The global aerosol-cloud first indirect effect estimated using MODIS, MERRA, and AeroCom

D. T. McCoy1,2, F. A.-M. Bender3, J. K. C. Mohrmann1, D. L. Hartmann1, R. Wood1, and D. P. Grosvenor2

1Department of Atmospheric Sciences, University of Washington, Seattle, USA, 2School of Earth and Environment, University of Leeds, Leeds, UK, 3Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Abstract Aerosol-cloud interactions (ACI) represent a significant source of forcing uncertainty in global climate models (GCMs). Estimates of radiative forcing due to ACI in Fifth Assessment Report range from −0.5 to −2.5 W m−2. A portion of this uncertainty is related to the first indirect, or Twomey, effect whereby aerosols act as nuclei for cloud droplets to condense upon. At constant liquid water content this increases the number of cloud droplets (Nd) and thus increases the cloud albedo. In this study we use remote-sensing estimates of Nd within stratocumulus regions in combination with state-of-the-art aerosol reanalysis from Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA2) to diagnose how aerosols affect Nd. As in previous studies, Nd is related to sulfate mass through a power law relationship. The slope of the log-log relationship between Nd and SO4 in maritime stratocumulus is found to be 0.31, which is similar to the range of 0.2–0.8 from previous in situ studies and remote-sensing studies in the pristine Southern Ocean. Using preindustrial emissions models, the change in Nd between preindustrial and present day is estimated. Nd is inferred to have more than tripled in some regions. Cloud properties from Moderate Resolution Imaging Spectroradiometer (MODIS) are used to estimate the radiative forcing due to this change in Nd. The Twomey effect operating in isolation is estimated to create a radiative forcing of −0.97 ± 0.23 W m−2 relative to the preindustrial era.

1. Introduction

Aerosols have the ability to alter the radiative flux at the top-of-atmosphere (TOA). The influence of aerosol on TOA radiative flux is separated into two categories: effective radiative forcing from aerosol-radiation interaction (ERFai) and effective radiative forcing from aerosol-cloud interactions (ERFaci) [Boucher et al., 2014]. Within the Fifth Assessment Report (AR5) more intermodel uncertainty was contributed by ERFaci [Boucher et al., 2014]. This adds significant uncertainty to observational estimates of the equilibrium climate sensitivity.

ERFaci may be further subdivided. Enhancement in cloud condensation nuclei (CCN) increases the number concentration of cloud droplets (Nd) [Twomey, 1977]. This is referred to as the first indirect, cloud albedo or Twomey effect. Also, because smaller cloud droplets precipitate less efficiently this may enhance the cloud liquid water path (LWP) and cloud cover [Albrecht, 1989], which is the second indirect effect. In this study we focus on the first indirect, or Twomey, effect.

Numerous studies of satellite and in situ data have offered estimates of ERFaci. In situ studies have the capability to differentiate aerosol species and examine their interaction with cloud properties but operate at high-spatial and temporal resolution of short time period and constrained study regions [Boucher and Lohmann, 1995; Lowenthal et al., 2004]. This makes it difficult to offer a top-down global estimate of the ERFaci due to anthropogenic aerosol. Satellite-based studies of ERFaci have the benefit of an extensive spatial and temporal domain and a top-down diagnosis of the effect of aerosol on cloud properties, but satellites are unable to differentiate aerosol species and thus these studies are restricted to examining either aerosol optical depth (AOD) or aerosol index (AI) [Bellouin et al., 2013; Bréon et al., 2002; Lebsock et al., 2008; Nakajima et al., 2001; Quaas et al., 2009; Sekiguchi et al., 2003; Storelvmo et al., 2006; Wetzel and Stowe, 1999]. Note that AI is a measure of both aerosol amount and size. Over oceans it is defined as the product of the AOD and Ångström coefficient [Bréon et al., 2002]. Use of column-integrated quantities such as AI and AOD must disregard the vertical structure of aerosol. When vertically resolved measures of aerosol and cloud from spaceborne lidar are used, it can be seen that location of aerosol and cloud layers is crucial...
In this study we employ near-surface measures of aerosol to use a CCN proxy that is vertically collocated with boundary layer cloud cover. Use of aerosol reanalysis also avoids the concern that aerosols are swelled by nearby cloud leading to a stronger aerosol scattering [Twomey et al., 2009]. It should be noted that in its calculation of AOD assimilation in Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA2) follows Chin et al. [2002]. Further, while Nd may depend on AOD or aerosol index in a certain way in the current atmosphere, there is no guarantee that the same aerosols contributed to these quantities in the same way in the pristine, preindustrial climate, making the interpretation of these observational constraints unclear [Penner et al., 2011].

In this work we present an alternate approach to examining how aerosols affect cloud albedo by leveraging the speciation and vertical resolution provided by aerosol reanalysis. In section 2 we will discuss the methodology used to create the observational and reanalysis data sets. In section 3 we will show that aerosol mass concentration predicts a significant amount of variability in cloud microphysical properties, and we will show that global climate model (GCM) aerosol fields are sufficiently well resolved to reproduce the climatology of observed cloud microphysical properties. Finally, we will use models of preindustrial emissions to estimate how much anthropogenic sulfate has altered Nd. Based on this we will offer a rough estimate of the radiative forcing induced by anthropogenic aerosol through the Twomey effect.

2. Methods

In this work we examine how cloud brightening is influenced by aerosol mass concentration in the boundary layer. All data are resolved at 1° × 1°-monthly resolution. Data describing cloud properties are collected from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument [King et al., 2003], and data describing aerosol properties are collected from Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2) [Buchard et al., 2015; Molod et al., 2015; Rienecker et al., 2011] and several aerosol models participating in AeroCom phase II [Myhre et al., 2013]. MODIS observations from 2001 to 2013 are examined in combination with MERRA2 reanalysis from the same time period. In the case of the AeroCom phase II models, only 1 year of data was available for each model and this year is compared to climatological data from MODIS. We will now discuss the cloud and aerosol data sets independently.

