Development towards a global operational aerosol consensus: Basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME)

Walter R. Sessions[1], Jeffrey S. Reid[2*], Angela Benedetti[3], Peter R. Colarco[4], Arlindo da Silva[4], Sarah Lu[5], Thomas Sekiyama[6], Taichu Y. Tanaka[6], J.M. Baldasano[7], Sara Basart[7], Malcome E. Brooks[8], Thomas F. Eck[9], Mark Iredell[5], James A. Hansen[2], Oriol C. Jorba[7], Hann-Ming Henry Juang[5], Peng Lynch[1], Jean-Jacques Morcrette[3], Shrinivas Moorthi[5], Jane Mulcahy[8], Yaswant Pradhan[8], Miha Razinger[3], Charles B. Sampson[2], Jun Wang[5,10], Douglas L. Westphal[2]

[1] {CSC Inc., Monterey CA}
[2] {Marine Meteorology Division, Naval Research Laboratory, Monterey CA}
[3] {European Centre for Medium-Range Weather Forecasts Reading, UK }
[4] {NASA Goddard SFC, Greenbelt, MD}
[5] {NOAA NCEP, College Park, MD}
[6] {Atmospheric Environment and Applied Meteorology Research Department, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba Japan}
[7] {Earth Sciences Department, Barcelona Supercomputing Center-Centro Nacional de Supercomputación, Barcelona, Spain}
[8] {Met Office, Exeter, UK}
[9] {USRA, NASA Goddard SFC, Greenbelt MD}
[10] I.M. Systems Group Inc., Rockville, MD

----
* Corresponding Author, Jeffrey S. Reid, Marine Meteorology Division, Naval Research Laboratory, 7 Grace Hopper Ave, Stop 2, Monterey, CA 93943-5502 Tel: 831-656-4725, Fax: 831 656-4769, Email: jeffrey.reid@nrlmry.navy.mil

Abstract: Here we present the first steps in developing a global multi-model aerosol forecasting ensemble intended for eventual operational and basic research use. Drawing from members of the International Cooperative for Aerosol Prediction (ICAP) latest generation of quasi-operational aerosol models, five day AOT forecasts are analyzed for December 2011 through November 2012 from four institutions: ECMWF, JMA, NASA GSFC, and NRL/FNMOC. For dust, we also include the NOAA NGAC product in our analysis. The Barcelona Supercomputing Centre and UK Met Office dust products have also recently become members of ICAP, but have insufficient data to be included in this analysis period. A simple consensus ensemble of member and mean AOT fields for modal species (e.g., fine & coarse mode, and a separate dust ensemble) is used to create the ICAP Multi-Model Ensemble (ICAP-MME). The ICAP-MME is run daily at 0Z for 6 hourly forecasts out to 120 hrs. Basing metrics on comparisons to 21 regionally representative Aerosol Robotic Network (AERONET) sites, all models generally captured the basic aerosol features of the globe. However, there is an overall AOT low bias among models, particularly for high AOT events. Biomass burning regions have the most diversity in seasonal average AOT. The southern oceans, though low in AOT, nevertheless also have high diversity. In regard to root mean square error, as expected the ICAP-MME placed first over all models worldwide, and was typically first or second in ranking against all models at individual sites. These results are encouraging; as more global operational aerosol models come on line, we expect their inclusion in a robust operational multi-model ensemble will provide valuable aerosol forecasting guidance.

1.0 Introduction

Aerosol modeling, once purely the domain of regional air quality and climate models, has seen recent rapid development at traditional Numerical Weather Prediction (NWP) centers (e.g., Tanaka et al., 2003; Morcrette et al., 2009; Westphal et al., 2009; Kukkonen et al., 2012). Applications are numerous, and include corrections for radiance assimilation systems for the NWP modeling systems themselves (Wang and Niu, 2013; Weaver et al., 2007). There is further mounting evidence that for heavily burdened atmospheres, inclusion of the radiative effects of aerosol particles improves overall NWP forecasts (e.g., Haywood et al., 2005; Perez et al., 2006; Wang et al., 2010; Mulcahy et al., 2014) and is even hypothesized to impact Tropical Cyclone (TC) development (e.g., from Karyampudi and Carlson, 1988; Karyampudi and Pierce, 2002; Dunion and Velden, 2004; to most recently Dunstone et al., 2013, Reale et al., 2011, 2014). Direct and indirect radiative effects have also been found to impact common NWP parameters such as temperature. For example, in response to large biomass burning events, surface temperatures clearly drop (Westphal and Toon, 1991). Smoke over the Indian Ocean in 1997 and 2006 may have resulted in a net cooling of sea surface temperatures (e.g. Thampi et al., 2009, Rajeev et al., 2008), with dust over the Atlantic Ocean similarly indicted both physically (Evan et al., 2008) and as an artifact (Merchant et al., 2006). Atmospheric transport and diffusion can expand aerosol impacts to continental and global scales thus posing further NWP impact questions (e.g., Colarco, 2004; Damoah et al., 2004 over North America and Koe et al., 2001 over Asia). For these reasons, most NWP centers with global modeling mandates have some form of aerosol prediction program. Indeed, increased accuracy in forecasting aerosol particles has benefits for mitigating human impacts: poor air quality negatively impacts biological processes including human cardiovascular and respiratory health (Seaton et al., 1995;

Poschl et al., 2005). Reduced visibility due to aerosols creates operational hazards on land, at sea and for aviation. Volcanoes represent a dramatic example, with SO_2 and ash reducing visibility, while silicate tephra induces aircraft engine stalls and flame outs (Miller and Casdevall, 2000; Carn, et al., 2008). Large volcanic eruptions that inject SO_2 in the stratosphere can also have a long-lasting cooling impact on surface temperature.

The path to the development of NWP aerosol capabilities has been quite different among centers. Certainly, the underlying meteorology driving aerosol models is from largely independent models. The aerosol source, microphysics and sink functions have also been developed or drawn from a variety of air quality and climate data sources. The differences in meteorology and assumed aerosol heritage when many aerosol parameterizations were developed lead to significant amounts of model tuning. Sometimes unphysical tuning parameters are required in order to get physical results against key metrics. Given the complexity of the aerosol and meteorological environment, this tuning can lead to high scoring in one metric (say Aerosol Optical Thickness-AOT) and poor scoring on another (say Particulate Matter d_p<2.5 μm, PM2.5). With the advent of AOT data assimilation, models are driving towards that metric (e.g., Reid et al., 2011) and AOT model analyses are dramatically improved. But even here, assimilation methods diverge significantly between centers (Reid et al., 2011; Benedetti 2014), and eventually this must be reconciled for multi-day forecasts.

Due to the stochastic nature of the atmosphere, for any NWP variable, aerosol species or dynamical, deterministic forecasts eventually reduce in quality with increasing forecast time no better than climatological values (or sometimes worse). There are many sources of forecast error, but there are two categories in particular that garner significant NWP attention; 1) Systemic errors from the imperfect nature of the model; 2) Sensitivity of models to initial conditions. Lorenz (1963, 1965, 1969a,b) showed in his classic papers that small errors in initial conditions produce large errors and divergence even within a perfect model. Errors ranging up to the synoptic scale have been found to not be the result of model deficiencies, but even small variation in initial states (Reynolds et al., 1994). To help control for these errors, ensemble-based prediction, single-model ensemble meteorological forecasts are used by nearly all the major operational weather centers (Buizza et al., 2005). However, while single-model probabilistic ensemble forecasting is clearly enhancing model solutions (particularly in data sparse regions), multi-model ensembles are an ever increasing tool for forecasters. Multi-model ensemble (or consensus) forecasting, using independent and skilled forecasts, has long proven valuable to atmospheric sciences. Ensemble techniques have been applied to the benefit of tropical cyclone track (Leslie and Fraedrich, 1990; Mundell and Rupp, 1995; Goerss, 2000; Sampson, 2010) and intensity forecasting (Kaplan and DeMaria 2001; DeMaria et al. 2006; Sampson et al., 2008). The consensus of cyclone track forecasts was found, on average, to be more accurate than the individual member deterministic models. Consensus style multi model ensembles and their interpretation are a mainstay of the climate change community (e.g., Meehl et al., 2007; Knutti et al., 2010). Fordham et al. (2012) used a multi-model ensemble of general circulation models to explore potential impacts of climate change following demonstration of GCM consensus values by Reichler and Kim (2008). Non-NWP methods also benefit from consensus techniques, as Sanders (1973) showed when the average forecast from a group of forecasters often proved better than any of the individual contributions given. Taken a step further, error weighting a multi-model ensemble leads to the development of the super-ensemble (e.g., Krishnamurti et al., 1999; Casanova and Ahrens, 2009). However, equal weighting in a consensus style appears to provide the most robust result overall for a host of forecasting applications (e.g., DelSole et al., 2013; Sansom et al., 2013), especially if model errors are not precisely known (Weigel et al., 2010). This final point is salient here as models develop rapidly, error estimates are often quickly out of date.

The rapid increase in the number of operational and quasi-operational global aerosol models coupled with the NWP community's wide experience of ensemble systems has resulted in an opportune moment to explore the development of a global operational multi model aerosol forecast consensus. The International Cooperative for Aerosol Prediction (ICAP), consisting of developers servicing aerosol programs at forecasting centers and remote sensing data providers began meeting in April 2010 to discuss issues germane to the operational aerosol forecasting community (Reid et al., 2011; Benedetti, 2012). As a relatively nascent community, ICAP has worked to build the standards for data protocols, validation, and verification between international centers. Data exchange for the purposes of consistent error analysis and consensus forecasting began in early 2011 and now includes four complete aerosol forecast models (ECMWF- Monitoring Atmospheric Composition and Climate Model, MACC; FNMOC/NRL Navy Aerosol Analysis and Prediction System; JMA- Model of Aerosol Species IN the Global AtmospheRe, MASINGAR; and NASA GMAO Goddard Earth Observing System Version 5, GEOS-5). Three dust-only models are also included (NMMB/BSC-CTM Non-hydrostatic Multi-scale Meteorological; NOAA NCEP NEMS GFS Aerosol Component, NGAC; UKMO Unified Model). In this paper we briefly describe the ICAP-MME framework and explore the first year of ensemble and ensemble member AOT products. For this study we only include the four complete aerosol models and NGAC (NMMB and the Unified Model will be in subsequent publications once sufficient data are incorporated for robust statistics). Forming an arithmetic mean of model parameters, the ICAP Multi-Model Ensemble (ICAP-MME) was generated. Climatological characteristics of the ensemble mean are presented. Verification statistics against Aerosol Robotic Network (AERONET) sun-sky radiometer data are presented including bias and root mean square error and we highlight areas of relative consensus and divergence. Finally, to set the stage for the next round of analyses, an example for Cape Verde dust is presented on ICAP-MME on issues to be addressed in predicting extreme aerosol events.

2.0 Methodology

For this introductory paper on the ICAP-MME, we briefly describe the included models and outline the fundamental metrics for model performance for AOT. The analysis period for this paper spans one year from December 2011 through November 2012. A further seasonal breakdown was also performed for boreal winter/spring (December-May) and summer/fall (June-November) periods. As per original ICAP agreements, we do not identify specific models to specific metrics other than the ensemble model itself. All such evaluations are to be performed and presented by the individual model's developers. Rather, we emphasize relative spread in skill for different sites. There are multiple reasons for this anonymous approach. These include the developmental nature of some of the input models and the very rapid updates the input models are receiving (e.g., by the time this paper is published, the model performance statistics will be certainly out of date). This paper is intended to demonstrate the usefulness of a multi-model ensemble in both forecasting applications, as well as a way to identify areas of common development needs in aerosol prediction.

2.1 Input Models

The ICAP-MME is currently based on four comprehensive global aerosol models (GEOS-5, NAAPS, MACC, MASINGAR), and three dust-only global models (NOAA NGAC, NMMB/BSC-CTM, UKMO Unified Model). Requirements for entry in the ICAP-MME are a global model with at least quasi-operational status and reliable data distribution from a large data center. During the development of this paper, there was insufficient data to fully evaluate two of the dust models (BSC and UKMO). Thus while we include these two models in the description, they are not used in this early evaluation.

We provide brief synopses of the input models in the current quasi-operational ICAP-MME consensus in Appendix A. As can be seen, these models tend to be quite independent in the parameterizations used for sources-particularly for dust and biomass burning. Although, biomass bringing emissions all have some lineage back to MODIS active fire hotspot counts. Sea salt is treated similarly in nature between models in terms of functional form, but tuning based on underlying meteorology and sink terms results in significant differences between the models. Perhaps the most similar aspect of the models is in emissions of anthropogenic emissions which are poorly constrained and thus similar inventories are developed. MACCity is used by both ECMWF and NRL. NASA GMAO GEOS-5 uses Edgar, which as similar components to MACCity. Finally, the NOAA NCEP NGAC dust model has the same GOCART foundation as GEOS-5, although with different driving meteorology and without data assimilation.

2.2 ICAP-MME

The International Cooperative for Aerosol Prediction Multi-model Ensemble (ICAP-MME) is a consensus style multi-model ensemble where all members are equally weighted. ICAP-MME was born out of a simple ICAP proposition that some uniform basis of AOT plotting be adopted across centers. This quickly led to data exchange and ultimately the formation of the AOT ensemble. Because of differences in member centers data policy, data availability of ICAP-MME member and consensus fields is limited to participating centers. However, consensus plots are available on the web (http://www.nrlmry.navy.mil/aerosol/_) with further expansion in the coming year.

The basic resolution of ICAP-MME is 1x1 degree, with member model data re-gridded through linear interpolation to 1x1 degree model grid. Three-dimensional aerosol and AOT fields are then generated in a member agreed NetCDF format. The ICAP consensus is the arithmetic mean of the interpolated fields. Because of latency constraints by some of the members, ICAP-MME is generated with a 24 hour lag. This will be reviewed as those constraints change. Forecasts are available 6 hourly out to 120 hours. At the moment the ensemble is limited to speciated AOT at a standard 550 nm wavelength. Data continuity for ICAP-MME for the current study period is presented in Figure 1(a). Because data are provided in an operational data stream, it was not always possible to back populate to make a completely contiguous data set. Outages could be due to a combination of network issues either at NRL, where the data are assembled, or at the production center. For this study, ICAP-MME is only generated when all 4 core models populate the ensemble, which holds data for 90% of forecasts.

ICAP-MME has 4 broad species, Dust, Sea Salt, Pollution/Sulfate and Biomass Burning/Smoke. The largest difficulty in combining model data is in the various member models' speciation. For example, NAAPS separates out species by source (as is done in the ICAP-MME). Other models, such as GEOS-5, carry species by chemical specie (e.g., sulfate, organic carbon, black carbon, etc.). Also, some models carry size information (MACC), and others ignore biogenic organic carbon emissions. In the case of coarse mode aerosol species such as sea salt or dust, the speciation versus source is easy to reconcile as the source and chemistry are one in the same, and size information can be integrated. The separation between anthropogenic pollution, biomass burning, and sometimes included biogenic emissions, is much more ambiguous. Therefore we developed the simple rubric that sulfate and biogenic are considered in the pollution/sulfate category, whereas organic carbon is listed with biomass burning-which if not physical, is in line with how the species are input and transformed into the models. Clearly, this is unsatisfying from multiple points of view. To clarify the situation, then, our analysis reduces the degrees of freedom further, and we largely analyze on a simple fine and coarse mode specie AOT (e.g., dust and sea salt is coarse, pollution, biomass burning, etc is fine). While there is some residual "coarse mode" material in sub micron size ranges, the SDA algorithm takes these tails into account.

There are a number of products that are then generated from ICAP-MME. Most commonly used is the consensus arithmetic mean coupled with the standard deviation for the so-called mean-spread plot. Similarly, the median is calculated and sometimes used, as it is robust in the face of a major outlier. For event based metrics, such as scores for dust storms, several cut points (e.g. thresholds) were used. The most notable and consistent is a 550 nm AOT of 0.8, high enough for the sky to have a complete haze color. Now that there are suitable data to develop a climatology, a dynamic event cut-point will be developed in the future based on multiples of regional standard deviations or geometric standard deviations (e.g., 1σ or 2σ event).

Figure 2 presents example data from the ICAP-MME for the 72 hour forecast for a particularly large dust event on June 29th, 2012 including contributions from all four core and the three dust members. Plots and data such as these are expected to be released to the public following the publication of this paper. Figure 2(a) presents the simple ICAP-MME AOT mean. In Figure 2(b), a 'mean/spread" plot is presented where the isopleths are AOT and the color is standard deviation. From these plots we can see that in many dust areas, the models are very consistent, whereas in the Sahel and Arabian Gulf there is more uncertainty. In figure 2(c), isopleths of AOT of 0.8 are presented, showing spatial differences in models, whereas in Figure 2(d), a simple warning area mask is plotted where at least half the models predict AOT>0.8. All of these products are designed for easy interpretation and verification.

2.3 Verification

For comparison with available observations, for this introductory paper we focus on the ICAP-MME 550 nm AOT apportioned into total, fine and coarse mode contributions, as well as some limited examination of the 5 member dust ensemble (ICAP-CORE+NGAC). Comparisons henceforth referencing ICAP values refer to that mean while individual member results remain anonymous. Core verification metrics here include mean bias, root mean square error, and fractional gross error.

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(c_i - o_i) \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)^2} \tag{2}$$

$$FGE = \frac{2}{n}\sum_{i=1}^{n}\left|\frac{c_i - o_i}{c_i + o_i}\right| \tag{3}$$

AOT data from the Aerosol Robotic Network (AERONET; Holben et al., 1998) are used to validate ICAP forecasts. AERONET level 2 data (cloud screened and quality assured with final calibrations; Smirnov et al., 2000) are used where available but can take upwards of twelve months to be processed. During periods where L2 data are not available, L1.5 data (cloud screened but without final calibration) are substituted after being hand filtered at NRL for clear outliers. Total, fine and coarse mode AOT at 550 nm were extracted using the O'Neill et al., (2003; 2008) spectral deconvolution method (SDA) from AERONET provided AOTs. Our extraction differs from the 500 nm extraction performed at AERONET. The AERONET Level 2 input spectral AOT to the SDA algorithm (380 nm to 870 nm) are accurate to ~0.01 to 0.02 (higher in the UV; Eck et al., 1999). These accuracies are for non-cloud contaminated data and comparison of AERONET field site AOT with independently calibrated sun photometers showed agreement to within ~0.015 (root mean square) or better (Schmid et al., 1999; Nyeki et al., 2012). Level 1.5 AOD may have typical accuracies of ~0.02-0.04 but is quite variable and uncertainty may be larger, depending primarily on the length of deployment since initial calibration and the amount of material deposited on the optics lenses (dust, sea salt, etc.). Instances of cirrus contamination (Chew et al., 2011) were evident in the level 1.5 and, to a lesser extent, level 2 products. Influence of these outliers was removed by hand for clear outliers, as well as trimming the top five percent of coarse observations in Northern Africa and East Asia, and the top fifteen percent elsewhere. The remaining observations are then binned by the median observation value within a six hour window centered on the model valid time. We focus on 21 sites chosen by the ensemble developers in consultation with AERONET before the analysis was conducted. Selection was based on regional representativeness (e.g., Shi et al., 2011) as well as a contiguous data record throughout the one year study period. These are listed in Table 1 and marked on Figure 1(b).

All quantitative comparisons to AERONET are pairwise, conducted only when AERONET and the ICAP-MME can be co-located. Because all four of the multi-species models invoke some form of data assimilation, and ECMWF does not generate an analysis field of AOT on the model grid, our primary model metric for global representation of aerosol loadings is the 6-24 hour forecast. Given AERONET only collects data on the sun side of the earth, this corresponds to 6-30 hours of forecast time for any "data day". For all calculations of forecasts out to 5 days, verification is performed +/- 3 hours of model valid time which is instantaneous for that time. For brevity, we group error statistics into data days to simplify the number of columns in data tables. With the once daily 0Z (GMT) production of ICAP some regions benefit from the availability of daytime only data for verification and assimilation. Thus over Asia verification and assimilation are at a shorter forecast time than say Europe and North America. This gives

Asia a beneficial regional verification bias. But, we do not believe this will impact any of our key results.

Rank histograms, also known as the "Talagrand diagram," were also generated for the models (Talagrand et al. 1997). These help determine if the ensemble members are drawn from the same distribution that produces the true state. While not a true verification tool, they are diagnostically useful to judge the ensemble reliability. Given an observation point, an n member ensemble is organized from highest to lowest and assigned a rank of 1 to n+1. If the ensemble is representative, the observation value is equally likely to be of any rank of the n+1 ranks, assuming a statistically significant number of independent observations, resulting in a flat histogram. Conversely, bias could be evaluated if the observation too often falls into the top or bottom bin. A U-shaped histogram potentially indicates insufficient ensemble spread, as all the forecasts consistently resolve too high or low. Care must be taken with interpretation, as uniform or U-shaped distributions can arise, such as when observational biases change sign by location. More detail on rank histograms can be found in Talagrand et al. (1997) and Hamill (2001).

