Development of the Ensemble Navy Aerosol Analysis Prediction System (ENAAPS) and its application of the Data Assimilation Research Testbed (DART) in Support of Aerosol Forecasting

Juli I. Rubin[1] , Jeffrey S. Reid[2], James A. Hansen[2], Jeffrey L. Anderson[3], Nancy Collins[3], Timothy J. Hoar[3], Timothy Hogan[2], Peng Lynch[4], Justin McLay[2], Carolyn A. Reynolds[2], Walter R. Sessions[4], Douglas L. Westphal[2], and Jianglong Zhang[5].

[1]{National Research Council, Washington D.C.; Naval Research Laboratory, Monterey, CA}

[2]{Marine Meteorology Division, Naval Research Laboratory, Monterey, CA}

[3]{National Center for Atmospheric Research, Boulder, CO}

[4]{CSC, Inc, Monterey, CA}

[5]{University of North Dakota, Grand Forks, ND}

Correspondence to: Juli I. Rubin (juli.rubin.ctr@nrlmry.navy.mil)

Abstract:

An ensemble-based forecast and data assimilation system has been developed for use in Navy aerosol forecasting. The system makes use of an ensemble of the Navy Aerosol Analysis Prediction System (ENAAPS) at 1x1 degree, combined with an Ensemble Adjustment Kalman Filter from NCAR's Data Assimilation Research Testbed (DART). The base ENAAPS-DART system discussed in this work utilizes the Navy Operational Global Analysis Prediction System (NOGAPS) meteorological ensemble to drive offline NAAPS simulations coupled with the DART Ensemble Kalman Filter architecture to assimilate bias-corrected MODIS Aerosol Optical Thickness (AOT) retrievals. This work outlines the optimization of the 20-member ensemble system, including consideration of meteorology and source-perturbed ensemble members as well as covariance inflation. Additional tests with 80 meteorological and source members were also performed. An important finding of this work is that an adaptive covariance inflation method, which has not been previously tested for aerosol applications, was found to perform better than a temporally and spatially constant covariance inflation. Problems were identified with the constant inflation in regions with limited observational coverage. The second major finding of this work is that combined meteorology and aerosol source ensembles are superior to either in

isolation and that both are necessary to produce a robust system with sufficient spread in the ensemble members as well as realistic correlation fields for spreading observational information. The inclusion of aerosol source ensembles improves correlation fields for large aerosol source regions such as smoke and dust in Africa, by statistically separating freshly emitted from transported aerosol species. However, the source ensembles have limited efficacy during long range transport. Conversely, the meteorological ensemble generates sufficient spread at the synoptic scale to enable observational impact through the ensemble data assimilation. The optimized ensemble system was compared to the Navy's current operational aerosol forecasting system which makes use of NAVDAS-AOD (NRL Atmospheric Variational Data Assimilation System for aerosol optical depth), a 2D variational data assimilation system. Overall, the two systems had statistically insignificant differences in RMSE, bias and correlation, relative to AERONET observed AOT. However, the ensemble system is able to better capture sharp gradients in aerosol features compared to the 2DVar system, which has a tendency to smooth out aerosol events. Such skill is not easily observable in bulk metrics. Further, the ENAAPS-DART system will allow for new avenues of model development, such as more efficient lidar and surface station assimilation as well as adaptive source functions. At this early stage of development, the parity with the current variational system is encouraging.

1 Introduction

In support of monitoring aerosol impacts on air quality and climate, many of the world's major weather and climate centers have engaged in the rapid development of operational aerosol data assimilation and forecasting capabilities (Tanaka et al. 2003; Zhang et al. 2008; Benedetti et al. 2009; Colarco et al. 2010; Sekiyama et al. 2010; Pérez et al., 2011). Operational forecasting centers are also making use of aerosol predictions to correct radiances for assimilation in numerical weather prediction (NWP) systems (e.g., Merchant et al., 2006; Wang and Niu, 2013; Bogdanoff et al., 2015), further motivating the development of aerosol forecasting and assimilation systems. As aerosol forecasting capabilities are further developed, many lessons can be learned from the NWP community. For example, forecast skill can be enhanced by moving from deterministic to ensemble-based simulations (Kalnay 2003). By using the ensemble average forecast, the most uncertain aspects of the forecast tend to be minimized, generally leading to an increase in skill (Kalnay 2003). Additionally, ensemble systems provide a means for quantifying forecast uncertainty. Finally, ensemble systems provide an opportunity to apply Ensemble Kalman Filter (EnKF) data assimilation technologies which are relatively easy to implement and allow for flow-dependent corrections to the predicted state fields (Evensen, 1994; Houtekamer and Mitchell, 1998). As a result, ensemble-based forecasts are used by nearly all the major operational weather centers (Buizza et al. 2005). The successful use of ensembles in the NWP community (Houtekamer et al., 2005, Whitaker et al. 2008, Szunyogh et al. 2008, Bowler et al. 2008, Miyoshi et al. 2010) has led to increased interest in the use of both single and

multi-model ensembles for aerosol forecasting systems (Sekiyama et al.,2010; Sessions et al. 2015).

Current operational aerosol forecasts for the United States Navy are made by the Fleet Numerical Meteorological and Oceanography Center (FNMOC) and use the deterministic Navy Aerosol Analysis Prediction System (NAAPS, Christensen et al. 1997; Witek et al. 2007; Reid et al. 2009) combined with the Navy Variational Data Assimilation System for Aerosol Optical Depth (NAVDAS-AOD) (Zhang et al. 2008; 2011). NAAPS is an offline aerosol model driven by Navy global meteorological models; formerly the Navy Operational Global Analysis Prediction System- NOGAPS (Hogan and Rosmand, 1991) and currently the Navy Global Environmental Model NAVGEM (Hogan et al., 2014). As an initial exploration of aerosol forecast uncertainty and its dependencies on underlying meteorology, a 1 degree resolution, 20-member ensemble version of NAAPS (ENAAPS) driven by the NOGAPS or NAVGEM meteorology ensemble was created. Forecasts using ENAAPS were initially run off of the analysis fields from the NAVDAS-AOD data assimilation system. Encouraged by successes using aerosol EnKF data assimilation within an NWP framework (e.g., Sekiyama et al., 2010; Schutgens et al., 2010a,b ; Pagowski and Grell, 2012; Khade et al., 2013), here we investigate the use of ENAAPS for operational aerosol forecasting purposes by replacing the NAVDAS-AOD data assimilation system with the NCAR Data Assimilation Research Testbed (DART) implementation of an EnKF. This system is referred to as the ENAAPS-DART system. In this paper, we describe the implementation of DART within the ENAAPS framework and document the initial tuning and evaluation using the operational 2D VAR system as a control for 2 month and 6 month simulation periods in 2013. In Section 2, we describe the model, the numerical experiments conducted, and the evaluation method. In Section 3, we describe results for the 2 month tuning period (six week valid simulation) followed by a 6 month run for more robust comparison of the optimized system to the current NAVDAS-AOD control. In Section 4, we discuss the nature of the outcomes, and the positive and negative aspects of adopting an ensemble data assimilation system. We conclude with key points and lessons learned from the experiments conducted.

2 Model and numerical experiment

2.1 NAAPS and ENAAPS

NAAPS is a global offline aerosol mass transport model based on the Danish Eulerian Hemispheric Model (Christensen et al. 1997) that produces deterministic 6 day forecasts of a combined anthropogenic and biogenic fine, smoke, sea salt, and dust aerosol on 25 vertical levels at 1/3 degree every six hours. While operational runs are generated at FNMOC, quasi-operational offline NAAPS runs are made in parallel at NRL with the latest model updates. A one degree reanalysis version of NAAPS for retrospective studies is also frequently employed and used as a baseline (Lynch et al. 2015). NAAPS and its reanalyses have historically been

driven by operational meteorological fields from the U.S. Navy Operational Global Analysis and Prediction System (NOGAPS; Hogan et al., 1991) with a late 2013 transition to the Navy Global Environment Model (NAVGEM; Hogan et al., 2014). Because this study occurs during the transition period where many changes to NAVGEM were taking place, here we solely utilize NOGAPS data fields. A thorough description of basic NAAPS characteristics can be found in Witek et al., (2007) and Reid et al., (2009), but a brief synopsis is provided here, including a few key differences between the NAAPS implementation used in this work and the literature. Smoke emissions from biomass burning are derived from satellite-based thermal anomaly data used to construct smoke source functions via the Fire Locating and Modeling of burning Emissions-FLAMBE database (Reid et al. 2009; Hyer et al. 2013). However, for simulations conducted in this work, a version of FLAMBE that derives smoke emissions from MODIS thermal anomaly data only is used, consistent with the NAAPS decadal reanalysis (Lynch et al. 2015). Dust is emitted dynamically as a function of friction velocity, surface wetness, and surface erodibility using NAAPS standard friction velocity to the fourth power method, but with the erodibility map of Ginoux et al. 2001. The sea salt aerosol source is dynamic in nature with emissions as a function of surface wind speed as described in Witek et al. 2007. A combined anthropogenic and biogenic fine aerosol species (ABF) is represented in NAAPS which accounts for a combined sulfate, primary organic aerosol and a first order approximation of secondary organic aerosol. Anthropogenic emissions come from the ECMWF MACC inventory (Lamarque et al. 2010). The Navy's current operational aerosol forecasting system uses NAAPS coupled to a 2-dimensional variational (2dVAR) data assimilation system (NAVDAS-AOD, Zhang et al. 2008; 2014) for assimilating AOT retrievals (Zhang et al. 2005; Zhang and Reid, 2006, 2009; Hyer et al. 2011; Shi et al. 2011) to generate forecast initial conditions every 6 hours. NAAPS with the NAVDAS-AOD data assimilation has been fully operational at FNMOC since 2010. The operational system serves as a member of the International Cooperative for Aerosol Prediction (ICAP) multi-model ensemble (Sessions et al. 2015) and is the baseline for comparison in this work.

With the exception of data assimilation (Section 2.2), the architecture of ENAAPS-DART is very similar to the deterministic version of NAAPS/NAVDAS-AOD. The model physical parameterizations are the same. However, instead of deterministic NOGAPS meteorology fields, NOGAPS ensemble meteorology fields are used. The NOGAPS ensemble meteorology fields (20 member) are generated operationally at FNMOC at 0.5 degree resolution out to six days. These fields are created by perturbing initial conditions (wind, temperature, specific humidity, and surface pressure) using an ensemble transform method as discussed in McLay et al. (2010). For ENAAPS, all twenty NOGAPS meteorology ensemble members are used for driving the model simulations, truncated to 1 degree to match the deterministic NAAPS reanalysis. As discussed in Section 2.3, both meteorology and source ensembles are tested in this work.

2.2 Ensemble data assimilation and DART

A core rationale for developing ENAAPS was to experiment with ensemble data assimilation techniques which have been successfully implemented at operational centers on an experimental basis (e.g., Sekiyama et al. 2010). For aerosol applications, a number of data assimilation methodologies have been tested both regionally and globally and shown to improve model performance (Collins et al. 2001; Yu et al 2003; Generoso et al. 2007; Adhikary et al. 2008; Zhang et al. 2008; Benedetti et al. 2009; Schutgens et al. 2010a,b, Zhang et al. 2011, Schwartz et al. 2012, Rubin et al. 2014, Sekiyama et al. 2010). While the premise of these different approaches is the same (ie. combine the model prediction and observations in a way that minimizes the analysis error), the representation of the model forecast error differs. The variational approach, which is used in the current NAVDAS-AOD system, uses a static model forecast error. On the other hand, the EnKF is based on the use of an ensemble of model forecasts to define the error where each forecast is considered to be a random draw from the probability distribution of the model's state given all previously used observations. The use of ensembles to sample the error allows the error to evolve non-linearly in time, with the flow-dependent covariances between different state components determining how observations impact the ensemble estimate. This is opposed to univariate NAVDAS-AOD assimilation which uses a static horizontal correlation model with an assumed lengthscale of 200km around an observation (Zhang et al. 2008). EnKF representation of flow dependencies and the model error should, in theory, provide a more accurate adjustment of forecasts to new observations, resulting in a reduced error in the analysis state (Hamill and Whitaker, 2005). The focus in this work is to put an EnKF assimilation system into place to take advantage of ENAAPS and the ability of the EnKF to correct aerosol fields with flow-dependent covariances. The Ensemble Adjustment Kalman Filter (EAKF) algorithm (Anderson 2001), a variant of the more traditional EnKF implementation, has been set up with a six hour cycle, with analyses generated at 0000, 0600, 1200, and 1800 UTC each day.

DART has been developed since 2002 at the National Center for Atmospheric Research (NCAR) and is an open-source community facility for ensemble-based data assimilation research and development (Anderson et al. 2009). DART has been successfully applied to a host of meteorological and atmospheric composition data assimilation problems (e.g., Arellano et al. 2007, Khade et al., 2012, Raeder et al. 2012 , Hacker et al. 2013 and many more) and provides the option to interface to a number of different filter types, including EAKF, EnKF, kernel and particle filters. ENAAPS was interfaced with DART to take advantage of its EAKF algorithm and is further referred to as the ENAAPS-DART system. ENAAPS passes aerosol mass concentrations for each species as well as model-predicted AOT to DART every 6 hours for assimilation of MODIS AOT retrievals. The posterior (analysis) aerosol mass concentrations are then passed back to ENAAPS to initialize the next model prediction cycle.

2.3 Experimental design

This study was conducted in two phases: a) a two month spin up and simulation period for the July and August 2013 period to develop and optimize the DART EAKF implementation in ENAAPS; and b) a six month April through September 2013 run to compare ENAAPS to a NAAPS baseline. These experiments are described in detail below.

2.3.1 DART EAKF implementation and optimization

As ensemble data assimilation systems can be sensitive to system design, a number of short experiments for July through August, 2013 were run with ENAAPS-DART for system optimization. This time period is coincident with the peak of the African dust season, significant pollution events, and continental scale boreal fire outbreaks. The application of ensemble data assimilation to atmospheric prediction is complicated as the model datasets are large, multivariate, and multidimensional (Anderson 2007). In atmospheric applications, it is always the case that the ensemble size is too small, resulting in sampling error and an under-prediction of the model uncertainty (Anderson and Lei, 2013). The under-prediction of model uncertainty, represented as insufficient variance in the ensemble members, can lead to poor performance and, in some cases, filter divergence in which the observations no longer impact the model state (Anderson 2007). Important considerations in the system setup include ensemble size and the means for generating the ensembles. Additionally, several tuning techniques have been developed for alleviating the sampling issue for large models, including covariance inflation for increasing ensemble spread (Anderson and Anderson 1999, Anderson 2007, Anderson 2009) and localization for spatially limiting the impact of an observation (Hamill et al. 2001, Houtekamer and Mitchell, 2001).

The effectiveness of the ensemble data assimilation system is highly dependent on having sufficient spread in the ensemble members in order for the observations to impact the model forecast. The method for generating the ensemble is an important consideration for an optimal aerosol forecasting system since the ensembles represent the uncertainty in the model forecast. For aerosol, sources of uncertainty include meteorology, sources, sinks, and any physics that impact aerosol concentration or intensive properties. Aerosol source ensembles are first tested since previous studies have relied on source perturbations alone (Schutgens et al. 2010a,b). Random perturbations with a 25% uncertainty are applied to the aerosol source functions for each species (ABF, smoke, sea salt, and dust). The random perturbation factor for ensemble member *n* and aerosol species *i* ($f_{i,n}$) is drawn from a normal distribution with a mean of 1 and a standard deviation of 0.25. The aerosol source for ensemble member *n* and species *i* ($S_{i,n}(x,y)$) is described as:

$$S_{i,n}(x,y) = f_{i,n} S_i(x,y) \quad (1)$$

where $S_i(x,y)$ is the initial aerosol source flux for aerosol type *i* at a given location (x,y). It should be noted that $f_{i,n}$ is independent of location. Grid by grid perturbations were initially tested and found to have no impact on ensemble spread, therefore, this method was excluded.

Meteorology ensembles are evaluated in addition to the source draws, using the 20-member NOGAPS meteorology ensemble.

A common method in ensemble data assimilation for increasing ensemble spread about the mean is multiplicative covariance inflation (Anderson 2007, Anderson and Anderson 1999). In multiplicative inflation, the difference between the ensemble mean and each ensemble member is increased, usually in the prior, by a predetermined factor that is greater than 1 (ie. 1.1 produces a 10 percent increase in the difference). Sekiyama et al. (2010) used a multiplicative inflation factor of 1.1 for aerosol predictions, while Schutgens et al. (2010b) conducted sensitivity tests on the inflation factor and used values ranging from 1.03 to 1.30. These inflation factors are applied uniformly in both space and time. An alternative method to a uniform multiplicative inflation is adaptive covariance inflation (Anderson 2009) which creates temporally and spatially varying inflation factors. This approach is based on a Bayesian algorithm that estimates the inflation with time as part of the state update, using a normally distributed inflation factor associated with each element of the model state vector. An initial inflation factor of 1 (ie. no inflation) was set for all locations and a fixed standard deviation of 0.4 was used. In this work, a uniform multiplicative covariance inflation of 1.1 (ie. 10%) in a fashion similar to Sekiyama et al. (2010) will be tested against the Anderson (2009) adaptive inflation (AI) algorithm. It should be noted that several initial tuning experiments were conducted with the 20 member ensemble in which a range of constant inflation factors were tested, in a similar fashion to Schutgens et al. (2010b). Due to the similarities across the experiments and the prior use of the 10% inflation in ensemble aerosol assimilation, only the 10% inflation results are presented to limit the number of experiments. AI has not been previously tested for aerosol applications.

