AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1002/2017MS000932

On the sensitivity of anthropogenic aerosol forcing to model-internal variability and parameterizing a Twomey effect

S. Fiedler[1], B. Stevens[1], and T. Mauritsen[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany

Key Points:
- Ensembles of atmosphere-only experiments with MACv2-SP allow a sensitivity assessment of instantaneous and effective radiative forcing (ERF)
- Global mean all-sky ERFs with aerosol patterns of mid-1970s and today difficult to distinguish, when atmospheric variability considered
- A moderate pattern effect on ERF could occur in models with presumably strong aerosol-cloud interaction

Correspondence to:
S. Fiedler,
stephanie.fiedler@mpimet.mpg.de

Citation:
Fiedler, S., B. Stevens, and T. Mauritsen (2017), On the sensitivity of anthropogenic aerosol forcing to model-internal variability and parameterizing a Twomey effect, *J. Adv. Model. Earth Syst.*, *9*, 1325–1341, doi:10.1002/2017MS000932.

Received 2 FEB 2017
Accepted 8 MAY 2017
Accepted article online 16 MAY 2017
Published online 1 JUN 2017

Abstract Despite efforts to accurately quantify the effective radiative forcing (ERF) of anthropogenic aerosol, the historical evolution of ERF remains uncertain. As a further step toward a better understanding of ERF uncertainty, the present study systematically investigates the sensitivity of the shortwave ERF at the top of the atmosphere to model-internal variability and spatial distributions of the monthly mean radiative effects of anthropogenic aerosol. For this, ensembles are generated with the atmospheric model ECHAM6.3 that uses monthly prescribed optical properties and changes in cloud-droplet number concentrations designed to mimic that associated with the anthropogenic aerosol using the new parameterization MACv2-SP. The results foremost highlight the small change in our best estimate of the global averaged all-sky ERF associated with a substantially different pattern of anthropogenic aerosol radiative effects from the mid-1970s (–0.51 Wm^{-2}) and present day (–0.50 Wm^{-2}). Such a small change in ERF is difficult to detect when model-internal year-to-year variability (0.32 Wm^{-2} standard deviation) is considered. A stable estimate of all-sky ERF requires ensemble simulations, the size of which depends on the targeted precision, confidence level, and the magnitude of model-internal variability. A larger effect of the pattern of the anthropogenic aerosol radiative effects on the globally averaged all-sky ERF (15%) occurs with a strong Twomey effect through lowering the background aerosol optical depth in regions downstream of major pollution sources. It suggests that models with strong aerosol-cloud interactions could show a moderate difference in the global mean ERF associated with the mid-1970s to present-day change in the anthropogenic aerosol pattern.

1. Introduction

In the course of recent history, anthropogenic aerosol emission has been reduced in Europe while emissions in East Asia increased. Associated with this spatial shift of anthropogenic emissions between the mid-1970s and today (e.g., sulfur dioxide) [*Smith et al.*, 2011], a majority of the prevailing downwind transport of aerosol laden air occurs over the North Pacific ocean instead of the Eurasian continent that are dominated by different cloud regimes and surface albedos. The associated differences in the radiation budget in combination with the changed anthropogenic aerosol pattern suggest different effective radiative forcing associated with anthropogenic aerosol (ERF), i.e., instantaneous radiative forcing (RF) plus rapid atmospheric adjustments. Models of different complexity disagree on the effect of such a spatial shift of major pollution on the global mean ERF [*Murphy et al.*, 2009; *Carslaw et al.*, 2013; *Shindell et al.*, 2013; *Kühn et al.*, 2014; *Stevens*, 2015]. For instance, the mean of ten models from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) shows a decrease by -0.27 Wm^{-2} in the global mean ERF between 1980 and 2000, but the ERF has multimodel standard deviations of 0.36 Wm^{-2} and 0.47 Wm^{-2} in these years, respectively (based on Table 8 in *Shindell et al.* [2013]).

Although many potential reasons for this multimodel spread in ERF have been suggested, the major factors are still unclear. Possible reasons include differences in the aerosol parameterizations, atmospheric processes, surface conditions, and their interactions as well as natural variability [e.g., *Quaas et al.*, 2009; *Stevens and Feingold*, 2009; *Stier et al.*, 2013; *Deser et al.*, 2014; *Rosenfeld et al.*, 2014; *Wilcox et al.*, 2015]. Several studies point to meteorological conditions as a reason for model diversity in aerosol simulations [e.g., *Penner et al.*, 2006; *Liu et al.*, 2007; *Stier et al.*, 2013; *Myhre et al.*, 2013; *Shepherd*, 2014; *Bony et al.*, 2015; *Fiedler et al.*, 2015], that can impact ERF estimates. For instance, simulating the emission of natural aerosol with comprehensive aerosol-climate models is tightly linked with near-surface wind speed [e.g., *Blanchard and*

Woodcock, 1980; *Shao*, 2001], the simulation of which is a known challenge in atmospheric modeling [e.g., *Fiedler et al*., 2013; *Sandu et al*., 2013; *Largeron et al*., 2015]. Moreover, realistically simulating clouds and precipitation is a long-standing problem in models with parameterized moist convection [*Cesana and Chepfer*, 2012; *Jiang et al*., 2012; *Li et al*., 2012; *Nam et al*., 2012; *Dolinar et al*., 2014], such that the parametric uncertainty in cloud microphysics is exploited for adjusting climate models to specified targets [*Mauritsen et al*., 2012] and can contribute to the spread in radiative forcing estimates [*Lohmann and Ferrachat*, 2010; *Golaz et al*., 2013]. In particular, it remains difficult to constrain the contributions of aerosol-cloud interactions to ERF due to parametric and structural uncertainties of simulated clouds [*Ghan et al*., 2016].

Considering the number of uncertainty sources, a systematic analysis of their relative contribution is required for better understanding and ultimately reducing the model spread in ERF. It is not our ambition to quantify the whole range of process uncertainty contributing to the model spread in ERF, but rather to study those sources of uncertainty that remain if we pretend to know the effective radiation perturbations due to anthropogenic aerosol. In the present work, we aim to study the sensitivity of ERF estimates to model-internal variability, here defined as primarily meteorological variability as simulated by the model and the boundary data, and the changing spatial distribution of monthly mean anthropogenic aerosol radiative effects between the mid-1970s and the present. To do so, we produce ensembles of 10 year simulations. These are conducted with the atmospheric model component ECHAM6.3 of the Max-Planck-Institute for Meteorology's (MPI-M) Earth system model version 1.2 (MPI-ESM1.2) with monthly prescribed sea surface temperatures, as well as anthropogenic aerosol optical properties and an associated Twomey effect from MACv2-SP. MACv2-SP mimics the spatio-temporal distribution of monthly means of anthropogenic aerosol optical properties and effective changes in cloud droplet number concentration (N) associated with anthropogenic aerosol. The computational efficiency of MACv2-SP allows us to quickly produce 180 annual ERF estimates with annually repeating monthly aerosol patterns. This approach has been designed to study the ERF sensitivity to a prescribed pattern of monthly mean aerosol radiative effects. In a multimodel context, the usage of MACv2-SP can help to isolate how meteorological processes influence the model spread in ERF estimates for a given pattern of anthropogenic aerosol radiative effects, e.g., within the" Radiative Forcing Model Intercomparison Project" (RFMIP) [*Pincus et al*., 2016] and the BACCHUS project. Two overarching questions guide the analysis of the present work: (1) How can we estimate ERF in a naturally variable atmosphere, and (2) does the spatial distribution of aerosol radiative effects change the global mean ERF? The methods for addressing these questions are described in section 2. We show the results for these questions in sections 3.1 and 3.2, respectively, followed by a discussion of ERF scalability as a metric for future model inter-comparison studies in section 3.3. Conclusions and an outlook are given at the end (section 4).

2. Method

2.1. ECHAM6.3

The present work uses the atmospheric model component ECHAM6.3 of the Earth system model MPI-ESM1.2 of the MPI-M, participating in CMIP6. ECHAM6.3 is an updated model version of the sixth generation general circulation model ECHAM6 that has been developed at MPI-M [*Stevens et al*., 2013]. Boundary conditions are specified by external data sets, e.g., for surface properties, trace gas concentrations, and natural aerosol. Natural aerosol is prescribed as a monthly mean climatology in the shortwave radiation calculation and thereby influences aerosol-radiation interaction (RF_{ari}).

2.2. MACv2-SP

The newly developed parameterization MACv2-SP [*Stevens et al*., 2017] is a tool for prescribing the spatio-temporal distributions of optical properties for anthropogenic aerosol and an associated Twomey effect to estimate radiative forcing. It is deliberately simple for the ease of use in different models to facilitate experimentation and test hypotheses. The aim of MACv2-SP is to induce radiative effects associated with patterns of anthropogenic aerosol optical properties. By prescribing patterns of anthropogenic aerosol radiative effects in that way, MACv2-SP aids in separating the problem of pattern influences on the climate system from the problem of aerosol-process contributions to ERF. In other words, if we knew the anthropogenic aerosol radiative effects, MACv2-SP induces them in our model and thereby allows us to study the impact of remaining sources of uncertainty on ERF. For the present study, we have tuned MACv2-SP to the historical evolution of anthropogenic aerosol optical properties using observational constraints. Other studies on

the climate response to aerosol could use MACv2-SP as a tool for prescribing different anthropogenic aerosol radiative effects, e.g., a setup that mimics those that have been produced by an aerosol-climate model.

2.2.1. Concept and Design

MACv2-SP determines the anthropogenic aerosol optical depth (AOD) τ_a, the asymmetry parameter g, the single scattering albedo ω_0 of anthropogenic fine-mode aerosol, and a prefactor η_N to be applied to the cloud droplet number concentration (section 2.2.2) as function of geographical location, time and wavelength. Anthropogenic aerosol is assumed to consist of particles smaller than 0.5 μm primarily from biomass burning and industrial pollution. We therefore neglect a direct effect on the thermal infrared radiation transfer and parameterize an anthropogenic aerosol radiative effect in the shortwave radiation transfer calculation of ECHAM6.3.

The model accounts for RF_{ari} and aerosol-cloud interaction (RF_{aci}) in the shortwave radiation spectrum as well as adjustments allowing us to diagnose the effective radiative forcing (ERF). Shortwave radiation (SW) captures wavelengths of 0.2–12.2 μm in the model. Although we do not prescribe a direct anthropogenic aerosol effect on radiation with longer wavelength, longwave radiative adjustments are accounted for in the model. For reasons outlined in section 2.2.2, we parameterize a Twomey effect in the radiation calculation, i.e., the model treats no adjustments to RF_{aci} in form of an aerosol perturbation in the parameterization of cloud microphysics. Adjustments to RF_{aci} are in our model radiatively induced through the increased cloud albedo.