2.1. Cloud Droplet Number Density (Nd)

The MODIS instrument offers remotely sensed observations of cloud microphysical properties. It does this by using radiances from the visible and near-IR 2.1 μm and 0.86 μm channels to simultaneously retrieve cloud optical depth (τ) and cloud droplet effective radius (rd) [Platnick et al., 2003]. In this work we utilize the Collection 5.1 MODIS data from the Terra platform, but this algorithm is effectively the same in MODIS Collection 6 [Cho et al., 2015]. The number concentration of cloud droplets may be calculated using τ and rd following the algorithm presented in Bennartz [2007] with modifications following Grosvenor and Wood [2014]. Nd is particularly useful because it (1) influences cloud albedo and (2) bears a direct relationship to the number of available CCN, thus linking aerosol sources to cloud albedo [Wood, 2012].

It is important to keep in mind that the retrieval of Nd is fairly complex and suffers from a variety of measurement biases [Cho et al., 2015; Grosvenor and Wood, 2014; Zhang and Platnick, 2011; Zhang et al., 2012]. The retrieval of Nd uses a plane-parallel radiative transfer model to create a lookup table for the 2.1 μm and 0.86 μm retrieval bands. In situations with heterogeneous clouds [Zhang et al., 2012] or when the Sun is near the horizon [Grosvenor and Wood, 2014] the retrieval becomes highly biased. Further, Nd must assume the cloud’s vertical condensate structure. Nd is calculated using the relationship

\[ N_d = \sqrt{\frac{\Gamma_{\text{eff}}}{\pi D_w}} \sqrt{\frac{10^\tau}{\pi^2 r_d^3}} \]  

where \( D_w \) is the density of liquid water, \( \tau \) is the optical depth, the constant term \( k \) is parameterized as detailed in Wood [2000], and \( \Gamma_{\text{eff}} \) is the effective lapse rate and can be expressed as the adiabatic lapse rate multiplied by an adiabaticity factor, i.e. \( \Gamma_{\text{eff}} = f_{\text{ad}} \Gamma_{\text{ad}} \). While \( f_{\text{ad}} \) is near unity in stratocumulus [Albrecht et al., 1990; Wood, 2012; Zuiderma et al., 2005], the constraint on this factor in other regions is still poor [Merk et al., 2016]. In this study, as in Bennartz [2007], we will assume that \( f_{\text{ad}} = 1 \) for the purpose of calculating Nd. This assumption will be discussed in more detail in section 3.3. Information from microwave radiometers may also be used to calculate Nd. Bennartz [2007] compared the retrieval of Nd that uses MODIS to measure
optical depth and effective radius to a retrieval where MODIS effective radius was used in combination with Advanced Microwave Scanning Radiometer-EOS (AMSR-E) liquid water path. Comparison of these two retrievals of Nd within overcast situations was found to be similar, but MODIS and AMSR-E LWP diverge from each other when cloud cover is less homogeneous [Horvath and Davies, 2007].

Covariance between Nd and aerosols is examined using monthly data. To create monthly Nd data, we first filter the daily Nd observations. We select daily data where the retrieval is being conducted in a situation that is roughly analogous to the uniform cloud cover assumed in the plane-parallel radiative transfer model. We exclude data when the fraction of liquid cloud cover is less than 0.8. This restricts the cloud cover to be both more extensive and homogeneous [Wood and Hartmann, 2006]. If fewer than 10 days in a given month meet this criterion, the mean Nd for that month is not considered representative and is disregarded. This restricts the MODIS Nd data set to the stratocumulus-dominated subtropics, midlatitude oceans, and a few continental cloud regimes (Figure 1). In the first part of this paper we focus on the maritime stratocumulus regions (centered on the regions in Klein and Hartmann [1993]) because near-horizon Sun angles rarely occur and the clouds are near adiabatic. This makes the Nd retrieval in this region relatively reliable.

In the main body of this paper we will only consider retrievals where liquid cloud fraction (CF) is greater than 80%. As noted above this criterion was selected to restrict retrievals of Nd to situations where the clouds were approximately plane parallel in keeping with the retrieval algorithm’s assumptions. While this makes the interpretation of our results somewhat easier in the sense of understanding the retrievals, it does complicate the physical interpretation of our results. Gryspeerdt and Stier [2012] showed substantially different sensitivities of Nd to AOD in different cloud regimes. To examine the sensitivity of our analysis to the CF constraint applied to the daily Nd retrievals, the analysis presented in this paper was repeated when only days where CF was greater than 50% were considered and when no CF restriction was applied. This is presented in the supporting information. Overall, the results presented in the main body of the paper do not change in a qualitative sense if a less stringent cloud fraction criterion is applied to examine Nd where the retrieval is nearer to the plane-parallel ideal.