In addition to an under-prediction of model uncertainty, sampling errors due to small ensemble size can lead to spurious correlations in the background error covariance at far distances. It has been shown that limiting the distance over which an observation impacts the state variables, or localizing, is effective in reducing the effects of these noisy correlations. For aerosol applications, state-space localization using the Gaspari and Cohn function (Gaspari and Cohn 1999) and observation-space localization in the Local Ensemble Transform Kalman Filter (LETKF) using patch size have been demonstrated (Sekiyama et al. 2010, Schutgens et al. 2010a,b). A Gaspari and Cohn (1999) localization function is used in this work where the covariance magnitude decreases to zero at two times the selected cutoff length scale from the observation location. Several length scales were tested in initial tuning runs of the 20 member ensemble and a length scale of 1000km is selected for use in this work. Since the findings from the localization tuning runs are consistent with previously mentioned studies, the impact of the localization lengthscale on data assimilation performance is not a focus of this work. The number of ensemble members is held fixed for all experiments (20 members) with the exception of a single 80-member simulation tested. It should be noted that the single 80-member simulation uses the same localization lengthscale as the 20-member ensemble. Optimization of

the 80-member ensemble was not conducted due to resource limitations and will be evaluated in future work.

The initial conditions for the ENAAPS-DART experiments are generated using a 24 hour ENAAPS forecast initialized with NAAPS/NAVDAS-AOD analysis fields, using the ensemble meteorology to allow some initial ensemble spread to develop. Subsequent forecast/assimilation cycles use the DART/EAKF data assimilation with the 6 hour cycling run out for the July and August, 2013 timeframe. The performance of the 2-month experimental simulations is evaluated in several ways. The first method is through examination of the prior 6-hour forecast against MODIS AOT observations, before assimilation occurs, using diagnostics such as RMSE, bias, ensemble and total spread, number of assimilated observations, and rank histograms. Rank histograms are generated by repeatedly tallying the rank of the observation relative to values from the ensemble sorted from lowest to highest and can be used for diagnosing errors in the mean and spread of the ensemble forecast (Hamill 2001). In order to account for the effect of observation error in the rank histograms, the forecast values are randomly perturbed for each ensemble members by the observation error (Anderson 1996, Hamill, 2001, Saetra et al. 2004). The focus of this observation-space evaluation relative to MODIS AOT is on the prior since this is a stronger indicator of how the assimilation is impacting the model forecast. Benchmarks of a good ensemble system include stability in ensemble spread, an RMSE that is small and comparable to the total spread, and rank histograms that indicate an ensemble distribution that is consistent with the observations (Anderson 1996). Since aerosol composition and characteristics are variable depending on the type of aerosol sources and the location-dependent processes that impact transport, transformation, and lifetime, the diagnostics are evaluated regionally. The experimental 6-hour AOT forecasts are evaluated over 15 land regions as indicated in Figure 1 as well as six ocean regions, including the northern and southern hemisphere Pacific and Atlantic Oceans, the Indian and the Southern Ocean. Additionally, it is important to evaluate the posterior fields since these serve as forecast initial conditions. The assimilation posterior fields are examined relative to ground-based 550 nm AOT fields based on NASA AErosol RObotic NETwork (AERONET) observations (Holben et al. 1998; O'Neill et al., 2003) since these observations are not assimilated and therefore, can be used as an independent evaluation of the data assimilation analysis fields. The 550nm AERONET AOT fields used for validation are interpolated based on AOT values from the 500 and 675nm spectral channels, and are derived using a method described in Zhang and Reid, 2006. A total of five short ensemble experiments for optimization are performed. These experiments are summarized in Table 1 and account for the method used for generating the ensemble members, number of ensemble members, and different covariance inflation methods. Using diagnostics, an ENAAPS-DART system configuration is selected and compared to the operational NAAPS/NAVDAS-AOD system.

2.3.2 Baseline evaluation of EAKF versus variational data assimilation

Once a good configuration was identified, the ENAAPS-DART system was run out for a six month (April 1, 2013 to September 31, 2013) period with 6 hour cycling. The analysis fields

(i.e. data assimilation posterior) from the six month ENAAPS-DART simulation are compared to ground-based AERONET AOT observations as an independent evaluation. Analysis fields from the NAAPS/NAVDAS-AOD system are similarly compared to AERONET AOT for the same six month time period. The NAAPS/NAVDAS-AOD simulations are run with a 1 degree resolution and assimilate the same MODIS AOT observational dataset with the same observational errors (Zhang et al. 2005; Zhang and Reid, 2006, 2009; Hyer et al. 2011; Shi et al. 2011) for consistency.

The impact of the analysis fields generated from the EAKF and 2DVar system on 24 hour forecasts are also examined. Due to inconsistencies in the NOGAPS deterministic and ensemble meteorology, including differences in precipitation and wind speed, the 24 hour forecast comparisons are conducted using the same meteorology. The deterministic 24-hour forecast is initialized with the NAVDAS-AOD aerosol fields or with the ensemble mean aerosol fields from the ENAAPS-DART system (DART deterministic). The ensemble 24-hour forecast is initialized with the same NAVDAS-AOD aerosol fields for all 20 ensemble members (ENAAPS-NAV) or with the ENAAPS-DART initial conditions.

3 Results

The results from this study are presented in three sections. First, the aerosol environment for the experimental time period is examined. This is followed by a section on the EAKF optimization for ENAAPS-DART over the six week mid-July through August, time period. Finally, an evaluation of the ENAAPS-DART system relative to the current operational system, NAAPS/NAVDAS-AOD, over the April through September time period is conducted.

3.1 Synopsis of Global Aerosol Features

Average ENAAPS-DART AOT fields (Met+Source, adaptive) for the Boreal Spring (April, May) and Boreal Summer (June-September), 2013 are shown in Figure 2. Seasonally-averaged AOT for ABF, smoke, dust, and seasalt aerosol are also presented. Variability in AOT is related to major monsoonal patterns and other climate shifts associated with the spring and summer time periods. Aerosol in Asia is heavily regulated by the monsoon with the pre-monsoon dry season exhibiting a peak in aerosol and an observed boreal summertime decrease due to removal by heavy precipitation. Smoke aerosol varies by region with the observed peaks coinciding with the regional dry seasons. Some key aerosol features are discussed for the boreal spring and the boreal summer seasons.

3.1.1 Boreal spring aerosol features

AOT attributed to smoke peaks in the Yucatan Peninsula in April and May, consistent with previous studies (Reid et al. 2004, Wang et al. 2006) and extends into the northern region of South America. During peak burning, smoke transport from these Central American fires impacted Texas and the Southeast United States. Biomass burning is also present in Asia during

the pre-monsoon months of April and early May and is concentrated in Peninsular Southeast Asia, including Thailand and Cambodia.

Dust aerosol in Asia, originating from the Gobi and Taklimakan Deserts, peaks in spring due to intense frontal activity that favors lofting and contributes to the observed long-range dust transport that impacts North America in April. India is found to have a greater dust loading in the northern/northwest part of the country, originating from the Thar Desert in northwestern India. Saharan dust, although not in its peak during the April and May, dominates the AOT signal over North Africa with some outflow over the Atlantic Ocean. Under conditions of southwesterly flow, North African dust is transported into Europe and the Mediterranean region. Dust AOT in the Arabian Peninsula is slightly higher in the northern/northeast part of the peninsula. This pattern is consistent with climatology which is attributed to a dominant high pressure system that produces transport from the south/west to the north/east (Shalaby et al. 2015).

The ABF combined aerosol, including both anthropogenic and biogenic species, is prevalent throughout the Northern Hemisphere. Peaks in ABF aerosol are observed over Asia in the boreal spring with plumes extending out over the Pacific and Indian Oceans. ABF is also observed over South America and is attributed to biogenic aerosol.

3.1.1 Boreal summer aerosol features

Although fires are present throughout the summer months, the largest Boreal fires occur in August in Siberia, with smoke aerosol transport from these events reaching western North America. The fires are attributed to a persistent high-pressure weather pattern in the Russian Arctic that resulted in unusually high temperatures and long periods of stable air. Wildfires are prevalent in the Western United States in July and August, with transport from these events impacting the Eastern US. This includes the California Rim Fire, one of the largest wildfires in California's history, which occurred during August 2013 (Peterson et al. 2015). Burning events also occur in the Amazonian basin in South America. South Africa is characterized by large, persistent biomass burning events that peak in June through September with smoke transport over the South Atlantic Ocean. In the boreal summer, biomass burning events in Southeast Asia move further south and are concentrated in Borneo, Sumatra, and the Malaysian Peninsula.

Dust AOT values peak in the summer months over the North Africa, Sahara desert region, consistent with the literature (Prospero et al. 2013). The dust from Africa is transported over the Atlantic Ocean and was found to impact Central America and parts of the Southeast United States, in June, July and August. This is consistent with satellite measurements (Hsu et al. 2012) as well as aerosol records accumulated at Barbados (Prospero and Lamb, 2003), Puerto Rico (Reid et al. 2003), and Miami (Prospero, 1999), showing dust transport from the coast of Africa into the Caribbean Basin. Some transport of Saharan dust into Europe and the Mediterranean region is also observed in the summer months. Over the Arabian Peninsula, dust AOT peaks in

the summer months, particularly in the Southern region, extending over the Arabian Sea. The dust loading in India is concentrated in the south/southwest, as a result of transport from the Arabian Peninsula. In East Asia, dust AOT is limited to northern China and Mongolia.

Peak build-up of anthropogenic and biogenic fine aerosol in the Eastern US occurs during the summer months, consistent with the literature (Hsu et al. 2012). ABF buildup occurs over Europe during the summer months as well and is prevalent throughout Asia.

3.2 Ensemble data assimilation optimization

The EAKF optimization experiments focus on an evaluation of covariance inflation methods as well as an evaluation of the method for generating the ensemble (Table 1). Monthly-averaged posterior AOT fields for the EAKF optimization experiments, as well as the average difference in the posterior AOT relative to the combined meteorology and source ensemble experiment (Met+Source, adaptive), are presented in Figure 3. Some key differences are that the experiments without ensemble meteorology forcing (Source, AI; Source, Const) tend to produce a smaller AOT, especially over the Siberian fire region and dust impacted regions, including North Africa, parts of the Arabian Peninsula, India, and East Asia. At the same time, higher AOT values are generated near select source regions such as smoke in South Africa and dust in parts of Africa, Arabian Peninsula, and Asia. With the meteorology ensemble (Met, AI), higher AOT values are predicted relative to the combined ensemble, especially in regions impacted by fires.

The following sections look in detail at the performance across the ENAAPS-DART experiments. In addition to bulk statistics, representative case studies pulled from Section 3.1 are used to further understand the impact of the configurations.

3.2.1 Evaluation of covariance inflation methods

Two covariance inflation methodologies, the constant 10% multiplicative inflation and the adaptive inflation, were tested with the source only ensemble simulation. Additional 10% constant covariance inflation experiments were not conducted since the results from the source only experiments demonstrated the advantage of the AI methodology. The advantage of the adaptive inflation over the constant covariance inflation will be discussed below. The AI method itself requires some tuning to create a stable system. As previously discussed, large persistent Siberian fires generated high smoke levels in the Eurasian Boreal region in August, 2013. This region provided particular trouble for adaptive inflation, which under several configurations resulted in a blow up of the inflation factor. The inflation factor blow up indicates that the discrepancy between the prior and observational distributions increased over time, producing unrealistic AOT values and aerosol mass concentrations, eventually leading the model to crash. This type of behavior is indicative of model shortcomings related to smoke aerosol. An important tuning parameter for the adaptive inflation algorithm is the inflation factor standard deviation (Anderson 2009). The selected standard deviation affects how quickly the inflation

factor changes, especially in places like Siberia where the observations and prior ensemble are inconsistent. Adaptive inflation was tested with inflation factor standard deviations of 0.2, 0.4, and 0.6, with a selected value of 0.4. Other means were used to prevent the inflation factor from growing too large, including an applied maximum inflation factor of 1.5, preventing the inflation from growing beyond 50%. Additionally, a spatially uniform damping factor of 0.9 is applied to the inflation factors before each assimilation cycle. In this implementation of the adaptive inflation algorithm, the prior estimates of the inflation factor are assumed to be equal to the posteriors from the previous cycle, multiplied by a 0.9 damping factor. The damping factor, therefore, serves as the time variation model for the inflation. The system was found to be stable even under the extreme burning conditions in Siberia with the standard deviation of 0.4, maximum inflation of 1.5, and a damping factor of 0.9. Results are shown for this stable AI configuration.

While the 10% constant covariance inflation and AI have similar results in well-observed regions, issues occur with the constant covariance inflation where there is limited observational coverage. For the experimental time period, the observation density for assimilated MODIS AOT is presented in Figure 4(e). Since the assimilated observations are heavily bias-corrected and cloud-screened, there are spatial gaps in the observational coverage, leaving many ocean and coastal regions with little observational constraint. If the observation density is compared to the prior ensemble spread, represented as the standard deviation of the ensemble AOT normalized by the mean, at the end of the constant inflation experiment (Figure 4a), it is apparent that large spread develops where there is limited observational information, including high latitudes and spots over the Pacific Ocean. The ensemble spread at the end of the constant inflation experiment is much greater than that from AI in the other source only ensemble experiment (Figure 4b). Figure 4 provides a sense of what the ensemble spread looks like spatially throughout the globe. The change in ensemble spread is also examined over time for a number of regions (Figure 5). For most of the regions shown, the ensemble spread as a function of time is approximately the same for the source ensemble experiments with constant and adaptive inflation (Source, const and Source, adaptive). On the other hand, a difference is observed between the two experiments for the Southern Hemisphere Pacific Ocean with a steady growth in spread found for the constant inflation (Source, const) and a stable spread for the adaptive inflation configuration (Source, adaptive). The Southern Hemisphere Pacific Ocean has very little observational coverage compared to the other regions shown in Figure 5. The growth in spread in the Southern Pacific Ocean for the constant inflation experiment is a result of having continuous inflation with no observations to bring the ensemble back to reality. This demonstrated growth in ensemble spread was also found across initial tuning experiments in which a range of constant inflation factors were tested (1.03-1.5). The only difference was the timescale over which the spread developed in under-observed regions. The average inflation factor for the source only adaptive inflation experiments is shown in Figure 4f. The spatial pattern of the inflation factor follows the observation density spatial pattern with almost no inflation in the Pacific and Southern Ocean where limited observations are available. Although

spatially and temporally constant covariance inflation has been the chosen method for aerosol applications in the past, it is not recommended since aerosol observations are spatially heterogeneous. On the other hand, adaptive inflation increases ensemble spread where there is observational information available, producing stability, a desirable characteristic for an ensemble system. These findings are consistent with idealized experiments and NWP applications of ensemble systems where a temporally and spatially varying inflation is recommended over a constant inflation approach (Anderson 2009b, Li et al. 2009, Miyoshi et al. 2011).

3.2 Evaluation of ensemble generation

In addition to evaluating the impact of the covariance inflation method, the impact of the ensemble generation approach is examined with a source-only, meteorology-only, and a combined meteorology and source ensemble experiment. One impact of using the source-only ensemble is that the ensemble itself has less spread (i.e. smaller standard deviation in ensemble AOT). The spatial differences between the experiment ensemble spreads are demonstrated in Figure 4a through d, although these differences will vary with time. When comparing the adaptive inflation experiments, it is clear that including the meteorology ensemble increases the spread globally (Figure 4b through d). This is especially true over the dusty Sahara and the entire Arabian Peninsula, where the standard deviation in AOT is on the order of 1 to 15 percent (Figure 4b) compared to the 5 to 50 percent range seen with the inclusion of the meteorology ensemble (Figure 4c,d). In particular, a large increase in spread is found at dust source regions. For example, the spread increases from approximately 20 to 50 percent in the Northern Arabian Peninsula. As discussed in Section 3.1, summertime dust aerosol in the Arabian Peninsula comes from the northern region and is transported south. Similar increases are observed in Northern Africa which coincide with large dust source regions, such as the Bodele depression. Since dust emissions are dynamically driven, the inclusion of the meteorology ensemble, either by itself or with the source ensemble, greatly increases the spread in dust aerosol. Likewise, the meteorology ensemble increases spread for sea salt aerosol, which is also dynamically driven, over the Southern Ocean for example.

Whether the ensemble includes only the NOGAPS meteorology members or includes both the meteorology and source members, the ensemble spread is quite comparable, both spatially and temporally (Figure 4, Figure 5). The meteorology ensemble appears to be the main driver of ensemble spread when included with a 25% source-perturbed ensemble. The adaptive inflation compensates for differences in spread that result from including the source ensemble with the meteorology. For example, in the Northwest United States, an inflation factor in the range of 1.25 to 1.3 is applied with the combined meteorology and source ensemble. However, with the meteorology only ensemble, the inflation factor is greater, in the range of 1.3-1.4 (Figure 4g,h). Occasionally, a larger inflation factor in the meteorology only ensemble experiment results in an ensemble spread that is greater than the spread in the combined ensemble, for example in the Eastern US and the Eurasian Boreal region in August. Additional diagnostics are needed to

understand how well the ensemble spread represents actual uncertainty. It should be noted that the ensemble spread stabilizes very quickly for the AI experiments, reflected by a stable baseline ensemble spread (Figure 5). This result indicates that only a short spin-up time is needed for these simulations.

