Reprint 2016-4

Metamodeling of Droplet Activation for Global Climate Models

Daniel Rothenberg and Chien Wang

Reprinted with permission from *Journal of the Atmospheric Sciences*, 73, 1255–1272.

The **MIT Joint Program on the Science and Policy of Global Change** combines cutting-edge scientific research with independent policy analysis to provide a solid foundation for the public and private decisions needed to mitigate and adapt to unavoidable global environmental changes. Being data-driven, the Joint Program uses extensive Earth system and economic data and models to produce quantitative analysis and predictions of the risks of climate change and the challenges of limiting human influence on the environment—essential knowledge for the international dialogue toward a global response to climate change.

To this end, the Joint Program brings together an interdisciplinary group from two established MIT research centers: the **Center for Global Change Science (CGCS)** and the **Center for Energy and Environmental Policy Research (CEEPR)**. These two centers—along with collaborators from the Marine Biology Laboratory (MBL) at Woods Hole and short- and long-term visitors—provide the united vision needed to solve global challenges.

At the heart of much of the program's work lies MIT's Integrated Global System Model. Through this integrated model, the program seeks to discover new interactions among natural and human climate system components; objectively assess uncertainty in economic and climate projections; critically and quantitatively analyze environmental management and policy proposals; understand complex connections among the many forces that will shape our future; and improve methods to model, monitor and verify greenhouse gas emissions and climatic impacts.

This reprint is intended to communicate research results and improve public understanding of global environment and energy challenges, thereby contributing to informed debate about climate change and the economic and social implications of policy alternatives.

—Ronald G. Prinn and John M. Reilly,
Joint Program Co-Directors

MIT Joint Program on the Science and Policy of Global Change

Massachusetts Institute of Technology
77 Massachusetts Ave., E19-411
Cambridge MA 02139-4307 (USA)

T (617) 253-7492 **F** (617) 253-9845
globalchange@mit.edu
http://globalchange.mit.edu

Metamodeling of Droplet Activation for Global Climate Models

DANIEL ROTHENBERG AND CHIEN WANG

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 4 August 2015, in final form 13 November 2015)

ABSTRACT

The nucleation of cloud droplets from the ambient aerosol is a critical physical process that must be resolved for global models to faithfully predict aerosol–cloud interactions and aerosol indirect effects on climate. To better represent droplet nucleation from a complex, multimodal, and multicomponent aerosol population within the context of a global model, a new metamodeling framework is applied to derive an efficient and accurate activation parameterization. The framework applies polynomial chaos expansion to a detailed parcel model in order to derive an emulator that maps thermodynamic and aerosol parameters to the supersaturation maximum achieved in an adiabatically ascending parcel and can be used to diagnose droplet number from a single lognormal aerosol mode. The emulator requires much less computational time to build, store, and evaluate than a high-dimensional lookup table. Compared to large sample sets from the detailed parcel model, the relative error in the predicted supersaturation maximum and activated droplet number computed with the best emulator is $-0.6\% \pm 9.9\%$ and $0.8\% \pm 17.8\%$ (one standard deviation), respectively. On average, the emulators constructed here are as accurate and between 10 and 17 times faster than a leading physically based activation parameterization. Because the underlying parcel model being emulated resolves size-dependent droplet growth factors, the emulator captures kinetic limitations on activation. The results discussed in this work suggest that this metamodeling framework can be extended to accurately account for the detailed activation of a complex aerosol population in an arbitrary coupled global aerosol–climate model.

1. Introduction

Interactions between aerosol and clouds yield one of the largest sources of uncertainty in understanding climate and future climate change on regional and global scales (Boucher et al. 2013). Within Earth's atmosphere, homogeneous liquid water droplet formation is not thermodynamically favorable (Pruppacher and Klett 1997); instead, the pathway to nucleating cloud droplets is aided by the presence of ambient aerosol, a subset of which possess physical and chemical characteristics that allow them to serve as cloud condensation nuclei (CCN). These CCN provide a linkage between the physiochemical processes of atmospheric particles and cloud microphysics.

Changes in the background aerosol population can directly affect the properties of a nascent cloud droplet population. For instance, holding liquid water content constant, an increase in the number of CCN would tend to increase the total cloud droplet number concentration (the ''Twomey'' effect) while necessarily reducing the average size of the droplets (Twomey 1974). Such a change could enhance a cloud's albedo, an effect that could be further amplified through microphysical feedbacks since smaller droplets impede the production of drizzle and thus lengthen cloud lifetime (Albrecht 1989). Mechanisms whereby aerosol influence the properties of clouds (and ultimately climate) are generally known as ''aerosol indirect effects'' (Haywood and Boucher 2000; Lohmann and Feichter 2005) and provide a path for changes in the ambient aerosol to produce cascading effects up to progressively larger scales of atmospheric motion (e.g., Wang 2005; Ekman et al. 2011; Morrison et al. 2011; Tao et al. 2011; Fan et al. 2012; Altaratz et al. 2014).

Aerosol indirect effects can either warm or cool the climate, but they all fundamentally depend on a subset of the ambient aerosols that function as CCN. The theory describing the dependency of cloud droplet nucleation

Corresponding author address: Daniel Rothenberg, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.
E-mail: darothen@mit.edu

DOI: 10.1175/JAS-D-15-0223.1

(also known as aerosol activation) on CCN availability and ambient aerosol has been rigorously developed using adiabatic and entraining parcel theory (Seinfeld and Pandis 2006; Pruppacher and Klett 1997) and depends on details of the heterogeneous chemical composition, number, size distribution(s), and mixing state of the background aerosol (Mcfiggans et al. 2006) as well as local meteorology (Morales and Nenes 2010). Under polluted conditions, effects relating to chemical composition could produce a climatic effect as large as the basic Twomey effect (Nenes et al. 2002; Lance et al. 2004).

The development of activation parameterizations was pioneered by Twomey (1959) and Squires and Twomey (1961), who derived a relationship between the number of activated particles and the environmental supersaturation based on an aerosol size distribution approximated by a power law. Ghan et al. (2011) presented a thorough overview of subsequent developments over the past five decades and an intercomparison of several modern parameterizations. However, there is still an active effort to improve these parameterizations, as they are increasingly called upon to mediate between ever more complex aerosol models and the climate models to which they are coupled. For instance, the parameterization initially developed by Nenes and Seinfeld (2003) has seen continuous development, including modifications to handle condensation onto insoluble but wettable particles using adsorption activation theory (Kumar et al. 2009), environmental entrainment (Barahona and Nenes 2007), and numerical improvement of the population-splitting technique (Barahona et al. 2010; Morales Betancourt and Nenes 2014). Similarly, Ghan et al. (2011) extended the parameterization of Abdul-Razzak and Ghan (2000) to account for nonunity values of the accommodation coefficient a_c. Beyond idealized testing and droplet closure studies (Meskhidze 2005; Fountoukis et al. 2007), these modern parameterizations have been implemented in coupled climate–aerosol models such as the Community Earth System Model (CESM) to predict online cloud droplet number concentrations, where they have been shown to correct biases in global-average cloud droplet number concentrations and improve agreement with cloud properties measured from satelliteborne instruments (Gantt et al. 2014). Furthermore, adjoints of these parameterizations have been derived and coupled to chemical transport and global models in order to study the sensitivities of cloud droplet number to aerosol, chemical, and microphysical factors (Karydis et al. 2012; Moore et al. 2013).

Additionally, following the original integral/geometric approach by Twomey (1959), analytical representations of supersaturation evolution from adiabatic parcel theory have been progressively generalized to relate aerosol distributions to activation kinetics (Cohard et al. 1998; Khvorostyanov and Curry 2006, 2008; Shipway and Abel 2010; Shipway 2015). Although fundamentally analytical parameterizations, schemes of this class typically must rely on expensive numerical operations, such as in the evaluation of hypergeometric functions and iterative loops.

While most of these recent efforts toward improving activation parameterizations have focused on building highly generalized, ''physically based'' tools, there is still an application for other parameterization approaches. Saleeby and Cotton (2004) parameterized droplet nucleation for a cloud-resolving model, the Regional Atmospheric Modeling System (RAMS), by constructing a four-dimensional lookup table based on temperature, vertical velocity, aerosol number concentration, and the median radius of a lognormal aerosol mode with assumed chemical composition. Ward et al. (2010) added a fifth dimension representing chemical composition via aerosol hygroscopicity [following κ–Köhler theory (Petters and Kreidenweis 2007)] to the lookup table and later generalized this dimension to aerosol soluble fraction (Saleeby and van den Heever 2013). Constructing lookup tables of detailed parcel model results can be considered a form of model emulation combining a cache of known results and local polynomial (linear) approximation.

As the degrees of freedom and number of parameters describing a given aerosol population in a model increase, the burden of saving enough known points to interpolate through the parameter space via lookup table to some reasonable accuracy increases algebraically. For instance, the CESM features a modal aerosol population with three predefined, internally mixed lognormal modes, each with a fixed geometric standard deviation (Liu et al. 2012). Each mode is uniquely described by two moments (total number and total mass concentration) and the chemical composition of the mode by a single prognostic hygroscopicity term. Thus, the entire aerosol population has $N = 9$ degrees of freedom—too many to build a lookup table of activation statistics. Physically based parameterizations were designed to accommodate these sorts of arbitrary mixtures of aerosol but have a tendency to systematically underestimate activated fractions and subsequently cloud droplet number (Simpson et al. 2014). This is because of the parameterizations' use of a set of assumptions that become increasingly likely to be violated as the aerosol population becomes more complex, specifically 1) that protodroplets grow in equilibrium with environmental changes in relative humidity and 2) that there are no kinetic or inertial limitations to droplet growth. The

presence of giant CCN (Barahona et al. 2010) and weak updrafts or excessively polluted conditions (Nenes et al. 2001) exacerbates this problem.

The goal of this study is to apply a surrogate modeling or emulation technique commonly used in the uncertainty quantification literature to a detailed parcel model capable of describing aerosol activation; this yields an efficient parameterization optimized for the high-dimensional parameter space affecting droplet nucleation in coupled aerosol–climate model. In essence, employing the derived emulator as an activation parameterization would be akin to directly coupling a detailed parcel model to a global model. Such a parameterization would be directly physically based but rely on fewer assumptions that affect the condensational growth of aerosol into CCN. However, it would also incorporate the efficiency of a lookup table, since the emulator would be designed to require a scarce amount of cached information and to be computationally cheap to evaluate. In this way, the emulator would improve upon the framework of a lookup table and be extensible to a very high-dimensional parameter space and thus be compatible with aerosol–climate models of increasing complexity.

Additionally, this study aims to better understand which parameters and inputs into droplet nucleation calculations are key to determining the resulting activated number concentration. Morales Betancourt and Nenes (2014) supplemented traditional error metrics by computing local sensitivities of the number concentration activated to key aerosol size distribution parameters using a detailed parcel model and comparing them to the adjoint of their parameterization. This study applies a related approach instead using global sensitivity analysis, which is suitable for identifying how uncertainty in inputs and parameters contributes to uncertainty in a model response. This analysis has not previously been applied to droplet activation and can provide additional metrics for evaluating parameterizations and their potential biases.

Section 2 describes the parcel model and the probabilistic collocation method (PCM) used to build its emulator. Section 3 presents results from applying the PCM to build a parcel model emulator designed to simulate the activation of a single lognormal aerosol mode under a wide variety of background environments and compares the new emulator to existing activation parameterizations. Section 4 motivates an extension of the technique to an emulator suitable of mediating aerosol activation in a coupled aerosol–climate model.

