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 ~~produces~~ 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 ~~clearly~~ able to better capture sharp gradients in aerosol features compared to the ~~variational~~ 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 ~~which~~ 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). ~~In order to increase understanding~~ As an initial exploration of aerosol forecast uncertainty and ~~aerosol forecasting~~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 ~~and what can be learned in~~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 ~~also~~ investigate the use of ENAAPS for operational aerosol forecasting ~~ensemble forecasting system for operational aerosol data assimilation~~ 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 th~~eis exercise~~ 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 ~~produced by~~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, ~~noting~~ including a few key differences between the NAAPS implementation used in this work and the literature~~a few key differences in the NAAPS implementation~~. 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 ~~this global reanalysis~~simulations conducted in this work, a ~~MODIS only~~ 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. ~~Likewise, the~~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 ~~the model~~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 ~~involves~~ a 2-dimensional variational (2dVAR) data assimilation system (NAVDAS-AOD, Zhang et al. 2008; 2014) ~~which~~ for ~~incorporates~~ assimilating AOT retrievals (Zhang et al. 2005; Zhang and Reid, 2006, 2009; Hyer et al. 2011; Shi et al. 2011) to ~~produce~~ 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). ~~Twenty NOGAPS members were produced operationally at FNMOC at half degree resolution out to six days.~~ 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. ~~The NOGAPS ensemble members are produced using perturbed initial conditions~~

Formatted: Font: Not Bold

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Normal

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

~~based on the local ensemble transform methodology of McLay et al. (2010).~~ 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~~l~~ ~~in the operational arena~~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. ~~requires a priori assumptions about the 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 ~~produced~~ 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~~a~~). 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:

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

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.

~~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). 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 2009b) and localization for spatially limiting the impact of an observation (Hamill et al. 2001, Houtekamer and Mitchell, 2001).~~

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~~b~~) which ~~produces~~ 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. ~~requires an additional assimilation step with an inflation factor associated with each element of the model state vector.~~ 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~~b~~) 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 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. While tuning methods such as covariance inflation and localization have been shown to be important for overcoming sampling error, the method for generating the ensembles themselves is an important consideration for an optimal aerosol forecasting system. 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:~~

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

~~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. Meteorology ensembles are evaluated in addition to the source draws, using the 20-member NOGAPS ensemble.~~ 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 ~~predictions~~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, ~~it is important to evaluate diagnostics~~the diagnostics are evaluated regionally. The experimental 6-hour AOT forecasts are evaluated over ~~13~~ 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. T~~In addition, t~~he 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 ~~performance~~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.

Formatted: Left

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

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 ~~produced~~ 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 ~~incorporate~~ 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 ~~produced~~ generated from the EAKF and 2DV~~v~~ar 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 ~~produced~~ 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 ~~clearly~~ 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 ~~produce~~create a stable system. As previously discussed, large persistent Siberian fires ~~produced~~ 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~~b~~). 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 ~~prior~~ inflation factors before each assimilation cycle. In ~~the~~ 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. ~~With the 0.9 damping factor, the prior inflation is assumed to be 90% of the posterior.~~ 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 ~~produce~~ 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 ~~normalized~~ 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 ~~produced~~ 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 ~~a~~ 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 ~~produces~~ 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 ~~produces~~ 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 ~~produced~~created, ~~creating~~ 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 ~~produced~~ 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. ~~The result~~It indicates that the observational error is ~~likely~~ 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, ~~an observational error that is too large~~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. ~~Likewise~~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 ~~may be~~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 ~~produced~~ 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~~n't~~ ~~fully representing the distribution~~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 ~~produces a good representation of the 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 ~~generated~~ 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~~performance gains~~ can be ~~made~~ 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 ~~produced~~ 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 ~~produced by~~ 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 ~~produces AOT fields~~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.~~Large differences in the Southern Ocean are attributed to differences in the ensemble and deterministic meteorology since there are few observations to assimilate in that region.~~ 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

Formatted: Font: (Default) Times New Roman, 12 pt

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 ~~produced a slight reduction in RMSE~~ 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 ~~produced~~ had a slightly smaller or the same RMSE as the EAKF with the largest difference in RMSE (0.016) ~~produced~~ 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~~variational~~ 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 ~~produced~~ from~~by~~ 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 ~~produced~~ 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 ~~variational~~ 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 ~~produced~~ had a smaller RMSE are also impacted by long-range transport events. EAKF appears to have an edge over the ~~variational~~ 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 ~~examining~~ an examination of the posterior fields ~~produced~~ 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 ~~clearly~~ 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 ~~optimal~~ 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~~ness~~ of the dust front that is ~~clearly~~ also seen in the MODIS image for this event (Figure 15). On the other hand, the ~~variational~~ 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 ~~produced~~ 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~~ness~~ as seen in the MODIS image. ~~However, the ensemble version of the 2dVAR forecast is smoother than the deterministic counterpart.~~ 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~~ with the ENAAPS-DART version being smoother than the deterministic counterpart along the dust front~~. This demonstrates that the sharp gradient~~ness~~ achieved in the ensemble data assimilation propagates in the forecast. This ~~and~~ is an advantage of using the EAKF initial conditions over the ~~variational~~ 2DVar initial conditions for the short-term forecast.

~~4.0 Discussion~~

~~The optimization of the EAKF data assimilation from DART for use with the ensemble version of NAAPS revealed several interesting insights about ensemble data assimilation for aerosol prediction. With respect to the ensemble, having both meteorology ensembles as well as perturbations to the aerosol source functions produce the best results. This is due to a combination of the meteorology ensembles being important for long range transport events, as demonstrated by the dust transport examples shown in the results section, and the source ensemble being important for large local aerosol sources with spatial correlations, as demonstrated by several smoke aerosol cases. There are caveats to this when dealing with~~

~~aerosol sources for the same species with behavior that is spatially independent. This is believed to be the case for fires in the North American boreal region during the experimental summer 2013 time period. The application of the source perturbation to fires in this region creates spatial correlations, due to the manner in which the perturbations were applied, that are not real if the fires are behaving independently. This can be tested by applying source perturbations that are not spatially correlated in this region and allow the remaining fires to be perturbed as usual. Likewise, performance issues were identified for the EAKF in the Eastern United States and Europe. This may be a result of pollution sources of aerosol, such as point sources, which would have independent behavior. Further investigation is needed to understand how to properly capture the ensemble statistics in regions dominated by independent aerosol sources in order to improve performance.~~

~~The evaluation of the ensemble diagnostics for the ENAAPS-DART optimization also highlighted some potential issues with the smoke emissions used in the simulations. In particular, regions dominated by smoke aerosol did not have sufficient spread at high AOT values, 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 with respect to smoke aerosol. The smoke emissions used in the simulations are based on MODIS. Smoke emissions are highly uncertain, often having several factors of uncertainty, which could be contributing to the observed bias. It is also known that remote sensing algorithms have difficulty in detecting small fires without a large enough thermal signal (Schroeder et al. 2008), and therefore, smoke aerosol from small fires may be underrepresented. The analysis of total spread to RMSE for smoke dominated regions indicated that the observational error may be too large for small AOT values, which could also contribute to the positive bias observed in smoke regions. Additional tuning of the smoke sources or including the aerosol sources as part of the state to be estimated by the data assimilation may be a means for overcoming this type of bias in the smoke emissions.~~

~~For overcoming sampling errors inherent in ensemble data assimilation, both spatially and temporally constant multiplicative covariance inflation and an adaptive covariance inflation algorithm were tested. Although the constant covariance inflation has been used in past applications of ensemble data assimilation for aerosol prediction, the results in this study show that this is not the best approach. The constant inflation produced an unstable system in regions without good observational coverage. This result is likely applicable to any data assimilation problem where the observation density is not spatially uniform (i.e. other atmospheric tracers). An adaptive inflation method from Anderson (2009b) was tested for the first time, to our knowledge, for an aerosol application. The results showed that the adaptive inflation increases ensemble spread only where observational information is available, preventing the issue seen with the constant inflation. The ENAAPS-DART experiments using adaptive inflation had stable ensemble spread with time, an indicator of a healthy ensemble system. This is the recommended inflation method for aerosol and potentially other atmospheric tracers.~~

~~Relative to the current operational 2dVAR data assimilation system, the EAKF produced analysis fields that had similar results in regional RMSE, bias, and R^2 against AERONET AOT. However, differences were more apparent at individual AERONET sites. The EAKF outperformed the variational assimilation at sites that were impacted by long-range transport, including several Eastern US, Europe, and Central America sites. This is not unexpected given that the EAKF uses flow-dependent covariances. Additionally, the source ensemble provides an advantage of producing more structured ensemble covariances for regions impacted by large aerosol sources that are spatially correlated. This provides an advantage over the 2DVar, particularly at several dust-impacted sites located in North Africa and the Mediterranean region. On the other hand, the univariate adjustment by the variational data assimilation performed better in regions where the sources behave independently, as was seen at several European, Eastern US, and North American Boreal sites. Further investigation is needed to understand how to better characterize statistics for regions impacted by independent sources in order to push the EAKF ahead of the 2dVAR for these types of regions. While the EAKF and 2DVAR were both capable of capturing aerosol features, reflected by the similarity in regional statistics, the EAKF provided an advantage in being able to better capture events spatially. This was demonstrated for a dust transport case off of the coast of Western Africa. By using the ensemble statistics to spread observational information, the EAKF is able to capture sharp gradients that are smoothed out in univariate assimilation methods, effectively reducing true model resolution. This provides a particularly important advantage to the EAKF, especially when moving to higher resolution simulations. Based on these results, the EAKF should be able to take advantage of resolution increases while the 2dVAR may smooth out any resolution advantage.~~

~~Forecasts out to 24 hours were conducted using the initial conditions from the DART and the NAVDAS-AOD data assimilation, in a deterministic and an ensemble configuration. The forecasts initialized with EAKF initial conditions had smaller RMSE, although not statistically different, in the 24 hour forecast than their variational counterparts. Also, the ensemble configurations had smaller RMSE relative to the deterministic configurations. 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 moved towards ensemble prediction systems and aerosol forecasting should move in the same direction. In order to evaluate the spatial impact of the different forecast configurations, the 24 hour forecast of the same Saharan dust transport case used to evaluate the analysis fields was examined. With the DART-EAKF initial conditions, the sharpness of the dust feature is predicted and even more so in the ENAAPS-DART configuration. The findings from this study show that an ensemble prediction system, including an EAKF data assimilation for producing initial conditions combined with a probabilistic forecast, demonstrate an advantage over the current operational deterministic system with a univariate variational data assimilation architecture. 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.~~

4~~5~~.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 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, ~~An~~ 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 .
- ~~Performance gains~~A reduction in RMSE can be achieved by increasing the ensemble size from 20 to 80 members. Further ~~gains~~ reductions may be achieved with optimization of the 80 member ensemble (ie. localization ~~and inflation~~).
- 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 t~~T~~he ensemble system underrepresents uncertainty for large smoke events ~~and has some positive bias relative to MODIS AOT observations~~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,

Formatted: Font: (Default) Times New Roman, 12 pt

the ensemble spread for smoke aerosol is likely too small at low AOT values. ~~Additional tuning~~ 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 ~~produced~~ from the DART-EAKF data assimilation system and the NAVDAS-AOD ~~2dVAR~~ 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 a~~ major~~n 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~~a~~.

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

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 thickness 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.

Formatted: Font: Italic

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.

Formatted: Space After: 9.1 pt, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Pattern: Clear (White)

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 ~~produced by~~from the ~~variational~~ 2DVar NAVDAS-AOD assimilation system and the EAKF data assimilation from ENAAPS-DART.

	~~Variational~~ 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

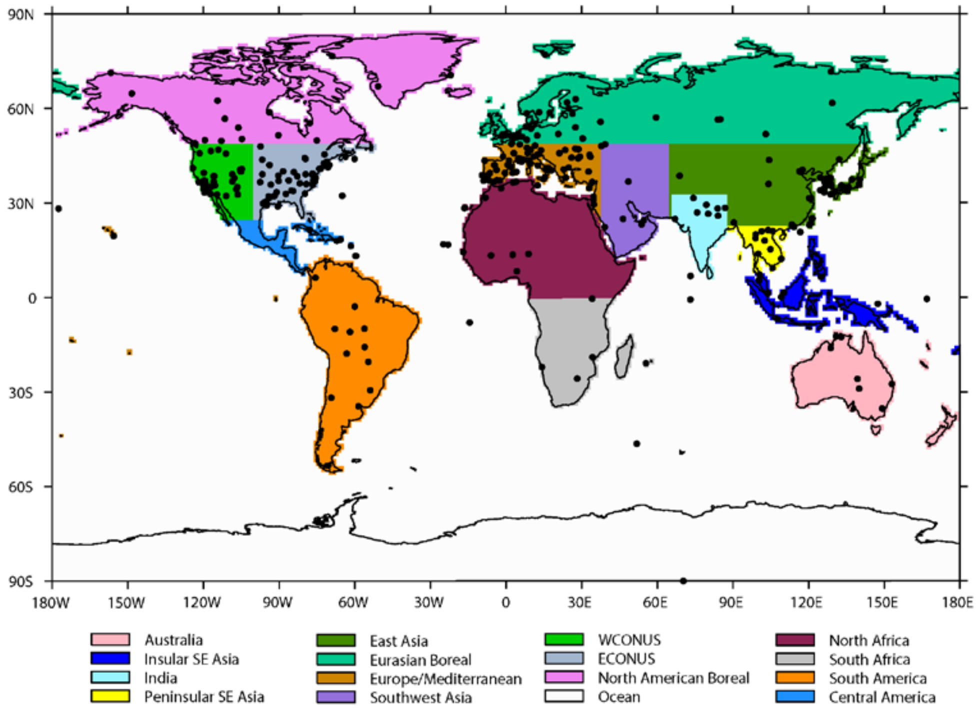


Figure 1. Diagnostic regions for evaluated ENAAPS-DART experiments. Black dots indicate AERONET sites with data available for 2013.