### 2.2. Aerosol Properties

In this work we only consider the mass concentration of aerosol species in the boundary layer. The majority of this analysis centers on four common aerosol species. These species are sulfate (SO$_4$), sea salt (SS), black carbon (BC), and dust (DU). Model output detailing the mass concentration of these species near the surface is available in MERRA2 and the phase II AeroCom models. In MERRA2 the 910 hPa level was taken as the near surface level, and in the AeroCom phase II models the model layer nearest the surface was used as the near surface level. MERRA2 also supplies concentrations of several species not available across the AeroCom phase II models: methane sulphonic acid (MSA), which acts as a diagnostic of biogenic sulfate production [Ayers and Gras, 1991], and organic carbon (OC), which has been recently shown to be effective as CCN due to surface activity [Ruehl et al., 2016]. MERRA2 mass mixing ratios for each aerosol species on the 910 hPa pressure level were downloaded at 3-hourly 0.5° × 0.626° resolution. The mass concentrations were averaged to a monthly

**Figure 1.** The fraction of months where retrievals of Nd are sufficiently common to be considered representative. In order to be considered a month must have at least 10 days where the fraction of liquid cloud was greater than 80%. Study regions are marked by white boxes and have been annotated.
Two regression models are considered in this study. The first model is a simple dependence of Nd on SO\textsubscript{4}, as in Boucher and Lohmann [1995]

\[ \log_{10}N_d = a_1 \log_{10}SO_4 + c \]  

(2)

The second model is a multiple linear regression on SO\textsubscript{4}, SS, BC, and DU

\[ \log_{10}N_d = a_1 \log_{10}SO_4 + a_2 \log_{10}SS + a_3 \log_{10}BC + a_4 \log_{10}DU + c \]  

(3)

To reduce the effects of colinearity in the predictor variables, partial least squares with tenfold cross validation is used when training the regression model [Kohavi, 1995]. Regression is performed on the 1°×1°-monthly data for the 2001–2013 time period when training the model using MODIS and MERRA2. When AeroCom data not included in the indirect experiment [S. Zhang et al., 2016] are used to describe aerosol properties, climatological cloud properties and climatological aerosol mass concentrations at monthly 1°×1° resolution are used to train the model.

### 3. Results

#### 3.1. Creation of a Regression Model

Because the stratocumulus regions offer the most reliable observations of Nd (see section 2.1) we will use these regions to train our regression model. To do this, we train MODIS Nd at 1°×1°-monthly resolution over the 2001–2013 data record on MERRA2 aerosol data. The regression model is trained separately in each of the stratocumulus regions and also using the aggregate data from the Peruvian, Californian, Namibian, Canarian, and Australian regions. This produces a total of six regression models. The coefficients relating aerosol concentrations from MERRA2 to MODIS Nd are shown in Figure 2. Except for dust, all the coefficients are distinct
However, the uncertainty of the coefficient relating dust to Nd may be due to the fact that significant amounts of dust are not present in all five stratocumulus regions and the estimate of the coefficient relating dust to Nd offered by training the model on all five regions may be more physically meaningful in this case. Organic carbon was also considered as a predictor in the regression model, but the coefficient relating OC to Nd was not found to be significantly different than zero (0.02 ± 0.20) and was not found to greatly increase the explained variance by the model. This does not prove that OC is incapable of acting as an efficient CCN. Model representation of OC may be flawed or retrieval of Nd in regions with significant OC may be biased. Because no strong dependence on OC can be determined from the data considered in this study, it will be neglected. Of the predictors considered, variability in SO4 contributed the most strongly to variability in Nd over the stratocumulus regions. Regression on SO4 alone captures a significant amount of the variability in stratocumulus Nd over the 2001–2013 record. The correlation coefficient between Nd predicted by SO4 and observed Nd was $r = 0.6 ± 0.01$ over the aggregate data from the stratocumulus regions. When DU, SS, and BC were considered as predictors in the regression model, the correlation was $r = 0.66 ± 0.07$.

Ultimately, the training of the regression model on the aggregate data from the five stratocumulus regimes yields the following model

$$\log_{10} N_d = 0.31 \log_{10} SO_4 - 0.19 \log_{10} SS + 0.057 \log_{10} BC + 0.031 \log_{10} DU + 1.78$$

(4)

where the units on all aerosol mass concentrations are $\mu g/m^3$ and the units on Nd is $cm^{-3}$. If only SO4 is considered, the following model is created

$$\log_{10} N_d = 0.41 \log_{10} SO_4 + 2.11$$

(5)

We argue that the former model offers a more realistic estimate of the dependence of Nd on SO4. Accounting for covariability between aerosol species should offer a more robust depiction of the sensitivity of Nd to SO4. In support of this, the sensitivity of Nd to SO4 is compared to the estimate offered by previous studies. This is shown in Figure 3. The sensitivities of Nd to SO4 diagnosed using in situ maritime observations in Lowenthal et al. [2004], Allen et al. [2011], and Boucher and Lohmann [1995] are compared to the sensitivities in equations (4) and (5). It should be noted that the in situ studies do not account for covariability between SO4 and other aerosol species, making the comparison to equation (5) more appropriate. The sensitivity diagnosed using in situ studies tends to be higher than diagnosed from the reanalysis and satellite data. This is reasonable given that the sampling resolution in these studies is much higher than offered by a satellite [Ma et al., 2015; McComiskey and Feingold, 2012]. That is to say, the in situ observations may be measuring the instantaneous effect of SO4 on Nd, while the satellite and reanalysis may offer a more integrated picture including other feedbacks between clouds and aerosols.

It is interesting to compare the sensitivity derived from the subtropical stratocumulus to the sensitivity derived in McCoy et al. [2015] over the pristine Southern Ocean, where sources of DU and BC are weak. The sensitivity derived from the Southern Ocean is in good agreement with the sensitivity derived from the subtropical low cloud regimes, as in equation (4), indicating that regressing upon BC, DU, and SS in addition to SO4 offers a more accurate estimate of the sensitivity.
It is also interesting to note that the coefficients relating BC, DU, and SO\textsubscript{4} to Nd are all positive in equation (4), while the coefficient relating SS to Nd is negative. This is unexpected. Higher concentrations of SS should lead to a larger CCN number, especially since larger diameter SS aerosols were not considered in this analysis (see section 2.2). One possible explanation is that because SS generation covaries strongly with wind speed [Grythe et al., 2014] the negative coefficient in equation (4) may be describing a dynamical effect.