2. Methodology

a. Parcel model

Adiabatic parcel models are commonly used to study droplet activation and its sensitivity to factors such as environmental conditions and ambient aerosol properties. For this work, a novel parcel model based on previous studies (Leaitch et al. 1986; Nenes et al. 2001; Seinfeld and Pandis 2006) was designed and implemented to accommodate diverse, chemically heterogeneous, and polydisperse aerosol populations. The model simulates droplet growth on the initial aerosol population due to condensation within a constant-speed adiabatic updraft.

Although an arbitrary aerosol size distribution function can be supplied as an input to the model, for the purposes of this study the initial aerosol distribution is assumed to be lognormal and described by the equation

$$n_N(r) = \frac{dN}{d\ln r} = \frac{N_t}{\sqrt{2\pi}\ln\sigma_g}\exp\left[-\frac{\ln^2(r/\mu_g)}{2\ln^2\sigma_g}\right], \quad (1)$$

where the parameter set (N_t, μ_g, σ_g) corresponds, respectively, to the total aerosol number concentration, the geometric mean radius, and the geometric standard deviation of the distribution. Within the model, this distribution is discretized into 200 size bins equally spaced over the logarithm of particle radius r and covers the size range $(\min[0.1\,\text{nm}, \mu_g/10\sigma_g], \mu_g \times 10\sigma_g)$. The mean radius in each bin grows as a result of condensation so that the activation of wetted aerosol into droplets is calculated in a Lagrangian sense. To relate size-dependent droplet growth to its embedded aerosol's chemical composition, each bin is prescribed a hygroscopicity following κ–Köhler theory (Petters and Kreidenweis 2007).

To simulate droplet activation, the parcel model first computes an equilibrium wet-size distribution from the given initial aerosol population and initial environmental temperature, pressure, and relative humidity. Then, a set of conservation equations that describe the evolution of the parcel temperature, supersaturation, liquid/vapor water content, and pressure are integrated forward in time using a solver suitable for stiff systems [variable-coefficient ordinary differential equation solver (VODE); Brown et al. (1989)]. The complete system of equations and further details on the parcel model can be found in appendix A.

b. Polynomial chaos expansion

We construct an emulator of the parcel model in order to assess droplet activation by applying the probabilistic collocation method (PCM; Tatang et al. 1997). The PCM maps a set of input parameters to an output from the parcel model by building a response surface using a polynomial chaos expansion. The polynomial that results from this process is a computationally efficient, high-fidelity reproduction of the detailed parcel model

simulation. Although often used for conducting global sensitivity analyses (Pan et al. 1997; Calbó et al. 1998; Mayer et al. 2000; Lucas and Prinn 2005; Anttila and Kerminen 2007) chaos expansion-based emulators have also been used to build deterministic parameterizations (Cohen and Prinn 2011). To apply and build the chaos expansions discussed here, the open-source Design Analysis Kit for Optimization and Terascale Applications (DAKOTA; Adams et al. 2014), version 6.1, was used, which automates the sampling of the PCM collocation points and the computation of the coefficients of the polynomial chaos expansion given a user-generated interface to a numerical model (the parcel model described in section 2a) and a description of the inputs and outputs to and from that interface.

A review of the theoretical basis of polynomial chaos expansion and its potential applications is provided by Sudret (2008); here, we highlight the important details of the technique as applied via PCM for the benefit of the reader. PCM is a nonintrusive polynomial chaos expansion technique; rather than require complex, significant modifications to the model being emulated, PCM instead considers the model to be a black box and constructs a map from an input parameter space to the model output parameter space. To accomplish this, PCM recasts the input parameters to a model as a set of M independent random variables, $\mathbf{X} = X_1, \ldots, X_M$, each with an associated probability density function. For each input in $\mathbf{X}$, the associated PDF is used as a weighting function to derive an orthogonal polynomial that adds to the bases for the polynomial chaos expansion ϕ_j. Using a finite number of these bases, the chaos expansion for a given model response R is then

$$R \approx \sum_{j=0}^{P} \alpha_j \phi_j(\mathbf{X}). \tag{2}$$

The complete basis of polynomials up to a fixed total-order p is retained in the expansions computed here. For such a total-order expansion, Eq. (2) has $N_t = P + 1 = (M + p)!/(M!p!)$ terms as it contains each of the $p + 1$ orthogonal basis polynomials for each input parameter. PCM provides an experimental design for determining the coefficients of the expansion α_j by evaluating the model response for a set of N_s total input parameter sets, $\mathbf{X}^1, \ldots, \mathbf{X}^{N_s}$, corresponding to the roots of ϕ_j and solving a regression problem

$$\mathbf{\Phi}\boldsymbol{\alpha} = \mathbf{R}, \tag{3}$$

where $\mathbf{R}$ is the vector of model responses, $\boldsymbol{\alpha}$ is the vector of expansion coefficients, and the matrix $\mathbf{\Phi}$ contains rows for each of the polynomial terms ϕ_j evaluated for a given input parameter set $\mathbf{X}^j$.

A practical consideration in applying the PCM to a particular problem is what subset of the N_s potential points to use in solving for the coefficients. In general, there exists a full factorial design of size $N_s = (p + 1)^M$ available for use (all the roots of the orthogonal basis polynomials for all inputs). However, for even moderately sized p and M, the number of potential model evaluations grows very rapidly. In our application we choose a subset of N_s' parameter sets when applying the PCM by using two rules of thumb:

1) choose parameter sets with roots closest to the origin (Sudret 2008) and
2) cross validate the regression result using $3N_t$ parameter sets chosen according to rule 1.

These rules will always produce an overdetermined system for Eq. (3). The accuracy of the resulting emulators derived in this study were not sensitive to increasing N_s', and $3N_t$ does not produce an excessive number of required parcel model simulations. Three different techniques were tested for solving this system: typical linear regression by ordinary least squares (OLS), least angle regression (LARS; Efron et al. 2004), and least absolute shrinkage and selection operator (LASSO; Tibshirani 2011). Both LARS and LASSO involve computing

$$\boldsymbol{\alpha} = \arg\min \|\mathbf{\Phi}\boldsymbol{\alpha} - \mathbf{R}\|_{l_2}^2 \quad \text{such that} \quad \|\boldsymbol{\alpha}\|_{l_1} \le \tau \tag{4}$$

in an iterative, greedy fashion with the potential to yield sparse solutions with some coefficients $\alpha_j = 0$. This would be desirable for high-order chaos expansions for many input parameters, as it would reduce the number of coefficients necessary to save for reusing the expansion as an emulator. Additionally, this greedy characteristic helps to avoid overfitting the chaos expansions. Each time a potential term is added to the trial expansion, an error estimate is calculated based on leave-one-out sampling (Blatman and Sudret 2011); if the error estimate increases, the potential term is rejected. Inspecting the resulting terms gives a metric to compare the OLS-derived chaos expansion. The same error calculation using leave-one-out sampling can be applied using the larger, independent sampling dataset used to evaluate the chaos expansions in section 2a for each higher-order chaos expansion to help identify when overfitting is occurring.

c. *Emulation of parcel model*

The PCM was applied to emulate the activation of a single, lognormal aerosol mode embedded in a constant-speed adiabatic updraft as simulated by the parcel model described in section 2a. Specifically, the model was used to predict the maximum supersaturation $S_{\max}$

TABLE 1. Input parameters and bounds used in computing chaos expansion for emulating droplet activation from a single, lognormal aerosol mode embedded in a constant-speed updraft.

Symbol	Name	Units	Bounds
$\log_{10}N$	Mode number concentration	$\log_{10}\,\mathrm{cm}^{-3}$	[1, 4]
$\log_{10}\mu_g$	Mode geometric mean radius	$\log_{10}\,\mu\mathrm{m}$	[−3.0, 1.0]
σ_g	Mode standard deviation	—	[1.2, 3.0]
κ	Mode hygroscopicity	—	[0.0, 1.2]
$\log_{10}V$	Updraft velocity	$\log_{10}\mathrm{ms}^{-1}$	[−2.0, 1.0]
T	Air temperature	K	[240, 310]
P	Air pressure	Pa	[50 000, 105 000]
a_c	Accommodation coefficient	—	[0.1, 1.0]

given aerosol of different lognormal size distributions and hygroscopicities for different environmental and thermodynamic conditions. The mechanics of the PCM—and polynomial chaos expansion more generally—permit the use of arbitrary PDFs to describe the input parameters over their physically relevant values. In this application, uniform distributions were chosen to emphasize that the derived chaos expansion should perform equally well anywhere within the input parameter space. Each uniform distribution is defined by minimum and maximum permissible bounds for each input parameter, a and b, such that its probability distribution is just given as $f(x) = 1/(b-a)$ for $a < x < b$.

To utilize the PCM, the uniform distribution for each parameter must be rescaled to the range $[-1, 1]$. This produces a new set of random variables for each parameter X_i:

$$Z_i = \frac{2(X_i - a_i)}{b_i - a_i} - 1. \tag{5}$$

The orthogonal polynomials used in the basis of the chaos expansion that correspond to a uniform PDF over the interval $[-1, 1]$ are the canonical Legendre polynomials that follow the three-term recurrence relation:

$$P_0(Z) = 1, \tag{6}$$

$$P_1(Z) = Z, \quad \text{and} \tag{7}$$

$$P_{n+1}(Z) = \frac{(2n+1)ZP_n(Z) - nP_{n-1}(Z)}{n+1}. \tag{8}$$

The roots of these Legendre polynomials can be inverted using Eq. (5) to determine values in the original, physical parameter space to use in sampling the parcel model.

The bounds for the physical parameters supplied to the PCM were chosen in order to characterize activation near cloud base (Table 1). The logarithm of several variables (aerosol number concentration, aerosol geometric mean radius, and updraft velocity) is used because the supersaturation maximum computed by the parcel model is sensitive to changes in these parameters over several orders of magnitude. Updraft velocity is permitted to vary between 0.01 and 10.0 $\mathrm{m\,s}^{-1}$; over this range (which covers a spectrum from weakly convecting, stratiform clouds to strong, deeply convecting ones) and the range of aerosol number concentration (which includes clean and very polluted regimes), activated fraction can range from virtually nothing to complete activation of the entire aerosol population. The lower bound of updraft speeds considered here is less than the minimal value allowed in many climate models (Golaz et al. 2011). The aerosol mode geometric mean radii μ_g span a variety of smaller Aitken-type modes to large, coarse aerosol modes and potentially giant CCN (Seinfeld and Pandis 2006). The mode geometric standard deviation is fixed in some global model aerosol schemes (e.g., Kim et al. 2008; Liu et al. 2012) and the range chosen here covers many potential values. Hygroscopicity values are based on Table 1 of Petters and Kreidenweis (2007) and span values for materials ranging from organic aerosol to highly hygroscopic salts. The accommodation coefficient was limited to a globally representative range based on observations of CCN activation kinetics from many campaigns (Raatikainen et al. 2013). Temperature and pressure ranges were chosen to reflect typical lower-troposphere values. Note that the bounds on the parameters considered here are expanded from those considered by Ghan et al. (2011) in their intercomparison of activation schemes.

Many parameterizations of droplet nucleation diagnose activation directly by applying equilibrium Köhler theory. To do this, the maximum supersaturation achieved by a cloudy parcel is used as a threshold; particles with a Köhler-predicted critical supersaturation lower than this maximum environmental supersaturation are considered to be activated. However, physically, for a droplet to activate it must grow beyond a critical size corresponding to this critical supersaturation. Because of kinetic limitations on droplet growth, this may

not be realized for droplets growing on very large CCN (Nenes et al. 2001). Ghan et al. (2011) suggests particles with radius larger than 0.1 μm or those whose critical supersaturations are close to the environmental maximum supersaturation are likely to suffer from this effect. By directly considering a detailed parcel model, the emulators constructed here consider kinetic limitations on droplet growth and their feedback on the evolving parcel supersaturation. In existing, physically based parameterizations in the literature, an instantaneous growth-rate assumption must be applied. This assumption causes parameterizations to underpredict supersaturation maximum because instantaneous growth will tend to condense water from the vapor phase too quickly and release surplus latent heat, which suppresses the increase of the supersaturation.

