4AOP (Scott & Chédin, 1981; Chéruy et al., 1995), provided by the Laboratoire de Météorologie Dynamique. Half the models use spectroscopic information from HITRAN 2012 (Rothman et al., 2013), while GRTCODE results are based on HITRAN 2016 (Gordon et al., 2017), 4AOP uses GEISA 2015 (Jacquinet-Husson et al., 2016), and LBLRTM employs 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 six models provide eighteen sets of longwave fluxes and nine 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: the change in net downward flux at the top-of-atmosphere (TOA) and the 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 (forcing at the top-of-atmosphere, with the atmosphere, and at the surface, for longwave and shortwave) is less than 0.025 $W\,m^{-2}$ with the exception of LW absorption, where the standard deviation is 0.033 $W\,m^{-2}$. There is no systematic variation across forcing variants, indicating that the equivalent concentrations accurately sum-

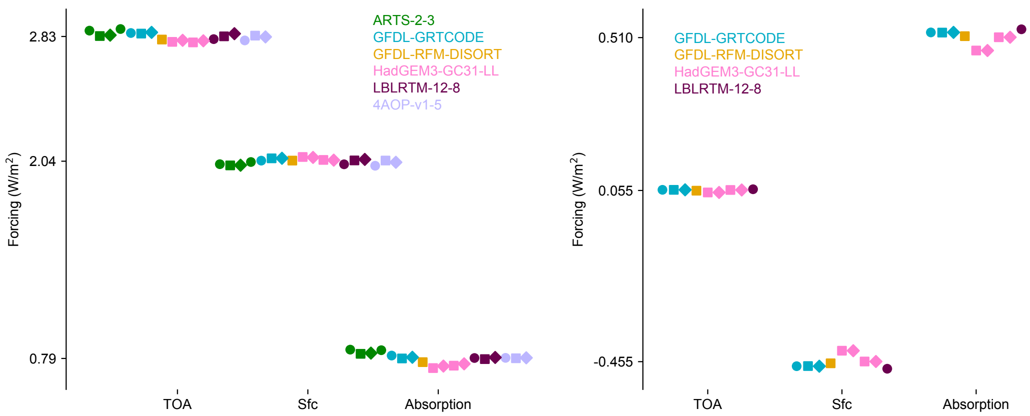

Figure 2. Global, annual mean instantaneous clear-sky radiative forcing by greenhouse gases at present-day, relative to pre-industrial conditions, as computed by benchmark radiative transfer models. Longwave results are on the left, shortwave results on the right, with the reference model denoted by 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.

marize 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 et al., 2011, 2013). Absorption 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 masks changes in opacity due to well-mixed greenhouse gases and reduces the forcing at the surface 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. Variability across models and forcing and physics variants, in both longwave and shortwave forcing calculations, increases with the magnitude of the forcing (Figure 3).

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. (Sampling er-

Table 2. Instantaneous radiative forcing computed as the mean across all available benchmark models, forcing variants, and physics variants, in $\mathrm{W\,m^{-2}}$. Forcing is defined as net downward flux under perturbed conditions minus net downward flux 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	2.830	2.040	0.055	-0.455
Future	7.377	5.542	0.355	-1.393
Last Glacial Maximum	-2.384	-1.416	-0.065	0.316
Computed as negative difference from perturbation "PD"				
Present-day CO_2	1.308	0.929	0.029	-0.165
Present-day CH_4	0.613	0.275	0.055	-0.242
Present-day N_2O	0.205	0.088	0.002	-0.011
Present-day O_3	0.129	0.325	-0.032	-0.033
Present-day halocarbons	0.534	0.393	0.000	-0.001
Computed as difference from perturbation "PI CO2"				
$\frac{1}{2}\times CO_2$	-2.695	-1.790	-0.050	0.274
$2\times CO_2$	2.709	1.978	0.064	-0.367
$3\times CO_2$	4.302	3.260	0.110	-0.629
$4\times CO_2$	5.436	4.252	0.146	-0.840
$8\times CO_2$	8.201	7.035	0.252	-1.442