For event forecasting, we use the Critical Success Index also known as the Threat Score (CSI or TS=hits/(hits+misses+false alarms) as a common and straightforward metric with scales that range from 0 (no skill) to 1 (perfect skill). For AOT, threat scores are somewhat subjective. If the bar for triggering a hit is too low, then the model forecast is without functional value. If it is set extremely high, then the TS gives a false optimism in system performance. To address this issue, we also use the Equitable Threat Score accounts for random change ETS = (hits – random chance hits) / (hits + misses + false alarms– random chance hits), where the random chance is (the total forecasts of the event* the total observation)/sample size. As discussed in Section 6, the use of threat scores are somewhat problematic, especially in regard to amplitude or displacement error (Baldwin and Kain, 2006).

3.0 Results: Climatological Characteristics of ICAP-MME

The mean and standard deviation of the ICAP-MME 6 hourly forecast mean is provided in Figure 3. Data are broken down into a seasonal and size mode degree of freedom. Seasonally, data are presented for the boreal winter/spring Dec 2011-May 2012 and boreal summer/fall June-November 2012 time periods. These bi-seasonal temporal stratifications account for the major monsoonal and climatic shifts in the atmosphere while preserving major aerosol events such as, for the boreal summer/fall, the August-October biomass burning seasons in Africa, South America, and Maritime Continent, the June-August African Dust Season, and the contiguous United States (CONUS) and European summer haze seasons. Similarly the boreal winter/spring period captures the March-May Asian dust season, and the Southeast Asia and Sahelian African biomass burning season. The next set of striation is by modal size, separating fine mode species (sulfate, organic carbon, black carbon etc.) from coarse (sea salt and dust). This stratification resolves speciation differences between models. Corresponding seasonal means of AERONET fine and coarse mode 550 nm AOT are also presented on the mean plots.

The ICAP-MME, as well as the entirety of the core model members, easily resolves the world’s largest aerosol features: Saharan dust, continental biomass burning, and the great Asian dust and pollution plume are well described. Associated standard deviations of the ICAP-MME 6 hourly means also highlight regions of more episodic aerosol events, with African and Asian dust being particularly noteworthy. The seasonal biomass burning features also stand out.

While in the next section we focus on member scores, from a climatological point of view it is worthwhile to examine the climatological variability between the models. In a manner similar to Figures 3, in Figure 4 we present bi-seasonal and size modal estimates of the point-wise maximum or minimum AOT of the ensemble members. That is, after generating seasonal and size modal mean AOT for each of the four core member models, for each 1x1 latitude and longitude point we select the highest and lowest AOT of the four. Such a minimum and maximum not only is indicative of differences in model amplitude, but also in plume location (if, for example, a zonal aerosol feature is shifted meridionally between models, then the minimum could be low across the region, missing the feature completely). Such a depiction can span the local seasonal range of coarse and fine mode AOT present in the models and identify which areas require attention. The largest area of difference between the models was clearly associated with biomass burning, with factors of three differences spanning springtime Sahelian and South American biomass burning and biomass burning over maritime continent. The coarse models generally tuned dust reasonably well, but sea salt maximum and minimum in the southern oceans spanned a factor of two.

We can become more quantitative through comparison to AERONET. Table 1 lists AERONET AOT and the model mean bias for the ICAP-MME and its core members for the two monsoonal periods. Table 2 presents a similar set of bias statistics for ICAP core models plus NGAC for those sites where the coarse mode is dominated by dust. In all of these tables, ICAP-MME ensemble mean is underlined. To improve visualization, in Figure 5 we present similar data in a scatter plot (similar in nature to a reliability diagram), where the ICAP-MME means are in bold.

Our interpretation of pairwise AERONET data are in agreement with our interpretation of plots in Figure 3 and 4. Overall the models have reasonable correlation and consistency across the AERONET sites. Cape Verde, perhaps the community's benchmark site for dust, was so well tuned in the models that it had virtually nonexistent dust biases for summer/fall and an insignificant 10% high bias for winter/spring. Most background sites performed equally well. The one exception was Crozet Island in the southern oceans for boreal summer/austral winter, where most models clearly overestimated sea salt production.

For higher AOT sites, all of the models have a clear and consistent low bias. A small part of this can be explained by the smoothing nature of a global model (models propose to represent the gridbox mean). High AOT plume or dust event amplitude simply is not captured either in the model physics or in data assimilation. Depending on how models screen their AOT data before data assimilation, bias could be a residual of the retrieval (e.g., see discussion in Zhang et al., 2008). However, some of the largest departures are clearly related to chemistry or sources. The highest single departure for the winter-spring period is Chiang Mai, Thailand, where all models seem to underestimate that season's biomass burning and pollution influence. Models also underestimate AOTs at Singapore. This is not surprising, as SE Asia has been identified as being perhaps the most challenging region in the world to observe and model (Reid et al., 2009; 2013) because, among other reasons, high cloud cover conspires to disrupt both fire detections and data assimilation. Most models rely in fact on retrieved products from the MODIS instruments, which have large biases in presence of clouds or are not available at all. The Sahel region in the winter/spring is another area of considerable difficulty for nearly all models. Ilorin in fact had the highest climatological AOT of any site (0.89) with consistent low biases in all models, on the order of 50%. This is likely due to underrepresentation of biomass burning in all models, although a correlated bias between models and smoke optical properties cannot be ruled out at this time. The Sahelian biomass burning system and its frequent mixing with dust and clouds makes it difficult to remotely monitor (Reid et. al., 2009). Finally, areas of very high pollution load, such as the sites on the Indo-Gangetic plain (Kanpur and Gandhi College) and Beijing also have persistent low biases (Table 1). Models that have secondary organic aerosol production have lower biases than those without. However, large uncertainties at this site also point at inadequacies of the emission inventories. Dust is also underrepresented for these sites.

A similar study of bias as above can be also conducted as a function of forecast day (Figure 6). After the generation of the forecast analysis through data assimilation and the forecast commences, both the meteorological and aerosol models will evolve into its free running behavior. Thus, in general we expect model biases to worsen in time as the model gets further and further away from the satellite observations that help initialize the run. In areas of poor natural model performance, the change in model bias with forecast time can be dramatic. Sometimes, site performance can completely reverse itself between monsoonal phases. Most notable is Ilorin in the African Sahel. Mean AOT biases become evermore negative in forecast time in the winter/spring period, reaching 50% of the mean value at 5 days. Similar biases are seen in Chiang Mai, Thailand. Bias change in the fine mode implicate biomass burning in this region. Again, these are the most complex burning regimes in the world (Reid et al., 2009; Reid et al., 2013a,b). However, for the summer/ fall, both sites do remarkably well. In the case of Ilorin, it is a result of a transition from mixed dust and biomass burning to dust dominated regime (Eck et al., 2010). For Chiang Mai, it is a result of the linear nature of consensus style ensembles-one model with a very large high bias counteracted three others with a moderate low bias. Also of note is Kanpur, India, which consistently demonstrates poor forecasting performance of all models-although its neighbor Gandhi College (not shown) only showed half the bias.

Some sites actually improve in time, such as, Baegnyeonng Korea, where there is statistically significant improvement in bias with forecast, in the winter spring. This could implicate bias in the analysis, as the free running forecasts relax into lower error states before being erroneously jarred into high error by the assimilation process. Other sites show little difference at all as forecast time increases, such as Beijing, Banizoumbou, Sahelian Africa, and winter/spring in Singapore, implicating the lesser impact of data assimilation in these regions.

4.0 Results: RMSE

While mean seasonal bias is important overall to many aerosol applications, for aerosol forecasting daily variability is equally if not more important. In this regard metrics such as the root mean square error (RMSE) become more appropriate for characterizing model skill. Since RMSE incorporates both bias and variance, additional steps can also be taken to perform bias removal in order to determine how well models capture aerosol variability. In a like manner to bias, Table 3 and 4, provide total AOT and dust RMSE for each site. These RMSEs are pictorially presented in Figure 7 against each site's mean AOT. Shown are each model's value (small dots color coded) and the RMSE for the ensemble mean (large blue data point). A likewise representation of RMSE for 4 day forecasts (84 to 108 hour model-AERONET matchups) is similarly presented in Figure 8 for total, and dust AOT. Total AOT RMSE and FGE as a function of forecast time for key sites are presented in Figure 9.

By definition, for biases the ICAP-MME ensemble mean provides no more information than the average of its members. As was clearly demonstrated, as all of the models tend to low bias average AOT, so does the ICAP-MME. If the model averages are evenly distributed around the true state the ICAP-MME will be without bias. For RMSE however, the situation is quite different, where typically we find the RMSE of the ensemble of skillful and independent models is superior than any individual members. We found this to be the case with ICAP MME. With RMSE (or mean absolute error, not shown), ICAP-MME provides the best performance. Examination of Figures 7 and Tables 3 and 4 shows that in nearly all cases the ICAP-MME RMSE is either the leader or the second best in RMSE. For dust in particular (in which all modeling groups emphasize development), the ICAP-MME is particularly skillful.

Based on the slope of RMSE against mean AOT value for each site in Figure 7, the RMSE's of the 1 day forecasts of ICAP-MME run approximately 50% of the climatological mean AOT value. Dust AOT forecasting is superior to overall fine and coarse mode AOT, running approximately 1/3rd of climatological AOT. Again, this is part reflects the importance of the dust species by centers. Further, the AERONET Cape Verde site (in which RMSE is particularly skillful) is a common benchmark site for Saharan dust-hence models are typically tuned for the region.

Regions of particular difficulty with RMSE are often the same as those with large biases. Chiang Mai and Singapore in their respective biomass burning seasons have some of the highest biases. Beijing China, Kanpur India and Ilorin in the Sahel have RMSE's that are more than half the mean AOT. But, if we account for mean AOT for the region in the Fractional Gross Error (FGE) errors at some low AOT sites become more pronounced. Perhaps most important of sites would be Baengnyeong Korea, a receptor for East Asia, with a normalized RMSE of 1.3, or a fractional gross error (FGE) of 0.55. Owing to the low baseline AOT and the difficulty with modeling and remote sensing in the southern Oceans, Crozet Island also appears to be poorly represented, with Normalized RMSE of 1.24, and a FGE of 0.67. Monterey CA, another marine site, also has FGEs in the 0.3-0.6 range.

Like bias, forecasting skill for all models and the ensemble mean degrades in time. Although the relative performance of the ICAP-MME mean relative to the member models increases in time, particularly for dust. In Figure 8, we show the RMSE versus AERONET AOT for 4 day forecasts, or three days after the first 24 hour baseline for the total and dust cases (Figure 7). In general, the RMSE's increase to 60% of the total AOT value from ~40% at one day. For dust, however, skills remain constant in time. These general trends can be seen even more clearly in RMSE and FGE as a function of forecast day of the ICAP-MME consensus (Figure 9).

5.0 Results: Rank Histograms

Thus far we have treated the ICAP-MME deterministically through comparisons of the ensemble mean to the individual members. Comparisons between models in bias and RMSE do tell us a general state of the modeling community. To move towards a goal of event driven applications of the ICAP-MME we can begin to view the ensemble members probabilistically and ask questions related to where individual observations fall relative to the model. The rank histogram (a.k.a Talagrand diagram) is a useful diagnostic to depict the relative distribution of observations and models (e.g., Hamill 2001). Rank histograms are constructed by repeatedly tallying the rank of the verifying observation relative to values from an ensemble sorted from lowest to highest. A flat rank histogram is usually taken as a sign of reliability, while a U-shaped rank histogram often indicates a lack of variability in the ensemble. In Figure 10 we present global rank histograms of the first forecast data day (6-24 hour forecast period) for all observations segregated biseasonally into the boreal winter (Dec-May) and summer(Jun-Nov) periods. Included are histograms for all AERONET matchups for our 22 sites (Figure 9a-d) as well as for those AERONET cases were AOT>0.6 (Figure 9e-h). This value of 0.6 is somewhat arbitrary, and was chosen to give balance between high AOT and enough data points to lend significance to the product. Plots are given for total, fine and coarse AOTs for the four core multispecies models (leading to 5 ranks), and dust for the core four plus NGAC (6 ranks).

As a rank histogram is a histogram as to where an observation falls relative to the models, it is useful to calculate and examine relative to the biases and RMSEs. For all data (Figure 10a), the histogram is relatively flat, with a slight slope with increasing rank. That is, the observations tend to be bigger than the individual members and the ICAP-MME mean. But, there is offsetting divergence in the individual aerosol particle size modes, with models generally overestimating fine mode AOT overall, and conversely underestimating coarse mode AOT. This is an agreement with the biases presented in Table 1. Thus, while the total AOT data histogram is relatively flat, it is flat for the wrong reasons with offsetting fine and coarse populations. If we examine more significant events for AOT>0.6 (Figure 10e-g) we see that overall the models are strongly low biased overall. That is, for dust, smoke, and pollution alike, the models are in general underestimating the most severe events.

These rank histograms are for all global observations and are generally representative for individual sites. In Figure 11 we present histograms for 15 of the 21 sites for the four multi-species models (to conserve space, plots that showed similar tendencies to neighbors were dropped). In the first column, sites of a background nature or as a long-range receptor are given. All of these sites are relatively clean and have average AOTs<0.15. In general, the histograms are relatively flat, although there is in general over prediction of AOT in the central United States, represented by the DOE CART site, and underrepresentation of dust at the Palma de Malloraca site in Spain as a receptor for dust. At Ragged Point, an African dust receptor in the Caribbean, the distribution is good. For sites with intermediate loadings or those that are taken as regionally representative of polluted areas (column 2), there is also a distribution of tendencies, with Singapore showing universal AOT under-representation in AOT, and Goddard Space flight center suggesting over representation. Most interesting are the heavily impacted sites (column 3 and 4), where we show all data plus those cases where AERONET AOT>0.6. Sites such as Beijing, China and Gandhi College, India for massive pollution, Baengnyeong Korea (an Asian receptor), and Cape Verde and Banizoubou for African dust, models have similar tendencies in regard to all data. But all models are strongly low biased for high AOT events. This shows that while the models are independent in the meteorology and parameterizations, they nevertheless succumb to correlated bias overall. As an example of typical behavior for a well characterized site we turn to Cape Verde as an example.

6.0 Results: Cape Verde and Kanpur as examples of issues related to forecasting significant events

A substantial motivation for operational aerosol forecasting is natural hazards and significant events forecasting. Thus, while it is important for models to generally reproduce the basic characteristics of the aerosol system via good bias and RMSE scores, it is perhaps equally important for the models to succeed in identifying significant and unusual events. Good RMSE scores by nature ensure the models have skill in predicting typical environments, but consistent bias and amplitude may cloud a model's value in more extreme situations. In the early stage of development we settled on an AOT of 0.8 to be a key benchmark for warning areas (e.g., Figure 2(c,d). For example, the MACC alert system which is aimed at detecting significant events for air quality exceedance, uses a threshold of 0.5, which can be shown to correspond to a particulate matter < 10 μm in diameter (PM_{10}) of approximately 50μg m^{-3}. The number of days during which this PM_{10} value is exceeded is used in European legislation as a threshold for fining EU countries. The value chosen for the ICAP-MME is largely subjective, and was agreed upon after an examination of AERONET data to find logical "2 sigma" events in heavily polluted regions. However, after deeper investigation, this became somewhat dissatisfying. In the context of a multi model ensemble, there are numerous subjective considerations in combining model products for the benefit of forecasters. For example, one model may have an amplitude consistent error (i.e., track AOT extremely well), but poor bias scores and threat scores. Others may have excellent amplitudes and biases overall, but have timing issues with significant events. As always, there is the potential for sampling bias in our observational dataset.

To conceptualize the above issues we considered two sites in detail: Cape Verde and Kanpur. The Cape Verde AERONET site is a long standing benchmark location for dust modeling. With more than 15 years of observations, it is one of AERONET's longest running providing not only satellite and model verification data, but also climatological aerosol trends. Given the significant amount of attention centers pay to modeling dust, it is no surprise that Cape Verde is a high scoring site for all models. In comparison, Kanpur is the lowest scoring site next to Beijing for the models. Given that Kanpur has a more contiguous data record than Beijing, we chose that site for further analysis.

6.1 Cape Verde

Cape Verde's location as a downwind receptor for African dust coupled with overall good model performance makes it a good location to study the nature of event scoring. The Boolean nature of threat scores is often problematic and there can be difficulty in this metric in first defining what constitutes an event. For air quality applications for example, an event can be referenced to a degree of violation. Near misses are frequently valuable from a forecasting point of view, both in magnitude and in temporal offset. Observations are also problematic, as clear sky bias can be a problem in both satellite and ground based observations, thus leading to a bias as to when one can verify. We can explore this further with the time series of AERONET coarse mode AOT and the ICAP-MME mean for the 1 year study period (Figure 12a). Differences in the dust seasonality are clear, with winter and spring months having a relatively low background with occasional significant events and a higher dust continuum during summer months with numerous high frequency events. An enlargement is provided in Figure 12(b) for the middle time series month of May. Examination of the data in combination of error statistics presented in Table 1 and 2 suggests that indeed the ICAP-MME is performing well. Scatter plots of the 12hr and 84hr forecasts for 00Z against AERONET (Figure 12c), representing 24 and 96 hours since the last satellite data assimilation cycle for the region, are quite good. However, there are clear outliers worth investigating from an events perspective.

In interpreting the regression of Figure 12(c), cases far to the right of the regression lines (say February 7, AERONET=1.15, ICAP MME=0.25) tend to be in association with residual cirrus contamination. Cases studies such as these were visually verified such as in the right satellite image in Figure 12(a). While such misses are infrequent, they nevertheless are reminders that no verification dataset is perfect, and in an unsupervised verification system, cases such as this can heavily affect scores. Data points far to the left of the regression line, are false alarm cases where presumably the models far over predicted a dust event that did not materialize. These cases are nearly all associated with the 84 hour forecast of isolated wintertime events, or 4 full days since the last satellite data assimilation cycle and thus are purely forecast meteorology driven. We found that errors dropped in half as forecast lengths decreased to about 2 days as the forecast meteorology became more accurate. However, there are also cases where when we track the peak in dust AOT, this peak arrives outside of AERONET verification data availability but within 12 hours. This artifact points to the necessity of loosening our verification criteria for the amplitude and timing for longer forecasts.

To help further describe the nature of the many Boolean skill scores it helps to provide an example for when we have the most confidence,: forecasts within 24 hours. Our first challenge is to define the threshold to be implemented. This can be done uniformally for all sites (say AOT>0.5, 0.8 or 1 etc), or it can be site specific, based on the probability of what is locally considered an extreme event. Figure 12(d) provides an AOT probability plot for the 1 year time series of the 12 hour and 84 hour forecasts. There is generally good agreement on the probability distribution of AOT between observations and corresponding 12 and 84 hour forecasts well past one geometric standard deviations (84.1 AOT percentile=0.50) to just short of two (97.7 AOT percentile=0.83). These lines are marked on Figure 12(a) and (b) as well as the common 1.5σ_g level (93 AOT percentile =0.62).

The difficulty in skill scores becomes apparent if we consider the 2σ_g level as a threshold. At 2σ_g there are 6 events recorded by AERONET (3 in May, Figure 12(b)), all of which were captured by the ICAP MME mean at 12 and 84 hours. However, at 12 hours, there were 6 false alarms. This leads to a TS or CSI of 0.5, and ETS of 0.48. This is a somewhat middling score. However, in five of the six false alarm cases, the observations reached at least 1.5 σ_g with remaining one was above the 1 sigma level. For 84 hour forecasts, the false alarm rate goes up to 11, but even here 6 reach the 1.5σ_g level. If we use 1.5σ_g as a threshold, the 12 hr TS goes up to 0.65 and the ETS to 0.58.

Between the above analysis and Figure 12(c) the model clearly has skill. However, the Boolean nature of the metrics can make interpretation difficult-particularly when one applies them uniformly over the globe. This situation is common in the Numerical Weather Prediction realm, and in response dozen of skill scores have been developed, including those with "fuzzy" neighborhood boundaries such as spatial multi-event contingency tables and fractional skill scores to (e.g., http://www.cawcr.gov.au/projects/verification/). If we move further to take advantage of the natural probabilistic applications of a multi- model ensemble, versions of Brier scores or the continuous rank probability score may also be appropriate. These are directions of research for the next set of multi-year ensemble data.