3.2. Can the Aerosol Modeled by GCMs Reproduce Nd in the Stratocumulus Regimes?

We have shown that MERRA2 reanalysis has the capacity to reproduce the observed variability in Nd across the stratocumulus cloud decks. We also showed that SO\textsubscript{4} explains a large fraction of the variability in Nd. This naturally raises the question: is the sulfate produced by GCMs sufficiently close to reality that it can reproduce a reasonable pattern of Nd? That is to say, do aerosol models have the capacity to produce a reasonable first indirect effect based on their aerosol modeling? In this section we evaluate several GCMs participating in phase II of the AeroCom comparison exercise [Myhre et al., 2013]. We supplement this analysis by examining the Nd calculated from the AeroCom phase II models and the indirect analysis experiment conducted by AeroCom [S. Zhang et al., 2016].

The sensitivity of Nd to SO\textsubscript{4} calculated from the AeroCom phase II models and the indirect analysis experiment is shown in Figure 4. Climatological SO\textsubscript{4} and Nd are used to train the regression model in equation (4) for both MERRA2 and AeroCom data. The correlation is also calculated using climatological data. This is done so that an apples-to-apples comparison can be made between MERRA2 and the AeroCom models. Overall, the AeroCom models do quite well at producing a reasonable sensitivity of Nd to SO\textsubscript{4} and can reproduce a significant fraction of the MODIS climatological Nd variability. This analysis was repeated in the ECHAM-HAM2 and GFDL-AM3 models, which output both aerosol fields and Nd while participating in the indirect effect exercise conducted by AeroCom. Interestingly, the regression models created using GFDL-AM3 aerosol to predict GFDL-AM3 Nd and using ECHAM-HAM2 aerosol to predict ECHAM-HAM2 Nd only correlated at, respectively, \(r = 0.42\) and \(r = 0.58\) (Figure 4). It should be noted that this analysis is only presented for illustrative purposes, and several caveats must be made: (1) two models do not represent an exhaustive survey of models with prognostic cloud microphysics and (2) the Nd diagnosed by the GCMs is not intended to be equivalent to MODIS-observed Nd. Given the strong dependence of Nd on SO\textsubscript{4} diagnosed using MODIS Nd and SO\textsubscript{4} from GCMs, however, it is surprising that the dependence of modeled Nd on modeled SO\textsubscript{4} is not noticeably
stronger than the dependence of MODIS Nd on modeled or reanalyzed sulfate. This result is consistent with the uncertainty of the dependence of Nd on CCN in GCMs \cite{Ghan2016, Quaas2009}.

### 3.3. The Prediction of Nd by MERRA2 Aerosol in Nonstratocumulus Regimes

As we noted earlier, the retrieval of Nd by MODIS is dependent on several assumptions, one of which is that the cloud liquid water content is adiabatic. Clouds can be subadiabatic if they are precipitating or are entraining heavily, or are partially frozen (for instance, a precipitating ice layer beneath the liquid cloud top detected by MODIS). However, we argue that in these regions the sensitivity of Nd to aerosol mass concentration should be similar to that of the stratocumulus regions. The degree to which a cloud is subadiabatic may be roughly inferred using the difference in the constant term of the regression model in that region relative to equation (4), which assumes an adiabatic cloud. Taking the log of equation (1) and substituting $\Gamma_{eff} = f_{ad}\Gamma_{ad}$ yields the following relationship:

$$\log_{10}Nd_{subadiabatic} = \frac{1}{2}\log_{10}f_{ad} + \log_{10}Nd$$  \hspace{1cm} (6)

Of course, this is a highly approximate method of examining the degree of adiabaticity of a cloud. It does not account for other factors that might bias the retrieval of Nd such as cloud top heterogeneity and sub-pixel variability. This approach also assumes that enhancement of aerosol does not alter cloud adiabaticity. In theory, increased CCN should suppress precipitation and increase adiabaticity. However, depending on cloud regimes, enhanced aerosol may either enhance or suppress entrainment and precipitation \cite{Berner2015}. Because the exclusion of this mechanism does not seem to introduce a systematic bias for simplicity, this effect is neglected in our analysis. In this paper we will refer to this factor ($f_{ad}$) naively as the subadiabaticity factor since it is the term that results from the derivation of Nd, although this factor could be more accurately described as a deviation from the ideal plane parallel, perfectly uniform, single-layer, adiabatic liquid cloud that the retrieval of Nd is predicated upon. It has been shown using simulated...
MODIS retrievals that subpixel cloud inhomogeneity may significantly bias the retrieval of effective radius [Marshak et al., 2006; Z. Zhang et al., 2016]. Occurrence of multilayer clouds has also been shown to bias retrieved Nd [Sourdeval et al., 2015; Sourdeval et al., 2016]. Differences in the droplet size distribution assumed by MODIS have also been shown to bias the retrieved effective radius [Zhang et al., 2012]. It is also possible that overlying layers of smoke and dust may bias retrieved cloud droplet sizes and optical depth [Haywood et al., 2004]. In addition to retrieval biases, physical differences in aerosol processing in different cloud regimes brought about by factors such as vertical motion and the vertical structure of droplet number concentration might also contribute to regional variation in this factor. This would be consistent with previous research contrasting different cloud regime sensitivities to aerosol [Gryspeerdt and Stier, 2012]. Overall, regional differences in this naïve adiabaticity factor should not alter the fairly consistent power law dependence of Nd on sulfate mass. Both assumed adiabaticity and retrieval errors related to subpixel heterogeneity and multilayer clouds should generally lead to an underestimate of Nd [Sourdeval et al., 2016; Z. Zhang et al., 2016].