A good means for determining how well the ensemble system represents uncertainty is a comparison of the prior total spread (the square root of the sum of the ensemble variance and the observational error variance) in AOT to the prior RMSE. The RMSE is calculated against the MODIS AOT observations, prior to assimilation. The total spread and the RMSE should have a ratio close to one if the ensemble is providing a good representation of model uncertainty. If the ratio is greater than one, the total spread is greater than the error and the uncertainty is overrepresented. For a ratio less than one, the uncertainty is being underrepresented. The RMSE of the 6 hour forecast relative to MODIS AOT and the average ratio between the total spread and the RMSE for the four experiments are presented in Table 2. The results are shown on a global and regional basis, including over-land and over-ocean regions. Globally, the experimental configuration with the smallest RMSE and a ratio closest to one is the combined meteorology and source ensemble experiment with adaptive inflation (Met+Source, AI). Performance varies by region for the different ENAAPS-DART configurations. The combined meteorology and source configuration (Met+Source,AI) has the smallest RMSE with the exception of East Asia, the Southern Hemisphere Atlantic and the Southern Ocean. In these identified regions, the source only configuration has a slightly smaller RMSE (Source, AI). The use of the source-perturbed ensemble is also beneficial in the North American Boreal and South Africa, both impacted by smoke aerosol, with the meteorology ensemble alone (Met, AI) having the worst performance. Additional investigation is required to understand the impact of the source ensemble in these regions. However, Central America is the only region where the difference in performance between the ENAAPS-DART configurations is statistically significant with the inclusion of the meteorology ensemble, either by itself or with the source ensemble, producing the smallest RMSE. Overall, the combined meteorology and source ensemble configuration has the smallest RMSE in the 6 hour forecast relative to MODIS AOT.

Further probing is required to understand the impact of the source ensemble on the RMSE for several identified regions, including South Africa and the North American Boreal region. Case studies were examined and it was found that including the source ensemble is beneficial for aerosol events that are large and spatially correlated, especially for cases where the observational information is limited due to heavy cloud cover. A smoke aerosol example for the Southern Africa burning region is presented in Figure 6a. In this case, the ensemble correlation fields relative to a point near the center of a smoke plume are shown for the three AI experiments, along with the MODIS AOT observations for the event. Burning events in South Africa are persistent throughout this time period and large in scale. For the source only ensemble experiment, a clear structure in the correlation fields is observed. This structure is a result of the ensemble source perturbations for smoke in this case. By perturbing the smoke emissions using

the same factor for a given ensemble member, a correlation between freshly emitted smoke aerosol is created, resulting in the observed structure. The source perturbations essentially create infinite correlation lengthscales for freshly emitted smoke aerosol (ie. all smoke emissions are correlated), only limited by localization. A very different relationship is observed for the meteorology-only ensemble with a much more spatially limited correlation field around the point of interest. When assimilating observations into these two experiments, the observational information will spread in a much different manner around the indicated point. The correlation fields for the combined meteorology and source ensemble experiment are a combination of the two. Since the presented South Africa case study is located within a large smoke source location, the ensemble correlations are mainly governed by the source perturbations with some influence by the meteorology. The structure from the source ensemble is present with more defined edges due to the inclusion of the meteorology ensemble, producing the smallest RMSE relative to MODIS AOT.

While in general the combined meteorology and source ensemble had the best performance, occasionally the source ensemble alone outperformed the combined ensemble. This is despite the fact that one would always expect the meteorology ensemble to improve performance. An example of this is shown in Figure 6b for a North American Boreal smoke event on August 15, 2013. Smoke events in this region are not persistent, like the South African region, and vary between large, transported plumes that occur when smoke is injected above the boundary layer, sometimes spreading over thousands of miles (Figure 6b), and less intense fire events that don't make it above the boundary layer and behave independently (Figure 6c). For the large transported plume shown in Figure 6b, the ensemble correlation fields for the source only ensemble are spatially larger than the other two configurations causing the sparse observational information in the region (due to heavy cloud cover) to be spread out, producing the smallest RMSE. In this case, it appears that the meteorology ensemble might not be accurately representing the aerosol transport for this event or perhaps is overspread, producing a slightly larger (although not statistically different) RMSE. Additional tests with increased ensemble size may shed light on why the meteorology ensemble has a slightly negative impact on the performance for this event.

On the other hand, the source ensemble occasionally had a negative impact on the systems performance. An example of this is the spatially independent North American Boreal fires on August 7, 2013, shown in Figure 6c. For this event, there are a cluster of fires (A) that coincide with the point around which the correlation fields are calculated. A second cluster of fires (B) is observed to the northeast of cluster A. These fires are much smaller and are independent of cluster A, as shown in the MODIS visible image. The meteorology ensemble has the most realistic correlation fields, statistically separating the two fire clusters, while the source ensemble configurations have correlation fields that statistically link the two fire regions. For this event, the meteorology ensemble alone has the smallest RMSE. Other spatially independent events, including pollution events in the Eastern United States, showed similar performance issues with

the source-perturbed ensemble, which statistically links emissions that may be independent of each other. For these types of independent events, the source perturbations need to be done in a way that better captures the spatial correlations. While occasionally the source ensemble alone or the meteorology ensemble alone had slightly better performance, the combined meteorology and source ensemble had the overall best performance in RMSE against MODIS AOT. The caveats to this are useful case studies to determine in what ways the ENAAPS-DART system can be improved.

In addition to producing the smallest RMSE overall, the combined meteorology and source ensemble configuration (Met+Source,AI) has a total spread to RMSE ratio closest to one globally as well as regionally for South Africa, Europe, Eurasian Boreal, and East Asia (Table 2). For the remaining regions, differences in the ratio are largely due to differences in the RMSE with the total spread being approximately the same across the experiments. However, for some regions the ratio of total spread to RMSE was found to be dependent on the AOT value (Figure 7). For example, in the North American Boreal region, the ratio tends to be greater than one for AOT values less than 0.1 with the ratio decreasing to approximately 0.5 as the AOT increases. At the lower end of the AOT distribution (< 0.1), the total spread (combined ensemble spread and observational error) exceeds the RMSE; however, it is found that the observational error dominates the total spread (Figure 7). This relationship is consistent across the experimental ENAAPS-DART configurations, represented by the different colors in Figure 7. It indicates that the observational error is too large relative to the ensemble spread for small AOT values, with similar results found for other fire-impacted regions (South America, Southern Hemisphere Atlantic). This relationship is likely caused by the ensemble spread being too small for small AOT values since aerosol mass is a positive-definite quantity. For data assimilation, this translates to a reduced impact of the observation on the model state for small AOT. For the case of large AOT in the North American Boreal for example, there is not enough spread and the uncertainty is underrepresented for all ENAAPS-DART experiments (Figure 7). This may be the result of not using large enough source perturbations for smoke or the result of not accounting for uncertainties in physical processes such as deposition. However, other regions impacted by summertime burning events such as South America, the Southern Hemisphere Atlantic Ocean (Figure 7), the Eurasian Boreal region, and the Western United States also have a tendency to underrepresent uncertainty for large AOT events. Smoke emissions have very large errors; often as large as an order of magnitude uncertainty (Reid et al. 2009, 2013; Hyer et al., 2013). As a result, a larger source perturbation (greater than the 25% standard deviation currently applied) for smoke emissions is likely needed to produce a better tuned system. This reasoning is bolstered by initial AI tests that were not capped by a maximum inflation and generated inflation factors exceeding 10 in smoke-dominated regions, indicating a large discrepancy between the prior and observational distributions.

Rank histograms for select regions with representative results are shown in Figure 8 for each of the four ENAAPS-DART configurations. The Eurasian Boreal smoke region rank histogram,

consistent with the evaluation of the total spread to RMSE ratio, shows that the ensemble is not capturing low AOT values in the observed distribution, with an excess of observations occurring for low ranks. The inclusion of the meteorology ensemble helps to reduce this excess, and even more so when both the meteorology and source ensemble are included. Similar results were found for other regions impacted by smoke (North American Boreal, South Africa, South America), indicating a positive bias associated with smoke aerosol and potential bias in the smoke emissions. The large observational errors relative to the ensemble spread found for small AOT values in smoke-dominated regions (Figure 7), reducing the impact of these observations on the model state, is likely another contributing factor to the observed positive bias in smoke regions. The increase in ensemble spread with the meteorology ensemble (Figure 4,5) helps to alleviate the bias in smoke-dominated regions. In the Eastern United States, the inclusion of the meteorology ensemble introduces some positive bias with a tendency to predict AOT that is greater than the observational MODIS AOT, however, the RMSE across configurations is the same. For dust dominated regions such as North Africa, the ENAAPS ensemble well represents the observational distribution with some negative bias in the source only configurations and a slight positive bias in the meteorology configurations. Regions such as Central America and India have a large negative bias in the source-only ensemble experiments. Including the meteorology ensemble greatly reduces this bias and helps to flatten the distribution. In general, an ensemble which is created using both source perturbations and the NOGAPS meteorology ensemble does a better job representing the distribution and producing a better tuned system.

Independent evaluation of the experiments was conducted through comparison to AERONET AOT observations, which are not assimilated. In this case, the posterior ensemble mean AOT is being compared to the observations, since they are independent. Statistics, including RMSE and bias, were calculated at each AERONET site over the July through August time. Scatterplots of the RMSE relative to AERONET AOT at each site between the experiments are shown in Figure 9 and are identified by region. With respect to the source only ensemble experiments (Source, constant vs Source, adaptive), the performance is approximately the same at most sites (Figure 9a). This is a result of having MODIS observational coverage in regions where AERONET sites are located, preventing issues with the constant inflation in under-observed locations as shown in the Southern Hemisphere Pacific Ocean. The adaptive inflation experiment outperforms the constant inflation at two Eurasian Boreal sites, likely due to the adaptive inflation factor being much greater than the constant 10 percent inflation. Additionally, the AI experiment outperforms at a single Southwest Asia site, a region lacking observational coverage. If deciding between a meteorology only ensemble and a source-perturbed ensemble, in general the meteorology ensemble has a smaller RMSE, especially over the Eastern United States, Central America, India, Southwest Asia, and Dakar, a dust-impacted site in North Africa (Figure 9b). Many sites in these regions are impacted by dust transport events during the experimental time period. Evaluation of the AOT time series at the individual sites reveals that with the source ensemble only, these transported dust events are completely missed, while the event is captured in both the meteorology configuration and the combined meteorology and source configuration.

The analysis AOT time series for one of the dust impacted sites (University of Houston) in the United States is shown in Figure 10 for all three adaptive inflation ensemble configurations (source only, met only, met+source). For these long-range dust transport sites, the combined ensemble and the meteorology ensemble alone perform approximately the same with a much smaller RMSE and bias than the source only configuration (Figure 10). This result demonstrates the importance of the meteorology ensemble for long-range transport. The western US sites and several South American sites, on the other hand, perform better when the source ensemble is included with the meteorology (Figure 9c). These sites are impacted by nearby smoke events such as the Rim Fire in the Western US. An AOT timeseries for the White Salmon AERONET site (Western US), including total and smoke AOT, is presented in Figure 11. The combined meteorology and source ENAAPS-DART simulation does the best job capturing the peak smoke AOT, reflected by the difference in RMSE and bias. The effect of the source ensemble on the correlations for large smoke events, as previously shown for the South African and North American Boreal regions, is applicable in the Western United States as well. The difference in RMSE was statistically significant for the Central American, Eastern US, and India sites impacted by dust transport (between source and the two meteorology configurations) and the smoke impacted Western US sites (between meteorology only and meteorology plus source). For these sites, the combined meteorology and source ENAAPS-DART configuration had the smallest RMSE or the same as the meteorology configuration.

Based on the diagnostics from the different ENAAPS-DART configurations, the NOGAPS meteorology ensemble combined with the perturbed aerosol source function had the best overall performance. One additional test was conducted to examine the impact of increasing the ensemble size from 20 members to 80 members. An additional ENAAPS-DART 80 member ensemble simulation was run with 80 meteorology members (NAVGEM) combined with the 25% source perturbations and adaptive inflation. The same localization was used, although the optimal localization length scale should increase with increasing ensemble members. Initial results show that further reductions in RMSE can be achieved by increasing the ensemble number at most AERONET sites, including Beijing in East Asia and many Eastern US, North African, European/Mediterranean, and Boreal sites (Figure 9d). A smaller RMSE was found with the 80 member ensemble for sites impacted by spatially large aerosol events, in which the source-perturbed ensemble had previously generated the smallest RMSE relative to observations. An example is shown for Sede Boker, a Mediterranean site impacted by dust and pollution aerosol (Figure 12). Relative to the 20 member combined ensemble, the posterior AOT bias is reduced by nearly 50% and the RMSE is reduced by approximately 35%. With the 80 member ensemble, both the RMSE and bias are now less than that of the source-only ensemble configuration. It is expected that further reductions in RMSE can be achieved by tuning the localization lengthscale for the 80 member ensemble. The 80 member ensemble is not currently available for simulations over longer time periods. As a result, the 20 member combined meteorology and source ENAAPS-DART is used for evaluation against the current operational system, based on its performance against both MODIS AOT in the 6 hour forecast and

AERONET in the posterior AOT relative to the other configurations. However, the 80 member ensemble is very promising and will be explored in future work.

3.3 Baseline comparison of ENAAPS-DART to NAAPS deterministic system

3.3.1 Comparison of data assimilation analysis

To objectively determine the efficacy of the ENAAPS-DART system, the data assimilation analysis fields from the EAKF were compared to analysis fields from the variational NAVDAS-AOD system over the six month April-September 2013 timeframe. Understanding the difference in the analysis is important as the aerosol fields from the data assimilation serve as the initial condition for aerosol forecasts. Average analysis fields by month for the DART-EAKF and the 2DVar NAVDAS-AOD data assimilation as well as the difference between the two are shown in Figure 13. They both capture the same large features, such as dust from the Saharan Desert and the Arabian Peninsula, springtime burning in Central America, and Boreal fires including the August Siberian fires. However, there are clear differences between the two with the ENAAPS-DART system having a tendency to produce AOT fields on the order of 0.02 greater than the NAAPS/NAVDAS-AOD system. The difference between the two systems is reflected in the analysis increments with the tendency of NAVDAS-AOD to increase AOT on the order of 0.01 and the ENAAPS-DART having a tendency to decrease AOT on the order of 0.001. The smaller increments in ENAAPS-DART could indicate that the base system is more consistent with the assimilated observations or could be due to differences in forecast error characterization between the systems. Regions where the AOT fields from the ENAAPS-DART system are less than the deterministic system include the South African and the August Siberian biomass burning regions, parts of the US and the tropical oceans, especially in the spring. Since there are very few AOT observations for assimilation in the Southern Ocean, any differences in this region are attributed to differences in the deterministic and ensemble meteorology fields (winds, humidity) that drive the models. For example, differences in wind would impact sea salt emissions and therefore, optical thickness in the region. Likewise, differences in humidity fields would impact the optical thickness. There is also a large positive difference in AOT off the Western coast of Africa, centered on the equator in September. Speciated AOT for this location shows the presence of ABF, dust and sea salt, in addition to smoke, with a similar spatial pattern (Figure 2). This is believed to be an artifact that developed from strong covariance inflation in this region, resulting in large ensemble spread that built up over time for all aerosol species. As previously discussed, large inflation develops with AI when there is a discrepancy between the observational and ensemble distributions. If consistency between model and observations can be achieved for this smoke-dominated region by further tuning smoke emissions, the adaptive inflation will be reduced and should alleviate this problem. The need for tuning of the smoke emissions is also supported by findings in the EAKF optimization section.

The analysis fields from the two systems are compared against AERONET AOT both regionally and by site. A summary of regional statistics including RMSE, mean bias and R^2 are shown in

Table 3. It was found that the regional RMSE values relative to AERONET AOT are not statistically different between the two data assimilation systems. The slight reduction in RMSE is found for the ENAAPS-DART system relative to NAVDAS-AOD in the North American Boreal region, Central America, India, Peninsular Southeast Asia, and over the oceans. The largest difference in performance occurred in Peninsular Southeast Asia with the EAKF producing an RMSE that is 0.023 less than NAVDAS-AOD. For the remaining regions, NAVDAS-AOD had a slightly smaller or the same RMSE as the EAKF with the largest difference in RMSE (0.016) found in East Asia. While regional statistics are similar between the two data assimilation systems, there is much more diversity in performance at individual AERONET sites. The AERONET site RMSE comparison between the EAKF and the 2DVar system are shown in Figure 14. The diversity in site performance is reflected by the scatter in site RMSE by region. For example, the analysis AOT from ENAAPS-DART had a smaller regional RMSE relative to AERONET over India. A nearly 50% reduction in RMSE is seen at two AERONET sites in India with the EAKF, however, there are several sites where NAVDAS-AOD has a smaller RMSE. The opposite is seen in South America where on a regional basis analysis AOT from NAVDAS-AOD had a smaller RMSE, but there are several sites in which a smaller RMSE is associated with ENAAPS-DART, including one site with a reduction in RMSE of approximately 70%.

Site by site differences in RMSE are useful in identifying ways to further improve the ENAAPS-DART performance. A good example of this is the Eastern United States in which the NAVDAS-AOD system had a smaller regional RMSE relative to AERONET (Table 3); however, performance varies by site (Figure 14). Upon further investigation, the Eastern US sites where EAKF does better are affected by long-range dust transport, including sites in the Houston area. For example, the 2DVar system had an RMSE of 0.065 at the University of Houston AERONET site, compared to the 0.060 RMSE from the EAKF system over the six-month time frame. Likewise, several of the European sites in which the EAKF had a smaller RMSE are also impacted by long-range transport events. EAKF appears to have an edge over the 2DVar system when it comes to capturing long-range transport. This is not unexpected given that ensemble data assimilation has flow-dependent covariances. On the other hand, having a 2.5 degree univariate adjustment around an observation as is done in the variational assimilation appears to perform better for complex local sources which behave independently, as is likely the case for many Eastern US and European cities (ie. local point sources, transportation) and the North American Boreal region (independent fires). Improvement in the EAKF performance for these types of sources may be achieved by decreasing the lengthscale associated with the source perturbations. A more in depth investigation is needed to understand how to get the ensemble statistics correct for these types of independent source. Additionally, increasing the ENAAPS-DART ensemble size may change the performance relative to NAVDAS-AOD since initial tests with the 80 member ensemble indicate that an increase in ensemble size can result in better performance at most AERONET sites (Figure 9d, Figure 12).