The optical properties of anthropogenic aerosol are designed to capture the monthly mean pattern of the anthropogenic aerosol radiative effect and are constrained by the MPI-M's Aerosol Climatology (MACv2.0), an updated version of MACv1.0 [*Kinne et al*., 2013] with more observations dating back to 2000. MACv2.0 merges monthly statistics of quality-controlled AERONET station observations [*Holben et al*., 1998] with the global distribution of the AeroCom model ensemble median [e.g., *Penner et al*., 2002; *Kinne et al*., 2006; *Shindell et al*., 2013]. MACv2-SP approximates the monthly distribution of τ_a from MACv2.0 by mathematical functions that generate three-dimensional plumes. The plumes are constructed at nine geographical positions that coincide with regional maxima in τ_a of MACv2.0. Figure 1 indicates the position of these maxima, namely five industrial regions and four regions with seasonally dominant biomass burning. *Stevens et al*. [2017] give a detailed description and validation of MACv2-SP, and outlines the standard settings of MACv2-SP. The present study uses additional setups with the following modifications.

2.2.2. Parameterization of Twomey Effect

We parameterize a Twomey effect, thus an increase in the cloud droplet number concentration (N) and associated reduction in droplet size under constant liquid water path [*Twomey*, 1974], designed to produce a monthly mean radiative effect from the anthropogenic aerosol perturbation. A prolonged lifetime of warm clouds has been proposed due to delayed rain formation associated with smaller cloud droplets for a constant liquid water path [*Albrecht*, 1989], but is difficult to underpin with observational evidence. Moreover, *Seifert et al*. [2015] identified a reduction in cloud cover associated with more cloud droplets. They found evaporation of cloud remnants due to the lower ambient relative humidity in response to deepening of clouds that have organized, i.e., a negative cloud-lifetime effect that comprehensive models with parameterized convection fail to represent. Resolving this controversy of the occurrence of an extended cloud

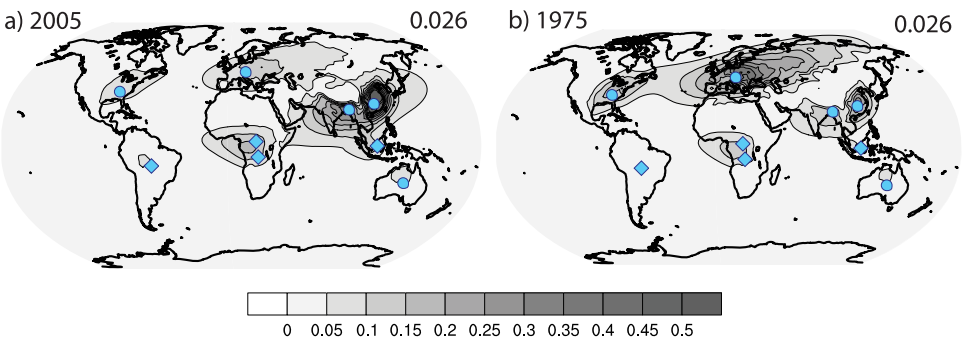

Figure 1. Global distribution of anthropogenic AOD in MACv2-SP. Shown are (shaded) the annual mean anthropogenic fine-mode AOD at 550nm for (a) 2005 and (b) mid-1970s. The geographical position of the centers of the aerosol plumes for MACv2-SP are indicated for regions of industrial pollution (circles) and biomass burning (rectangles). Right corners give globally averaged anthropogenic AODs.

lifetime due to anthropogenic aerosol [e.g., *Albrecht*, 1989; *Seifert et al*., 2015] is subject of ongoing research. Given this poor understanding of even the sign of other anthropogenic aerosol effects on clouds than a Twomey effect, uncertainties in cloud microphysics parameterizations [e.g., *Ghan et al*., 2011], and evidence that the Twomey effect might dominate [*McGibbon and Bretherton*, 2017], we limit our consideration to effects that can be mimicked by prescribed changes in the monthly mean cloud radiative properties using a prefactor η_N. For conciseness, we describe this effective change of cloud optical properties as a Twomey effect. However, the factor η_N could be tuned to achieve any desired net contribution from ERF_{aci} to ERF. For instance, if a potential contribution from an extended cloud lifetime would have a similar pattern like the anthropogenic aerosol radiative effect and would be expected to generate a more negative ERF, such an effect could be represented as an additional increase in N by MACv2-SP. This capability will be used to study the sensitivity of ERF to the strength of ERF_{aci}.

Twomey effects are induced by parameterizing a prefactor η_N, which is in ECHAM6.3 multiplied with the cloud droplet number concentration N. Along with the cloud liquid water path and a prescribed cloud droplet number concentration, η_N is used to derive the cloud optical properties for the purpose of the shortwave radiation transfer calculation. Monthly mean satellite data suggest that η_N can be parameterized as function of τ_a and a fine-mode background AOD τ_{bg} [*Stevens et al*., 2017]:

$$\eta_N = 1 + \frac{dN}{N} = \frac{\ln\left[1000\ \left(\tau_a + \tau_{bg}\right) + 1\right]}{\ln\left[1000\ \tau_{bg} + 1\right]}. \tag{1}$$

This is the original parameterization implemented in MACv2-SP. Although the Twomey effect [*Twomey*, 1974] is qualitatively understood, it remains difficult to constrain the magnitude due to the shortage of representative observations, the regime-dependent behavior of cloud adjustments, and different estimates of how much variability in the covariability of clouds and aerosol influences the monthly mean anthropogenic aerosol effect [e.g., *Stevens and Feingold*, 2009; *Rosenfeld et al*., 2014] that we attempt to parameterize. These difficulties have led to different proposals of functions that relate changes in N with aerosol perturbations [*Quaas et al*., 2006; *Andreae*, 2009; *Carslaw et al*., 2013; *Stevens et al*., 2017]. To test the sensitivity of ERF to the functional form of the parameterization, an alternative relationship is additionally defined in the present study. Based on the functional fit from temporally higher-resolved satellite data [*Quaas et al*., 2006] compared to *Stevens et al*. [2017], η_N is written as:

$$\eta_N = 1 + \frac{dN}{N} = \frac{\exp\left[5 + 0.3\ \ln\left(\tau_a + \tau_{bg}\right)\right]}{\exp\left[5 + 0.3\ \ln\ \tau_{bg}\right]}. \tag{2}$$

Figure 2 shows typical values of η_N as function of τ_{bg} and τ_a calculated with the two parameterizations. In both cases, η_N increases with larger τ_a and decreases with larger τ_{bg}. At any fixed τ_{bg} above 0.02 (standard, Figure 2), η_N increases more quickly with τ_a in equation (2), implying a stronger brightening of liquid-water clouds associated with aerosol perturbations in naturally aerosol-laden air. At small τ_{bg} and τ_a (low background and anthropogenic pristine, Figure 2), the behavior is opposite, namely a stronger increase of η_N with τ_a in equation (1).

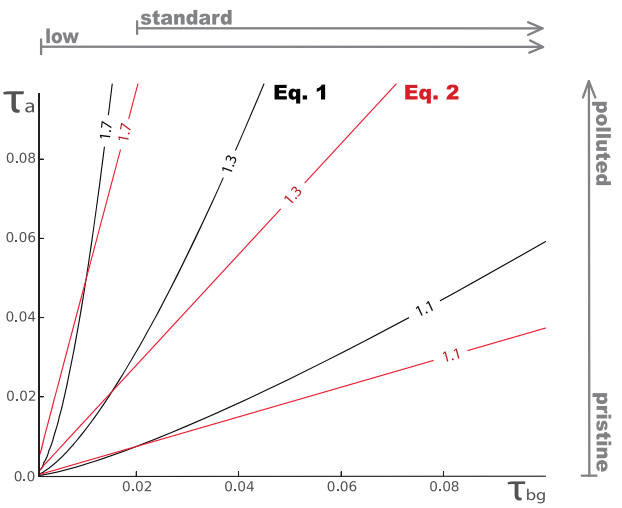

Figure 2. η_N as function of τ_a and τ_{bg}. Shown are isolines of constant η_N calculated with (black) equation (1) and (red) equation (2).

Both parameterization approaches use an idealized τ_{bg} climatology that is independent of the natural background used for representing RF_{ari}. This choice allows us to decouple the specification of the anthropogenic perturbation from a host model's natural background aerosol and also makes it easier to modify the strength of RF_{aci} without directly changing RF_{ari}. For η_N, τ_{bg} is a function of longitude ϕ, latitude λ and time of year t:

$$\tau_{bg}(\phi,\lambda,t)=\tau_{pl}(\phi,\lambda,t)+\tau_{gl}. \quad (3)$$

The plume-approximated background τ_{pl} is derived by adopting the spatial approximation of τ_a. Regions far from plume centers have very low values of τ_{pl}, motivating the introduction of a superimposed global background τ_{gl}. Evaluation against the natural fine-mode background of MACv2.0 over oceans suggests that using $\tau_{gl}=0.02$ gives a reasonable approximation. The standard settings of τ_{bg} lie therefore in the parameter space where equation (1) has typically smaller η_N than equation (2) for the same τ_{bg} and τ_a (standard, Figure 2). To investigate the sensitivity of ERF to the assumed background, we performed experiments that also include the parameter space with lowered τ_{bg} by setting $\tau_{gl}=0.002$ (low background, Figure 2). The impact of the parameterization choice and the aerosol patterns will be investigated with the following ensemble simulations.

2.3. Experiment Strategy

All model experiments are run with prescribed sea-surface properties for 2000–2010 inclusive, consistent with the observational period of the underlying observation climatology MACv2.0. Monthly means of sea surface temperatures (SSTs) and sea ice herein change from year to year, which we refer to as AMIP style. The model setups of this study are summarized in Table 1. Natural variability motivates the usage of ensemble simulations that are produced as follows. A simulation with observed SSTs and varying τ_a from MACv2-SP starting in 1976 produced initialization fields at the end of 1999. These are used to initialize the first ensemble member with an annual cycle of τ_a that does not change from year to year. Initial conditions for additional members are started from model states taken from different days at the beginning of January 2000. The first year of each member is considered as additional spin-up period after perturbing the model, so that all results are shown for 2001–2010. The control present-day ensemble with the standard setup of MACv2-SP (SP) uses the constant annual cycle of τ_a for 2005 and is composed of three members. Likewise, the sensitivity ensembles have three members each. These follow the setup of SP, but use modifications in MACv2-SP, introduced next.

For assessing the relative importance of the ERF_{ari} for ERF, the Twomey effect is switched off in setup NTE by setting η_N to one. Uncertainty in ERF associated with the parameterization form of the Twomey effect is quantified by the Twomey effect setup that uses equation (2) instead of equation (1) (QTE). QTE has a qualitatively similar distribution of η_N like SP, but some quantitative differences close to plume centers (Figures 3a and 3b), e.g., a larger maximum in East Asia.