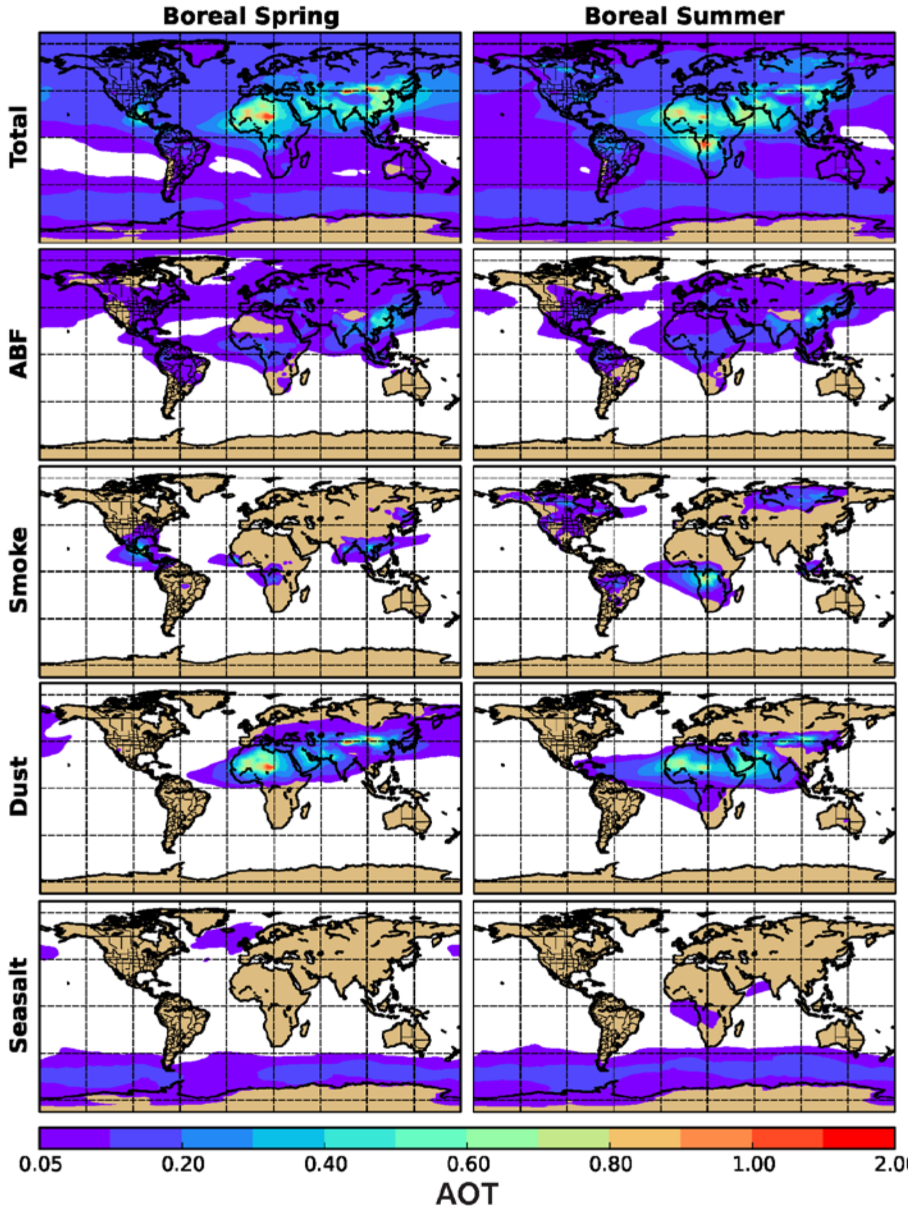


Figure 2. Seasonally averaged AOT (550nm) fields (posterior), predicted by the ENAAPS-DART system (Met+Source, adaptive), for the Boreal Spring (April, May) and Summer (June-September), 2013. Results are shown for total AOT and AOT attributed to combined anthropogenic and biogenic fine (ABF), smoke,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

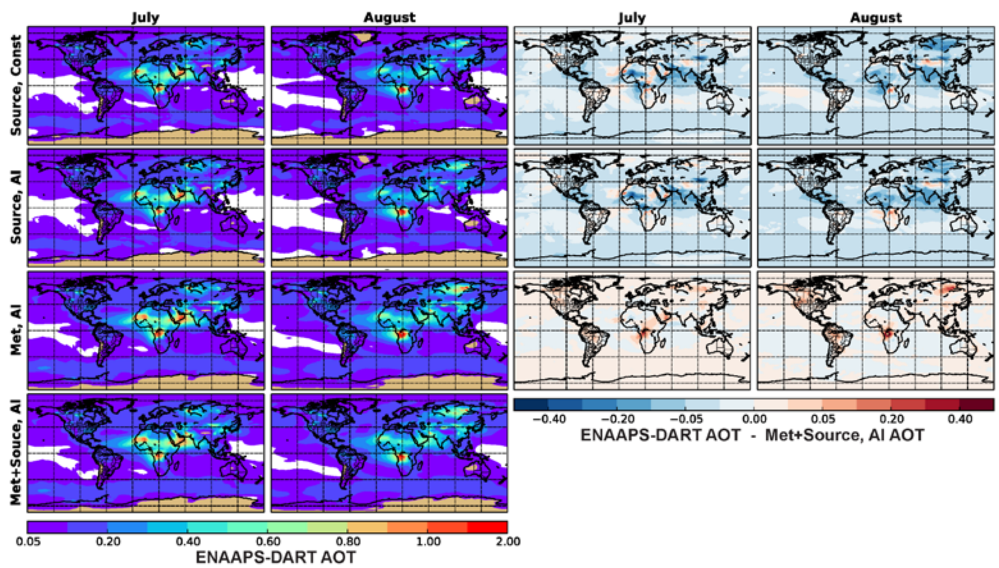
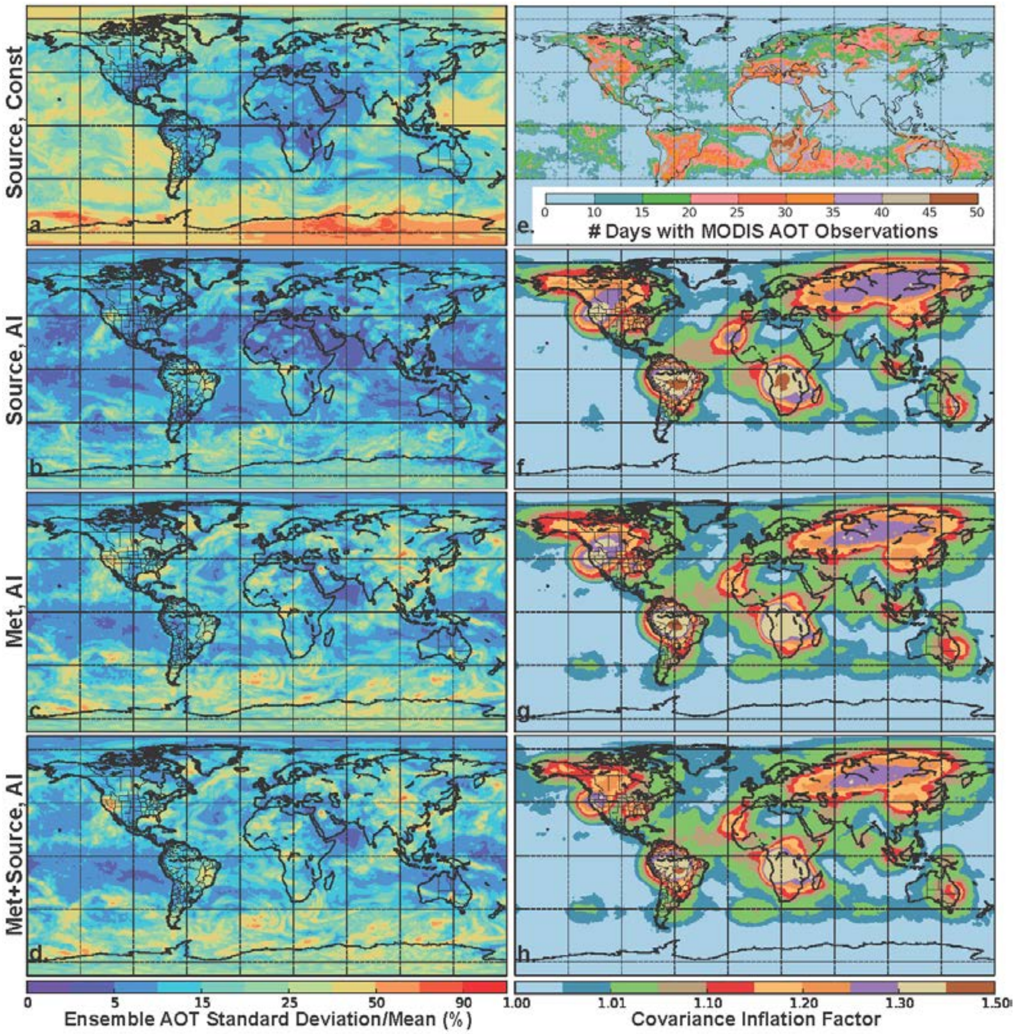


Figure 3. Monthly averaged AOT (550nm) for four ENAAPS-DART EAKF optimization experiments, including a source ensemble with constant inflation (Source, Const), a source ensemble with adaptive inflation (Source, AI), a meteorology ensemble with adaptive inflation (Met, AI), and a combined meteorology 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).



Source, Const
Source, AI
Met, AI
Met+Source, AI
a.
b.
c.
d.
e.
f.
g.
h.
0 10 15 20 25 30 35 40 45 50
Days with MODIS AOT Observations
0 5 15 25 50 90
Ensemble AOT Standard Deviation/Mean (%)
1.00 1.01 1.10 1.20 1.30 1.50
Covariance Inflation Factor

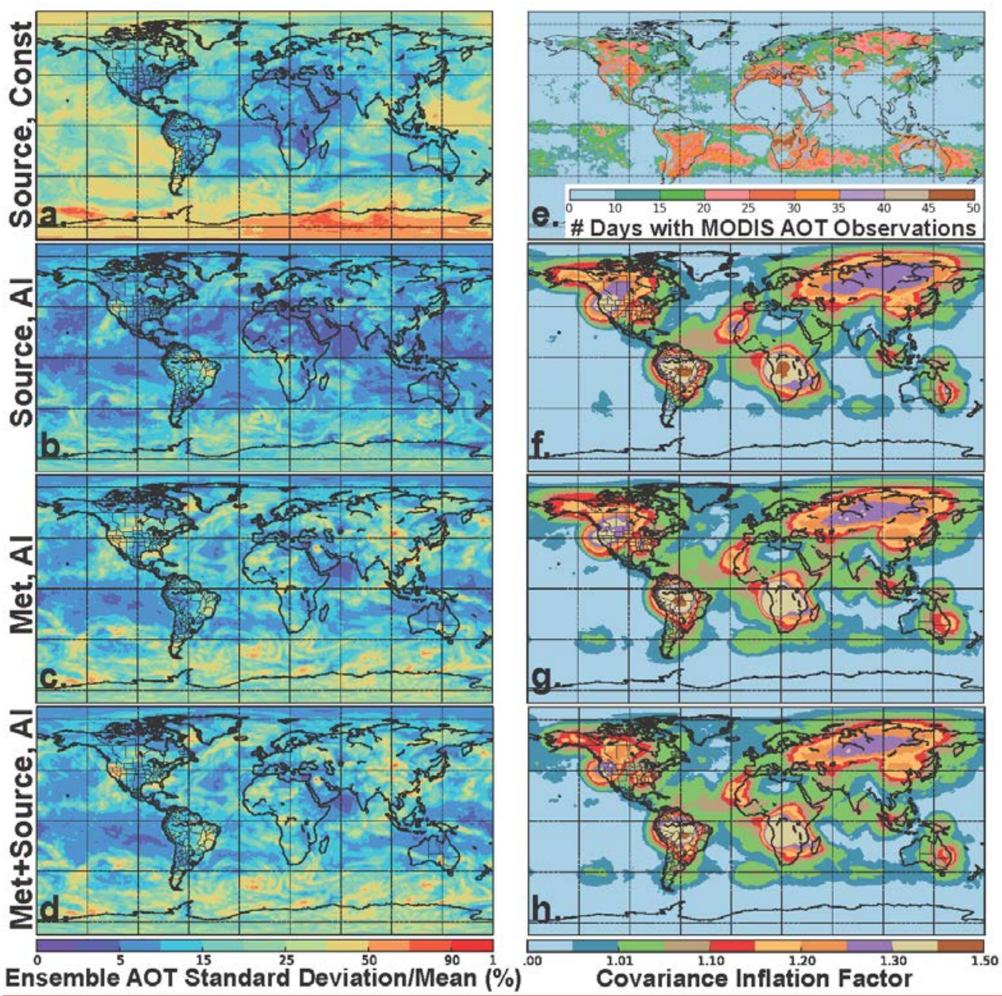


Figure 4. The standard deviation of the prior ensemble aerosol optical thickness normalized by the ensemble mean at the end of the experimental time period (August 31st, 1800) for four ENAAPS-DART experiments: a) source only ensemble with spatially and temporally constant 10 percent covariance inflation 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.

