

## Satellite-derived estimates of ultrafine particle concentrations over eastern North America

Paola Crippa,<sup>1</sup> Dominick Spracklen,<sup>2</sup> and S. C. Pryor<sup>3</sup>

Received 25 February 2013; revised 30 July 2013; accepted 2 August 2013; published 9 September 2013.

[1] High concentrations of ultrafine particles (UFP, i.e., particles with diameter  $< 100$  nm) impact both human health and Earth's climate. Recent innovations in remote sensing technologies and data retrievals offer the potential for predicting UFP concentrations based on data from satellite-borne instrumentation. Herein we present a physically based statistical algorithm to estimate UFP concentrations across eastern North America using remotely sensed aerosol optical depth, Ångström exponent, ultraviolet solar radiation flux, and ammonia and sulfur dioxide concentrations. The proposed algorithm is built and independently evaluated using an array of in situ observations. The algorithm is able to capture up to 60% of the variability in daily measured UFP number concentrations at a regionally representative reference site and is thus applied to generate seasonal UFP concentration estimates across eastern North America. The resulting UFP concentrations are cross-evaluated with simulations from a global aerosol microphysics model. There is a negative bias in the model output relative to the satellite-driven proxy, which is largest (up to 76%) in summer and may be due to overestimation of UFP from the satellite-based algorithm derived herein, due to the higher availability of remote sensing data in clear-sky conditions or uncertainty in the model simulation of new particle formation. Nevertheless, the model and algorithm indicate similar spatial and seasonal variability (spatial correlation coefficients of 0.10 to 0.56), indicating the value of the satellite-based UFP proxy in global and regional model evaluation exercises and in efforts to identify regions where future in situ data collection should be prioritized.

**Citation:** Crippa, P., D. Spracklen, and S. C. Pryor (2013), Satellite-derived estimates of ultrafine particle concentrations over eastern North America, *J. Geophys. Res. Atmos.*, 118, 9968–9981, doi:10.1002/jgrd.50707.

### 1. Introduction and Objectives

[2] High ultrafine aerosol particle (UFP) concentrations are associated with increased human morbidity and mortality [e.g., *Knol et al.*, 2009] and contribute to indirect climate forcing [e.g., *Bauer and Menon*, 2012]. UFP (defined here as particles having diameters ( $D_p$ )  $< 100$  nm) have been observed at very high concentrations across an array of data sets collected at or near the Earth's surface and in more limited data sets collected in the free troposphere [e.g., *Kulmala et al.*, 2004]. These particles may play a critical role in determining the abundance of cloud condensation nuclei and hence indirect climate forcing [*Kuang et al.*, 2009; *Spracklen et al.*, 2008] and ultimately contribute to the concentration

of particles responsible for direct climate forcing (via enhanced backscattering) [*Spracklen et al.*, 2006]. Hence, accurate temporally and spatially resolved UFP concentrations are needed at the global scale to enhance understanding of the physiochemical controls on their concentrations and the climate impacts of aerosol particles and ultimately to design effective policies to regulate climate forcing and protect human health. However, the relative paucity of surface observing networks prohibits quantification of the aerosol particle climate forcing [*Kulmala et al.*, 2011] and severely constrains our ability to evaluate numerical models [*Spracklen et al.*, 2010]. Thus, there is a clear need for new approaches to quantify particle size distributions (PSDs) (including the abundance of UFP) at both the regional and global scales.

[3] Recent technological advances in satellite-borne instrumentation and data retrieval protocols mean satellite-derived aerosol particle and gas concentrations now represent a unique and mature tool for investigating the spatial and temporal variabilities of aerosol particle concentrations with almost global, homogeneous, and near-real-time spatial coverage. Prior studies have reported analyses of satellite-derived aerosol particle properties in terms of their ability to represent parameters regulated by national and international air quality standards such as the mass concentration of  $PM_{2.5}$  (i.e., the concentration of particles  $D_p < 2.5$   $\mu m$ ) [e.g., *Gupta and Christopher*, 2008; *van Donkelaar et al.*, 2010; *Koelemeijer*

Additional supporting information may be found in the online version of this article.

<sup>1</sup>Atmospheric Science Program, Department of Geological Sciences, Indiana University, Bloomington, Indiana, USA.

<sup>2</sup>School of Earth and Environment, University of Leeds, Leeds, UK.

<sup>3</sup>Atmospheric Science Program/Department of Geological Sciences, Indiana University, Bloomington, Indiana, USA.

Corresponding author: P. Crippa, Atmospheric Science Program, Department of Geological Sciences, Indiana University, Bloomington, IN 47405, USA. (pcrippa@indiana.edu)

©2013. American Geophysical Union. All Rights Reserved.  
2169-897X/13/10.1002/jgrd.50707

*et al.*, 2006] and to undertake analyses of local versus distant sources of those aerosol particles [e.g., *Yu et al.*, 2012]. These studies exploit the correlation between aerosol optical depth (AOD) (i.e., the integral of the extinction coefficient due to light scattering and absorption, which scales with particle cross-sectional area) and the particle volume (and mass), which is maximized for particle diameters of  $\sim 1 \mu\text{m}$ . However, UFP are by definition not of a diameter that strongly interacts with radiation in the visible and infrared wavelengths and thus cannot be directly observed by instruments currently deployed on satellites. The current research is motivated by a recent first attempt to estimate nucleation mode particle concentrations (defined as  $D_p < 25 \text{ nm}$ ) over the entire globe using a proxy algorithm based on satellite data [*Kulmala et al.*, 2011]. The work by *Kulmala et al.* [2011] was highly innovative but exhibits several limitations such as the absence of a driving gas-phase mechanism and incomplete evaluation of the proxy and related assumptions (e.g., in the description of the nucleation rate and in the approximation of the condensational sink (CS)). Herein we propose and evaluate a new algorithm to estimate UFP number concentrations using data from satellite-borne radiometers. Our specific research objectives are to

[4] 1. Formulate a new proxy algorithm for UFP number concentrations based on satellite-derived properties that is more statistically robust and physically interpretable.

[5] 2. Evaluate the new UFP proxy algorithm using in situ measurements.

[6] 3. Apply the new algorithm to analyze the seasonality and spatial variability of UFP concentrations over eastern North America and cross-evaluate those results with output from a detailed global climate-chemistry model.

## 2. Data Used: Particle Size Distribution Measurements, Satellite Retrievals, and Model Simulations

### 2.1. Ground-Based PSD Measurements

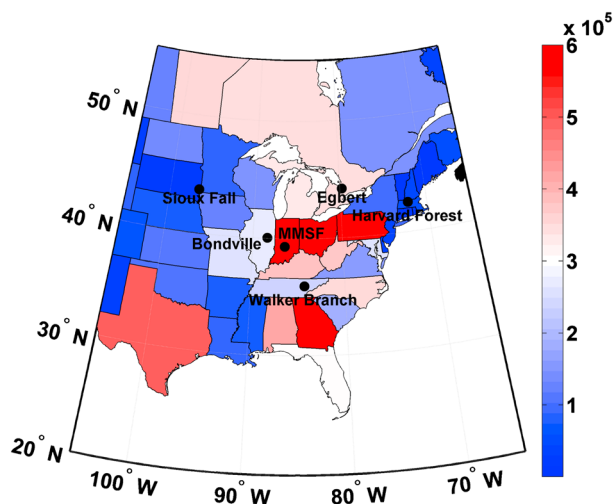
[7] The algorithm used to describe and predict UFP number concentrations is developed and evaluated primarily using ground-based particle size distribution (PSD) measurements collected above the canopy of a deciduous forest in southern Indiana (at Morgan Monroe State Forest (MMSF),  $39.317^\circ\text{N}$ ,  $86.417^\circ\text{W}$ ) and at the Center for Atmospheric Research Experiments in Egbert, Canada ( $44.23^\circ\text{N}$ ,  $79.78^\circ\text{W}$ ) (Figure 1). We focus most of the model development and evaluation on data from MMSF because that site is more regionally representative than Egbert and thus new particle formation (NPF) and the total UFP number concentrations are less affected by local emissions [*Crippa and Pryor*, 2013]. Additionally, the instrumentation deployed at MMSF has a lower size cut-point for particle detection and thus better characterizes the products of recent NPF and the total UFP burden. PSD measurements at MMSF derive from a Scanning Mobility Particle Sizer (SMPS) system (TSI SMPS3936: Electrostatic Classifier (EC) model 3080, nano-DMA model 3085, and a Condensation Particle Counter (CPC) model 3786). This system measures in the diameter range  $D_p = 3.22\text{--}105.5 \text{ nm}$  in 80 logarithmically spaced size channels. Data are also presented from a Fast Mobility Particle Sizer (TSI FMPS 3091) operated at MMSF

which measures PSD in 32 logarithmically spaced size channels  $D_p = 6.04\text{--}523.3 \text{ nm}$  (for further details of the site and instrumentation, see *Pryor et al.* [2010]). PSD from Egbert were also measured using a TSI SMPS (EC 3071 and CPC 3010) and cover the diameter range  $D_p = 10.75\text{--}392.42 \text{ nm}$  in 26 logarithmically spaced size channels (for further details of the site and instrumentation, see *Riipinen et al.* [2011]). Simultaneous measurements collected at the two sites during May 2007 to May 2008 are used for the UFP proxy development, and an independent evaluation of the resulting UFP proxy algorithm is undertaken using observations at MMSF during December 2006 to April 2007 and June 2008 to August 2009.

[8] The lack of other comparable data sets to those from Egbert and MMSF (i.e., ground-based ultrafine PSD measurements) over eastern North America limits our ability to undertake a rigorous evaluation of the proxy; thus, we also employ data collected during 2006–2008 at five rural Aerosol Robotic Network (AERONET) sites (Figure 1) and output from a global model. The AERONET Sun-scanning spectral radiometers measure direct Sun and diffuse-sky radiances at eight wavelengths between 340 and 1020 nm (usually 340, 380, 440, 500, 670, 870, 940, and 1020 nm) at known angular distances from the Sun. After applying an inversion algorithm to these data, the columnar values of PSD can be derived and discretized into 22 logarithmically spaced size bins ( $D_p = 100 \text{ nm}$  to  $30 \mu\text{m}$ ) [*Holben et al.*, 1998]. The accuracy of the retrieval algorithm for the AERONET volume size distribution has been assessed in several theoretical studies [*Dubovik et al.*, 2002; *Dubovik et al.*, 2000; *Dubovik and King*, 2000; *O'Neill et al.*, 2001a; *O'Neill et al.*, 2001b]. The average retrieval error for the volume size distribution in the size range  $0.1 \mu\text{m} \leq r \leq 7 \mu\text{m}$  under cloud-free conditions is estimated to be  $\sim 15\%$ . However, the accuracy of the retrieval depends on many factors including particle size, and for particles smaller than  $0.1 \mu\text{m}$  and larger than  $7 \mu\text{m}$  it is estimated to vary between 15 and 100% for water-soluble aerosols because of the small volumes usually associated with those particles [*Dubovik et al.*, 2000; *Dubovik and King*, 2000]. More recently, several experimental studies have shown the applicability of AERONET-retrieved variables for providing size-resolved characteristics of particles in terms of composition, absorption, and scattering properties [*Costabile et al.*, 2013; *Gobbi et al.*, 2007; *Russell et al.*, 2010; *Sinyuk et al.*, 2012]. Despite the uncertainty of the retrieved PSD in the smallest diameters and in the absence of other sources of measurements, in this analysis we use the total number concentration of aerosol particles summed over the three smallest diameter sections (i.e.,  $D_p = 100\text{--}200 \text{ nm}$ ) retrieved from the AERONET radiometers to assess the dependence of fine particle concentrations as derived from AERONET measurements and the satellite-derived estimates using the UFP proxy.