The effect of the spatial pattern of anthropogenic aerosol is studied by replacing τ_a by values representative for the mid-1970s in the anthropogenic pattern setup (PAT). In East Asia, η_N is reduced with the global distribution of τ_a of the mid-1970s, but Eurasia, and the North Atlantic have larger η_N than SP (Figures 3a and 3c), consistent with differences in the spatial distributions of τ_a (compare Figure 1). Such a substantial spatial shift in the pattern of τ_a provides a test of the relative importance of the pattern versus the same global aerosol optical depth for the global mean ERF.

In addition to testing the importance of the aerosol pattern, the role of the natural aerosol burden in influencing the strength of the Twomey effect is investigated with the low background setup (LBG). Here, the global background τ_{gl}, used to calculate η_N with equations (1) and (3), is lowered by a factor of ten. This has no effect on the natural aerosol climatology for RF_{ari}, which remains identical to SP. Reducing τ_{gl} has a strong impact on the magnitude of η_N away from major anthropogenic sources (Figures 3a and 3d). Close to aerosol sources, however, τ_{gl} has a relatively small effect on η_N owing to τ_{pl} being an order of magnitude

Table 1. Experiment Setups[a]

Name	τ_a	η_N	τ_{gl}	Description
SP	Present-day	Equation (1)	0.02	Standard setup
NTE		1	none	No Twomey effect
PAT	mid-1970s			Different anthropogenic pattern
LBG			0.002	Lowered aerosol background in equation (1)
PAT-LBG	mid-1970s		0.002	Combination of PAT and LBG
QTE		Equation (2)		Alternative parameterization equation (2)
REF	none	1	none	Reference: natural aerosol only

[a]Listed are settings from SP and relative changes in the sensitivity experiments.

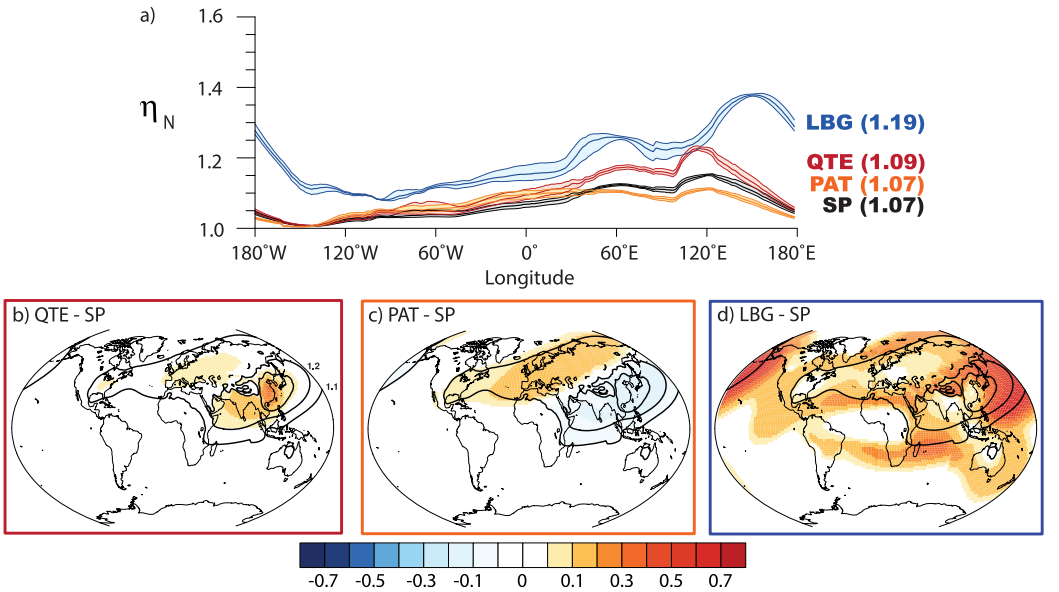

Figure 3. Differences in annual mean η_N in ensembles. Shown are (a) meridionally averaged latitude-weighted η_N of the annual (thick lines) mean and (shaded) range of color-coded setups, and (b–d) spatial distribution of (black lines) annual mean η_N from SP in steps of 0.1 and (shaded) difference to SP for (b) QTE, (c) PAT, and (d) LBG.

larger than τ_{gl}. We thereby induce stronger RF_{aci} in rather remote areas that is in contrast to the other setups, where differences always occur in plume centers (Figures 3b and 3c). The LBG ensemble simulations serve to explore the possibility that RF_{aci} are (1) much too weak with MACv2-SP in ECHAM6.3, or (2) much too strong in other modeling studies. For a climatological mean, the low background aerosol would be a rather extreme assumption, such that LBG should rather be interpreted as a hypothetical case when episodes of very low natural-aerosol burden would be sampled, or when RF_{aci} is presumably very strong. LBG will be compared to PAT to assess the relative importance of the aerosol background burden and anthropogenic aerosol patterns for the magnitude of ERF for present-day conditions. Additionally, we simulate an ensemble with combined settings of PAT and LBG (PAT-LBG) that will be compared to LBG to investigate what the anthropogenic pattern change between the mid-1970s and today would be in the case of much stronger RF_{aci}. This allows us to estimate the magnitude of the anthropogenic pattern effect on ERF for a model with strong RF_{aci}.

ERFs are calculated from SP and sensitivity ensembles relative to a preindustrial reference setup (REF) that uses only natural AOD for 1850 from MACv2.0, but is otherwise set up like SP. REF has six ensemble members so as to increase the number of ERF estimates for all the experimental setups, i.e., 6x3 time series of monthly values of ERF estimates over 10 years per setup. Spatial distributions of ERF estimates and summary statistics from the ensembles consider annual means from all years, thus 180 annual ERF estimates each. RF is calculated in the simulations with anthropogenic aerosol by calling the radiation module twice, i.e., once with and once without the prescribed anthropogenic aerosol radiative effects each time it is called through the course of a simulation. This gives three 10 year time series per ensemble, thus 30 values for statistics of annual mean RFs each. Although this data volume is smaller than for ERF, estimating RF is much less influenced by model-internal variability, so that the number of RF estimates is sufficient for detecting significant signals at a confidence level of 1% (not shown). Significance tests are performed with *t*-tests. The spread in ERF estimates associated with model-internal variability is illustrated as the standard deviation of yearly estimates.

3. Results

3.1. Estimating Present-Day ERF

3.1.1. Global Mean

The standard setup SP yields an ensemble-averaged annual mean ERF in clear-sky conditions of -0.67 ± 0.0004 Wm^{-2} (mean $\pm$ standard error). The year-to-year variability in the clear-sky ERF estimate is

Table 2. Globally Averaged ERF in [Wm^{-2}][a]

	Dec–Feb	Mar–May	Jun–Aug	Sep–Nov	Year
SP	−0.33 (–0.51)	−0.55 (–0.66)	−0.63 (–0.83)	−0.48 (–0.70)	−0.50 (–0.67)
NTE	−0.04 (–0.50)	−0.31 (–0.60)	−0.15 (–0.80)	−0.43 (–0.69)	−0.23 (–0.65)
LBG	−0.77 (–0.53)	−0.97 (–0.69)	−1.45 (–0.86)	−0.95 (–0.70)	−1.03 (–0.69)
PAT	−0.32 (–0.41)	−0.48 (–0.71)	−0.72 (–0.92)	−0.50 (–0.62)	−0.51 (–0.67)
PAT-LBG	−0.45 (–0.43)	−0.86 (–0.70)	−1.28 (–0.95)	−0.89 (–0.64)	−0.87 (–0.68)
QTE	−0.28 (–0.55)	−0.52 (–0.70)	−0.75 (–0.83)	−0.76 (–0.70)	−0.58 (–0.69)

[a]Shown are global ensemble averages of ERF in all- and (brackets) clear-sky for the seasons and the entire year.

0.07 Wm^{-2} (standard deviation). Clear-sky ERF variability could be associated with differences in the state of the atmosphere and surface, e.g., temperature. The clear-sky RF is -0.65 Wm^{-2} and changes little from one year to the next (0.002 Wm^{-2} standard deviation) such that clear-sky RF can be diagnosed with high precision after few simulation years.

In all-sky, the annual mean ERF is -0.50 ± 0.002 Wm^{-2}, that is less negative than the estimate of -1.17 Wm^{-2} (–1.44 Wm^{-2} to -0.71 Wm^{-2}) for 1850–2000 from models participating in ACCMIP [*Shindell et al*., 2013] and -0.9 Wm^{-2} (–1.9 Wm^{-2} to -0.1 Wm^{-2}) for 1750–2011 from models in the fifth assessment report of the IPCC [*Myhre et al*., 2013]. Most estimates from these models are more negative than our estimate, but are believed to represent too strong RF_{aci} and associated cloud adjustments for reasons discussed by *Stevens* [2015].

The all-sky ERF is smaller than our estimate for clear sky. This reduction, due to clouds, is found in all seasons (Table 2). Although a Twomey effect makes clouds brighter, they also mask clear-sky ERF, which is in our simulation more important (section 3.2.3). Moreover, rapid atmospheric adjustments cause a slight net reduction of all-sky ERF (section 3.1.4), compared to the all-sky RF of -0.60 Wm^{-2}. The year-to-year standard deviation of all-sky RF is relatively small (0.004 Wm^{-2}) compared to the variability of all-sky ERF (0.32 Wm^{-2}) so that the former can be more quickly diagnosed with the same precision. The year-to-year variability of all-sky ERF and the impact on the precision is further investigated in the following.

Year-to-year variability in the reflected SW TOA irradiance is independent of whether the simulations are performed with or without an anthropogenic aerosol perturbation τ_a. This is established by comparing the year-to-year standard deviation in the all-sky SW irradiance (Figure 4a) for simulations without (σ=0.31 Wm^{-2}) and with τ_a (σ=0.32 Wm^{-2}). The similarity allows us to use the variability in the observed SW TOA radiation budget for validating the model next. The year-to-year variability in our model is larger than what is retrieved from the energy balanced and filled (EBAF) TOA SW radiation fluxes for all-sky from the Clouds and Earth's Radiant Energy System (CERES) [*Wielicki et al*., 1996; *Loeb et al*., 2009]. We estimate a year-to-year variability in the SW TOA irradiance of 0.18 Wm^{-2} for 2001–2010 from CERES. The reason for the difference is not well understood. One could think that the difference arises because the prescribed SSTs imply a nonclosed surface energy budget that may distort the magnitude of variability at TOA, but using a hundred-member ensemble of coupled atmosphere-ocean experiments with MPI-ESM1.1 suggests a year-to-year standard deviation of 0.76 Wm^{-2} for the overlapping time period 2001–2005, a value that is roughly twice as large than all-sky σ from our simulations using fixed SSTs. Another reason could be the few annual estimates from CERES (10) that must not span the entire possible range of year-to-year variability that a model ensemble can simulate.