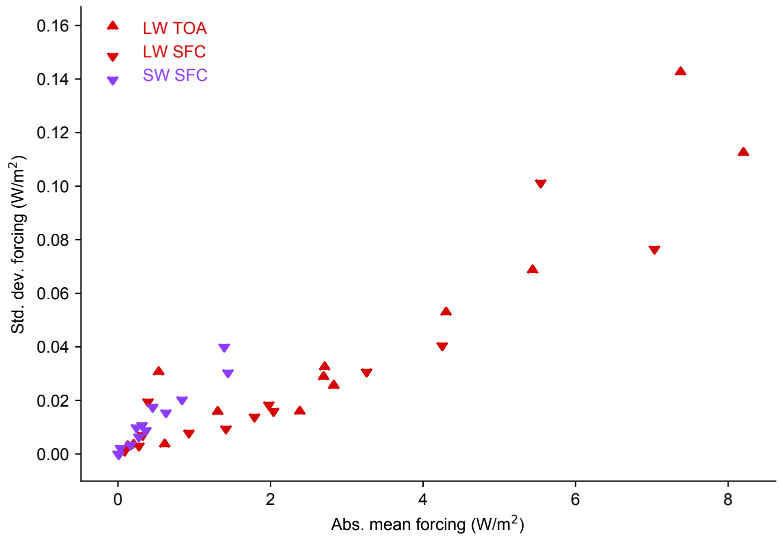

Figure 3. Standard deviation in estimates of global-mean instantaneous radiative forcing by greenhouse gases as a function of the absolute value of mean forcing across 18 benchmark calculations in the longwave (red) and nine in the shortwave (purple). Top-of-atmosphere forcing is indicated with an upward-pointing triangle; forcing at the surface with a downward-pointing triangle. Only forcing at the surface is shown for the shortwave. Agreement across models, forcing variants, and model physics variants increases with the mean forcing but it roughly two orders of magnitude smaller than the mean forcing across longwave experiments. Shortwave experiments are a factor of 2-3 more variable, partly driven by different treatments of near-infrared water vapor continuum. The figure illustrates agreement with respect to changed greenhouse gas concentrations; perturbations in experiment *rad-irf* in which temperature and/or humidity changes are omitted.

ror is smaller than the difference across reference models; see Figure 1). Figure 4, which compares error from two 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 spectral-resolution version of SOCRATES, as used in the HadGEM model, and for the newly-developed RTE+RRTMGP code (Pincus et al., 2019) which is trained on calculations with LBLRTM v12.8. These parameterizations 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.

4 Towards effective radiative forcing

RFMIP experiment *rad-irf* was designed to assess parameterization error but the benchmark calculations might also be exploited to refine knowledge of the radiative forcing experienced by Earth due to various composition changes. Two conceptually different steps are required, both of which are likely to make the estimate substantially less certain. One is accounting for the impact of clouds, which requires radiative calculations over the large range of imperfectly-characterized cloud properties. The other is account-

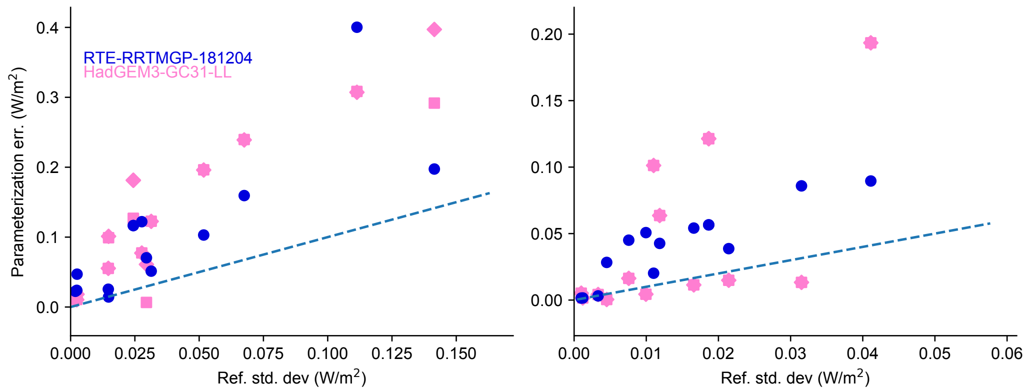

Figure 4. Absolute error in instantaneous radiative forcing (longwave at the top of atmosphere on the left, shortwave at the surface on the right) as computed by two parameterizations, both based on current spectroscopic information, as a function of amount of disagreement across the reference models. 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 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.

ing for adjustments (see Section 1) which introduces conceptually more uncertain non-radiative calculations. The long history of efforts to establish high-precision estimate of forcing by greenhouse gases (e.g., most recently, Myhre et al., 2006; Etminan et al., 2016) provides a point of reference for any efforts to leverage RFMIP calculations.