While inferring the degree of adiabaticity in a given regime based on the difference in the constant term of the log-log fit is a very simplistic approach, it does appear that the sensitivity of Nd to SO$_4$ across regimes is roughly the same. This is shown in Figure 5. Equation (4) is used to remove variability due to SS, DU, and BC, leaving the sensitivity to SO$_4$ Log10Nd in the subtropical stratocumulus and the Southern Ocean all appear to have the same linear relation with log10SO$_4$, indicating that the Southern Ocean is relatively near to the adiabaticity of the stratocumulus regions. The most subadiabatic clouds appear to be continental. The degree of adiabaticity inferred by the intercept in equation (4) over the Amazon and continental Chinese stratus regimes is near 0.2. Again, it is possible that some of these differences between regimes are due to remote-sensing biases or differences in meteorology and may not be due to deviation from the adiabaticity in the stratocumulus regions. However, it is interesting to note that if we naïvely assume that the offset is purely due to adiabaticity the inferred range of adiabaticity presented in the inset of Figure 5 is within the range of ground-based observations of adiabaticity factor [Merk et al., 2016]. We present Figure 5 to support the notion that the sensitivity of Nd to SO$_4$ is similar across a variety of regions.

One particularly interesting feature is the contrast in the degree of inferred adiabaticity between the northern and southern midlatitudes (regions 7–12 in Figure 1). The northern midlatitude regions appear to be significantly less adiabatic than the southern midlatitudes. One possible explanation of the difference between
northern and southern midlatitudes might be that the latter is more dominated by supercooled liquid clouds, as opposed to ice clouds. The Northern Hemisphere has a much greater availability of ice nuclei (IN), allowing clouds to glaciate nearer 0°C [Atkinson et al., 2013; Hu et al., 2010; Tan et al., 2014], while the Southern Hemisphere has a much smaller availability of IN and more persistent supercooled liquid cloud cover [Chubb et al., 2013; Kanitz et al., 2011; Morrison et al., 2010].

Comparison of MODIS observations to in situ observations offers a further piece of evidence supporting relatively constant global sensitivity of Nd to SO4. While Nd observed by MODIS near Peru and near China are very similar to each other, airborne observations show that Nd near China should be almost twice as large [Allen et al., 2011; Koike et al., 2012]. To determine how much the MODIS observations would need to be rescaled to match in situ observations, the MODIS Nd was subset to approximately match the regions and seasons of the in situ observations from A-Force and (VOCALS) VAMOS Ocean-Cloud-Atmosphere-Land Study/VAMOS (Variability of the American Monsoon Systems) (April 25°–35°N, 125°–135°E and October–November 15°–25°S, 85°–70°W, respectively) [Allen et al., 2011; Koike et al., 2012]. The ratio of near China to near Peru MODIS Nd in these subsets was 1.03. The ratio implied by in situ observations was 1.86 [Allen et al., 2011; Koike et al., 2012]. All Nd observed near China by MODIS was rescaled by this factor, and it was found that the relation between Nd and SO4 in the stratocumulus regions and downwind from China all fell on the same curve (Figure 6).

Again, this analysis is presented to support the idea that the difference in the constant term of the log-log relationship between Nd and SO4 among regimes is due to limitations in the remote sensing of Nd. It is not possible to tell whether the apparent underestimation of Nd near China is due to subpixel heterogeneity, multilayer cloud cover, or subadiabatic cloud cover. Based on in situ measurement it does not appear to be due to intrinsic differences in the relation between Nd and sulfate in these regimes.

### 3.4. Evaluation of Nd as Modeled by MERRA2 Aerosols

We will now evaluate the ability of MERRA2 aerosol to reproduce the observed record of Nd from MODIS in the regions examined in this study. Equation (4) will be used in combination with MERRA2 aerosol mass concentrations. As discussed in the previous section, nonstratocumulus regimes are likely not to be adiabatic. We hypothesize that this manifests itself as a constant offset in the log-log fit. We do not consider this difference in our evaluation of the ability of MERRA2 to reproduce MODIS Nd. This constant offset in some regions does not affect the correlation coefficient. It should be noted that the regression model in equation (4) is trained on the aggregated data from the stratocumulus regions. Because of this we are not really evaluating the regression model in the stratocumulus regions because the regression model is not being confronted with entirely new data. Calculation of Nd outside of the stratocumulus regions is an evaluation of the regression model because the model was not trained on those regions.

First, the ability of MERRA2 and equation (4) to explain the total variance in Nd at 1° × 1°-monthly resolution is evaluated in each region, as quantified by the correlation coefficient (r). This is shown for each region in Figure 7. Correlations between all available 1° × 1°-monthly observations and predictions of Nd are positive and can be quite high in some regions.

Next, we evaluate the ability of MERRA2 data and equation (4) to match the observed seasonal cycle. The mean Nd in each region for each month is calculated, and the correlation between the time series of

---

**Figure 6.** As in Figure 5 but the Nd measured by MODIS over the ocean near China has been rescaled relative to data near Peru using in situ data collected by Koike et al. [2012] and Allen et al. [2011]. The solid line is the regression model from equation (4) (note that equation (4) is only trained on Sc regions and is not trained on the data from the Chinese maritime clouds). Data are shown as a linear plot rather than log-log space and that variability due to SS, DU, and BC has been removed using equation (4).
Figure 7. The time series of Nd in each of the regions shown in Figure 1. Each plot shows the observed Nd from MODIS (black) and the modeled Nd predicted using equation (4) and MERRA2 data (red). Two correlation diagnostics are given in the title of each plot. How well does the regression model recreate the seasonal cycle and year-to-year variability over the whole region? To answer this, the correlation between the time series of monthly values averaged over each box is given. This is written as $r_{ts}$. How well does the regression model capture all the variability in Nd that we see in each region over the data record? To answer this, the correlation is calculated using the predicted Nd and every available observation at monthly 1 × 1° resolution in the period from 2001 to 2013. This is written as $r$. 