The variable and small globally, annually and ensemble averaged all-sky ERF (ERF_{all}) lead to the question of how long a time series needs to be for estimating ERF_{all} with a certain precision. Addressing this question is simplified by the normally distributed and weakly autocorrelated time series of annually and globally averaged ERF_{all}. In order to systematically account for various year-to-year changes, 18,000 alternative time series are generated by shuffling all annual and global mean ERF_{all} from SP and averaging over the values that have occurred. The more years for averaging are used, the smaller the width of the fitted Gaussian distributions to ERF_{all} frequency histograms becomes, as one would expect. Using the upper percentiles of the distributions after time t allows us to calculate the precision of the mean all-sky ERF estimate (Figure 4b), assuming the long-term mean ERF_{all} is the target value. The nonlinear improvement of precision approximately behaves similar to the standard error, but the latter predicts a quicker improvement in the precision of the all-sky ERF estimate. Typical values using the 99% percentile as measure are 6 years for a precision of

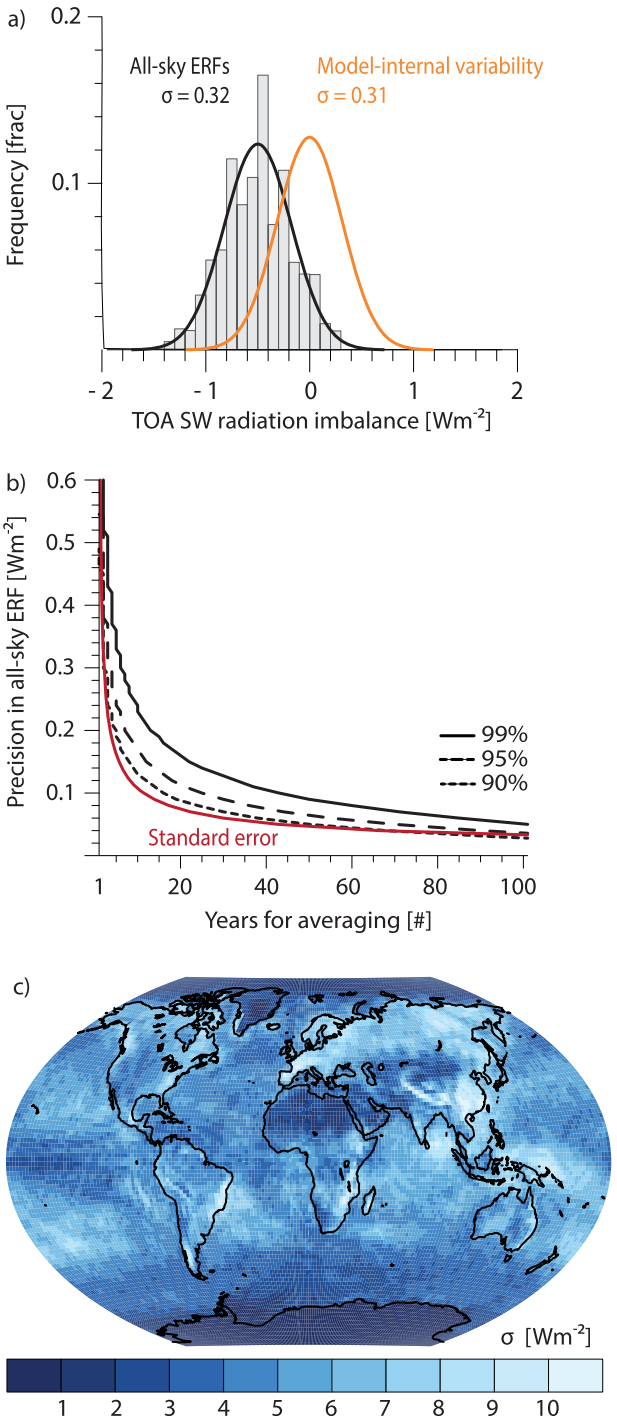

Figure 4. Variability and precision of global mean all-sky ERF from SP. (a) Gaussian distributions of (orange) natural variations in SW TOA radiation budget, and (black) one-year means of SW TOA ERF_{all} fitted to (gray bars) original frequency histogram from SP. (b) Years for certain precision in SW TOA ERF_{all}, using percentiles of distributions from shuffled time series (black, refer to section 3.1.1) and standard error (red). (c) Standard deviation in SW TOA ERF_{all}.

0.3 Wm^{-2} and 36 years for a three times higher precision of 0.1 Wm^{-2}, using the same monthly aerosol climatology without year-to-year changes. Note that the pairs of years and precision depend on the chosen percentile for evaluating the deviation from the long-term mean and the magnitude of model-internal variability that is likely model dependent.

Irrespectively of the exact number of years for a certain precision, the spread in one year ERF_{all} estimates suggests that their precision can be poor. For instance, a 1 year estimate can deviate from the long-term mean by as much as 0.5 Wm^{-2}. Such one year estimates have been used in past model inter-comparison studies to study transient aerosol changes [e.g., *Shindell et al.*, 2013], such that model-internal variability likely explains a part of the multimodel spread in ERF_{all}, e.g., about half if σ from our simulations was representative.

3.1.2. Spatial Pattern

The ensemble-averaged ERF_{clr} of SP has regional maximum contributions of around -5 Wm^{-2} (Figure 5a), resembles the pattern of τ_a (Figure 1a), and is in many areas statistically significant (Figure 5a), even when a conservative confidence level of 1% is applied. The spatial distribution of ERF_{all} contributions, shown in Figure 5b, is more noisy than in clear sky. Although ERF_{all} contributions of similar magnitudes occur, e.g., over Southeast Asia, they are in overall fewer regions statistically significant (Figure 5b). This can be attributed to the larger natural atmospheric variability under all-sky conditions than in clear-sky (compare Figure 4c) due to the variability in clouds.

Natural variability impacts the chances to detect regional contributions to the global mean ERF under present-day conditions. Assessing the number of annual estimates that depart from zero by at least one σ indicates a distinction of 1 year mean ERF_{clr} contributions in areas with large τ_a only. We here choose the regional standard deviation as a threshold to represent natural variability, that would make the detectability of regional contributions to ERF also challenging in nature. In all-sky conditions, it is more difficult to estimate regional contributions to ERF_{all} due to the larger variability, e.g., regionally σ mostly exceeds 1 Wm^{-2} (Figure 4c). The best chances for a distinction are typically found in areas offshore

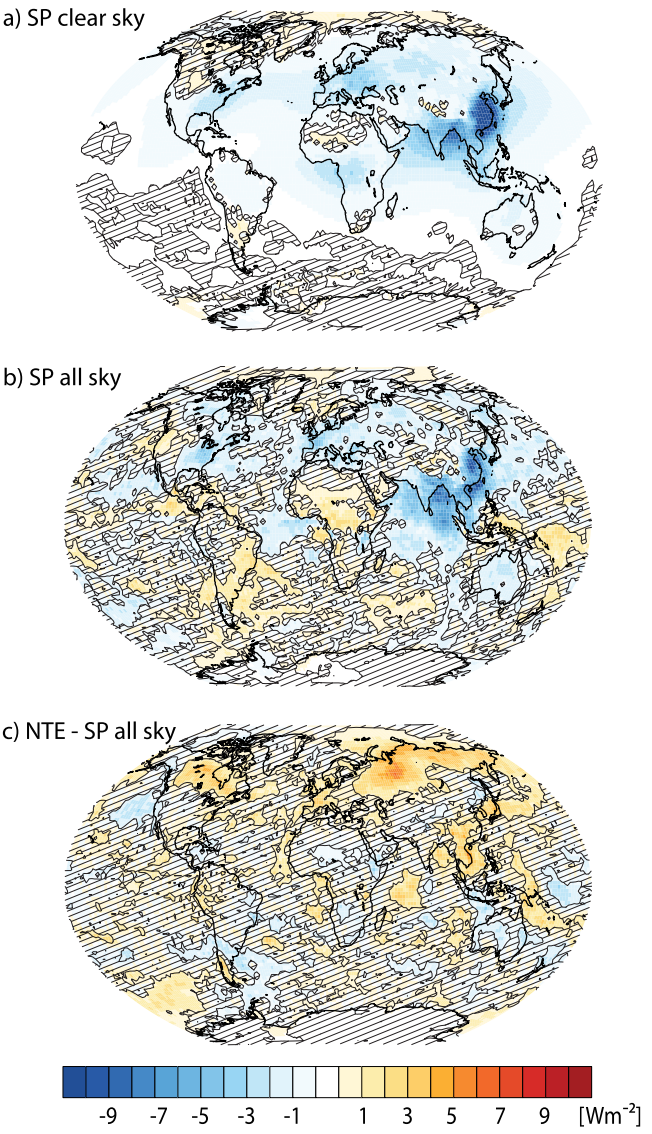

Figure 5. ERF from SP. Shown is the annually and ensemble averaged ERFs (SW, TOA) for (a) clear and (b) all sky from SP, as well as (c) the difference in all sky without Twomey effect (NTE) relative to SP. ERFs not significantly different from (a, b) zero or (c) SP are masked out by hatching, adopting a confidence level of 10%.

of some major pollution sources, such as Europe, East Asia, India, and Central Africa. Repeating the same statistical shuffling and averaging as for global mean ERFs regionally and requiring a distinction of the upper percentile of averaged ERF_{all} from zero by one σ, suggests necessary periods for averaging on the order of decades, even though the regional contributions to ERF_{all} is here comparably large (Figure 5b). Assessments of regional contributions to ERF_{all} with different patterns of anthropogenic aerosol radiative effects (section 3.2) use therefore ensemble averages and focus on areas where signals are statistically significant.

3.1.3. Role of Twomey Effect

Ensemble-averaged ERF_{all} in NTE is with -0.23 ± 0.002 Wm^{-2} roughly half as strong compared to the standard estimate wherein aerosols are allowed to modify the cloud droplet number in the calculation of the cloud optical properties. The partitioning of ERF_{ari} and ERF_{aci} is similar to the relative contributions stated in the fifth assessment report of the IPCC [*Myhre et al.*, 2013]. Variability affects the ability to distinguish ERF_{all} samples of NTE and SP. To illustrate the effect of model-internal variability, ten subsamples of 18 annually and globally averaged ERF_{all} with the same boundary conditions are compared in Figure 6a. These indicate that only six out of ten subsamples are significantly different, which underlines again that several decades for averaging are needed for yielding a stable ERF estimate.

Regionally, accounting for Twomey effects changes ERF_{all}, not only in terms of the magnitude but even in sign. For instance, northern regions of Asia show significant positive ERF_{all} without a Twomey effect (Figure 5c), in contrast to mostly negative and partly no ERF_{all} with RF_{aci} (Figure 5b). These differences are an indicator for the relevance of RF_{aci} and atmospheric adjustments, e.g., radiative heating perturbations that change the atmospheric stability, influencing circulation and clouds, and a near-surface temperature change, affecting snow coverage and thereby land-surface albedo. The net contribution of such rapid adjustments to ERF is more systematically analyzed next.