### 2.2. Satellite-Derived Observations

[9] In this study we focus on number concentrations of particles with  $D_p < 100 \text{ nm}$ . The concentrations of these UFP are, in non-urban environments, highly dependent on NPF [*Pirjola et al.*, 2004] which, in turn, is causally linked to the availability of semivolatile and low volatility vapors and competition for the vapors between condensation to



**Figure 1.** Location of the stations from which in situ or AERONET observations are analyzed (shown by the dots). The shading denotes annual  $\text{SO}_2$  emissions (tons) by state or Canadian province, averaged over the years 2006–2009. US data were obtained from the EPA Air Markets Program Data (AMPD) website (U.S. EPA, Air Markets Program Data (AMPD), 2013, <http://ampd.epa.gov/ampd/>), whereas Canadian data are from the National Pollutant Release Inventory (NPRI) website (National Pollutant Release Inventory (NPRI), 2013, [http://www.ec.gc.ca/inrp-npri/default.asp?lang=En&n=0EC58-C98-1#Emission\\_Summaries](http://www.ec.gc.ca/inrp-npri/default.asp?lang=En&n=0EC58-C98-1#Emission_Summaries)).

existing aerosol particles (often described in terms of the CS) versus development of critical clusters that can undergo a phase change and generate new aerosol particles. In many environments over eastern North America, ammonium sulfate significantly contributes to the composition of freshly nucleated particles, which indicates that ternary nucleation may be responsible for NPF, and thus ammonia ( $\text{NH}_3$ ), sulfuric acid ( $\text{H}_2\text{SO}_4$ ), and water may play a role in forming the critical clusters [Jung *et al.*, 2010; Pryor *et al.*, 2011; Smith *et al.*, 2005].  $\text{H}_2\text{SO}_4$  is largely produced from sulfur dioxide ( $\text{SO}_2$ ) via photochemical reactions, and CS is a product of the in situ PSD which can be described using the AOD and Ångström exponent ( $\text{Å}$ ). Thus, the following parameters as measured from satellite-borne instrumentation were selected a priori for inclusion in the UFP proxy development: AOD and  $\text{Å}$ , ultraviolet (UV) solar radiation flux, and  $\text{NH}_3$  and  $\text{SO}_2$  concentrations. In the following subsections we describe the sources of these data and summarize evaluations relative to ground-based observations and model output.

[10] Prior to describing the remote sensing data products used herein it is important to mention that the accuracy of satellite retrievals is dependent on several factors such as the surface type over which measurements are taken and specifically accurate characterization of the surface reflectance, because the basic principle of aerosol particle and gases detection is the difference between the amount of light reflected by the surface and the compounds of interest. Oceans, which are detected by satellites as dark targets, allow for maximum contrast and thus most accurate data retrievals [e.g., Remer *et al.*, 2005], whereas land surfaces are brighter and more heterogeneous, thus making data retrieval more difficult. Nevertheless, as discussed below, the observations we use have

been subject to evaluation, and the results imply some degree of skill. For example, over the United States where the emissions are well characterized, Ozone Monitoring Instrument (OMI)-retrieved  $\text{SO}_2$  values exhibit high agreement with observations ( $r=0.94$ ) [Lee *et al.*, 2011].

### 2.2.1. Aerosol Optical Properties: Aerosol Optical Depth and Ångström Exponent

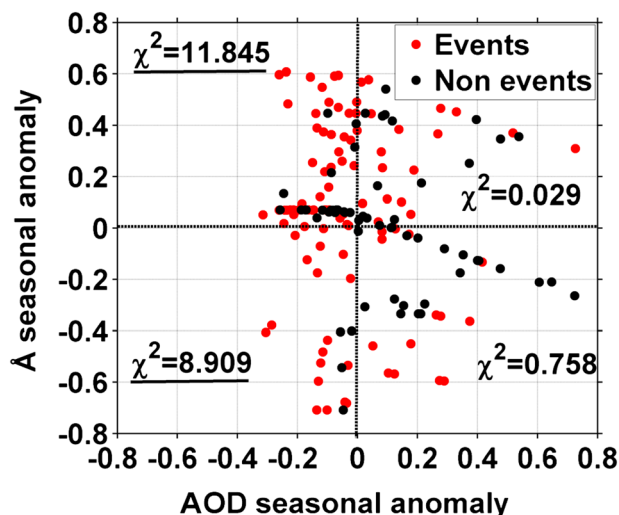
[11] The first attempt to quantify global nucleation mode aerosol concentrations from satellite-derived quantities used only AOD as a proxy of the CS [Kulmala *et al.*, 2011]. Here we extend that initial work by including the cross product of AOD and  $\text{Å}$  as measured by the Moderate Resolution Imaging Spectrometer (MODIS) on the Terra satellite as a proxy of the CS. AOD is a measure of the total columnar absorption and scattering of solar radiation and thus reflects the integrated aerosol burden, while  $\text{Å}$  expresses the spectral dependence of aerosol optical depth on the wavelength of the incident light; thus, high values of  $\text{Å}$  are associated with a larger abundance of small particles. We postulate that the combination of these two parameters allows for a more complete description of the CS and thus likelihood of NPF (see supporting information for detailed justification of use of both AOD and  $\text{Å}$ ).

[12] MODIS retrievals of AOD and  $\text{Å}$  have been extensively validated relative to ground-based observations over eastern North America [Engel-Cox *et al.*, 2004; Hyer *et al.*, 2011; Levy *et al.*, 2010]. In this work we use level-3 ( $1^\circ \times 1^\circ$ ) MODIS daily values of AOD (at  $\lambda=550$  nm) and  $\text{Å}$  (at 470–660 nm) as obtained from the GIOVANNI (Goddard Earth Sciences Data and Information Services Center Interactive Online Visualization and Analysis Infrastructure) website. The Terra satellite operates in a Sun-synchronous, near-polar, circular orbit with local equatorial crossing time in the descending mode of approximately 10:30 A.M. (LST). Thus, the resulting MODIS measurements are a good surrogate of conditions just prior to the time (10:00–12:00 LST) at which highest UFP concentrations are typically observed at MMSF [Pryor *et al.*, 2010].

[13] The frequency of NPF and high UFP concentrations at MMSF exhibit high seasonality, as do both AOD and  $\text{Å}$ . An overview of the relationship between MODIS-derived AOD and  $\text{Å}$  and the occurrence of NPF and elevated UFP concentrations is shown in Figure 2 using daily seasonal anomaly values. Negative AOD anomalies are representative of lower total aerosol burdens, while positive  $\text{Å}$  anomaly values indicate a higher fraction of fine mode particles. Thus, the four quadrants in Figure 2 can be treated as a contingency table and used to test the distribution of NPF events and non-event days (and thus high and low UFP concentrations) at MMSF (where the occurrence of NPF is defined as described in Crippa and Pryor [2013]) using a chi-square analysis ( $\chi^2$  test) as follows:

$$\chi^2 = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 90.42 \quad (1)$$

where  $O_i$  is the number of events (i.e., 17, 50, 18, 14), and  $E_i$  (i.e., 18, 21, 4, 19) is the number of non-events in each quadrant  $i$ . The critical value at 5% significance level is 7.815 for 3 degrees of freedom; thus, the null hypothesis of equal distribution of events among the four quadrants can be rejected. We also tested if the proportion of events and non-events is different in each quadrant. Quadrants associated with negative AOD



**Figure 2.** Seasonal anomaly of aerosol optical depth (AOD) and Ångström exponent ( $\text{\AA}$ ) values for NPF event and non-event days during 1 May 2007 to 31 May 2008 at MMSF ( $n=161$ ). The seasonal anomaly values of AOD and  $\text{\AA}$  are computed as a difference between each daily value and the seasonal mean.  $\chi^2$  values for the test of proportions of NPF event and non-event days in each quadrant are reported (underlined values show the quadrants in which the null hypothesis of equal distribution of event classification is rejected).

anomalies are characterized by a high occurrence of NPF events (and high UFP concentrations) implying low aerosol burdens (and thus CS) favors NPF (Figure 2). The Ångström exponent appears to exert a weaker control on nucleation occurrence, but given the proof offered in the supporting information and the analysis described in section 3.2, we explore the possibility of predicting UFP concentrations using both satellite products.

### 2.2.2. Trace Gases Concentrations

[14] Given the importance of  $\text{H}_2\text{SO}_4$  (and its precursor  $\text{SO}_2$ ) and  $\text{NH}_3$  to NPF over eastern North America, satellite retrievals of these gases and the UV flux were also included in the variable suite for construction of the proxy algorithm.

[15] Herein we use level-3 (quality assured) daily average column planetary boundary layer [ $\text{SO}_2$ ] in Dobson Units ( $1 \text{ DU} = 2.69 \times 10^{16}$  molecules  $\text{cm}^{-2}$ ) at a spatial resolution of  $0.25^\circ \times 25^\circ$  from the Ozone Monitoring Instrument (OMI) OMSO2e product on the GIOVANNI interface [Krotkov *et al.*, 2006; Lee *et al.*, 2009]. OMI is deployed onboard NASA's Aura satellite, and evaluation studies have shown that OMI-retrieved [ $\text{SO}_2$ ] correlate well with ground-based and aircraft measurements, as well as with simulated values over the United States, although the measurement uncertainty is largest when small spatial and temporal scales are considered [Lee *et al.*, 2009; Lee *et al.*, 2011].

[16] Ammonia concentrations are taken from the Tropospheric Emission Spectrometer (TES) observations. TES is a high-resolution ( $0.06 \text{ cm}^{-1}$ ) Fourier transform spectrometer that measures infrared radiation onboard the Aura satellite [Bowman *et al.*, 2006; Clough *et al.*, 2006; Shephard *et al.*, 2009]. Herein we use level-2  $\text{NH}_3$  representative volume mixing ratios at a horizontal resolution of  $5 \times 8 \text{ km}$  in the layer where TES exhibits highest sensitivity (700–900 hPa, i.e.,  $\sim 1\text{--}3 \text{ km}$  above sea level). The level of

detectability has been estimated to be approximately 0.4 ppbv for a profile with peak concentration of 1 ppbv [Shephard *et al.*, 2011]. Ammonia concentrations derived from TES have also been subject to evaluation at both the local and global scales relative to in situ measurements and model output [Pinder *et al.*, 2011; Shephard *et al.*, 2011]. The measurements are relatively sparsely distributed in both space and time; thus, in the following analyses the observations were used to calculate seasonal average [ $\text{NH}_3$ ] in each grid cell over eastern North America using an inverse squared distance weighted (i.e., ISDW) interpolation technique. The resulting seasonal pattern computed using data collected in May 2007 to May 2008, indicate peak  $\text{NH}_3$  concentrations during spring and summer seasons, which is consistent with other observations and simulations over eastern North America [Heald *et al.*, 2012; Mathur and Dennis, 2003] and measurements at MMSF [Pryor *et al.*, 2001]. The seasonal mean values of  $\text{NH}_3$  at 700–900 hPa from TES for the MMSF location in winter, spring, and summer are 0.1, 0.53, and 0.45 ppb, while in situ observations above the forest canopy (at 40 m) indicate similar seasonal variability but higher values of 0.3, 1.4 and 0.5 ppb, consistent with the importance of surface  $\text{NH}_3$  sources [Pryor *et al.*, 2001].