Because the computed supersaturation maximum in a parcel model activation simulation can also vary over several orders of magnitude, we use $\log_{10}(S_{\max})$ as the response function emulated by the PCM. However, in order to apply this transform to the response function, it must be assumed that the cloudy parcel always supersaturates with respect to water vapor (i.e., $S_{\max} > 0$). To ensure this, all simulations performed during sampling by the PCM start with an aerosol population equilibrated to 100% relative humidity and an initial environmental supersaturation of 0. Many existing parameterizations in the literature implicitly make this same assumption by representing the aerosol population with respect to a coordinate derived from the critical supersaturation for a given size [Ghan et al. (2011); Eqs. (12)–(17)]; in this case the integral over the size distribution spans $0 \leq S \leq S_{\max}$ and, thus, considers the same situation with respect to the growth of the nascent droplet population as Eq. (3) of Ghan et al. (2011).

For the eight-parameter input space governing single-mode activation considered here, the third- and fourth-order chaos expansions produced by the PCM have 165 and 495 terms, respectively. The number of terms is equivalent to the number of coefficients one must store in order to reuse a given chaos expansion. This small memory footprint affords chaos expansions a huge advantage over similar parameterizations based on detailed lookup tables. An isotropic lookup table with M parameters and n sample points for each parameter would require n^M values to be stored—a value that for even small numbers of parameters can be several orders of magnitude larger than even a high-order chaos expansion. A more detailed description of how the chaos expansions are saved and later evaluated is given in appendix B.

The parcel model emulator yields $\log_{10}(S_{\max})$ as a function of an input parameter set drawing from the terms defined in Table 1:

$$\log_{10}(S_{\max}) = f(\log_{10}N, \log_{10}\mu_g, \sigma_g, \kappa, \log_{10}V, T, P, a_c). \tag{9}$$

From this value of $\log_{10}(S_{\max})$ the number concentration of cloud droplets activated, N_{act}, can be obtained by integrating over the original lognormal aerosol size distribution reexpressed as a function of critical supersaturation rather than droplet radius (Ghan et al. 2011), yielding the expression

$$N_{\text{act}} = \frac{N}{2}\left\{1 - \text{erf}\left[2\ln\left(\frac{S_m}{S_{\max}}\right)\Big/(3\sqrt{2}\ln\sigma_g)\right]\right\}, \tag{10}$$

where S_m is the critical supersaturation for the geometric mean radius μ_g.

d. Global sensitivity analysis

We supplement the assessment of our new droplet activation emulator by calculating a set of global sensitivity metrics not previously applied to this problem. The method deployed here is a variance-based decomposition, which seeks to assign uncertainty in a model response to uncertainty in both individual model input parameters and their interactions with one another. Two different quantities, called Sobol' indices (Sobol' 2001), are produced by this method: main (S_i) and total (T_i) effect indices. Sobol' indices can be used to rank the relative importance of model inputs in influencing its response and for identifying potential inputs that are unimportant and also candidates to be held fixed without grossly biasing the accuracy of a model emulator (Sobol' 2001).

The main effect index indicates what fraction of the uncertainty in a given model response R is attributable to a single member of the model parameter set X_i by comparing the variance of the model response conditioned on X_i against the total variance in R. That is,

$$S_i = \frac{\text{Var}_{X_i}(E[R\,|\,X_i])}{\text{Var}(R)}. \tag{11}$$

This is in contrast with the total effect index, which instead compares the variance of R conditioned on all the input parameters save for X_i (notated as $X_{\sim i}$). The index T_i quantifies the variance of R attributable to X_i and the sum of its interaction with other input terms. Similar to the main effect index,

$$T_i = \frac{\text{Var}(R) - \text{Var}(E[R\,|\,X_{\sim i}])}{\text{Var}(R)}. \tag{12}$$

Sudret (2008) derives alternative equations that further clarify the meanings of these terms in the context of studying model emulators. Critically, these terms can be

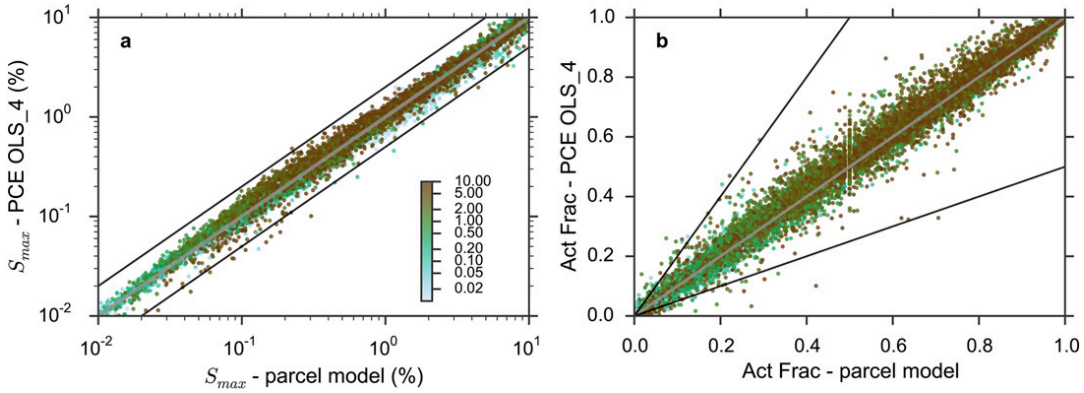

FIG. 1. One–one plot comparing (a) predicted supersaturation maximum and (b) diagnosed equilibrium droplet activated fraction between parcel model and a polynomial chaos expansion of order $p = 4$, with coefficients computed using ordinary least squares. Black lines denote a factor-of-2 difference between predicted values using parcel model and those computed with the parameterization. Glyph shading denotes updraft velocity V with corresponding scale in (a).

expressed as multidimensional integrals over model input parameters that can be approximated via Monte Carlo or other sampling techniques, although it is very computationally expensive to do so. We adopt the column swap-out sampling method of Weirs et al. (2012), which combines Latin hypercube sampling with perturbed combinations of parameters to efficiently approximate S_i and T_i [see also Saltelli et al. (2010)]. However, the sampling procedure employed still requires a large number of model evaluations; for an initial n-size sample of the M parameters being studied, the method requires $(2 + M) \times n$ evaluations of the full-complexity model. We found that the computed S_i and T_i converged to stable values for the parcel model (and other parameterizations studied here) for $n \sim O(10^3)$ and used $n = 1280$ to derive the values reported here.

Although the sampling procedure can be repeated for the emulators derived via polynomial chaos expansion, the orthogonality of the basis terms that constitute each expansion lends itself to a more direct computation of S_i and T_i. Following Sudret (2008), we compute these indices directly from the coefficients of the derived chaos expansions. The sampling technique used to derive Sobol' indices for the parcel model, applied to the chaos expansions, produces similar estimates to those computed from the coefficients.

3. Results

a. Evaluation of emulators

To assess the performance of the emulator, two sets of $n =$ 10 000 samples were drawn using maximum Latin hypercube sampling from the parameter space defined in Table 1. This randomized design ensured that representative, equal numbers of samples were drawn from across the multidimensional parameter space. In the first set, variables whose logarithms were used to build the emulator were sampled in logarithmic space; in the second set, these variables were transformed back to their original values (e.g., from $\log_{10}N$ to N) before the sample was constructed. The two independent sets were blended together to assess the emulator. This helps ensure that both very high and very low values of the log transformed are thoroughly represented within the sample. The set of sample parameter sets were run through all the derived chaos expansions of all orders, as well as the detailed parcel model as a reference benchmark for activation dynamics.

Figure 1 illustrates the performance of a fourth-order expansion whose coefficients were derived using ordinary least squares. The large range of initial temperatures, pressures, aerosol populations, and updraft speeds sampled here leads to a very large range of supersaturation maxima achieved by the ascending parcel. Weaker updraft speeds are generally associated with lower maximum supersaturations and corresponding to lower aerosol activated fraction; the opposite is true when strong updrafts are present, although there are some cases where a strong updraft activates a small fraction of aerosol. This typically occurs when initial aerosol size distribution is shifted toward larger radii and under polluted conditions with aerosol number concentrations greater than 3000 cm^{-3}. However, over the large parameter space sampled, the chaos expansion accurately reproduces the parcel model's determination of $S_{\max}$ and corresponding activated fraction. This is even true for predictions of small $S_{\max}$, which could potentially have a larger bias since the predicted error is expected to be uniform in $\log_{10}(S_{\max})$.

TABLE 2. Summary statistics for supersaturation maxima predicted by chaos expansions derived in this study compared to detailed parcel model calculations. The chaos expansions are organized by the method used to derive their coefficients and the expansion order in the two leftmost columns. (left to right) The reported statistics are the normalized root-mean-square error (NRMSE), coefficient of determination r^2, mean relative error (MRE), and standard deviation of the mean relative error (MRE std dev).

Method	Expansion order	NRMSE	r^2	MRE	MRE std dev
LASSO	2	0.292 766	0.877 600	−4.963 330	31.098 459
	3	0.193 247	0.946 670	−1.012 509	19.797 674
	4	0.112 323	0.981 983	3.941 698	13.396 632
	5	0.119 231	0.979 699	0.772 525	9.904 727
LARS	2	0.325 442	0.848 752	3.055 232	39.776 865
	3	0.250 264	0.910 559	−5.433 902	17.853 369
	4	0.104 401	0.984 435	−0.517 486	13.092 187
	5	0.135 634	0.973 729	−0.572 383	9.962 430
OLS	2	0.266 774	0.898 368	7.956 622	40.762 613
	3	0.220 034	0.930 861	−1.232 172	20.441 521
	4	0.201 934	0.941 768	4.243 836	14.483 854
	5	0.128 687	0.976 351	−0.259 165	10.430 909

For parameter sets leading to a large activated fraction of 0.8–1.0, the relative error of the chaos expansion (compared to the parcel model) rarely exceeds 5% and on average (for all activated fractions) is 5.7%. While the mean relative error in each activated fraction decile is close to 0, the standard deviation in the relative error tends to increase for the lower ones; the standard deviation in relative error decreases from 19.7% for activated fractions in the range 0.1–0.2 to 3.5% for those in the range 0.8–0.9. This suggests that there is a nonlinear component in the mapping from the input parameter space to the emulated maximum supersaturation and diagnosed droplet number concentration that is prevalent in the weak droplet activation regime; the predicted activated fraction is more sensitive to small changes in the input parameters in this regime than in others.

Increasing the order of the chaos expansion tends to improve the accuracy of the predicted $S_{\max}$, as recorded in Table 2. However, there is not much difference between the methods used to compute the coefficients of the expansion beyond expansion order. For example, for the fourth- and fifth-order expansions, the expansions perform equally well regardless of what method (OLS, LARS, or LASSO) was used to compute the coefficients when considering the mean and spread of the relative error to the parcel model reference simulations. In all cases, the chaos expansions produce very large r^2 values and small normalized fractional root-mean-square errors [RMSE/$\sum_{i=1}^{n}(X_i^2/n)$], which decrease as the order of the expansion increases.