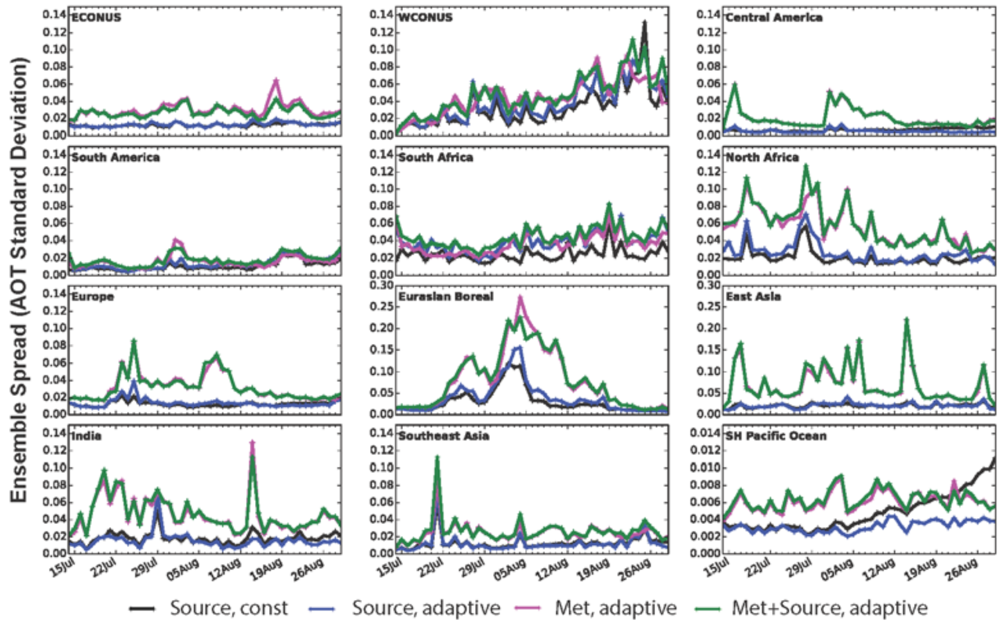
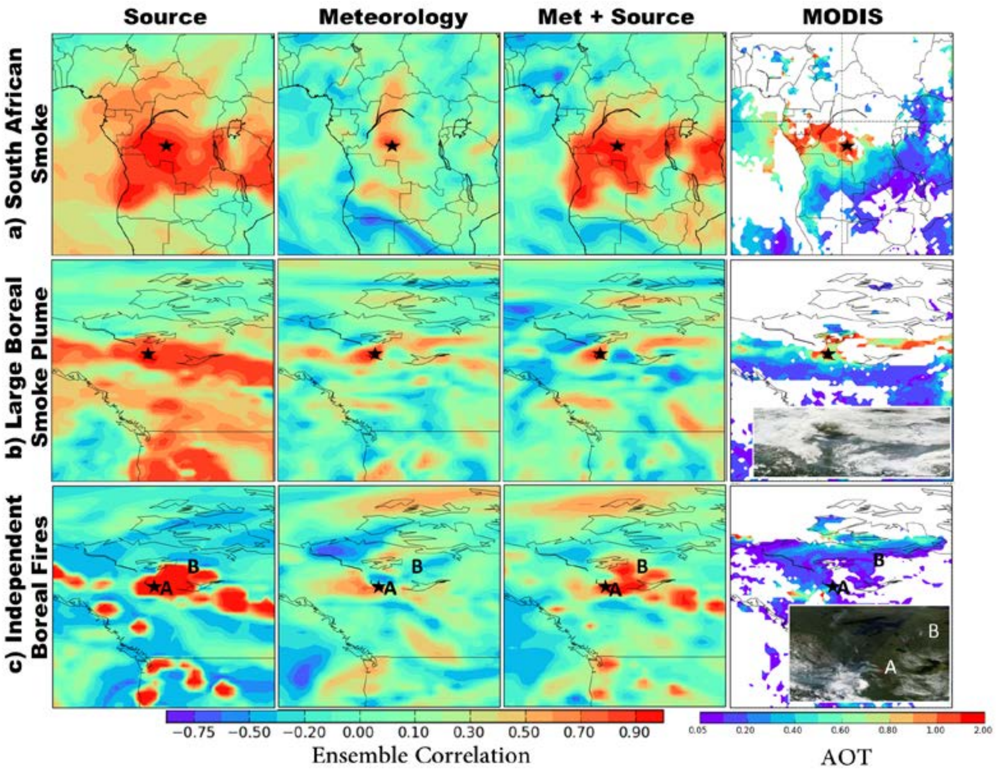


Figure 5. Timeseries of ensemble spread (AOT standard deviation) for 4 ENAAPS-DART experiments over the July 15 through August, 2013 time period. Results are shown for 12 regions, including the Eastern United States, the Western US, Central America, South America , South Africa, North Africa, Europe,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

1


Source
Meteorology
Met + Source
MODIS
a) South African Smoke
b) Large Boreal Smoke Plume
c) Independent Boreal Fires
A
B
−0.75 −0.50 −0.20 0.00 0.10 0.30 0.50 0.70 0.90
Ensemble Correlation
AOT

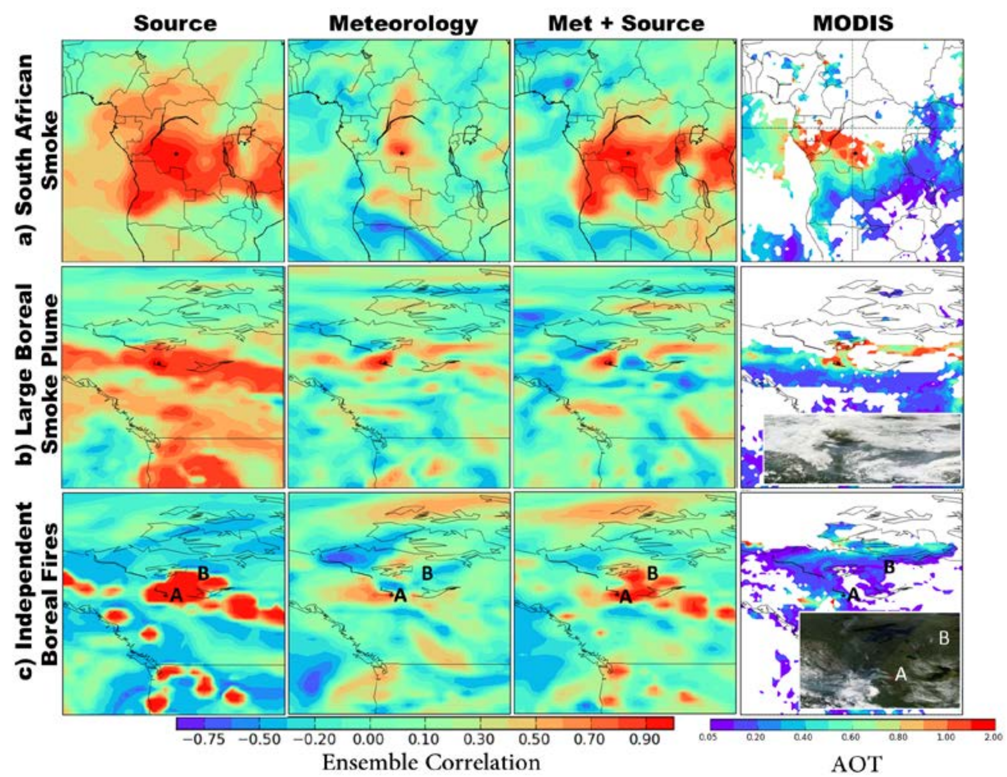


Figure 6. Ensemble correlation fields in the prior AOT relative to a point indicated by a black star for three different aerosol events: a) a South African smoke event on August 2, 2013 b) a large North American Boreal smoke plume on August 15, 2013 and c) small independent Boreal fires in North America 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.

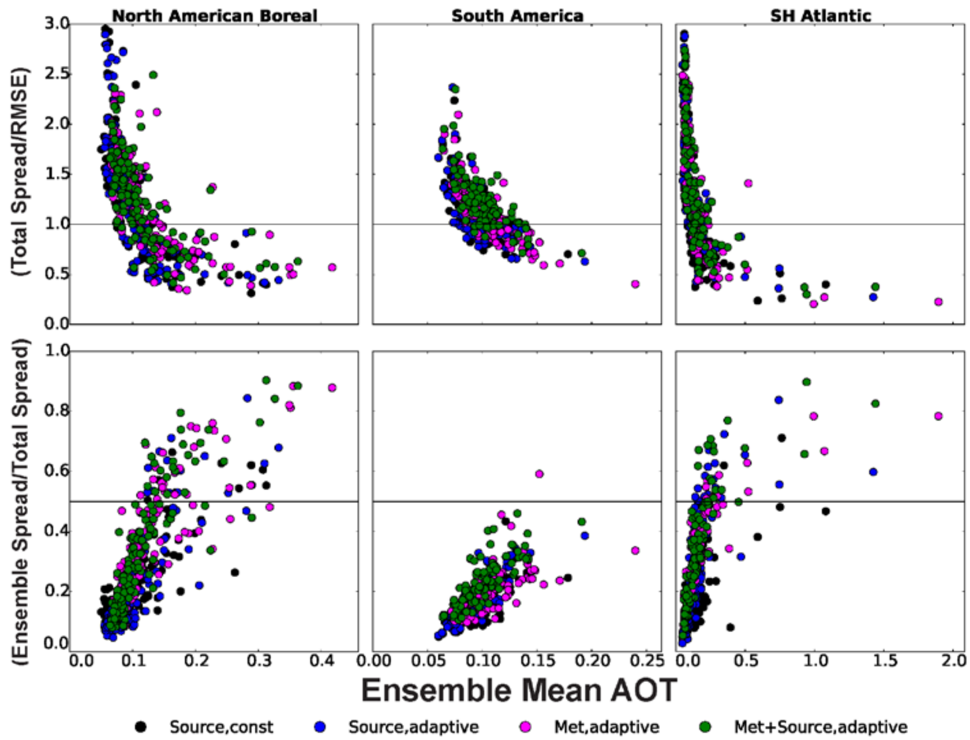


Figure 7. Regional scatterplots of the ratio of total spread (combined ensemble AOT spread and MODIS AOT error) to RMSE against the ensemble mean AOT (550nm) (top row) and the ratio of ensemble spread to total spread against the mean AOT (550nm) (bottom row). Results are shown for four ENAAPS-DART configurations 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.

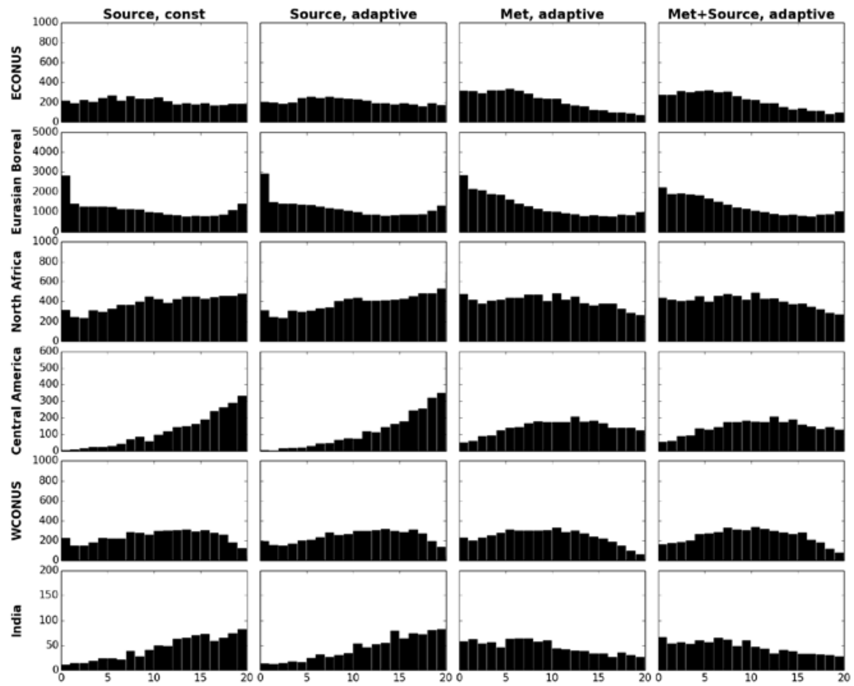


Figure 8. Rank histograms for select regions for the four ENAAPS-DART experiments, including source only ensemble with constant and adaptive inflation (Source, const; Source, adaptive), meteorology only ensemble with adaptive inflation (Met, adaptive), and meteorology plus source ensemble with adaptive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

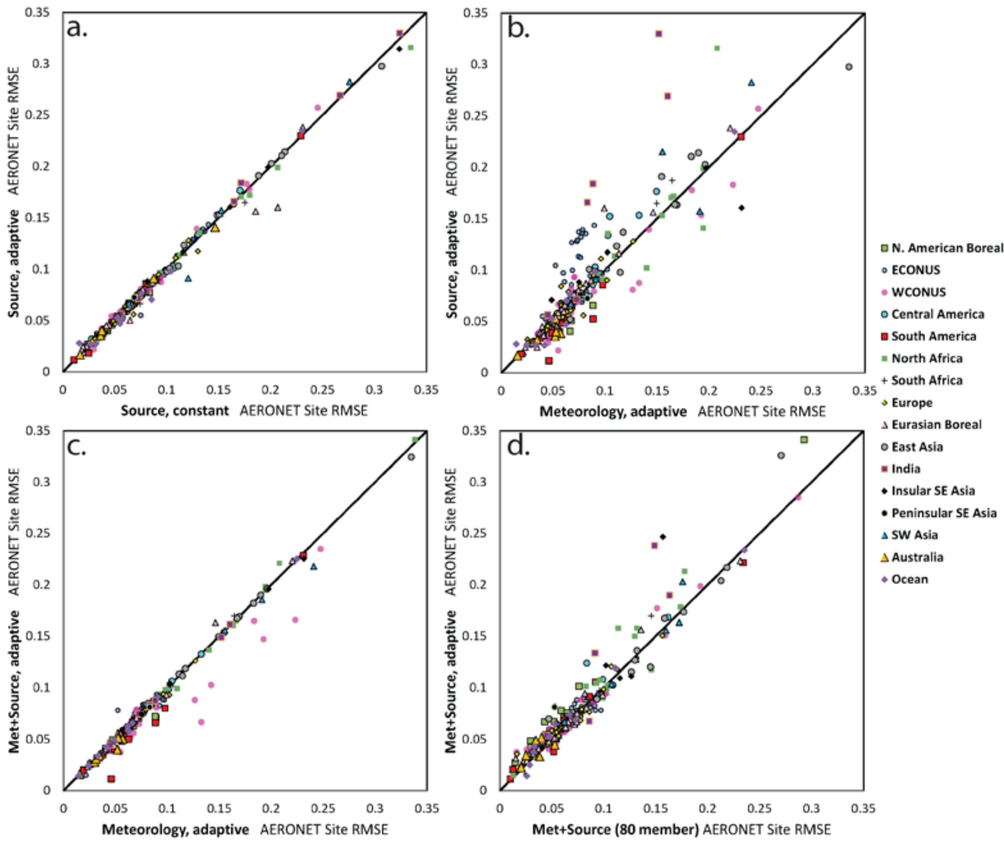


Figure 9. Scatterplots of ENAAPS-DART RMSE relative to AERONET AOT (550nm, Zhang and Reid, 2006) by site between different ENAAPS-DART experiments. Sites are identified by region. Results are shown for a) source only with constant covariance inflation versus adaptive inflation b) meteorology only versus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