### 2.2.3. Solar Radiation

[17] Local noontime surface UV irradiance at 310 nm (UV-B) was selected for possible inclusion in the predictive algorithm for UFP concentrations because of the key role played by photochemistry in producing nucleation precursors [Kulmala, 2003; Zhang *et al.*, 2011]. The level-3 OMI local noontime surface UV irradiances at a spatial resolution of  $1^\circ \times 1^\circ$  were evaluated relative to data from a Kipp and Zonen UV-S-A-T radiometer deployed at MMSF which measures in the wavelength range 280–400 nm (and thus includes UV-A, UV-B, and part of UV-C), and the results indicated a strong positive association. For a regression equation with zero intercept,  $\text{UV}_{\text{OMI}} = 1.565 \times \text{UV}_{\text{meas}}$ , with  $r^2 = 0.732$  for a sample size ( $n$ ) of 240.

### 2.3. GLOMAP Simulations

[18] In this work we evaluate the satellite-derived estimates of UFP concentrations relative to in situ observations and also cross-evaluate the seasonal patterns of UFP concentrations derived from our proxy based on satellite observations with simulations conducted with the Global Model of Aerosol Processes (GLOMAP) [Spracklen *et al.*, 2010]. GLOMAP is an aerosol microphysics model applied with a horizontal resolution of  $2.8^\circ \times 2.8^\circ$  and 31 vertical levels from the surface to 10 hPa. Meteorological fields are specified from the six hourly analyses of the European Centre for Medium-Range Weather Forecasts for 2000. The PSD is described using a sectional scheme with 20 size sections in the dry diameter range 3 nm to 20  $\mu\text{m}$ . Total particle number concentrations are computed as the sum of a hydrophobic distribution, including freshly emitted carbonaceous particles (elemental carbon (EC) and organic carbon (OC)), and a hydrophilic distribution comprising sulfate, sea salt, EC, and OC particles. In this work we analyze several different GLOMAP experiments which vary in terms of the treatment of NPF (i.e., ACT1, ACT2, ACT3, KIN1, and KIN2) as described in Spracklen *et al.* [2010]. Each of these experiments includes primary particle emissions (from

**Table 1.** Multivariate Linear Regression Fit Parameters (and Standard Errors Computed Using QR Decomposition or Orthogonal Matrix Triangularization in Which a Matrix A is Decomposed Into an Orthogonal Matrix (Q) and an Upper Triangular Matrix (R)) in the Original Fit to the Data From MMSF (i.e., Equation (2)) and for MMSF and Egbert After a Threshold Was Applied to the Satellite-Derived  $[\text{SO}_2]$  ( $>0.5 \text{ DU}$ )<sup>a</sup>

$r^2$	MMSF Original Fit			MMSF Fit Using a Threshold of $[\text{SO}_2] > 0.5 \text{ DU}$		Egbert Fit Using a Threshold of $[\text{SO}_2] > 0.5 \text{ DU}$		Test for Whether the Regression Coefficients Derived From MMSF Differ From Those From Egbert	
	Estimate	Std. Error	Range	Estimate	Std. Error	Estimate	Std. Error	$H_0$	$p$ value
$r^2$	0.359			0.588		0.149			
# Obs (n)	63			36		34			
Parameter	Estimate	Std. Error	Range	Estimate	Std. Error	Estimate	Std. Error	$H_0$	$p$ value
$\alpha_{(\text{SO}_2)}$	-0.011	0.056	0.02-2.89	<b>0.039</b>	0.189	<b>0.640</b>	0.438	$\alpha_{(\text{SO}_2)} = 0.04$	0.107
$\alpha_{(\text{UV})}$	<b>0.452</b>	0.163	21.97-127.58	<b>0.535</b>	0.143	-0.777	0.932	$\alpha_{(\text{UV})} = 0.54$	0.084
$\alpha_{(\text{NH}_3)}$	0.088	0.079	0.02-1.80	<b>0.188</b>	0.080	-0.090	0.121	$\alpha_{(\text{NH}_3)} = 0.19$	0.030
$\alpha_{(\text{AOD})}$	<b>-0.246</b>	0.058	0-1.04	<b>-0.198</b>	0.057	-0.114	0.136	$\alpha_{(\text{AOD})} = -0.20$	0.284
$\alpha_{(\text{Å})}$	<b>-0.588</b>	0.248	0.59-1.78	<b>-0.834</b>	0.226	-0.101	1.388	$\alpha_{(\text{Å})} = -0.83$	0.302
$\beta$	<b>2.948</b>	0.324	-	<b>2.938</b>	0.281	<b>4.832</b>	1.898	$\beta = 2.94$	0.164

<sup>a</sup>The range of values used to develop the proxy at MMSF is also reported in column 4. Estimated values of the regression coefficients are shown in bold and underlined if they are significantly different from 0 at the 95% confidence level, while values shown only in bold are significantly different from 0 at the 80% confidence level. Based on an analysis to determine whether the estimated coefficients from independent fitting of the two sites (shown in the last two columns of the table), it appears that the model parameters from MMSF are not statistically different from those for Egbert based on a confidence level of 95% with the exception of the coefficient for  $\text{NH}_3$ .

anthropogenic and natural sources), binary homogeneous nucleation, and an empirical nucleation mechanism (either activation (ACT) or kinetic (KIN)).

### 3. Development and Evaluation of a Proxy for Ultrafine Particle Concentrations Over Eastern North America

#### 3.1. Formulation of the Proxy

[19] Direct in situ observations and modeling studies have shown that NPF and resulting high UFP number concentrations are a function of the photochemical production of nucleation precursors and concentration of preexisting aerosols in the atmosphere. However, since the precise nature of the functional dependence between these variables is uncertain [Zhang *et al.*, 2011], we employed a multivariate linear regression (MLR) approach to determine the appropriate form of our proxy algorithm, using log-transformed values of the candidate variables (described in section 2) as follows:

$$\text{Log}N_{6-100} = \alpha_1 \text{LogSO}_2 + \alpha_2 \text{LogUV} + \alpha_3 \text{LogNH}_3 + \alpha_4 \text{LogAOD} + \alpha_5 \text{LogÅ} + \beta \quad (2)$$

where  $N_{6-100}$  is the daily mean number concentration of particles for  $D_p = 6-100 \text{ nm}$ , and the independent variables are satellite-retrieved daily values of AOD and Å from MODIS, UV, and  $[\text{SO}_2]$  values from OMI-OMSO2e, and seasonally averaged  $[\text{NH}_3]$  from TES computed for each site using ISDW interpolation. The parameter space sampled in constructing the proxy for MMSF is shown in Table 1 and spans at least a factor of 5 for all predictors. The estimated proxy describes 36% of the variability in observed daily mean UFP number concentrations at MMSF (i.e., the  $r^2$  for equation (2) is 0.359 for  $n = 63$ ). A proxy equation was also developed independently for data from Egbert, but the proxy predictive skill is much lower than at MMSF ( $r^2$  for equation (2) is 0.172 for  $n = 53$ ) (see Table 2). Proxy algorithms derived using alternative diameter cut-points (e.g.,  $D_p < 30 \text{ nm}$ ) were evaluated, but using a definition of UFP with  $D_p = 6-100 \text{ nm}$  provided slightly better proxy performance, and

using this size range allows us to integrate the total number concentration over both the nucleation and Aitken mode and thus to capture both formation and initial growth.

[20] Given the difficulty of retrieving accurate  $[\text{SO}_2]$  from OMI when  $\text{SO}_2$  is present at low concentrations, we analyzed the sensitivity of the model fit to different threshold values of  $[\text{SO}_2]$  using a bootstrapping technique (Table 2), although this procedure might induce a bias toward representing more NPF than non-events since higher  $[\text{SO}_2]$  are more likely to be associated with nucleation occurrence. We assumed a reference sample size ( $n_{\text{ref}}$ ) of 35 and of 32 for MMSF and Egbert, respectively, based on the number of observations associated with  $\text{SO}_2 \geq 0.6 \text{ DU}$ . A bootstrapping technique was then applied to randomly sample  $n_{\text{ref}}$  days from the entire data set (i.e., 63 and 53 days for MMSF and Egbert, respectively), to determine the  $[\text{SO}_2]$  threshold value (i.e.,  $[\text{SO}_2]_{\text{min}}$ ) that maximized the mean coefficient of determination ( $r^2$ ) from 10,000 iterations (Table 2). Setting  $[\text{SO}_2]_{\text{min}} = 0.5 \text{ DU}$  increases the mean  $r^2$

**Table 2.** Mean Coefficient of Determination ( $r^2$ ) for Regression Equation (2) as a Function of the Threshold Value for  $[\text{SO}_2]$  Applied (i.e.,  $[\text{SO}_2]_{\text{min}}$ )<sup>a</sup>

$[\text{SO}_2]$ Threshold in DU	MMSF		Egbert	
	Sample Size	$r^2$	Sample Size	$r^2$
0	63	0.394	53	0.172
0.05	59	0.413	53	0.171
0.1	56	0.415	51	0.157
0.15	55	0.441	50	0.169
0.2	52	0.452	49	0.169
0.25	49	0.492	45	0.154
0.3	47	0.477	43	0.129
0.35	46	0.509	42	0.121
0.4	42	0.471	38	0.158
0.45	37	0.539	35	0.150
0.5	36	0.589	34	0.158
0.55	35	0.588	33	0.117
0.6	35	0.588	32	0.114

<sup>a</sup>The analysis was undertaken setting a sample size of 35 and 32 for MMSF and Egbert, respectively, and using 10,000 iterations in the bootstrapping. The sample size of the original data set after setting  $[\text{SO}_2]$  lower bounds (column 1) is also reported.

