Benchmark Calculations of Radiative Forcing by Greenhouse Gases

Robert Pincus, Stefan Alexander Buehler, Manfred Brath, Cyril Crevoisier, Omar Jamil, K. Franklin Evans, James Manners, Raymond L. Menzel, Eli J. Mlawer, David Paynter, et al.

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033483

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], and Yoann Tellier[4]

[1]Cooperative Institute for Environmental Studies, University of Colorado Boulder, Boulder, CO, USA, [2]NOAA Physical Sciences Lab, Boulder, CO, 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, Paris, France, [5]Met Office, Exeter, UK, [6]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO, USA, [7]Global Systems Institute, Exeter University, Exeter, UK, [8]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [9]University Corporation for Atmospheric Research, Princeton, NJ, USA, [10]Atmospheric and Environmental Research, Lexington, MA, 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

Correspondence to:
R. Pincus,
Robert.Pincus@colorado.edu

Citation:
Pincus, R., Buehler, S. A., Brath, M., Crevoisier, C., Jamil, O., Franklin Evans, K., et al. (2020) Benchmark calculations of radiative forcing by greenhouse gases. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD033483. https://doi.org/10.1029/2020JD033483

Received 8 JUL 2020
Accepted 1 NOV 2020
Accepted article online 17 NOV 2020

Abstract Changes in concentrations of greenhouse gases lead to changes in radiative fluxes throughout the atmosphere. The value of this change, 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. Clear-sky instantaneous forcing by greenhouse gases is computed using a set of 100 profiles, selected from a reanalysis of present-day conditions, that represent the global annual mean forcing from preindustrial times to the present day with sampling errors of less than 0.01 W m^{-2}. Six contributing line-by-line models agree in their estimate of this forcing to within 0.025 W m^{-2} while even recently developed parameterizations have typical errors 4 or more times larger, suggesting both that the samples reveal true differences among line-by-line models and that parameterization error will be readily identifiable. Agreement among line-by-line models is better in the longwave than in the shortwave where differing treatments of the water 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, and the adjustment due to stratospheric temperature changes estimated by assuming fixed dynamical heating. Adjustments are large only for ozone and for carbon dioxide, for which stratospheric cooling introduces modest nonlinearity.

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 flux caused by 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 (Hansen, 2005; Rotstayn & Penner, 2001) 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); initial results (Smith et al., 2020)

suggest that this diversity persists in CMIP6 models. Some of this variability is due to 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*, that is, the change in flux in the absence of adjustments, as illustrated by comparisons that use prescribed atmospheric conditions to eliminate other causes of disagreement (Collins et al., 2006; Ellingson et al., 1991; Oreopoulos et al., 2012; Pincus et al., 2015).

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. 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, 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 toward 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 preindustrial 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 both true variability across line-by-line models and parameterization error. We then cautiously extend these benchmark estimates toward 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 the geographic distributions of temperature and water vapor are well characterized 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, then choosing a subset of these conditions that minimizes the sampling error across a range of measures. As we demonstrate below, the same set of profiles also provides an accurate sample of the parameterization or approximation error in radiative forcing.

Table 1
Perturbations Around Present-Day (PD) Conditions Used to Identify Representative Profiles

	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 +4 K 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

Note. 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 preindustrial (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.

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 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*.

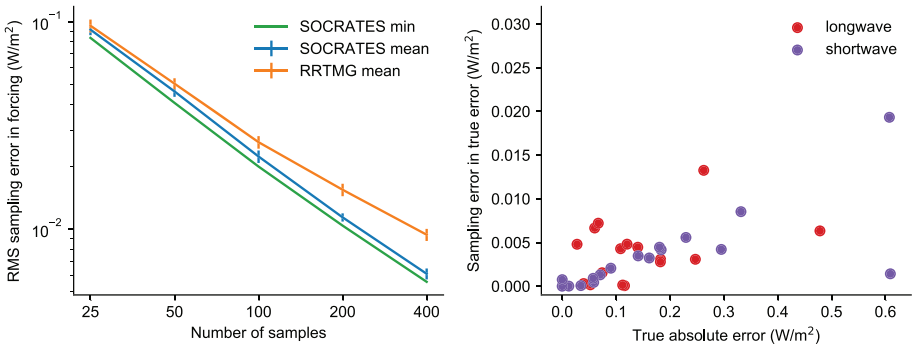

Figure 1. Left: values of the cost function O, an aggregate measure of error across regions, changes in atmospheric conditions, and measures of flux (Equation 1), as a function of the number of optimal profiles. The simulated annealing used to chose the profiles is stochastic; the mean and standard deviation across realizations are 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 W m^{-2} in the global, annual mean; sampling error is almost always less than 0.01 W m^{-2}.