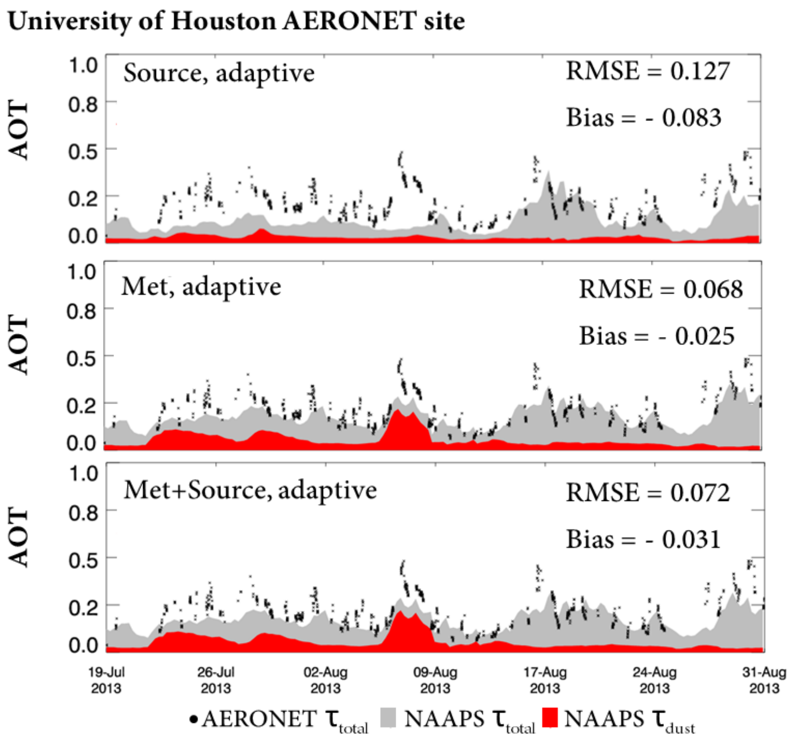


Figure 10. Timeseries of model predicted total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the University of Houston AERONET site. Results are shown for adaptive inflation experiments with source only ensemble, NOGAPS meteorology ensemble, and a combined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

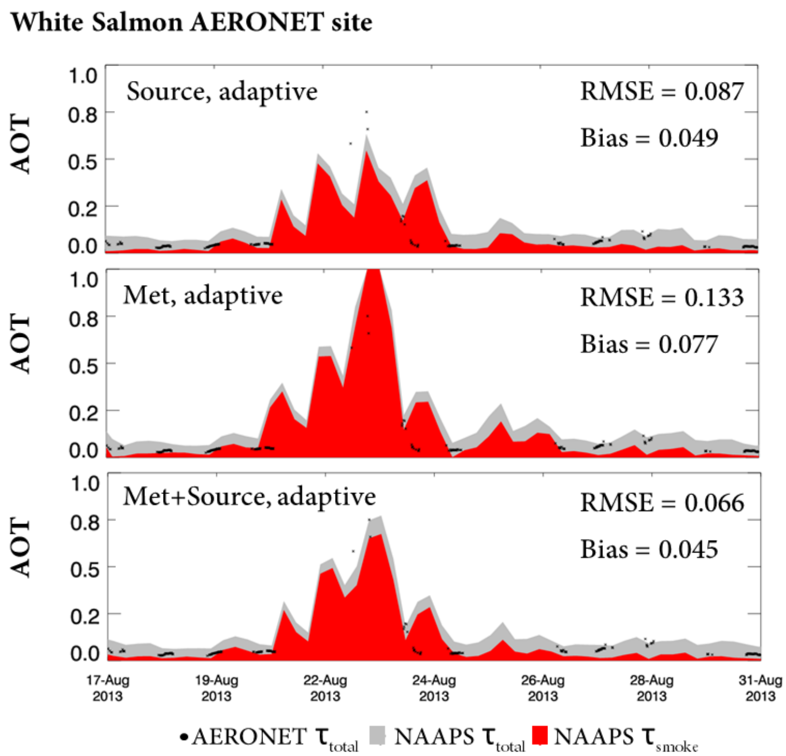


Figure 11. Timeseries of analysis ~~model predicted~~ total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the White Salmon AERONET site in the Western United States. Results are shown for adaptive inflation experiments with source only ensemble, NOGAPS meteorology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

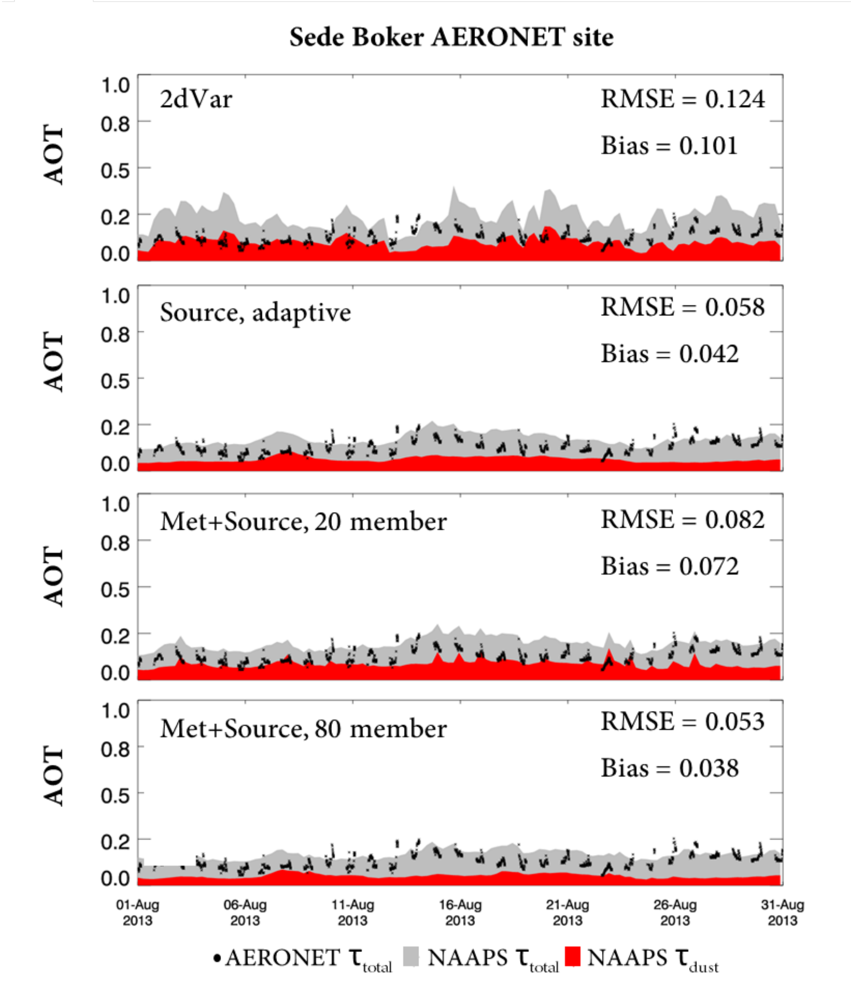


Figure 12. Timeseries of ~~model predicted~~analysis total AOT (grey) and dust AOT (red) with AERONET AOT (Zhang and Reid, 2006) (black) at 550nm at the Sede Boker AERONET site, a Mediterranean site in the Negev Desert. Results are shown for the NAVDAS-AOD 2dVar data assimilation as well as the ENAAPS-DART 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.

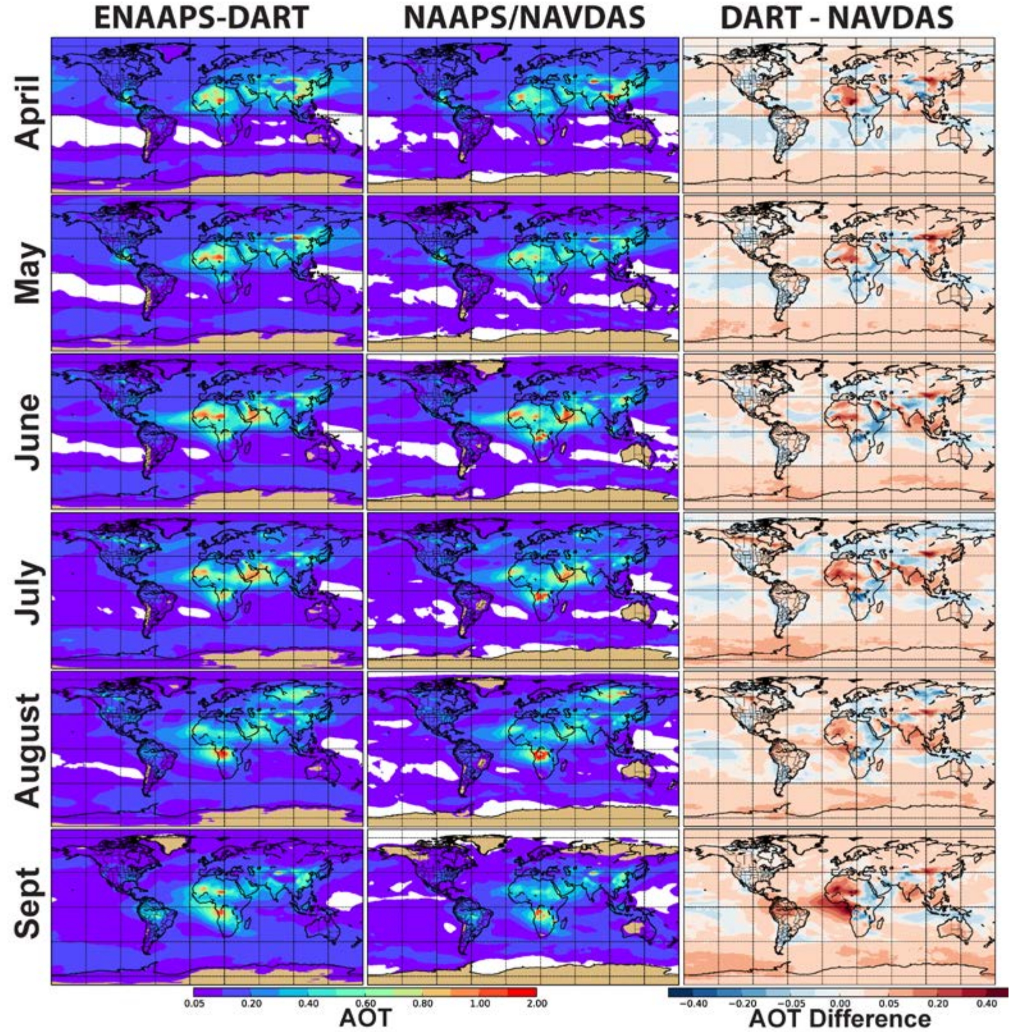


Figure 13 Monthly averaged AOT fields (550nm) from the ENAAPS-DART system and the NAAPS/NAVDAS-AOD system. Also shown is the monthly averaged AOT difference between ENAAPS-DART and NAAPS/NAVDAS-AOD.

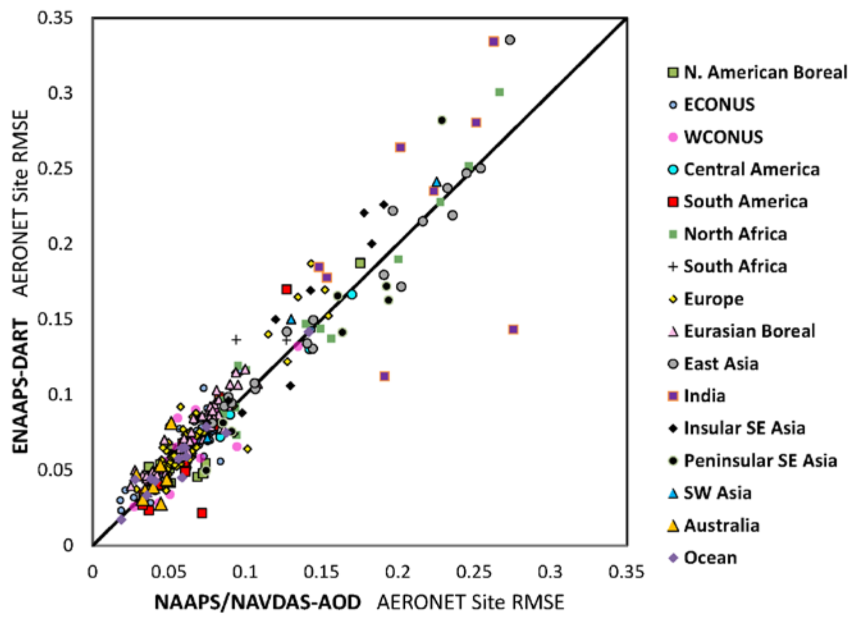


Figure 14. Comparison of AERONET site RMSE (AOT, 550nm) between ENAAPS-DART AOT analysis fields and NAAPS/NAVDAS-AOD analysis fields for simulations run over a six month time period (April through September, 2013). Sites are identified by region.