3.1.4. Role of Rapid Adjustments

Decomposing ERF into a net contribution of rapid adjustments and RF suggests that the global and annual mean ERF is mainly determined by RF in our model. The contribution of rapid adjustments to ERF in models with more comprehensive aerosol parameterizations might be different, since they account for cloud microphysical adjustments (section 2.2.2). Seasonal and annual means in our net contributions of rapid adjustments to ERF_{all} range from 0.06 to 0.13 Wm^{-2} (Table 3) corresponding to a net weakening by -10% to -35% of ERF_{all} contributions to the global mean.

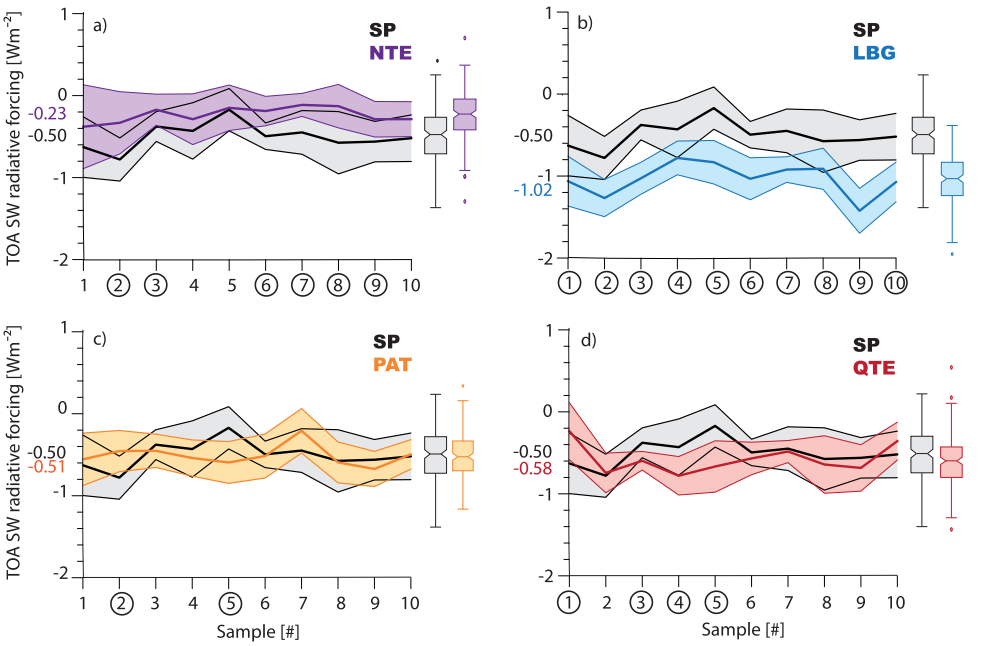

Figure 6. Ensemble statistics of all-sky ERF. Shown are (thick line) means and (shaded) standard deviation of samples of annually and globally averaged ERF_{all} (SW, TOA) that is associated with anthropogenic aerosol drawn from SP and (a) NTE, (b) LBG, (c) PAT, and (d) QTE. Each sample contains 18 annual means of the same boundary conditions, such that variations reflect model-internal variability. Circles indicate significantly different ERF_{all} statistics, measured with a *t* test and a confidence level of 1%. Box-and-whisker on the right show statistics of the entire ensemble. Numbers at the ordinate are ensemble means.

Zonally averaged RFs are largest in the NH (Figures 7a and 7d), relatively smooth with little differences amongst ensemble members, and in most zones significant, even at a conservative confidence level of 1%. RF_{clr} is the same as in SP for all sensitivity ensembles (Figure 7a), except PAT with a more northern maximum of τ_a (compare Figure 1). Incorporating clouds causes RF_{all} to differ more strongly amongst the ensemble means (Figure 7d), due to the differently prescribed Twomey effects.

Zonal mean ERF contributions are comparably noisy (Figures 7b and 7e) and in less zones significant compared to RF, owing to rapid adjustments. Comparing the mean of the ensembles of each sensitivity setup shows that the ensemble averaged ERF_{clr} contributions differ in the NH extra-tropics, where rapid adjustments in clear sky typically reach up to -0.5 Wm^{-2} (Figure 7c) and are in some zones significantly different. In all-sky, zonal net contributions of rapid adjustments are typically larger than in clear sky, e.g., up to 1 Wm^{-2} in the NH extra-tropics (Figure 7f). The net contributions of rapid adjustments to ERF are in their majority small compared to the magnitude of RF and ERF. We therefore only assess ERF in the following.

3.2. Impact of Spatial Distributions

3.2.1. Reducing Natural Background

A low natural aerosol burden efficiently enhances ERF. A substantial reduction of τ_{gl} results in a widespread increase in η_N by up to 0.5 toward the edges of pollution plumes (LBG, Figure 3). The LBG ensemble produces a globally averaged ERF_{all} of -1.03 ± 0.002 Wm^{-2}, more than twice as strong compared to SP. Drawing ten samples indicates that model-internal variability has little impact on the detectability of the difference, since all samples are significantly different, even at a strict confidence level of 1% (Figure 6b). ERF_{clr} from

Table 3. Global Averages From SP[a]

	Dec–Feb	Mar–May	Jun–Aug	Sep–Nov	Year
RF (Wm^{-2})	−0.45 (–0.50)	−0.60 (–0.62)	−0.76 (–0.81)	−0.59 (–0.69)	−0.60 (–0.66)
Rapid adjustment (Wm^{-2})	0.12 (–0.01)	0.06 (–0.03)	0.13 (–0.02)	0.11 (–0.01)	0.10 (–0.02)
Rapid adjustment/ERF (%)	−36 (2)	−11 (5)	−21 (2)	−23 (1)	−20 (3)

[a]Shown are SP ensemble averages of RF and net contributions from rapid adjustments in all- and (brackets) clear-sky for the seasons and the entire year. Should be compared to ERF from SP in Table 2.

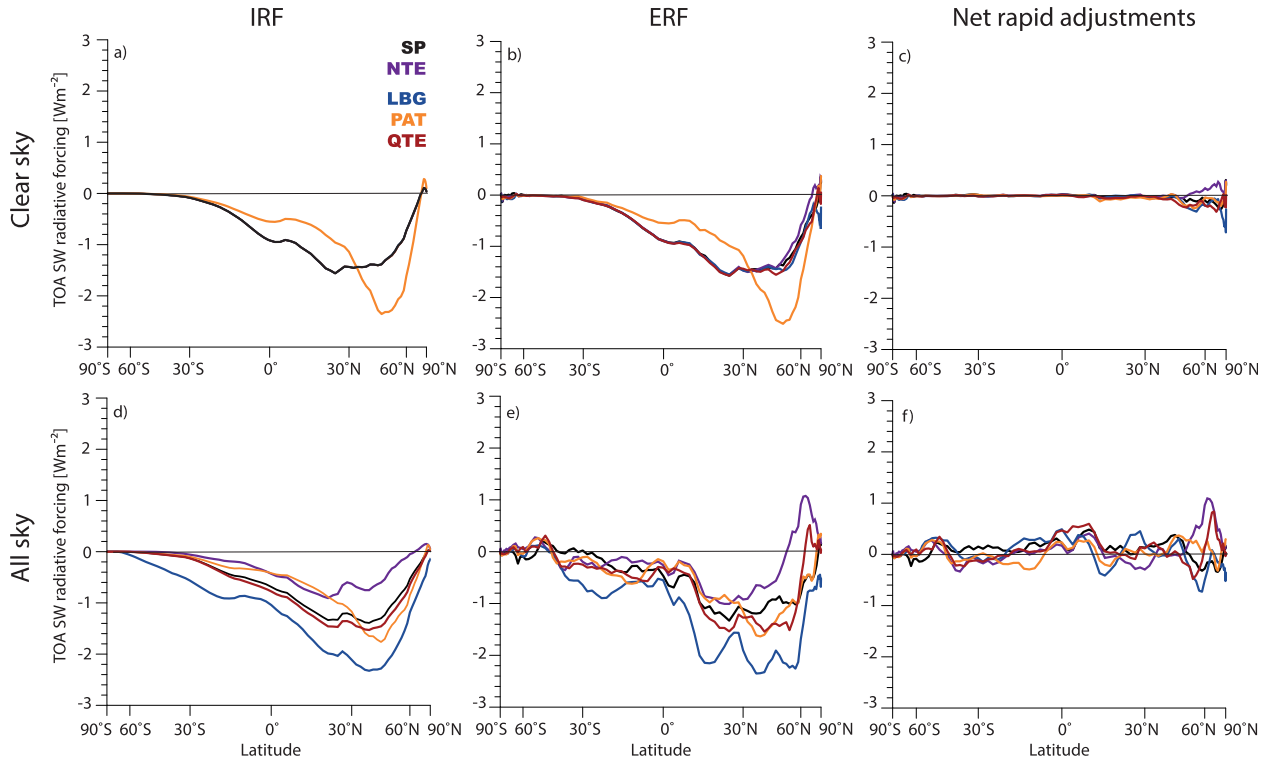

Figure 7. Net contributions of RF and rapid adjustments to ERF. Shown are zonal means of annually and zonally averaged (left to right) RF, ERF, and net contribution from rapid adjustments to SW TOA ERF in (top) clear- and (bottom) all-sky as ensemble means from different setups that are color-coded. Note cosine weighted x axis to account for meridional extent of zones.

LBG (-0.69 ± 0.0004 Wm^{-2}) and SP are similar, since the natural background is only changed in the parameterization of η_N. In addition to the relevance for the global mean, the natural background aerosol in η_N regionally emphasizes contributions to the global mean ERF_{all}, primarily away from source regions over the North Pacific (Figure 8a). The strongly negative contributions to ERF_{all} in LBG over the Oceans is consistent with large η_N downstream of Asian pollution sources, which causes brightening and a strong radiative impact over dark ocean surfaces.

The large impact of the natural background on ERF_{all} is consistent with the understanding of the importance of accurately estimating natural aerosol [e.g., *Carslaw et al*., 2013] and the spatial extent of transport of anthropogenic aerosol. Modeling natural aerosol, however, remains challenging. On the one hand, the global occurrence, mass and size distribution of natural aerosol are poorly constrained owing to the shortage of quality-controlled long-term observations from uninhabited regions. On the other hand, simulating the life cycle of natural aerosol is fraught with uncertainties both in association with aerosol and meteorological processes. For instance, winds differ amongst models [e.g., *Fiedler et al*., 2015; *Huneeus et al*., 2016], that affect natural aerosol emission.