6.2 Kanpur

In contrast to Cape Verde, Kanpur represents a site with overall poor event scoring by all models for the common metrics as bias, RMSE and threat score. In this case, Kanpur provides a complex overall environment over land in opposition to the more simplified dust environment at the nominally oceanic site of Cape Verde. Kanpur district has a high population density (~4.5 million), has high industrial and biofuel emissions, is a receptor for dust from all along the Indo-Gangetic plane and as is key here, and a complex aerosol meteorology-particularly in wintertime (e.g., Nair et al., 2007; Gautam et al., 2007, 2009; 2011; Kar et al., 2010; Arola et al., 2013). Given such complexity, it is little wonder that the global models have great difficulty with the region in the context of common metrics. But, after further examination and consideration of the nature of global modeling, we find that bulk metrics do not entirely describe model performance-particularly in regard to extreme events.

Figure 13 provides data of a similar nature as shown in Figure 12 for Cape Verde. Although here we provide fine mode data for the four multi-species models, and all five models under current analysis with dust. Beginning with fine mode comparison, we find that the 12 hour forecast nominally tracks the overall nature of the regions aerosol pattern-although with a significant low bias in the winter months. Also in the winter months is when we find significant spikes in fine mode AOT. These are quite often haze events created during the evaporation of winter time stratocumulus (Eck et al., 2010). Under such circumstances, global models are unlikely to cope with such strong boundary layer meteorological forcing. In contrast, we see that in the spring, when pollution events are more regional, the models have some skill in at least simulating event onset-albeit with a significant low bias. When taken as a whole skill scores for correlations are reasonable for 12 hour forecasts, or nominally 18-24 hours since the last satellite observations were assimilated (r^2=0.58). However by the time forecasts reach three to four days, models appear to lose all fine mode skill.

For dust, the models appear in general to perform better. Regressions are decent at both 12 and 84 hours (r^2=~0.6). But in this case, there are no "events." The distribution of coarse mode AOT observations are so tight, there are few to no observations past the 1.5 standard deviation level. At one geometric standard deviation, exceedances are in a continuum. Thus, a threat score does not provide sufficient context to evaluate models in this environment.

Finally, Kanpur highlights a further situation with verification data. While the SDA algorithm does an admirable job separating fine and coarse mode AOT, in this case coarse mode is a combination of aeolian dust (which is generally the context of dust in the global models), and regional coarse mode species, including agriculture, industrial or road dust, as well as perhaps droplets in the cloud burn off phase. This seems to be particularly true in the winter periods.

Thus, is the use of the term coarse mode cannot be used synonymously with “dust” in classical terms, and thus provides an additional challenge to the global models.

7.0 Discussion and Conclusions

This paper describes the basic climatological characteristics and evaluation of the world’s first global multi-model aerosol forecast model-the International Cooperative for Aerosol Research Multi Model ensemble: ICAP-MME. At the writing of this paper, there are 4 core multi species models (ECMWF, JMA MASINGAR, NASA GEOS-5, NRL NAAPS) and seven dust models (aforementioned four, plus NMMB/BSC-CTM, NOAA NGAC, and UKMO Unified Model) running daily at 00Z with 24 hour latency. Here we focus on the first year of data, from December 1, 2011 through November 30, 2012 when all four multi-species models plus NGAC were providing data in near real time. We expect rapid evolution in the individual member models based on these results and similar exercises with ICAP-MME products. Thus, the error metrics are likely out of date for the better at the publishing of this initial research. Further, as models are added to the ICAP-MME we expect better performance. The initial state of the ICAP-MME is worth documenting for base lining purposes, and the general tendencies in the state of global aerosol forecasting models are worth discussing. These are listed here:

1. *Overall performance via RMSE:* As we expected when we first constructed the ICAP-MME, the ensemble mean outperforms all of its individual members in RMSE against AERONET globally throughout the forecast period. Typically RMSE runs 40-60% of the mean AOT with coarse mode prediction outperforming fine mode. Given that RMSE has both a bias and variance component, and the ensemble mean bias is by definition in the middle of the

members, the improvement in variance prediction is significant. Like other ensemble based systems like the tropical cyclones (Leslie and Fraedrich, 1990; Mundell and Rupp, 1995; Goerss, 2000; DeMaria et al. 2006; Kaplan and DeMaria 2001; Sampson, 2010) and GCMs (e.g., Meehl et al., 2007; Knutti et al., 2010; Reichler and Kim 2008), we expect that as individual models improve and are added, so will the consensus. Indeed, even though NGAC has average performance relative to other dust models, it did improve the overall RMSE of the ICAP-MME for dust.

2. Overall performance via bias: In general all models and thus the ensemble mean capture the major climatological aerosol features around the globe. However, while the models perform well in RMSE, there is a tendency for the modeling community to have a low bias in AOT, particularly for significant events. Conversely, for more moderate or clean conditions, fine mode AOT is overestimated. These biases seem to be persistent in the modeling community, and documentation dates back to the AeroCom comparisons of Kinne et al. (2006). This persistence in low bias dust in models is perplexing, because one would think the community would tune around the observation. In the case of the forecast models, the assimilation of MODIS and verification via AERONET are ubiquitous. Further, regression is probably the most commonly used tuning metric, and as it is driven by the largest magnitude values we were surprised to find the under representation of AOT. We can surmise that in some heavily polluted urban site like Beijing, large scale models cannot represent fine scale features nor are there observations for extreme events (the maximum AOT measureable by AERONET is ~5). But regional polluted sites compared to urban counterparts (such as Gandhi College versus Kanpur) the biases remain. It may represent an overall reluctance by model developers to perturb or tune static emissions inventories. Thus, this persistent bias among models might also have a psychological supporting factor too in the way scientists interpret pollution data versus other species such as dust and biomass burning. In regard to ICAP-MME, all core models have satellite data assimilation in some cases, remote sensing biases can then work their way into forecast climatological biases. Even so, some species remain problematic. There is more diversity in climatological biomass burning AOTs than any other species. Despite the low AOTs, diversity in sea salt AOTs in the high mid latitudes is also large. Tracking this effect is a goal of future efforts.

3. *Site specific performance:* AERONET sites were picked by mutual agreement by the model developers based on data representativeness and availability. There are clearly regions of relative high and low model performance. Cape Verde is a widely used AERONET site for monitoring dust emissions from the Sahara, and models in general tune to this site to great effect even though the benefit of data assimilation is marginal outside of the analysis period. Aerosol receptor sites or those sites which will have the benefit of data assimilation also tend to score well such as Palma de Mallorca and Ragged Point. There are also sites with universal difficulty. Models clearly have more difficulty with sites in the mixed fine and coarse mode environments of the Sahel, India and polluted cities of Asia. Cloud cover impacts on data assimilation are also likely a factor in sites such as Singapore and Chiang Mai.

4. *Future directions:* This is the first paper on the ICAP-MME and there are clearly many directions in which studies may proceed. Perhaps the most common question received by developers on future direction is whether we intend to convert the ICAP-MME to a super ensemble where models are weighted by their scores (e.g., Krishnamurti et al., 1999; Casanova and Ahrens, 2009). Experience has shown however that equal weighting in a consensus style

appears to provide the most robust results overall, and this is backed up on both practical and theoretical grounds (DelSole et al., 2013). Further, we frequently see regional improvements to ensemble members as the models develop, and different models score differently by region or type of event. In the operational realm reanalyses cannot always be generated and significant events by nature are so rare that tuning will likely be unrepresentative. Thus, for all of these reasons an operational super ensemble is impractical at this time. Although in the future adaptive systems may be possible. But, the underlying premise that individual models be continuously scored uniformly is highly relevant to the field-particularly for major events. Now that a common dataset has been generated developers are now in a position to agree upon standard metrics and protocols to ensure that performance improvement and best practices are cleanly documented across models. A second area for future direction related to metrics is to take advantage of the probabilistic nature of the ICAP-MME. Already consensus threat scores and warning areas have been defined. These clearly need to be explored further. Finally, given that the ICAP-MME members share some development legacy and at times exhibit similar forecast outcomes we intend to probe the relative independence of the models.

9.0 Acknowledgments

The authors are greatly indebted to their individual programs for supporting ICAP and the development of the multi-model ensemble. We recognize and appreciate the countless researchers and computer engineers whose work supports the development and distribution of aerosol forecasts. As data assimilation is key to model performance, we are grateful to NASA LANCE-MODIS for providing MODIS near real time data used in nearly all of the models here. We also acknowledge the effort of the AERONET team (project leader Brent Holben) and the various site principal investigators and site managers of the numerous AERONET sites utilized in this study. Funding for the development of the construction of ICAP-MME was provided by the Office of Naval Research, Code 322. Angela Benedetti, Jean-Jacques Morcrette and Miha Razinger were supported through the MACC-II project, which is funded by the European Commission under the EU Seventh Research Framework Programme, contract number 283576. MASINGAR is developed in the Meteorological Research Institute of Japan Meteorological Agency, and a part of the development was funded by the Environmental Research and

Technology Development Fund (B-1202) of the Ministry of the Environment (MOE) of Japan. NAAPS development is supported by the Office of Naval Research Code 322, and PMW-120. NGAC development has been supported by Joint Center for Satellite Data Assimilation, NASA Applied Science Program, and NOAA National Weather Service. NMMB/BSC-CTM development is supported by the Spanish Government under grants CGL2010/19652, CSD2007-0050 and the grant SEV-2011-00067 of Severo Ochoa Program.

10.0 References:

Arola, A., Eck, T. F., Huttunen, J., Lehtinen, K. E. J., Lindfors, A. V., Myhre, G., Smirnov, A., Tripathi, S. N., and Yu, H.: Influence of observed diurnal cycles of aerosol optical depth on aerosol direct radiative effect, Atmos. Chem. Phys., 13, 7895-7901, doi:10.5194/acp-13-7895-2013, 2013.

Baldwin, M. E. and Kain, J. S.: Sensitivity of several performance measures to displacement error, bias, and event frequency. Weather Forecasting, 21, 636-648, 2006.

Benedetti, A., and Fisher, M.: Background error statistics for aerosols, Q. J. R. Meteorol. Soc., 133, 391– 405, doi: 10.1002/qj.37, 2007.

Benedetti, A., Morcrette, J.-J, Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, doi: 10.1029/2008JD011115, 2009

Benedetti, A., Baldasano, J. M., S. Basart3, S., Benincasa. F., O.Boucher, O., Brooks, M., Chen12, J.-P., Colarco, P. R., S., Gong, S., Huneeus, N., Jones, L., Lu, S., Menut, L., Morcrette1, J. J., Mulcahy, J., Nickovic, S., Pérez, C., Reid, J. S., Sekiyama, T. T., T.Y. Tanaka T. Y., Terradellas, E., Westphal, D. L., X.-Y. Zhang, X. Y., Zhou, C. H.,Operational Dust Prediction in: Knippertz, P.; Stuut, J.-B. (eds.), /Mineral Dust – A Key Player in the Earth System/, Springer Netherlands, 121–148, ISBN 978-94-017-8977-6. doi:10.1007/978-94-017-8978-3_6 <http://dx.doi.org/10.1007/978-94-017-8978-3_10, 2014.

Black, T., Juang, H. M. H., and Iredell, M.: The NOAA Environmental Modeling System atNCEP, Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conferenceon Numerical Weather Prediction, Omaga, NE, Ameri. Met. Soc., 2 A.6, available at:https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154223.htm (last access: 29 May 2014), 2009.

Boucher, O., Pham, M., and Venkataraman, C.: Simulation of the atmospheric sulfur cycle in the LMD GCM: Model description, model evaluation, and global and European budgets, Note 23, 26 pp., Inst. Pierre-Simon Laplace, Paris, France, available at: http://icmc.ipsl.fr/images/ publications/scientific_notes/note23.pdf (last access: 6 June 2014), 2002.Buizza, R., Houtekamer, P, L., Pellerin, G., Toth, Z., Zhu, Y., and Wei, M.: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. Mon. Wea. Rev., 133, 1076–1097, 2005.

Carn, S. A., Krueger, A. J., Krotkov, N. A., Yang, K., and Evans, K.: Tracking volcanic sulfur dioxide clouds for aviation hazard mitigation. Natural hazards, 51(2), 325-343, 2009

Casanova S, and Ahrens B.: On the weighting of multimodel ensembles in seasonal and short-range weather forecasting. Mon. Weather Rev., 137, 3811–3822, 2009.

Chew, B. N., Campbell, J.R., Reid, J. S., Giles, D. M., Welton, E. J., Salinas, S. V., and Liew, S. C.: Tropical cirrus cloud contamination in sun photometer data, Atmos. Environ., 45, 6724-6731, doi:10.1016/j.atmosenv.2011.08.017, 2011.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, J. Atmos. Sci., 59, 461-483, 2002.

Christensen, J. H.: The Danish eulerian hemispheric model—A three dimensional air pollution model used for the Arctic, Atmos. Environ., 31, 4169– 4191, 1997.

Colarco, P. R., Schoeberl, M. R., Doddridge, B. G., Marufu. L. T., Torres, O., and Welton, E. J.: Transport of smoke from Canadian forest fires to the surface near Washington, D.C.: Injection height, entrainment, and optical properties, J. Geophys. Res., 109, D06203, doi:10.1029/2003JD004248, 2004.

Colarco, P., da Silva, A., Chin, M. and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, *J. Geophy. Res.*, 115, D14207, doi:10.1029/2009JD012820, 2010.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051–1075, doi:10.5194/gmd-4- 1051-2011, 2011

Darmenov, A., and da Silva, A.: The Quick Fire Emissions Dataset (QFED) - Documentation of versions 2.1, 2.2 and 2.4. NASA Technical Report Series on Global Modeling and Data Assimilation. NASA TM-2013-104606, Vol. 32, 183 pp., 2013.

Damoah, R., Spichtinger, N., Forster, C., James, P., Mattis, I., Wandinger, U., Beirle, S., Wagner, T., Stohl, A. Around the world in 17 days - hemispheric - scale transport of forest fire smoke from Russia in May 2003, Atmos. Chem. Phys., 4, 1311-1321, 2004

Dee, D.P. and Uppala, S.: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. Quart.J. Roy.Meteor.Soc., 135,1830-1841,2009.

Dee, D., and da Silva, A.: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I:Methodology. Mon. Wea. Rev., 124, 1669–1694, 1999.

Dee, D. P., Rukhovets, L., Todling, R., Da Silva, A. M., and Larson, J.W.: An Adaptive Buddy Check for Observational Quality Control, Quarterly Journal of the Royal Meteorological Society 127 (577) (January 1): 2451–2471, 2001.

DelSole, T., Yang, X., and Tippett, M. K.: Is unequal weighting seignificantly better than equal weighting for multi-model forecasting?, Quart. J. Royal Meteorol. Soc., 139, 176-183, 2013.

DeMaria, M., Knalff, J. A., and Kaplan, J.: On the decay of tropical cyclone winds crossing narrow landmasses, J. Appl. Meteor. and Clim., 45, 491-499, 2006.

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J. J., Ito, A., Marelli, L., Penner, J. E., Putaud, J.-P., Textor, C., Schulz, M., van der Werf, G. R., and Wilson, J.: Emissions of primary aerosol and precursor gases in

the years 2000 and 1750 prescribed data-sets for AeroCom, Atmos. Chem. Phys., 6, 4321 – 4344, doi:10.5194/acp-6-4321-2006, 2006.
Dunion. J. P., and Velden, C.: The impact of the Saharan air layer on Atlantic cyclone activity, Bull. Amer. Meteor. Soc., 85, 353-365, 2004.
Dunstone, N.J., Smith, D. M., Booth, B. B. B., Hermanson, L., and Eade, R.: Anthropogenic aerosol forcing of Atlantic tropical storms, Nature Geosci., 6, 534-539, 2013.
Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophys. Res., 104, D24, 31,333–31,349, 1999.
Eck, T. F., Holben, B. N., Sinyuk, A., Pinker, R. T., Goloub, P., Chen, H., Chatenet, B., Li, Z., ingh, R. P., Tripathi, S. N., Reid, J. S., Giles, D. M., Dubovik, O., O'Neill, N. T., Smirnov, A., Wang, P., and Xia, X : Climatological aspects of the optical properties of fine/coarse mode aerosol mixtures, J. Geophys. Res., 115, D19205, doi:10.1029/2010JD014002, 2010.
Evan, A. T., Heidinger, A. K., Bennartzm, R., Bennington, V., Hamhwald, N. M., Corrada-Bravo, H., Velden, C. S., Myhre, G., and Kossin, J. P.: Ocean temperatures forcing by aerosols across the Atlantic tropical cyclone development region. Geochem. Geophys. Geosys., 9, Q05V04, doi: 10.1029/2007GC001774, 2008.
Fecan, F. and Marticorena, B., and Bergametti, G.: Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil mosture for arid and semi-arid areas. Ann. Geophysicae. 17, 149-157 doi:10.1007/s00585-999-0149-7., 1999.
Fordham, D. A., Wigley, T. M., Watts, M. J., & Brook, B. W.: Strengthening forecasts of climate change impacts with multi‐model ensemble averaged projections using MAGICC/SCENGEN 5.3. Ecography, 35(1), 4-8., 2012.
Gautam, R., Hsu, N. C., Kafatos, M., and Tsay, S.-C.: Influences of winter haze on fog/low cloud cover over the Indo-Gangetic plains, J. Geophys. Res., 1112, D05207, doi:10.1029/2005JD007036, 2007.
Gautam, R., Liu, Z., Singh, R. P., and Hsu, N. C. : Two contrasting dust-dominant periods over India observed from MODIS and CALIPSO data, Geophys. Res. Lett., 36, L06813, doi:10.1029/2008GL036967, 2009.
Gillette, D.: A wind tunnel simulation of the erosion of soil: Effect of soil texture, sandblasting, wind speed, and soil consolidation on dust production. Atmos. Environ. 12, 1735-1743, 1978.
Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B. N., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106(D17), 20255–20273, doi:10.1029/2000JD000053, 2001.
Goerss, J. S.: Tropical cyclone track forecasts using an ensemble of dynamical models. Mon. Wea. Rev., 128, 1187–1193, 2000.
Gong, S. L.: A parameterization of sea-salt aerosol source function for sub and super-micron particles, Global Biogeochem. Cy., 17, 1097, doi:10.1029/2003GB002079, 2003.
Guelle, W., Schulz, M., Balkanski, Y., and Dentener, F.: Influence of the source formulation on modeling the atmospheric global distribution of the sea salt aerosol, J. Geophys. Res., 106, 27,509– 27,524, 2001.
Kaplan, J., and DeMaria, M.: On the decay of tropical cyclone winds after landfall in the New England area. J. Appl. Meteor., 40, 280-286, 2001.
Hamill, T. M.: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. Mon. Wea. Rev., 129, 550–560, 2001.

Haustein, K., Pérez, C., Baldasano, J. M., Jorba, O., Basart, S., Miller, R. L., Janjic, Z., Black, T., Nickovic, S., Todd, M.C., Washington, R., Muller, D., Tesche, M., Weinzierl, B., Esselborn, M., and Schladitz, A.: Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model - Part 2: Experimental campaigns in Northern Africa, Atmospheric Chemistry and Physics, 12, 2933-2958, doi:10.5194/acp-252 12-2933-2012, 2012.

Haywood, J. M., Allan, R. P., Culverwell, I., Slingo, T., Milton, S., Edwards, J., and Clerbaux, N.: Can desert dust explain the outgoing longwave radiation anomaly over the Sahara during July 2003, J. Geophys. Res., 110, doi:10.1029/2004JD005232, 2005.

Hess, M., Koepke, P., and Schult, I.: Optical properties of aerosols and clouds: The software package OPAC. Bulletin Of The American Meteorological Society, 79, 831-844, 1998.

Hill, C., DeLuca, C., Balaji, V., Suarez, M., da Silva, A., and the ESMF Joint Specification Team: The Architecture of the Earth System Modeling Framework. IEEE Computing in Science and Engineering, Grand Challenges in Earth System Modeling, 6, 19–28, 2004.

Hogan, T. F., and Rosmond, T. E.: The description of the Navy Operational Global Atmospheric Prediction System's Spectral Forecast Model. Mon. Wea. Rev., 119, 1786–1815, 1991.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A.,Vermote, E., Reagan, J. A., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, 66, 1 –16, 1998.

Hsu N., Tsay S. C., King M. D., Herman J. R.: Aerosol properties over bright-reflecting source regions. IEEE Trans. on Geosci. and Rem. Sens. 43: 557-569, 2004.

Hsu N, Tsay S. C., King M.: Deep blue retrievals of Asian aerosol properties during ACE-Asia. IEEE Trans. on Geosci. and Rem. Sens. 44: 3180-3195, 2006.

Hyer, E. J., Reid, J. S., and Zhang, J.: 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, 379-408, doi:10.5194/amt-4-379-2011, 2011.

Hyer, E. J., Reid, J. S., Prins, E., Hoffman, J. P., Schmidt, C. C., Miettinen, J. I., and Giglio, L.: Patterns of Fire Activity over Indonesia and Malaysia from Polar and Geostationary Satellite Observations. Atmos. Res., 122, 504-519, doi: 10.1016/j.atmosres.2012.06.011, 2013.

Ignatov, A. and Gutman, G.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, Int. J. Remote Sens., 19, 1533–1543, doi:10.1080/014311698215333, 1998.