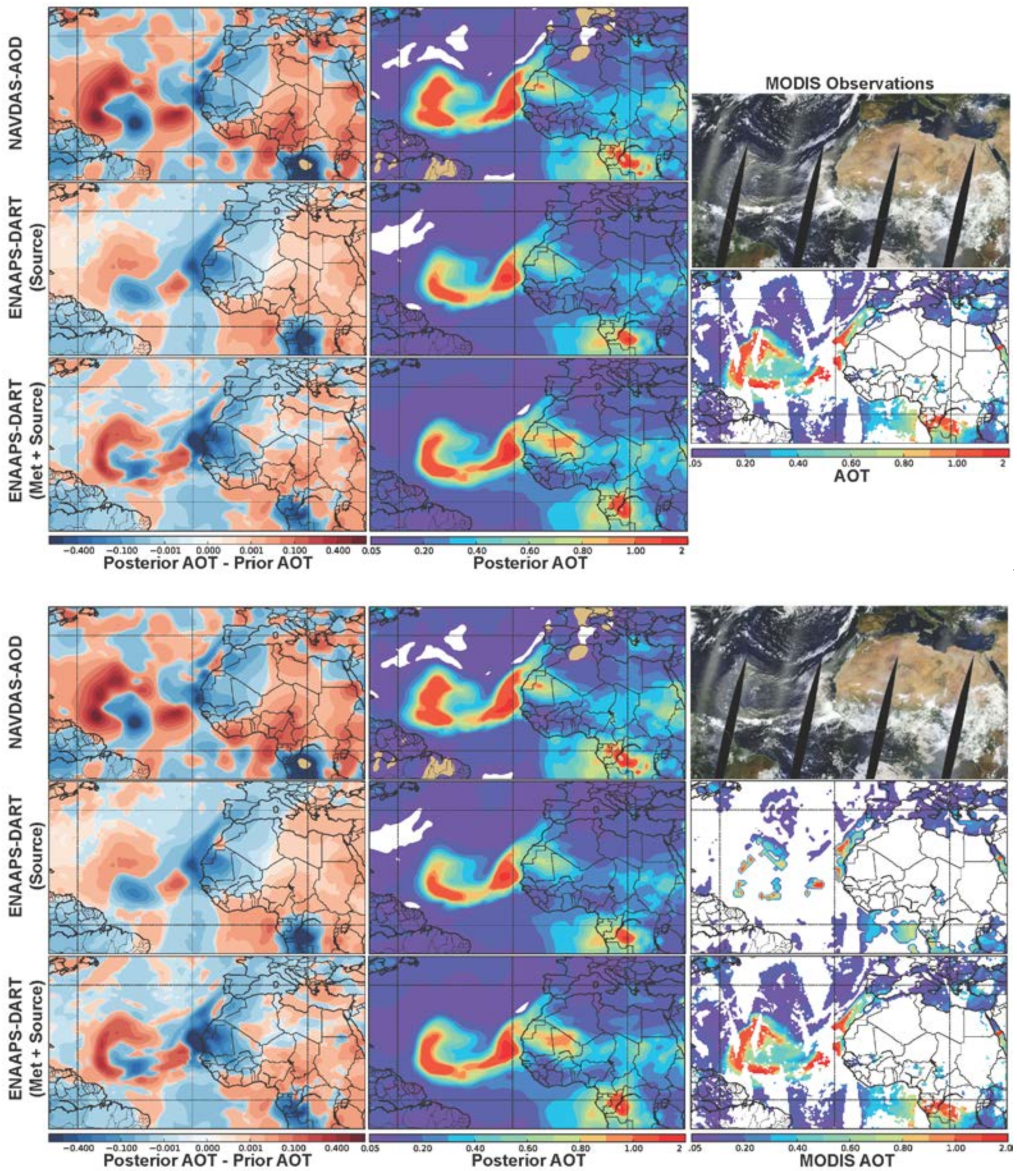


Figure 15. An example dust transport case off the coast of West Africa (August 1, 2013). Analysis increments (posterior AOT-Prior AOT) and posterior AOT (550nm) are shown for the variational NAVDAS-AOD (first row), EAKF for ENAAPS-DART with source ensemble only and adaptive inflation (second row), and 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 ~~available~~ Terra and Aqua MODIS AOT (550nm) observations for the event ~~from Terra and Aqua~~(bottom)~~, below~~. ~~It should be noted that not all available MODIS AOT observations are assimilated.~~

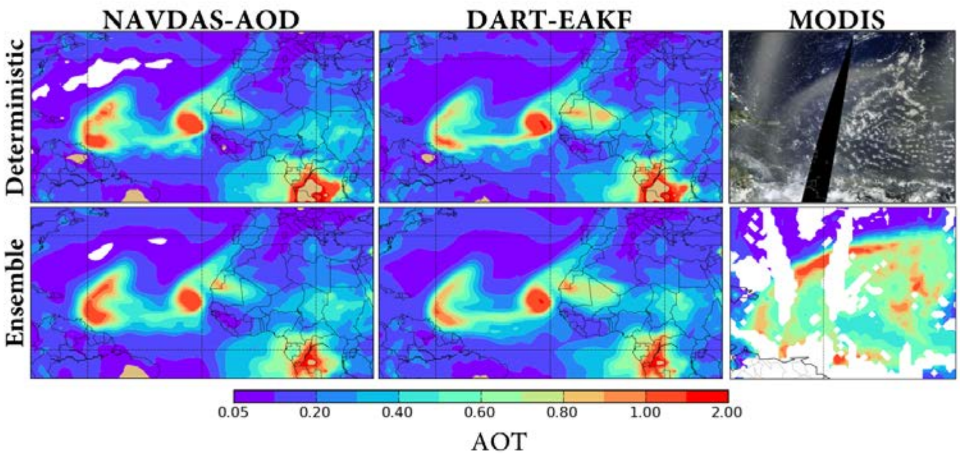


Figure 16. Example dust transport case off the coast of West Africa, initialized with analysis fields from Figure 15, and forecasted out to 24 hours. AOT (550nm) results are shown for four different forecast configurations: a deterministic forecast initialized with NAVDAS-AOD fields (2dVAR); a deterministic 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.

Response to Anonymous Referee #1

Thank you for your thorough comments. Please see our responses:

1. First, no results are shown for a standard run of the model without assimilation. Hence the improvement due to assimilation is unknown and the differences between various assimilation setups cannot be properly judged.

   Response: The goal of our study was to see how the new ensemble system performs relative to the current operational prediction system (NAAPS with NAVDAS-AOD); as a result, this was considered our control. (page 28075, lines 5-8)"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." Subsequent papers will show more detailed comparison of the different methods relative to a no DA control.

2. Second, no proper attempt at filter tuning is done. In particular, ensemble size and localisation length-scale are not systematically varied and their effects studied. In this respect Fig 9 is slightly worrying: panel d (which shows differences between a 20 and 80 member run) shows similar or larger differences than the sensitivity experiments for a 20-member ensemble (a,b and c).

   Response: There was a lot of tuning that went into setting up ENAAPS-DART, we will work to make this point more clear in the manuscript. With regards to localization, several tests were run, but the results were not presented in the paper. This is discussed on page 28078 (lines 26-28)-28079 (lines 1-2). What we found with regards to localization tests was that the 1000km lengthscale performed the best. Since these results were consistent with previously published studies (Schutgens et al. 2010), we didn't feel showing these additional tests introduced anything new and would just add to an already long paper. Instead, we wanted to focus on the experiments that introduced something new to aerosol data assimilation. That is why we chose to focus on looking at constant versus adaptive inflation as well as looking at methods for generating the ensemble members. We felt these experiments were both informative and introduced something new. With regards to ensemble size, we chose 20 members because this is the size that is run operationally out to 6 days, and hence is our basis set. There are some resource limitations in place for running a system operationally that we cannot control. However, we wanted to show one test of what an increase in ensemble size could buy, and a limited time period enhanced run was acquired (single 80 member ensemble run, page 28095 (lines 1-23)). As we expected, you can get a big payoff with increasing ensemble size because we are likely doing a lot better in capturing a realistic background error covariance. I'm unsure of why the scatterplot for 20 versus 80 members is troubling, except that it shows there is a lot of room for improvement in future development of this system. With our resources, the 20 member system will serve as the base system with potential for moving to larger ensemble sizes in the future. As we mentioned in the paper, we have plans for future studies in ensemble size as well as model resolution (page 28095, lines 18-24; page 28105, lines 23-26). It should be noted that the optimization tests were conducted on the 20 member ensemble, and therefore, things like localization are not necessarily optimal for the 80 member (less localization would be needed for the larger ensemble size). We don't have the resources for these optimization runs at the moment, but expect to do more work on this in the future.

Manuscript changes: To make it clear that the optimization tests were conducted on the 20-member ensemble, we made changes to the following sentences:

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).

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.

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.

3. The authors at times generalize too much from their own (limited set of) experiments: while the possible problem due to constant inflation is worth mention and analysis, no other authors have come across this and it is possible this is entirely due to very a specific system (ENAAPS-DART).

Response: We would argue that this finding is most likely not system specific. For idealized experiments and NWP applications, similar findings with regards to constant and a varying inflation were identified. This was mentioned in the manuscript on page 28087 (lines 6-9).

While this has never been directly discussed for aerosol applications, there have been hints to this issue. For example, Schutgens et al. (2010) ran sensitivity studies for a one month simulation (July 2005) for aerosol assimilation of AOT. One of the sensitivity experiments conducted was varying the inflation factor for a constant multiplicative inflation. They found instabilities developing for an inflation factor of 1.20 and 1.30 where unrealistic aerosol mass mixing ratios developed. This result was for a short one-month simulation, so the instability can be seen for large inflation factors. We suspect that if the simulation was run out for a longer time period, issues would have developed for smaller constant inflation factors as well. However, we will change the strength of the wording to indicate that we suspect this result is applicable to other systems.

Manuscript changes: **Based on the results in this work,** an 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.**

4. The relative importance of source vs meteorology perturbation is hard to assess given that source perturbations are always generated with a 25% spread. This uncertainty seems optimistic at hourly and gridbox scales.

Response: We agree that the 25% uncertainty applied to the source perturbations might be optimistic as we know emissions can be highly uncertain, especially for boreal fires. However, the system behavior indicates that regardless of the perturbation applied, the spatial impact (or lack therefore) of the data assimilation using only a source-perturbed ensemble would be the same. By perturbing the sources for smoke as an example, the impact on the system is to create large correlations at all distances between smoke emissions, only limited by the localization lengthscale. While increasing the source-perturbations would increase the size of the analysis increment, it wouldn't impact the area of influence (ie. near source regions). The same problem of not being able to impact aerosol transport events (away from source-regions) as discussed in the manuscript would hold for a source-only ensemble. While we know that some changes need to be made to how the source perturbations are generated as discussed in the manuscript (page 28091, lines 13-14; page 28092, lines 13-14;page 28104, lines 16-17), our conclusion of needing both source perturbations for data assimilation near-source regions and meteorology ensemble for transport events would hold.

5. Sometimes there are quite lengthy descriptions of results, region by region, while the same results are efficiently summarised in Figures and Tables. Maybe the authors can try to make their text more concise

   Response: Thank you, we will work to make the text more concise.

6. Apparently inconsistent acronyms: AOT and NAVDAS-AOD

   Response: Aerosol optical thickness (AOT) is the more appropriate term to use for aerosol extinction in the vertical, therefore, we choose to use AOT instead of AOD throughout the manuscript. Since its development, the variational data assimilation system used with NAAPS (ie. NAVDAS-AOD) has always been referred to in this manner (Zhang et al. 2008); therefore, we choose not to change the legacy name of this system.

7. The paper by Schwartz et al JGR 2014 deserves mention as it also compares 3D-VAR and ensemble Kalman filter methods for aerosol assimilation.

   Response: We agree and will add this reference to our manuscript.

   Manuscript changes: 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).

8. Introduction: a major advantage of ensemble DA systems over others is the relative ease of implementation and maintenance, especially in view of the fact that many aerosol and aerosol-cloud processes can be modelled in different ways

   Response: Thank you, we will add this point to the introduction.

   Manuscript changes: 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).

9. p 28073, l 13: "In order to increase understanding of forecast uncertainty and aerosol forecasting dependencies on underlying meteorology, a 1 resolution, 20 member ensemble version of NAAPS (ENAAPS) was created". The exact meaning eludes me. Does this refer to a one-off experiment or is it an on-going activity? What was learned from this?

   Response: As an initial exploration of forecast uncertainty, an ensemble version of NAAPS driven purely 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 and were available on the NRL aerosol webpage. However, we wanted to take full advantage of the ensemble and set up ENAAPS forecasts to be initialized with analysis field from an ensemble data assimilation system, the focus of this work. We will clarify this point in the introduction. Thanks.

   Manuscript change:) **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.**

10. p 28074, l 17: "a brief synopsis is provided here, noting a few key differences". While I agree with this level of detail, I think the text might be clearer in specifying what are the differences. E.g. "Likewise, the sea salt source is dynamic in nature with emissions as a function of surface wind speed (Witek et al., 2007)." suggests there are no differences wrt seasalt so why mention it? It doesn’t help that a brief (and necessary) explanation of basic aerosol description is interjected ("A combined anthropogenic and biogenic fine aerosol species (ABF) is represented in the model which accounts for a combined sulfate, primary organic aerosol and a first order approximation of secondary organic aerosol."). I suggest to reorganise this in two paragraphs: the first a very brief overview of essential NAAPS characteristics (e.g. basic aerosol description + emission datasets and parametrisations), the second the key differences of the version used in this paper

    Response: Thank you for your feedback on this. We will edit the description of NAAPS and ENAAPS to make it clearer.

    Manuscript changes: 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 produce forecast initial conditions every 6 hours.

11. What is meant by a MODIS-only version? FLAMBE is completely ignored? Or only MODIS data are used for a specific FLAMBE version?

    Response: Here we are using a version of FLAMBE that only uses MODIS data. We use this version of FLAMBE as it is used in the NAAPS decadal reanalysis which serves as an internal benchmark. We will clarify this point in the description.

    Manuscript change: 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).**

12. ENAAPS is in principle independent of (aerosol) assimilation, no? So the "exception of data assimilation" is a bit confusing. The distinction between 'deterministic' and 'ensemble' meteorology fields is also confusing. I'm guessing this is in-house jargon? The ensemble meteorology fields are also the result of deterministic models. How is this ensemble produced (e.g. what is perturbed, a very brief description of McLay et al would be good)? What does "truncated to 1 degree" mean (is NOGAPS a spectral grid model)? Why match the deterministic (!) NAAPS reanalysis? It will beused with ENAAPS, not?