3.2.2. Shifting NH τ_a

Substantially changing the pattern of τ_a (PAT) has a surprisingly small effect on the global mean ERF. The PAT setup uses a spatial distribution of aerosol representative for the mid-1970s, when major pollution occurred over Europe and North America instead of Asia (Figure 1). The global mean ERF_{all} from PAT is difficult to distinguish from SP. Although the ensemble average is -0.51 ± 0.002 Wm^{-2}, slightly more negative in PAT (Table 2), the difference of most of individual subsamples are not significant (Figure 6c). The only two subsamples, which are significantly different in magnitude, show an increase and decrease in ERF_{all} due to the shift in τ_a. This illustrates that model-internal variability can be a possible explanation for uncertainty in changes of the global mean ERF_{all} associated with a spatial shift of τ_a as seen in some past studies [e.g., *Shindell et al*., 2013]. Despite the small differences in the global mean, regional contributions to ERF_{all} are strongly modified by shifting τ_a (Figure 8b).

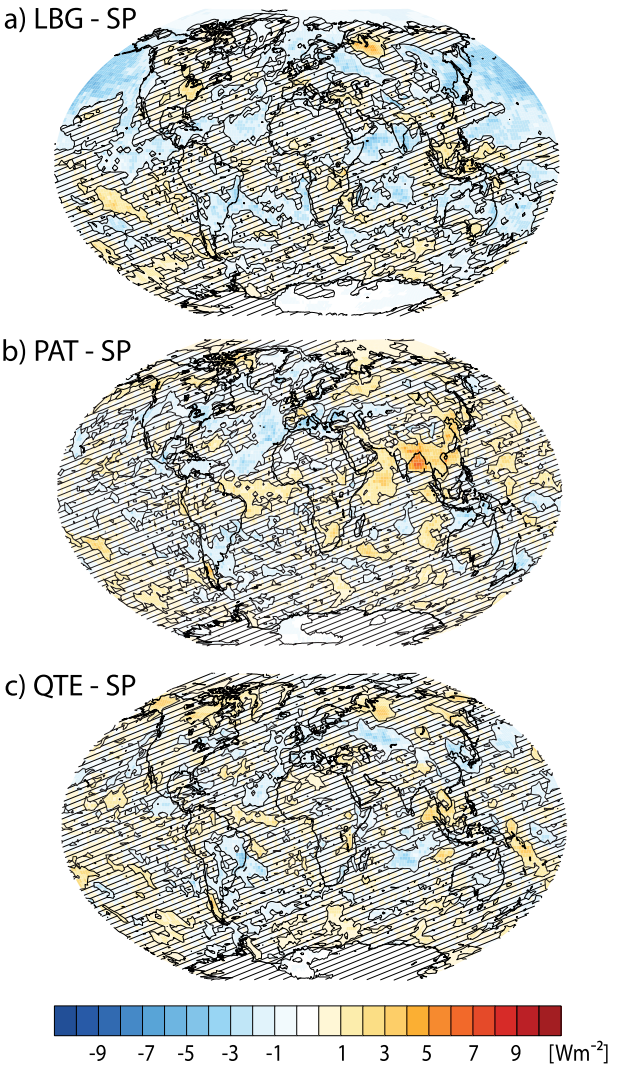

Figure 8. Difference in all-sky ERF. Shown are the differences in annual mean SW TOA ERF_{all} for (a) LBG, (b) PAT, and (c) QTE, relative to SP. Hatching masks out ERF_{all} not significantly different from SP at a confidence level of 10%.

What if we are wrong about the applied natural aerosol background or if the net contribution of RF_{aci} and adjustments to ERF were stronger than the current evidence suggests? To model this, we combine the anthropogenic pattern of the mid-1970s with the low natural background in an additional ensemble (PAT-LBG). The resulting globally and ensemble averaged ERF_{all} of -0.87 Wm^{-2} is slightly smaller than the one for LBG with present-day τ_a, consistent with the smaller mean η_N of 1.18 in PAT-LBG. This illustrates that a different aerosol background can indeed cause a difference in ERF_{all} associated with shifting NH τ_a other than by model-internal variability, but this difference is moderate (15%). This result suggests that a change in ERF_{all} due to the NH shift in τ_a between the mid-1970s and present-day may arise in models that underestimate the natural aerosol burden downwind of East Asia or parameterize by other means strong RF_{aci} and adjustments.

3.2.3. Relative Importance of Background and τ_a

LBG, PAT and PAT-LBG could be interpreted as extreme present-day patterns of heating anomalies due to anthropogenic and natural aerosol. In doing so our results suggest that the magnitude of ERF_{all} is more sensitive to the natural background (compare LBG versus SP and PAT-LBG versus. PAT) than to τ_a (compare SP versus. PAT and LBG versus. PAT-LBG). Accurately constraining the natural background burden might therefore be important to estimate the global mean ERF_{all}, consistent with *Carslaw et al.* [2013]. In the following, we attempt to better understand why the anthropogenic pattern (SP versus. PAT) has such a small impact on the global mean ERF_{all}, compared to the natural background (SP versus. LBG).

LBG uses modified background aerosol in η_N, affecting the cloudy fraction of ERF_{all}, while PAT uses changed τ_a altering both clear- and cloudy-sky ERFs. In order to better estimate the relative importance of τ_a and τ_{bg}, cloudy-sky ERF_{cld} is calculated from annual ensemble averages of ERF_{all}, ERF_{clr}, and total cloud fraction f_{cld} [e.g., *Schulz et al.*, 2006]:

$$f_{cld}ERF_{cld} = ERF_{all} - (1 - f_{cld})ERF_{clr}. \quad (4)$$

Decomposing ERF_{all} into the relative contributions from ERF_{cld} and ERF_{clr} (left and right terms in equation (4)) shows only small changes in the spatial distribution of ERF_{cld} contribution to ERF_{all} between SP and PAT (Figures 9d and 9f). Their similarity is consistent with fairly small changes in η_N between PAT and SP (Figure 3c), e.g., in comparison to LBG that shows a much stronger change in regional cloudy-sky contributions to ERF_{all}. The similarly large cloud cover of up to 80% over the maritime continent and southeast Asia substantially contributes to the strong cloudy-sky contribution over that region.

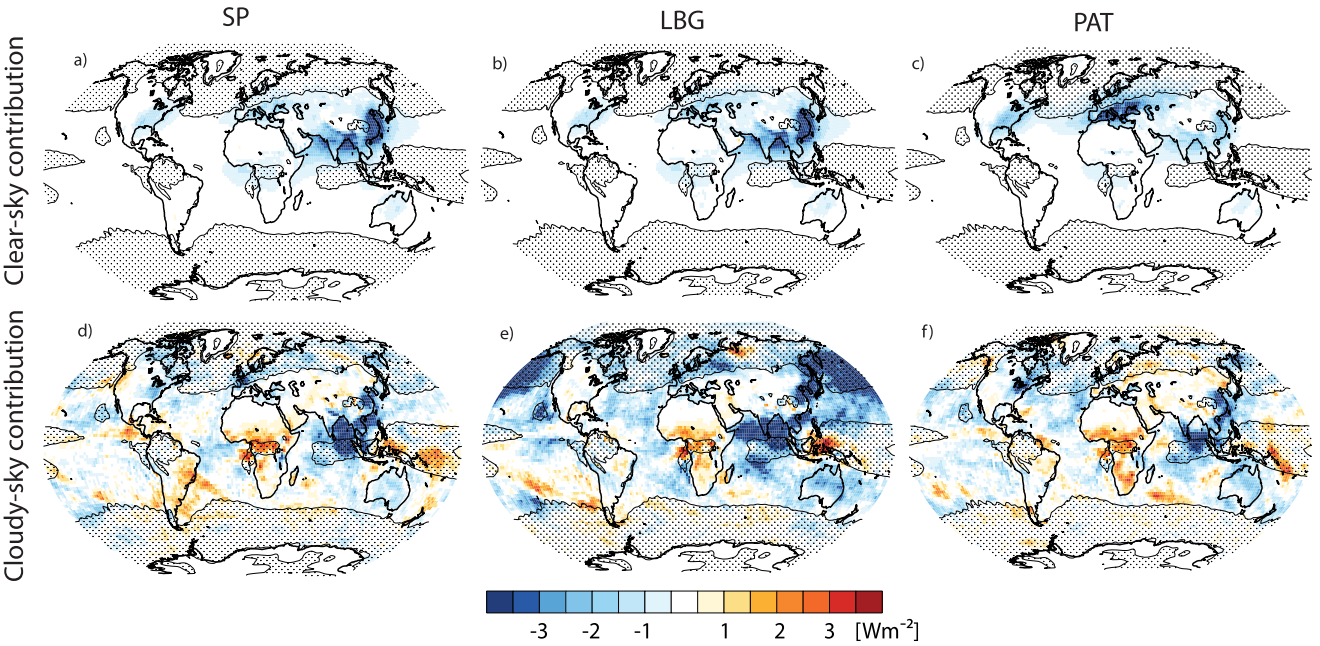

Figure 9. Relative contributions of clear- and cloudy-sky to all-sky ERF. Shown are annually and ensemble averaged (shaded) ERF contributions from (top) clear- and (bottom) cloudy-sky fractions (right and left terms in equation (4)), and (dotted) total cloud cover exceeding 6 okta from (left to right) SP, LBG, and PAT.

Most of the detectable changes in ERF_{all} between SP and PAT are associated with the changing pattern in the ERF_{clr} contributions (Figures 9a and 9c), rather than the effect of τ_a on η_N. Even though τ_a is large over wide areas of Eurasia (Figure 1b), cloud cover reduces the contribution of ERF_{clr} to ERF_{all} (Figure 9c). Cloud masking of ERF_{clr} in PAT particularly occurs at NH high latitudes, where the mean cloud fraction exceeds six okta. In combination with rather small changes in regional contributions to ERF_{cld}, clouds efficiently limit the impact of the changed spatial distribution of τ_a on global mean ERF_{all}.

Lowering the background τ_{gl} substantially increases the contributions from ERF_{cld} to ERF_{all} due to the prescription of a stronger net contribution of RF_{aci} and adjustments. Contributions from ERF_{cld} typically further enhance the negative ERF_{all} in regions situated away from plume centers (Figure 9e). The spatial distribution of ERF_{cld} contributions strongly resembles the one of ERF_{all} (Figure 8a), since ERF_{clr} contributions change little between SP and LBG (Figures 9a and 9b). Large negative ERF_{cld} over the Northern Pacific coincide with areas of typically large total cloud cover, exceeding six okta in the climatological mean. This implies that the ERF_{cld} contribution is here enhanced due to large cloud fractions. The Indian Ocean has far fewer clouds, such that the impact of η_N on ERF_{cld} must more strongly increase the cloudy-sky contribution to ERF_{all}. These regions overlap with areas where ERF_{all} is large enough for detection when sufficiently long time periods for averaging are used (section 3.1.2). Studying anthropogenic aerosol-effects on climate might therefore be most favorable over oceans downwind of major anthropogenic pollution sources when data from periods with relatively low natural-aerosol burden over decadal time scales are sampled. To what extent these results might be affected by uncertainty in the parameterization form of the Twomey effect is addressed next.