While comparing the statistics at individual sites provides some insight into differences between the EAKF and the 2dVar, it doesn't provide any insight into what is happening spatially. From an examination of the posterior fields from the two data assimilation methodologies, it is clear that while both methods are able to capture important aerosol features, the EAKF has an ability to capture sharp gradients. On the other hand, the 2dVar, with its 2.5 degree univariate adjustment around an observation, tends to have a smoothing effect. This point is demonstrated in an example of a dust plume transported over the Atlantic Ocean, off of the Sahara Desert. The example, shown in Figure 15, shows the analysis increments for the NAVDAS-AOD 2dVar system as well as analysis increments for ENAAPS-DART, both for the source only and the combined meteorology and source ensemble. Even though the focus is now on the combined meteorology and source ensemble, the analysis increments for the source-only ensemble further demonstrate why the meteorology ensemble is so important for these transported events. The univariate adjustments from the 2dVar can be seen as circular bullets. On the other hand, the EAKF adjustments are more realistic and occur along the dust plume. The result is a dust plume which captures the sharp gradient of the dust front that is also seen in the MODIS image for this event (Figure 15). On the other hand, the 2DVar system produces a dust plume feature that is smoothed out. This dust case demonstrates a major advantage of the EAKF system over the 2dVar in its ability to spread information in a realistic manner and as a result, capture sharp gradients. It is anticipated that the ability of the EAKF to generate more realistic corrections to the state field will become more important as additional observational information is introduced into the system, such as Lidar and other spatially limited pieces of information.

3.3.2 Impact of initial condition on short-term forecast

To investigate how the impact of data assimilation persists in the forecast, four sets of 24 hour forecasts were run with the initial conditions generated from the DART-EAKF or the NAVDAS-2dVar system. Each set of initial conditions were run in a deterministic and an ensemble configuration. This is done so that the initial conditions can be tested with the same NOGAPS meteorological fields driving the model simulations. For the deterministic version of the EAKF, the forecast is initialized with the ensemble mean (DART deterministic). For the ensemble version of NAVDAS-AOD, each of the 20 ensembles is initialized with the same aerosol initial condition and run using the meteorology ensemble (ENAAPS-NAV). The forecasts were compared to AERONET AOT. The 24-hour forecast global RMSE against AERONET AOT with bootstrapped 95% confidence intervals are 0.108(0.103-0.113), 0.107(0.102-0.112), 0.100(0.097-0.104), and 0.099(0.095-0.103) for the NAVDAS-AOD deterministic, DART deterministic, ENAAPS-NAV, and ENAAPS-DART, respectively. The RMSE from the forecasts initialized with the EAKF analysis fields is less than its variational counterpart in deterministic or ensemble forecast mode, although the RMSE values are not statistically different. It should be noted that running the forecasts as ensembles produces a smaller RMSE than a deterministic configuration. This result is in line with the general knowledge about

ensembles from NWP that ensembles tend to average out the most uncertain aspects of a forecast and therefore, reduce error.

Similar to the finding with respect to the analysis fields, the comparison to site AOT from AERONET provides valuable information, but does not provide a spatial picture of the forecast behavior. The same Saharan dust transport case shown in Figure 15 is examined in Figure 16, however, now the plume is forecasted out to 24 hours. These results are initialized with either the NAVDAS-AOD or the DART-EAKF analysis fields. Results are shown for the four forecast configurations, including deterministic and ensemble forecasts. The MODIS visible image and MODIS AOT for the dust case is also included and shows a narrow band of high optical thickness at the leading edge of the dust front. All four configurations predict the dust plume, although the Northern portion of the plume is missing for all cases. The missing portion of the plume is likely attributed to the model physics since this is consistent in NAAPS and ENAAPS. Both of the forecasts initialized with the 2dVAR fields capture the event, but like the analysis fields, don't capture the sharp gradient as seen in the MODIS image. On the other hand, the forecasts initialized with the EAKF fields do a better job capturing the AOT gradient at the leading edge of the dust front. This demonstrates that the sharp gradient achieved in the ensemble data assimilation propagates in the forecast. This is an advantage of using the EAKF initial conditions over the 2DVar initial conditions for the short-term forecast.

4.0 Conclusion

This study evaluates the performance of an ensemble aerosol prediction system, ENAAPS-DART, for Navy applications under several configurations, as well as against the current operational system (NAAPS/NAVDAS-AOD). The major findings from this work are:

- Having both meteorology ensembles and perturbations to the aerosol source functions generated the best results. The use of the meteorology ensemble is essential for capturing long-range aerosol transport events. This was demonstrated for dust transport cases off the coast of Africa, as well as at dust impacted AERONET sites in Central America and the United States. The source ensemble is beneficial for capturing spatially large aerosol events, including smoke and dust cases. This was demonstrated for large burning events over Southern Africa and the North American Boreal region.
- The source ensemble can also have a negative impact for regions with sources that behave independently. This is the case for many North American boreal fires that are small and independent. This is also believed to be the case for pollution dominated sites in the United States and Europe. Source ensembles which better represent the statistics for these independent cases are needed.
- An adaptive inflation method from Anderson (2009) was tested for the first time, to our knowledge, for an aerosol application. Based on the results in this work, the adaptive covariance inflation is recommended over a spatially and temporally uniform covariance inflation. The adaptive approach overcomes instability issues that arise due to spatially

heterogeneous observations with the constant inflation approach and it is expected the same finding will apply to other systems. It is also expected that this finding will apply to data assimilation for other atmospheric tracers where the observation density is not spatially uniform .

- A reduction in RMSE can be achieved by increasing the ensemble size from 20 to 80 members. Further reductions may be achieved with optimization of the 80 member ensemble (ie. localization).
- The evaluation of the ensemble diagnostics for the ENAAPS-DART optimization highlighted some potential issues with the smoke emissions used in the simulations. It was found that the ensemble system underrepresents uncertainty for large smoke events as indicated by the total spread (ensemble spread combined with observational error) being much less than the RMSE. Likewise, the rank histograms show an excess at the lower ranks, indicating a positive bias in smoke aerosol relative to MODIS AOT. These findings are supported by the behavior of the AI algorithm in smoke dominated regions, which indicated a large discrepancy between the model predicted and observational distributions. Additionally, the ensemble spread for smoke aerosol is likely too small at low AOT values. Tuning of smoke aerosol emissions is needed to address the identified issues.
- Positive bias in the Eastern United States was also found with the ensemble system. Further work needs to be conducted to determine how to better capture complicated pollution aerosol sources.
- The aerosol analysis fields from the DART-EAKF data assimilation system and the NAVDAS-AOD 2DVAR data assimilation system have similar RMSE and bias relative to AERONET sites on a regional basis. This indicates that both data assimilation systems are able to capture similar aerosol features. However, spatially, the EAKF does a better job of capturing sharp gradients while the 2dVAR system has a smoothing effect. This is a result of the EAKF being able to spread observational information in a flow-dependent manner.
- The ENAAPS-DART system and the NAAPS/NAVDAS-AOD system also had similar RMSE statistics relative to AERONET AOT in the 24 hour forecast. However, the sharpness of features is maintained in the 24-hour forecast with the ENAAPS-DART system, as demonstrated for the Saharan dust transport case. This is an advantage over the current operational system.
- An additional advantage of the ensemble configuration is that uncertainty information in the forecast can be extracted at a given time using the ensemble members. This is an important reason why many NWP forecasting centers have implemented ensemble prediction systems and aerosol forecasting should consider doing the same. With some further tuning for the ENAAPS-DART system based on the findings from this study, additional advantages over the NAAPS/NAVDAS-AOD system can likely be attained.

The ENAAPS-DART system outlined in this work will serve as the base ensemble aerosol prediction system for Navy applications and will serve as a testbed for assimilation of additional, spatially-limited observations, such as ground-based and Lidar observations. ENAAPS-DART will also be used to evaluate aerosol forecast uncertainty, an additional advantage over the current deterministic system. Means for evaluating ensemble system performance were outlined in this work and may provide a useful guideline for future ensemble system developers, particularly with aerosol or other atmospheric tracers. Based on the results from this study, work is underway to understand how additional performance gains can be made in the ENAAPS-DART system through source tuning, increases in the number of ensemble members, and increases in model resolution.

Acknowledgements

This work was conducted as part of a postdoctoral research fellowship from the National Research Council and funded by the Office of Naval Research Code 322. The authors would like to thank the MODIS aerosol team and all of the investigators that participate in the Aerosol Robotic NETwork and make the network of data available. DART is developed and maintained at the National Center for Atmospheric Research which is sponsored by the National Science Foundation. We would also like to thank others for helpful discussion and support. We are grateful to Edward Hyer of NRL for the use of his AERONET verification code. We would also like to thank Cindy Curtis of NRL and Greg Ramos of IMRI Inc. for their software engineering support of the groups modeling and remote sensing architecture. Finally, we would like to thank Arthur Mizzi of NCAR for a helpful discussion.

References

Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L.R., Qian, Y., Chung, C.E., Ramanathan, V., and Carmichael, G.R.: A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique. *Atmos. Environ.*, 42 (37), 8600-8615, doi:10.1016/j.atmosenv.2008.08.031, 2008.

Anderson, J.L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, 9, 1518-1530, 1996.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev*. 127, 2741–2758, 1999.

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev*. 129, 2894–2903, 2001.

Anderson, J.L.: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, 59A, 210-224, 2007.

Anderson, J.L, Hoar, T., Raeder, K., Liu, H., and Collins, N.: The Data Assimilation Research Testbed: A Community Facility. *Bull. Amer. Meteor.Soc*., 90, 1283-1296, 2009.

Anderson, J.L.: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, 61A, 72-83, 2009.

Anderson, J.L., and Lei, L.: Empirical localization of observation impact in Ensemble Kalman Filters. *Mon. Wea. Rev*., 141, 4140-4153. doi: 10.1175/MWR-D-12-00330.1, 2013.

Arellano, A.F., Raeder, K., Anderson, J. L., Hess, P. G., Emmons, L. K., Edwards, D. P., Pfister, G. G., Campos, T. L., and Sachse, G. W.: Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission, *Atmos. Chem. Phys.*, 7, 5695-5710, doi:10.5194/acp-7-5695-2007, 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.

Bogdanoff, A. S., Westphal, D, L., Campbell, J. R., Cummings, J. A., Hyer, E. J., Reid, J. S., and Clayson, C. A.: Sensitivity of infrared sea surface temperature retrievals to the vertical distribution of airborne dust aerosol, *Remote Sens of Environ*., 159, 1-13, doi:10.1016/j.rse.2014.12.002 , 2015.

Bowler, N. E., Arribas, A., Mylne, K.R., Robertson, K.B., and Beare, S.E.: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc*., 134, 703–722, 2008.

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. Weather Rev*., 133, 1076– 1097, 2005.

Christensen, J. H.: The Danish eulerian hemispheric model—A three-dimensional air pollution model used for the arctic, *Atmos. Environ*., 31, 4169–4191, doi:10.1016/S1352-2310(97)00264-1,1997.

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. Geophys. Res.*, 115, D14207, doi:10.1029/2009JD012820, 2010.

Collins, W.D, Rasch, P.J., Eaton, B.E., Khattatov, B.V., Lamarque, J-F., and Zender, C.S.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX. *J. Geophys. Res.*, 106(D7), 7313-7336, doi: 10.1029/2000JD900507, 2001.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to do forecast error statistics. *J. Geophys. Res*. 99(C5), 10143–10162, 1994.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc*. 125, 723–757, 1999.

Generoso, S., Br´eon, F.-M., Chevallier, F., Balkanski, Y., Schulz, M., and Bey, I.: Assimilation of POLDER aerosol optical thickiness into the LMDz-INCA model: Implications for the Arctic aerosol burden, *J. Geophys. Res*., 112, D02311,doi:10.1029/2005JD006954, 2007.

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)**,** 20,255-20,273, 2001.

Hacker, J., and Angevine, W.M.: Ensemble data assimilation to characterize surface-layer errors in numerical weather prediction models. *Mon. Wea. Rev*, 141, 1804-1821, doi:http://dx.doi.org/10.1175/MWR-D-12-00280.1, 2013.

Hamill, T.M.: Interpretation Of Rank Histograms For Verifying Ensemble Forecasts. ***Mon. Wea. Rev,*** 129, 3, 550-560, doi: 10.1175/1520-0493, 2001.

Hamill, T. M., Whitaker, J. S. and Snyder C.: Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev*. 129, 2776–2790, 2001.

Hamill, T. M. and Whitaker, J. S.: Accounting for error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev*. 133, 3132–3147, 2005.

Hogan, T. F., and Rosmond, T.E.: The description of the Navy Operational Global Atmospheric Prediction Systems spectral forecast model, *Mon. Wea. Rev.*, 119, 1786–1815, doi:10.1175/1520-0493, 1991.

Hogan, T.F., Liu, M., Ridout, J.S., Peng, M.S., Whitcomb, T.R., Ruston, B.C., Reynolds, C.A., Eckermann, S.D., Moskaitis, J.R., Baker, N.L., McCormack, J.P., Viner, K.C., McLay, J.G., Flatau, M.K., Xu, L., Chen, C., and Chang, S.W.: The Navy Global Environmental Model. *Oceanography, Special Issue on Navy Operational Models*, 27, No. 3, 2014.

Holben, B. N., et al.: AERONET—A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, 66, 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.

Houtekamer, P.L. and Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique, *Mon. Wea. Rev*, 126, p 796-811, 1998.

Houtekamer, P.L. and Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev*, 126, 796-811, 2001.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon.Weather Rev.* 133: 604–620, 2005.

Hsu, N. C., Gautam, R., Sayer, A.M., Bettenhausen, C., Li, C., Jeong, M.J., Tsay, S.-C., and Holben, B.N.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. *Atmos. Chem. Phys*., 12, 8037–8053, 2012.

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.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, 2003.

Khade, V.M., Hansen, J.A., Reid, J.S., and Westphal, D.L.: Ensemble filter based estimation of spatially distributed parameters in a mesoscale dust model: experiments with simulated and real data, *Atmos. Chem and Phys.*, 13, 3481-3500. 10.5194/acp-13-3481-2013, 2013.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of

reactive gases and aerosols: methodology and application, *Atmos. Chem. Phys*., 10, 7017–7039, doi:10.5194/acp-10-7017-2010, 2010.

Li, H., Kalnay, E., and T. Miyoshi, T.: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc*., 135, 523–533, 2009.

Lynch, P., Reid, J.S., Westphal, D.L., Zhang, J., Hogan, T.F., Hyer, E.J., Curtis, C.A., Hegg, D.A., Shi, Y., Campbell, J.R., Rubin, J.I., Sessions, W.R., Turk, F.J., and Walker, A.L.: Development studies towards an 11-year global gridded aerosol optical thickness reanalysis for climate and applied applications. *Geosci. Model. Dev. Discuss*., 8, 10455-10538, doi:10.5194/gmdd-8-10455-2015, 2015.

McLay, J.G., Bishop, C.H., and Reynolds, C.A.: A local formulation of the ensemble transform (ET) analysis perturbation scheme. The ensemble-transform scheme adapted for the generation of stochastic forecast perturbations. *Wea. Forecasting*. 25, 985-993, 2010.

Merchant, C. J., Embury, O., Le Borgne, P., and Bellc, B.: Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature, *Rem Sens. of Environ*., 104, 15-30. doi: 10.1016/j.rse.2006.03.007, 2006.

Miyoshi, T., Sato, Y., and Kadowaki, T.: Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Mon. Wea. Rev*., 138, 2846–2866, 2010.

Miyoshi, T.: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. *Mon. Wea. Rev*. 139, 1519-1535. DOI: 10.1175/2010MWR3570.1, 2011.

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 (D17), 4559-4573, doi: 10.1029/2002JD002975, 2003.

Pérez, C., Haustein, K., Janjic, Z., Jorba, O., Huneeus, N., Baldasano, J. M., Black, T., Basart, S., Nickovic, S., Miller, R. L., Perlwitz, J. P., Schulz, M., and Thomson, M.: Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model – Part 1: Model description, annual simulations and evaluation, *Atmos. Chem. Phys*., 11, 13001–13027, doi:10.5194/acp-11-13001-2011, 2011.

Peterson, D.A., Hyer, E.J. Campbell, J.R., Fromm, M.D., Hair, J.W., Butler, C.F., and M.A. Fenn, M.A.: The 2013 Rim Fire: Implications for Predicting Extreme Fire Spread, Pyroconvection, and Smoke Emissions. *Bull. Amer. Meteor.Soc*. http://dx.doi.org/10.1175/BAMS-D-14-00060.1, 2015.

Pagowski, M., and Grell, G.A.: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, *J. Geophys. Res*., 117, D21302, doi:10.1029/2012JD018333, 2012.

Prospero, J. M.: Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality. J.Geophys. Res., 104, 15 917–15 928, 1999.

Prospero, J, and Lamb, P.J.: African droughts and dust transport to the Caribbean: Climate change implications. *Science*, 302, 1024–1027, 2003.

Prospero, J.M., and Mayol-Bracero, O.L.: Understanding the transport and impact of Africa dust on the Caribbean basin. *Bull. Amer. Meteor. Soc.*, 94, 1329-1337, doi:http://dx.doi.org/10.1175/BAMS-D-12-00142.1, 2013.

Raeder, K., Anderson, J.L., Collins, N., Hoar, T.J., Kay, J.E., Lauritzen, P.H., and Pincus, R.: DART/CAM: An Ensemble Data Assimilation for CESM Atmospheric Models. *J. Climate*, 25 6304-6317 doi:10.1175/JCLI-D-11-00395.1, 2012.