4.1 Accounting for clouds

Clouds modulate radiative forcing by greenhouse gases by screening changes in concentration behind the cloud. The degree to which clouds obscure greenhouse gas forcing depends primarily on the cloud optical depth (though longwave emissivity and shortwave reflectance and transmittance). Top of atmosphere forcing is also modulated by surface properties and cloud top height or pressure; surface forcing is modulated by cloud base height. Accounting for clouds in estimates of radiative forcing by greenhouse gases requires characterizing the wide variation in these properties in space and time. Observations from passive satellite sensors offer the best sampling of global variations but provide much stronger constraints on the quantities that affect top-of-atmosphere forcing than surface forcing. Previous efforts to establish benchmarks for radiative forcing (e.g. Etminan et al., 2016; Myhre et al., 2006) have used 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 at the top of the atmosphere are thought to be of order 1% although this error estimate has not been revisited since the 1990s (Myhre & Stordal, 1997; Freckleton et al., 1998). Errors in cloud impacts on surface forcing have not been assessed.

We hope to revisit this question in future work. One important question will be whether computational effort is better spent in sampling the co-variability of cloud properties with other atmospheric and surface properties or in high-spectral resolution cal-

Table 3. Ratio of all-sky to clear-sky instantaneous radiative forcing, at the top-of-atmosphere and the surface, across a range of models and experiments in CMIP6. Clear-sky and all-sky (including clouds) fluxes are computed using a second radiative transfer calculation in which the forcing agents are modified for diagnostics purposes. Results from HadGEM3 and IPSL-CM6A use diagnostic calculations requested for CFMIP in which CO_2 concentrations are quadrupled from pre-industrial values. Values from GFDL-CM4, performed for this work, are computed by setting forcing agents to pre-industrial values in three RFMIP fixed-SST integrations. Results from HadGEM3 are preliminary and may be revised before they are made publicly available. Shortwave forcing at the top of atmosphere is so small that inferences of cloud masking are quite uneven across models.

	HadGEM3-GC31-LL	IPSL-CM6A-LR		GFDL-CM4	
experiment	amip	historical	4xCO2	GHG	anthro
LW TOA	0.764	0.735	0.763	0.757	0.767
LW SFC	0.622	0.608	0.696	0.689	0.680
SW SFC	0.718	0.732	0.711	0.853	0.714

culations to limit approximation errors. These questions, though, are beyond the scope of what can be accomplished with reference model calculations to *rad-irf*. 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 for experiment *amip* and IPSL-CM6A for experiment *historical*). We have also made diagnostic radiation calculations in GFDL's AM4 model (Zhao et al., 2018) using pre-industrial greenhouse gas concentrations during RFMIP "fixed-SST" experiments in which these concentrations are normally held constant at present-day values; these follow the protocol described by (Lin et al., 2017).

Results are provided in Table 4.1. A decade ago Andrews & 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 (e.g. 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 4.1 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 adjustments from temperature changes in the stratosphere

As explained in Section 1 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. Gregory & Webb, 2008; Bretherton et al., 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 (e.g. Smith et al., 2018; Smith, Kramer, & Sima, 2020). This cooling was first noted by Manabe & Wetherald (1967) and identified as an adjustment to longwave forcing by Hansen et al. (1997). As Shine & Myhre (2020) explain, increased concentrations of well-mixed greenhouse gases increase both emission by the stratosphere and absorption of radiation emitted from the troposphere. If the background atmosphere is optically thick in the spectral region in which the gas is active (e.g. for CO_2) additional warming from tropospheric emission is small and the stratosphere cools, enhancing instantaneous forcing at the top of the atmosphere, but if the the background atmosphere is optically thin (as for most halocarbons) the stratosphere may warm, damping the instantaneous forcing.