**Table 3.** Results From the ANOVA at MMSF When a Threshold of 0.5 DU is Applied to [SO<sub>2</sub>]

Model	r <sup>2</sup>	Variables	Df	Sum of Squares	F Value	p Value		
Model 1	0.1599	Log <sub>10</sub> UV	1	0.3631	6.4698	0.01569		
		residuals	34	1.90817				
Model 2	0.3356	Log <sub>10</sub> UV	1	0.3631	7.9408	0.008104		
		Log <sub>10</sub> AOD	1	0.39921	8.7304	0.005735		
Model 3	0.5117	residuals	33	1.50896	10.476	0.002812		
		Log <sub>10</sub> UV	1	0.3631				
		Log <sub>10</sub> AOD	1	0.39921			11.518	0.001853
		Log <sub>10</sub> Å	1	0.39987			11.537	0.001839
Model 4	0.5120	residuals	32	1.10909	10.156	0.003274		
		Log <sub>10</sub> UV	1	0.3631				
		Log <sub>10</sub> AOD	1	0.39921			11.166	0.002184
		Log <sub>10</sub> Å	1	0.39987			11.185	0.002168
		Log <sub>10</sub> SO <sub>2</sub>	1	0.00082			0.023	0.880518
Model 5	0.5878	residuals	31	1.10827	11.635	0.001869		
		Log <sub>10</sub> UV	1	0.3631				
		Log <sub>10</sub> AOD	1	0.39921			12.7921	0.001204
		Log <sub>10</sub> Å	1	0.39987			12.8132	0.001195
		Log <sub>10</sub> SO <sub>2</sub>	1	0.00082			0.0263	0.872224
		Log <sub>10</sub> NH <sub>3</sub>	1	0.17205			5.513	0.025662
		residuals	30	0.93622				

from 0.39 to 0.59 at MMSF, and the result equation for UFP concentration as a function of the satellite-derived data has the following form:

$$N_{6-100} = 868 \times \frac{\text{SO}_2^{0.04} \times \text{UV}^{0.54} \times \text{NH}_3^{0.19}}{\text{AOD}^{0.20} \times \alpha^{0.83}} \quad (3)$$

[21] The form of, and coefficients in, the proxy algorithm shown in Table 1 and Equation (3) can be physically interpreted. The numerator contains variables that play a role in the formation of nucleation precursors (i.e., UV, SO<sub>2</sub>, and NH<sub>3</sub>), whereas the denominator contains terms associated with suppression of NPF due to reduction of the concentration of semivolatile gases by condensation on existing aerosols (i.e., AOD and Å as representative of CS). Further, as expected from the complex system of chemical reactions involving SO<sub>2</sub> and NH<sub>3</sub> in the atmosphere, the exponents associated with these trace gases are lower than 1. Similarly, the quantum yield for photochemical production of OH which is involved in the formation of H<sub>2</sub>SO<sub>4</sub> is less than unity, which is consistent with the estimated coefficient for UV [Fountoukis and Nenes, 2007; Gerhard and Johnstone, 1955]. To quantify the importance of the predictors in explaining the variability in UFP number concentrations, we applied an analysis of variance (ANOVA) test adding the variables presented in equation (2) one at a time. Comparing the sum of squares of the regression model with that associated with the unexplained variance, we tested if adding more variables reduced the unexplained variance and thus increased r<sup>2</sup>. Further, in order to assess if adding new variables significantly improves the estimates, we applied an *F* test. As shown by Table 3, adding each variable improves the predictive skill of the model (i.e., constantly increasing r<sup>2</sup>) and all the variables, except for SO<sub>2</sub>, significantly improve the model skill at a 5% confidence level. Although the estimated coefficient for SO<sub>2</sub> is not statistically different from 0 and does not improve significantly the fit (see Tables 1 and 3), we maintain this variable in our proxy formulation for physical reasons; H<sub>2</sub>SO<sub>4</sub> and its precursor

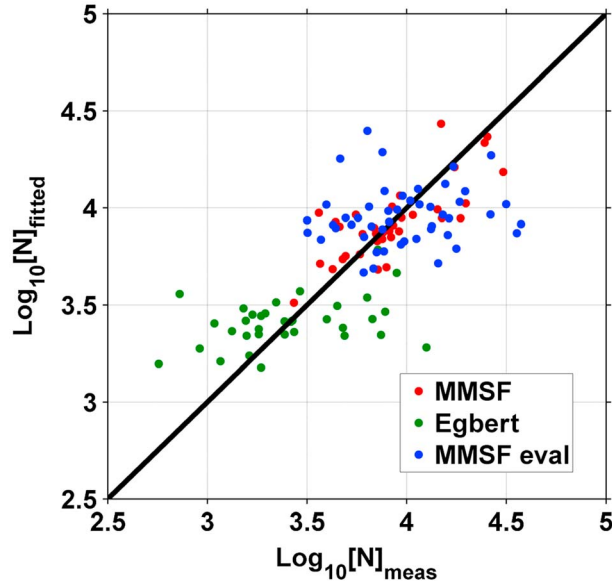
SO<sub>2</sub> have shown a strong correlation with the nucleation rate and thus with NPF events at several locations worldwide including the United States [Kuang *et al.*, 2008]. Further, ammonium sulfate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) was the main constituent of aerosol particle mass for D<sub>p</sub> < 50 nm at MMSF [Pryor *et al.*, 2011], and SO<sub>2</sub> has been demonstrated to be an essential control on nucleation at this site [Crippa *et al.*, 2012; Pryor *et al.*, 2011]. The prefactor (i.e., β), which derives from the logarithm of the intercept, may be interpreted as the background UFP concentration (the minimum daily mean UFP concentration at MMSF was 857 cm<sup>-3</sup> during the days used to fit the regression model).

[22] To provide an independent evaluation of the regression fit to number concentration data for D<sub>p</sub> < 100 nm from MMSF, equation (3) was independently applied to data from December 2006 to April 2007 and June 2008 to August 2009.

The mean absolute error (MAE = |Log<sub>10</sub>obs<sub>*i*</sub> - Log<sub>10</sub>sat<sub>*i*</sub>|) between the base-10 logarithm of observed (obs) and estimated (sat) UFP number concentrations in the training data sets from MMSF and Egbert (i.e., from May 2007 to May 2008) and in the independent evaluation data from MMSF are similar (i.e., 0.12, 0.23, and 0.24, respectively). Thus, the fit described in (3) exhibits predictive skill for the independent data sample from MMSF, and the general form of the equation appears to be similar (though not identical) at MMSF and Egbert, implying some degree of generalizability (see Figure 3 and Table 1). The coefficient values (α) in a fit to the independent data (data from December 2006 to April 2007 and June 2008 to August 2009) from MMSF differ from those in Table 1 for UV, Å, and the intercept, but when data from the entire period December 2006 to August 2009 are used to determine the coefficients, the values are not different from those shown in equation (3). This indicates that the form of equation (3) is relatively robust but that the daily variability in emissions and meteorological conditions may affect the functional form of the proxy, and thus equation (3) may not always provide optimal estimates for short-term predictions.

[23] Although the proxy predictive skill in data from Egbert is much lower than at MMSF for all ranges of [SO<sub>2</sub>]<sub>min</sub>





**Figure 3.** Comparison between the base-10 logarithm of the daily average UFP number concentration and fitted values computed from equation (3) using data from MMSF ( $r^2=0.588$  and  $n=36$ ) and Egbert ( $r^2=0.149$  and  $n=34$ ) during May 2007 to May 2008. The blue dots show the proxy evaluation based on independent data collected at MMSF during December 2006 to April 2007 and June 2008 to August 2009. For a regression equation with forced-zero intercept, this proxy evaluation with the independent data set ( $n=48$ ) has an  $r^2=0.985$  (see *Eisenhauer* [2003] for a discussion of calculating  $r^2$  values for regression through the origin; here we use  $R^2 = \frac{\sum \hat{Y}_i^2}{\sum Y_i^2}$ ).

(see Table 2), the estimated coefficients when  $[\text{SO}_2] > 0.5$  DU is used as a threshold are not statistically significantly different from those derived at for MMSF (Table 1). This may indicate that a similar mechanism is responsible for NPF at both sites, although more definitive conclusions might be drawn if more data were available. The lack of sensitivity to  $[\text{SO}_2]_{\text{min}}$  at Egbert may be indicative of a major role played by local emission sources [*Riipinen et al.*, 2011], and it is also observed in data from the Egbert AERONET site as discussed below.

### 3.2. Evaluation of the Proxy Relative to Prior Research

[24] The previously developed proxy for NPF (i.e., aerosol particle number concentrations for  $D_p < 25$  nm) using satellite-derived observations is based on an assumption that NPF may be described using an activation mechanism (i.e., with a linear relationship between the nucleation rate and  $\text{H}_2\text{SO}_4$  concentration) and has the following form [*Kulmala et al.*, 2011]:

$$N_{<25\text{nm}} = k \times \frac{\text{SO}_2^1 \times \text{UV}^1}{\text{AOD}^2} \quad (4)$$

[25] To test the generalizability of that model (equation (4)) and assess its performance relative to the formulation

**Table 4.** Regression Coefficient Estimates and Standard Errors for the Multivariate Regression Analysis (See Equation (5)) Based on UFP Concentration Data for MMSF and Using Only the Parameters Considered by *Kulmala et al.* [2011] Without Applying a Threshold on  $[\text{SO}_2]^a$

Parameter	Estimate	Std. Error	$H_0^b$	$p$ Value
$\alpha_{(\text{SO}_2)}$	0.006	0.058	$\alpha_{(\text{SO}_2)}=1$	$<10^{-16}$
$\alpha_{(\text{UV})}$	<b><u>0.350</u></b>	0.123	$\alpha_{(\text{UV})}=1$	$<0.001$
$\alpha_{(\text{AOD})}$	<b><u>-0.251</u></b>	0.060	$\alpha_{(\text{AOD})}=2$	$<10^{-16}$
$\beta$	<b><u>3.005</u></b>	0.251	-	-

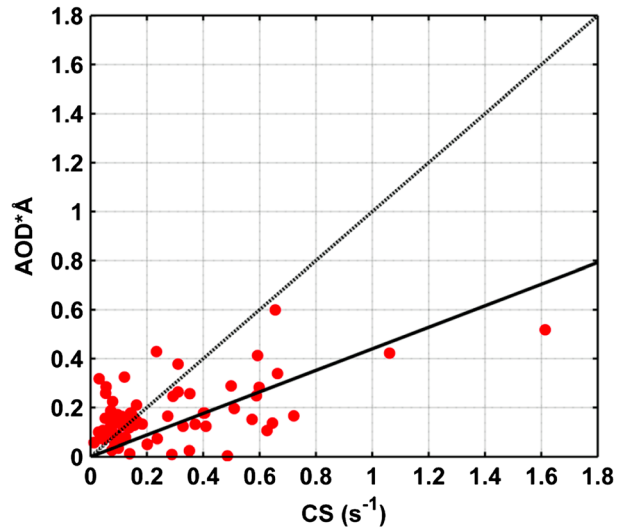
<sup>a</sup>Estimated values shown in bold and underlined are significantly different from 0 at the 95% confidence level. The last two columns indicate that a null hypothesis that the regression coefficients are equal to those set a priori by *Kulmala et al.* [2011] is rejected at a confidence level  $>99\%$  (we do not report the intercept since it is not directly comparable with least-squares fit approach of *Kulmala et al.* [2011]).