Reid, J.S., Kinney, J.E., Westphal, D.L., Holben, B.N., Welton, E.J., Tsay, S.-C., Eleuterio, D.P., Campbell, J.R., Christopher, S.A., Colarco, P.R., Jonsson, H.H., Livingston, J.M., Maring, H.B., Meier, M.L., Pilewskie, P., Prospero, J.M., Reid, E.A., Remer, L.A., Russell, P.B., Savoie, D.L., Smirnov, A., and Tanre, D.: Analysis of measurements of Saharan dust by airborne and ground-based remote sensing methods during the Puerto Rico Dust Experiment (PRIDE). *J. Geophys. Res*., 108, 8586, doi:10.1029/2002JD002493, 2003.

Reid, J.S., Prins, E.M., Westphal, D.L., Schmidt, C.C., Richardson, K.A., Christopher, S.A., Eck, T.F., Reid E.A., Curtis, C.A., and Hoffman J.P.: Real-time monitoring of South American smoke particle emissions and transport using a coupled remote sensing/box-model approach. *Geophys. Res. Lett*. 31 L06107, 2004.

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., and Hoffman, J.P.: Global monitoring and forecasting of biomass burning smoke: Description of and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) program, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*., 2, 144–162, doi:10.1109/ JSTARS.2009.2027443, 2009.

Reid, J.S., Hyer, E.J., Johnson, R.S., Holben, B.N., Yokelson, R., Zhang, J., Campbell, J.R., Christopher, S., Di Girolamo, L., Giglio, L., Holz, R.E., Kearney, C., Miettinen, J., Reid, E.A., Turk, J., Wang, J., Xian, P., Zhao, G., Balasubramanian, R., Chew, B.N., Janjai, S., Lagrosas, N., Lestari, P., Lin, N., Mahmud, M., Nguyen, A.X., Norris, B., Oanh, N.T.K., Oo, M., Salinas, S.V., Welton, J. and Liew, S.C. 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, 2013.

Rubin, J. I., and Collins, W. D.: Global simulations of aerosol amount and size using MODIS observations assimilated with an Ensemble Kalman Filter, *J. Geophys. Res. Atmos*., 119, 12,780–12,806, doi:10.1002/2014JD021627, 2014.

Saetra, O., Hersbach, H., Bidlot, J.R., and Richardson, D.S.: Effects of observation errors on the statistics for ensemble spread and reliability. *Mon. Wea. Rev.,* 132: 1487-1501, 2004.

Schroeder, W., Prins, E.M., Giglio, L., Csiszar, I., Schmidt, C., Morissette, J., and Morton, D.: Validation of GOES and MODIS active fire detection products using ASTER and ETM plus data. *Remote Sens. Environ.,* 112, 2711–2726, 2008.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Applying an ensemble Kalman filter to the assimilation of AERONET observations in a global aerosol transport model, *Atmos. Chem. Phys*., 10, 2561–2576, doi:10.5194/acp-10-2561-2010, 2010a.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, *Atmos. Chem. Phys*., 10, 6583-6600, doi:10.5194/acp-10-6583-2010, 2010b.

Schwartz, C.S., Liu, Z., Lin, H-C, and Cetola, J.D.: Assimilating aerosol observations with a "hybrid" variational-ensemble data assimilation system. *J. Geophys. Res. Atmos*., 119, 7,4043-4069, doi: 10.1002/2013JD020937, 2014.

Sekiyama, T. T., Tanaka, T.Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, *Atmos. Chem. Phys*.,10, 39–49, doi:10.5194/acp-10-39-2010, 2010.

Sessions, W.R., Reid, J.S., Benedetti, A., Colarco, P.R., da Silva, A., et al. (2015). Development towards a global operational aerosol consensus: basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME). *Atmos. Chem. Phys.,* 15, 335-362, doi:10.5194/acp-15-335-2015, 2015.

Shalaby, A., Rappenglueck, B. and Elthir, E.A.B.: The climatology of dust aerosol over the Arabian pensinula. *Atmos. Chem. Phys.,* 15, 1523-1571, 2015.

Shi, Y., Zhang, J., Reid, J.S., Holben, B.N., Hyer, E.J., and Curtis, C.: An analysis of the collection 5 MODIS over-ocean aerosol optical depth product for its implication in aerosol assimilation, *Atmos. Chem. Phys*., 11, 557–565, doi:10.5194/acp-11-557-2011, 2011.

Szunyogh, I., Kostelich, E.J., Gyarmati, G., Kalnay, E., Hunt, B.R., Ott, E., Satterfield, E., Yorke, J.A.: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, 60A: 113–130, 2008.

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. Meteorol. Geophys.,* 53, 119–138, 2003.

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, J., and Christopher, S.A.: Mesoscale modeling of Central American smoke transport to the United States, part II: smoke radiative impact on regional surface energy budget and boundary layer process. *J. Geophys. Res.* 111, D14S92, 2006.

Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y., Toth, Z.: Ensemble data assimilation with the NCEP global forecasting system. *Mon. Weather Rev*. 136: 463–482, 2008.

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

Yu, H., Dickinson, R. E., Chin, M., Kaufman, Y. J., Geogdzhayev, B., and Mishchenko, M. I.: Annual cycle of global distributions of aerosol optical depth from integration of MODIS retrievals and GOCART model simulations. *J. Geophys. Res*., 108, 4128,doi:10.1029/2002JD002717, 2003.

Zhang, J., Reid, J.S., and Holben, B.N.: An analysis of potential cloud artifacts in MODIS over ocean aerosol optical thickness products. *Geophys. Res. Lett*., 32, L15803, doi:10.1029/2005GL023254, 2005.

Zhang, J., and Reid, J.S.: MODIS Aerosol Product Analysis for Data Assimilation: Assessment of 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.

Zhang, J., Campbell, J.R., Reid, J.S., Westphal, D.L., Baker, N.L., Campbell, W.F., and Hyer, E.J.: Evaluating the impact of assimilating CALIOP-derived aerosol extinction profiles on a global mass transport model. *Geophys. Res. Lett*. 38, L14801, doi:10.1029/2011GL047737, 2011.

Zhang, J., Campbell, J. R., Hyer, E. J., Reid, J. S., Westphal, D. L., and Johnson R. S.: Evaluating the impact of multisensor data assimilation on a global aerosol particle transport model, J. Geophys. Res. Atmos., 119, 4674–4689, doi:10.1002/2013JD020975, 2014.

Table 1. Summary of five ENAAPS-DART experiments conducted for EAKF optimization. The experiments include variations in ensemble generation (meteorology or source only, meteorology with source ensemble), number of ensemble members, and the covariance inflation method. The meteorology ensemble uses NOGAPS ensemble meteorology fields and the source ensembles use a 25% random Gaussian perturbation to the aerosol source functions.

Experiment Name	Ensembles	Inflation
Source, const	Source, 20 member	10% Constant Covariance Inflation
Source, adaptive	Source, 20 member	Adaptive Inflation
Meteorology, adaptive	Meteorology Only, 20 member	Adaptive Inflation
Met+Source, adaptive	Meteorology + Source, 20 member	Adaptive Inflation
Met+Source, 80	Meteorology + Source, 80 member	Adaptive Inflation

Table 2. Global and regional diagnostics for four EAKF optimization experiments conducted during the July through August, 2013 timeperiod. The diagnostics are computed using the ENAAPS-DART 6-hour AOT (550nm) forecasts against MODIS AOT (550nm), prior to assimilation. The root mean squared error (RMSE) is shown as well as the average ratio between the total spread (ensemble spread in AOT + observational AOT error) and the RMSE. Well-tuned ensemble systems should have a small RMSE that is approximately equal to the total spread.

	RMSE (Standard Deviation)				Mean (Total Spread/RMSE) Ratio			
Region	Source, const	Source, AI	Met, AI	Met+Source, AI	Source, const	Source, AI	Met, AI	Met+Source, AI
Global	0.127 (0.095)	0.123 (0.086)	0.122 (0.083)	0.115 (0.077)	0.802	0.82	0.875	0.925
North American Boreal	0.084 (0.074)	0.084 (0.074)	0.091 (0.079)	0.085 (0.072)	1.387	1.355	1.254	1.298
ECONUS	0.071 (0.04)	0.071 (0.038)	0.069 (0.031)	0.069 (0.033)	1.298	1.28	1.225	1.234
WCONUS	0.152 (0.119)	0.153 (0.123)	0.15 (0.114)	0.139 (0.111)	0.956	0.965	1.017	1.084
Central America	0.094 (0.052)	0.099 (0.051)	0.064 (0.038)	0.064 (0.038)	1.142	1.041	1.662	1.661
South America	0.069 (0.019)	0.071 (0.021)	0.076 (0.025)	0.067 (0.018)	1.158	1.149	1.091	1.214
South Africa	0.133 (0.048)	0.128 (0.043)	0.14 (0.065)	0.124 (0.046)	0.69	0.745	0.721	0.8
North Africa	0.174 (0.111)	0.176 (0.099)	0.166 (0.086)	0.163 (0.082)	0.837	0.806	0.911	0.918
Europe	0.098 (0.045)	0.094 (0.039)	0.09 (0.036)	0.09 (0.037)	0.863	0.889	0.989	0.994
Eurasian Boreal	0.176 (0.211)	0.166 (0.193)	0.155 (0.181)	0.15 (0.167)	0.799	0.819	0.925	0.934
East Asia	0.143 (0.055)	0.141 (0.055)	0.165 (0.094)	0.161 (0.09)	0.951	0.956	0.958	0.975
India	0.149 (0.076)	0.158 (0.076)	0.134 (0.069)	0.134 (0.07)	1.131	1.007	1.322	1.501
Southeast Asia	0.083 (0.036)	0.085 (0.037)	0.08 (0.036)	0.079 (0.035)	1.075	1.037	1.144	1.155
Australia	0.04 (0.006)	0.04 (0.006)	0.044 (0.009)	0.042 (0.007)	1.505	1.482	1.395	1.447
NH Pacific	0.089 (0.056)	0.091 (0.056)	0.088 (0.061)	0.082 (0.053)	1.242	1.237	1.333	1.386
SH Pacific	0.035 (0.013)	0.037 (0.013)	0.034 (0.011)	0.034 (0.011)	2.134	2.003	2.098	2.106
NH Atlantic	0.099 (0.061)	0.099 (0.061)	0.093 (0.058)	0.092 (0.058)	0.979	0.99	1.145	1.163
SH Atlantic	0.088 (0.086)	0.085 (0.093)	0.099 (0.147)	0.088 (0.11)	1.291	1.304	1.318	1.366
Indian Ocean	0.079 (0.036)	0.085 (0.036)	0.074 (0.033)	0.073 (0.031)	1.16	1.076	1.279	1.291
Southern Ocean	0.04 (0.018)	0.04 (0.016)	0.047 (0.021)	0.047 (0.021)	2.08	1.997	1.732	1.759

Table 3. Regional statistics of the analysis AOT against AERONET AOT (550nm) (Zhang and Reid, 2006) for a six month simulation (April-September, 2013). The statistics are shown for the analysis AOT from the 2DVar NAVDAS-AOD assimilation system and the EAKF data assimilation from ENAAPS-DART.

	2DVar (NAVDAS-AOD)				EAKF (ENAAPS-DART)				AERONET
Region	R^2	Bias	RMSE	Mean AOT	R^2	Bias	RMSE	Mean AOT	Mean AOT
North American Boreal	0.38	0.021	0.068	0.094	0.43	0.026	0.067	0.098	0.072
ECONUS	0.55	-0.001	0.066	0.147	0.53	0.013	0.068	0.162	0.147
WCONUS	0.32	0.024	0.07	0.116	0.27	0.02	0.07	0.112	0.093
Central America	0.58	-0.023	0.107	0.18	0.61	0.016	0.102	0.189	0.205
South America	0.33	0.001	0.074	0.09	0.23	-0.01	0.081	0.079	0.088
North Africa	0.58	0.002	0.161	0.259	0.59	0.044	0.167	0.301	0.257
Europe	0.55	0.01	0.092	0.166	0.49	0.011	0.097	0.167	0.156
Eurasian Boreal	0.65	-0.005	0.068	0.132	0.58	-0.004	0.076	0.134	0.137
East Asia	0.65	-0.04	0.168	0.289	0.60	-0.044	0.184	0.286	0.33
India	0.38	-0.016	0.252	0.402	0.39	-0.058	0.25	0.359	0.418
Insular SE Asia	0.52	-0.017	0.13	0.166	0.52	0.005	0.15	0.186	0.182
Peninsular SE Asia	0.64	-0.016	0.194	0.351	0.72	-0.024	0.171	0.343	0.367
Southwest Asia	0.61	0.019	0.15	0.355	0.48	-0.001	0.166	0.338	0.339
Australia	0.43	-0.008	0.043	0.055	0.21	0.01	0.048	0.072	0.062
Ocean	0.64	0.017	0.064	0.127	0.67	0.022	0.062	0.131	0.109

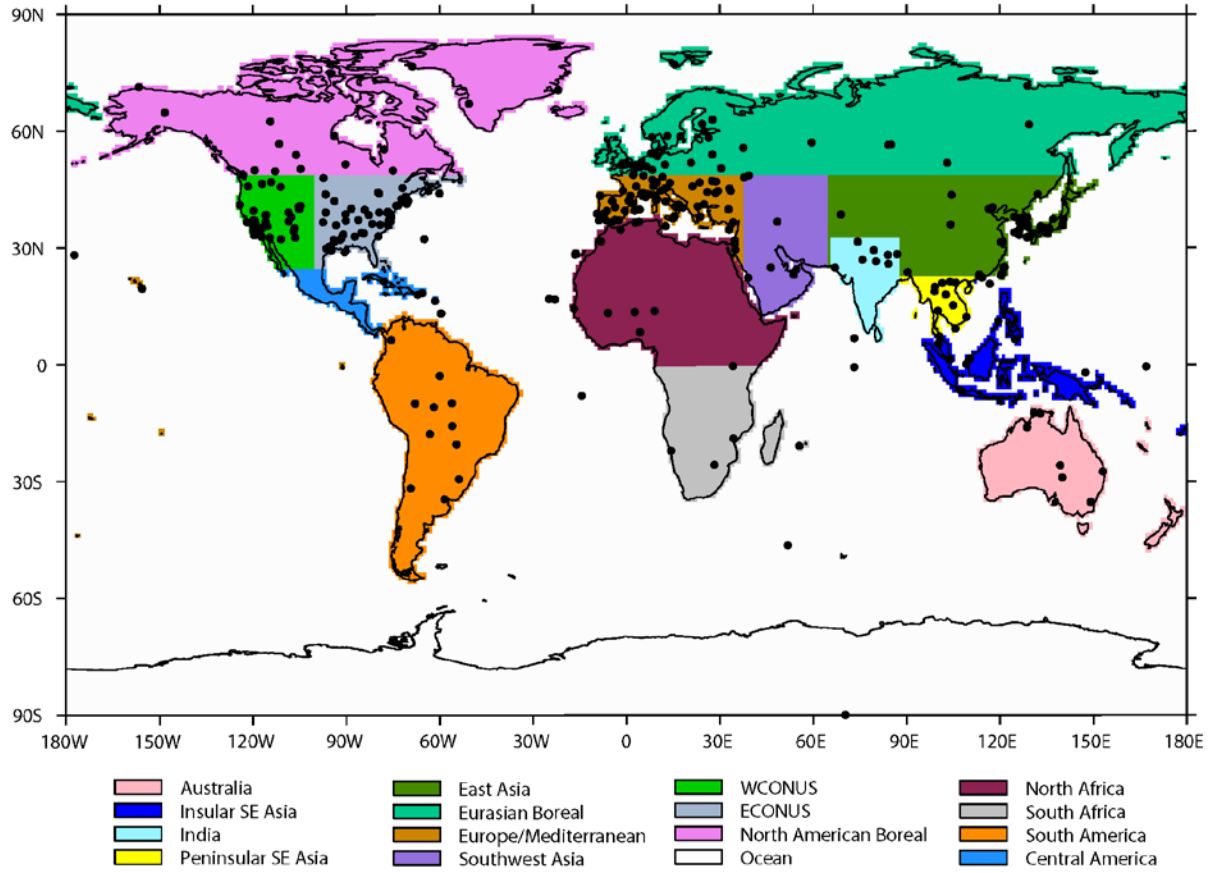

Figure 1. Diagnostic regions for evaluated ENAAPS-DART experiments. Black dots indicate AERONET sites with data available for 2013.