Jaeglé, L., Quinn, P. K., Bates, T. S., Alexander, B., and Lin, J.-T.: Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations, Atmos. Chem. Phys., 11, 3137–3157, doi:10.5194/acp-11-3137-2011, 2011.

Janjic, Z., and Gall R.: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB), Part 1 Dynamics, Tech. rep. NCAR/TN-268 489+STR, 2012. available at: http://nldr.library.ucar.edu/repository/collections/TECH269NOTE-000-000-000-857 (last access: December 2013), 2012.

Janjic, Z., Janjic, T., and Vasic R.:A class of conservative fourth order advection schemes and impact of enhanced formal accuracy on extended range forecast, Monthly Weather Review, 139, 1556-1568, doi:10.1175/2010MWR3448.1, 2011.

Janjic, Z. I.: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes, Mon. Weather Rev., 122, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2, 1994.

Johnson B.T., Brooks M.E., Walters D., Christopher S., and Schepanski, K.: Assessment of the met office dust forecast model using observations from the GERBILS campaign. Quart. J. Roy. Meteor. Soc. 137: 1131–1148.doi:10.1002/qj.736, 2011.

Jorba, O., D. Dabdub, D., Blaszczak-Boxe, C., Péerez, C., Janjic, Z., Baldasano, J. M., Spada, M., Badia, A. and Gonçalves M.: Potential Significance of Photoexcited NO2 on Global Air Quality with the NMMB/BSC Chemical Transport Model, Journal of Geophysical Research, 117, D13301, doi:10.1029/2012JD017730, 2012.

Kaiser, J.W., A. Heil, M.O. Andreae, A. Benedetti, N. Chubarova, L. Jones, J.-J.Morcrette, M. Razinger: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, 2012.

Kaplan, J., and Demaria, M.: On the decay of tropical cyclone winds after landfall in New England area, J. Appl. Meteor., 40, 280-286, 2001.

Kar, J., Deeter, M. N., Fishman, J., Liu, Z., Omar, A., Creilson, J. K., Trepte, C. R., Vaughan, M. A., and Winker, D. M.: Wintertime pollution over the Eastern Indo-Gangetic Plains as observed from MOPITT, CALIPSO and tropospheric ozone residual data, Atmos. Chem. Phys., 10, 12273-12283, doi:10.5194/acp-10-12273-2010, 2010.

Karyampudi, V. M., and Carlson, T. N.: Analysis and numerical simulations of the Saharan air layer and its effect on easterly wave disturbances, *J. Atmos. Sci*., 45, 3103 – 3136, 1988.

Karyampudi, V. M., and Pierce, H. F.: Synoptic-scale influence of the Saharan air layer on tropical cyclogenesis over the Eastern Atlantic. Mon. Wea. Rev., 130, 3100-3128, 2002.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R., Flynn, L.: The MODIS 2.1 μm channel—Correlation with visible reflectance for use in remote sensing of aerosol, IEEE Trans. Geos. and Remote Sens., 35(5), 1286–1298, 1997a.

Kaufman, Y. J., Tanre, D., Remer, L. A., Vermote, E. F., Chu, A., and Holben, B. N.: Operational remote sensing of tropospheric aerosol over the land from EOS-MODIS, J. Geophys. Res., 102, 17,051–17,061, 1997b.

Kettle, A., Andreae, M., Amouroux, D., Andreae, T., Bates, T., Berresheim, H., Bingemer, H., Boniforti, R., Curran, M., DiTullio, G., Helas, G., Jones, G., Keller, M., Kiene, R., Leck, C., Levasseur, M., Malin, G., Maspero, M., Matrai, P., McTaggart, A., Mihalopoulos, N., Nguyen, B., Novo, A., Putaud, J., Rapsomanikis, S., Roberts, G., Schebeske, G., Sharma, S., Simo, R., Staubes, R., Turner, S. and Uher, G.: A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month, Global Biogeochem Cy, 13(2), 399–444, 1999.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Berntsen, T., Berglen, T. F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment – optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815-1834, 10.5194/acp-6-1815-2006, 2006.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models. J. Climate, 23, 2739–2758, 2010.

Koe, L. C. C., Arellano Jr., A. F., and McGregor, J. L.: Investigating the haze transport from 1997 biomass burning in Southeast Asia: its impact upon Singapore, Atmos. Environ., 35, 2723–2734, 2001.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S. , Surendran, S.: Improved weather and seasonal climate forecasts from multimodel superensemble, Science, 285, 1548-1550, 1999.

Kukkonen, J., Olsson, T., Schultz, D. M., Baklanov, A., Klein, T., Miranda, A. I., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, A., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K. E. J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., and Eben, K.: A review of operational, regional-scale, chemical weather forecasting models in Europe, Atmos. Chem. Phys., 12, 1-87, doi:10.5194/acp-12-1-2012, 2012.

Lary, D., Remer, L. A., MacNeil, D., Roscoe, B., and Paradise, S.: Machine Learning and Bias Correction of MODIS Aerosol Optical Depth. Ieee Geosci. Temote Sens. Lett., 6, 694–698, doi:10.1109/LGRS.2009.2023605, 2010..

Leslie, L. M., and Fraedrich, K.: Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. Weather and Forecasting, 5(1), 158-161, 1990

Levy, R., Remer, L., Mattoo, S., Vermote, E., and Kaufman, Y. J.: Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, J. Geophys. Res.-Atmos., 112(D13), D13211, 10.1029/2006JD007811, 2007.

Levy, R., Remer, L., Tanre, D., Mattoo, S., and Kaufman, Y.: Algorithm for remote sensing of tropospheric aerosol over dark targets from MODIS: Collections 005 and 051: Revsion 2, February 2009, MODIS Algorithm Theoretical Basis Document, available at: http: //modis-atmos.gsfc.nLorenz, E. N.: Deterministic nonperiodic flow. Journal of the atmospheric sciences, 20(2), 130-141, 1963.

Lorenz, E. N.: A study of the predictability of a 28 ‐ variable atmospheric model. Tellus, 17(3), 321-333, 1965.

Lorenz, E. N.: Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric sciences, 26(4), 636-646, 1969a.

Lorenz, E. N.: The predictability of a flow which possesses many scales of motion. Tellus, 21(3), 289-307, 1969b.

Lu, S., Huang, H.-C., Hou, Y.-T., Tang, Y., McQueen, J., da Silva, A., Chin, M., Joseph, E., and Stockwell, W.: Development of NCEP Global Aerosol Forecasting System: an Overview and its Application for Improving Weather and Air Quality Forecasts, NATO Science for Peace and Security Series: Air Pollution Modeling and Its Application XX, available at: http://www.jcsda.noaa.gov/documents/meetings/wkshp2009/Session-3/3.11.Poster-Sarah.Lu.pdf (last access: 29 May 2014), 451–454, doi:10.1007/978-90-481-3812-8, 2010

Lu, S., Iredell, M., Wang, J., Moorthi, S., McQueen, J., Chuang, H.-Y., Hou, Y. –T., H. Juang, H., Yang, W., da Silva, A., Chin, M.: The NEMS GFS Aerosol Component: NCEP's global aerosol forecast system, NCEP Office Note 472, available at: http://www.lib.ncep.noaa.gov/ ncepofficenotes/files/on472.pdf (last access: 29 May 2014), Washington D.C., 2013.

Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H., and Hansson, H.-C.: Laboratory simulations and parameterization of the primary marine aerosol production, J. Geophys. Res., 108, 4297, doi:10.1029/2002JD002263, 2003.

Marticorena, B. and Bergametti, G.: Modeling the atmospheric dust cycle: 1. design of a soil-derived dust emission scheme, J. Geophys. Res., 100, 16415–16430, doi:10.1029/95JD00690, 1995.

Marticorena, B., Bergametti, G., Aumont, B., Callot, Y., N' Doumé, C., and Legrand, M.: Modeling the atmospheric dust cycle 2. simulation of Saharan dust sources, J. Geophys. Res., 102, 4387–4404, doi:10.1029/96JD02964, 1997.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchel, J. F. B., Stouffer, R. J., and Taylor, K. E.: The WCRP CMIP3 multimodel dataset-A new era in climate change research, Bull. Amer. Meteorol. Soc., 88, 1383-1394, 2007.

Meng, Z., Yang, P., Kattawar, G. W., Bi, L., Liou, K. N., and I. Laszlo, I.: Single-scattering properties of tri-axial ellipsoidal mineral dust aerosols: A database for application to radiative transfer calculations. J Aerosol Sci, vol. 41, no. 5, pp. 501-512, 2010.

Merchant, C. J., Embury, O., Le Borgne, P., Bellec, B.: Saharan dust in nightime thermal imagery: Detection and reduction of the related biases in retrieved sea surfacetemperature, Remote Sens. Of Environ., 104, 15-30, 2006.

Miller, T. P., and Casadevall, T. J.: Volcanic Ash Hazards to Aviation, in Encyclopedia of Volcanoes, Sigurdsson, H. (ed.), Academic Press, San Diego, California, USA, p. 915-930, 2000.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102, 16663–16682, doi:10.1029/97JD00237, 1997.

Monahan, E. C., Spiel, D. E., and Davidson, K. L.: A model of marine aerosol generation via whitecaps and wave disruption, in Oceanic Whitecaps, edited by E. C. Monahan and G. MacNiocaill, pp. 167–174, D. Reidel, Norwell, Mass, 1986.

Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M., Serrar, S., Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M., and Untch, A.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modeling, *J. Geophys. Res*., 114, D06206, doi:10.1029/2008JD011235, 2009.

Mundell, D. B., and Rupp, J. A.: Hybrid forecast aids at the Joint Typhoon Warning Center: Applications and results. Preprints, 21st Conf. on Hurricanes and Tropical Meteorology, Miami, FL, Amer. Meteor. Soc., 216–218, 1995.

Nickovic, S., Kallos, G., Papadopoulos, A., and Kakaliagou, O.: A model for prediction of desert dust cycle in the atmosphere, J. Geophys. Res., 106, 18113–18130, doi:10.1029/2000JD900794, 2001.

Nair, V. S., Moorthy, K. K., Alappattu, D. P., Kunhikrishnan, P.K., George, S., Nair, P. R., Babu, S. S., Abish, B., Satheesh, S. K., Tripathi, S. N., Niranjan, K., Madhavan, B. L., Srikant, V., Dutt, C. B. S., Badarinath, K. V. S., and Reddy, R. R.: Wintertime aerosol characteristics over the Indo-Gangetic Plain (IGP): Impacts of local boundary layer processes and long-range transport, J. Geophys. Res.-Atmos., 112, D13205,doi:10.1029/2006jd008099, 2007.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman S.: Spectral discrimination of coarse and fine mode optical depth. J. Geophys. Res., 108, D05212, doi:10.1029/2002JD002975, 2003.

O'Neill, N. T., Eck, T. F., Reid, J. S., Smirnov, A., and Pancrati O.: Coarse mode optical information retrievable using ultraviolet to short-wave infrared Sun photometry: Application to United Arab Emirates Unified Aerosol Experiment data. J. Geophys. Res., 113, D05212, doi:10.1029/2007JD009052, 2008.

Nyeki, S., C. H. Halios, W. Baum, K. Eleftheriadis, H. Flentje, J. Gröbner, L. Vuilleumier, and C. Wehrli: Ground-based aerosol optical depth trends at three high-altitude sites in Switzerland and southern Germany from 1995 to 2010, J. Geophys. Res., 117, D18202, doi:10.1029/2012JD017493, 2012.

Morcrette, J.-J., Beljaars, A., Benedetti, A., Jones, L., and Boucher, O.: Sea-salt and dust aerosols in the ECMWF IFS mode, Geophys. Res. Lett., 35, L24813, doi:10.1029/2008GL036041., 2008.

Mulcahy, J. P., Walters, D. N., Bellouin, N., and Milton, S. F.: Impacts of increasing the aerosol complexity in the Met Office global numerical weather prediction model, Atmos. Chem. Phys., 14, 4749-4778, doi:10.5194/acp-14-4749-2014, 2014.

Park, R., Jacob, D., Chin, M. and Martin, R.: Sources of carbonaceous aerosols over the United States and implications for natural visibility, J Geophys Res-Atmos, 108(D12), 4355, doi:10.1029/2002JD003190, 2003.

Parrish, D. F., and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Mon. Weather Rev.,120, 1747– 1763, 1992.

Pawson, S., Stolarski, R. S., Douglass, A. R., Newman, P. A., Nielsen, J. E., Frith, S. M., and Gupta, M. L.: Goddard Earth Observing System chemistry-climate model simulations of stratospheric ozone- temperature coupling between 1950 and 2005. Journal Of Geophysical Research-Atmospheres , 113, D12103, doi:10.1029/2007JD0095112008, 2008.

Pérez, C., Nickovic, S. Pejanovic, G., Baldasano, J. M., and Özsoy, E.: Interactive dust-radiation modeling: A step to improve weather forecasts, J. Geophys. Res., 111, D16206, doi:10.1029/2005JD006717, 2006.

Pérez, C., Haustein, K., Janjic, Z., Jorba, O., Huneeus, N., Baldasano, J. M., Black, T., Basart, S., Nickovic, S., Miller, R. L., and Perlwitz, J.: An online mineral dust aerosol model for meso to global scales: 1. Model description, annual simulations and validation, Atmos. Chem. Phys., 11, 13001–13027, doi:10.5194/acp-11-13001-2011, 2011..

Poschl, U.: Atmospheric aerosols: Composition, transformation, climate and health effects, Angewandte Chemie – International Edition, 44, 7520–7540, 2005.

Rajeev, K., Parameswaran, K., Nair, S. K.,and Meenu, S.: Observational evidence for the radiative impact of Indonesian smoke in modulating the sea surface temperature of the equatorial Indian Ocean, J. Geophys. Res., 113, D17201, doi: 10.1029/2007JD009611., 2008.

Reale, O., Lau, K. M., da Silva, A. M.: Impact of interactive aerosol on the African Easterly Jet in the NASA GEOS-5 global forecasting system. Wea. Forecasting. 26, 504–519. doi: 10.1175/WAF-D-10-05025.1, 2011.

Reale, O., Lau, K. M., da Silva, A., and Matsui, T.: Impact of assimilated and interactive aerosol on Tropical Cyclogenesis, Geophys. Res. Lett., 41, doi:10.1002/2014GL059918, 2014.

Reddy, M. S., Boucher, O., Bellouin, N., Schulz, M., Balkanski, Y., Dufresne J.-L., and Pham M.: Estimates of global multi-component aerosol optical depth and direct radiative perturbation in the Laboratoire de Météorologie Dynamique general circulation model, Journal of Geophysical Research, 110, D10S16, doi:10.1029/2004JD004757, 2005.

Reichler, T., and Kim, J.: How well do coupled models simulate today's climate?. Bulletin of the American Meteorological Society, 89, 303–311, doi:10.1175/BAMS-89-3-303, 2008.

Reid, J.S., Hyer, E. J., Prins, E. M., Westphal, D. L., Zhang, J., Wang, J., Christopher, S. A., Curtis, C. A., Schmidt, C. C., Eleuterio, D. P., Richardson, K. A., Hoffman, J. P.:Global monitoring and forecasting of biomass-burning smoke: Description and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) program, *J of Sel. Topics in Appl. Earth Obs. and Rem. Sens, 2,* 144-162, 2009.

Reid, J. S., Holben, B., and Lin, L.: Aerosol and Climate Studies in the Maritime Continent: First assessments from the 7 Southeast Asian Studies (7SEAS) effort. Atmos. Res. 122, 401-402. Editorial., 2013a.

Reid, J. S., Hyer, E. J., Johnson, R. S., Holben, B. N., Yokelson, R. J., Zhang, J., Campbell, J. R., Christopher, S. A., Di Girolamo, L., Giglio, L., Holz, R. E., Kearney, C., Miettinen, J., Reid, E. A., Turk, F. J., Wang, J., and Xian, P.: Observing and Understanding the Southeast Asian Aerosol System by Remote Sensing: An Initial Review and Analysis for the Seven Southeast Asian Studies (7SEAS) Program. Atmos. Res. 122, 403-468. doi: 10.1016/j.atmosres.2012.06.005, 2013b.

Reynolds, R. W. and Smith, T. M., Improved global sea surface temperature analyses, J. Climate, 7, 929–948, 1994.

Rienecker, M. M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R. D., Gelaro, R., Stajner, I., and Nielsen, E.: The GEOS-5 data assimilation system – Documentation of versions 5.0.1, 5.1.0, and 5.2.0. NASA Tech. Memo. 2007-104606, vol. 27, edited by: Suarez, M. J., 2008.

Sampson, C. R., Franklin, J. L., Knaff, J. A., & DeMaria, M.: Experiments with a Simple Tropical Cyclone Intensity Consensus. Weather & Forecasting, 23(2), 304-312, 2008.

Sanders, F.: Skill in forecasting daily temperature and precipitation: Some experimental results. Bulletin of the American Meteorological Society, 54(11), 1171-1178, 1973.

Sansom, P. G., Stephenson, D. B., Ferro, C. A., Zappa, G., and Shaffrey, L.: Simple uncertainty frameworks fo rslecting weighting schemes and interpreting multi-model ensemble climate change experiments, J. of Climate, 26, 4017-4037, 2013.

Schmid, B., Michalsky, J., Halthore, R., Beauharnois, M., Harrison, L., Livingston, J., Russell, P., Holben, B. N., Eck, T. F., and Smirnov, A.: Comparison of aerosol optical depth from four solar radiometers during the fall 1997 ARM intensive observation period, Geophys. Res. Lett., 26, 2725– 2728, 1999.

Schulz, M., de Leeuw, G., and Balkanski, Y.: Sea-salt aerosol source functions and emissions, in Emission of Atmospheric Trace Compounds, edited by C. Granier, P. Artaxo, and C. E. Reeves, pp. 333–354, Kluwer Acad., Norwell, Mass. 2004.

Schultz, M. G., Suttie, M. and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. Biogeosciences Discuss., 8(4), 7339-7398, 2011.

Seaton A., MacNee W., Donaldson K., Godden D.: Particulate air pollution and acute health effects. Lancet 345, 176–178. doi:10.1016/S0140-6736(95)90173-6, 1995.

Shi, Y., Zhang, J., Reid, J. S., Holben, B., Hyer, E. J., & Curtis, C. An analysis of the collection 5 MODIS over-ocean aerosol optical depth product for its implication in aerosol assimilation. Atmospheric Chemistry and Physics, 11(2), 557-565, 2011.

Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker, Cloud screening and quality control algorithms for the AERONET data base, Remote Sens. Environ., 73, 337– 349, 2000.

Smith, M. H., Park, P. M., and Consterdine, I. E.: Marine aerosol concentrations and estimated fluxes over the sea, Q. J. Roy. Meteor. Soc., 119, 809–824, 1993.

Smith, M. H., and N. M. Harrison, The sea spray generation function, J. Aerosol Sci., 29, Suppl. 1, S189–S190, 1998.

Spada, M., Jorba, O., Pérez Garcia-Pando, C., Z. Janjic, Z., and. Baldasano. J. M.: Modeling and evaluation of the global sea-salt aerosol distribution: sensitivity to size resolved and sea-surface temperature dependent emission schemes, Atmospheric Chemistry and Physics, 13, 11735-11755, doi:10.5194/acp-13-11735-2013, 2013.

Streets, D. G., Yan, F., Chin, M., Diehl, T., Mahowald, N., Schultz, M., Wild, M., Wu, Y. and Yu, C.: Anthropogenic and natural contributions to regional trends in aerosol optical depth, 1980-2006, J Geophys Res-Atmos, 114, D00D18, doi:10.1029/2008JD011624, 2009.

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of probabilistic prediction systems. Proceedings, ECMWF Workshop on Predictability, ECMWF, 1–25, 1997.

Tanaka, T. Y., Orito, K., Sekiyama, T. T., Shibata, K., Chiba, M., and Tanaka, H.: MASINGAR, a global tropospheric aerosol chemical transport model coupled with MRI/JMA98 GCM: Model description, Pap. Meteor. Geophys., 53, 119-138, 2003.

Tanaka, T. Y. and Chiba, M: Global simulation of dust aerosol with a chemical transport model, MASINGAR. J. Meteor. Soc. Japan 83A, 255–278, 2005.

Tegen, I., Harrison, S. P., Kohfeld, K., Prentice, I. C., Coe, M., and Heimann, M.: Impact of vegetation and preferential source areas on global dust aerosol: results from a model study, J. Geophys. Res., 107, 4576, doi:10.1029/2001JD000963, 2002.

Tegen, I., & Lacis, A. A.: Modeling of particle size distribution and its influence on the radiative properties of mineral dust aerosol. Journal of Geophysical Research: Atmospheres (1984–2012), 101(D14), 19237-19244, 1996.