3.2.4. Parameterization Uncertainty

Using an alternative parameterization approach for the Twomey effect as a measure of uncertainty in formulating η_N (QTE) has an overall small impact on the global mean ERF. The mean ERF_{clr} is with -0.69 ± 0.0004 Wm^{-2} very similar to SP, while ERF_{all} is with -0.58 ± 0.002 Wm^{-2} a little more negative than SP (Table 2), consistent with the prescribed stronger Twomey effect near plume centers. Drawing ten samples, however, illustrates that variability only allows a distinction in four of ten cases, adopting a 1% confidence level (Figure 6d).

Despite the small impact on the global mean ERF, regional radiative effects are partly sensitive to the functional form of η_N underlining the challenge to constrain regional radiative effects. QTE shows more negative ERFs near centers of the Asian, North American, and European plumes as well as regions in the Indian

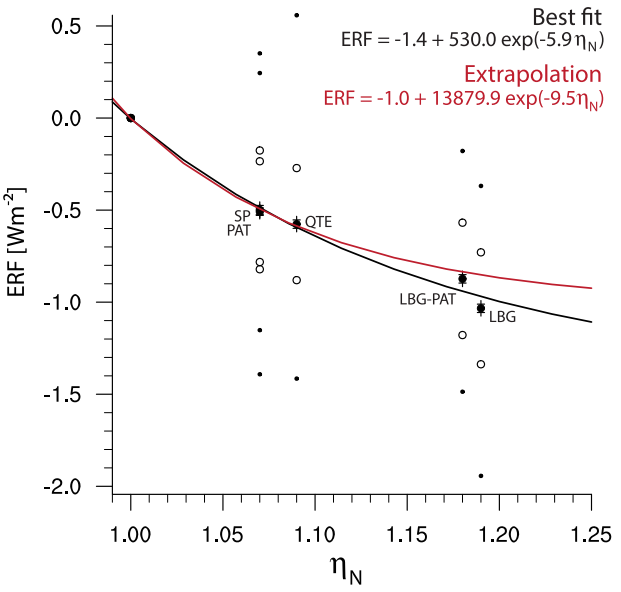

Figure 10. Global mean all-sky ERF as function of η_N. Shown are (filled circles) ensemble means and (error bars) standard error, as well as (open circles) standard deviation and (points) extremes of annual estimates of SW TOA ERF_{all} from all experiment setups with RF_{aci} (compare Table 1). Lines show exponential functions fitted with method of least squares using (black) all data and (red) data for $\eta_N < 1.1$, assuming no ERF_{all} at $\eta_N = 1$.

Ocean (Figure 8c), consistent with larger η_N (Figure 3b) and a larger ERF_{cld} contribution (not shown). Since these have the same sign as the differences with LBG, one could expect that reducing the background aerosol in a setup with the parameterization from QTE would regionally generate even more negative contributions to ERF_{cld} and ERF_{all} than in LBG. This sensitivity is, however, small compared to the differences between LBG and SP and the majority of the globe show no significant differences with QTE (Figure 8c).

3.3. ERF Scalability

As an aid for future model intercomparison studies using MACv2-SP the analysis of a relationship between ERF_{all} and η_N is proposed. The present work suggests that different spatial distributions in QTE, PAT and SP have rather similar global mean ERF_{all} in stark contrast to LBG and LBG-PAT with more negative global mean ERF_{all} (Figure 6). Using the method of least squares, an exponential function is fitted to all ensemble statistics of ERF_{all}, $ERF_{all} = -1.4 + 530 \exp(-5.9\ \eta_N)$ shown in Figure 10. An exponential function is chosen rather than a linear regression in analogy to the parameterization of η_N that flatten with increasing τ_a and thereby bounds the magnitude of ERF_{all}. A flattening is reasonable since the formation of cloud droplets is limited by supersaturation [*Carslaw et al.*, 2013]. Moreover, η_N ranges between one and an upper bound, determined by τ_a and τ_{bg} (Figure 2) that are constrained by emission and deposition of aerosol. It could be questioned whether the few data points are sufficient to fit a curve, but it is hard to imagine outliers from ECHAM6.3 since drastically different spatial distributions of η_N, and possible uncertainty in the functional form of the parameterized Twomey effect have already been considered.

If we had not known ERF_{all} associated with the lowered natural aerosol background, would we have been able to predict it? For addressing this question, the fitting of the curve is repeated only considering ERF_{all} from SP, QTE, and PAT and extrapolating to η_N used in LBG and LBG-PAT. Comparing it to the best fit of all data (lines in Figure 10) gives a similar relationship with a particularly close agreement for global mean $\eta_N < 1.1$. Although the extrapolated and fitted functions differ more for $\eta_N > 1.1$, the difference in ERF_{all} of roughly 0.1 around $\eta_N = 1.19$ is smaller than differences associated with model-internal variability, independent of the choice of the ensemble setup. This suggests that η_N is a predictor of ERF_{all}, thereby providing a framework for more comprehensive studies. It would be interesting to study whether a similar relationship exists in other models, including those with a detailed treatment of aerosol processes. Previous assessments of the effective change in cloud droplet number concentrations associated with anthropogenic aerosol [*Stevens*, 2015] are limited by the current shortage of required information from more comprehensive aerosol modeling studies.

4. Conclusion and Outlook

The present study uses ECHAM6.3 ensembles with fixed climatologies of aerosol optical properties, effective changes in cloud-droplet numbers, and sea-surface temperatures to study the sensitivity of effective radiative forcing of anthropogenic aerosol (ERF) to prescribed patterns of anthropogenic aerosol radiative effects. For estimating how year-to-year meteorological variability influences the ERF estimate, we calculate the annually and globally averaged ERF 180 times, accounting for meteorological variability in the decade 2001–2010. Considering both ERF_{ari} and ERF_{aci} yields a TOA globally and annually averaged ERF in the

shortwave spectrum of -0.50 ± 0.002 Wm^{-2} (ensemble mean $\pm$ standard error) for all sky, and -0.67 ± 0.0004 Wm^{-2} for clear sky.

Natural variability affects the significance and precision of all-sky ERF estimates. Our analysis suggests that the natural year-to-year variability about the precise long-term mean in all-sky ERF is 0.32 Wm^{-2} (standard deviation). A statistical assessment indicates that simulation lengths on the order of decades are needed for estimating a global and annual mean all-sky ERF with a precision of 0.1 Wm^{-2}. Other models might require different lengths, e.g., those with different model-internal variability in the SW TOA radiation budget. Assuming our variability was representative, a three-member ensemble, as proposed in the protocol of the radiative forcing model inter-comparison project (RFMIP) [*Pincus et al*., 2016], could at best determine the global mean all-sky ERF of a specific year with a precision of 0.2 Wm^{-2}. If only one simulation per model was performed, the contribution from model-internal variability to uncertainty in all-sky ERF could not be estimated. Moreover, ECHAM6.3 has a small net contribution of rapid adjustments to ERF compared to the magnitude of RF from double radiation calls. Since RF is less variable and directly comparable to offline radiation calculations, it might be better suited for the comparison of the rather small ensembles to be performed in RFMIP. Estimating ERF should, however, not be discarded since RF is difficult to deduce from observations that always include adjustments.

The sensitity of all-sky ERF is investigated with ensembles using targeted parameter perturbations. Substantially different global distributions of anthropogenic aerosol, namely present-day versus mid-1970s, give global and annual mean all-sky ERFs that are surprisingly challenging to distinguish when model-internal variability is considered. Reducing the sample size can even completely eliminate the chances of a separation, here in 8 out of 10 subsamples of the same boundary conditions. This finding suggests that the uncertainty in the global mean of all-sky ERF differences between the mid-1970s and present day from past studies [e.g., *Shindell et al*., 2013] might in large part measure the noise of meteorological variability, rather than a meaningful difference in the global mean ERF. Hypothesizing the anthropogenic aerosol distribution of the mid-1970s as an extreme change in the pattern of the present-day anthropogenic aerosol radiative effects suggests that the pattern of the anthropogenic aerosol burden is less important for the global mean ERF than the total global burden. By implication comparably smaller uncertainty in the present-day distribution of anthropogenic aerosol is unlikely to meaningfully impact the globally and annually averaged all-sky ERF. In contrast to this small effect, an extremely low natural aerosol-burden allows the Twomey effect to have a much stronger impact on all-sky ERF. The more negative all-sky ERF, even detectable in rather small sample sizes, is explained by larger cloudy-sky ERF contributions to all-sky ERFs, similar to what has been found by other studies [*Carslaw et al*., 2013; *Wilcox et al*., 2015]. Uncertainty in the functional form for parameterizing the Twomey effect has a small impact (16%) on all-sky ERF. Interpreting the low background as a tool for achieving a strong net ERF_{aci} allows us to construct a testable hypothesis: the anthropogenic pattern effect on the all-sky ERF is moderate in models with strong ERF_{aci} and/or small natural aerosol burden. Testing this hypothesis with existing output from CMIP models might be difficult due to the complexity of processes involved. However, experiments with aerosol-climate models illustrate that allowing such models to reach particularly low N results in considerably stronger ERF_{aci} [*Hoose et al*., 2009].

Aerosol-climate models are known to produce diverse amounts of natural aerosol, that are partly associated with uncertainties in the atmospheric model component. In addition to natural aerosol, cloud masking affects all-sky ERF. Since representing both clouds and natural aerosol are known challenges in climate modeling, a coordinated inter-comparison study with the same aerosol properties, using MACv2-SP and the suggested relationship of ERF to η_N, would help to advance the understanding of their relative importance. MACv2-SP could be used to quantify the relative contributions from model components other than the treatment of aerosol to the model spread in ERF, avoiding uncertainties associated with coupling a complex aerosol parameterization. The simplicity of MACv2-SP would herein help us to identify basic problems that affect ERF and would otherwise stay hidden in the model complexity. Such knowledge would be useful to pinpoint the most urgent requirements for model development for improved climate simulations.

Acknowledgments
This work is funded by the Max Planck Society and the FP7 project "BACCHUS" (603445). We thank Stefan Kinne, Max Planck Institute for Meteorology, the associate editor Chris Bretherton, the chief editor Robert Pincus, and two anonymous reviewers for their constructive comments for improving our manuscript. We further acknowledge the usage of the DKRZ supercomputer facilities and the CERES EBAF Ed2.7 data set obtained from the obs4MIPs (https://www.earthsystemcog.org/projects/obs4mips/) project hosted on the Earth System Grid Federation (http://esgf.llnl.gov). Model data and scripts of this study are archived by the Max Planck Institute for Meteorology and can be made accessible by contacting publications@mpimet.mpg.de.

References

Albrecht, B. (1989), Aerosols, cloud microphysics and fractional cloudiness, *Science*, *245*, 1227–1230.
Andreae, M. O. (2009), Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions, *Atmos. Chem. Phys.*, *9*(2), 543–556, doi:10.5194/acp-9-543-2009.