We also compute fluxes for these sets of atmospheric conditions with an approximate model: RRTMG (Iacono et al., 2000; Mlawer et al., 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 minimizes sampling error in the global, annual mean obtained from the full calculation. To identify such a subset, we define 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, the change in top-of-atmosphere flux 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 nine roughly equal-area latitude bands centered on the equator. We choose an l^2 norm so that

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

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 $\Delta\tilde{F}$ is a computation made with an approximate model. Figure 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 W m^{-2} in the global, annual mean; sampling error in these estimates is almost always less than 0.01 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 W m^{-2} (cf. the true mean of $\sim$1,361/4 = 340.25 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 13 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 section 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 (Chéruy et al., 1995; Scott & Chédin, 1981), 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 18 sets of longwave fluxes and 9 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-134a. (Concentrations of water vapor and ozone are drawn from reanalysis, as described in section 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 preindustrial 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

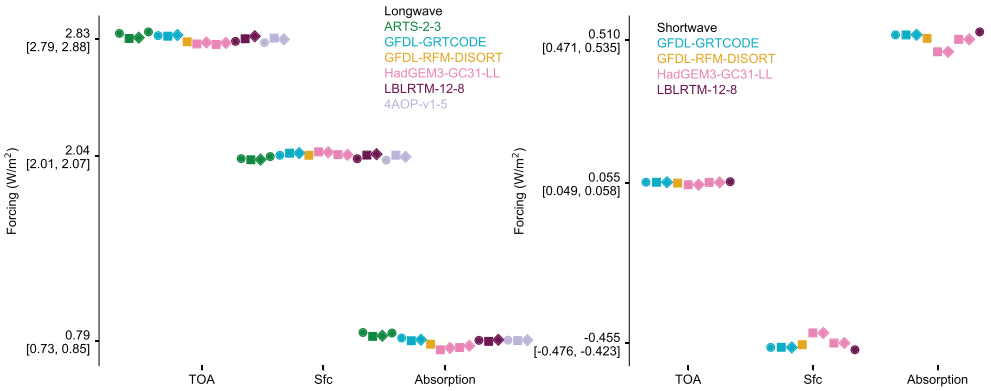

Figure 2. Global, annual mean instantaneous clear-sky radiative forcing by greenhouse gases at present day, relative to preindustrial conditions, as computed by benchmark radiative transfer models. Longwave results are on the left, shortwave results on the right, with the reference model denoted by the color. Tick labels show the mean value and range of all model results. 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.

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 is 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 for each of the six quantities (forcing at the TOA, 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 summarize the radiative impact of the neglected gases in the transition from preindustrial to present-day conditions.

Changes in shortwave flux between preindustrial 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 continuum 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 preindustrial 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 2. Variability across models, forcing variants, and model physics variants increases with the mean forcing (Figure 3) but is roughly 2 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.

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 error 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

Table 2
Instantaneous Radiative Forcing Computed as the Mean Across All Available Benchmark Models, Forcing Variants, and Physics Variants, in W m^{-2}

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 -("PI X" -"PD") for species X				
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"				
$1/2\times CO_2$	−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

Note. Forcing is defined as net downward flux under perturbed conditions minus net downward flux under preindustrial (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.

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, for the parameterization used in GFDL's AM4 model (Zhao et al., 2018), 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.