Thampi, B.V., Rajeev, K., Parameswaran, K. and Mishra, M.K: Spatial distribution of the Southeast Asian smoke plume over the Indian Ocean and its radiative heating in the atmosphere during the major fire event of 2006. Geophysical Research Letters, L16808, doi:10.1029/2009GL039316, 2009.

Tompkins, A. M.: A revised cloud scheme to reduce the sensitivity to vertical resolution, Tech. Memo. 0599, 25 pp., Res. Dep., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K., 2005.

Wang, H., and Niu, T.: Sensitivity studies of aerosol data assimilation and direct radiative feedbacks in modeling dust aerosols, Atmos. Environ., 64, 208-218, 2013.

Wang, H., Zhang, X., Gong, S., Chen, Y., Shi, G., and Li, W.: Radiative feedback of dust aerosols on the East Asian dust storms, *J. Geophys. Res*., 115, D23214, doi:10.1029/2009JD013430, 2010.

Weaver, C., da Silva, A., Chin, M., Ginoux, P., Dubovik, O., Flittner, D., Zia, A., Remer, L., Holben, B. and Gregg, W.: Direct insertion of MODIS radiances in a global aerosol transport model, J Atmos Sci, 64(3), 808–827, 2007.

Weigel, A. P., Knutti, R., Liniger, M.A., Appenzeller, C.: Risks of model weighting in multimodel climate projections, J. of Climate, 23, 4175-4192, 2010.

Westphal, D. L., Curtis, C. A., Liu, M., Walker, A. L., Operational aerosol and dust storm forecasting, WMO/GEO Expert Meeting on an International Sand and Dust Storm Warning System, IOP Conf.Series: Earth and Environmental Science, 7, 012007, doi:10.1088/1755-1307/7/1/012007, 2009

Westphal, D. L., and Toon, O. N.: Simulations of the microphysical, radiative, and dynamic processes in a continental-scale forest-fire smoke plume, J. Geophys. Res, 96, 22379-22400, 1991.

White, B. R.: Soil transport by winds on Mars, J. Geophys. Res., 84, 4643–4651, doi:10.1029/JB084iB09p04643, 1979.

Witek, M. L., Flatau, P. J., Quinn, P. K., and Westphal, D. L.: Global sea-salt modeling: Results and validation against multicampaign shipboard measurements, J. Geophys. Res., 112, D08215, doi:10.1029/2006JD007779, 2007.

Woodward, S.: Modeling the atmospheric life cycle and radiative impact of mineral dust in the Hadley Centre climate model. J. Geophys. Res. 106: 18155–18166. doi:10.1029/2000JD900795, 2001.

Woodward, S.: Mineral dust in HadGEM2. Technical Report 87, Met Office Hadley Centre for Climate Change, Exeter, UK, available at: http://sds-was.aemet.es/forecast-products/dust-forecasts/Woodward_2011_HadGEM2.pdf (last access: 29 May 2014), 2

Xian, P., Reid, J. S., Turk, J. F., Hyer, E. J., and Westphal, D. L.: Impact of modeled versus satellite measured tropical precipitation on regional smoke optical thickness in an aerosol transport model, *Geophys. Res. Lett.*, 36, L16805, doi:10.1029/2009GL038823, 2009.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T.Y., Shindo, E., Tsujino, H., Deushi, M., Mizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T., and Kitoh, A.: A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 —Model Description and Basic Performance —. J. Meteorol. Soc. Jpn., 90A, 23-64, doi:10.2151/jmsj.2012-A02, 2012.

Zhang, J. and Reid, J. S.: MODIS aerosol product analysis for data assimilation: assessment of over-ocean level 2 aerosol optical thickness retrievals, J. Geophys. Res., 111, D22207, doi:10.1029/2005JD006898, 2006.

Zhang, J., Reid, J. S., Westphal, D. L., Baker, N. L., and Hyer, E. J.: A system for operational aerosol optical depth data assimilation over global oceans, J. Geophys. Res., 113, D10208, doi:10.1029/2007JD009065, 2008

Zhang, J. and Reid, J. S.: An analysis of clear sky and contextual biases using an operational over ocean MODIS aerosol product, Geophys. Res. Lett., 36, L15824, doi:10.1029/2009GL038723, 2009.

Appendix A: Member Model Descriptions

Provided in the Appendix are short narratives of individual model descriptions provided by their developers. We begin with the four core multi species model developers (ECMWF MACC, FNMOC/NRL NAAPS, JMA- MASINGAR, NASA GMAO GEOS-5) followed by the three dust only models (NMMB/BSC-CTM, NOAA NCEP NGAC , and UKMO Unified Model).

A.1 Multi species models

A.1.1 ECMWF MACC

Starting in 2008, ECMWF has been providing daily aerosol forecasts including dust as part of the EU-funded projects GEMS, MACC and MACC-II. All data are publicly available online at http://www.copernicus-atmosphere.eu. In the near future, these forecasts will be available operationally as part of the EU Copernicus Atmospheric Services which provides predictions of global atmospheric composition and regional European air pollution. The current model resolution is ~80km, and it is envisaged that this will be increased to ~40km in the operational phase expected to start in 2015.

A detailed description of the ECMWF forecast and analysis model including aerosol processes is given in Morcrette et al. (2009) and Benedetti et al. (2009). The initial package of ECMWF physical parameterizations dedicated to aerosol processes mainly follows the aerosol treatment in the LOA/LMD-Z model (Boucher et al. 2002; Reddy et al. 2005). Five types of tropospheric aerosols are considered: sea salt, dust, organic and black carbon, and sulfate aerosols. Prognostic aerosols of natural origin, such as mineral dust and sea salt are described using three size bins. For dust bin limits are at 0.03, 0.55, 0.9, and 20 microns while for sea-salt bin limits are at 0.03, 0.5, 5 and 20 microns. Emissions of dust depend on the 10-m wind, soil moisture, the UV-visible component of the surface albedo and the fraction of land covered by vegetation when the surface is snow-free. A correction to the 10-m wind to account for gustiness is also included (Morcrette et al. 2008). Sea-salt emissions are diagnosed using a source function based on work by Guelle et al. (2001) and Schulz et al. (2004). In this formulation, wet sea-salt mass fluxes at 80% relative humidity are integrated for the three size bins, merging work by Monahan et al. (1986) and Smith and Harrison (1998) between 2 and 4 mm. Sources for the other aerosol types which are linked to emissions from domestic, industrial, power generation, transport and shipping activities, are taken from the SPEW (Speciated Particulate Emission Wizard), and EDGAR (Emission Database for Global Atmospheric Research) annual- or monthly-mean climatologies. More details on the sources of these aerosols are given in Dentener et al. (2006). Emissions of OM, BC and SO2 linked to fire emissions are obtained using the Global Fire Assimilation System (GFAS) based on MODIS satellite observations of fire radiative power, as described in Kaiser et al. (2012).

Several types of removal processes are considered: dry deposition including the turbulent transfer to the surface, gravitational settling, and wet deposition including rainout by large-scale and convective precipitation and washout of aerosol particles in and below the clouds. The wet and dry deposition schemes are standard, whereas the sedimentation of aerosols follows closely what was introduced by Tompkins (2005) for the sedimentation of ice particles. Hygroscopic effects are also considered for organic matter and black carbon aerosols.

MODIS AOT data at 550 nm are routinely assimilated in a 4D-Var framework which has been extended to include aerosol total mixing ratio as extra control variable (Benedetti et al. 2009). A variational bias correction for MODIS AOD is implemented based on the operational set-up for assimilated radiances following the developments by Dee and Uppala (2009). The bias model for the MODIS data consists of a global constant that is adjusted variationally in the minimization based on the first-guess departures. Although simple, this bias correction works well in the sense that the MACC analysis matches well the debiased MODIS observations. The observation error covariance matrix is assumed to be diagonal, to simplify the problem. The errors have been chosen based on the departure statistics and are prescribed as fixed values over land and ocean for the assimilated observations. The aerosol background error covariance matrix used for aerosol analysis was derived using the Parrish and Derber method (also known as NMC method; Parrish and Derber, 1992) as detailed by Benedetti and Fisher (2007). This method was long used for the definition of the background error statistics for the meteorological variables and is based on the assumption that the forecast differences between the 48-h and the 24-h forecasts are a good statistical proxy to estimate the model background errors.

A.1.2 FNMOC/NRL NAAPS

The Navy Aerosol Analysis and Prediction System (NAAPS) is the US Navy's offline chemical transport model running with dust, smoke, sulfate, and sea salt at 1x1 degrees/ 27 levels based on the Danish Eulerian Hemispheric Model (Christensen, 1997; Witek et al., 2007). NAAPS has generated quasi-operational forecasts since 1999 at the Naval Research Laboratory (NRL; www.nrlmry.navy.mil/aerosol), but in 2008 became fully operational global at Fleet Numerical Meteorology and Oceanography Center (FNMOC; http://www.usno.navy.mil/FNMOC/). At the writing of this paper, NAAPS is in the process of a major revision change, including an increase in resolution to 1/3 degree, new meteorology through NAVGEM, updated data assimilation, and improved fire emissions. For this study, an intermediate version of the model is used for consistency. The 1x1 degree model is driven by the 0.5 degree Navy Operational Global

Analysis and Prediction System (NOGAPS; Hogan and Rosmond, 1991). A 1st order approximation of secondary organic aerosol (SOA) processes is adopted in which production of SOA from its precursors is assumed to be instant and included with the original sulfate specie to form a combined pollution specie. Anthropogenic emissions come from the ECMWF MACC inventory (Lamarque et al., 2010). Smoke from biomass burning is derived from near-real time satellite based thermal anomaly data used to construct smoke source functions (Reid et al., 2009; Hyer et al., 2013). In the NAAPS version for the ensemble, dust is emitted dynamically and is a function of modeled friction velocity to the fourth power, surface wetness and surface erodability, which in this model run is adopted from Ginoux (2001) with regional tuning. Sea salt modeling in ensemble version of NAAPS is the same as Witek et al. (2007) and sea salt emission is driven dynamically by sea surface wind. Analysis fields assimilate quality controlled collection 5 MODIS AOT (Zhang and Reid, 2006; Zhang et al, 2008, Hyer et al, 2011) with minor corrections from Multi-angle Imaging SpectroRadiometer (MISR). Aerosol wet deposition is constrained at analysis time with satellite retrieved precipitation within the tropics (Xian et al., 2009).

A.1.3 JMA MASINGAR

The Japan Meteorological Agency (JMA) has been providing the "Aeolian Dust Information" to the general public via its website (http://www.jma.go.jp/en/kosa/) since January 2004. The operational numerical dust forecast in JMA is based on the Model of Aerosol Species in the Global Atmosphere (MASINGAR) (Tanaka et al., 2003), which is coupled with the MRI/JMA98 AGCM. The model includes five aerosol species, namely sulfate (and its precursors), black carbon, organic aerosols, sea salt, and mineral dust. The model resolutions were set to a T106

Gaussian horizontal grid (approximately 1.125° × 1.125°) and 30 vertical layers from the surface to a height of 0.4 hPa. Dust and sea salt particles are logarithmically divided into 10 discrete size-bins from 0.1 to 10 μm in radius. The operational version of MASINGAR calculates the emission flux of dust as a function of the third-power of 10-m wind velocity (Gillette, 1978), soil moisture, soil type, snow cover and vegetation cover. Anthropogenic emissions curing this study period are taken from the Representative Concentration Pathways Database (RCP), but have since transitioned to using MACCity. The ICAP-MME version of MASINGAR used updated dust aerosol module based on the saltation-bombardment dust emission theory, which is described in Tanaka and Chiba (2005). The transport of aerosol is calculated with 3D semi-Lagrangian advection, subgrid vertical diffusion, moist convective transport and gravitational settling. Removal processes of aerosol include rainout, washout and dry deposition. JMA is planning to update the operational dust forecast model to be based on the latest global climate model MRI-CGCM3 (Yukimoto et al., 2012).

A.1.4 NASA GEOS-5

The Goddard Earth Observing System model, version 5 (GEOS-5), is the latest version of the NASA Global Modeling and Assimilation Office (GMAO) Earth system model (Rienecker et al., 2008). GEOS-5 serves NASA (1) as a state-of-the-art modeling tool to study climate variability and change, (2) as a provider of research quality reanalyses for use by NASA instrument teams and the scientific community at large, and (3) as a source of near real-time forecasts of aerosol and atmospheric constituents in support of NASA aircraft campaigns (e.g., SEAC4RS, ARCTAS, HS3, DISCOVER-AQ). GEOS-5 includes components for atmospheric circulation and composition (including atmospheric data assimilation), ocean circulation and biogeochemistry, and land surface processes. Components and individual parameterizations within components are coupled under the Earth System Modeling Framework (ESMF, Hill et al. 2004). GEOS-5 has a mature atmospheric data assimilation system that builds upon the Grid-point Statistical Interpolation (GSI) algorithm jointly developed with NCEP (Rienecker *et al*. 2008) and is currently evolving into a hybrid ensemble-variational assimilation system. The version of GEOS-5 documented here is run in near real-time at a 0.25 x 0.3125 degree latitude x longitude horizontal spatial resolution on 72 hybrid sigma levels from the surface to approximately 85 km. In addition to traditional meteorological parameters (winds, temperatures, etc., Rienecker et al. 2008), GEOS-5 includes modules to represent aerosols (Colarco et al. 2010) and tropospheric/stratospheric chemical constituents (Pawson et al. 2008), and their respective radiative feedback. Aerosols are handled through a version of the Goddard Chemistry, Aerosol, Radiation, and Transport model (GOCART, Chin et al. 2002) run online and radiatively coupled in GEOS-5. GOCART treats the sources, sinks, and chemistry of dust, sulfate, sea salt, and black and organic carbon aerosols. Aerosol species are assumed to be external mixtures. Aerosol and precursor emissions are based on a number of sources. Biofuel emissions of black and organic carbon are based on Park et al. (2003) with emissions from shipping based on EDGAR. Other anthropogenic sources follow from Streets et al. (2009). For SO2 we have anthropogenic emissions from EDGAR except for aircraft emissions, which are based on the NASA AEAP program. Natural sources of organic carbon are derived from the GEIA terpene inventory (assuming 10% conversion to secondary organic aerosol). DMS emissions (converted to SO2 and then to sulfate) are based on Kettle et al (1999). Dust and sea salt emissions are as in Colarco et al. (2010). Total mass of sulfate and hydrophobic and hydrophilic modes of carbonaceous aerosols are tracked, while for dust and sea salt the particle size distribution is explicitly resolved across five non-interacting size bins for each. Both dust and sea-salt have wind-speed dependent emission functions, while sulfate and carbonaceous species have emissions principally from fossil fuel combustion, biomass burning, and biofuel consumption, with additional biogenic sources of organic carbon. Sulfate has additional chemical production from oxidation of SO_2 and dimethylsulfide (DMS), as well as a database of volcanic SO_2 emissions and injection heights. For all aerosol species, optical properties are primarily from the commonly used Optical Properties of Aerosols and Clouds data set (OPAC, Hess et al., 1998). Except for dust, optical properties are derived under the assumption of spherical particles. Our dust optical properties dataset incorporates non-spherical dust properties based on Meng et al. (2010). GEOS-5 is driven by biomass burning emissions from the Quick Fire Emission Dataset (QFED, Darmenov and da Silva 2013.) In near-real time, GEOS-5 includes assimilation of AOT observations from the MODIS sensors on both Terra and Aqua satellites. Based on the work of Zhang and Reid (2006) and Lary (2010), we originally developed a back-propagation neural network to correct observational biases related to cloud contamination, surface parameterization, and aerosol microphysics. This empirical algorithm has been adapted to retrieve AOT directly from cloud-cleared MODIS reflectances. On-line quality control is performed with the adaptive buddy check of Dee et al. (2001), with observation and background errors estimated using the maximum likelihood approach of Dee and da Silva (1999). Following a multi-channel AOT analysis, three-dimensional analysis increments are produced exploring the Lagrangian characteristics of the problem, generating local displacement ensembles intended to represent misplacements of the aerosol plumes.

A.2 Dust Only Models

A.2.1 NMMB/BSC-CTM

The NMMB/BSC-CTM (Pérez et al., 2011; Jorba et al., 2012; Spada et al., 2013) is an online chemical weather prediction system for meso- to global-scale applications, developed at the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS) in collaboration with NOAA/NCEP, NASA Goddard Institute for Space Studies, the International Research Institute for Climate and Society (IRI) and the University of California Irvine. BSC-CNS maintains global and regional dust and sea-salt aerosol forecasts based on NMMB/BSC-CTM. The BSC-Dust module is fully embedded into the Non-hydrostatic Multiscale Model NMMB developed at NCEP (Janjic et al., 2011; Janjic and Gall, 2012). It includes a physically based dust emission scheme, which explicitly takes account of saltation and sandblasting processes (White, 1979; Marticorena and Bergametti, 1995; Marticorena et al., 1997) and assumes a viscous sublayer between the smooth desert surface and the lowest model layer (Janjic, 1994; Nickovic et al., 2001). For the source function, the model uses the topographic preferential source approach after Ginoux et al. (2001) and the NESDIS vegetation fraction climatology (Ignatov and Gutman, 1998). It includes an 8-bins size distribution within the 0.1–10 microns radius range according to Tegen and Lacis (1996) and radiative interactions (Mlawer et al., 1997). The NMMB/BSC-Dust model has been evaluated at regional and global scales (Pérez et al., 2011; Haustein et al., 2012). Complementing the dust atmospheric aerosol, a sea-salt module (Spada et al., 2013) is implemented through 8 bins in the dry radius interval (0.1 − 15 microns) to describe mass concentrations and optical depth. A sub-bin lognormal approach is assumed to calculate the optical properties of the particles. Several open-ocean emission schemes are implemented, accounting for bubble-bursting and spume production (Gong, 2003; Monahan et al., 1986; Smith et al., 1993; Martensson et al., 2003; Jaeglé et al., 2011). The water uptake is taken into account by using prescribed growth factors for different relative humidity values following Chin et al. (2002). The parameterizations of the aerosol processes affected by the water-uptake (i.e. sedimentation, dry deposition, wet deposition, etc.) have been extended to wet particles from those implemented in the dust module. These developments are steps forward towards a unified multiscale chemical-weather prediction system at BSC-CNS. This sea salt component is not in the ICAP-MME but may be included at a later date.

A.2.2 NOAA NCEP NGAC

Since September 2012 NOAA NCEP has been providing 5-day global dust forecasts at 1x1 deg/64 levels once per day (at 00 UTC cycle) from the NEMS GFS Aerosol Component (NGAC) system. It includes a 5-bins size distribution with effective radius at 1, 1.8, 3, 6, and 10 microns. The NGAC is an on-line global atmospheric aerosol model developed at NCEP in collaboration with NASA GMAO (Lu et al., 2010, 2013). The forecast model is the NCEP's Global Forecast System (GFS) within the NOAA Environmental Modeling System (NEMS) infrastructure (Black et al., 2009). The aerosol component is NASA's GOCART within GMAO's GEOS-5 earth system model (Colarco et al., 2010). While NGAC has the capability to forecast dust, sea salt, sulfate, and carbonaceous aerosols, the initial NGAC operational production in 2012 only generates global dust forecasts. NCEP is planning to upgrade the operational NGAC in 2015 to include the full suite of aerosols using real-time fire emissions from satellites observations.

A.2.3 UKMO Unified Model

The dust forecasts from the UK Met Office are produced by the global NWP configuration of the Met Office Unified Model (MetUM). The dust scheme is essentially that of Woodward (2001) with modifications as described in Woodward (2011) and Collins et al. (2011). The dust emission scheme is based on Marticorena and Bergametti (1995) and represents an initial horizontal/saltation flux in a number of size bins with subsequent vertical flux of bare soil particles from the surface into the atmosphere. The global NWP model uses only 2 bins (0.1-2 microns and 2-10 microns) from the original 9 bins. The magnitude of the emission is a cubic function of the exceedance of the friction velocity over bare soil with respect to a threshold value, where this friction velocity is determined from the model wind field and boundary layer structure and the threshold friction velocity is increased by the presence of soil moisture according to Fecan (1999). The conversion from the horizontal flux to the vertical flux is first limited using the clay fraction in the soil texture dataset, according to Gillette (1978), and then partitioned into the new bins by prescribing the emitted size distribution. Once the dust is lifted into the atmosphere it is transported as a set of tracers by the model 3D wind field. Johnson et al. (2011) gives in-depth description and evaluation of the Met Office dust forecasts, in a local area model over North Africa. Dust is assimilated in a 4D-Var framework following Benedetti et al. (2009), using aerosol observations from MODIS (Collection 5.1) on-board NASA's Aqua platform. MODIS observations (best quality, dust-filtered) are assimilated only over the land based on MODIS Dark Target (Kaufman et al., 1997a,b; Levy et al., 2007; 2009) and Deep Blue (Hsu et al., 2004; 2006) retrievals.