    Response: Yes, ENAAPS can be independent from data assimilation. Here we are referring to the ENAAPS-DART system versus the NAAPS/NAVDAS-AOD system. We will change ENAAPS to ENAAPS-DART and NAAPS to NAAPS/NAVDAS-AOD to make this point clear. With respect to the meteorology fields, we are referring to a single set of meteorology fields produced from the deterministic model for the 'deterministic' fields. For the ensemble meteorology fields, these are produced using an ensemble transform to perturb the analysis fields (wind, temperature, specific humidity, and surface pressure) as discussed in McLay et al. (2010). Yes, NOGAPS is a spectral model with a higher resolution than ENAAPS, therefore, the NOGAPS output is truncated to produce a one-degree resolution output for the ENAAPS simulations. We chose to match the 1 degree resolution used here in the ENAAPS-DART base system with the NAAPS reanalysis to have aerosol product lines that can be easily compared. However, as mentioned in the manuscript, we plan to do additional studies on model resolution.

    Manuscript changes: 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 produced operationally at FNMOC at 0.5 degree resolution out to six days. These fields are produced 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 ensemble members are used **for driving the model simulations**, truncated to 1 degree to match the deterministic NAAPS reanalysis **(Lynch et al. 2015)**.

13. "requires a priori assumptions". It can be argued that ensemble DA methods also require a-priori assumptions on the model forecast error, in that they assume a-priori uncertainties in meteorlogy and emissions and from that calculate the ensemble forecast.

    Response: Yes, this is true that there are assumptions in ensemble data assimilation about Gaussian distributions etc. This is certainly a limitation of ensemble data assimilation. Here we are trying to make the point that the error covariance is produced a priori and is static. While the ensemble covariance will of course not be perfect, it provides a means for allowing the uncertainty to vary with time and with processes that occur in the model simulations. We will be more specific in making this point. Thank you.

    Manuscript changes: The variational approach, which is used in the current NAVDAS-AOD system, **uses a static model forecast error**.

14. p 28076, l 4: "is considered to be a random draw from the probability distribution of the model's state given all previously used observations." This sentence completely ignores a-priori error sources in the ensemble, even though they are the essence of the system

    Response: This is the premise of ensemble prediction systems and the formulation of EnKF is based on this principle. While the analytical theory is based on this, ensemble DA systems have been found to work well even when these assumptions are violated. In particular, ensembles have been found to work well with heavily biased model forecasts when using the adaptive inflation (Anderson 2009).

15. p 28076, l 5: "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" Shouldn't there be a comma after 'in time'?

    Response: Thank you, we will add a comma.

16. p 28076, l 17: It is not entirely clear how EAKF and DART relate? EAKF is part of DART, and I think it is the only ensemble DA in DART. What does DART offer beyond EAKF?

    Response: EAKF is one of the filter options available in DART. There are several different filter types including an EnKF, Kernel filter, Particle filter and several other options as described in DART documentation. We will make this point more clearly in the text.

Manuscript change: 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**.

17. p 28076, l 20-25: Apparently DART does not include an observation operator H, but uses ENAAPS calculations of AOT. As AOT will depend on humidity (which will be different in different ENAAPS members), doesn't this imply that the effective observation operator used in DART is non-linear instead of the linear operator assumed in a Kalman filter? (That is: across the ensemble, AOT cannot be generated from a form like H x, with x the aerosol state vector and H a matrix).

    Response: This is up to the person implementing DART on whether they want to use an observation operator that acts on the state variables as they are read into DART or as done here, apply an observation operator outside of DART. Yes, there are nonlinearities due to humidity, which does vary between ensemble members. This is always an issue with data assimilation. However, DART applies forward operators sequentially, so arbitrary nonlinear h are trivial to implement.

18. p 28077, l 27: Why usually in the prior? Won't this distort any covariances that have been built up during the short-term forecast? Can't it be applied to the posterior? I thought that was the more common way to use inflation.

    Response: Priors that are unrealistically confident result in the observations having insufficient weight in the data assimilation update and over time, lead to filter divergence. Because of this, the covariance inflation is typically applied to the prior (Anderson and Anderson, 1999) and this is especially the case for EnKF systems for weather prediction. However, the inflation can be applied to the posterior as well (ie. Whitaker and Hamill, 2012). The inflation increases the spread about the mean, so it doesn't impact the sample correlations between components.

19. p 28079, l 3-5: "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." This suggests that the biggest issue is to have as large a spread as possible. I would argue instead that the spread should be an indication of forecast uncertainty (both know uncertainties, ie meteorology and emissions and unknown uncertainties, e.g. due to model errors).

    Response: Note that we aren't saying that we want the most spread possible here, we are saying there must be <u>sufficient</u> spread. This means we need adequate or enough spread (ie. to represent the uncertainty). Often times with ensembles, they are spread deficient which can lead to filter divergence and the observations won't have an impact. Here we are saying we want sufficient spread that represents the system. The adaptive inflation algorithm used in this work is designed to try and make the spread consistent with the RMSE as you suggest.

20. p 28079, l 5-15: Maybe the generation of the emission ensemble should be discussed before the inflation/localization? The latter are after all solutions to limitations in the first.

Response: We reordered this section to make it clearer. Thanks.

21. p 28079, l 13: Why 25% and not 10 or 100%? For sea-salt and dust, arguably perturbing emitted particle size/windspeeds can be just as important?

    Response: The impact of wind speed on sea salt and dust emissions is accounted for when the meteorology ensemble is used (ie. through differences in the wind fields across the ensemble members). While 25% uncertainty may be optimistic as discussed in a response to a previous comment, we thought this was a good first estimate of the source-perturbation. A means for evaluating if this is sufficient is to look at whether or not the system as a whole has enough spread. This is done in this work by evaluating how the pooled spread (combined ensemble spread and observational error) compare to the RMSE of the prior relative to the observations. These should be approximately the same if the system is well tuned. What we found is that the system was pretty well tuned with the exception of fire-impacted regions with not enough spread for high AOT events. This indicates we don't have enough spread and we need to potentially change how the fire emissions are represented in the ensemble (page 28101, lines 27-29). This could be done by increasing the source perturbations to the fire emissions (page 28092, lines 13-15). So in conclusion, we selected a conservative perturbation for the sources and based on the results from this study, have recommendations on how to move forward and improve the system.

22. p 28080, l 6: It would be good to have a brief explanation how rank histograms are created and what their purpose is? They are not a standard test in aerosol ensemble DA (but possibly should be).

    Response: Yes, we can add a few sentences to better explain the purpose of the rank histogram and how it is generated.

    Manuscript change: (page 28076, lines 34-36) 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).**

23. p 28080, l 8: Why is the prior a stronger indication of assimilation? I guess because they show how well a previous analysis pulled the system to the truth. An analysis will agree (fairly) well with observations by construction. Still, a bit more explanation or references are welcome. Do your data actually bear this out: i.e. does the prior show stronger signal to variation in experimental setup than the posterior? This would be very interesting to show.

    Response: It is much harder to compare MODIS AOT observations to the posterior AOT because they are no longer independent. It has been assimilated and therefore, you would expect better agreement. Here we are saying to use the 6-hour forecast AOT (ie. Prior) and compare that against MODIS AOT before assimilation. This gives us an indication if the model is doing a better job in predicting the state relative to the observations (before they are combined) and provides a means for evaluating how well the system is doing in representing forecast uncertainty. This is common practice in evaluating a data assimilation system. This section was updated to clarify the points being made.

Manuscript change: 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 predictions. 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 13 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**.

24. p 28081, l 8: Maybe change "incorporate" to "assimilate"?

    Response: Ok, thanks.

    Manuscript change: The NAAPS/NAVDAS-AOD simulations are run with a 1 degree resolution and **assimilate** the same MODIS AOT observational dataset for consistency.

25. p 28082, l 1: So which ENAAPS-DART assimilation experiment is shown here? What has been perturbed here? Has the system been optimised or not (inflation/localization)? What is the purpose of this Section? If it is to show global aerosol features, isn't this better shown during the comparison with NAAPS/NAVDAS? It might be clearer to first discuss the optimization experiments and only then discuss the global features seen in the best setup.

    Response: This result is for the meteorology and source perturbed ensemble with adaptive inflation. The purpose of this section was to present what aerosol features are being predicted during this time period so that they can be discussed in evaluating the system optimization as well as during the comparison between the deterministic and ensemble systems. We will work to make this clearer.

Manuscript change: 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.

26. p 28084, l 13: Why now the posterior AOT? Earlier you argued that the prior AOT should be used for comparison against observations.

    Response: Evaluating the prior is a good way to evaluate and diagnose the performance of the system relative to the observations that will be assimilated (MODIS AOT). This provides a means for evaluating how well we are doing in representing forecast uncertainty in our system and the overall health of the system. It is fair game to evaluate the posterior AOT against independent observations (AERONET) that are not assimilated. This is also an important evaluation since the posterior serves as initial conditions for our aerosol forecasts. We use both methods of evaluating the system performance in this work. This is discussed in the methods section on page 28080 (lines 1 through 21). The section describing diagnostics was updated as shown in response to comment 23 to clarify.

27. p 28084, l 17: Higher dust AOT is probably due to some higher windspeeds in the meteorology ensemble and the threshold windspeed for dust emission? What drives the increased AOT over wildfires?

    Response: Yes, the higher dust AOT is due to the introduction of different wind speeds across the ensemble members with the inclusion of the meteorology ensemble. For fire-impacted regions, the model generally produces a positive bias. With more spread in the simulations that include the meteorology ensemble, the observations have more weight in the analysis and the AOT is reduced.

28. p 28085, l 11: This is an interesting discussion of the role of inflation. It seems to me that the discrepancy between prior and observations is due to either: 1) observational biases; 2) model biases. A Kalman filter assumes that both are unbiased. Your results suggests that adaptive inflation serves to camouflage such biases (unless they become too big and the syetm crashes). This warrants some discussion by the authors.

    Response: Inflation is one of several means used to help overcome errors in ensemble systems. While it is one method for improving system performance, careful evaluation of how the algorithm behaves is also a means for better understanding the system and in ways that it can be improved. Case in point is the example you pointed out on page 28085, line 11. This is an issue that indicates a potential problem with the model as you suggested and in particular, fire-dominated regions. There were several issues related to smoke-dominated regions highlighted and the case is made throughout the manuscript (page 28092, lines 12-18; page 28096, lines 23-28) that issues in smoke-dominated regions indicate a need for re-tuning of the smoke emissions which we expect would alleviate the problems seen in the adaptive inflation algorithm for the Eurasian Boreal fire impacted region. One of our major concluding points is that work needs to be done in smoke-dominated regions to improve the system (page 28104, line 25-26 to page 28105, line 1).

29. p 28085, l 18: prior of inflation equals its posterior from previous cycle: this is also known as persistence modelling.

Response: Yes, we agree with you. In our implementation of adaptive inflation, a damping factor of 0.9 is applied to the posterior from the previous cycle to produce the prior for the next cycle (page 28085, line 21). So the damping is the time variation model for the inflation.

30. p 28086, l 2: "issues occur with the constant covariance inflation where there is limited observational coverage". See my previous comment, I believe this could be equally due to biases in observations or models than coverage.

    Response: Covariance inflation does help overcome underrepresented variance in the ensemble due to model bias and sampling error caused by the small ensemble size. We certainly agree that model bias will vary with location and time and therefore, an inflation factor at one location might not be appropriate at another. This can certainly be an issue with constant covariance inflation. However, when you have a non-uniform observing network, the result of applying a uniform inflation is that you end up with unreasonable solutions in regions that have limited observations (ie. Southern Ocean) because the ensemble is continuously inflated and there are no observations to constrain the state fields. This is a bigger issue in these under-observed regions because it can lead to the simulation crashing. This is the point we are making here.

31. p 28086, l 8: "the normalized standard deviation", that is: 1 ? Ah, Figure 4 suggests it is normalised by the mean. Please indicate this in the text as well.

    Response: Thank you, we will do that.

    Manuscript change: 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.

32. p 280866, l 22: "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". I think it is important here to note that this may be a feature solely found in DART-EAKF. To my knowledge, no other studies (e.g. Sekiyama et al, Schutgens et al, Dai et al) have found this growing ensemble spread. It may be related to the fact that in DART, inflation is applied 1) to the prior; 2) even when there is no reason for inflation (i.e. when there are no observations). p 28087, l 2: "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". This statement is far too bold with little evidence to back it up. Your analysis suggests this to be true for DART-EAKF but as I said before, it hasn’t be noticed by other authors. I suggest rephrasing this to something like: "It is suggested that particular attention is paid to the temporal evolution of ensemble spread in case a constant inflation factor is used, because our results suggest."

Response: There is a lot of evidence of this occurring in atmospheric data assimilation and hints of this in aerosol data assimilation as discussed in the response to comment 3. This is related to inflation without observations as previously discussed and is not specific to DART EAKF.

33. p 28087, l7: "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, 2009; Li et al., 2009; Miyoshi et al., 2011)." Obviously there are other reasons why AI may be preferential to a constant inflation factor. I believe the listed authors discuss the issue of model biases that are effectively dealt with by AI. Note that model biases are really the bane of DA and AI is essentially a way to sweep them under the carpet (or conversely: a way of studying them by tracking the evolution of the inflation factor).

    Response: For example, in Li et al. 2009: "we have used a globally uniform inflation factor, which is clearly not a good assumption in reality where the observations are non-uniformly distributed. With a spatially dependent inflation, we may be able to better deal with an irregularly observing network". Likewise, in Anderson 2009: "A more serious problem occurs when a single value of inflation is not appropriate for all state variables. Assimilation of in situ observations, like radiosonde and aircraft observations, in a global numerical weather prediction model provides an example. In densely observed regions like the upper troposphere over North America, ensemble variance can be inappropriately small due to model bias and sampling error. Inflation can reduce this problem. However, over the Southern Ocean, there are very few observations to constrain the model. Repeated application of inflation values large enough to correct problems over North America can systematically increase the variance of the ensemble over the Southern Ocean. Eventually, this can lead to values that are inconsistent with climatological values, and in the worst case, incompatible with the model's numerical methods. The result is ridiculous solutions, at best, and model failure, at worst. "

34. p 28087, l 23: "In particular, a large increase in spread is found at dust source regions." Presumably because of the windspeed threshold for dust emission? How much bigger than 25% is the spread?