<sup>b</sup>Values are as in *Kulmala*.

proposed herein (equation (3)), we applied the following multivariate regression to data from MMSF:

$$\text{Log}N_{6-100} = \alpha_1 \text{LogSO}_2 + \alpha_2 \text{LogUV} + \alpha_3 \text{LogAOD} + \beta \quad (5)$$

[26] The parameter fits from multiple linear regression (i.e.,  $\alpha$  coefficients and  $\beta$ ) differ substantially from those in equation (4), and the variance explanation for equation (5) is lower than the proxy algorithm proposed here (equation (3)) (Table 4). Thus, we assert that the proxy formulation shown in equation (3) is an improvement relative



**Figure 4.** Comparison between  $\text{AOD} \times \hat{A}$  and condensational sink (CS) for  $D_p = 6$  nm to  $10 \mu\text{m}$  estimated using data from the FMPS operated at MMSF. Data were collected during 1 November 2007 to 31 May 2008. The CS was estimated by applying multiple lognormal distributions to the average PSD measured during 10 A.M. TO 12 P.M. (LST) by the FMPS. The regression through the origin is shown by the solid line:  $y=0.44 \times x$  ( $n=69$ , and the  $r^2$  for a forced-zero intercept

regression line is computed from  $R^2 = \frac{\sum \hat{Y}_i^2}{\sum Y_i^2} = 0.6357$  [*Eisenhauer*, 2003]). For comparison, the 1:1 line is also reported (dotted line).

**Table 5.** Variance Explanation ( $r^2$ ) and Sample Size (n) for Each of the AERONET Sites<sup>a</sup>

Parameter	Bondville	Sioux Falls	Walker Branch	Egbert	Harvard Forest
	2006–2008	2006–2008	2006–2008	2006–2008	2007–2008
$r^2$	0.616	0.585	0.497	0.327	0.405
n	63	72	77	80	39
$\alpha_{(\text{SO}_2)}$	<b><u>-0.077</u></b>	<b><u>0.011</u></b>	0.253	-0.007	0.266
$\alpha_{(\text{UV})}$	<b><u>0.793</u></b>	<i>0.269</i>	<i>0.410</i>	<i>1.078</i>	<i>0.211</i>
$\alpha_{(\text{NH}_3)}$	0.392	0.080	0.246	<b>-0.770</b>	<b>0.033</b>
$\alpha_{(\text{AOD})}$	<b>0.303</b>	<i>0.306</i>	<i>0.369</i>	<i>0.362</i>	<i>0.293</i>
$\alpha_{(\lambda)}$	0.105	<i>0.172</i>	0.149	0.199	0.451
$\beta$	<b>4.961</b>	0.133	0.202	3.833	0.160

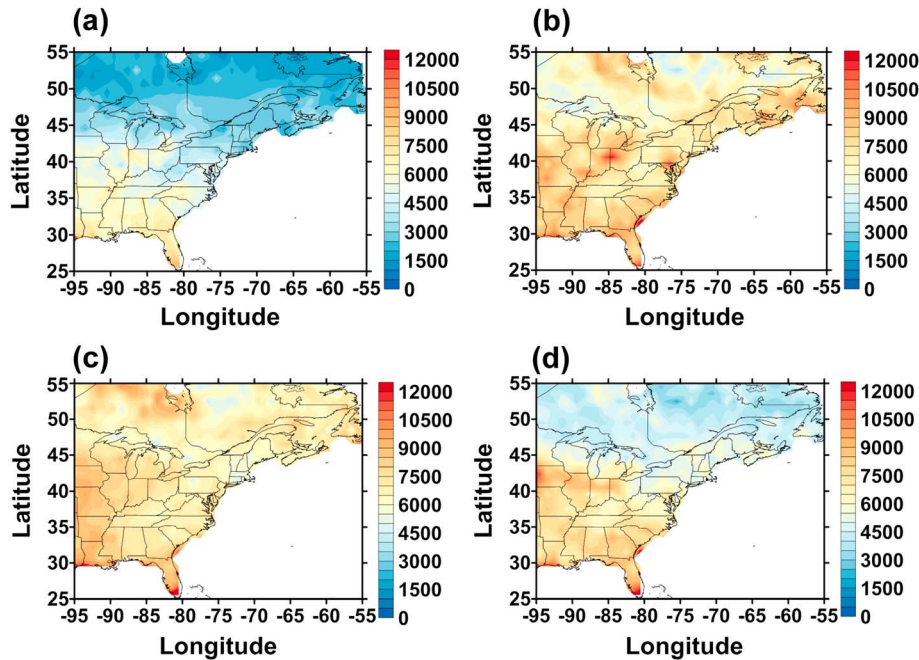
<sup>a</sup>Also shown are the coefficient estimates from multivariate linear regression applied to AERONET-derived total number concentrations  $D_p = 100\text{--}200\text{ nm}$ . Estimated values shown in bold and underlined in column 2 (i.e., for the fit from Bondville) are significantly different from 0 at the 95% confidence level, while values shown only in bold are significantly different from 0 at the 80% confidence level. In the other columns values shown in bold and underlined are significantly different from coefficient derived for Bondville again at the 95% confidence level. Values in italics indicate variables contributing significantly to the regression fit at 95% confidence level as tested by ANOVA.

to that postulated by *Kulmala et al.* [2011] for the following reasons:

[27] 1. The formulation appears statistically robust. The predictive skill of our proxy is higher than the results presented for the Hyytiälä forest site. *Kulmala et al.* [2011] report  $r = 0.25$  for equation (4) relative to observed particle concentrations for  $D_p < 25\text{ nm}$  at Hyytiälä, which equates to an  $r^2 = 0.06$ . For all variables, our estimated exponents from equation (5) are significantly different from those in equation (4) (Table 4), indicating the model proposed in equation (4) is not optimal to describe the variability of UFP concentrations measured at MMSF. Further, including the gas-phase  $\text{NH}_3$  concentration and the Ångström exponent significantly improves the model fit (i.e.,  $r^2$  for the model shown in equation (3)) = 0.588 if the  $\text{SO}_2$  threshold is applied and 0.359 otherwise; see Table 1).

[28] 2. Derivation of the proxy presented herein is not based on as many a priori assumptions. Further, the formulation given in equation (3) is consistent with physical reasoning. We investigated the relationship between  $(\text{AOD} \times \text{Å})$  and CS as derived from a lognormal fit to the FMPS data at MMSF, measuring the PSD in the size range of 25–523.3 nm and extrapolating it up to 10  $\mu\text{m}$ . This analysis showed that  $(\text{AOD} \times \text{Å})$  is a good proxy for the CS of 6 nm to 10  $\mu\text{m}$  particles (Figure 4,  $r^2 = 0.636$  when using  $\text{AOD} \times \text{Å}$  and  $r^2 = 0.503$  when using AOD only). Although the variance explained by using  $\text{AOD} \times \text{Å}$  is only slightly higher than using AOD only, the experimental data provide consistent results with our mathematical proof showing that  $\text{AOD} \times \text{Å}$  provides a better approximation for CS than AOD alone (see supporting information).

[29] 3. We develop and apply a proxy to estimate time-averaged UFP concentrations since temporal variability in



**Figure 5.** Seasonally average UFP number concentrations ( $\text{cm}^{-3}$ ) for 2006–2009 in (a) winter, (b) spring, (c) summer, and (d) fall derived using the satellite-based proxy shown in equation (3) (i.e.,  $N = 868 \times \frac{[\text{SO}_2]^{0.04} \times UV^{0.54} \times [\text{NH}_3]^{0.19}}{[\text{AOD}]^{0.20} \times [\alpha]^{0.83}}$ ).



**Table 6.** Seasonal Normalized Mean Bias (NMB) and Pearson's Correlation Coefficient ( $r$ ) of UFP Concentrations as Simulated by GLOMAP for Different Experiments (ACT1, ACT2, ACT3, KIN1, KIN2) Relative to Our Proxy Estimates<sup>a</sup>

Experiment (Nucleation Coefficient)	Season	NMB	$r$	95% CI
ACT1 ( $A = 2 \times 10^{-7} \text{ s}^{-1}$ )	Winter	-0.776	<b>0.567</b>	[0.537, 0.596]
	Spring	-0.802	<b>0.177</b>	[0.135, 0.219]
	Summer	-0.882	<b>0.091</b>	[0.047, 0.134]
	Fall	-0.840	<b>0.449</b>	[0.413, 0.483]
ACT2 ( $A = 2 \times 10^{-6} \text{ s}^{-1}$ )	Winter	-0.643	<b>0.533</b>	[0.501, 0.564]
	Spring	-0.770	<b>0.129</b>	[0.086, 0.172]
	Summer	-0.827	<b>0.156</b>	[0.112, 0.198]
	Fall	-0.740	<b>0.556</b>	[0.525, 0.586]
ACT3 ( $A = 2 \times 10^{-5} \text{ s}^{-1}$ )	Winter	-0.441	<b>0.516</b>	[0.483, 0.547]
	Spring	-0.345	0.044	[0.001, 0.088]
	Summer	-0.744	<b>0.270</b>	[0.229, 0.310]
	Fall	-0.579	<b>0.557</b>	[0.526, 0.586]
KIN1 ( $K = 2 \times 10^{-13} \text{ cm}^3 \text{ s}^{-1}$ )	Winter	-0.675	<b>0.552</b>	[0.521, 0.582]
	Spring	-0.788	<b>0.167</b>	[0.124, 0.209]
	Summer	-0.830	<b>0.168</b>	[0.125, 0.210]
	Fall	-0.747	<b>0.525</b>	[0.493, 0.556]
KIN2 ( $K = 2 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$ )	Winter	-0.512	<b>0.540</b>	[0.508, 0.570]
	Spring	-0.689	<b>0.105</b>	[0.061, 0.148]
	Summer	-0.762	<b>0.254</b>	[0.213, 0.295]
	Fall	-0.613	<b>0.560</b>	[0.529, 0.589]

<sup>a</sup>Each experiment simulates primary emissions, binary homogeneous nucleation, and an empirical particle formation mechanism where the nucleation coefficients for either the activation (A) or the kinetic (K) approach are reported in column 1. The highest agreement between satellite-derived and GLOMAP-simulated values occurs for experiment KIN2; thus, spatial patterns for that numerical experiment are shown in Figure 6. Bold  $r$  values indicate correlations significantly different from 0 at a confidence level of 95%. The 95% confidence interval (CI) for  $r$  is also reported.

number concentrations at high resolution (e.g., 30 min averages are used in *Kulmala et al.* [2011]) cannot be captured using once daily satellite observations.