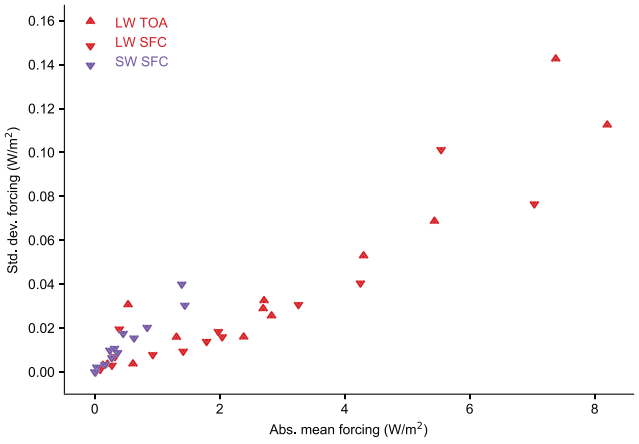

Figure 3. Standard deviation in estimates of global mean instantaneous radiative forcing by greenhouse gases as a function of the absolute value of mean forcing across 18 benchmark calculations in the longwave (red) and 9 in the shortwave (purple). Top-of-atmosphere forcing is indicated with an upward pointing triangle, forcing at the surface with a downward pointing triangle. Only forcing at the surface is shown for the shortwave. 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.

4. Toward Effective Radiative Forcing

RFMIP experiment *rad-irf* was designed to assess parameterization error but 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 accounting for adjustments (see section 1) which introduces conceptually more uncertain nonradiative calculations. The long history of efforts to establish high-precision estimate of forcing by greenhouse gases (e.g., most recently, Etminan et al., 2016; Myhre et al., 2006) provides a point of reference for any efforts to leverage RFMIP calculations.

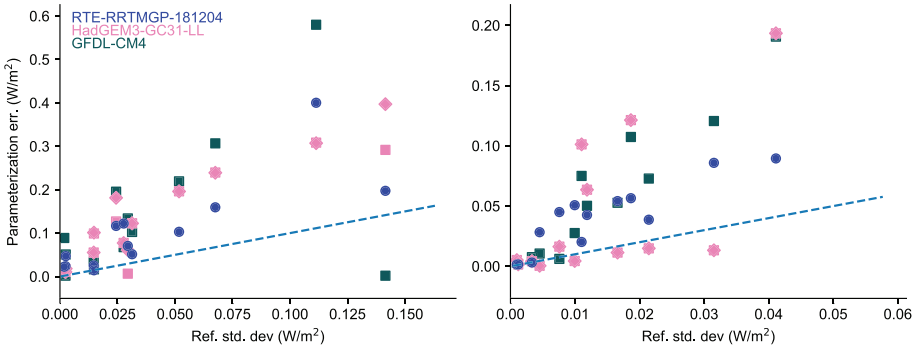

Figure 4. Absolute error in global mean instantaneous radiative forcing (longwave at the top of atmosphere on the left, shortwave at the surface on the right) as computed by two parameterizations, both based on current spectroscopic information, as a function of amount of disagreement across the reference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

4.1. Accounting for Clouds

Clouds modulate radiative forcing by greenhouse gases by screening the effects of changes in concentration behind the cloud. The degree to which clouds obscure greenhouse gas forcing depends primarily on the cloud optical depth (through longwave emissivity and shortwave reflectance and transmittance). Top-of-atmosphere forcing is also modulated by surface properties and, in the longwave, by cloud top height or pressure; longwave 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 section 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 (Freckleton et al., 1998; Myhre & Stordal, 1997). 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 covariability of cloud properties with other atmospheric and surface properties or in high-spectral resolution calculations 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 preindustrial 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 preindustrial 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 3. 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 (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 3 it is 23.6% to 26.5%). Clouds have a similar impact on shortwave forcing at the surface and an even

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

	HadGEM3-GC31-LL	IPSL-CM6A-LR		GFDL-CM4	
Experiment	Amip	Historical	4 x CO2	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

Note. 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 preindustrial values. Values from GFDL-CM4, performed for this work, are computed by setting forcing agents to preindustrial 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.