The magnitude of this adjustment can be computed to a good approximation by assuming that dynamical heating in the stratosphere is fixed (Ramanathan & Dickinson, 1979; Fels et al., 1980): computing the radiative cooling rate of the stratosphere under baseline (present-day) conditions, assuming that this cooling is balanced by dynamical heating, and then finding the temperature profile necessary to obtain the same net 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. We compute the adjustment caused by stratospheric temperature re-equilibration, assuming fixed dynamical heating, by iterating with GRTCODE model at reduced spectral resolution until radiative heating rates reach their values in the present-day atmosphere. The calculations assume a uniform tropopause pressure of 200 Pa and account for changes in both longwave and shortwave heating rates. For well-mixed greenhouse gases the impact of stratospheric temperature adjustment depends primarily on the spectral region in which the gas absorbs.

The impact of stratospheric temperature adjustment, expressed as the ratio of the change in flux due to temperature equilibration to the instantaneous longwave radiative forcing, is shown for a range of concentrations at present-day relative to pre-industrial conditions in Table 4.2. Stratospheric temperature changes from well-mixed greenhouse gases amplify (CO_2, N_2O) or damp (CH_4, halocarbons) forcing at the top of the atmosphere; for all gases but CO_2 the impact is just a few percent. Surface forcing is damped by a similar amount.

Carbon dioxide is a notable exception: the amplification of top-of-atmosphere forcing at present-day is more than 55%. This large adjustment occurs because the total forcing at the top-of-the-atmosphere is a balance between contributions from distinct spectral regions. Near the center of the 15 µm absorption band of CO_2 the atmosphere is optically thick and emission to space occurs in the stratosphere; increases CO_2 concentrations tends to increase outgoing longwave radiation because stratospheric temperature increases with height. Away from the band center the atmosphere is optically thin, emission is from the troposphere, and increasing concentrations acts to decrease outgoing longwave radiation. Net forcing is negative (see Table 3.2) because the the tropospheric contribution dominates. Stratospheric cooling damps the instantaneous forcing from the band center, allowing the optically-thin regions to dominate the change in top-of-atmosphere flux even more effectively. The adjustment also increases by 1.8% per $\mathrm{W\,m^{-2}}$ (Figure 5) so that effective radiative forcing may be modestly super-logarithmic in CO_2 concentrations even though the instantaneous radiative forcing is nearly perfectly logarithmic.

Stratospheric temperature adjustment nearly doubles the top-of-atmosphere instantaneous forcing from ozone but for quite different reasons. Ozone concentrations at present-

Table 4. Ratio of adjustment due to stratospheric temperature equilibration under the fixed dynamical heating assumption to instantaneous clear-sky longwave radiative forcing at the top of atmosphere and the surface for a range of forcing agents. Both forcing and stratospheric adjustment are computed using GFDL GRTCODE line-by-line model. Shortwave adjustments are all essentially zero.

Experiment	TOA	SFC
Present-day	0.31	-0.03
Present-day CO_2	0.57	-0.05
Present-day CH_4	-0.05	0.01
Present-day N_2O	0.03	-0.01
Present-day O_3	1.90	-0.06
Present-day halocarbons	-0.11	0.01

day vary substantially in the vertical, peaking in the stratosphere. As one consequence ozone acts to heat the stratosphere near the center of the 10 µm band and increases in ozone concentration in either the troposphere or stratosphere tend to decrease net radiation at the top of the atmosphere. The vertical distribution of change is also non-uniform: relative to pre-industrial conditions ozone concentrations have increased in the troposphere but decreased in the stratosphere. The modest positive forcing from present-day ozone relative to pre-industrial conditions results from a slightly larger decrease in outgoing radiation from tropospheric emission than can be balanced by increased emission from concentration reductions in the stratosphere. The stratosphere cools modestly despite because reduced concentrations of ozone because decreases in absorption of shortwave radiation are larger than the increases from enhanced longwave emission. This cooling, too, reduces the stratospheric contribution to forcing. Stratospheric adjustment of ozone is larger than for carbon dioxide, in a relative sense, only because the balance between stratosphere and troposphere is more even for instantaneous forcing.