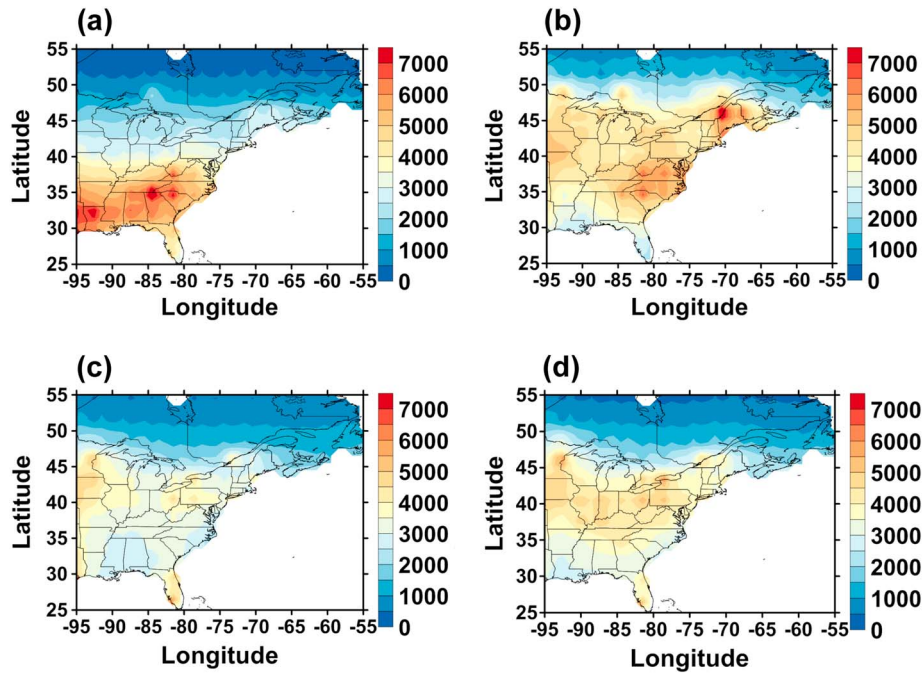
### 3.3. Evaluation of the Relationship Between Satellite Retrievals and Data From AERONET Sites

[30] Daily average observations from five AERONET sites were used to further investigate the relationship between the satellite-derived properties used to build the UFP proxy and properties of aerosol particles of climate relevant sizes over eastern North America. In this analysis the number concentrations for  $D_p = 100\text{--}200 \text{ nm}$  (i.e., for particles that significantly contribute to cloud condensation nuclei (CCN)) were used to derive a fit similar to that shown in equation (3), and we use the Bondville site as a reference location for the statistical intercomparison given it is the closest AERONET site to MMSF. Although we do not expect the proxy developed for the AERONET data to maintain the same mathematical and physical structure of equation (3) given different aerosol dynamics dictate the concentration of larger particles, the purpose of this analysis is to test the spatial variability of the relationship between aerosol particle concentrations and the link to physical parameters used in the UFP proxy. The a priori hypothesis is that at least some UFP will grow to  $D_p = 100\text{--}200 \text{ nm}$ , thus that there will be a positive association between particle concentrations in the diameter space  $100\text{--}200 \text{ nm}$  and the UFP proxy predictors, and that the form of the relationship will show some spatial consistency, allowing the inference to be drawn that we can spatially extrapolate from the proxy algorithm as conditioned on data from MMSF. Moreover, since both satellite-derived variables and AERONET measurements are columnar values, we anticipate covariability. The results of this analysis support these hypotheses since they indicate that (i) the satellite-derived predictors (AOD,  $\text{\AA}$ , UV,  $\text{NH}_3$ , and  $\text{SO}_2$ ) are able to capture at least some of

the day-to-day variability in number concentrations for  $D_p = 100\text{--}200 \text{ nm}$  (i.e.,  $r^2$  values for the regression fits to data from the five AERONET sites range from 0.33 to 0.62), and that (ii) there is some consistency between the coefficients derived for the MLR fits to independent data from all sites (i.e., coefficient estimates from the regression equation fitted at Bondville are not significantly different at 95% confidence level from those estimated at the other sites for most variables) (Table 5). In order to assess the relative contribution of each variable to the regression fit, an ANOVA test was conducted (Table 5). ANOVA results indicate that AOD and UV are the most significant variables for explaining the daily variability in CCN at all sites, whereas only at Sioux Falls,  $\text{SO}_2$  and  $\text{\AA}$  also appear to play a major role. These results indicate that the Ångström exponent and  $\text{NH}_3$ , identified as significant players in describing UFP concentrations at MMSF, become less relevant as particles grow to a larger size. The positive association between MODIS AOD and AERONET number concentrations further supports that these particles have already grown to a size where they are part of the mode that determines AOD and the Ångström exponent.

### 4. Application of the Proxy to ‘Predict’ Seasonal Average Ultrafine Particle Concentrations Over Eastern North America

[31] The evaluation exercises presented above indicate some level of skill in the proxy presented in equation (3) in reproducing the temporal and spatial variability of NPF and resulting UFP concentrations at a range of stations across eastern North America. Thus, we applied equation (3) across the entire eastern North America region using the mean during 2006–2009 of the 3 month averaged satellite retrievals interpolated to a common  $1^\circ \times 1^\circ$  grid. Use of seasonally averaged satellite retrievals is constrained by the format in



**Figure 6.** Seasonally average total number concentrations ( $D_p > 3 \text{ nm}$ ) ( $\text{cm}^{-3}$ ) as simulated by GLOMAP (KIN2 experiment) including primary emissions, binary homogeneous, and kinetic ( $K = 2 \times 10^{-12} \text{ cm}^{-3} \text{ s}^{-1}$ ) nucleation mechanisms above and within the boundary layer, respectively, in (a) winter, (b) spring, (c) summer, and (d) fall during 2000 [Spracklen *et al.*, 2010]. Note that in this depiction the data have been regridded from the native grid of  $2.8^\circ \times 2.8^\circ$  to a resolution of  $1^\circ \times 1^\circ$  using ISDW interpolation.

which data for large domains are available on the GIOVANNI interface. To evaluate the uncertainty induced by applying the proxy that was developed on daily values to seasonally averaged values, a sensitivity analysis was conducted on data from MMSF during May 2007 to May 2008. The results showed UFP concentrations computed applying the proxy in equation (3) to seasonally averaged satellite retrievals are underestimated by approximately 16%, 24%, 7%, and 28% for winter, spring, summer, and fall, respectively, relative to the observed number concentrations. Although the sample size on which this sensitivity analysis was conducted is not very large ( $n_{\text{total}} = 63$ ), we thus expect our seasonal estimates (shown in Figure 5) to be low biased relative to observations.

[32] Application of the proxy generates results which indicate high UFP concentrations over large areas in eastern North America during spring and summer (Figure 5b and 5c), with peak concentrations during the spring over central Indiana and western Ohio. This is consistent with long-term observations at MMSF that show NPF is most frequently observed during spring (when NPF is observed on about 1 day in 5) and least frequent during winter (when NPF was observed on less than 1 day in 20) [Pryor *et al.*, 2010]. Our proxy estimates do not exhibit pronounced interannual variability over the 4 year period, and a similar seasonal pattern is observed during each year. The fall peak in NPF over eastern North America predicted by the proxy formulated by Kulmala *et al.* [2011] is not supported by observations from MMSF and Egbert and is not replicated in results from the proxy developed in this current work (Figure 5).

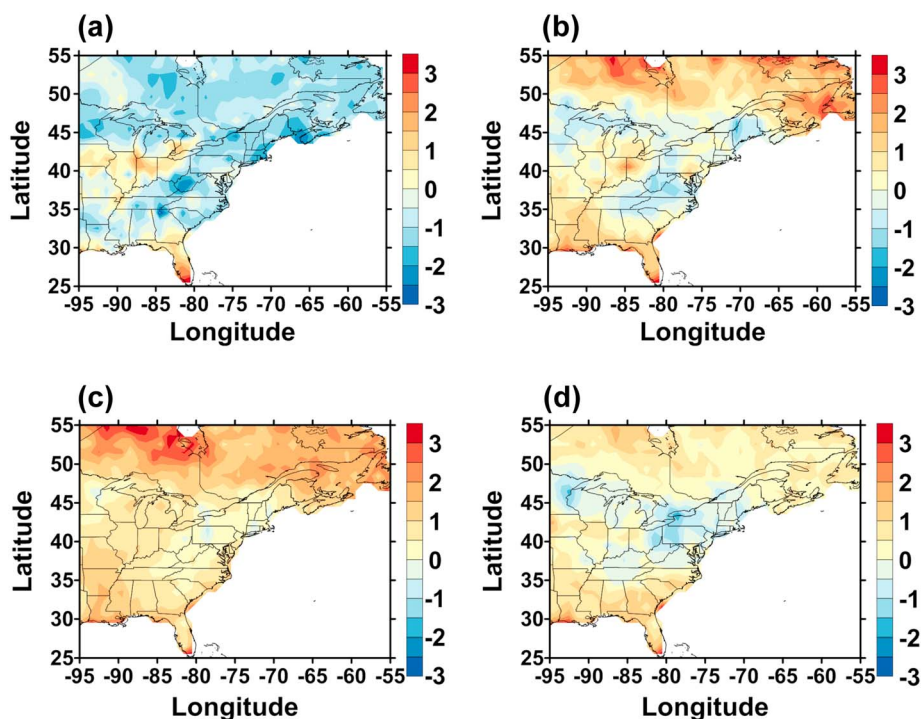
[33] Analysis of the importance of each of the variables in equation (3) in determining the resulting seasonality in

UFP concentrations indicates that summer maxima in UV over the entire domain explain the high predicted UFP concentrations, whereas the lack of spatial variability appears to reflect the spatially homogeneous values of both UV and Å. The seasonal variability of  $\text{NH}_3$  is most pronounced in states to the west of Indiana and is partially responsible for the higher spatial variability during the spring and fall seasons. Satellite-derived estimates of  $\text{SO}_2$  peak during winter, but the relatively small weighting coefficient, and offsetting effect of the other parameters, means  $\text{SO}_2$  does not appear to be responsible for much of the observed seasonality in the UFP proxy.

[34] We also cross-evaluated the predicted seasonal pattern of UFP from our satellite-based proxy with total particle concentrations from a range of simulations conducted using GLOMAP for the year 2000. In this analysis the normalized mean bias (NMB) is used to quantify the bias between GLOMAP-simulated values (sim) and our satellite-based estimates (sat) for each  $1^\circ \times 1^\circ$  grid cell  $i$  over the analyzed domain (where GLOMAP values have been regridded to a common  $1^\circ \times 1^\circ$  grid using ISDW interpolation):

$$NMB = \frac{\sum_{i=1}^n (sim_i - sat_i)}{\sum_{i=1}^n sat_i} \quad (6)$$

[35] NMB, computed for each season, indicates a systematic underprediction of total particle number concentrations from GLOMAP relative to those from the satellite proxy (Table 6), which is consistent with previous comparisons of



**Figure 7.** Seasonal spatial bias of standardized particle number concentrations ( $\text{cm}^{-3}$ ) computed as the difference between proxy values from equation (3) (shown in Figure 5) and GLOMAP simulations (shown in Figure 6) during (a) winter, (b) spring, (c) summer, and (d) fall. Values  $x_i$  have been standardized using  $\frac{x_i - \bar{x}}{\sigma_x}$ , where  $x_i$  is the seasonal average particle number concentration in the grid cell  $i$ , and  $\bar{x}$  and  $\sigma_x$  are the seasonal spatial mean and standard deviation, respectively. For a specific grid cell, a value of the bias equal to 0 indicates the two data sets exhibit the same departure from the mean. The 95% confidence interval is approximately  $[-2\sqrt{2}, +2\sqrt{2}]$  (the distribution of the difference is  $\approx N(0, 2)$ ) and includes almost all the spatial bias values calculated across the entire eastern North America.