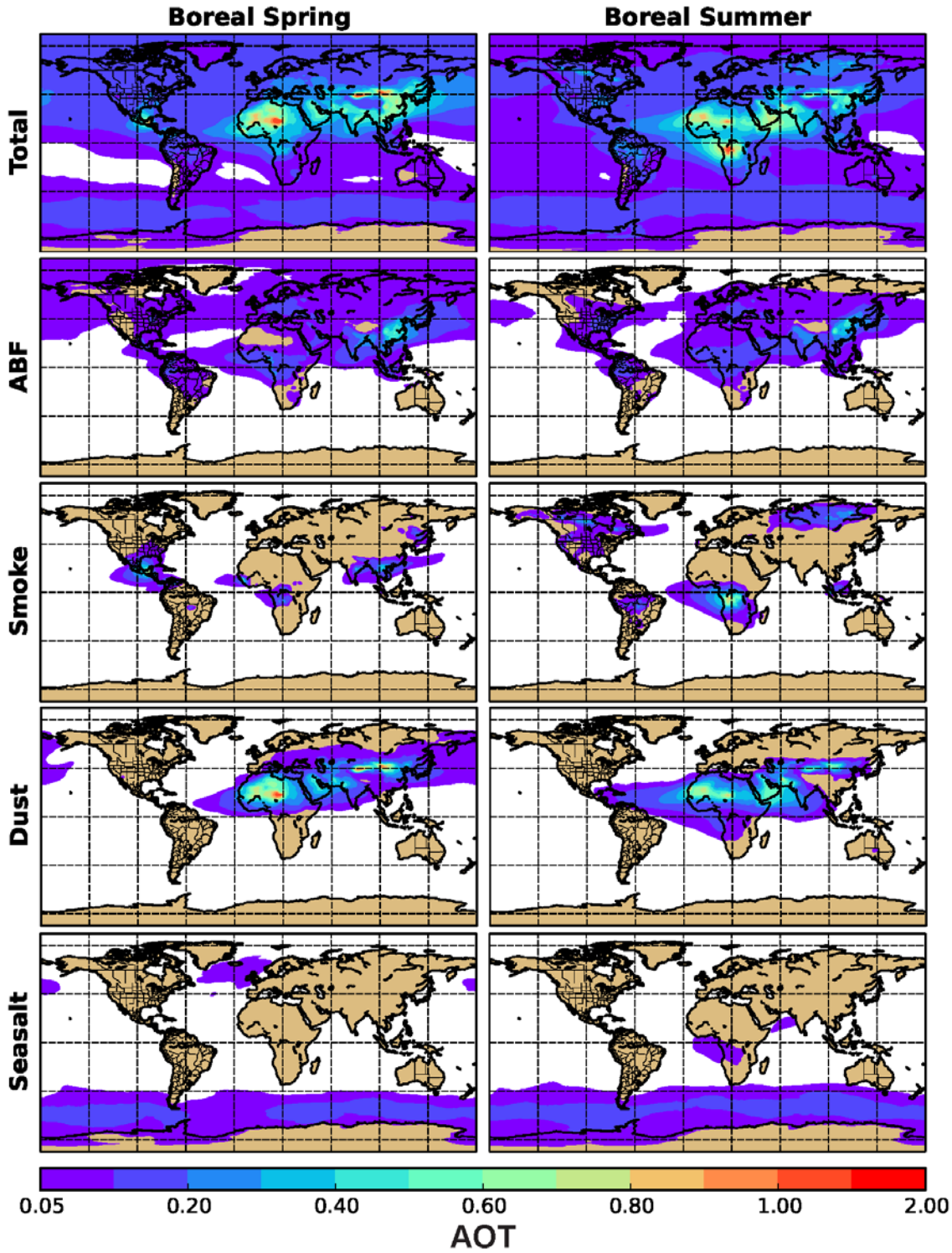

Figure 2. Seasonally averaged AOT (550nm) fields (posterior), predicted by the ENAAPS-DART system (Met+Source, adaptive), for the Boreal Spring (April, May) and Summer (June-September), 2013. Results are shown for total AOT and AOT attributed to combined anthropogenic and biogenic fine (ABF), smoke, dust, and seasalt aerosol, respectively.

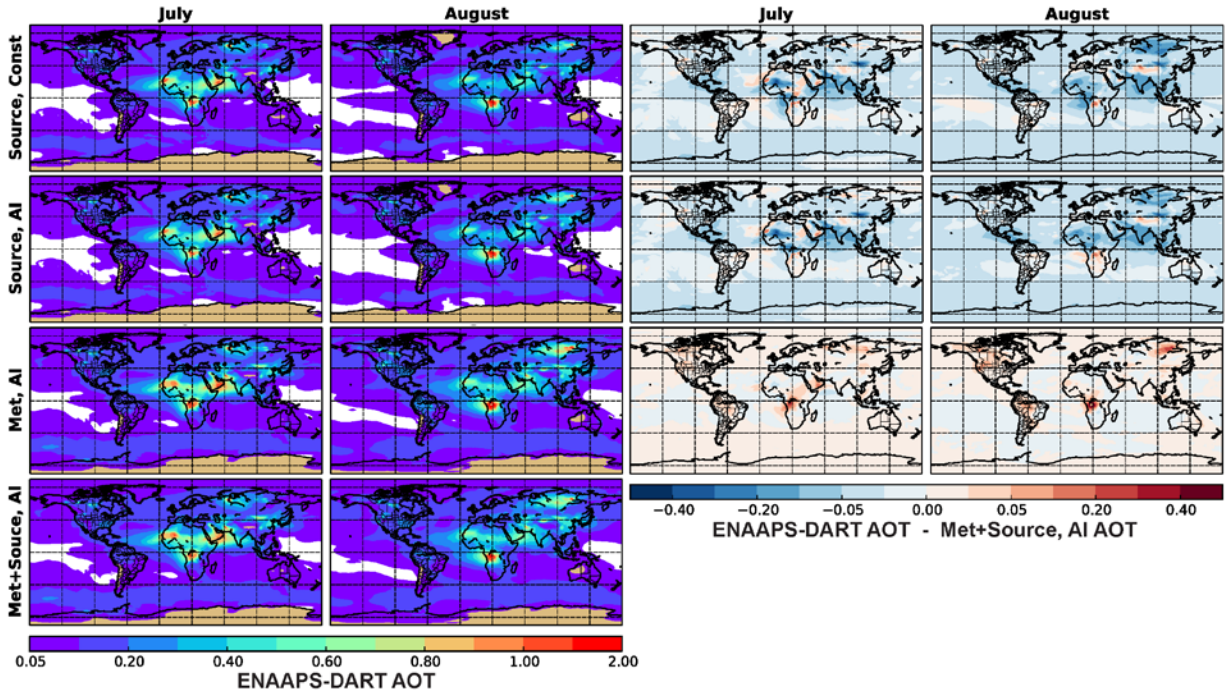

Figure 3. Monthly averaged AOT (550nm) for four ENAAPS-DART EAKF optimization experiments, including a source ensemble with constant inflation (Source, Const), a source ensemble with adaptive inflation (Source, AI), a meteorology ensemble with adaptive inflation (Met, AI), and a combined meteorology and source ensemble with adaptive inflation (Met+Source, AI). Also shown is the average difference in AOT between the identified ENAAPS-DART experiment and the combined meteorology and source ensemble experiment (Met+Source, AI).

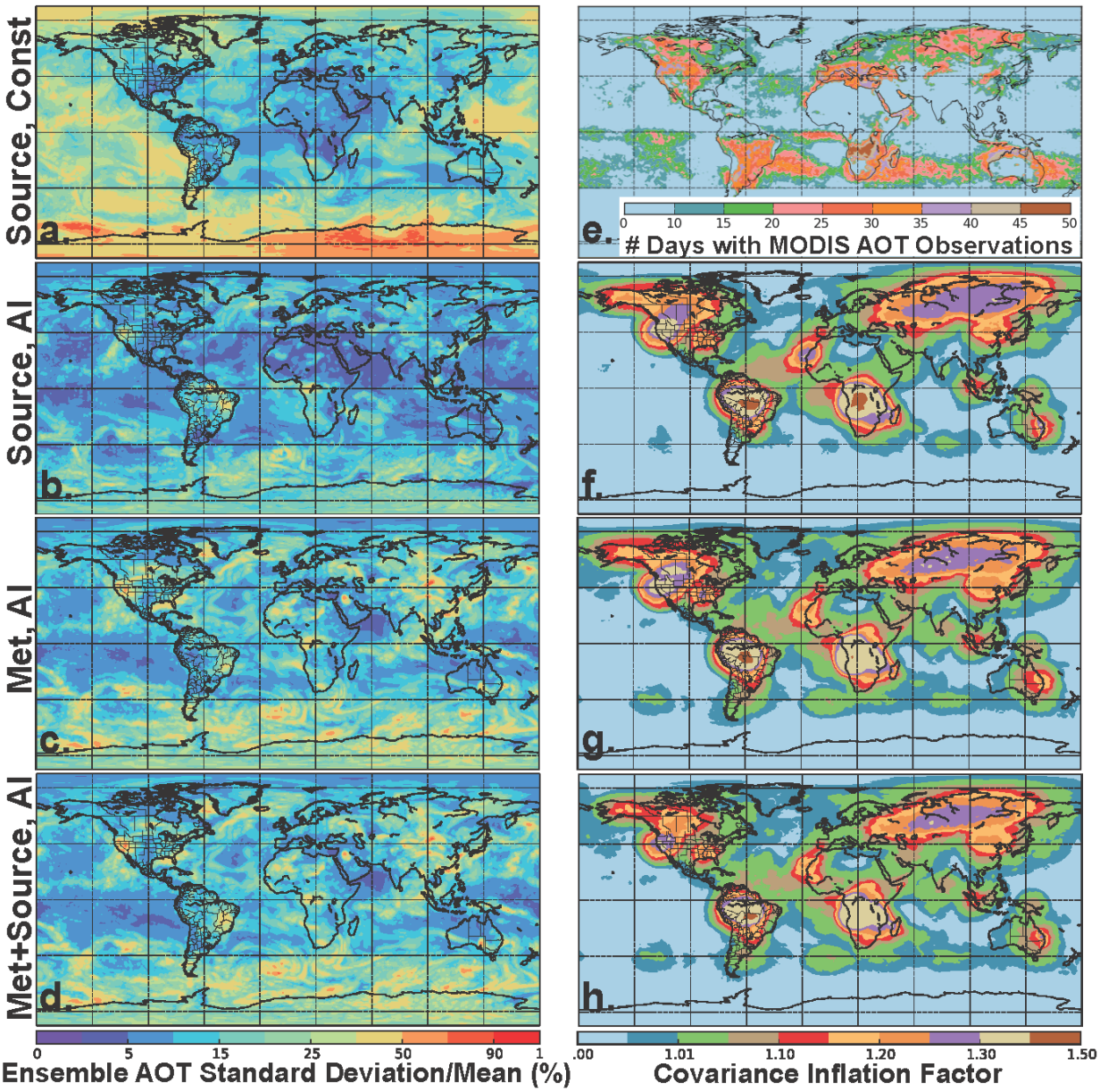

Figure 4. The standard deviation of the prior ensemble aerosol optical thickness normalized by the ensemble mean at the end of the experimental time period (August 31st, 1800) for four ENAAPS-DART experiments: a) source only ensemble with spatially and temporally constant 10 percent covariance inflation b) source only ensemble with adaptive inflation c) meteorology ensemble only with adaptive inflation and d) combined meteorology and source ensemble with adaptive inflation. Also shown are e) The count of days with MODIS 1-degree gridded data assimilation quality AOT observations (Zhang et al. 2005, 2006; Hyer et al., 2011) available for assimilation during the July 15 to August 31, 2013 time period and f) the average inflation factor for the source only adaptive inflation g) the average inflation factor for the meteorology only adaptive inflation experiment and h) the average inflation factor for the combined meteorology and source ensemble adaptive inflation experiment. For adaptive covariance inflation, regions with high observation density are coincident with inflation regions.

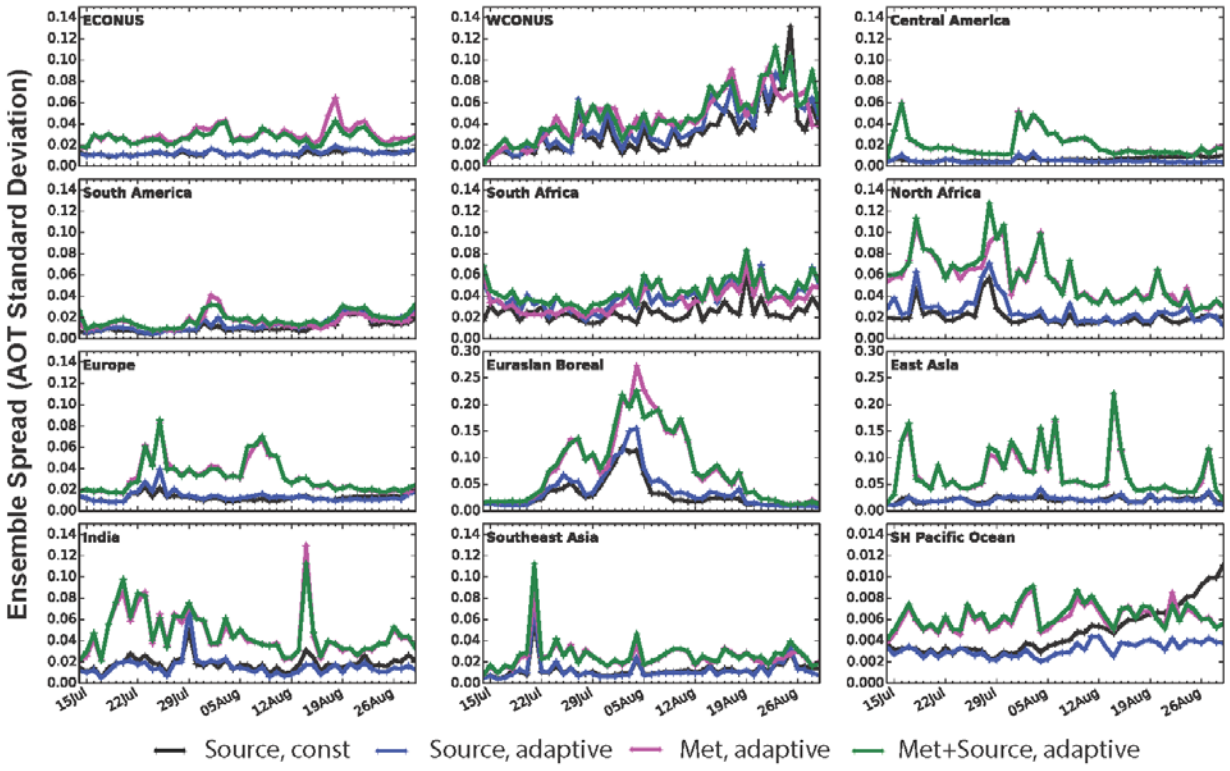

Figure 5. Timeseries of ensemble spread (AOT standard deviation) for 4 ENAAPS-DART experiments over the July 15 through August, 2013 time period. Results are shown for 12 regions, including the Eastern United States, the Western US, Central America, South America , South Africa, North Africa, Europe, Eurasian Boreal, East Asia, India, Southeast Asia, and the Southern Hemisphere Pacific Ocean.

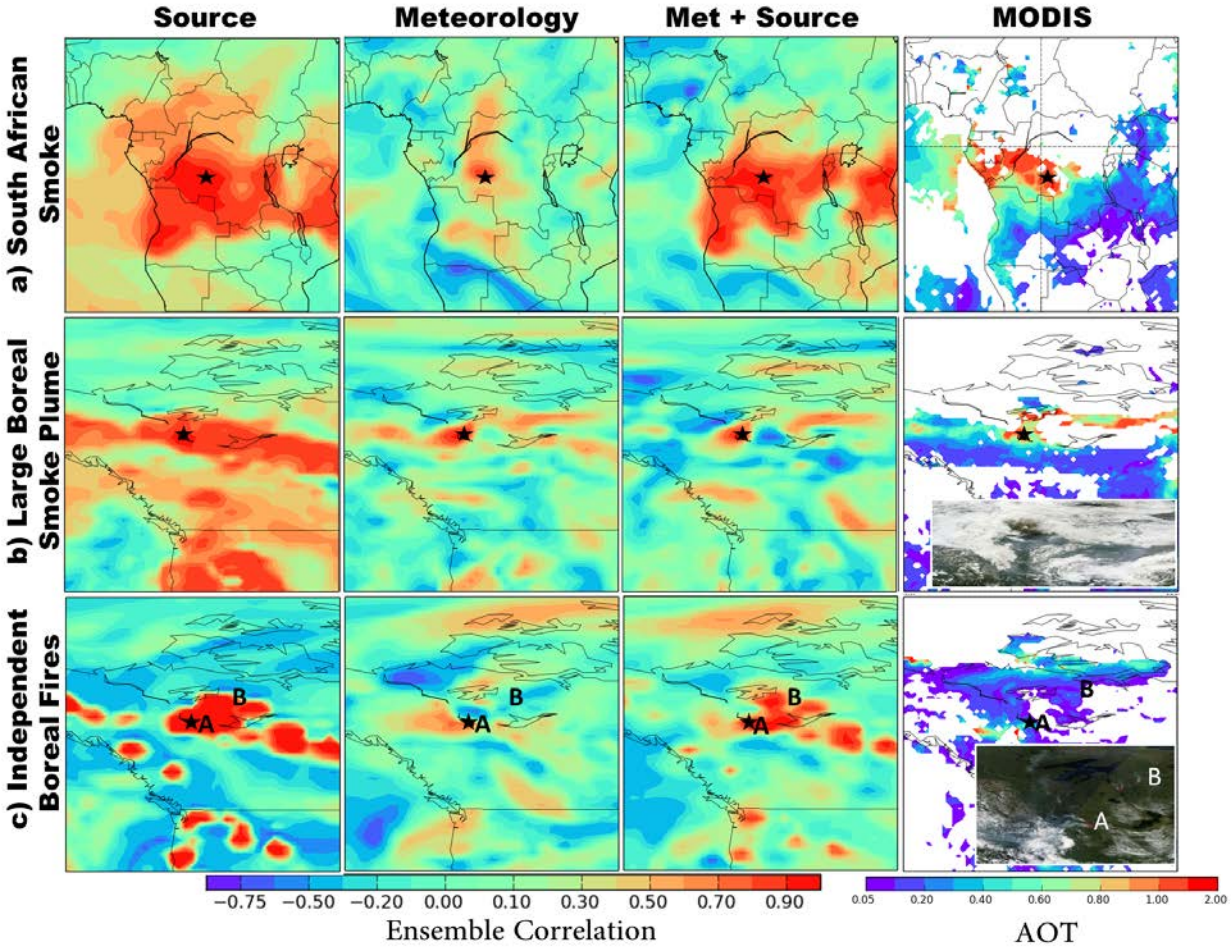

Figure 6. Ensemble correlation fields in the prior AOT relative to a point indicated by a black star for three different aerosol events: a) a South African smoke event on August 2, 2013 b) a large North American Boreal smoke plume on August 15, 2013 and c) small independent Boreal fires in North America on August 7, 2013. Correlation fields are shown for three ENAAPS-DART configurations, source ensemble (Source), NOGAPS meteorology ensemble (Meteorology), and a combined meteorology and source ensemble (Met + Source). Also included are the MODIS AOT (550nm) observations for the smoke events, as well as a zoomed in look at the MODIS visible image with MODIS fire detections in red for the two North American Boreal cases.