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 caused change of 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 et al., 2013; Gregory & Webb, 2008; 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, 2020). This cooling, which is driven by increased concentrations of CO_2, was first noted by Manabe and Wetherald (1967) and identified as an adjustment to longwave forcing by Hansen et al. (1997). As Shine and 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 (Fels et al., 1980; Ramanathan & Dickinson, 1979): 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. The estimate relies on assumptions which will be violated if stratospheric circulation changes very much, but the calculation does not rely on a dynamical model, so 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 reequilibration, 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 species at present-day relative to preindustrial conditions in Table 4. Stratospheric temperature changes from well-

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

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

Note. Both forcing and stratospheric adjustment are computed using GFDL GRTCODE line-by-line model. Shortwave adjustments are all essentially zero.

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; increased 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 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 W m^{-2} (Figure 5) so that effective radiative forcing is modestly superlogarithmic 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 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 nonuniform: Relative to preindustrial conditions, ozone concentrations have increased in the troposphere but decreased in the stratosphere. The modest positive forcing from present-day ozone relative to preindustrial 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 for reduced concentrations of ozone because absorption of both incoming solar shortwave radiation and upwelling terrestrial longwave radiation decreases. 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.

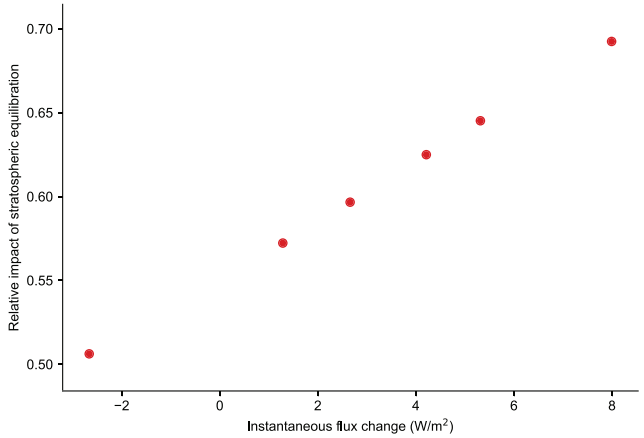

Figure 5. Ratio of stratospheric temperature adjustment to instantaneous radiative forcing at the top of the atmosphere for CO_2 perturbations ranging from 0.5× to 8× preindustrial concentrations. Assuming that heating from atmospheric dynamics stays constant allows the computation of a new equilibrium 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 is slightly superlogarithmic in concentration even if the instantaneous radiative forcing is not.

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. RFMIP experiment *rad-irf* and the benchmarks reported here make it possible to quantify parameterization error in instantaneous radiative forcing accurately, so that these two sources of diversity can be disentangled. But the diversity of climate model estimates is far larger than the true uncertainty. By using accurate radiative transfer models across a representative set of observed conditions, we have shown that the value of clear-sky instantaneous radiative forcing can be determined quite precisely. All-sky estimates are limited primarily by challenges in representing the covariability of clouds and atmospheric state. Adjustments arising from greenhouse gas forcing, because they reflect changes in circulation and atmospheric state that cannot be determined without using dynamical models, 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 preindustrial conditions as the base state. 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 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 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 W m^{-2} in the longwave and 0.7 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.

Data Availability Statement

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 are archived online (Zenodo via DOI: 10.5281/zenodo.4267190). ERA-Interim data were obtained online (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 3 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.

Acknowledgments
This work was financially supported by the Regional and Global Climate Modeling Program of the U.S. Department of Energy Office of Environmental and Biological Sciences (Grants DE-SC0012549 to R. P. and DE-SC0012399 to E. J. M.). S. B.'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 390683824 and a contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. Y. T. was financially supported by Thales Services and by the Centre National d'Études Spatiales. C. C. and Y. T. gratefully acknowledge fruitful discussions with Raymond Armante, Nicolas Meilhac, and Jean-Louis Dufresne.