GLOMAP relative to in situ observations [Spracklen *et al.*, 2010]. This is true for all GLOMAP treatments of NPF in which the boundary-layer nucleation rate is controlled by either a linear or a squared dependence on  $[\text{H}_2\text{SO}_4]$  (as represented by the activation and kinetic approach, respectively) and is true for the range of nucleation coefficients applied. Causes of the positive bias in the satellite-based proxy relative to GLOMAP include (i) the proxy might be biased toward representing more nucleation events and thus higher UFP concentrations since the satellite data are available only for cloud-free pixels; (ii) the proxy implicitly represents NPF from a ternary nucleation mechanism, which may lead to higher UFP number concentrations in response to the role played by  $\text{NH}_3$  in the formation of a charged cluster and in increasing the nucleation rate and thus the subsequent particle growth; (iii) NPF and subsequent growth may be affected by the presence of VOCs and other condensable species which are implicitly included in the proxy but may be underestimated in GLOMAP simulations given the only secondary organic aerosol source is from monoterpenes; and (iv) the global annual mean background surface concentration associated with primary emissions from GLOMAP is  $640 \text{ cm}^{-3}$  over the continents [Spracklen *et al.*, 2010], and thus a systematic positive bias of at least 35% is expected from primary emissions (our proxy intercept is  $868 \text{ cm}^{-3}$ ). We should also point out that although our proxy is tuned relative to near-surface observations, it is based on satellite total column values, and thus spatial variability in

NPF (or primary emissions) not captured by the proxy may also result in a positive bias. Further, given the effect of applying different nucleation coefficients does not reduce significantly the bias (Table 6), it is possible that this bias arises from errors in the estimated condensation and coagulation sinks (e.g., due to incorrect specification of aerosol particle density). In the following intercomparison we primarily focus on simulated patterns including primary emissions, binary homogeneous nucleation, and a kinetic approach for nucleation (experiment KIN2, with  $K = 2 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$ ), which is the GLOMAP experiment with greatest agreement with our proxy (highest Pearson's coefficients and lowest NMB) (Table 6). Despite the bias in absolute UFP concentrations from GLOMAP relative to the satellite proxy, a similar seasonal cycle is observed in the two data sets, with highest concentrations over the entire domain during spring (Figures 5b and 6b). The spatial patterns shown in Figures 5 and 6 exhibit relatively high Pearson's correlation coefficients particularly in winter and fall and lowest values in spring and summer (Table 6), which is consistent with comparison of GLOMAP with in situ observations [Spracklen *et al.*, 2010]. The overall spatial variability as measured by the spatial standard deviation of UFP concentrations in Figures 5 and 6 normalized by the seasonal mean is higher for GLOMAP simulations than our proxy ranging from about 9% during winter to 77% in spring, thus indicating a possible role of local emissions in affecting the spatial variability of UFP.

[36] Although our predicted values based on the satellite observations show a systematic positive bias when compared to GLOMAP results, a more complex spatial and seasonal pattern is observed for the bias computed using values standardized to a distribution with mean 0 and variance 1 (this allows for a proper comparison of data sets with different means) (Figure 7). Our proxy estimates higher concentrations over Iowa, Illinois, Indiana, and Ohio than GLOMAP during all seasons. During the summer, use of the proxy results in an almost homogeneous overprediction over the entire domain relative to GLOMAP (Figure 7c), whereas the spatial pattern in winter is dominated by a negative bias. However, it is worthy of note that these biases are within the 95% confidence interval  $[-2\sqrt{2}, +2\sqrt{2}]$  over almost the entire domain, thus implying that the actual spatial variability as described by our proxy and GLOMAP is similar and that the systematic bias may be due to other factors, such as differences in background concentrations, vertical emission profiles, and the nucleation mechanism.

## 5. Concluding Remarks

[37] In this work we explore the possibility of using data from satellite-borne radiometers to estimate UFP number concentrations. We applied a multivariate linear regression approach to develop our proxy for total number concentration of particles with diameters below 100 nm using as predictors a range of satellite-observed properties (ultraviolet solar radiation, sulfur dioxide and ammonia concentrations, aerosol optical depth, and Ångström exponent) that were selected to represent the major physical and chemical controls on new particle formation. The functional dependencies expressed in the resulting statistical model are both physically interpretable and capable of representing aspects of both the temporal and spatial variabilities of UFP concentrations over eastern North America.

[38] Major findings from this work include the following:

[39] 1. UFP number concentrations can be predicted at remote locations using data from satellite-borne instrumentation. The accuracy of prediction is maximized for remote locations since they are less affected by local sources which cannot be captured by the relatively coarse resolution of satellite retrievals.

[40] 2. Consistent with prior research [Crippa and Pryor, 2013], there is evidence that UFP concentrations, and by association NPF events, are relatively coherent across large spatial scales over eastern North America.

[41] 3. NPF intensity peaks during spring and summer months when the photochemical production of nucleation precursors is maximized. These findings are consistent with long-term ground-based observations [Pryor et al., 2010].

[42] 4. Satellite-derived estimates of UFP concentrations exhibit a high degree of coherence with GLOMAP simulations in terms of seasonality and spatial patterns. The systematic positive bias in our proxy estimates might be related to differences in the nucleation mechanism applied, in background concentrations associated with primary emissions, and errors in the condensation and coagulation sinks. Part of the bias may be also dictated by differences in sampling periods (only cloud-free for the satellite-derived estimates).

[43] This work thus represents an advance in the use of satellite retrievals to investigate aerosol processes in the

atmosphere and potentially distinguish between background emissions and in situ production of freshly nucleated particles. We envisage that this technique may be a useful tool for developing global near-real-time aerosol particle number concentrations, which may thus provide unique insights when planning for strategies to reduce human exposure to high concentrations of atmospheric particles and manage climate forcing. Satellite-derived estimates of UFP concentrations and our ability to investigate aerosol dynamics will benefit from improvements in the spatial and temporal coverage of observations from space. A more complete diagnostic evaluation of the temporal variability of the proxy and thus of mechanisms responsible for NPF and UFP in different seasons would benefit from development of a transcontinent network for UFP measurements (such as has been undertaken over Europe [Asmi et al., 2011]), and deeper understanding about the relative contribution of NPF and primary emissions to the observed particle number concentrations would also benefit from development of size-resolved particle number emission inventory for North America similar to one generated for Europe [Paasonen et al., 2013].

[44] Although this study shows the possibility of integrating observations to accurate satellite retrievals for investigating aerosol issues and the generalizability of our approach, some limitations remain. There is relatively large uncertainty in the absolute magnitude of predicted UFP concentrations at least partially due to the uncertainty associated with the retrieval of the nucleation precursors when they are present at low concentrations (e.g., SO<sub>2</sub>) and the relatively sparse availability of NH<sub>3</sub> observations from space. Thus, future work will be directed to (i) assessing the sensitivity on the spatial and vertical resolution and number of size sections adopted in GLOMAP for simulating aerosol dynamics, (ii) investigating the vertical distribution of nucleation precursors (e.g., [SO<sub>2</sub>] and [NH<sub>3</sub>]) by comparison with observations and satellite data to better identify the atmospheric layer in which NPF is likely occurring, and (iii) a detailed investigation of the possible role of organics in dictating NPF and subsequent growth using, for example, satellite-retrieved formaldehyde columns as a proxy for isoprene emissions [Palmer et al., 2006].

[45] **Acknowledgments.** This research was funded by grants to S.C.P. from NSF (0544745 and 1102309). Richard Leitch is acknowledged for providing the PSD data for Egbert. TES ammonia data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. We also acknowledge the MODIS and OMI mission scientists and associated NASA personnel for the production of the data used in this research effort. AERONET data were obtained from the AERONET Data Display Interface website (Aerosol Inversion). Thanks to the AERONET group, in particular to Brent Holben, Norm O'Neill, Gregory Stensaa, and Mike Caughey.

## References

- Asmi, A., et al. (2011), Number size distributions and seasonality of submicron particles in Europe 2008–2009, *Atmos. Chem. Phys.*, 11(11), 5505–5538, doi:10.5194/acp-11-5505-2011.
- Bauer, S. E., and S. Menon (2012), Aerosol direct, indirect, semidirect, and surface albedo effects from sector contributions based on the IPCC AR5 emissions for preindustrial and present-day conditions, *J. Geophys. Res.*, 115, D01206, doi:10.1029/2011JD016816.
- Bowman, K. W., et al. (2006), Tropospheric emission spectrometer: Retrieval method and error analysis, *IEEE Trans. Geosci. Remote Sens.*, 44(5), 1297–1307, doi:10.1109/tgrs.2006.871234.
- Clough, S. A., et al. (2006), Forward model and Jacobians for tropospheric emission spectrometer retrievals, *IEEE Trans. Geosci. Remote Sens.*, 44(5), 1308–1323, doi:10.1109/tgrs.2005.860986.
- Costabile, F., F. Barnaba, F. Angelini, and G. P. Gobbi (2013), Identification of key aerosol populations through their size and composition resolved