5 Constraints on radiative forcing

Previous work (e.g. Chung & Soden, 2015; Soden et al., 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 error in perturbations around present-day conditions, so that our estimates of instantaneous radiative forcing are based on present-day distributions of temperature and humidity. Forcing depends modestly on both quantities (Huang et al., 2016) so 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

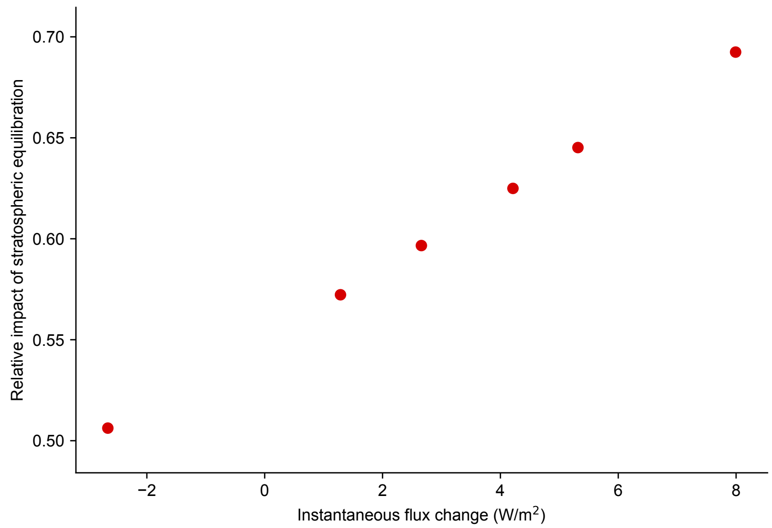

Figure 5. Ratio of stratospheric temperature adjustment to instantaneous radiative forcing at the top of the atmosphere for CO_2 perturbations ranging from 0.5× to 8× pre-industrial concentrations. Assuming that heating from atmospheric dynamic stays constant allows the computation of a new equilibrium temperature profile to be computed; this profile is colder (because the stratosphere is a more effective emitter) so the adjustment amplifies instantaneous radiative forcing. The magnitude of the adjustment depends modestly on the magntiude of the forcing itself, suggesting that effective radiative forcing by CO_2 may be slightly super-logarithmic in concentration even if the instantaneous radiative forcing is not.

surface albedo and emissivity as obtained from ERA-Interim. Shortwave forcing at the top of the atmosphere, which arises from the sensitivity to greenhouse gases of radiation reflected at the surface and transmitted through the atmosphere, 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 under present-day conditions varies by less than 0.6 $\mathrm{W\,m^{-2}}$ in the longwave and 0.7 $\mathrm{W\,m^{-2}}$ in the shortwave - smaller than the variability in forcing estimates, in a relative sense, by an order of magnitude. 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 demonstrably more complete than it was 30 years ago (Mlawer & Turner, 2016), small variations in forcing estimates – high precision – should be understood as being conditioned on this knowledge rather than evidence of true accuracy.

Acknowledgments

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's work is a contribution to Excellence Cluster CLICSS "Climate, Climatic Change, and Society", supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany?s Excellence Strategy - EXC 2037 - Project Number 390683824, and a contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. YT was financially supported by Thales Services and by the Centre National d'Études Spatiales. CC and YT gratefully acknowledge fruitful discussions with Raymond Armante, Nicolas Meilhac and Jean-Louis Dufresne.

Data availability

All results for RFMIP experiment *rad-irf* are available on the Earth System Grid Federation (searching for the experiment name is an effective way to find the data). Python scripts and Jupyter notebooks to produce the paper are available at `https://github.com/RobertPincus/rfmip-benchmark-paper-figures` and will be archived at Zenodo, with a DOI, on 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. Preliminary data for Table 4.1 were provided by Tim Andrews and Alejandro Bodas-Salcedo of the UK Met Office but will be derivable through data provided on the Earth System Grid.

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.