Table 1. List of + 1 day biases from study core AERONET sites. Included are the 550 nm Total AOTs. These are followed by list of model biases for the four core ICAP members listed sequentially low to high for each site. The ICAP-MME ensemble mean bias is underscored.

Site	Location	550 nm Winter Total AOT	550 nm Total AOT Dec-May Model Biases (ensemble underscored)	550 nm Summer Total AOT	550 nm Total AOT Jun- Nov Model Biases (ensemble underscored)
Alta Floresta	1. Brazil: 9 S; 56 W	0.12	-0.06, **-0.01**,-0.01, +0.02,+0.02	0.24	-0.04, -0.03, **-0.02**, -0.01, 0.00
Baengnyeong	2. Yellow Sea: 37 N; 124 E	0.40	-0.01,+0.01,+**0.05**,+0.08,+0.12	0.36	+0.03,+0.04,+**0.06**,+0.06,+0.10
Banizoumbou	3. Sahel: 13 N; 2 E	0.65	-0.17, -0.09, **-0.09**, -0.06, -0.04	0.45	-0.24, -0.16, **-0.15**, -0.12, -0.10
Beijing	4. China: 39 N; 116 E	0.61	-0.09, -0.07, **-0.07,** -0.01, , +0.01	0.72	-0.19, -0.14, **-0.10**, -0.07, 0.01
Cape Verde	5. Subtrop. Atlantic 16N; 22 W	0.34	+0.01,+0.06,+**0.07**,+0.07,+0.13	0.36	-0.05, 0.00, **0.00**, 0.00, +0.05
CART Site	6. Great Plains: 36 N; 97 W	0.12	-0.02,+0.02,+**0.02**,+0.02,+0.07	0.15	-0.04,+0.01,+**0.01**,+0.02,+0.04
Chapais	7. Quebec: 49 N; 74 W	0.17	-0.08, -0.04, **-0.04**, -0.02, -0.02	0.13	-0.02, 0.00,+**0.01**,+0.03,+0.04
Chiang Mai	8. Thailand: 18 N; 98 E	0.59	-0.41, **-0.26**, -0.25, -0.22, -0.17	0.21	-0.07, -0.06, **0.00**, 0.00, +0.12
Crozet Island	9. Southern Oceans: 46 S; 51 E	0.12	-0.02, -0.02,+**0.01**,+0.01,+0.05	0.10	+0.02,+0.03,+**0.05**,+0.06,+0.09
Gandhi College	10. Rural India: 25 N; 84 E	0.63	-0.28,-0.19, **-0.17**, -0.11, -0.09	0.66	-0.27, -0.19, **-0.18**, -0.17, -0.07
GSFC	11. E. CONUS: 38 N; 76 W	0.12	-0.01,+**0.03**,+0.03,+0.04,+0.04	0.17	0.00,+0.02,+**0.02**,+0.02,+0.06
Ilorin	12. Sahel: 8 N; 4 E	0.89	-0.38, -0.26, **-0.26**,-0.20, -0.20	0.30	-0.11, **-0.04**, -0.03, -0.02,+0.01
Kanpur	13. Urban India: 26 N; 80 E	0.60	-0.28, -0.19, **-0.16**, -0.11, -0.07	0.67	-0.32, -0.19, **-0.16**, -0.16, 0.00
Minsk	14. Western Asia: 53 N; 27 E	0.18	-0.01, **0.0**, +0.01,+0.01,+0.02	0.16	0.00, 0.01,+**0.01**,+0.02,+0.03
Moldova	15. Eastern Europe: 47 N; 28 E	0.19	-0.01, 0.00, +**0.01**, +0.02,+0.03	0.18	0.00, 0.00,+**0.01**,+0.01,+0.03
Monterey	16. W. CONUS: 36 N; 121 W	0.09	0.0,+0.01, +**0.02**, +0.03, +0.03	0.09	-0.02, -0.01, **-0.01**,0.00, 0.00
Palma de Malloraca	17. Mediterranean: 39 W; 2 E	0.19	-0.07, -0.05, **-0.05**, -0.04, -0.03	0.19	-0.01,-0.01,+**0.01**,+0.03,+0.04
Ragged Point	18. Subtr. Atlantic: 13 N; 59 W	0.15	-0.02, -0.01,+**0.01**,+0.03,+0.04	0.16	-0.02, -0.01, **-0.01**,0.00,+0.02
Rio Branco	19. South America:9 S 67 W	0.10	-0.04, 0.00, **0.00**,+0.02,+0.04	0.21	-0.07, -0.03, **-0.03**, -0.03, 0.00
Singapore	20. Maritime Cont.: 1 N; 103 E	0.33	-0.16, -0.11, **-0.11**, -0.10, -0.05	0.43	-0.21, -0.17, **-0.14**, -0.13, -0.04
Solar Village	21. Southwest Asia:24 N; 46 E	0.47	-0.15, -0.14, -0.03, **-0.02**,+0.23	0.39	-0.10,+0.01,+**0.01**,+0.02,+0.13

Table 2. Same as Table 1 but for coarse mode AOT, and for those sites in which the coarse mode is dominated by dust. This includes the ICAP core and NGAC models.

Site	550 nm Dec-May Dust AOT	550 nm Coarse AOT Dec-May Model Dust Biases (ensemble underscored)	550 nm Jun-Nov Dust AOT	550 nm Coarse AOT Jun-Nov Model Dust Biases (ensemble underscored)
Baengnyeong	0.10	-0.09, -0.02, **-0.01**,+0.01,+0.01,+0.02	0.09	-0.08, -0.03, **-0.02,** 0.00, +0.01,+0.01
Banizoumbou	0.43	-0.13, -0.08, -0.06, **-0.04**, -0.03,+0.09	0.36	-0.25, -0.19, -0.13, -0.13, **-0.12**,+0.02
Beijing	0.16	-0.12, -0.04, **-0.03**, -0.01,+0.01,+0.02	0.14	-0.12, -0.06, **-0.03**, -0.03, 0.00, +0.01
Cape Verde	0.28	-0.03,+0.01,+0.01,+**0.03**,+0.03,+0.11	0.31	-0.11, -0.03, -0.03, **0.00**,+0.01, +0.14
Gandhi College	0.20	-0.09, 0.09, -0.09, -0.09, **-0.06**, -0.01	0.18	-0.11, -0.09, -0.08, -0.07, **-0.06**, -0.01
Ilorin	0.38	-0.13, -0.11, -0.10, **-0.07**, -0.06,+0.01	0.15	-0.10, -0.07, -0.06, -0.04, **-0.03**,+0.06
Kanpur	0.24	-0.10, -0.10, -0.09, **-0.07**,-0.09, -0.01	0.26	-0.15, -0.12, -0.11, -0.11, **-0.09**, -0.02
Palma de Malloraca	0.11	-0.07, -0.06, -0.05, -0.05, **-0.04**, -0.01	0.12	-0.04, -0.03, -0.02, **-0.01**,-0.01,+0.03
Ragged Point	0.13	-0.07, -0.04, **-0.04**,-0.04, -0.03, -0.03	0.14	-0.05, -0.05, -0.03, **-0.03**, -0.02,+0.01
Solar Village	0.30	-0.05, -0.04, **-0.04**, -0.01, 0.01, 0.14	0.27	-0.02, -0.02,+0.02, +0.02,+**0.02**,+0.10

Table 3. List of +1 day forecast 550 nm total AOT RMSE from study core AERONET sites. These are followed by list of model biases for the four core ICAP members listed sequentially low to high for each site. The ICAP-MME ensemble mean bias is underscored.

Site	550 nm Dec-May Total AOT	550 nm Total AOT Dec-May Model RMSE (ensemble underscored)	550 nm Jun-Nov Total AOT	550 nm Total AOT Jun- Nov Model RMSE (ensemble underscored)
Alta Floresta	0.12	<u>**0.05**</u>, 0.06, 0.07, 0.08,0.09	0.24	0.08, <u>**0.12**</u>, 0.14, 0.15, 0.21
Baengnyeong	0.40	<u>**0.18**</u>, 0.21, 0.22, 0.22, 0.32	0.36	<u>**0.18**</u>, 0.19, 0.19, 0.25, 0.29
Banizoumbou	0.65	<u>**0.29**</u>, 0.29, 0.29, 0.36, 0.4	0.45	0.20, 0.20, <u>**0.22**</u>, 0.24, 0.30
Beijing	0.61	0.38, <u>**0.40**</u>, 0.41, 0.46, 0.55	0.72	0.44, <u>**0.46**</u>, 0.49, 0.53, 0.62
Cape Verde	0.34	0.11, <u>**0.13**</u>, 0.13, 0.14, 0.30	0.36	<u>**0.10**</u>, 0.12, 0.12, 0.13, 0.15
CART Site	0.12	<u>**0.04**</u>, 0.04, 0.04, 0.04, 0.10	0.15	0.05, <u>**0.07**</u>, 0.08, 0.09, 0.17
Chapais	0.17	<u>**0.17**</u>, 0.17, 0.18, 0.18, 0.27	0.13	<u>**0.05**</u>, 0.05, 0.07, 0.06, 0.09
Chiang Mai	0.59	0.37, <u>**0.43**</u>, 0.43, 0.47, 0.64	0.21	<u>**0.10**</u>, 0.14, 0.14, 0.16, 0.26
Crozet Island	0.12	0.06, <u>**0.07**</u>, 0.07, 0.080.11	0.10	0.05, 0.06, <u>**0.07**</u>, 0.10, 0.11
Gandhi College	0.63	0.17, 0.20, <u>**0.23**</u>, 0.25, 0.36	0.66	0.27,<u>**0.31**</u>, 0.32, 0.33, 0.48
GSFC	0.12	<u>**0.05**</u>, 0.05, 0.05, 0.07, 0.08	0.17	0.05, <u>**0.07**</u>, 0.08, 0.10, 0.12
Ilorin	0.89	0.36, 0.38, <u>**0.40**</u>, 0.42, 0.55	0.30	<u>**0.11**</u>, 0.12, 0.13, 0.14, 0.16
Kanpur	0.60	0.18, 0.24,<u>**0.26**</u>, 0.29, 0.29	0.67	<u>**0.30**</u>, 0.30, 0.31, 0.34, 0.48
Minsk	0.18	<u>**0.04**</u>, 0.04, 0.05, 0.05, 0.10	0.16	<u>**0.07**</u>, 0.07, 0.08, 0.09, 0.10
Moldova	0.19	0.08, <u>**0.09**</u>, 0.09, 0.11, 0.18	0.18	0.05, <u>**0.08**</u>, 0.09, 0.11, 0.18
Monterey	0.09	<u>**0.04**</u>, 0.04, 0.05, 0.05, 0.06	0.09	<u>**0.03**</u>, 0.03, 0.04, 0.04, 0.05
Palma de Malloraca	0.19	<u>**0.06**</u>, 0.06, 0.06, 0.08, 0.12	0.19	0.05, <u>**0.06**</u>, 0.06, 0.08, 0.10
Ragged Point	0.15	<u>**0.05**</u>, 0.05, 0.05, 0.06, 0.11	0.16	<u>**0.05**</u>, 0.05, 0.06, 0.06, 0.09
Rio Branco	0.10	<u>**0.03**</u>, 0.04, 0.04, 0.05, 0.07	0.21	0.08, <u>**0.09**</u>, 0.10, 0.11, 0.15
Singapore	0.33	<u>**0.18**</u>, 0.19, 0.20, 0.22, 0.23	0.43	0.19, <u>**0.23**</u>, 0.26, 0.27, 0.32
Solar Village	0.47	<u>**0.13**</u>, 0.19, 0.20, 0.21, 0.29	0.39	<u>**0.09**</u>, 0.11, 0.14, 0.18, 0.19
1st day rank (21 pos)		13, 5, 3, 0, 0		10, 9, 2, 0, 0
4th day rank (21 pos)		9, 11, 1, 0, 0		10, 9, 1, 1, 0

Table 4. Same as Table 3 but for AERONET coarse mode AOT and model dust RMSE, and for those sites in which the coarse mode is dominated by dust. This includes the ICAP core and NGAC models.

Site	550 nm Dec-May Dust AOT	550 nm Coarse AOT Dec-May Model Dust RMSE (ensemble underscored)	550 nm Jun-Nov Dust AOT	550 nm Coarse AOT Jun-Nov Model Dust RMSE (ensemble underscored)
Baengnyeong	0.10	**0.06**, 0.07, 0.07, 0.10, 0.10, 0.16	0.09	**0.05**, 0.06, 0.06, 0.09, 0.09, 0.11
Banizoumbou	0.43	**0.20**, 0.22, 0.24, 0.25, 0.29, 0.30	0.36	**0.18**, 0.18, 0.18, 0.21, 0.24, 0.28
Beijing	0.16	**0.12**, 0.13, 0.16, 0.17, 0.17, 0.32	0.14	**0.13**, 0.13, 0.14, 0.14, 0.20, 0.36
Cape Verde	0.28	**0.11**, 0.11, 0.11, 0.13, 0.17, 0.18	0.31	**0.10**, 0.12, 0.13, 0.13, 0.19, 0.19
Gandhi College	0.20	**0.09**, 0.09. 0.11, 0.12, 0.12, 0.15	0.18	**0.08**, 0.08, 0.09, 0.10, 0.10, 0.12
Ilorin	0.38	**0.20**, 0.20, 0.24, 0.25, 0.30, 0.32	0.15	**0.08**, 0.09, 0.11, 0.12, 0.13, 0.13
Kanpur	0.24	0.09, **0.10**, 0.13, 0.13, 0.14, 0.16	0.26	**0.11**, 0.12, 0.14, 0.14, 0.15, 0.17
Palma de Malloraca	0.11	**0.06**, 0.07, 0.07, 0.07, 0.08, 0.08	0.12	**0.04**, 0.04, 0.05, 0.06, 0.07, 0.08
Ragged Point	0.13	**0.06**, 0.07, 0.07, 0.08, 0.08, 0.09	0.14	**0.06**, 0.06, 0.06, 0.07, 0.07, 0.09
Solar Village	0.30	**0.09**, 0.11, 0.12, 0.14, 0.22, 0.22	0.27	**0.09**, 0.10, 0.10, 0.13, 0.16, 0.23
24 hr rank (10 pos)		9, 1, 0, 0, 0, 0		10, 0, 0, 0, 0,0
96 hr rank (10 pos)		10, 0, 0, 0, 0, 0		10, 0, 0, 0, 0, 0

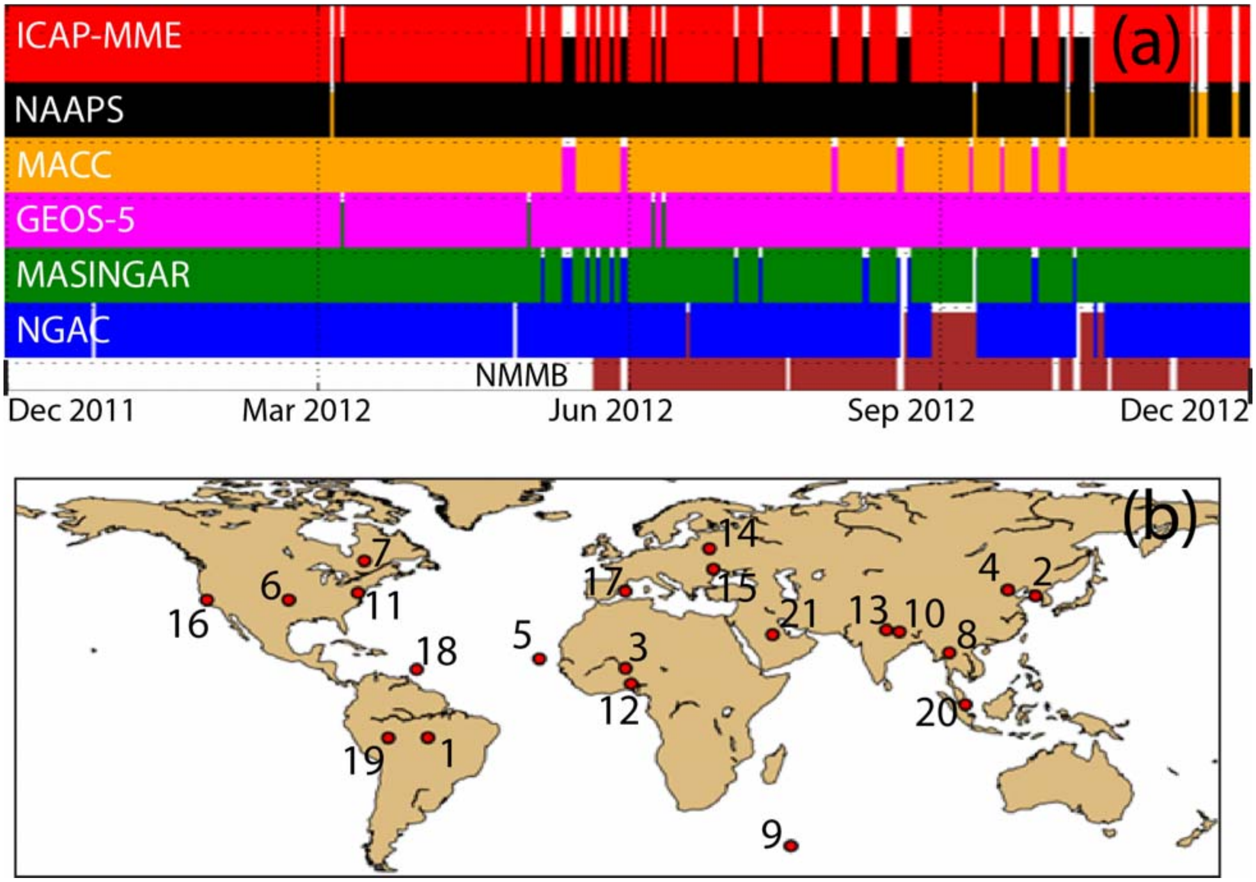

Figure 1. (a) Timeline of available data within this paper's study period. (b) Location of AERONET sites used for verification. Labels are listed in Table 1.

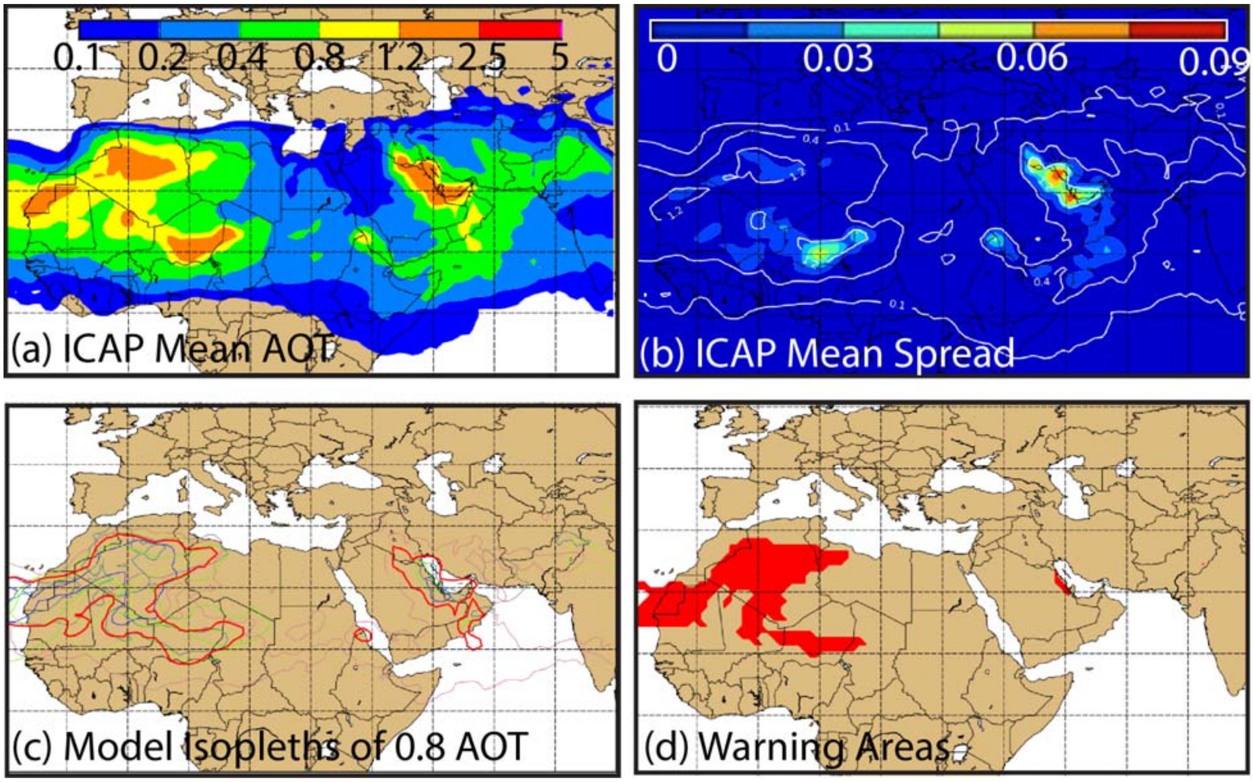

Figure 2. Examples of ICAP-MME products expected to be released to the public at publication of this paper for an example 72 hour forecast of 2012's most significant dust events plus a secondary event over the Arabian Gulf using all 6 dust members. (a) Ensemble mean 550 nm AOT; (b) "Mean/Spread" of the 6 ensemble members, with the standard deviation as color and AOT isopleths; (c) "Spaghetti plot" of AOT 0.8 isopleth; (d) Dust warning areas where more than half of the models predict AOT>0.8.