Blanchard, D. C., and A. H. Woodcock (1980), The production, concentration, and vertical distribution of the sea-salt aerosol*, *Ann. N. Y. Acad. Sci.*, *338*(1), 330–347, doi:10.1111/j.1749-6632.1980.tb17130.x.
Bony, S., et al. (2015), Clouds, circulation and climate sensitivity, *Nat. Geosci.*, *8*(4), 261–268, doi:10.1038/ngeo2398.
Carslaw, K., et al. (2013), Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, *503*(7474), 67–71.
Cesana, G., and H. Chepfer (2012), How well do climate models simulate cloud vertical structure? A comparison between CALIPSO-GOCCP satellite observations and CMIP5 models, *Geophys. Res. Lett.*, *39*, L20803, doi:10.1029/2012GL053153.
Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak (2014), Projecting north American climate over the next 50 years: Uncertainty due to internal variability, *J. Clim.*, *27*(6), 2271–2296, doi:10.1175/JCLI-D-13-00451.1.
Dolinar, E. K., X. Dong, B. Xi, J. H. Jiang, and H. Su (2014), Evaluation of CMIP5 simulated clouds and TOA radiation budgets using NASA satellite observations, *Clim. Dyn.*, *44*(7), 2229–2247, doi:10.1007/s00382-014-2158-9.
Fiedler, S., K. Schepanski, B. Heinold, P. Knippertz, and I. Tegen (2013), Climatology of nocturnal low-level jets over North Africa and implications for modeling mineral dust emission, *J. Geophys. Res. - Atmos.*, *118*, 6100–6121, doi:10.1002/jgrd.50394.
Fiedler, S., P. Knippertz, S. Woodward, G. Martin, N. Bellouin, A. Ross, B. Heinold, K. Schepanski, C. Birch, and I. Tegen (2015), A process-based evaluation of dust-emitting winds in the CMIP5 simulation of HadGEM2-ES, *Clim. Dyn.*, *46*(3), 1107–1130, doi:10.1007/s00382-015-2635-9.
Ghan, S., et al. (2016), Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability, *Proc. Natl. Acad. Sci. U. S. A.*, *113*(21), 5804–5811, doi:10.1073/pnas.1514036113.
Ghan, S. J., H. Abdul-Razzak, A. Nenes, Y. Ming, X. Liu, M. Ovchinnikov, B. Shipway, N. Meskhidze, J. Xu, and X. Shi (2011), Droplet nucleation: Physically-based parameterizations and comparative evaluation, *J. Adv. Model. Earth Syst.*, *3*, M10001, doi:10.1029/2011MS000074.
Golaz, J.-C., L. W. Horowitz, and H. Levy (2013), Cloud tuning in a coupled climate model: Impact on 20th century warming, *Geophys. Res. Lett.*, *40*, 2246–2251, doi:10.1002/grl.50232.
Holben, B., et al. (1998), AERONET––A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*(1), 1–16, doi:10.1016/S0034-4257(98)00031-5.
Hoose, C., J. E. Kristjánsson, T. Iversen, A. Kirkevåg, Ø. Seland, and A. Gettelman (2009), Constraining cloud droplet number concentration in GCMs suppresses the aerosol indirect effect, *Geophys. Res. Lett.*, *36*, L12807, doi:10.1029/2009GL038568.
Huneeus, N., et al. (2016), Forecasting the northern African dust outbreak towards Europe in April 2011: A model intercomparison, *Atmos. Chem. Phys.*, *16*(8), 4967–4986, doi:10.5194/acp-16-4967-2016.
Jiang, J. H., et al. (2012), Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations, *J. Geophys. Res.*, *117*, D14105, doi:10.1029/2011JD017237.
Kinne, S., et al. (2006), An AeroCom initial assessment––Optical properties in aerosol component modules of global models, *Atmos. Chem. Phys.*, *6*(7), 1815–1834, doi:10.5194/acp-6-1815-2006.
Kinne, S., D. O'Donnel, P. Stier, S. Kloster, K. Zhang, H. Schmidt, S. Rast, M. Giorgetta, T. F. Eck, and B. Stevens (2013), MAC-v1: A new global aerosol climatology for climate studies, *J. Adv. Model. Earth Syst.*, *5*, 704–740, doi:10.1002/jame.20035.
Kühn, T., et al. (2014), Climate impacts of changing aerosol emissions since 1996, *Geophys. Res. Lett.*, *41*, 4711–4718, doi:10.1002/2014GL060349.
Largeron, Y., F. Guichard, D. Bouniol, F. Couvreux, L. Kergoat, and B. Marticorena (2015), Can we use surface wind fields from meteorological reanalyses for sahelian dust emission simulations?, *Geophys. Res. Lett.*, *42*, 2490–2499, doi:10.1002/2014GL062938.
Li, J.-L. F., et al. (2012), An observationally based evaluation of cloud ice water in CMIP3 and CMIP5 GCMs and contemporary reanalyses using contemporary satellite data, *J. Geophys. Res.*, *117*, D16105, doi:10.1029/2012JD017640.
Liu, X., J. E. Penner, B. Das, D. Bergmann, J. M. Rodriguez, S. Strahan, M. Wang, and Y. Feng (2007), Uncertainties in global aerosol simulations: Assessment using three meteorological data sets, *J. Geophys. Res.*, *112*, D11212, doi:10.1029/2006JD008216.
Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009), Toward optimal closure of the earth's top-of-atmosphere radiation budget, *J. Clim.*, *22*(3), 748–766, doi:10.1175/2008JCLI2637.1.
Lohmann, U., and S. Ferrachat (2010), Impact of parametric uncertainties on the present-day climate and on the anthropogenic aerosol effect, *Atmos. Chem. Phys.*, *10*(23), 11,373–11,383, doi:10.5194/acp-10-11373-2010.
Mauritsen, T., et al. (2012), Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, *4*, M00A01, doi:10.1029/2012MS000154.
McGibbon, J., and C. S. Bretherton (2017), Skill of ship-following large-eddy simulations in reproducing MAGIC observations across the Northeast Pacific stratocumulus to cumulus transition region, *J. Adv. Mod. Earth Syst.*, *9*, doi:10.1002/2017MS000924, in press.
Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, and T. Wong (2009), An observationally based energy balance for the Earth since 1950, *J. Geophys. Res.*, *114*, D17107, doi:10.1029/2009JD012105.
Myhre, G., et al. (2013), Anthropogenic and natural radiative forcing, in *Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., Cambridge Univ. Press, Cambridge, U. K.
Nam, C., S. Bony, J.-L. Dufresne, and H. Chepfer (2012), The 'too few, too bright' tropical low-cloud problem in CMIP5 models, *Geophys. Res. Lett.*, *39*, L21801, doi:10.1029/2012GL053421.
Penner, J. E., et al. (2002), A comparison of model- and satellite-derived aerosol optical depth and reflectivity, *J. Atmos. Sci.*, *59*(3), 441–460.
Penner, J. E., J. Quaas, T. Storelvmo, T. Takemura, O. Boucher, H. Guo, A. Kirkevåg, J. E. Kristjánsson, and Ø. Seland (2006), Model intercomparison of indirect aerosol effects, *Atmos. Chem. Phys.*, *6*(11), 3391–3405, doi:10.5194/acp-6-3391-2006.
Pincus, R., P. M. Forster, and B. Stevens (2016), The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6, *Geosci. Model Dev. Discuss.*, *2016*, 1–18, doi:10.5194/gmd-2016-88.
Quaas, J., O. Boucher, and U. Lohmann (2006), Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data, *Atmos. Chem. Phys.*, *6*(4), 947–955, doi:10.5194/acp-6-947-2006.
Quaas, J., et al. (2009), Aerosol indirect effects––General circulation model intercomparison and evaluation with satellite data, *Atmos. Chem. Phys.*, *9*(22), 8697–8717.
Rosenfeld, D., S. Sherwood, R. Wood, and L. Donner (2014), Climate effects of aerosol-cloud interactions, *Science*, *343*(6169), 379–380, doi:10.1126/science.1247490.
Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo (2013), Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models?, *J. Adv. Model. Earth Syst.*, *5*, 117–133, doi:10.1002/jame.20013.
Schulz, M., et al. (2006), Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, *Atmos. Chem. Phys.*, *6*(12), 5225–5246, doi:10.5194/acp-6-5225-2006.
Seifert, A., T. Heus, R. Pincus, and B. Stevens (2015), Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection, *J. Adv. Model. Earth Syst.*, *7*, 1918–1937, doi:10.1002/2015MS000489.

Shao, Y. (2001), A model for mineral dust emission, *J. Geophys. Res.*, *106*, 20,239–20,254.
Shepherd, T. G. (2014), Atmospheric circulation as a source of uncertainty in climate change projections, *Nat. Geosci.*, *7*(10), 703–708.
Shindell, D. T., et al. (2013), Radiative forcing in the accmip historical and future climate simulations, *Atmos. Chem. Phys.*, *13*(6), 2939–2974, doi:10.5194/acp-13-2939-2013.
Smith, S. J., J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias (2011), Anthropogenic sulfur dioxide emissions: 1850–2005, *Atmos. Chem. Phys.*, *11*(3), 1101–1116, doi:10.5194/acp-11-1101-2011.
Stevens, B. (2015), Rethinking the lower bound on aerosol radiative forcing, *J. Clim.*, *28*(12), 4794–4819, doi:10.1175/JCLI-D-14-00656.1.
Stevens, B., and G. Feingold (2009), Untangling aerosol effects on clouds and precipitation in a buffered system, *Nature*, *461*, 607–613.
Stevens, B., et al. (2013), Atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model. Earth Syst.*, *5*, 146–172, doi:10.1002/jame.20015.
Stevens, B., S. Fiedler, S. Kinne, K. Peters, S. Rast, J. Müsse, S. J. Smith, and T. Mauritsen (2017), MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, *Geosci. Model Dev.*, *10*(1), 433–452, doi:10.5194/gmd-10-433-2017.
Stier, P., et al. (2013), Host model uncertainties in aerosol radiative forcing estimates: Results from the AeroCom Prescribed intercomparison study, *Atmos. Chem. Phys.*, *13*(6), 3245–3270, doi:10.5194/acp-13-3245-2013.
Twomey, S. (1974), Pollution and the planetary albedo, *Atmos. Environ.*, *8*, 1251–1256.
Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper (1996), Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment, *Bull. Am. Meteorol. Soc.*, *77*(5), 853–868, doi:10.1175/1520-0477(1996)077 <0853:CATERE >2.0.CO;2.
Wilcox, L. J., E. J. Highwood, B. B. B. Booth, and K. S. Carslaw (2015), Quantifying sources of inter-model diversity in the cloud albedo effect, *Geophys. Res. Lett.*, *42*, 1568–1575, doi:10.1002/2015GL063301.