References

Alvarado, M. J., Payne, V. H., Mlawer, E. J., Uymin, G., Shephard, M. W., Cady-Pereira, K. E., et al. (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. *Atmospheric Chemistry and Physics*, *13*(14), 6687–6711. https://doi.org/10.5194/acp-13-6687-2013

Andrews, T., & Forster, P. M. (2008). CO2 forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophysical Research Letters*, *35*, L04802. https://doi.org/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. *Journal of Advances in Modeling Earth Systems*, *5*, 316–337. https://doi.org/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. *Geoscientific Model Development*, *11*(4), 1537–1556. https://doi.org/10.5194/gmd-11-1537-2018

Chéruy, F., Scott, N. A., Armante, R., Tournier, B., & Chedin, A. (1995). 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. https://doi.org/10.1016/0022-4073(95)00026-H

Chung, E.-S., & Soden, B. J. (2015). An assessment of methods for computing radiative forcing in climate models. *Environmental Research Letters*, *10*(7), 074004. https://doi.org/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., et al. (2005). Atmospheric radiative transfer modeling: A summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *91*(2), 233–244. https://doi.org/10.1016/j.jqsrt.2004.05.058

Collins, W. D., Ramaswamy, V., Schwarzkopf, M. D., Sun, Y., Portmann, R. W., Fu, Q., et al. (2006). Radiative forcing by well?mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). *Journal of Geophysical Research*, *111*, D14317. https://doi.org/10.1029/2005JD006713

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dudhia, A. (2017). The reference forward model (RFM). *Journal of Quantitative Spectroscopy and Radiative Transfer*, *186*, 243–253. https://doi.org/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. *Quarterly Journal of the Royal Meteorological Society*, *122*(531), 689–719. https://doi.org/10.1002/qj.49712253107

Ellingson, R. G., Ellis, J., & Fels, S. (1991). The intercomparison of radiation codes used in climate models: Long wave results. *Journal of Geophysical Research*, *96*(D5), 8929–8953. https://doi.org/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. *Geophysical Research Letters*, *43*, 12,614–12,623. https://doi.org/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. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D., & Sinclair, R. W. (1980). Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response. *Journal of the Atmospheric Sciences*, *37*(10), 2265–2297. https://doi.org/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., et al. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, *121*, 12,460–12,475. https://doi.org/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. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 2099–2127. https://doi.org/10.1002/qj.49712455014

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., et al. (2017). The HITRAN2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *203*, 3–69. https://doi.org/10.1016/j.jqsrt.2017.06.038

Gregory, J., & Webb, M. (2008). Tropospheric adjustment induces a cloud component in CO2 forcing. *Journal of Climate*, *21*(1), 58–71. https://doi.org/10.1175/2007JCLI1834.1

Hansen, J. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*, 1042. https://doi.org/10.1029/2005JD005776

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research*, *102*(D6), 6831–6864. https://doi.org/10.1029/96JD03436

Huang, Y., Tan, X., & Xia, Y. (2016). Inhomogeneous radiative forcing of homogeneous greenhouse gases: Inhomogeneous forcing of homogeneous gas. *Journal of Geophysical Research: Atmospheres*, *121*, 2780–2789. https://doi.org/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. *Journal of Geophysical Research*, *105*(D11), 14,873–14,890. https://doi.org/10.1029/2000JD900091

Jacquinet-Husson, N., Armante, R., Scott, N. A., Chédin, A., Crépeau, L., Boutammine, C., et al. (2016). The 2015 edition of the GEISA spectroscopic database. *Journal of Molecular Spectroscopy*, *327*, 31–72. https://doi.org/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). A new paradigm for diagnosing contributions to model aerosol forcing error. *Geophysical Research Letters*, *44*, 12,004–12,012. https://doi.org/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. *Atmospheric Measurement Techniques*, *9*(2), 669–682. https://doi.org/10.5194/amt-9-669-2016

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. https://doi.org/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. *Journal of Geophysical Research: Atmospheres*, *118*, 1329–1342. https://doi.org/10.1002/jgrd.50141

Lin, P., Paynter, D., Ming, Y., & Ramaswamy, V. (2017). Changes of the tropical tropopause layer under global warming. *Journal of Climate*, *30*(4), 1245–1258. https://doi.org/10.1175/JCLI-D-16-0457.1

Manabe, S., & Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*, *24*(3), 241–259. https://doi.org/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., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Merlis, T. M. (2015). Direct weakening of tropical circulations from masked CO2 radiative forcing. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13167. https://doi.org/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. *Philosophical Transactions of the Royal Society A*, *370*(1968), 2520–2556. https://doi.org/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. *Journal of Geophysical Research*, *102*(D14), 16,663–16,682. https://doi.org/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. https://doi.org/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., et al. (2016). The spectroscopic foundation of radiative forcing of climate by carbon dioxide. *Geophysical Research Letters*, *43*, 5318–5325. https://doi.org/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. *Geophysical Research Letters*, *25*(14), 2715–2718. https://doi.org/10.1029/98GL01908