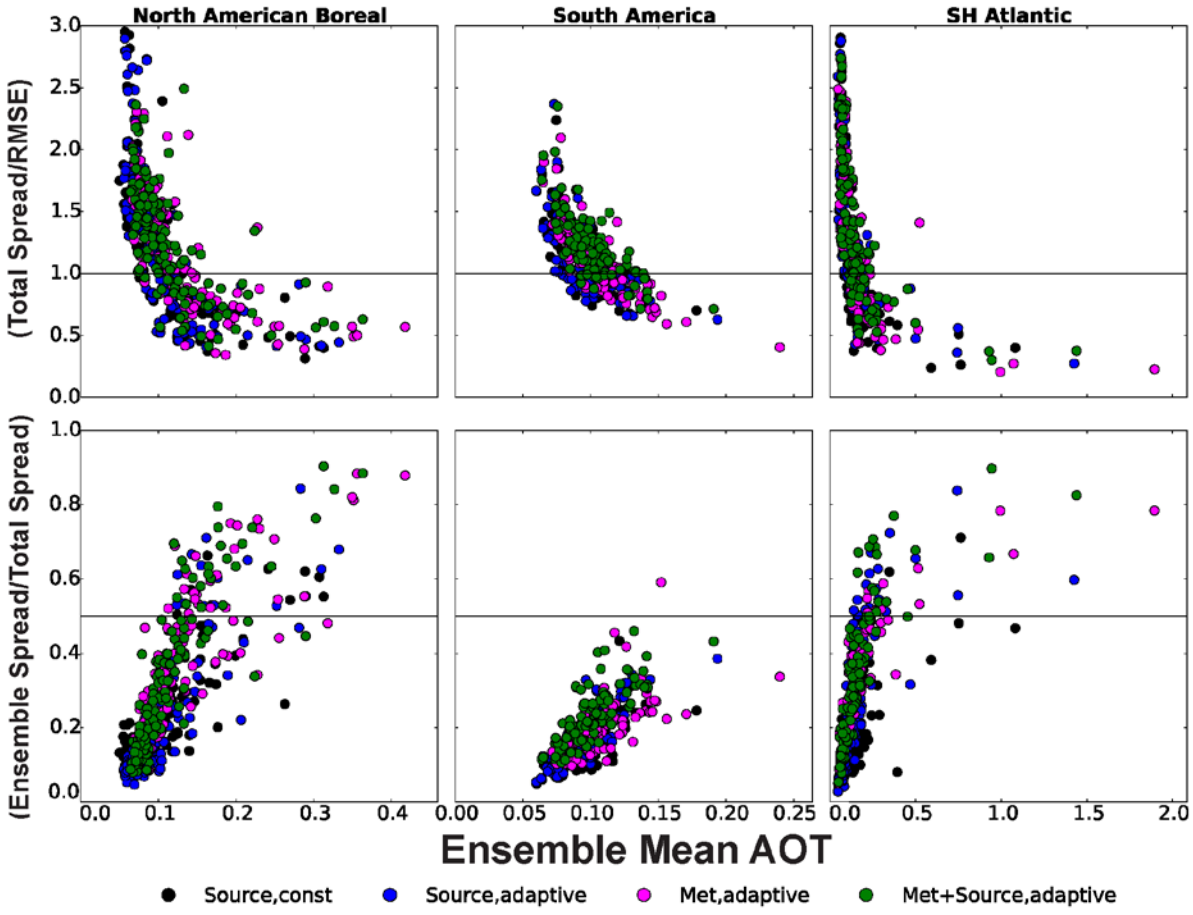

Figure 7. Regional scatterplots of the ratio of total spread (combined ensemble AOT spread and MODIS AOT error) to RMSE against the ensemble mean AOT (550nm) (top row) and the ratio of ensemble spread to total spread against the mean AOT (550nm) (bottom row). Results are shown for four ENAAPS-DART configurations including source ensemble with constant covariance inflation, source ensemble with adaptive inflation, meteorology ensemble with adaptive inflation, and a combined meteorology and source ensemble with adaptive inflation.

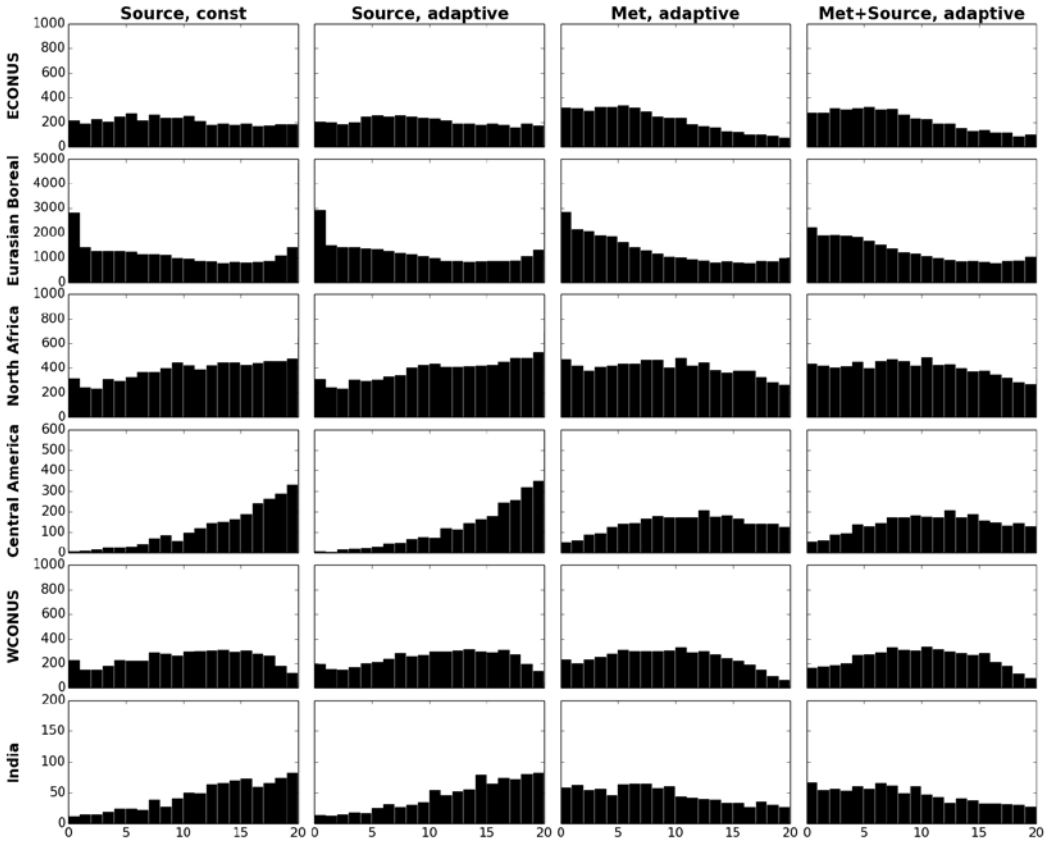

Figure 8. Rank histograms for select regions for the four ENAAPS-DART experiments, including source only ensemble with constant and adaptive inflation (Source, const; Source, adaptive), meteorology only ensemble with adaptive inflation (Met, adaptive), and meteorology plus source ensemble with adaptive inflation (Met+Source, adaptive).

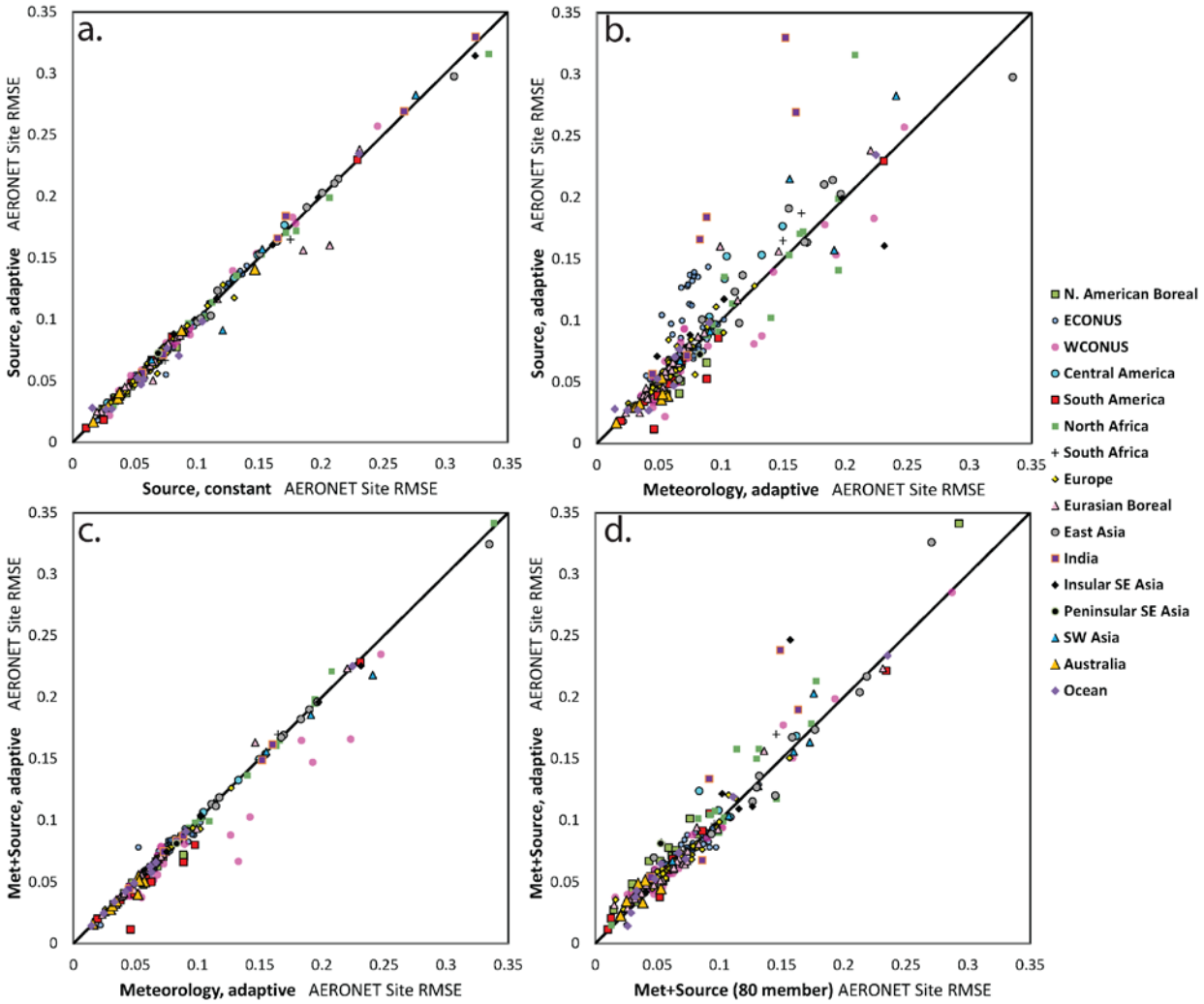

Figure 9. Scatterplots of ENAAPS-DART RMSE relative to AERONET AOT (550nm, Zhang and Reid, 2006) by site between different ENAAPS-DART experiments. Sites are identified by region. Results are shown for a) source only with constant covariance inflation versus adaptive inflation b) meteorology only versus source only ensemble c) meteorology only versus meteorology+source ensemble and d) meteorology+source 20 member ensemble against a meteorology+source 80 member ensemble.

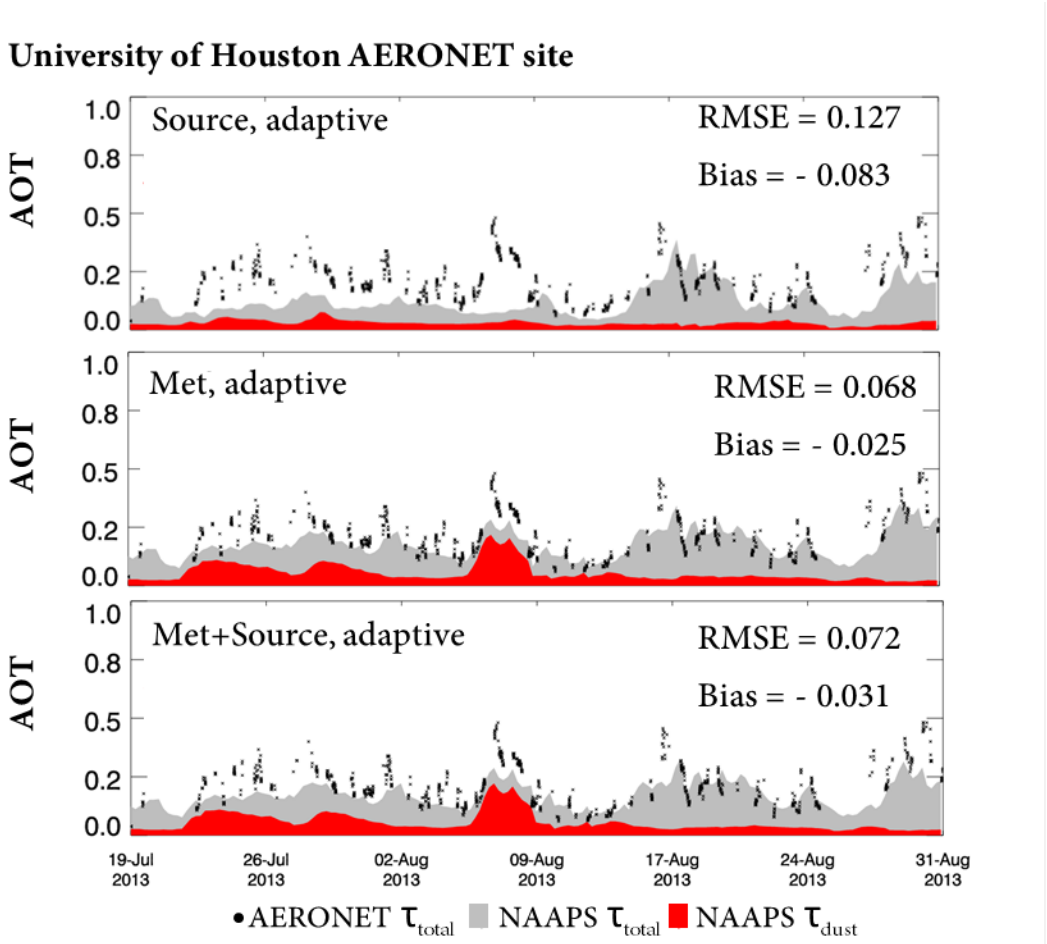

Figure 10. Timeseries of model predicted total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the University of Houston AERONET site. Results are shown for adaptive inflation experiments with source only ensemble, NOGAPS meteorology ensemble, and a combined meteorology and source ensemble.

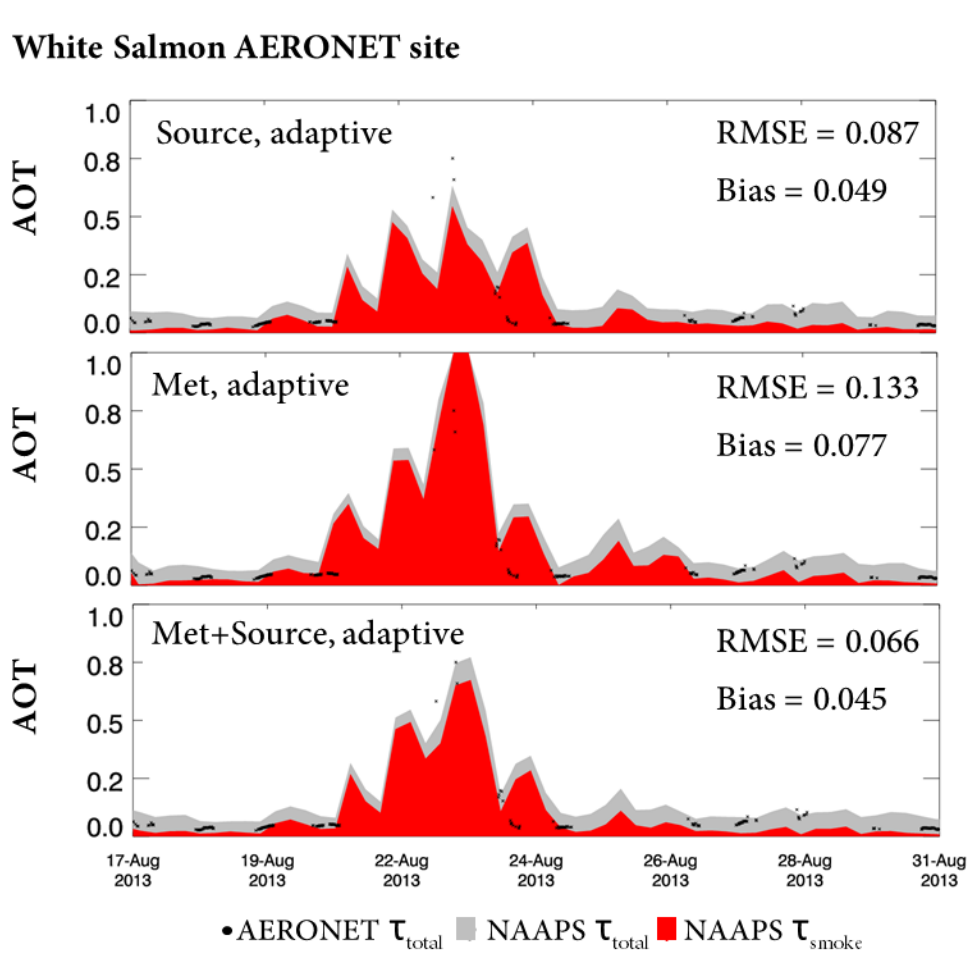

Figure 11. Timeseries of analysis total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the White Salmon AERONET site in the Western United States. Results are shown for adaptive inflation experiments with source only ensemble, NOGAPS meteorology ensemble, and a combined meteorology and source ensemble.