- spectral scattering and absorption, *Atmos. Chem. Phys.*, 13, 2455–2470, doi:10.5194/acp-13-2455-2013.
- Crippa, P., and S. C. Pryor (2013), Spatial and temporal scales of new particle formation events in eastern North America, *Atmos. Environ.*, 75, 257–264, doi:10.1016/j.atmosenv.2013.04.051.
- Crippa, P., T. Petäjä, H. Korhonen, G. S. E. Afandi, and S. C. Pryor (2012), Evidence of an elevated source of nucleation based on model simulations and data from the NIFTy experiment, *Atmos. Chem. Phys.*, 12(17), 8021–8036, doi:10.5194/acp-12-8021-2012.
- van Donkelaar, A., R. V. Martin, M. Brauer, R. Kahn, R. Levy, C. Verduzco, and P. J. Villeneuve (2010), Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application, *Environ. Health Perspect.*, 118(10), 847–855, doi:10.1289/ehp.0901623.
- Dubovik, O., and M. D. King (2000), A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *J. Geophys. Res.*, 105(D16), 20,673–20,696, doi:10.1029/2000JD900282.
- Dubovik, O., A. Smirnov, B. N. Holben, M. D. King, Y. J. Kaufman, T. F. Eck, and I. Slutsker (2000), Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, *J. Geophys. Res.*, 105(D8), 9791–9806, doi:10.1029/2000JD900040.
- Dubovik, O., B. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanré, and I. Slutsker (2002), Variability of absorption and optical properties of key aerosol types observed in worldwide locations, *J. Atmos. Sci.*, 59(3), 590–608, doi:10.1175/1520-0469(2002)059<0590:voaaop>2.0.co;2.
- Eisenhauer, J. G. (2003), Regression through the origin, *Teach. Stat.*, 25(3), 76–80, doi:10.1111/1467-9639.00136.
- Engel-Cox, J. A., C. H. Holloman, B. W. Coutant, and R. M. Hoff (2004), Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality, *Atmos. Environ.*, 38(16), 2495–2509, doi:10.1016/j.atmosenv.2004.2401.2039.
- Fountoukis, C., and A. Nenes (2007), ISORROPIA II: A computationally efficient thermodynamic equilibrium model for  $K^+$ – $Ca^{2+}$ – $Mg^{2+}$ – $NH_4^+$ – $Na^+$ – $SO_4^{2-}$ – $NO_3^-$ – $Cl^-$ – $H_2O$  aerosols, *Atmos. Chem. Phys.*, 7(17), 4639–4659, doi:10.5194/acp-7-4639-2007.
- Gerhard, E. R., and H. F. Johnstone (1955), Air pollution studies—Photochemical oxidation of sulfur dioxide in air, *Ind. Eng. Chem.*, 47(5), 972–976, doi:10.1021/ie50545a50027.
- Gobbi, G. P., Y. J. Kaufman, I. Koren, and T. F. Eck (2007), Classification of aerosol properties derived from AERONET direct sun data, *Atmos. Chem. Phys.*, 7, 453–458, doi:10.5194/acp-7-453-2007.
- Gupta, P., and S. A. Christopher (2008), Seven year particulate matter air quality assessment from surface and satellite measurements, *Atmos. Chem. Phys.*, 8(12), 3311–3324, doi:10.5194/acp-8-3311-2008.
- Heald, C. L. et al. (2012), Atmospheric ammonia and particulate inorganic nitrogen over the United States, *Atmos. Chem. Phys.*, 12(8), 10,295–10,312, doi:10.5194/acp-12-10295-2012.
- Holben, B. N., 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(1098)00031-00035.
- Hyer, E. J., J. S. Reid, and J. Zhang (2011), An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, *Atmos. Meas. Tech.*, 4(3), 379–408, doi:10.5194/amt-4-379-2011.
- Jung, J. G., C. Fountoukis, P. J. Adams, and S. N. Pandis (2010), Simulation of in situ ultrafine particle formation in the eastern United States using PMCAMx-UF, *J. Geophys. Res.*, 115, D03203, doi:10.1029/2009JD012313.
- Knol, A. B., et al. (2009), Expert elicitation on ultrafine particles: Likelihood of health effects and causal pathways, *Part. Fibre Toxicol.*, 6, 16, doi:10.1186/1743-8977-6-19.
- Koelemeijer, R. B. A., C. D. Homan, and J. Matthijsen (2006), Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe, *Atmos. Environ.*, 40(27), 5304–5315, doi:10.1016/j.atmosenv.2006.04.044.
- Krotkov, N. A., S. A. Carn, A. J. Krueger, P. K. Bhartia, and K. Yang (2006), Band residual difference algorithm for retrieval of SO<sub>2</sub> from the aura Ozone Monitoring Instrument (OMI), *IEEE Trans. Geosci. Remote Sens.*, 44(5), 1259–1266, doi:10.1109/tgrs.2005.861932.
- Kuang, C., P. H. McMurry, and A. V. McCormick (2008), Dependence of nucleation rates on sulfuric acid vapor concentration in diverse atmospheric locations, *J. Geophys. Res.*, 113, D10209, doi:10.1029/2007JD009253.
- Kuang, C., P. H. McMurry, and A. V. McCormick (2009), Determination of cloud condensation nuclei production from measured new particle formation events, *Geophys. Res. Lett.*, 36, L09822, doi:10.1029/2009GL037584.
- Kulmala, M. (2003), How particles nucleate and grow, *Science*, 302(5647), 1000–1001, doi:10.1126/science.1090848.
- Kulmala, M., H. Vehkamäki, T. Petaja, M. Dal Maso, A. Lauri, V. M. Kerminen, W. Birmili, and P. H. McMurry (2004), Formation and growth rates of ultrafine atmospheric particles: A review of observations, *J. Aerosol. Sci.*, 35, 143–176, doi:10.1016/j.jaerosci.2003.10.003.
- Kulmala, M., A. Arola, T. Nieminen, L. Riuttanen, L. Sogacheva, G. de Leeuw, V.-M. Kerminen, and K. E. J. Lehtinen (2011), The first estimates of global nucleation mode aerosol concentrations based on satellite measurements, *Atmos. Chem. Phys.*, 11(21), 10,791–10,801, doi:10.5194/acp-11-10791-2011.
- Lee, C., R. V. Martin, A. van Donkelaar, G. O’Byrne, N. Krotkov, A. Richter, L. G. Huey, and J. S. Holloway (2009), Retrieval of vertical columns of sulfur dioxide from SCIAMACHY and OMI: Air mass factor algorithm development, validation, and error analysis, *J. Geophys. Res.*, 114, D22303, doi:10.1029/2009JD012123.
- Lee, C., R. V. Martin, A. van Donkelaar, H. Lee, R. R. Dickerson, J. C. Hains, N. Krotkov, A. Richter, K. Vinnikov, and J. J. Schwab (2011), SO<sub>2</sub> emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, *J. Geophys. Res.*, 116, D06304, doi:10.1029/2010JD014758.
- Levy, R. C., L. A. Remer, R. G. Kleidman, S. Mattoo, C. Ichoku, R. Kahn, and T. F. Eck (2010), Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, *Atmos. Chem. Phys.*, 10(21), 10,399–10,420, doi:10.5194/acp-10-10399-2010.
- Mathur, R., and R. L. Dennis (2003), Seasonal and annual modeling of reduced nitrogen compounds over the eastern United States: Emissions, ambient levels, and deposition amounts, *J. Geophys. Res.*, 108(D15), 4481, doi:10.1029/2002JD002794.
- O’Neill, N. T., T. F. Eck, B. N. Holben, A. Smirnov, O. Dubovik, and A. Royer (2001a), Bimodal size distribution influences on the variation of Angstrom derivatives in spectral and optical depth space, *J. Geophys. Res.*, 106(D9), 9787–9806, doi:10.1029/2000JD900245.
- O’Neill, N. T., O. Dubovik, and T. F. Eck (2001b), Modified angstrom exponent for the characterization of submicrometer aerosols, *Appl. Optics*, 40(15), 2368–2375, doi:10.1364/ao.40.002368.
- Paasonen, P., et al. (2013), Warming-induced increase in aerosol number concentration likely to moderate climate change, *Nat. Geosci.*, 6, 438–442, doi:10.1038/ngeo1800.
- Palmer, P. L., et al. (2006), Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, *J. Geophys. Res.*, 111, D12315, doi:10.1029/2005JD006689.
- Pinder, R. W., J. T. Walker, J. O. Bash, K. E. Cady-Pereira, D. K. Henze, M. Luo, G. B. Osterman, and M. W. Shephard (2011), Quantifying spatial and seasonal variability in atmospheric ammonia with in situ and space-based observations, *Geophys. Res. Lett.*, 38, L04802, doi:10.1029/2010GL046146.
- Pirjola, L., K. E. J. Lehtinen, H.-C. Hansson, and M. Kulmala (2004), How important is nucleation in regional/global modelling, *Geophys. Res. Lett.*, 31, L12109, doi:10.1029/2004GL019525.
- Pryor, S. C., R. J. Barthelme, L. L. Sørensen, and B. Jensen (2001), Ammonia concentrations and fluxes over a forest in the midwestern USA, *Atmos. Environ.*, 35(32), 5645–5656, doi:10.1016/S1352-2310(01)00259-x.
- Pryor, S. C., A. M. Spaulding, and R. J. Barthelme (2010), New particle formation in the Midwestern USA: Event characteristics, meteorological context and vertical profiles, *Atmos. Environ.*, 44(35), 4413–4425, doi:10.1016/j.atmosenv.2010.07.045.
- Pryor, S. C., R. J. Barthelme, L. L. Sørensen, J. G. McGrath, P. Hopke, and T. Petäjä (2011), Spatial and vertical extent of nucleation events in the Midwestern USA: Insights from the Nucleation In Forests (NIFTy) experiment, *Atmos. Chem. Phys.*, 11(4), 1641–1657, doi:10.5194/acp-11-1641-2011.
- Remer, L. A., et al. (2005), The MODIS aerosol algorithm, products, and validation, *J. Atmos. Sci.*, 62(4), 947–973, doi:10.1175/jas3385.1.
- Riipinen, I., et al. (2011), Organic condensation: A vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations, *Atmos. Chem. Phys.*, 11(8), 3865–3878, doi:10.5194/acp-11-3865-2011.
- Russell, P. B., R. W. Bergstrom, Y. Shinozuka, A. D. Clarke, P. F. DeCarlo, J. L. Jimenez, J. M. Livingston, J. Redemann, O. Dubovik, and A. Strawa (2010), Absorption Angstrom exponent in AERONET and related data as an indicator of aerosol composition, *Atmos. Chem. Phys.*, 10(3), 1155–1169, doi:10.5194/acp-10-1155-2010.
- Shephard, M. W., S. A. Clough, V. H. Payne, W. L. Smith, S. Kireev, and K. E. Cady-Pereira (2009), Performance of the line-by-line radiative transfer model (LBLRTM) for temperature and species retrievals: IASI case studies from JAIVEx, *Atmos. Chem. Phys.*, 9(19), 7397–7417, doi:10.5194/acp-9-7397-2009.
- Shephard, M. W., et al. (2011), TES ammonia retrieval strategy and global observations of the spatial and seasonal variability of ammonia, *Atmos. Chem. Phys.*, 11(20), 10,743–10,763, doi:10.5194/acp-11-10743-2011.



- Sinyuk, A., B. N. Holben, A. Smirnov, T. F. Eck, I. Slutsker, J. S. Schafer, D. M. Giles, and M. Sorokin (2012), Assessment of error in aerosol optical depth measured by AERONET due to aerosol forward scattering, *Geophys. Res. Lett.*, *39*, L23806, doi:10.1029/2012GL053894.
- Smith, J. N., K. F. Moore, F. L. Eisele, D. Voisin, A. K. Ghimire, H. Sakurai, and P. H. McMurry (2005), Chemical composition of atmospheric nanoparticles during nucleation events in Atlanta, *J. Geophys. Res.*, *110*, D22S03, doi:10.1029/2005JD005912.
- Spracklen, D. V., K. S. Carslaw, M. Kulmala, V.-M. Kerminen, G. W. Mann, and S.-L. Sihto (2006), The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales, *Atmos. Chem. Phys.*, *6*, 5631–5648.
- Spracklen, D. V., et al. (2008), Contribution of particle formation to global cloud condensation nuclei concentrations, *Geophys. Res. Lett.*, *35*, L06808, doi:10.1029/2007GL033038.
- Spracklen, D. V., et al. (2010), Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation, *Atmos. Chem. Phys.*, *10*(10), 4775–4793, doi:10.5194/acp-10-4775-2010.
- Yu, H., L. Remer, M. Chin, H. Bian, Q. Tan, T. Yuan, and Y. Zhang (2012), Aerosols from overseas rival domestic emissions over North America, *Science*, *337*(6094), 566–569.
- Zhang, R. Y., A. Khalizov, L. Wang, M. Hu, and W. Xu (2011), Nucleation and growth of nanoparticles in the atmosphere, *Chem. Rev.*, *112*(3), 1957–2011, doi:10.1021/cr2001756.