These same statistics, computed for the diagnosed droplet number concentration given the predicted supersaturation maximum, are summarized in Table 3. Here, the trend is similar to before; increasing the order of the expansion tends to improve the accuracy of the diagnosed number concentration in terms of mean relative error and also tends to decrease spread around that value. Third-order expansions tend to produce more accurate results with respect to the mean relative error, but this is overshadowed by the fact that there is far more variance in their predicted values as indicated by the standard deviation of their relative errors, which are almost twice as large as those of the higher-order expansions.

b. *Comparison with other parameterizations*

We compare the performance of the chaos expansion-based emulators to two existing parameterizations from the literature. The scheme by Abdul-Razzak and Ghan (2000) (ARG)—which is widely used in global models—utilizes a psuedoanalytical solution to an integro-differential equation derived from the adiabatic parcel system with embedded aerosol growing via condensation. This is in contrast to the scheme by Morales Betancourt and Nenes (2014) (MBN), which is based on an iterative scheme to separate the aerosol population into subsets whose growth is inertially limited or not and uses this information to derive a maximum supersaturation for a given parcel system. The MBN scheme is generally more expensive to evaluate than the ARG scheme because of its iterative nature but is often more accurate owing to its consideration of the potentially important effect of large albeit unactivated aerosol particles (Simpson et al. 2014). In contrast with these schemes, our emulators simulate the activation process based on the explicit numerical solution obtained from a detailed parcel model, which is similar to the one used to build and evaluate the MBN scheme (e.g., Nenes and Seinfeld 2003).

The parameter sets used in section 3a were also used to compute droplet activation with the ARG and MBN parameterizations. The relative errors between

TABLE 3. As in Table 2, but for predicted droplet number concentration.

Method	Expansion order	NRMSE	r^2	MRE	MRE std dev
LASSO	2	0.165 043	0.929 439	5.697 161	49.813 356
	3	0.108 790	0.969 342	−0.930 886	20.270 639
	4	0.074 406	0.985 659	2.545 998	19.834 428
	5	0.058 872	0.991 022	1.441 665	14.497 129
LARS	2	0.167 168	0.927 611	3.524 631	39.031 227
	3	0.121 896	0.961 510	−0.030 537	34.613 813
	4	0.075 984	0.985 044	−0.280 769	16.430 850
	5	0.058 961	0.990 995	0.884 227	17.817 749
OLS	2	0.174 550	0.921 076	7.640 377	43.978 774
	3	0.125 045	0.959 496	0.380 943	28.562 715
	4	0.079 971	0.983 434	2.556 756	20.929 598
	5	0.061 762	0.990 119	1.295 857	17.903 202

the supersaturation maximum and droplet number predicted by these schemes and the chaos expansions compared to the parcel model are illustrated in Fig. 2. Both the ARG and MBN parameterizations are more accurate than the second- and third-order chaos expansions. The ARG scheme tends to underpredict the maximum supersaturation, which is consistent with previous investigations into its performance (Abdul-Razzak and Ghan 2000; Ghan et al. 2011; Simpson et al. 2014). This tends to produce a bias toward underprediction of droplet number. The MBN scheme tends to yield more accurate predictions of both maximum supersaturation and droplet number. However, both schemes are outperformed by the fourth- and fifth-order chaos expansions, both on average and in terms of the variance of the predictions; for instance, the OLS-derived fourth- and fifth-order expansions yield relative error in predicted S_{max} with a mean plus or minus one standard deviation of 4.2% ± 14.4% and −0.32% ± 10.4%, whereas the ARG and MBN schemes yield −13.7% ± 18.2% and

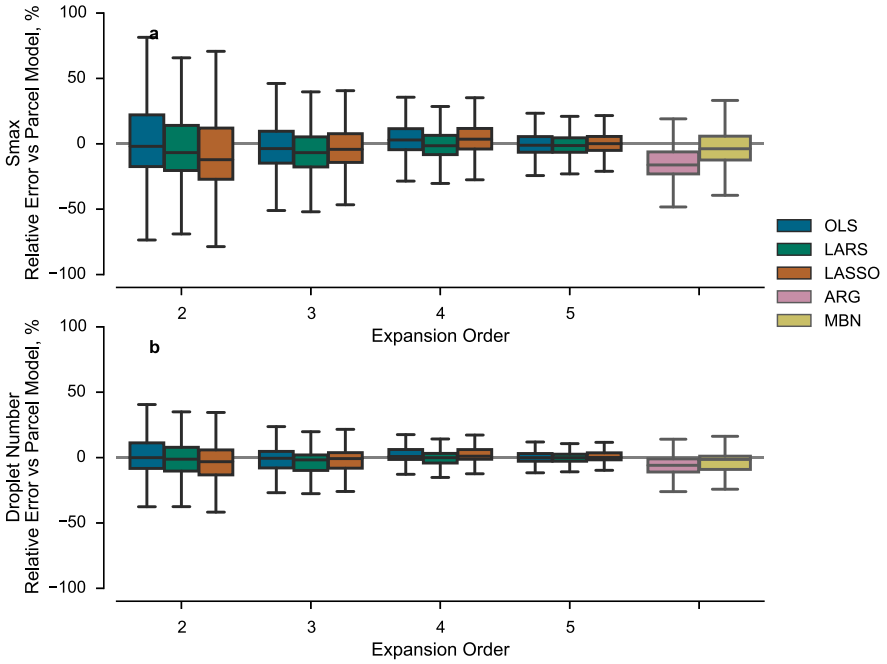

FIG. 2. Box plots illustrating mean relative error between (a) supersaturation max and (b) droplet number concentration predicted by chaos expansions and parameterizations vs detailed parcel model. The chaos expansions have been grouped by expansion order (x axis) and method for computing their coefficients (OLS, LARS, and LASSO; hue). "ARG" refers to the scheme of Abdul-Razzak and Ghan (2000); "MBN" refers to the scheme of Morales Betancourt and Nenes (2014).

$-2.4\% \pm 15.8\%$, respectively. This is larger than other studies have reported, but we explore a much larger parameter space in our sampling for the purposes of deriving the chaos expansion.

As a consequence of tending to slightly underpredict S_{max}, both the ARG and MBN schemes underpredict the number of activated droplets in the framework considered here. The mean relative error in droplet number predicted by the ARG and MBN schemes for the samples here are $-9.6\% \pm 23.4\%$ and $-4.9\% \pm 16.8\%$, respectively. All of the chaos expansions outperform the mean relative error of the ARG scheme, and those of order $p \geq 3$ do so with less variance.

In addition to producing low mean relative error in predicted S_{max} and droplet number activated, the chaos expansions also reproduce the dependence of activation dynamics on aerosol physical properties and updraft speed, as illustrated in Fig. 3. In response to increasing aerosol number concentration, S_{max} reached by an ascending parcel tends to decrease because there is a larger surface area available where condensation can occur, producing a larger source of latent heat release that limits the production of supersaturation. Overall, though, the droplet number concentration increases despite this effect as the aerosol activated fraction only decreases by a factor of 4 when the total number of initial aerosol increases by an order of magnitude (Figs. 3a,b). Shifting the aerosol population to larger sizes (Figs. 3c,d) produces a similar effect in inhibiting the increase in a parcel's supersaturation; however, Köhler theory predicts that these larger particles will more easily activate, which offsets the increase in S_{max} and yields larger droplet number concentrations. A similar effect occurs as aerosol hygroscopicity increases (Figs. 3e,f).

The chaos expansions, as well as both the ARG and MBN schemes, capture these subtleties of activation dynamics as well as the detailed parcel model. More importantly, the expansions reproduce the sensitivity of activation to updraft speed (Figs. 3g,h), which is an important factor controlling S_{max} and setting the droplet number. At the largest updraft speeds of a few meters per second—indicative of deep, vigorous convection—the MBN scheme outperforms both the chaos expansions and the ARG scheme. However, for the aerosol population considered in Fig. 3g,h (with $N = 1000\,\text{cm}^{-3}$, $\mu = 0.05\,\mu\text{m}$, and $\sigma = 2.0$), the relative error in predicted S_{max} by the chaos expansions at high updraft speeds does not substantially affect the diagnosed droplet number concentration, since in this case all but the smallest aerosol particles activate under equilibrium considerations. Note that this S_{max} overprediction coupled with an accurate assessment of activated fraction occurs for many different single-mode, lognormal aerosol populations.

Although all the chaos expansion results in Fig. 3 appear biased high compared to the parcel model, this bias does not hold true in general. Note that the ARG scheme is biased low in this particular analysis; this is generally just an artifact of fixing seven of the eight parameters and analyzing one-dimensional transects. Fixing the nonvarying parameters at different values tends to shift the bias positive or negative in a nonsystematic way. Critically, the choice of values for these parameters does not affect the chaos expansion's ability to reproduce the sensitivity to the varying parameter, which lends confidence that the expansions accurately reproduce the behavior of the parcel model.

Since Fig. 3 highlights the fact that different schemes potentially perform better in different parts of the parameter space governing droplet activation, we stratified the sampling results based on level of pollution and updraft-speed strength and computed activated fraction relative error statistics in each of these bins as shown in Fig. 4. All of the schemes are accurate in clean and lightly polluted conditions (with aerosol number concentration $N < 1000\,\text{cm}^{-3}$). However, there is a tendency for both the ARG and MBN schemes to underpredict droplet number in heavily polluted conditions ($N > 2500\,\text{cm}^{-3}$).

The fourth-order OLS-derived chaos expansion is plotted in Fig. 4a as a representative example of the chaos expansions, and it retains its accuracy across the pollution level–updraft strength spectrum. The combination of light updrafts and heavy pollution tends to produce the largest underprediction in activated droplet number, ranging from 10% to 30% for the ARG and MBN schemes. Unsurprisingly, relative error in activated fraction tends to be least sensitive to increasing aerosol number concentration in the strong updraft regime. In this case, the vigorous updraft produces strong adiabatic cooling that overwhelms latent heat release from condensation as the droplets in the parcel grow, contributing to a rapid and large S_{max} [Eq. (A1)] and thereby activating a significant fraction of the aerosol.

It should be noted that parts of the parameter space we considered in applying the PCM, evaluating its output, and comparing to existing parameterizations may not be typical of real atmospheric cases. We chose a large parameter space in order to derive the most general emulator possible for this particular single-mode aerosol case. In the real world, there should be some correlation between the ambient temperatures, pressures, and updraft speeds used when diagnosing aerosol activation, while we sample these factors as if they were independent from one another. The output from applying the parcel model