Myhre, G., & Stordal, F. (1997). Role of spatial and temporal variations in the computation of radiative forcing and GWP. *Journal of Geophysical Research*, *102*(D10), 11,181–11,200. https://doi.org/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. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *97*(3), 317–331. https://doi.org/10.1016/j.jqsrt.2005.04.015

Oreopoulos, L., Mlawer, E., Delamere, J., Shippert, T., Cole, J., Fomin, B., et al. (2012). The continual intercomparison of radiation codes: Results from phase I. *Journal of Geophysical Research*, *117*, D06118. https://doi.org/10.1029/2011JD016821

Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model intercomparison project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*, 3447–3460. https://doi.org/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. *Journal of Advances in Modeling Earth Systems*, *6*, 3074–3089. https://doi.org/10.1029/2019MS001621

Pincus, R., Mlawer, E. J., Oreopoulos, L., Ackerman, A. S., Baek, S., Brath, M., et al. (2015). Radiative flux and forcing parameterization error in aerosol-free clear skies. *Geophysical Research Letters*, *42*, 5485–5492. https://doi.org/10.1002/2015GL064291

Ptashnik, I. V., McPheat, R. A., Shine, K. P., Smith, K. M., & Williams, R. G. (2011). Water vapor self-continuum absorption in near-infrared windows derived from laboratory measurements. *Journal of Geophysical Research*, *116*, 488. https://doi.org/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. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *120*, 23–35. https://doi.org/10.1016/j.jqsrt.2013.02.016

Ramanathan, V., & Dickinson, R. E. (1979). The role of stratospheric ozone in the zonal and seasonal radiative energy balance of the earth-troposphere system. *Journal of the Atmospheric Sciences*, *36*(6), 1084–1104. https://doi.org/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., et al. (2013). The HITRAN2012 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *130*(0), 4–50. https://doi.org/10.1016/j.jqsrt.2013.07.002

Rotstayn, L. D., & Penner, J. E. (2001). Indirect aerosol forcing, quasi forcing, and climate response. *Journal of Climate*, *14*(13), 2960–2975. https://doi.org/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. https://doi.org/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). Adjustments in the forcing-feedback framework for understanding climate change. *Bulletin of the American Meteorological Society*, *96*, 217–228. https://doi.org/10.1175/BAMS-D-13-00167.1

Shine, K. P., & Myhre, G. (2020). The spectral nature of stratospheric temperature adjustment and its application to halocarbon radiative forcing. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001951. https://doi.org/10.1029/2019MS001951

Smith, C. J., Kramer, R. J., Myhre, G., Forster, P. M., Soden, B. J., Andrews, T., et al. (2018). Understanding rapid adjustments to diverse forcing agents. *Geophysical Research Letters*, *45*, 12,023–12,031. https://doi.org/10.1029/2018GL079826

Smith, C. J., Kramer, R. J., & Sima, A. (2020). The HadGEM3-GA7.1 radiative kernel: the importance of a well-resolved stratosphere. *Earth System Science Data*, *12*(3), 2157–2168. https://doi.org/10.5194/essd-12-2157-2020

Soden, B. J., Collins, W. D., & Feldman, D. R. (2018). Reducing uncertainties in climate models. *Science*, *361*(6400), 326–327. https://doi.org/10.1126/science.aau1864

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., et al. (2019). The Met office unified model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geoscientific Model Development*, *12*(5), 1909–1963. https://doi.org/10.5194/gmd-12-1909-2019

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The cloud feedback model intercomparison project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, *10*(1), 359–384. https://doi.org/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. *Quarterly Journal of the Royal Meteorological Society*, *138*(664), 561–575. https://doi.org/10.1002/qj.951

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018). The GFDL global atmosphere and land model AM4. 0/LM4. 0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*, 735–769. https://doi.org/10.1002/2017MS001209