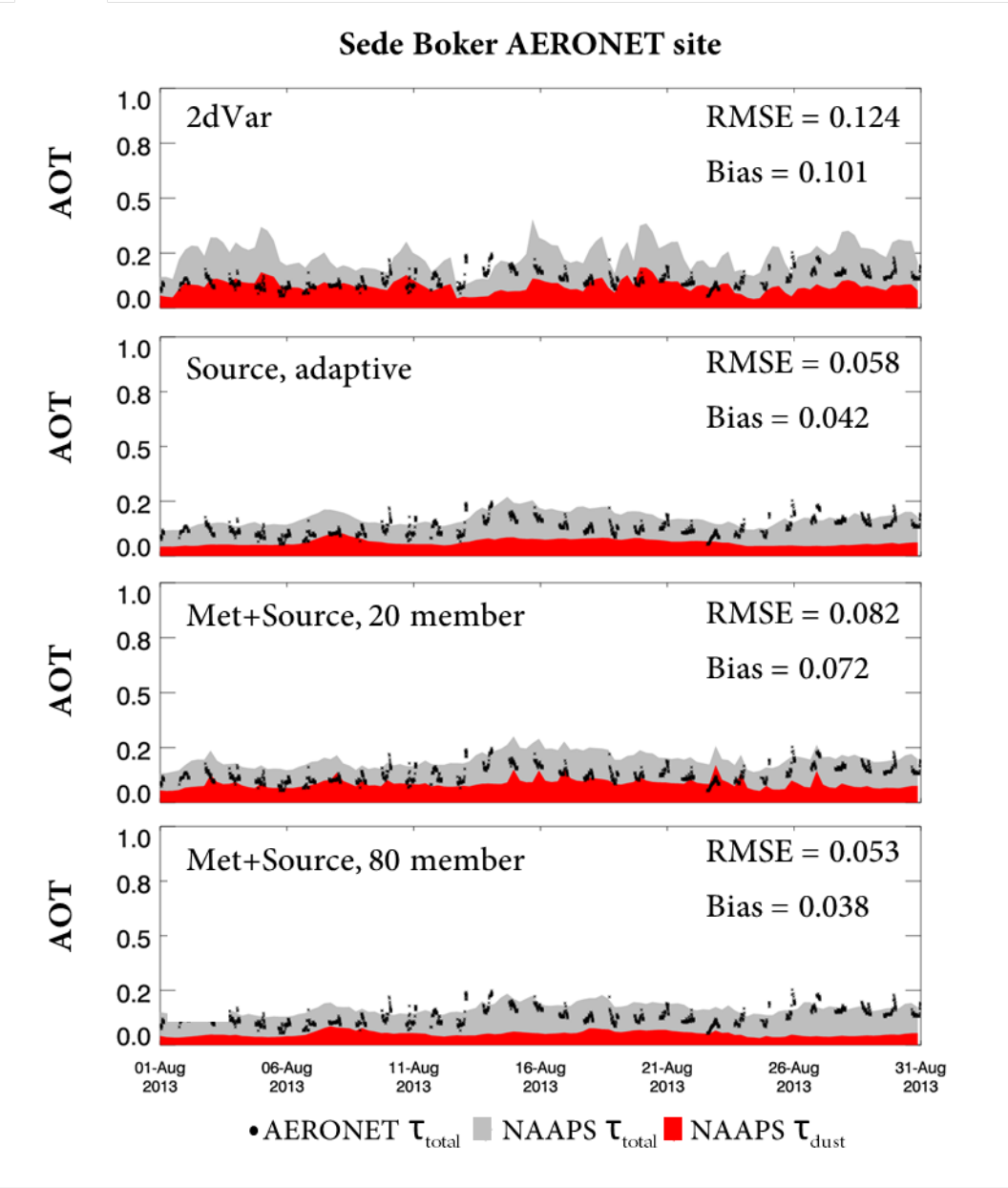

Figure 12. Timeseries of analysis total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the Sede Boker AERONET site, a Mediterranean site in the Negev Desert. Results are shown for the NAVDAS-AOD 2dVar data assimilation as well as the ENAAPS-DART for the source only ensemble and the combined source and meteorology ensemble with 20 and 80 ensemble members. RMSE and bias relative to AERONET AOT are included.

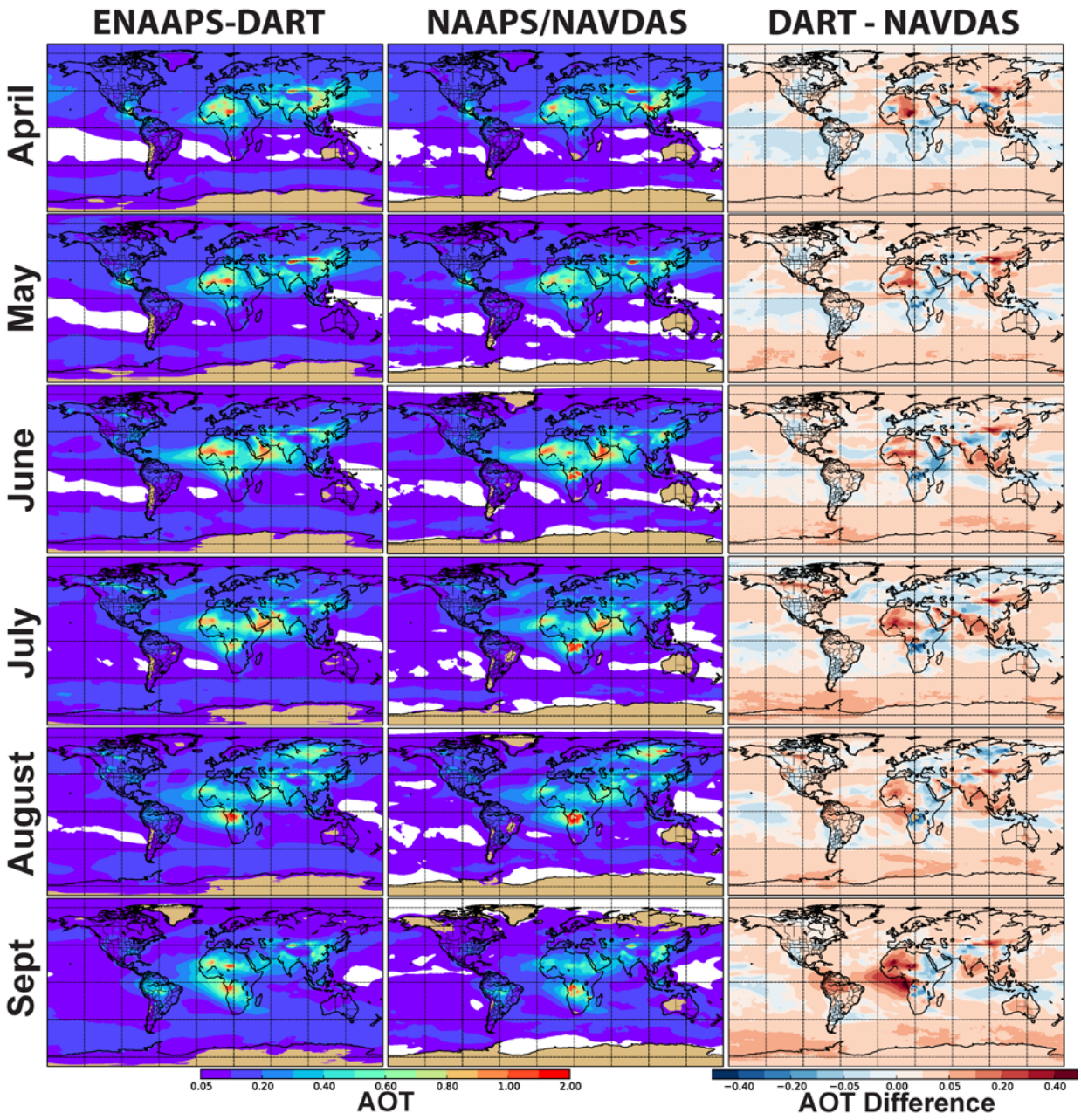

Figure 13 Monthly averaged AOT fields (550nm) from the ENAAPS-DART system and the NAAPS/NAVDAS-AOD system. Also shown is the monthly averaged AOT difference between ENAAPS-DART and NAAPS/NAVDAS-AOD.

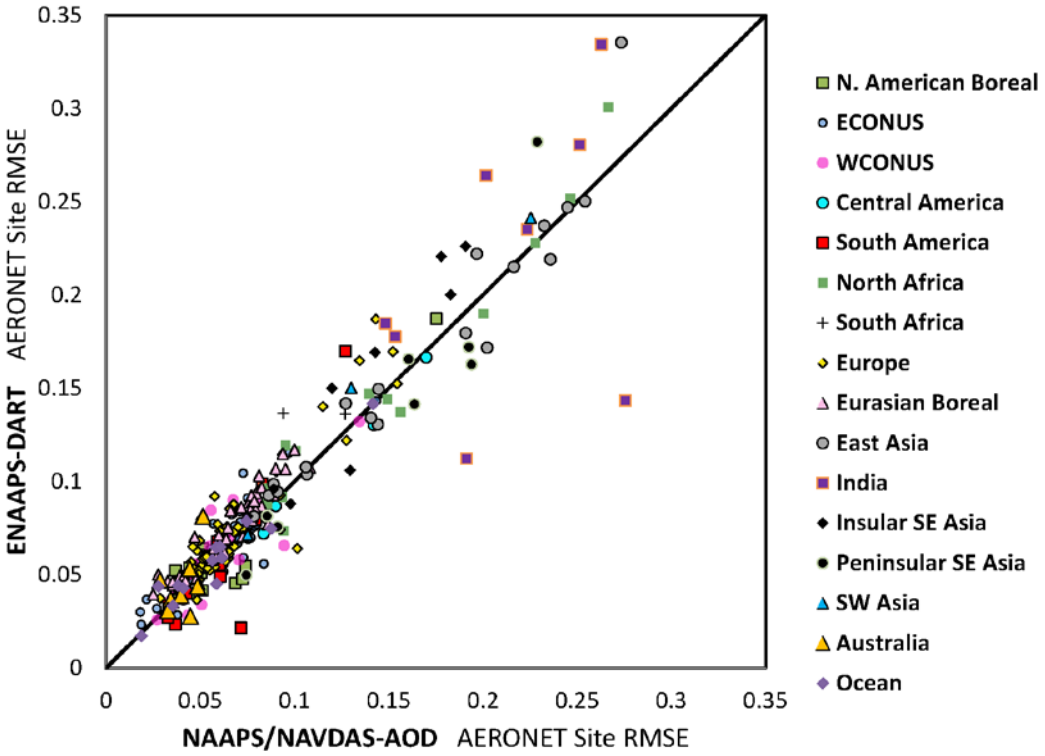

Figure 14. Comparison of AERONET site RMSE (AOT, 550nm) between ENAAPS-DART AOT analysis fields and NAAPS/NAVDAS-AOD analysis fields for simulations run over a six month time period (April through September, 2013). Sites are identified by region.

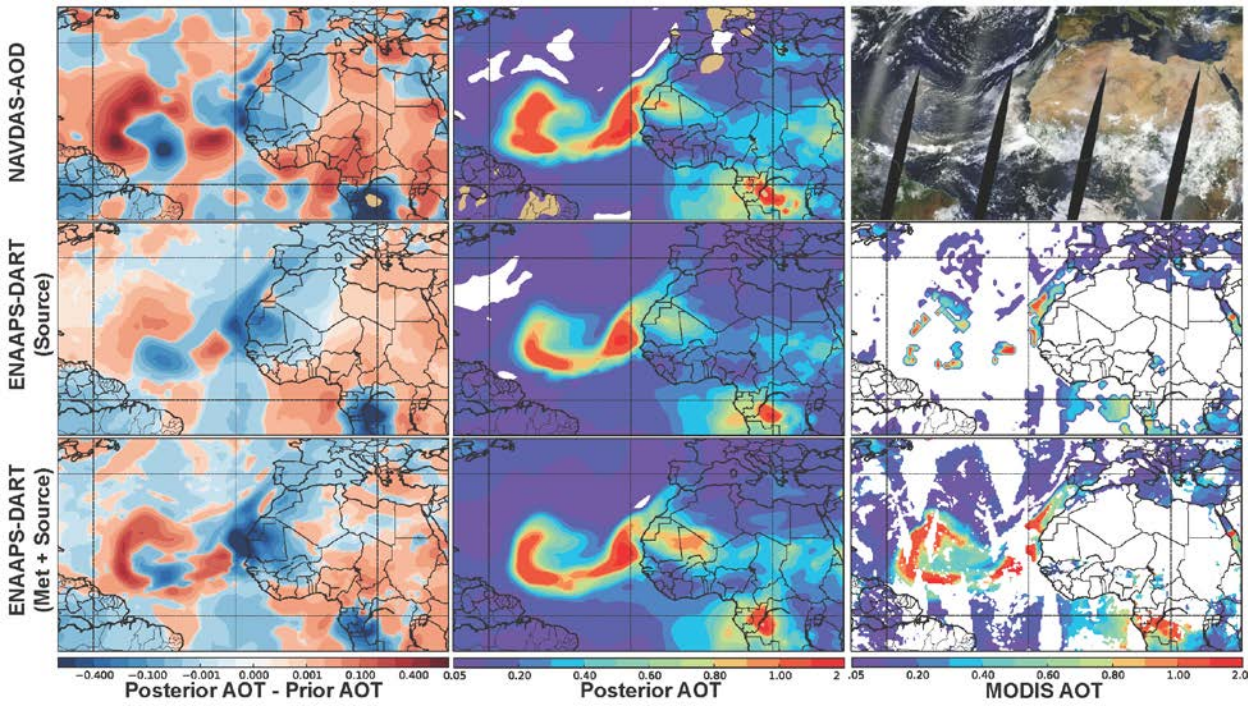

Figure 15. An example dust transport case off the coast of West Africa (August 1, 2013). Analysis increments (posterior AOT-Prior AOT) and posterior AOT (550nm) are shown for the variational NAVDAS-AOD (first row), EAKF for ENAAPS-DART with source ensemble only and adaptive inflation (second row), and EAKF for ENAAPS-DART with the combined meteorology and source ensemble and adaptive inflation (third row). Also shown are MODIS observations in the third column, including a MODIS visible image of the dust event (top), a plot of assimilated MODIS AOT observations (middle), and a plot of all Terra and Aqua MODIS AOT (550nm) observations for the event (bottom).

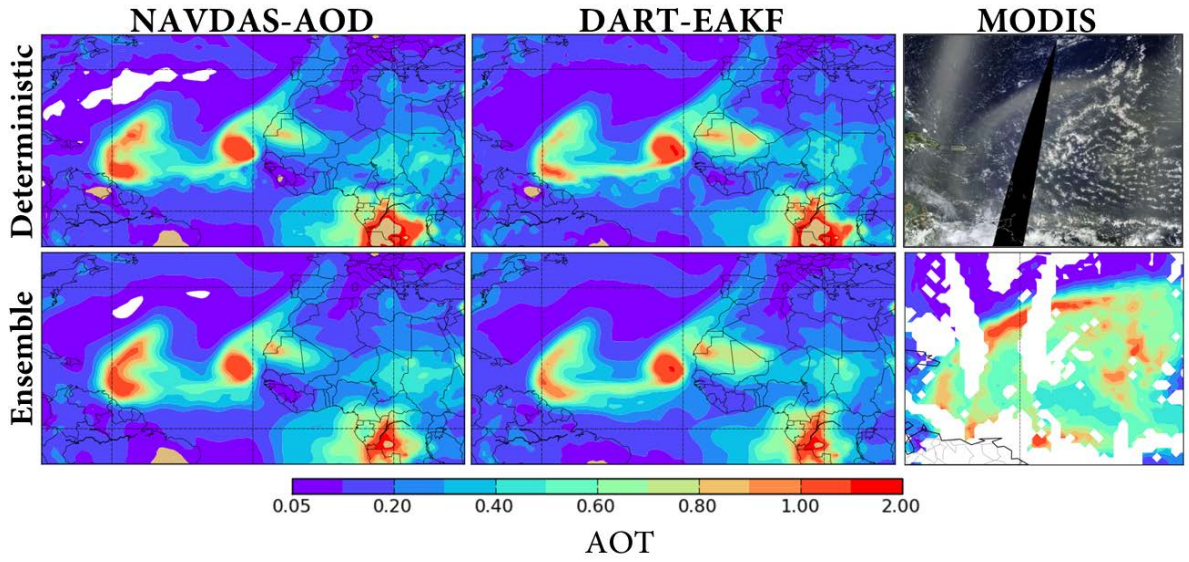

Figure 16. Example dust transport case off the coast of West Africa, initialized with analysis fields from Figure 15, and forecasted out to 24 hours. AOT (550nm) results are shown for four different forecast configurations: a deterministic forecast initialized with NAVDAS-AOD fields (2dVAR); a deterministic forecast initialized with DART-EAKF fields (ensemble mean); an ensemble forecast initialized with NAVDAS-AOD fields; an ensemble forecast initialized with DART-EAKF fields. A zoomed in MODIS true color image of the leading edge of the dust plume is also shown as well as MODIS AOT (550nm) observations.