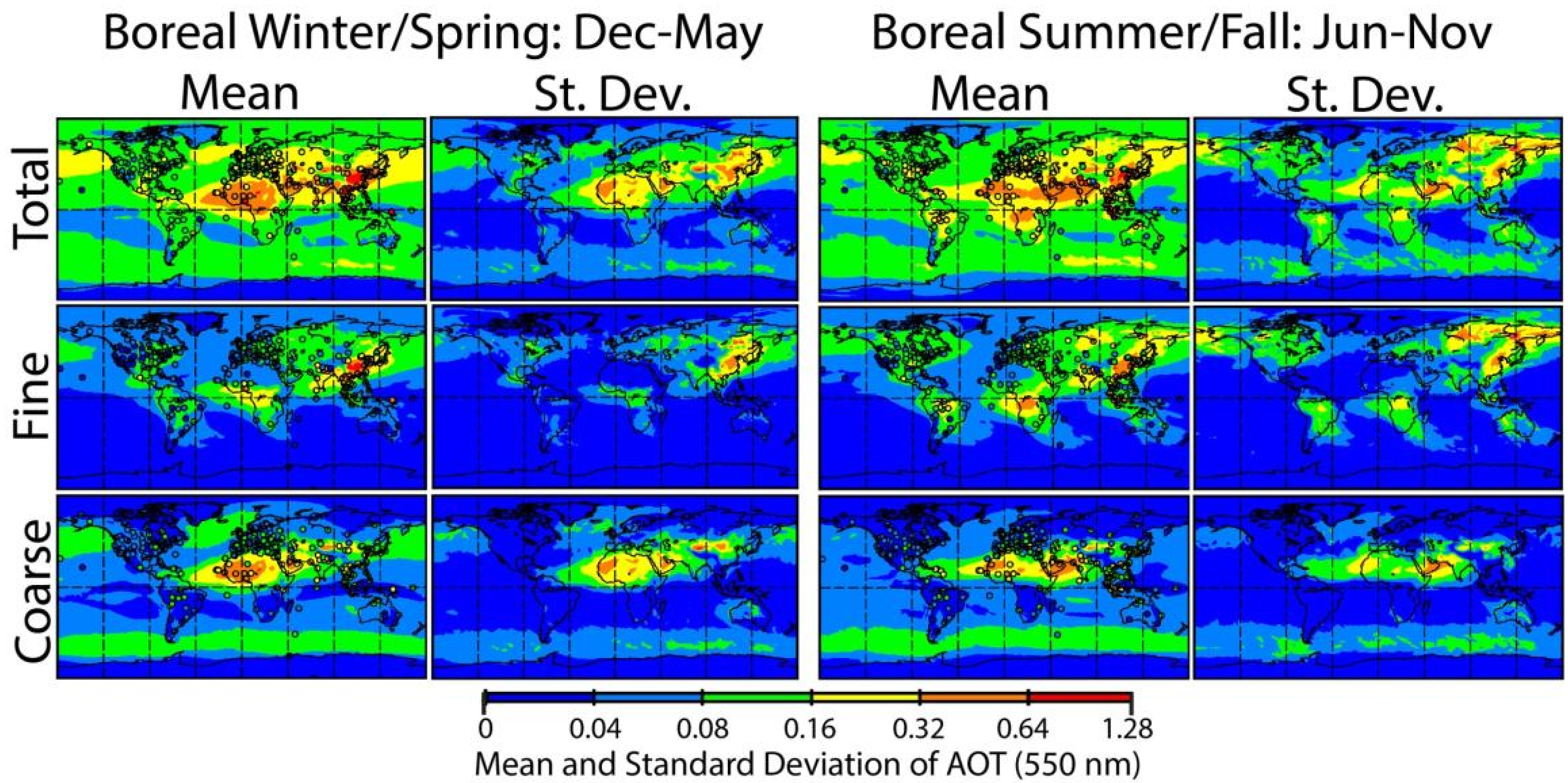

Figure 3. Mean and standard deviation of the ICAP-MME 550 nm AOT ensemble consensus for the December 2011-November 2012 time period. Included are the 4 core models of ECMWF MACC, FNMOC/NRL NAAPS, JMA MASINGAR, and NASA GMAO GEOS-5. Breakout is by boreal winter/spring (December-May) and summer/fall (June-November). Further striations are for total, fine and coarse mode optical depth. Provided in dots are the AERONET means for the same time period-although these are not pairwise with the model data.

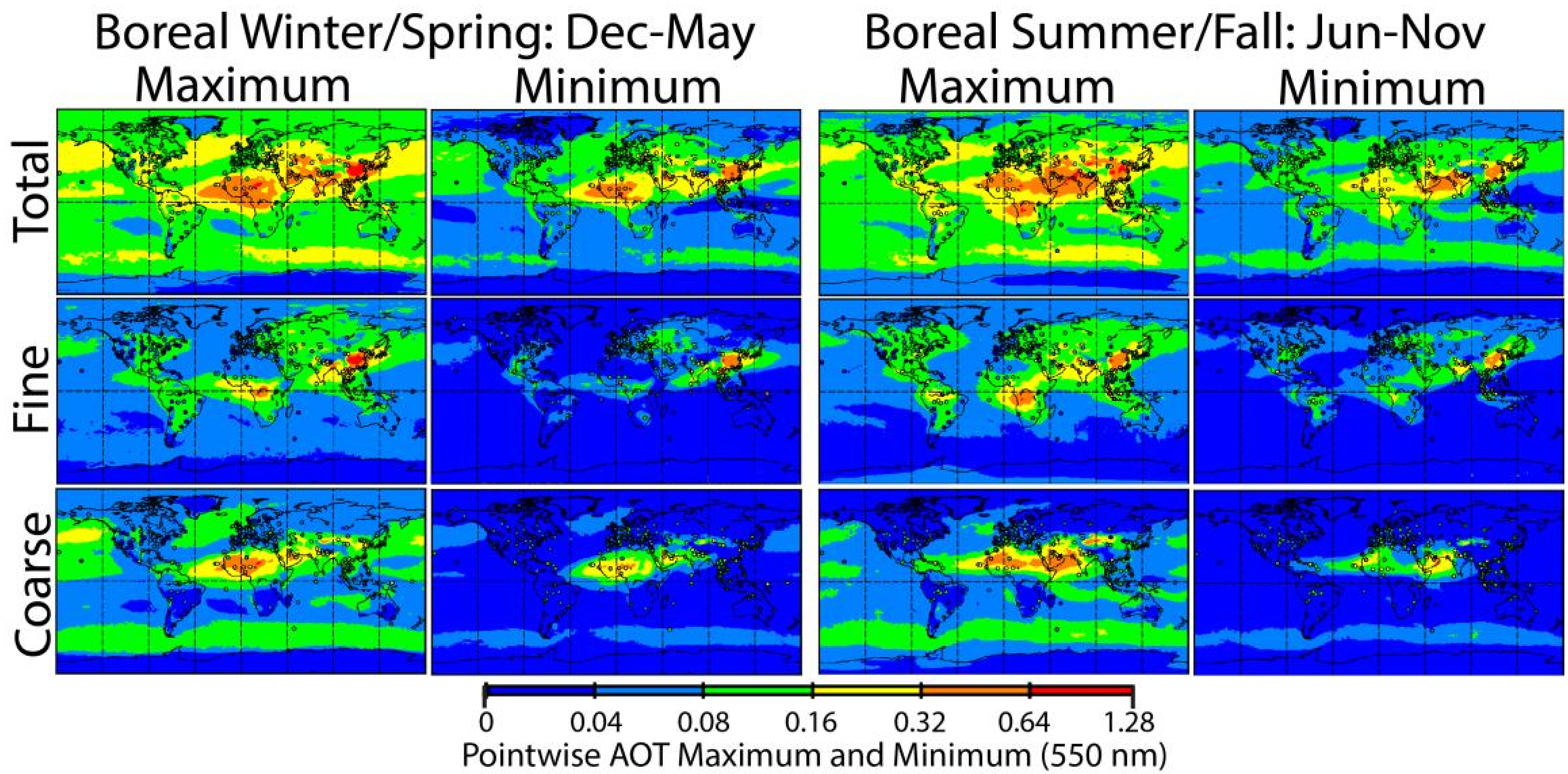

Figure 4. Same as Figure 3, but for point wise maximum and minimum 550 nm AOTs drawn from the ICAP-MME's four core member seasonally averaged AOT fields. AERONET circles represent AOT means.

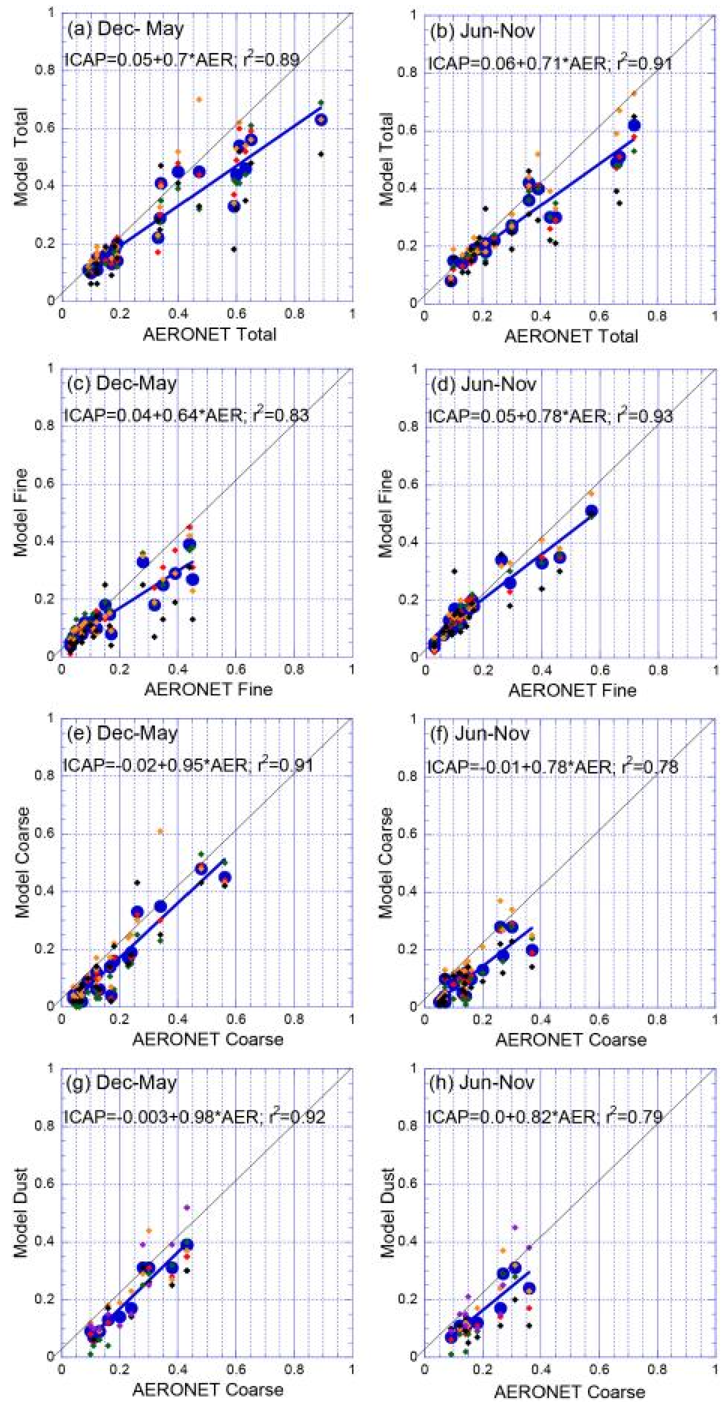

Figure 5. Bi-seasonal comparisons of model 550 nm AOT means with 21 core AERONET verification sites listed in Table 1. Large blue circles are ICAP-MME means. Other models are small colored diamonds. Data are stratified (left column) for December-May 2011, (right column) June-November. (a) & (b) Core models and ensemble mean comparisons to total AERONET derived 550 nm AOT. (c) & (d) Model versus AERONET for fine mode particles. (e) & (f) Models versus AERONET for coarse mode particles. (g) & (h) Model dust versus AERONET Coarse for dust stations listed in Table 2. NGAC is included in the dust comparison.

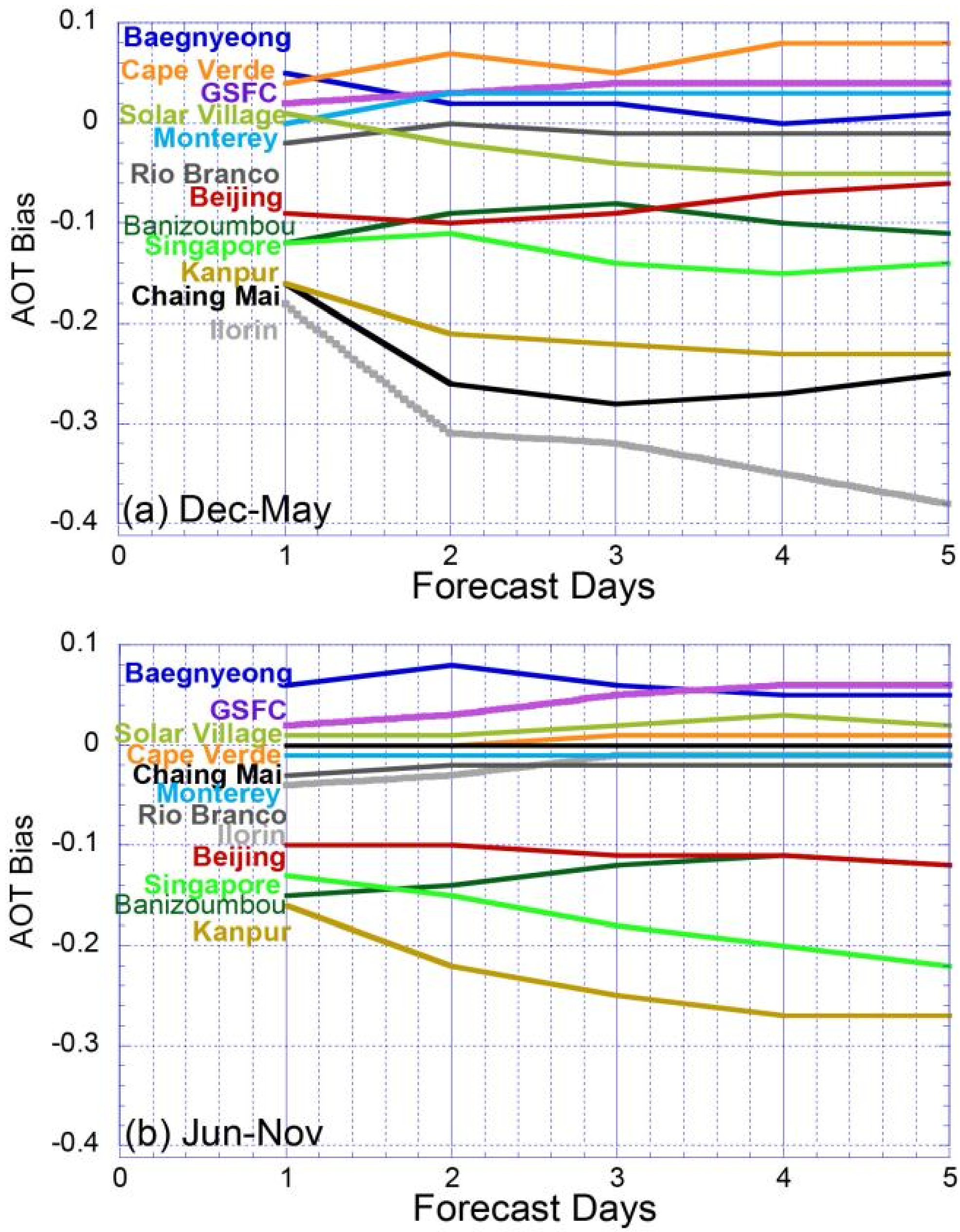

Figure 6. ICAP-MME 550 nm total AOT model bias as a function of forecast hour for key AERONET sites. (a) December-May boreal winter/spring period; (b)June-November boreal summer/fall.

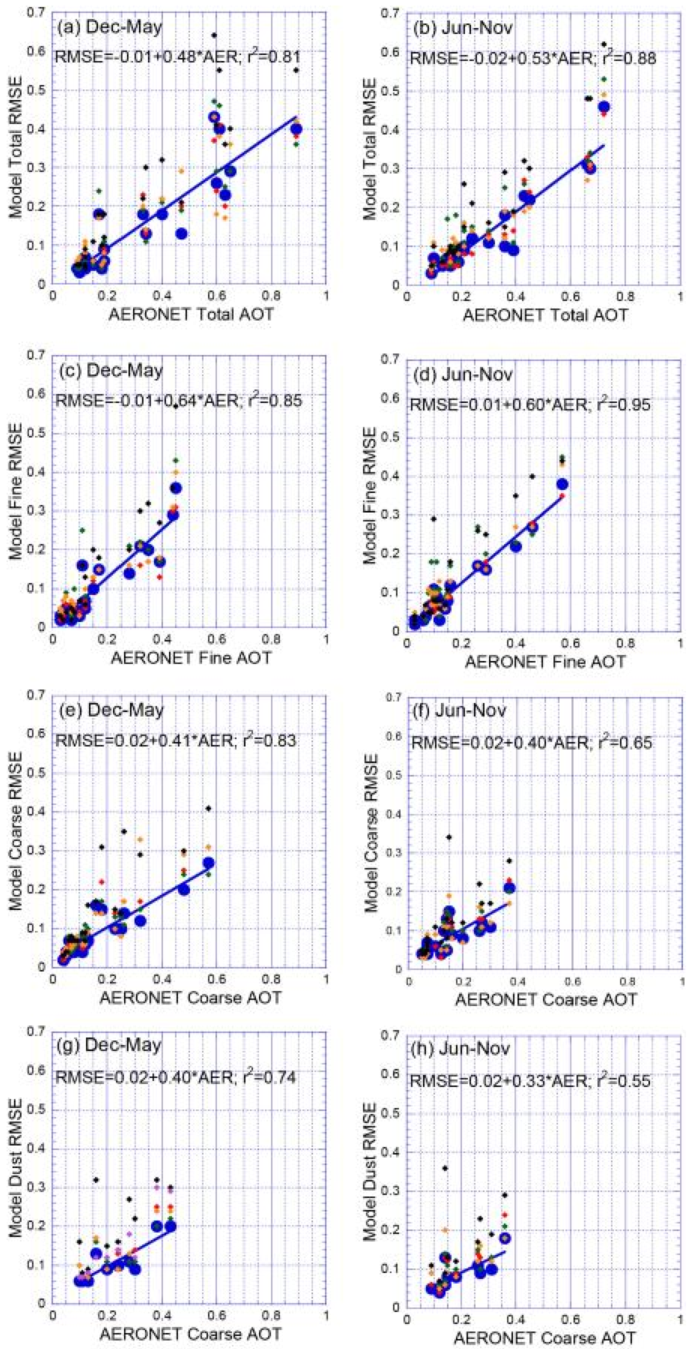

Figure 7. Bi-seasonal comparisons of +1 day model 550 nm AOT RMSE with 21 core AERONET verification sites listed in Table 1. Large blue circles are ICAP-MME means. Other models are small colored diamonds. Data is stratified (left column) for December-May 2011, (right column) June-November. (a) & (b) Core models and ensemble mean comparisons to total AERONET derived 550 nm AOT. (c) & (d) Model versus AERONET for fine mode particles. (e) & (f) Models versus AERONET for coarse mode particles. (g) & (h) Model dust versus AERONET Coarse for dust stations listed in Table 2. NGAC is included in the dust comparison.