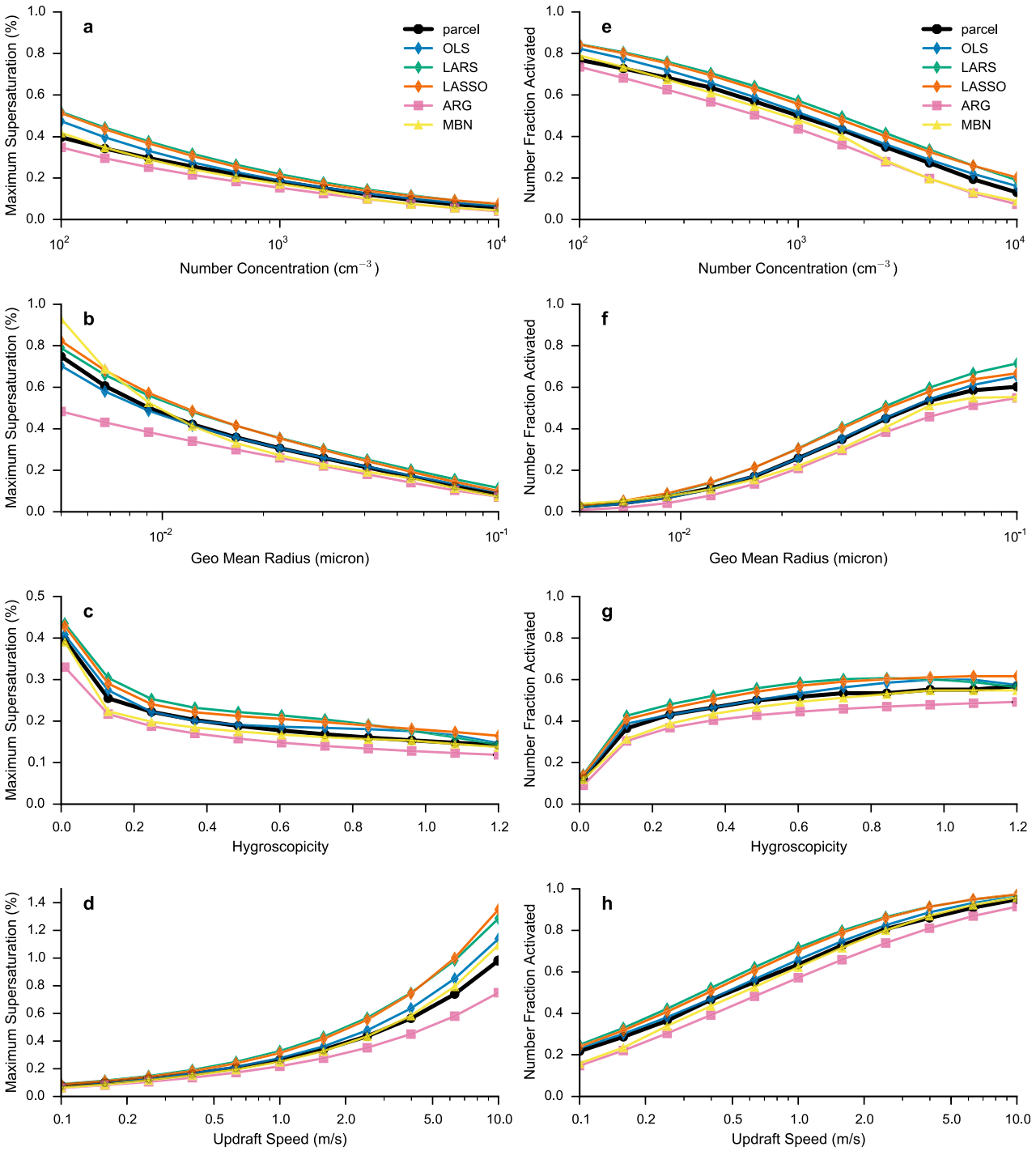

FIG. 3. Sensitivity of (a)–(d) parameterized and simulated maximum supersaturation and (e)–(h) activated number fraction to changes in mode number concentration, mode geometric mean radius, mode hygroscopicity, and updraft speed with all other parameters held fixed at the values $T = 283\,\mathrm{K}$, $P = 850\,\mathrm{hPa}$, $V = 0.5\,\mathrm{m\,s^{-1}}$, $a_c = 1.0$, $\mu = 0.05\,\mu\mathrm{m}$, $\kappa = 0.54$, $N = 1000\,\mathrm{cm^{-3}}$, and $\sigma = 2.0$. "MBN" and "ARG" correspond to the schemes of Morales Betancourt and Nenes (2014) and the update by Ghan et al. (2011) to Abdul-Razzak and Ghan (2000), respectively; the curves correspond to fourth-order chaos expansions with coefficients derived using the named method.

to nonrealistic activation scenarios could tend to inflate the computed relative errors and exacerbate differences between the parcel model and the existing, physically based parameterizations.

c. *Global sensitivity analysis*

Total Sobol' indices T_i for each of the input parameters summarized in Table 1 corresponding to the prediction of

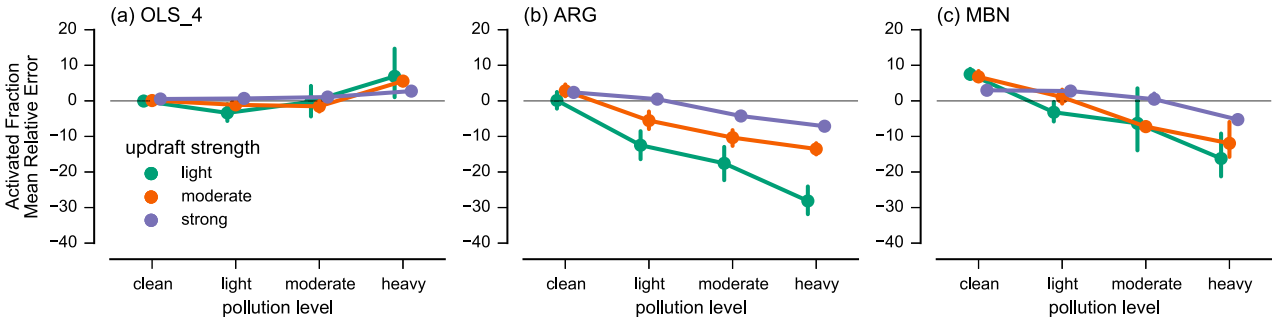

FIG. 4. Mean relative error in activated fraction for (a) fourth-order OLS-derived chaos expansion, (b) ARG, and (c) MBN schemes relative to detailed parcel model. Updraft speeds are light (10–50 cm s^{-1}), moderate (0.5–2.0 m s^{-1}), and strong (2.0–10.0 m s^{-1}); pollution levels are clean (10–250 cm^{-3}), light (250–1000 cm^{-3}), moderate (1000–2500 cm^{-3}), and heavy (2500–10 000 cm^{-3}). Error bars denote 95% confidence interval on mean relative error from in-bin samples; samples with $\mu < 10$ nm were omitted from these calculations.

S_{max} are plotted for each of the chaos expansions, parameterizations, and parcel model in Fig. 5. The order of the chaos expansion does not impact the relative importance of each term between different schemes or for higher-order expansions of the same scheme. Both updraft speed and the aerosol distribution size parameter most strongly contribute to the variance in S_{max}, followed by the total number of aerosols. This is in contrast with the chemical parameters in the model, a_c and κ, which provide the weakest constraints, suggesting the importance of the total aerosol surface area in dominating the potential for droplet activation by controlling S_{max}.

Indices derived using the chaos expansions very closely approximate those derived by sampling the full parcel model. This is expected since they simply provide an alternative framework for calculating the indices from the parcel model. However, the differences between the ARG and MBN schemes and the parcel model highlight the potential for biases in these schemes due to oversensitivity to particular model parameters. For instance, the ARG scheme is less sensitive to variations in σ_g than the full parcel model; however, the dependence of S_{max} in that scheme on σ_g is tuned to their own numerical calculations, which may differ from ours (Abdul-Razzak et al. 1998). Furthermore, the ARG scheme is more sensitive to variations in a_c than our parcel model and emulators, which parameterize the dependence of the condensational growth coefficient by a simple rescaling against a reference value computed for $a_c = 1$ rather than explicitly account for it.

Maximum supersaturations produced by the MBN scheme are more sensitive to the geometric mean size of the aerosol than those from the parcel model or ARG scheme and generally less sensitive to the strength of the updraft speed. The relative importance of each term for both the ARG and MBN scheme generally agrees with the estimates from the parcel model and chaos expansions, although the ARG scheme is most sensitive to

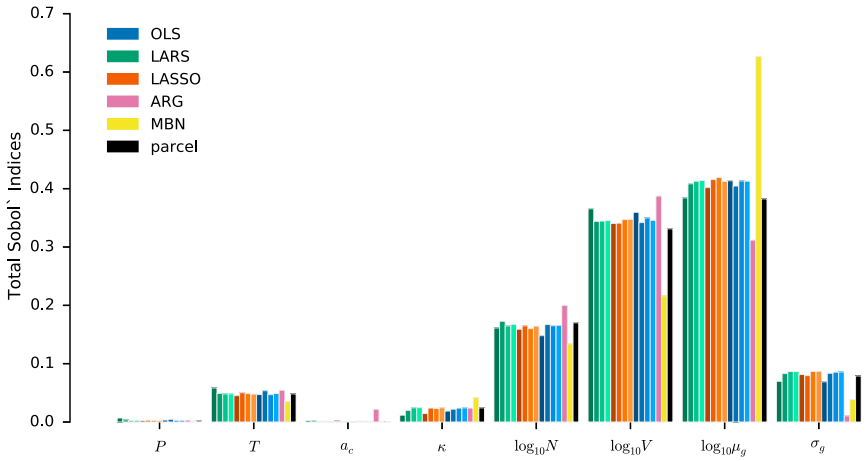

FIG. 5. Total Sobol' indices corresponding to the prediction of S_{max} for each parameter in Table 1 for each chaos expansion (OLS, LARS, and LASSO), parameterization (ARG and MBN), and parcel model. For the chaos expansions, the lighter colors indicate successively higher-order expansions.

TABLE 4. Input parameters and combinations ranked by main Sobol' index for the OLS-based chaos expansions.

	Order							
	2		3		4		5	
Rank	Term	Main S_i	Term	Main S_i	Term	Main S_i	Term	Main S_i
1	$\log_{10}\mu_g$	0.379	$\log_{10}\mu_g$	0.358	$\log_{10}\mu_g$	0.360	$\log_{10}\mu_g$	0.359
2	$\log_{10}V$	0.331	$\log_{10}V$	0.311	$\log_{10}V$	0.315	$\log_{10}V$	0.312
3	$\log_{10}N$	0.122	$\log_{10}N$	0.129	$\log_{10}N$	0.127	$\log_{10}N$	0.129
4	σ_g	0.049	σ_g	0.058	σ_g	0.055	σ_g	0.056
5	T	0.042	T	0.052	T	0.045	T	0.047
6	$\log_{10}V\log_{10}\mu_g$	0.016	$\log_{10}N\log_{10}\mu_g$	0.022	$\log_{10}N\log_{10}\mu_g$	0.021	$\log_{10}N\log_{10}\mu_g$	0.021
7	$\log_{10}N\log_{10}\mu_g$	0.016	κ	0.017	κ	0.019	κ	0.019
8	κ	0.015	$\log_{10}V\log_{10}\mu_g$	0.017	$\log_{10}V\log_{10}\mu_g$	0.018	$\log_{10}V\log_{10}\mu_g$	0.017

updraft speed rather than geometric mean aerosol size. By means of those parameters' higher T_i, the aerosol size distribution parameters exert far more influence over $S_{\max}$ robustly for both parameterizations and the parcel model.

The chaos expansions help rank the relative importance of each input parameter with the potential for important interactions between terms. The leading eight terms ranked by main Sobol' index computed using the OLS-derived chaos expansions are summarized in Table 4. For all four (second- through fifth-order) chaos expansions summarized, all the terms with rank greater than eight were an order of magnitude less important than those in the table. With the exception of the second-order scheme, the rankings of the top eight terms do not change relative order, and the main terms dominate the higher-order ones. The only higher-order terms that contribute grossly to the variance in $S_{\max}$ are combinations of the updraft speed and aerosol size distribution parameters, reiterating their importance compared to the chemistry terms (only κ appears in the top eight).

d. *Computational efficiency of chaos expansions*

As detailed in appendix B, evaluating the emulator produced by the chaos expansion requires two sets of straightforward floating-point operations. The first set requires the projection of the input parameters into the vector space spanned by the basis polynomials of the chaos expansion, which can then be used to evaluate the basis polynomials up to the required order. The remaining operations simply multiply these intermediate evaluations together and sum them to evaluate the full expansion. In general, this procedure should lie in between the ARG and MBN schemes in terms of computational complexity. The ARG scheme relies on straightforward floating-point operations to derive an estimate for $S_{\max}$, which at worst involves evaluating a logarithm. However, the MBN scheme requires sets of iterations, each of which necessitates a costly evaluation of the error function and the complementary error function.

On average, the second- and fifth-order OLS-derived chaos expansion was 10–17 times faster than the MBN scheme given the same single-mode aerosol population. The exact speedup depended on the background velocity; for weak updraft speeds, the performance of the MBN scheme fared better, although it became much worse for updrafts where $V < 2\,\mathrm{m\,s^{-1}}$. The ARG scheme was consistently 1–3 times faster than those same chaos expansions. Since the pathway for evaluating either the ARG or chaos expansion schemes do not change depending on the input parameters, their performance was the same regardless of what inputs were provided.