MCCOY ET AL. MODIS-MERRA INDIRECT EFFECT 10
observed and predicted Nd is calculated. The time series are shown in Figure 7. It is important to note that the correlation between observed and modeled Nd can be quite high, but the bias can also be large. Unfortunately, at this time it is difficult to evaluate what combination of effects of subadiabaticity, heterogeneity, and multilayer clouds may be creating the bias between observed Nd and the Nd calculated by the regression model in each region. Based on the evaluation presented in Figure 6 it is reasonable to suppose that this constant bias is related to remote-sensing retrieval errors.

Finally, the ability of the MERRA2 aerosol reanalysis and equation (4) to reproduce the spatial pattern of Nd in a given month is evaluated. For each month from 2001 to 2013 the spatial correlation between observed and predicted Nd is calculated. This is shown in Figure 8. Spatial correlations tend to be positive in regions where the sources of CCN are anthropogenic point sources, as opposed to diffuse natural sources. Contrasting the ability of the regression model to reproduce a region’s seasonal variability and its ability to reproduce spatial variability reinforces this. For pristine regions in the Southern Hemisphere midlatitudes the seasonal correlation is quite high, but the spatial correlation can be quite low. This is consistent with most of the sulfate in these regions originating from phytoplanktonic DMS, which has a relatively poorly constrained spatial structure, but a well-known seasonal cycle that is driven by insolation [Ayers and Gras, 1991; Lana et al., 2012; Lana et al., 2011; McCoy et al., 2015; Meskhidze and Nenes, 2006, 2010; Vallina and Simó, 2007; Vallina et al., 2006].

Overall, it appears that the MERRA2 aerosol reanalysis, combined with equation (4), can reproduce a significant portion of the variability in Nd as observed by MODIS from 2001 to 2013. It is worth repeating that, while MERRA2 is a reanalysis product and is nudged by MODIS aerosol optical depth, it does not ingest MODIS cloud properties into its algorithm. This makes MERRA2’s aerosol reanalysis independent from MODIS Nd.

3.5. Estimates of the Change in Nd Since the Industrial Era

As noted above, ERFaci represents a major source of uncertainty in the anthropogenic radiative forcing [Boucher et al., 2014; Storelmo et al., 2009; Zelinka et al., 2014]. Individual model estimates of ERFaci in AR5 range from ~0.5 to ~2.5 W m\(^{-2}\) [Boucher et al., 2014; Zelinka et al., 2014]. ERFaci is contributed to by both changes in cloud microphysics and macrophysics. In this work we only consider changes in cloud microphysics. Based on the robust dependence of Nd on sulfate shown in the current climate we can estimate how much Nd has changed since the preindustrial era.

To calculate the change in Nd since industrialization, we utilize the sulfate surface mass concentration from the preindustrial and present-day harmonized emission simulations from AeroCom PRE and B [Dentener et al., 2006; Kinne et al., 2006; Schulz et al., 2006]. Only three models provide sulfate mass concentration near the surface: LOA (GCM LMDzT 3.3), LSCE (GCM LMDzT 3.3-INCA), and Goddard Institute for Space Studies (GISS) (GCM model E). LOA and LSCE models share the same core model (LMDzT 3.3), but their aerosol microphysics schemes are different [Textor et al., 2006]. An important caveat of this method is that by using these harmonized emission simulations the uncertainty range in change in Nd between preindustrial (PI) and present day (PD) does not relate to uncertainty in sulfate sources between this time period and only considers uncertainty in aerosol processing by the AeroCom models. The hope is that the uncertainty in sulfate...
emissions between PI and PD is small [Smith et al., 2011] relative to the uncertainty in aerosol processing and the sensitivity of Nd to SO$_4$$_2$. The change in sulfate mass concentration since 1750 A.D. provided by these models is combined with the sensitivity of Nd to sulfate mass concentration estimated in this study (Figure 6). The ratio of PD to PI Nd is calculated as

$$\frac{Nd_{PD}}{Nd_{PI}} = \left( \frac{SO_{4, PD}}{SO_{4, PI}} \right)^{a_1}$$

where $a_1$ is the coefficient relating log$_{10}$(SO$_4$$_2$) to log$(10Nd)$ from equation (4). We may use this to calculate the fractional change in Nd between preindustrial (PI) and present day (PD) by assuming that the relation between SO$_4$ and Nd that we calculate over the stratocumulus regions represents the sensitivity everywhere. In defense of this assumption, the relationship estimated from the stratocumulus regions appears to hold in the midlatitudes and near China (Figures 5 and 6). However, in situ studies indicate a stronger sensitivity to sulfate over land [Lowenthal et al., 2004], potentially making the change in Nd between PI and PD estimated by our study conservative in these areas. In keeping with the results from in situ studies, it appears that the sensitivity of Nd to SO$_4$$_2$ is slightly higher in the continental regions examined in this study (Figure 5). On the other hand, this may also be due to increasing CCN suppressing precipitation and moving the clouds toward adiabaticity.

The ratio of PD to PI Nd calculated using equation (6) is shown in Figure 9. Nd is inferred to have tripled over highly polluted regions during the industrial era, while it remains effectively unchanged throughout most of the Southern Hemisphere. This is consistent with the predominance of biogenic sources of sulfate in this region (Figure 10) [Hamilton et al., 2014].

Figure 9. The ratio of PD to PI Nd contingent on the change in sulfate mass concentration estimated from the multimodel mean of AeroCom PRE and B simulations from GISS, LOA, and LSCE.