    Response: With the meteorology ensemble, we now have different wind speeds associated with each ensemble member. This produces different amounts of dust for each ensemble member since dust emissions are a function of wind speed, therefore, increasing the ensemble spread in these regions. "In particular, a large increase in spread is found at dust source regions. For example, the spread increases from approximately 20 to 50 % in the Northern Arabian Peninsula" page 28087, lines 21-23.

35. p 28088, l 1: "the meteorology ensemble increases spread for sea salt aerosol" Seasalt emission is presumably not governed by a windspeed threshold, although it will have a non-linear dependence on windspeed. Is this effect therefore larger for dust than seasalt?

Response: We see a pretty good increase in spread for both dust and sea salt. The increase in spread is determined by how much the wind speed varies across the ensemble for a particular region and at a particular time of interest and how that difference across the ensemble translates to sources via the emission function. For dust, the emissions are represented as the surface friction velocity to the fourth power. For sea salt, the emissions are a function of the 10 meter wind speed raised to the 3.41 as described in Witek et al. 2007.

36. p 28088 l 5: "The meteorology ensemble appears to be the main driver of ensemble spread." It may be good to remind the reader that you have assumed a 25% uncertainty in emissions. I find this rather low especially because this is uncertainty on short time-scales (hourly, daily). Already at longer time-scales (months, year) Granier et al 2011 and Huneeus et al. 2011 find larger uncertainties over large regions.

    Response: We will add this point to the above sentence.

    Manuscript change: The meteorology ensemble appears to be the main driver of ensemble spread **when included with a 25% source-perturbed ensemble**.

37. p 28088, l 15: Regarding stabilization of ensemble spread, this is not obvious for WCONUS

    Response: It's hard to see in this region because there are large wildfires impacting WCONUS during the end of the simulation period. With larger AOT being produced due to the fires, the ensemble spread will increase as well. However, for longer simulations that have been conducted, we see no problems in this region with stabilization.

38. p 28088, l 20: I suggest using brackets instead of commas to delineate "the square root of the sum of the ensemble variance and the observational error variance"

    Response: Ok, thank you. We will change this.

    Manuscript change: 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.

39. p 28092, l 3: couldn't this be due to insufficient ensemble spread at low AOT? Several authors have pointed out that a positive variable like AOT can only have a large spread at small values if the distribution is allowed to be very skewed (i.e. non-Gaussian, contradicting a basic assumption in a Kalman filter). The small spreads that occur in ensemble runs are a direct result of small source perturbation at low mean source values. I believe this is an unresolved issue.

    Response: Yes, we agree on this point and will include this in our discussion of the results.

    Manuscript change: 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.

40. p 28092, l 7: The case of too small a spread at high AOT may also be the result of missing causes of uncertainty. E.g. you don't perturb deposition processes. Perturbing them will have a bigger impact at high AOT than at low AOT because (again) AOT cannot go below zero.

    Response: Yes, we agree that not having enough spread means that we aren't capturing all the uncertainties.

    Manuscript change: 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.

41. p 28093, l 17: "since they are independent." The prior and the observations are also independent so this cannot be the reason to choose the posterior.

    Response: Please see the response to comments 23 and 26.

42. p 29095, l 7: "performance gains" The authors are undoubtably aware that this comes at a hefty cost: 4x more CPU requirements. I think that 'performance' may not be the best word here as it implicitly suggests some optimal cost/benefit ratio.

    Response: Thank you, we will reword this statement.

    Manuscript change: 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).

43. p 28096, l 1: It would be good at this stage to point out that NAVDAS-AOD does not include perturbed meteorology (as far as I understand it). I.e. something like Fig 10 is unlikely to be seen for NAVDAS-AOD

    Response: NAVDAS-AOD and NAAPS can't have a perturbed meteorology because it is a deterministic simulation.

44. Sect 3.3 & 3.4 and Table 3 etc: an evaluation of a base model run (control) should be part of this analysis. Is there even a substantial improvement in AOT due to assimilation (either 3DVAR or EAKF)?

Response: Please see our response to comment 1 on this topic. Our focus in this work is to see how the ENAAPS-DART system performs relative to the current operational system which serves as our baseline. Including DA does produce an improvement in AOT. We have subsequent work that will show this in more detail.

45. p 28100, l 14-19: "On the other hand, the forecasts initialized with the EAKF fields do a better job capturing the leading edge of the dust front with the ENAAPS-DART version being smoother than the deterministic counterpart along the dust front. This demonstrates that the sharpness achieved in the ensemble data assimilation propagates in the forecast and is an advantage of using the EAKF initial conditions over the variational initial conditions for the short-term forecast." The use of ’sharpness’ and ’smooth’ confused me initially. Unless I am mistaken, they are not juxtaposed but describe different aspects. Consider repharsing this sentence.

    Response: Yes, we can see where your confusion comes from in this statement. We were referring to different aspect of the predicted dust front, which makes it confusing. We will reword this discussion to make it clearer. Thanks.

    Manuscript change: 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 variational initial conditions for the short-term forecast.

46. p 28100, l 5-19: I think it should be pointed out that a substantial part of the plume (eg the northern edge) is missed by all four forecasts. Please discuss possible causes.

    Response: Since this is consistent across all forecasts, this is likely attributed by model physics which is consistent across these configurations.

    Manuscript change: 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 produce 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.**

47. Section 4, Discussion: I suggest removing this Section in its entirety. It is not really a discussion but an extended summary. Its main points have already been discussed (in detail) in the main text. Important conclusions in this Discussion that are not yet in the Summary should be moved there and phrased more consisely.

    Response: Thank you, we will rework the discussion.

48. Section 5, Summary: consider my general comments.

Response: Ok

49. Fig 6: Not quite clear what is shown here. This is essentially the model forecast covariance? So it is with respect to a single location? Presumably the black dot in the top row (there are no dots in the lower rows)? It is the correlation in the AOT fields?

    Response: This is the spatial correlation in the prior AOT relative to a point indicated by the black star. This is meant to show how observational information will spread in different configurations of the ENAAPS-DART system. The figure caption will be updated and the size of the black star is now increased in the figure.

    Manuscript change: Figure 6. Ensemble correlation fields **in the prior AOT relative to a point indicated by a black star** for three different aerosol events:

50. Fig 15: What does "Not all available MODIS observations are assimilated" refer to? I realise that the NRL-MODIS dataset is a subset of the official Col 5 product. But why show here a different product than that which you have assimilated?

    Response: In this figure, we were trying to show the sharp gradient in the dust front that is produced in the ENAAPS-DART system is also seen in MODIS observations. You can see this clearly when you look at all the observations. That is why we included this figure, however, we will include an additional plot of just the assimilated observations (which are a subset of what has been shown already).

References:

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

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

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, Weather Forecast., 25, 985–993, 2010.

Miyoshi, T.: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter, Mon. Weather Rev., 139, 1519–1535, doi:10.1175/2010MWR3570.1, 2011.

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, 2010

Whitaker, J.S., and Hamill, T.M.: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. Monthly Weather Review, Volume 140, pp 3078-3089, 2012.

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

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.

Reply to comments by T. Sekiyama:

Thank you Thomas for the thoughtful review. Below are our responses to your comments.

1. I was surprised that the adaptive inflation worked well for aerosol in this study because my adaptive inflation failed and diverged when I used a method other than Anderson 2009. I have thought that the adaptive inflation for aerosol is unstable due to a large uncertainty of aerosol modeling compared to NWP. What do you think?

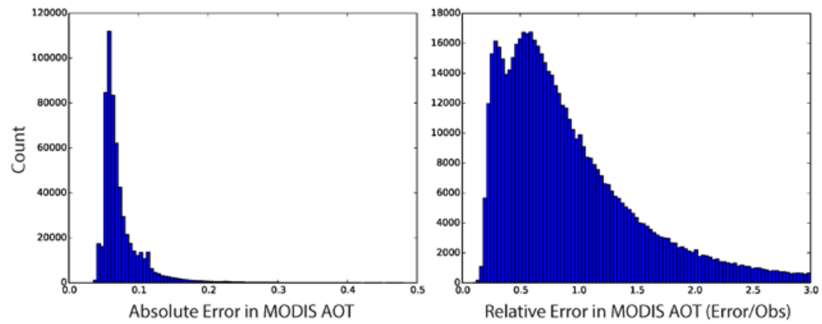
   Response: The adaptive inflation worked quite well (ie. stable) with the exception of regions impacted by fires. Without any measures to control inflation in these regions, the adaptive inflation did in fact blow-up with inflation factors exceeding 10. Eventually, unrealistic aerosol concentrations were produced and the simulations crashed. This behavior of the adaptive inflation algorithm for fire-impacted regions indicates that there is an inconsistency between the observational and the background distribution in optical thickness. Fire emissions have very large uncertainties and we thought were the likely drivers of the inconsistencies generating the unstable growth in the inflation factor since large and persistent fires were occurring during the simulation time period. In order to create stability in the simulations, we tuned the standard deviation of the inflation factor and defined a maximum inflation factor (1.5). However, we think that doing some tuning to the smoke emissions in the future would allow for the adaptive inflation to run without a maximum inflation needed. These stability problems were discussed in the results section (page 28085, lines 7-27).

2. It was not described in this paper how (and how much) the observation errors were estimated. Even though it is described in the references, the estimation method and size of observation errors are crucial for data assimilation. It is better to show the validity of the method (and error size) in the manuscript, if possible. Generally speaking, “observation errors” are underestimated because it is difficult to estimate spatial representativeness and remote-sensing bias. I am afraid that observation errors are unnaturally underestimated in this paper too.

   Response: Yes, we agree that the observational errors are a crucial component of data assimilation. The observational error estimates are based on long-term comparisons of MODIS Terra and Aqua AOT to AERONET AOT for over-ocean (Zhang and Reid, 2006, 2009) and over-land (Hyer et al. 2011). The observational error covariances are treated as diagonal matrices, so no accounting for correlated errors. We can add more discussion on this in the paper. In this study, since we are using the current operational system as a baseline for comparison, we wanted to assimilate the exact same product as is used in the NAVDAS-AOD system, so we made no changes to how the observational error is represented. However, we can include some additional plots in the supplementary material to show what the observational error looks like. In future work, we may reevaluate the observational error.

   Manuscript change: 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.



3. All the emissions in this study were perturbed using the same factor for a given ensemble member. Actually, it is not a good way as the authors mentioned. Instead of that, there are some alternative techniques to reduce correlations between independent sources. For example, make perturbation factors one-by-one randomly each grid-point, 2) smooth out the distribution of the factors using a 3-dimensional smoothing filter, 3) and use the smoothed factors to perturb emission sources. Usually, the ensemble mean of the perturbed emission flux is not very shifted by this method.

   Response: In this study, the same perturbation factor is applied for a given ensemble member for each source type. As an example, smoke emissions for ensemble member n are all perturbed with the same randomly produced perturbation factor. This essentially creates an infinite correlation lengthscale for smoke emissions that is only limited by the localization lengthscale. However, for ensemble member n, the perturbation factor for smoke, dust, sea salt, and anthropogenic and biogenic fine are not the same. Thus, given our localization of X, we do have a regional smoothing parameter in a way built in. We preferred this methodology to the moving Gaussian method in that method predefines the maximum length scales. Here we wanted to see what naturally and reasonably covarried, and then look at how those covariance fields looked. We will make these points more clear in the manuscript. But, as noted in the manuscript (page 28079, lines 20-21), we did initially try grid-by-grid source perturbations as you suggested. We found this had no impact on ensemble spread, therefore, ruled this method out. Indeed, the strategy used in this work for perturbing source functions worked well when the emission correlation lengthscale is greater than the localization lengthscale (ie. large spatially correlated aerosol events). For source types in which the emission correlation lengthscale is less than the localization lengthscale (ie. spatially independent sources such as small boreal forest fires), we plan to test a perturbation function as you suggested. We think this should provide substantial improvement in some problem regions identified in the manuscript (ie. Eastern united states, North American boreal regions). This problem, and the point you made above regarding adaptive inflation, are of course intertwined with your comment 1.

4. The authors are using a maximum inflation limit, but I am afraid that the maximum value (=1.30) is too small because adaptive inflation factors more than 2 or 3 are acceptable for NWP without any problem.

   Response: The maximum inflation used in this work is actually 1.50 (page 28085, line 20). When we tested the free-running adaptive inflation without any constraints on the maximum inflation, the maximum inflation did not exceed 1.5 with the exception of fire-impacted regions where the adaptive inflation became unstable. We believe this instability is due to the persistent nature of the fires during the simulation time period and an inconsistency between the background (dominated by emissions for these large fire events) and the observations. This was the value for which we found the adaptive inflation was stable for fire regions; however, we think with some tuning of the fire emissions, we can let the adaptive inflation algorithm run freely without a maximum inflation constraint. This is work planned for future studies.

5. The authors say, “the ensemble isn’t fully representing the distribution with an excess of observations occurring of low ranks,” but when the rank histogram shows a one-side peak, it is only certain that the ensemble members have a large bias. With only the information of “one-side peak,” we don’t know whether the ensemble spread is small or not.

   Response: When the majority of observations are below the lowest bin, this indicates bias in the ensemble relative to the observations, just as you stated. Yes, we agree that since the observations are mostly below the ensemble members, you can’t state much about the actual spread of the ensemble relative to the observational spread. What we meant with this statement is that the ensemble members aren’t capturing the low AOT values of the observed distribution. We will revise this statement to make it clearer.

   Manuscript change: The Eurasian Boreal smoke region rank histogram, consistent with the evaluation of the total spread to RMSE ratio, shows that the ensemble isn’t **capturing low AOT values of the observed distribution**, with an excess of observations occurring for low ranks.

6. If AOT values are small, it’s no wonder AOT observational errors are relatively large because the error of remote sensing is almost independent from the AOT (=retrieved) value. On the other hand, when AOT values are small, it’s impossible to make a large ensemble spread. This is a disadvantage of ensemble data assimilation.

   Response: Yes, we agree with your statement. Since we have a positive-definite state variable, the ensemble spread can only be so large for small AOT values. The result is that the observational error is much greater than the forecast error and the assimilation would weight the analysis mostly to the background. So if there is a bias present, the assimilation won’t be able to correct for this. We will include the fact that this is a limitation of ensemble data assimilation as you mentioned.