4. Summary and conclusions

An efficient parameterization of droplet activation for a single aerosol model under a wide variety of different physico-chemical properties and thermodynamic conditions was developed via statistical emulation of a detailed parcel model using polynomial chaos expansion. The emulators predict the maximum supersaturation achieved by a parcel, which is then used to diagnose activated droplet number using Köhler theory in a similar framework to existing activation parameterizations. The fourth- and fifth-order chaos expansions derived from the detailed parcel model are more accurate on average than two commonly used, physically based parameterizations from the literature (Abdul-Razzak and Ghan 2000; Morales Betancourt and Nenes 2014). Additionally, the chaos expansions are all at least 10 times faster to evaluate than the MBN scheme and only about twice as expensive as the ARG scheme. A simple algorithm was suggested for evaluating a chaos expansion that requires a minimal amount of data about the expansion (such as the basis polynomials and the coefficients of the expansion terms) to be saved; in this way, the chaos expansions offer a method for extending lookup tables to very high dimensionalities without suffering from exponentially rising storage costs.

Based on the large set of aerosol properties and thermodynamic conditions we sampled in order to derive and evaluate the chaos expansions, we observed that our emulators particularly outperform the existing schemes in conditions where a light updraft and heavy aerosol pollution (with respect to number concentration) are present. Because the ultimate goal of an activation parameterization is to couple the aerosol physics and chemistry to the cloud microphysics of a global-scale model, this deficiency in the existing parameterizations could be particularly important. Few global models have aerosol–cloud microphysics connections in their deep convection parameterizations, but many source potential cloud droplet formation based on a detailed aerosol activation calculation for their shallow convection and stratiform cloud microphysics schemes. These schemes sometimes artificially restrict the lowest possible updraft speed available for estimating droplet activation, but as a consequence they ensure that weak updrafts make up a large portion of the activation conditions considered during a model run. In regions of the world with heavy anthropogenic aerosol pollution—such as southern and eastern Asia—this provides a recipe for systematically underpredicting droplet number and potentially impacting either a global model's simulated aerosol indirect effect on climate or the modeled aerosol–cloud interaction's sensitivity to changes in anthropogenic aerosol emissions.

The global sensitivity analysis framed on the input parameter set used to derive the new chaos expansion emulators provides an additional, new check on the performance of existing activation schemes compared to the detailed parcel model. The breakdown of main Sobol' indices calculated using the chaos expansions provides insight into the importance of interactions between the dominant first-order terms (updraft speed, number concentration, and geometric mean size of the aerosol distribution), which are further summarized by the total Sobol' indices derived for the parcel model and both ARG and MBN schemes. The oversensitivity of the MBN scheme to the geometric mean radius and the undersensitivity of the ARG scheme to the updraft speed contribute to their disagreement with the parcel model across the range of pollution levels and updraft speeds studied here. Further, such sensitivity analyses could shed additional light on the potential biases of activation schemes and could provide useful metrics for evaluating the improvement of parameterizations more generally than simple ones based on relative or absolute error alone.

Critically, the framework from which the chaos expansions reported here are derived is extensible to the case where a complex, multispecies/multimodal aerosol population is tracked by a global model; in that case, the number of parameters describing the aerosol size distribution and chemical composition simply increases. Future work will derive chaos expansions emulating activation for a multimodal aerosol distribution specific to a particular global aerosol–climate model. Additionally, physical processes not considered here can also be introduced into the chaos expansion framework. For instance, entrainment can be incorporated into the parcel model following Seinfeld and Pandis (2006) and Barahona and Nenes (2007). Subgrid-scale variability in updraft speeds due to the coarse resolution of global model grids and the distribution of these updrafts can be represented either by a characteristic value (Morales and Nenes 2010) or by numerical integration over a distribution (Lohmann et al. 1999; Golaz et al. 2011). In the latter case, many activation calculations must be performed, incurring a large computational cost. However, the entire integration over a spectrum of droplet speeds could be parameterized in the chaos expansion framework, greatly reducing the cost of this calculation and potentially improving the accuracy of diagnosed cloud droplet number.

As the complexity of global aerosol–climate models increases with respect to the number of aerosol modes and species tracked by the model, there is a pressing need to understand how biases in activation calculations across the high-dimensional parameter spaces defining the aerosol–climate model affect cloud properties and ultimately impact modeled climate. This work highlights a novel way to build efficient, accurate activation schemes for this purpose akin to customized lookup tables, which cannot themselves extend to cover the necessary parameters. Employing such schemes should help improve simulated cloud microphysical properties and constrain modeled aerosol indirect effects on climate.

Acknowledgments. The work in this study was supported by the National Science Foundation Graduate Research Fellowship Program under NSF Grant 1122374, NSF Grant AGS-1339264, the National Research Foundation Singapore through the Singapore–MIT Alliance for Research and Technology and the interdisciplinary research group of the Center for Environmental Sensing and Modeling, and the U.S. Department of Energy, Office of Science (DE-FG02-94ER61937). We thank Steve Ghan (PNNL) and Athanasios Nenes (Georgia Tech) for reference implementations of their activation parameterizations.

APPENDIX A

Parcel Model Description

The adiabatic cloud parcel model implemented for this study follows the basic equations of Pruppacher and Klett (1997) and adopts the framework used by Nenes et al. (2001) to account for kinetic limitations on droplet growth. Fundamentally, the model integrates a system

of coupled ordinary differential equations that describe the thermodynamic evolution of an adiabatically lifted, nonentraining parcel. In all the simulations described here, we use the variable-coefficient ordinary differential equation solver (VODE; Brown et al. 1989) to integrate the system forward in time.

The model tracks the evolution of supersaturation S with respect to water as

$$\frac{dS}{dt} = \alpha(T,P)V - \gamma(T,P)\frac{dw_c}{dt}, \tag{A1}$$

where $\alpha(T, P) = (gM_wL/c_pRT^2) - (gM_a/RT)$ and $\gamma(T, P) = (PM_a/e_sM_w) + (M_wL^2/c_pRT^2)$ are functions that are weakly dependent on temperature and pressure (Leaitch et al. 1986), M_w and M_a are the molecular weights of water and air, L is the latent heat of evaporation of water, c_p is the specific heat of dry air at constant pressure, R is the universal gas constant, g is the acceleration due to gravity, e_s is the saturation vapor pressure, and w_c is the liquid cloud water mass mixing ratio. Equation (A1) expresses the supersaturation as a balance between production due to adiabatic cooling and loss due to latent heat release. This same framework describes the parcel's change in temperature over time,

$$\frac{dT}{dt} = -\frac{gV}{c_p} - \frac{L}{c_p}\frac{dw_v}{dt}, \tag{A2}$$

where V is the updraft velocity and w_v is water vapor mass mixing ratio. Water mass is conserved as vapor condenses into cloud water,

$$\frac{dw_v}{dt} + \frac{dw_c}{dt} = 0. \tag{A3}$$

Equations (A1)–(A3) are linked through the growth of the cloud droplet population from the initial aerosol. Given n bin sizes, each associated with a number concentration N and a radius r, the change in cloud water can be written as

$$\frac{dw_c}{dt} = \frac{4\pi\rho_w}{\rho_a}\sum_{i=1}^{n} N_i r_i^2 \frac{dr_i}{dt}, \tag{A4}$$

where ρ_w and ρ_a denote the density of water and air, respectively.

The diffusional growth rate for droplets in the ith bin is calculated by

$$\frac{dr_i}{dt} = \frac{G}{r_i}(S - S_{eq}), \tag{A5}$$

where S is the environmental supersaturation, S_{eq} is the Köhler-predicted equilibrium supersaturation of the droplet, and G is a growth coefficient which is a function of both the physical and chemical properties of the particle receiving condensate,

$$G = \left\{\frac{\rho_w RT}{e_s D_v' M_w} + \frac{L\rho_w[(LM_w/RT) - 1]}{k_a' T}\right\}^{-1}. \tag{A6}$$

Noncontinuum effects on the diffusivity D_v' and thermal conductivity k_a' factors are accounted for with the corrections

$$D_v' = D_v \Big/ \left(1 + \frac{D_v}{a_c r}\sqrt{\frac{2\pi M_w}{RT}}\right) \tag{A7}$$

and

$$k_a' = k_a \Big/ \left(1 + \frac{k_a}{a_T r \rho_a c_p}\sqrt{\frac{2\pi M_a}{RT}}\right). \tag{A8}$$

In these correction terms, the thermal accommodation coefficient a_T is assumed to be 0.96; the condensation coefficient a_C is allowed to vary as observations suggest it could take values between 0.1 and 1.0 (Raatikainen et al. 2013). The instantaneous droplet growth rate is further modulated by the difference between the environmental supersaturation S and the saturation ratio over the surface of the aqueous droplet S_{eq}. We treat the droplet-dependent S_{eq} following Petters and Kreidenweis (2007), who employ a single-term κ to parameterize particle hygroscopicity; values of κ can be derived from laboratory experiments. Under the framework of κ–Köhler theory the curvature effect term remains the same, while the solute effect term is rewritten such that

$$S_{eq} = \frac{r^3 - r_d^3}{r^3 - r_d^3(1-\kappa)}\exp\left(\frac{2M_w\sigma_w}{RT\rho_w r}\right), \tag{A9}$$

where r and r_d are the droplet radius and the dry radius of its embedded aerosol particle (which is tracked for each initial aerosol size in the model) and σ_w is the droplet surface tension, which we take to be independent of the droplet solution composition and described following the recommendation of Pruppacher and Klett (1997), $\sigma_w = 0.0761 - 1.55\times10^{-4}(T - 273)$. A limitation of the this approach for computing S_{eq} is that it is not convenient to derive analytical expressions for the critical supersaturation and radius; they must be computed numerically by finding the value r_{crit} such that

$$\left.\frac{\partial S_{eq}}{\partial r}\right|_{r_{crit}} = 0 \tag{A10}$$

and then computing $S_{crit} = S_{eq}(r_{crit})$ for a given κ and r_d. This is accomplished using Brent's method (Brent 1973) and by bounding r_{crit} from below with the observation that $r_{crit} > r_d$.

Finally, the parcel thermodynamic description is closed by predicting the pressure change within the ascending parcel following the hydrostatic relationship, which can be written using the ideal gas law as

$$\frac{dP}{dt} = -\frac{gPV}{R_d T_v}, \tag{A11}$$

where T_v is the virtual temperature, which is employed to account for changes in air density due to loss of water vapor to condensate. Equations (A1)–(A4), (A11), and (A5) applied to each n droplet size bins form a closed system which conserves total water mass.

APPENDIX B

Chaos Expansion Emulator Evaluation

The PCM as applied here produces two outputs: a P-length vector of coefficients α comprising real values and a $P \times M$ matrix of orthogonal polynomial orders $\mathbf{\Phi}$ comprising integers. Each term in matrix $\mathbf{\Phi}$ contains a multi-index component for each term in the chaos expansion and indicates the order of the orthogonal polynomial corresponding to term $1 \le j \le M$ for expansion term $0 \le i \le P$. For any expansion, $\max(\mathbf{\Phi}) = p$, the desired order of the chaos expansion. Algorithm 1 describes the evaluation of a chaos expansion.