Figure 10. The correlation between anomalies relative to the seasonal climatology in MSA and SO$_4$$_2$ from MERRA 2001–2013. This gives a rough measure of the predominance of biogenic sulfate sources.
We note that Penner et al. [2011] showed that using the observed dependence of cloud properties on aerosol optical depth or index to infer aerosol-cloud interactions (ACI) results in an underestimate of ACI during the industrial era. However, we argue that this is not applicable to our study. Our study utilizes reanalysis that provides speculation of aerosol, making it insensitive to differences in aerosol speculation in the PI and PD. Further, the sensitivity of Nd to SO4 appears to hold in both the polluted Northern Hemisphere (NH) and the pristine Southern Ocean (when covariation with other species in polluted regions is accounted for). We argue that the Southern Ocean serves as an analogy for the PI atmosphere. If our result holds in a pristine region, it seems reasonable to suppose that it would hold in an entirely pristine atmosphere.

The change in Nd since the PI era can be quite significant in polluted areas. It is useful to put this change in Nd in the context of the change in upwelling SW. The ACI radiative forcing due to a change in Nd is modulated by the cloud area and cloud LWP. In order to put our inferred change in Nd in the context of ACI radiative forcing, we offer a simple estimate of the change in SW due to ACI [see Charlson et al., 1992; Meskhidze and Nenes, 2006]. The ACI radiative forcing is calculated as

$$\Delta SW_{ACI} = -\frac{1}{3} SW_{\text{Liquid}} \cdot CF_{\text{Liquid}} \cdot (1 - a_{\text{Liquid}}) \left[ 1 - \frac{Nd_{\text{PI}}}{Nd_{\text{PD}}} \right]$$

(7)

where CF is the liquid cloud fraction detected by MODIS, $a_{\text{Liquid}}$ is the albedo of liquid clouds observed by MODIS approximated as $a_{\text{Liquid}} = 0.98 [\text{Lacis and Hansen, 1974}]$, and the downwelling SW is the clear-sky downwelling SW at the surface from Clouds and the Earth’s Radiant Energy System (CERES) [Kato et al., 2013]. This calculation is conceptually similar to Storelamo et al. [2009], which showed that when different observational parameterizations of Nd as a function of aerosol amount were implemented in a single GCM the aerosol indirect effect forcing was between $-0.62$ and $-1.94$ W m$^{-2}$. The calculation presented here is much more idealized and is only intended to show an approximate range consistent with the regression model shown in equation (4). A few potential caveats to these calculations must be noted. First, it is unclear to what extent overlying cloud cover would damp-out increases in upwelling SW due to the brightening of underlying clouds. Because of this uncertainty the radiative forcing due to the Twomey effect is calculated using the liquid-topped cloud fraction. This is likely to produce a conservative estimate of the change in upwelling SW because low liquid clouds often exist under upper level ice clouds, and these clouds are neglected by only considering regions with liquid cloud tops (as detected by MODIS) [Li et al., 2015; McCoy et al., 2014]. Second, because of attenuation by water vapor, the downwelling SW at the surface is smaller than the downwelling SW at the cloud top. Both of these approximations tend to make the change in SW due to ACI calculated here conservative. Finally, the cloud macrophysical properties are assumed not to have changed since the preindustrial era.

Assuming that cloud macrophysical properties are unchanged since the PI era represents a major caveat in our calculations. It is likely that clouds have changed since the PI. Changes in cloud cover have been observed within the satellite era alone [Bender et al., 2011; Norris et al., 2016; Seethala et al., 2015]. Further, changes in aerosol and changes in cloud macrophysics may be related [Albrecht, 1989]. It remains unclear if neglecting indirect effects beyond the first indirect effect would tend to decrease or increase the radiative forcing [Gryspeerdt et al., 2016; Stevens and Feingold, 2009; Wood, 2007]. Comparison of monthly and daily anomalies in cloud and scene albedo indicates that on longer time scales the brightening due to the first indirect effect does not dominate the cloud albedo [Bender et al., 2016]. Overall, it remains unclear what the precise effect of the first indirect effect on the radiative forcing is on longer time scales. It is entirely possible that compensating feedbacks negate it entirely. The calculation of change in upwelling SW due to changes in Nd due to sulfate aerosol is presented here to put these relations in the context of the energy budget. It is important to note that by using LWP and CF from observations this calculation integrates interactions between cloud macrophysics and microphysics into the calculation of ACI [McComiskey and Feingold, 2012]. If a true ERFaci was to be calculated, the interactions of cloud microphysics and macrophysics would need to be corrected for. Overall, the purpose of this calculation is to contextualize where changes in Nd have the most effect, for instance, in regions with relatively thin cloud cover or weak downwelling, SW large changes in Nd would have fairly little impact on albedo.