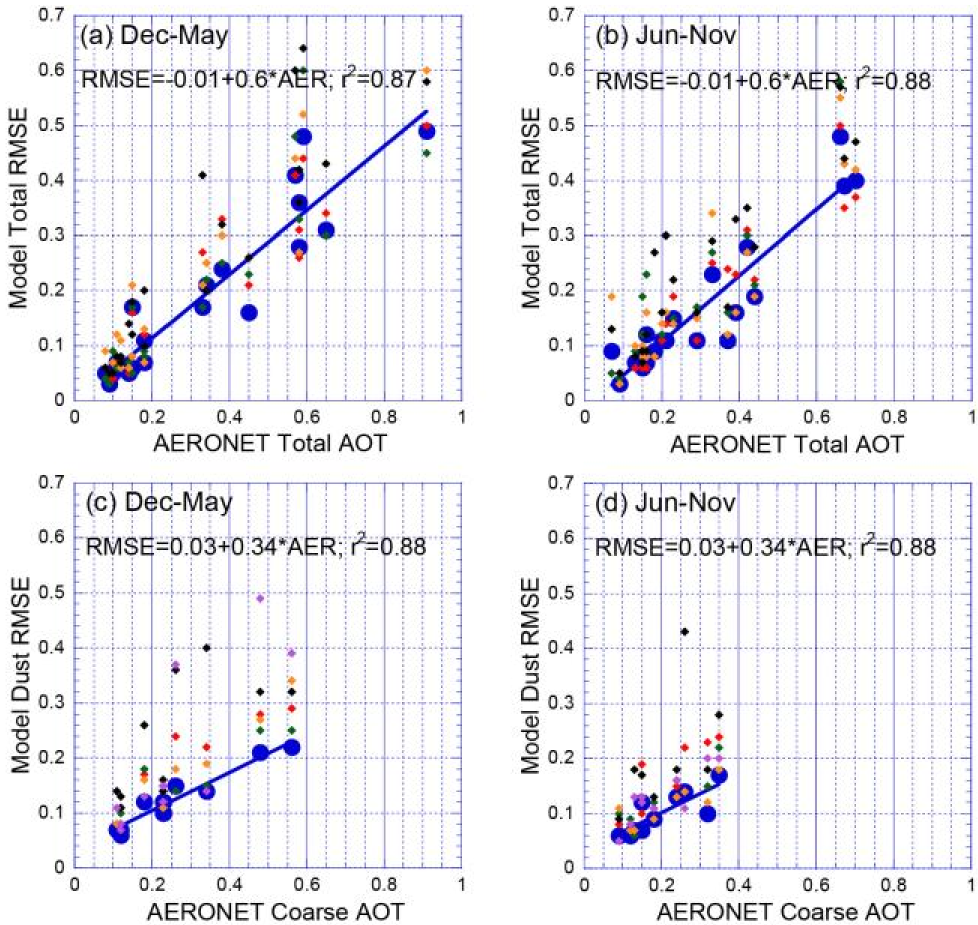

Figure 8. Same as Figure 7 for Total and Dust AOT, with +4 day RMSEs.

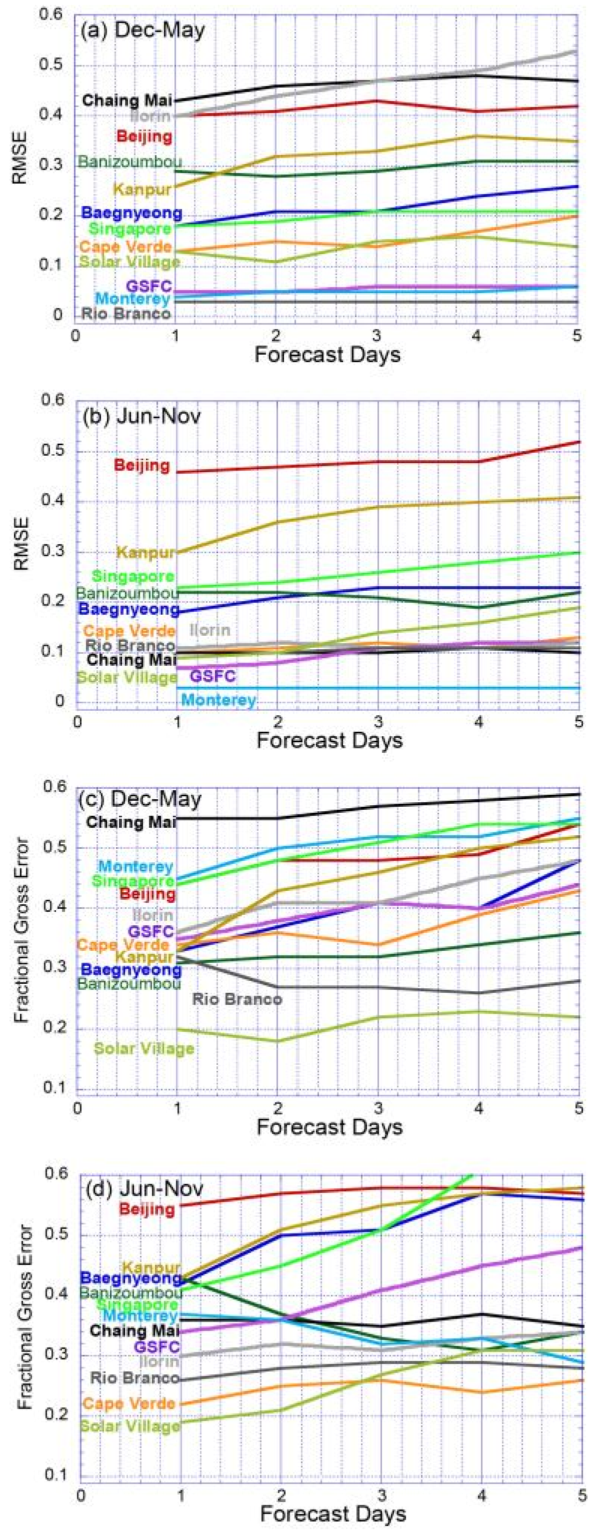

Figure 9. ICAP-MME consensus Root Mean Square Error (RMSE) and Fractional Gross Error as a function of forecast day for selected AERONET sites shown in Figure 1b.

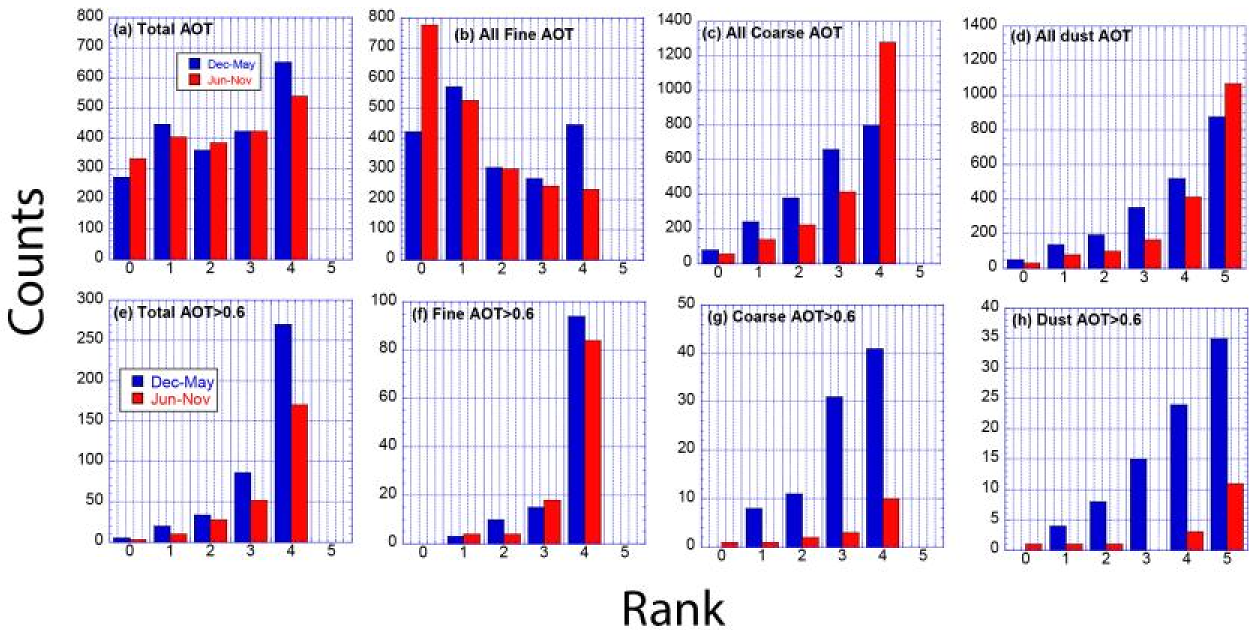

Figure 10. (a)-(d), bi seasonal rank histograms of ICAP-MME members and the ensemble mean total, fine, coarse and dust AOT for all data, respectively. (e)-(h) same as previous for cases where AERONET AOT>0.6.

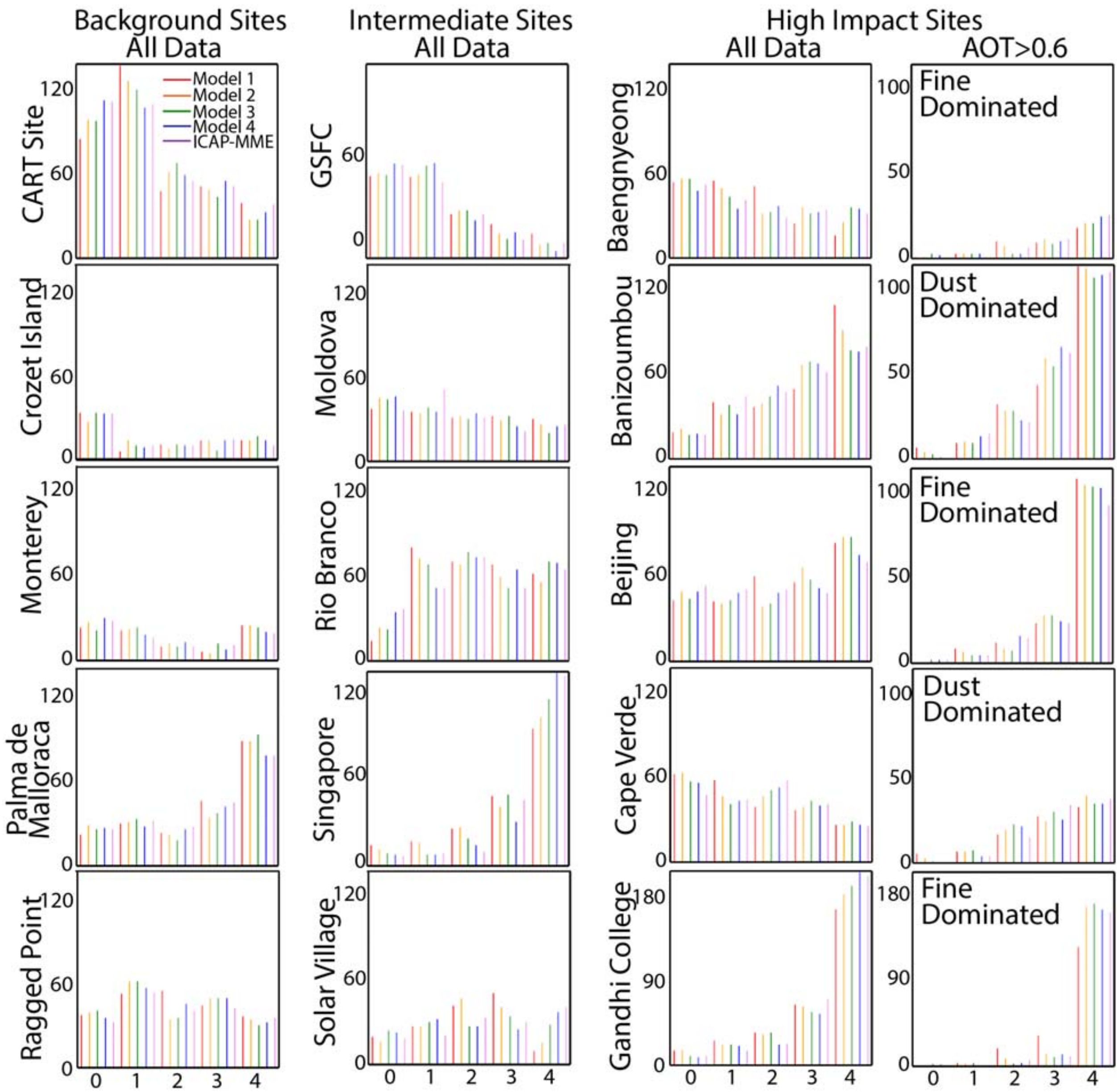

Figure 11. Rank histograms for selected sites the entire 1 year study period. Included are sites considered as background or long range receptor sites (Column 1) ; sites with intermediate loadings (Column 2), and sites with high aerosol impact, segregated into all data (Column 3) and those cases with AERONET AOT>0.6 (Column 4). The dominant aerosol type leading to AOTs>0.6 are listed for sites in Column 4.

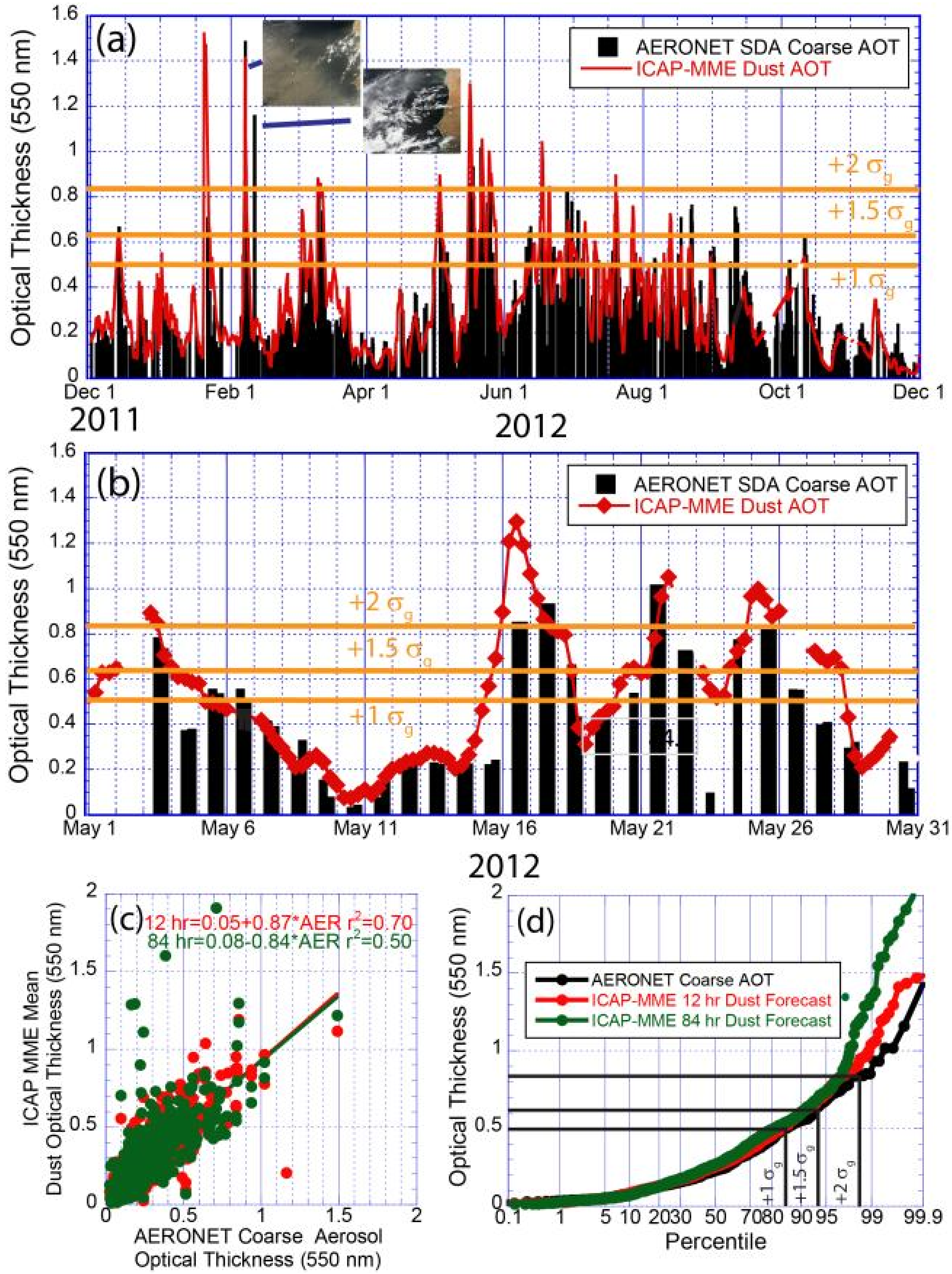

Figure 12. An example of the derivation of threat scores for the CapeV site. (a) One year Time series of first day forecasted ICAP-MME mean AOT with corresponding AERONET coarse mode AOT. Insets are MODIS RGB images for an actual and artifact dust event. (b) enlargement of (a) for the month of May, 2012. (c) Scatterplot of forecasted AOT against AERONET; (d) probability distribution of AERONET and forecasted AOT.

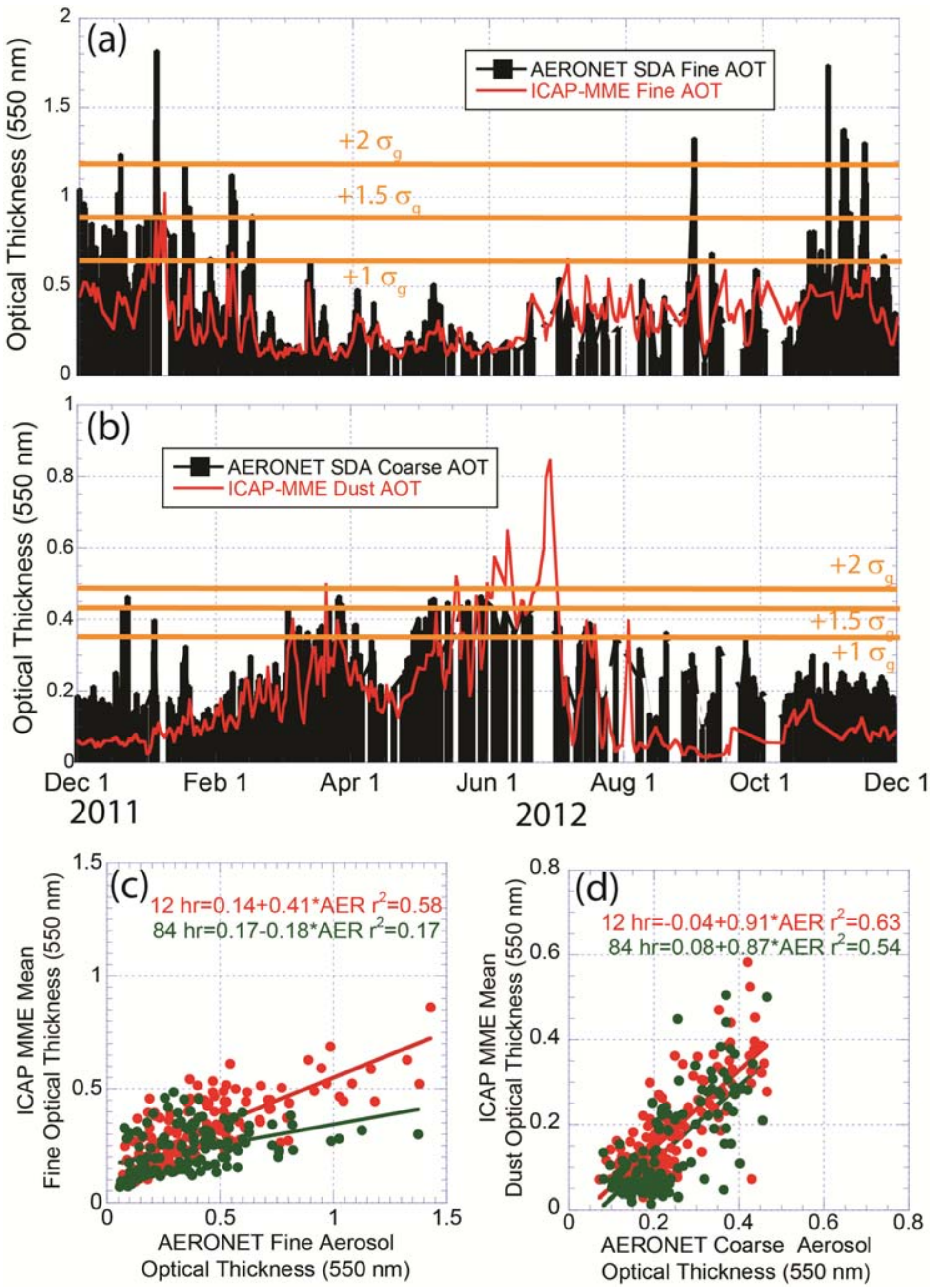

Figure 13. ICAP MME-AERONET comparisons for the Kanpur India site. Included are the (a) fine mode and (b) dust components. Marked are the 1, 1.5 and 2 geometric standard deviation lines. Also shown are scatter plots against 12 and 84 hr forecasts for (c) fine mode and (d) dust, respectively.