Algorithm 1: Psuedocode for evaluating a polynomial chaos expansion of the form given in Eq. (2), applied to the computation of $S_{\max}$

```
for all X_j do
    Z_j ← project X_j
end for
Ŝ ← 0
for row i = 0; i ≤ P do
    Ŝ_i ← 1
    for column j = 0; j ≤ M do
        k ← Φ(i, j)
        P_{j,k} ← φ_j^k[Z(j)]
        Ŝ_i = Ŝ_i × P_{j,k}
    end for
    Ŝ = Ŝ + α(i) × Ŝ_i
end for
S_max ← 10^Ŝ
```

Evaluating the chaos expansion involves two parts. First, the input parameters must be projected to conform to the space supported by the PDFs associated with each basis polynomial type, producing a set of parameters Z_j. In general, a set of mixed orthogonal polynomials could be used to derive a chaos expansion, but here only Legendre polynomials were used, and each parameter can be projected using Eq. (5). Second, the polynomial can be evaluated by treating $\mathbf{\Phi}$ as a lookup table for the orders of each basis orthogonal polynomial. In practice, the evaluation of these orthogonal polynomials at Z_j for orders up to k can be efficiently precomputed (before the polynomial evaluation loop) by existing orthogonal polynomial libraries (Gautschi 1994).

In general, the only computationally complex part of the chaos expansion evaluation algorithm is the projection from X_j to Z_j; given certain basis orthogonal polynomials and their associated PDFs, this procedure could involve numerical integration or otherwise complicated function evaluations. In the case of simple uniform PDFs, though, the process is achieved entirely by rescaling the parameters, with little computational overhead. Furthermore, although evaluating a chaos expansion requires looping over each of its terms, each term can be computed independently from one another and efficiently optimized. This is in contrast with an iterative scheme that could involve numerical integration or other costly operations and must be performed in sequence.

REFERENCES

Abdul-Razzak, H., and S. J. Ghan, 2000: A parameterization of aerosol activation: 2. Multiple aerosol types. *J. Geophys. Res.*, **105**, 6837–6844, doi:10.1029/1999JD901161.

——, ——, C. Rivera-Carpio, H. A. Razzak, and C. R. Carpio, 1998: A parameterization of aerosol activation: 1. Single aerosol type. *J. Geophys. Res.*, **103**, 6123–6131, doi:10.1029/97JD03735.

Adams, B. M., and Coauthors, 2014: Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.0 user's manual. Sandia National Laboratories Tech. Rep., 333 pp. [Available online at https://dakota.sandia.gov/sites/default/files/docs/6.0/Users-6.0.0.pdf.]

Albrecht, B. A., 1989: Aerosols, cloud microphysics, and fractional cloudiness. *Science*, **245**, 1227–1230, doi:10.1126/science.245.4923.1227.

Altaratz, O., I. Koren, L. Remer, and E. Hirsch, 2014: Review: Cloud invigoration by aerosols—Coupling between microphysics and dynamics. *Atmos. Res.*, **140–141**, 38–60, doi:10.1016/j.atmosres.2014.01.009.

Anttila, T., and V.-M. Kerminen, 2007: On the contribution of Aitken mode particles to cloud droplet populations at continental background areas—A parametric sensitivity study. *Atmos. Chem. Phys.*, **7**, 4625–4637, doi:10.5194/acp-7-4625-2007.

Barahona, D., and A. Nenes, 2007: Parameterization of cloud droplet formation in large-scale models: Including effects of

entrainment. *J. Geophys. Res.*, **112**, D16206, doi:10.1029/2007JD008473.

——, R. E. L. West, P. Stier, S. Romakkaniemi, H. Kokkola, and A. Nenes, 2010: Comprehensively accounting for the effect of giant CCN in cloud activation parameterizations. *Atmos. Chem. Phys.*, **10**, 2467–2473, doi:10.5194/acp-10-2467-2010.

Blatman, G., and B. Sudret, 2011: Adaptive sparse polynomial chaos expansion based on *least angle regression*. *J. Comput. Phys.*, **230**, 2345–2367, doi:10.1016/j.jcp.2010.12.021.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657.

Brent, R. P., 1973: An algorithm with guaranteed convergence for finding a zero of a function. *Algorithms for Minimization without Derivatives*, Prentice-Hall, 19–57.

Brown, P. N., G. D. Byrne, and A. C. Hindmarsh, 1989: VODE: A variable-coefficient ODE solver. *SIAM J. Sci. Stat. Comput.*, **10**, 1038–1051, doi:10.1137/0910062.

Calbó, J., W. Pan, M. Webster, R. G. Prinn, and G. J. McRae, 1998: Parameterization of urban subgrid scale processes in global atmospheric chemistry models. *J. Geophys. Res.*, **103**, 3437–3451, doi:10.1029/97JD02654.

Cohard, J.-M., J.-P. Pinty, and C. Bedos, 1998: Extending Twomey's analytical estimate of nucleated cloud droplet concentrations from CCN spectra. *J. Atmos. Sci.*, **55**, 3348–3357, doi:10.1175/1520-0469(1998)055<3348:ETSAEO>2.0.CO;2.

Cohen, J. B., and R. G. Prinn, 2011: Development of a fast, urban chemistry metamodel for inclusion in global models. *Atmos. Chem. Phys.*, **11**, 7629–7656, doi:10.5194/acp-11-7629-2011.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004: Least angle regression. *Ann. Stat.*, **32**, 407–499, doi:10.1214/009053604000000067.

Ekman, A. M. L., A. Engström, and A. Söderberg, 2011: Impact of two-way aerosol–cloud interaction and changes in aerosol size distribution on simulated aerosol-induced deep convective cloud sensitivity. *J. Atmos. Sci.*, **68**, 685–698, doi:10.1175/2010JAS3651.1.

Fan, J., L. R. Leung, Z. Li, H. Morrison, H. Chen, Y. Zhou, Y. Qian, and Y. Wang, 2012: Aerosol impacts on clouds and precipitation in eastern China: Results from bin and bulk microphysics. *J. Geophys. Res.*, **117**, D00K36, doi:10.1029/2011JD016537.

Fountoukis, C., and Coauthors, 2007: Aerosol–cloud drop concentration closure for clouds sampled during the International Consortium for Atmospheric Research on Transport and Transformation 2004 campaign. *J. Geophys. Res.*, **112**, D10S30, doi:10.1029/2006JD007272.

Gantt, B., J. He, X. Zhang, Y. Zhang, and A. Nenes, 2014: Incorporation of advanced aerosol activation treatments into CESM/CAM5: Model evaluation and impacts on aerosol indirect effects. *Atmos. Chem. Phys.*, **14**, 7485–7497, doi:10.5194/acp-14-7485-2014.

Gautschi, W., 1994: Remark on algorithm 726: ORTHPOL—A package of routines for generating orthogonal polynomials and Gauss-type quadrature rules. *ACM Trans. Math. Software*, **24**, 355–356, doi:10.1145/292395.292467.

Ghan, S. J., and Coauthors, 2011: Droplet nucleation: Physically-based parameterizations and comparative evaluation. *J. Adv. Model. Earth Syst.*, **3**, M10001, doi:10.1029/2011MS000074.

Golaz, J.-C., M. Salzmann, L. J. Donner, L. W. Horowitz, Y. Ming, and M. Zhao, 2011: Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3. *J. Climate*, **24**, 3145–3160, doi:10.1175/2010JCLI3945.1.

Haywood, J., and O. Boucher, 2000: Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. *Rev. Geophys.*, **38**, 513–543, doi:10.1029/1999RG000078.

Karydis, V. A., S. L. Capps, A. G. Russell, and A. Nenes, 2012: Adjoint sensitivity of global cloud droplet number to aerosol and dynamical parameters. *Atmos. Chem. Phys.*, **12**, 9041–9055, doi:10.5194/acp-12-9041-2012.

Khvorostyanov, V. I., and J. A. Curry, 2006: Aerosol size spectra and CCN activity spectra: Reconciling the lognormal, algebraic, and power laws. *J. Geophys. Res.*, **111**, D12202, doi:10.1029/2005JD006532.

——, and ——, 2008: Kinetics of cloud drop formation and its parameterization for cloud and climate models. *J. Atmos. Sci.*, **65**, 2784–2802, doi:10.1175/2008JAS2606.1.

Kim, D., C. Wang, A. M. L. Ekman, M. C. Barth, and P. J. Rasch, 2008: Distribution and direct radiative forcing of carbonaceous and sulfate aerosols in an interactive size-resolving aerosol–climate model. *J. Geophys. Res.*, **113**, D16309, doi:10.1029/2007JD009756.

Kumar, P., I. N. Sokolik, and A. Nenes, 2009: Parameterization of cloud droplet formation for global and regional models: Including adsorption activation from insoluble CCN. *Atmos. Chem. Phys.*, **9**, 2517–2532, doi:10.5194/acp-9-2517-2009.

Lance, S., A. Nenes, and T. A. Rissman, 2004: Chemical and dynamical effects on cloud droplet number: Implications for estimates of the aerosol indirect effect. *J. Geophys. Res.*, **109**, D22208, doi:10.1029/2004JD004596.

Leaitch, W. R., J. W. Strapp, G. A. Isaac, and J. G. Hudson, 1986: Cloud droplet nucleation and cloud scavenging of aerosol sulphate in polluted atmospheres. *Tellus*, **38B**, 328–344, doi:10.1111/j.1600-0889.1986.tb00258.x.

Liu, X., and Coauthors, 2012: Toward a minimal representation of aerosols in climate models: Description and evaluation in the Community Atmosphere Model CAM5. *Geosci. Model Dev.*, **5**, 709–739, doi:10.5194/gmd-5-709-2012.

Lohmann, U., and J. Feichter, 2005: Global indirect aerosol effects: A review. *Atmos. Chem. Phys.*, **5**, 715–737, doi:10.5194/acp-5-715-2005.

——, ——, C. C. Chuang, and J. E. Penner, 1999: Prediction of the number of cloud droplets in the ECHAM GCM. *J. Geophys. Res.*, **104**, 9169–9198, doi:10.1029/1999JD900046.

Lucas, D. D., and R. G. Prinn, 2005: Parametric sensitivity and uncertainty analysis of dimethylsulfide oxidation in the clear-sky remote marine boundary layer. *Atmos. Chem. Phys.*, **5**, 1505–1525, doi:10.5194/acp-5-1505-2005.

Mayer, M., C. Wang, M. Webster, and R. G. R. Prinn, 2000: Linking local air pollution to global chemistry and climate. *J. Geophys. Res.*, **105**, 22 869–22 896, doi:10.1029/2000JD900307.

McFiggans, G., and Coauthors, 2006: The effect of physical and chemical aerosol properties on warm cloud droplet activation. *Atmos. Chem. Phys.*, **6**, 2593–2649, doi:10.5194/acp-6-2593-2006.

Meskhidze, N., 2005: Evaluation of a new cloud droplet activation parameterization with in situ data from CRYSTAL-FACE and CSTRIPE. *J. Geophys. Res.*, **110**, D16202, doi:10.1029/2004JD005703.

Moore, R. H., V. A. Karydis, S. L. Capps, T. L. Lathem, and A. Nenes, 2013: Droplet number uncertainties associated with CCN: An assessment using observations and a global model adjoint. *Atmos. Chem. Phys.*, **13**, 4235–4251, doi:10.5194/acp-13-4235-2013.

Morales, R., and A. Nenes, 2010: Characteristic updrafts for computing distribution-averaged cloud droplet number and stratocumulus cloud properties. *J. Geophys. Res.*, **115**, D18220, doi:10.1029/2009JD013233.

Morales Betancourt, R., and A. Nenes, 2014: Droplet activation parameterization: The population-splitting concept revisited. *Geosci. Model Dev.*, **7**, 2345–2357, doi:10.5194/gmd-7-2345-2014.

Morrison, H., G. de Boer, G. Feingold, J. Harrington, M. D. Shupe, and K. Sulia, 2011: Resilience of persistent Arctic mixed-phase clouds. *Nat. Geosci.*, **5**, 11–17, doi:10.1038/ngeo1332.

Nenes, A., and J. H. Seinfeld, 2003: Parameterization of cloud droplet formation in global climate models. *J. Geophys. Res.*, **108**, 4415, doi:10.1029/2002JD002911.