Keeping the caveats stated above in mind, the inferred change in radiative forcing due to the first indirect effect during the industrial era is shown in Figure 11. One of the interesting features to note in comparing Figures 9–11 is how a small fractional change in Nd in the stratocumulus regions translates to a significant change in reflected SW. Globally, if the sensitivity of Nd to SO4 in equation (4) is the best estimate of the
The relation between SO$_4$ and Nd, the global mean change in SW ranges from $-0.84$ W m$^{-2}$ to $-0.92$ W m$^{-2}$ as a function of the varying estimates of PI and PD SO$_4$. To evaluate the impact of sampling uncertainty, the change in SW is calculated again using the sensitivity of Nd to SO$_4$ estimated in each stratocumulus region (Figure 4, empty symbols for MERRA2). Combined with the three estimates of change in SO$_4$ between the PI and PD, this yields 15 estimates of the change in SW. The change in SW ranges from $-1.3$ W m$^{-2}$ to $-0.61$ W m$^{-2}$, and the mean change in SW is $-0.97$ W m$^{-2}$ with a standard deviation of 0.23 W m$^{-2}$. This is well within the range of model-estimated ERFaci ($-0.3$–$0.75$ W m$^{-2}$) [Boucher et al., 2014; Zelinka et al., 2014]. It is also interesting to note that this change in SW overlaps the lower bound on ERFaci of $-0.75$ W m$^{-2}$ that Stevens [2015] calculates to be consistent with pre-1950 instrumental records. However, the estimate of the change in SW made by this study is among the stronger estimates made by previous evaluations of the first indirect effect. Empirical studies of cloud-aerosol interactions using CERES and MODIS observations conducted by Quaas et al. [2008] estimated a contribution of $-0.2 \pm 0.1$ W m$^{-2}$ by the cloud albedo effect. Lebsock et al. [2008] estimated a somewhat larger first indirect effect of $-0.42$ W m$^{-2}$ utilizing collocated data from the CERES, Cloudsat, MODIS, and AMSR-E instruments aboard the A-Train constellation. Bellouin et al. [2013] utilized (MACC) monitoring atmospheric composition and climate reanalysis AOD coupled with CERES and MODIS observations to estimate that the first indirect effect was between $-0.1$ and $-2.0$ W m$^{-2}$ but with a nonsymmetric distribution leading to a best estimate of $-0.6 \pm 0.4$ W m$^{-2}$. The estimate of the strength of the first indirect effect in this study compared to GCMs-based estimates seems to also be on the strong side. Storelvmo et al. [2009] found aerosol indirect effects ranging between $-0.62$ and $-1.94$ W m$^{-2}$, depending on what observationally derived parameterization was implemented in the GCM evaluated in their study. When only sulfate was perturbed in the GCMs studied by Zelinka et al. [2014], estimates of ERFaci ranged from $-0.18$ to $-1.05$ W m$^{-2}$.

In summary, we have examined the dependence of Nd on SO$_4$ in the current climate, accounting for covariability with other aerosol species, and we find a consistent dependence across polluted and pristine regions, as well as midlatitude and subtropical regimes. The sensitivity of Nd to sulfate diagnosed by this study is smaller than the dependence diagnosed by previous in situ studies (Figure 3). Estimates of the change in sulfate between the PI and PD combined with the dependence of Nd on sulfate indicate that Nd may have increased more than threefold in some regions in the NH. The magnitude of radiative forcing due to the changing Nd while holding cloud macrophysical properties constant at present-day values is estimated to be $-0.97 \pm 0.23$ W m$^{-2}$. Again, this is in the absence of any cancellation due to changes in cloud macrophysical properties. Radiative forcing due only to the Twomey effect is provided to indicate the importance of understanding to what degree other indirect or semidirect effects might either amplify or damp the Twomey effect.
4. Conclusions

In this work we have presented a top-down observational estimate of the dependence of Nd on aerosol mass concentration from MERRA2. We summarize the central points of this paper as follows:

1. The dependence of Nd on sulfate seems to be fairly invariant across geographical regions and between pristine and polluted regions (Figures 2, 5, and 7). An important caveat to this result is that the Nd we consider has been restricted to overcast conditions, and this result may not apply in situations with more broken cloud cover.

2. The sensitivity of Nd to SO4 yields by this study is on the lower end of previous in situ studies but agrees with satellite studies over the remote Southern Ocean (Figure 3). It is possible that some of this difference is due to differences in the spatial and temporal aggregation of these data sets [Grandey and Stier, 2010; McComiskey and Feingold, 2012].

3. Utilizing state-of-the-art aerosol reanalysis from MERRA2 yields a very robust relationship between near-surface SO4 and Nd. Over 40% of the variability in MODIS-observed Nd over the stratocumulus regimes may be explained by MERRA2 aerosol mass concentrations. Given the uncertainties in both observations of Nd and aerosol reanalysis, this is a significant amount of variance to explain.

4. Climatological aerosol fields from AeroCom models also appear to be able to replicate the dependence of Nd on sulfate (Figure 4). Evidently, GCM aerosol mass concentrations posses the ability to represent the first indirect effect. It is important to keep in mind that even if GCMs perfectly reproduce the dependence of Nd on aerosol mass concentration in the current climate, they may still fail to properly represent buffering in the system due to aerosol-induced variation in cloud macrophysics or radiation [Stevens and Feingold, 2009].

It appears that anthropogenic aerosol substantially affects cloud microphysics on a monthly and regional time scale through the Twomey effect. On its own, this effect would have the ability to obscure a significant fraction of the CO2-induced warming signal in the industrial era. It is unclear whether indirect effects beyond the Twomey effect would amplify or damp its radiative effects. Studies of aerosol indirect effects do not have a strong consensus on whether other aerosol-cloud interactions damp or enhance its radiative effects [Albrecht, 1989; Bender et al., 2016; Berner et al., 2015; Gysel and Feingold, 2016; Stevens and Feingold, 2009]. The strength of the Twomey effect diagnosed by this study indicates the importance of investigation of cloud-aerosol indirect effects to constrain the overall cloud brightening during the industrial era.

Acknowledgments

The authors wish to thank Steve Ghan for editing the manuscript and the anonymous reviewers for their time and efforts improving this article. D.T.M. and D.L.H. were supported under NASA grant NNX14AG26G. The data used are listed in the references. All data used in this study are freely available online. The MODIS data are available from modis.gsfc.nasa.gov. Aerosol model data are available from the AeroCom project upon request. MERRA2 data are available from gmao.gsfc.nasa.gov/reanalysis/MERRA-2. CERES data are available from ceres.larc.nasa.gov/order_data.php.

References


Lana, A., R. Simo, S. M. Vallina, and J. Dachs (2012), Potential for a biogenic in...