Andrews, T., & Forster, P. M. (2008, February). CO_2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.*, *35*(4), L04802. doi: 10.1029/2007GL032273

Bretherton, C. S., Blossey, P. N., & Jones, C. R. (2013, May). 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, January). 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

Chéruy, F., Scott, N., Armante, R., Tournier, B., & Chedin, A. (1995, June). Contribution to the development of radiative transfer models for high spectral resolution observations in the infrared. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *53*(6), 597–611. doi: 10.1016/0022-4073(95)00026-H

Chung, E.-S., & Soden, B. J. (2015, January). 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, March). 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, July). 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.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011, January). 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, January). 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, April). 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.

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016, December). 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, May). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958.

Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D., & Sinclair, R. W. (1980, October). Stratospheric 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, October). 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, July). 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, January). The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*, *203*, 3–69.

Gregory, J., & Webb, M. (2008, January). Tropospheric Adjustment Induces a Cloud Component in CO2 Forcing. *J. Climate*, *21*(1), 58–71. doi: 10.1175/2007JCLI1834.1

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

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

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

Iacono, M. J., Mlawer, E. J., Clough, S. A., & Morcrette, J.-J. (2000, June). 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

Jacquinet-Husson, N., Armante, R., Scott, N., Chédin, A., Crépeau, L., Boutammine, C., … Makie, A. (2016, September). The 2015 edition of the GEISA spectroscopic database. *Journal of Molecular Spectroscopy*, *327*, 31–72. doi: 10.1016/j.jms.2016.06.007

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

Kiel, M., Wunch, D., Wennberg, P. O., Toon, G. C., Hase, F., & Blumenstock, T. (2016, January). 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, May). 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, February). 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

Lin, P., Paynter, D., Ming, Y., & Ramaswamy, V. (2017, February). Changes of the Tropical Tropopause Layer under Global Warming. *Journal of Climate*, *30*(4), 1245–1258. doi: 10.1175/JCLI-D-16-0457.1

Manabe, S., & Wetherald, R. T. (1967, May). 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, May). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.*, *10*(5), 2057–2116.

Merlis, T. M. (2015, October). Direct weakening of tropical circulations from masked CO_2 radiative forcing. *Proc Natl Acad Sci USA*, *112*(43), 13167.

Mlawer, E. J., Payne, V. H., Moncet, J. L., Delamere, J. S., Alvarado, M. J., & Tobin, D. C. (2012, April). 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, July). 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, April). 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, May). 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, July). 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. (1997, May). Role of spatial and temporal variations in the computation of radiative forcing and GWP. *J. Geophys. Res.*, *102*(D10), 11181–11200. doi: 10.1029/97JD00148

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.

Oreopoulos, L., Mlawer, E., Delamere, J., Shippert, T., Cole, J., Fomin, B., …

Rossow, W. B. (2012, March). 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, January). 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.*, *6*(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, July). 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, May). 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

Ramanathan, V., & Dickinson, R. E. (1979, June). The Role of Stratospheric Ozone in the Zonal and Seasonal Radiative Energy Balance of the Earth-Troposphere System. *J. Atmos. Sci.*, *36*(6), 1084–1104. doi: 10.1175/1520-0469(1979)036⟨1084:TROSOI⟩2.0.CO;2

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P. F., ... Wagner, G. (2013, November). 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, July). 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

Scott, N. A., & Chédin, A. (1981). A fast line-by-line method for atmospheric absorption computations: The Automatized Atmospheric Absorption Atlas. *Journal of Applied Meteorology*, *20*, 802–812. doi: 10.1175/1520-0450(1981)020⟨0802:AFLBLM⟩2.0.CO;2

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., & Stevens, B. (2015, January). 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

Shine, K. P., & Myhre, G. (2020, March). The Spectral Nature of Stratospheric Temperature Adjustment and its Application to Halocarbon Radiative Forcing. *J. Adv. Model. Earth Syst.*, *12*(3). doi: 10.1029/2019MS001951

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., ... Forster, P. M. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics Discussions*, *2020*, 1–37. doi: 10.5194/acp-2019-1212

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

Smith, C. J., Kramer, R. J., & Sima, A. (2020, March). The HadGEM3-GA7.1 radiative kernel: The importance of awell-resolved stratosphere. *Earth System Science Data Discussions*. doi: 10.5194/essd-2019-254

Soden, B. J., Collins, W. D., & Feldman, D. R. (2018, July). 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, January). 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, January). 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, October). 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, March). 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