——, S. Ghan, H. Abdul-Razzak, P. Y. Chuang, and J. H. Seinfeld, 2001: Kinetic limitations on cloud droplet formation and impact on cloud albedo. *Tellus*, **53B**, 133–149, doi:10.1034/j.1600-0889.2001.d01-12.x.

——, R. J. Charlson, M. C. Facchini, M. Kulmala, A. Laaksonen, and J. H. Seinfeld, 2002: Can chemical effects on cloud droplet number rival the first indirect effect? *Geophys. Res. Lett.*, **29**, 1848, doi:10.1029/2002GL015295.

Pan, W., M. A. Tatang, G. J. McRae, and R. G. Prinn, 1997: Uncertainty analysis of direct radiative forcing by anthropogenic sulfate aerosols. *J. Geophys. Res.*, **102**, 21 915–21 924, doi:10.1029/97JD01653.

Petters, M. D., and S. M. Kreidenweis, 2007: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmos. Chem. Phys.*, **7**, 1961–1971, doi:10.5194/acp-7-1961-2007.

Pruppacher, H. R., and J. D. Klett, 1997: *Microphysics of Clouds and Precipitation*. 2nd ed. Kluwer Academic Publishers, 954 pp.

Raatikainen, T., and Coauthors, 2013: Worldwide data sets constrain the water vapor uptake coefficient in cloud formation. *Proc. Natl. Acad. Sci. USA*, **110**, 3760–3764, doi:10.1073/pnas.1219591110.

Saleeby, S. M., and W. R. Cotton, 2004: A large-droplet mode and prognostic number concentration of cloud droplets in the Colorado State University Regional Atmospheric Modeling System (RAMS). Part I: Module descriptions and supercell test simulations. *J. Appl. Meteor.*, **43**, 182–195, doi:10.1175/1520-0450(2004)043<0182:ALMAPN>2.0.CO;2.

——, and S. C. van den Heever, 2013: Developments in the CSU-RAMS aerosol model: Emissions, nucleation, regeneration, deposition, and radiation. *J. Appl. Meteor. Climatol.*, **52**, 2601–2622, doi:10.1175/JAMC-D-12-0312.1.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, 2010: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.*, **181**, 259–270, doi:10.1016/j.cpc.2009.09.018.

Seinfeld, J. H., and S. N. Pandis, 2006: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. 2nd ed. Wiley, 1232 pp.

Shipway, B. J., 2015: Revisiting Twomey's approximation for peak supersaturation. *Atmos. Chem. Phys.*, **15**, 3803–3814, doi:10.5194/acp-15-3803-2015.

——, and S. J. Abel, 2010: Analytical estimation of cloud droplet nucleation based on an underlying aerosol population. *Atmos. Res.*, **96**, 344–355, doi:10.1016/j.atmosres.2009.10.005.

Simpson, E., P. Connolly, and G. McFiggans, 2014: An investigation into the performance of four cloud droplet activation parameterisations. *Geosci. Model Dev.*, **7**, 1535–1542, doi:10.5194/gmd-7-1535-2014.

Sobol', I. M., 2001: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.*, **55**, 271–280, doi:10.1016/S0378-4754(00)00270-6.

Squires, P., and S. Twomey, 1961: The relation between cloud drop numbers and the spectrum of cloud nuclei. *Physics of Precipitation, Geophys. Monogr.*, Vol. 5, Amer. Geophys. Union, 211–219.

Sudret, B., 2008: Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.*, **93**, 964–979, doi:10.1016/j.ress.2007.04.002.

Tao, W.-K., J.-P. Chen, Z. Li, C. Wang, and C. Zhang, 2011: Impact of aerosols on convective clouds and precipitation. *Rev. Geophys.*, **50**, RG2001, doi:10.1029/2011RG000369.

Tatang, M. A., W. Pan, R. G. Prinn, and G. J. McRae, 1997: An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophys. Res.*, **102**, 21 925–21 932, doi:10.1029/97JD01654.

Tibshirani, R., 2011: Regression shrinkage and selection via the lasso: A retrospective. *J. Roy. Stat. Soc.*, **73B**, 273–282, doi:10.1111/j.1467-9868.2011.00771.x.

Twomey, S., 1959: The nuclei of natural cloud formation. Part II: The supersaturation in natural clouds and the variation of droplet concentration. *Pure Appl. Geophys.*, **43**, 243–249, doi:10.1007/BF01993560.

——, 1974: Pollution and the planetary albedo. *Atmos. Environ.*, **8**, 1251–1256, doi:10.1016/0004-6981(74)90004-3.

Wang, C., 2005: A modeling study of the response of tropical deep convection to the increase of cloud condensation nuclei concentration: 1. Dynamics and microphysics. *J. Geophys. Res.*, **110**, D21211, doi:10.1029/2004JD005720.

Ward, D. S., T. Eidhammer, W. R. Cotton, and S. M. Kreidenweis, 2010: The role of the particle size distribution in assessing aerosol composition effects on simulated droplet activation. *Atmos. Chem. Phys.*, **10**, 5435–5447, doi:10.5194/acp-10-5435-2010.

Weirs, V. G., J. R. Kamm, L. P. Swiler, S. Tarantola, M. Ratto, B. M. Adams, W. J. Rider, and M. S. Eldred, 2012: Sensitivity analysis techniques applied to a system of hyperbolic conservation laws. *Reliab. Eng. Syst. Saf.*, **107**, 157–170, doi:10.1016/j.ress.2011.12.008.

Joint Program Reprint Series - Recent Articles

For limited quantities, Joint Program publications are available free of charge. Contact the Joint Program office to order.
Complete list: http://globalchange.mit.edu/publications

2016-4 Metamodeling of Droplet Activation for Global Climate Models. Rothenberg, D. and C. Wang, *Journal of the Atmospheric Sciences*, 73, 1255–1272 (2016)

2016-3 Climate extremes and ozone pollution: a growing threat to China's food security. Tian, H., W. Ren, B. Tao, G. Sun, A. Chappelka, X. Wang, S. Pan, J. Yang, J. Liu, B.S. Felzer, J.M. Melillo and J. Reilly, *Ecosystem Health and Sustainability*, 2(1): 2332–8878 (2016)

2016-2 Carbon emissions in China: How far can new efforts bend the curve? Zhang, X., V.J. Karplus, T. Qi, D. Zhang and J. He, *Energy Economics*, 54 (Feb 2016): 388–395 (2016)

2016-1 Long-term economic modeling for climate change assessment. Y.-H.H. Chen, S. Paltsev, J.M. Reilly, J.F. Morris and M.H. Babiker, *Economic Modelling*, 52(Part B): 867–883 (2016)

2015-33 The Response of the South Asian Summer Monsoon to Temporal and Spatial Variations in Absorbing Aerosol Radiative Forcing. Lee, S.-Y. and C. Wang, *Journal of Climate*, 28, 6626–6646 (2016)

2015-32 The Observed State of the Water Cycle in the Early Twenty-First Century. Rodell, M., H.K. Beaudoing, T.S. L'Ecuyer, W.S. Olson, J.S. Famiglietti, P.R. Houser, R. Adler, M.G. Bosilovich, C.A. Clayson, D. Chambers, E. Clark, E.J. Fetzer, X. Gao, G. Gu, K. Hilburn, G.J. Huffman, D.P. Lettenmaier, W.T. Liu, F.R. Robertson, C.A. Schlosser, J. Sheffield, and E.F. Wood, *Journal of Climate*, 28, 8289–8318 (2016)

2015-31 Anthropogenic aerosols and the distribution of past large-scale precipitation change. Wang, C., *Geophysical Research Letters*, 42, 10,876–10,884 (2015)

2015-30 Consumption-Based Adjustment of Emissions-Intensity Targets: An Economic Analysis for China's Provinces. Springmann, M., D. Zhang and V.J. Karplus, *Environmental and Resource Economics*, 61(4): 615–640 (2014)

2015-29 Emulating maize yields from global gridded crop models using statistical estimates. Blanc, É and B. Sultan, *Agricultural and Forest Meteorology*, 214-215: 134–147 (2015)

2015-28 Reconciling reported and unreported HFC emissions with atmospheric observations. Lunt, M.F., M. Rigby, A.L. Ganesan, A.J. Manning, R.G. Prinn, S. O'Doherty, J. Mühle, C.M. Harth, P.K. Salameh, T. Arnold, R.F. Weiss, T. Saito, Y. Yokouchi, P.B. Krummel, L.P. Steele, P.J. Fraser, S. Li, S. Park, S. Reimann, M.K. Vollmer, C. Lunder, O. Hermansen, N. Schmidbauer, M. Maione, J. Arduini, D. Young and P.G. Simmonds, *PNAS*, 112(19): 5927–5931 (2015)

2015-27 The Observed State of the Energy Budget in the Early 21st Century. L'Ecuyer, T.S., H.K. Beaudoing, M. Rodell, W. Olson, B. Lin, S. Kato, C.A. Clayson, E. Wood, J. Sheffield, R. Adler, G. Huffman, M. Bosilovich, G. Gu, F. Robertson, P.R. Houser, D. Chambers, J.S. Famiglietti, E. Fetzer, W.T. Liu, X. Gao, C.A. Schlosser, E. Clark, D.P. Lettenmaier and K. Hilburnn, *J. Climate,* 28, 8319–8346 (2015)

2015-26 US major crops' uncertain climate change risks and greenhouse gas mitigation benefits. Sue Wing, I., E. Monier, A. Stern and A. Mundra, *Environmental Research Letters*, 10(2015): 115002 (2015)

2015-25 Benefits of mercury controls for the United States. Giang, A. and N.E. Selin, *PNAS*, online first (doi: 10.1073/pnas.1514395113) (2015)

2015-24 Electricity generation costs of concentrated solar power technologies in China based on operational plants. Zhu, Z., D. Zhang, P. Mischke and X. Zhang, *Energy*, 89(September): 65–74 (2015)

2015-23 An analysis of China's climate policy using the China-in-Global Energy Model. Qi, T., N. Winchester, V.J. Karplus, D. Zhang and X. Zhang, *Economic Modelling*, 52(Part B): 650–660 (2015).

2015-22 Protected areas' role in climate-change mitigation. Melillo, J.M., X. Lu, D.W. Kicklighter, J.M. Reilly, Y. Cai and A.P. Sokolov, *Ambio*, online first (doi: 10.1007/s13280-015-0693-1)

2015-21 The impact of climate change on wind and solar resources in southern Africa. Fant, C., C.A. Schlosser and K. Strzepek, *Applied Energy* 161, online first (doi: 10.1016/j.apenergy.2015.03.042) (2015)

2015-20 Enhanced marine sulphur emissions offset global warming and impact rainfall. Grandey, B.S. and C. Wang, *Scientific Reports* 5 (Article: 13055) (2015)

2015-19 Climate change impacts and greenhouse gas mitigation effects on U.S. water quality. Boehlert, B., K.M. Strzepek, S.C. Chapra, C. Fant, Y. Gebretsadik, M. Lickley, R. Swanson, A. McCluskey, J.E. Neumann and J. Martinich, *Journal of Advances in Modeling Earth Systems* 7(3): 1326–1338 (2015)

2015-18 Quantitative Assessment of Parametric Uncertainty in Northern Hemisphere PAH Concentrations. Thackray, C.P., C.L. Friedman, Y. Zhang and N.E. Selin, *Environ. Sci. Technol.* 49, 9185−9193 (2015)

2015-17 The feasibility, costs, and environmental implications of large-scale biomass energy. Winchester, N. and J.M. Reilly, *Energy Economics* 51(September): 188–203 (2015)

MIT Joint Program on the Science and Policy of Global Change

Massachusetts Institute of Technology
77 Massachusetts Ave., E19-411
Cambridge MA 02139-4307 (USA)

T (617) 253-7492 F (617) 253-9845
globalchange@mit.edu
http://globalchange.mit.edu