   Manuscript change: 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.

7. Do the authors mean that there is a large difference between meteorological analyses since there are few meteorological observations in the Southern Ocean? If so, this sentence (Line 18-20) is a little confusing.

   Response: NAAPS and ENAAPS are offline, so here we are assimilating only aerosol-related observations. There are no AOT observations being assimilated in the Southern Ocean. Between deterministic NAAPS and ensemble NAAPS (ENAAPS), the only difference is the data assimilation system and the meteorology fields (deterministic NOGAPS and the ensemble NOGAPS fields) they are run on. Since there are no AOT observations being assimilated in the Southern Ocean, any differences between the NAAPS/NAVDAS-AOD simulation and the ENAAPS-DART simulation are due to differences in the meteorology fields used to drive the simulations. For example, sea salt emissions are parameterized as a function of wind speed. Differences in wind speed between deterministic and ensemble meteorology fields would impact sea salt emissions and therefore, the optical thickness in the region. Likewise, differences in humidity fields would impact the optical thickness. We will add to the discussion in the manuscript to make this point more clear.

   Manuscript change: 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.**

8. The authors often use the term "variational" (assimilation, system, initial condition, etc.) as an inferior method to the EAKF, but the "variational" method is the 2D Var in this paper. We have another variational method, the 4D Var, which is comparable or superior to the EAKF. It is better to always use the term "2D Var" in this paper to avoid confusion

   Response: Yes, we agree with your statement and will update the manuscript to be clear that we refer to 2D Var and not all variational methods.

9. I could not understand the meaning of "the optimal combined meteorology and source ensemble". What is optimal?

   Response: Here we meant that the combined source and meteorology ensemble performed better than source-perturbed or meteorology ensemble alone and was the chosen approach. We will revise this statement to say the chosen configuration instead of optimal.

   Manuscript change: 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.

10. In this study, the EAKF system captures sharp gradients while the 2D Var smooths plume distributions. However, the EAKF and 2D Var have similar RMSE and bias. That means, probably, although the EAKF result looks realistic, the plume location is slightly shifted from the real one. It is difficult to judge which is better "sharp but slightly-shifted plumes" or "blunt but broadly-covering plumes" as operational prediction/warning, I think.

    Response: Yes, we agree that it is difficult to define which is better in this instance depending on what the application of the forecast is (ie. Smoothed out event would give a larger warning region). However, we think the real advantage of the ensemble approach is that we can produce more realistic corrections to the state fields (which produce sharper gradients that are consistent with what is seen from satellite) which will become more important as additional observational information is introduced into the system, such as Lidar and other spatially limited pieces of information.

    Manuscript change: 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 produce 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.**

11. Are these RMSE global?

    Response: Yes, these are global.

    Manuscript change: The 24-hour forecast **global** RMSE against AERONET AOT with bootstrapped 95% confidence intervals

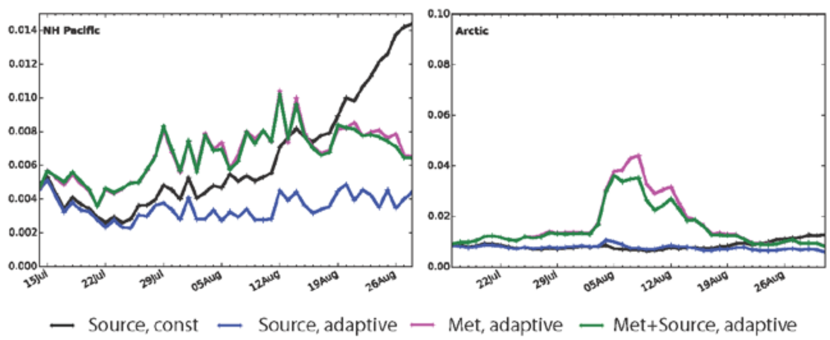
12. The authors say, "the observational error may be too large for small AOT values, which could also contribute to the positive bias", but I don't think so. Generally speaking, it is extremely difficult to assimilate zero or almost zero values like small AOT. It is because a population that contains a lot of zeros (or almost zeros) and is not allowed to be negative values (e.g., radar-measured precipitation) is not Gaussian-distributed. Fundamentally, it is nonsense to quantify the error of non-Gaussian-distributed values using a standard deviation. However, data assimilation assumes everything Gaussian. It is the reason why zero-value assimilation is difficult. The positive bias observed in smoke regions may be relevant to non-Gaussian AOT distribution and irrelevant to the size of AOT observational error.

    Response: Thank you for your input on this. We agree with your statement and will add discussion on this issue. There has been discussion here on to what extent this is a real problem, and what is the best way to cope with, ranging from complex transforms to something simple, like assimilate in log space.

    Manuscript change: The discussion and conclusion were consolidated, but we have changed our discussion in the manuscript to talk about dealing with small AOT values (such as changes to comment 6 above) as well as in the conclusions.

13. I am very interested in NH Pacific Ocean, Arctic, and Antarctic.

Response: We can add additional plots to the supplementary material for these regions.



14. Why did the authors plot MODIS AOT that was not quality-controlled? I would like to see the comparison between assimilated observations (= quality-controlled AOT) and assimilation results.

Response: We were trying to show the sharp-gradient present in the MODIS AOT observations. This can be seen pretty clearly when all AOT values are shown, however, we added an additional plot with assimilated AOT only.

15. I am very interested in why the AOT over the Sahara is largely changed by the 2DVar although there is almost no observation over the Sahara. The influence radius of observations in 2D Var is only 250 km or so, right?

Response: The radius of influence for the variational system is determined through an exponential function as defined in Zhang et al. 2008. If R is the distance between observation and background location and L is the defined 200km lengthscale, the function is (1+R/L)*exp(-R/L). An influence can be present beyond the defined 200km lengthscale; however, the impact will decrease with distance.

16. Page 28080, Line 16: The description “over 13 land regions” is actually “over 15 land regions”?

Response: Thank you, we updated this to 15.

Manuscript change: 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.

17. Page 28085, Line 11: There are two spellings “blow-up” and “blowup” in this manuscript. Choose either one

Manuscript change: We changed this to “blow up”. Thank you.

18. Page 28090, Line 18: Is it necessary after “large” to put a comma?

    Response: I think with multiple adjectives, we need to separate them with a comma. We could be wrong though.

19. Page 28092, Line 22: isn’t -> is not

    Manuscript change: 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.

20. Page 28093, Lines 10 and 12, etc.:There are two expressions “meteorology ensemble” and “NOGAPS ensemble” in this manuscript. Choose either one.

    Manuscript change: These were changed to either “meteorology ensemble” or “NOGAPS meteorology ensemble”. The NOGAPS is included at times to be specific about where the meteorology fields come from.

21. Page 28093, Line 29: There are two spellings “source-perturbed” and “source perturbed” in this manuscript. Choose either one

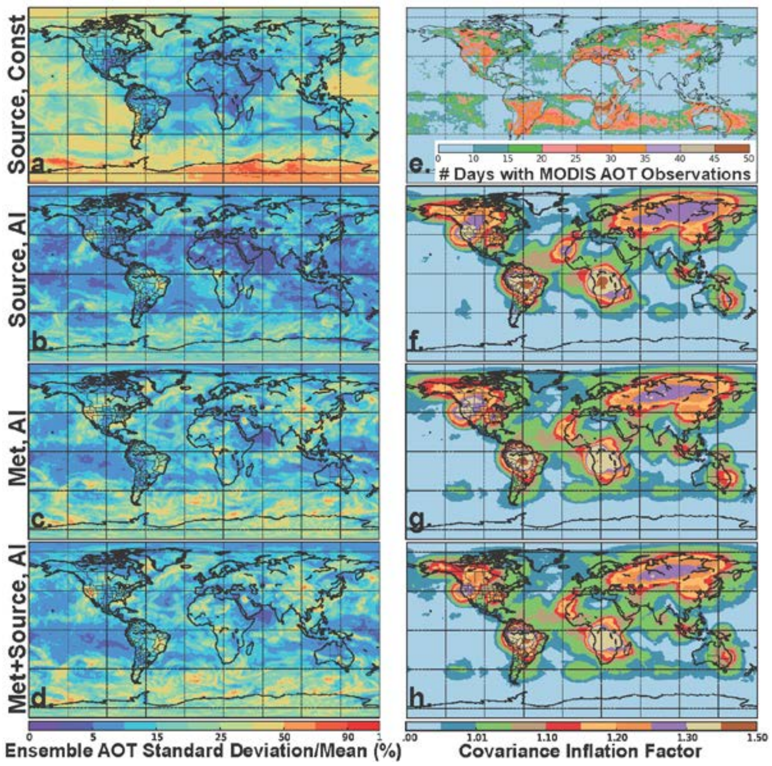
    Manuscript change: These were all updated to “source-perturbed”. Thanks.

22. Page 28094, Line 27: Putting “(Table 2)” at the end of this sentence, it becomes easy understandable.

    Response: This paragraph is talking about the evaluation of the posterior relative to AERONET AOT. Table 2 is the evaluation of the prior to MODIS AOT.

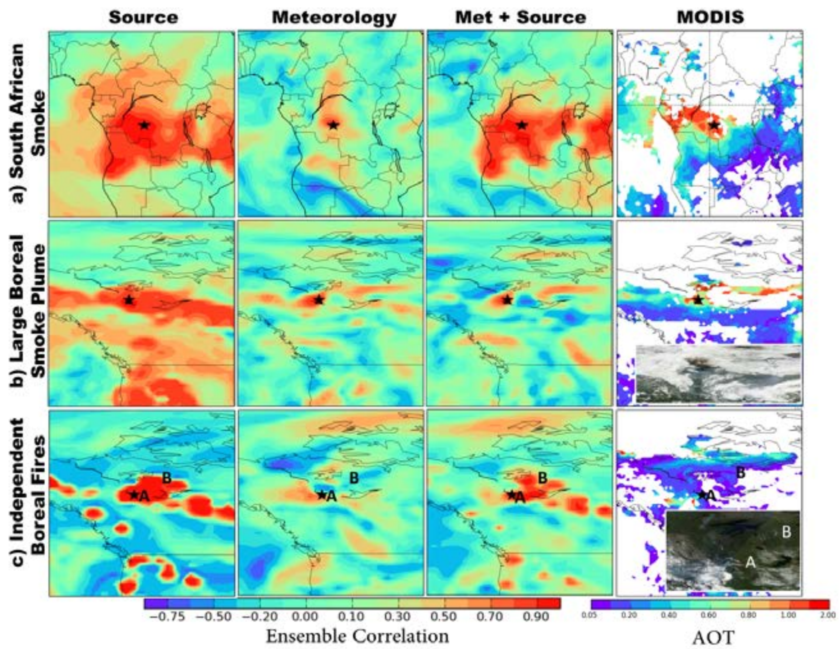
23. Figure 4: The characters “a” “h” in the figure panels are too small and extremely unreadable

    Manuscript change: This figure was updated with larger font.



24. Figure 6: It is very difficult to find a “point”, especially in (b) panels

Manuscript change: This figure was updated with larger black stars.



25. Figure 11: Some of the AOT observation plots are illegible, especially on 22 August

    Response: In the AERONET AOT timeseries plot, there aren't any observations on August 22.

26. Caption of Figures 11 and 12: The "analysis" is plotted here, I think. But the caption says, "predicted total AOT".

    Response: Yes, you are correct. We changed the caption to be more specific. Thank you.

    Manuscript change: Timeseries of **analysis** total AOT (grey

27. Figure 16: It is hard to find the area where the MODIS plot indicates, at a glance

    Response: We were trying to zoom in to the leading edge to show how sharp the gradient is. We will update the caption to make it clearer.

    Manuscript change: 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.

Response to Referee A. Benedetti's Comments

Thank you, Angela for taking the time to review the paper. We very much appreciate all of your comments. Please see our responses below.

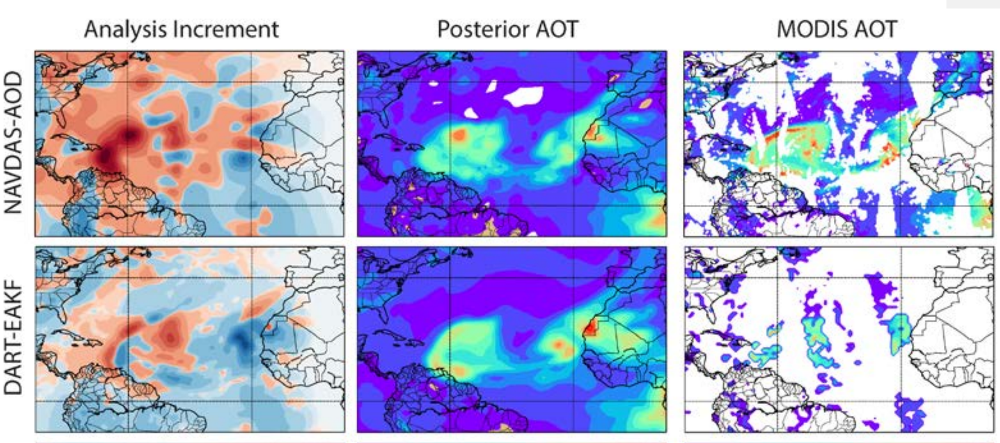
1. Page 4 Line 33.I would say "research" arena rather than "operational" arena as to my knowledge at the moment there are no operational ensemble systems for aerosols (although the situation may soon change).

   Response: We agree, we will update this statement to make it accurate. Thanks.

   Manuscript change: 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).

2. Page 5 Line 10. Here, like elsewhere where the comparison between the ensemble and variational systems was made, I thought it would be good to see the background error covariance matrices for the ensemble and the variational system side by side. Perhaps, if possible, for future work as well, it would be interesting showing the increments from a single observation experiment to show how the different background error statistics affect the distribution of the increments and spread to neighbouring points the information from a single observation

   Response: We agree that showing the analysis increments and error covariances would be helpful. We do show analysis increments for our Saharan dust case in Figure 15, but we can add a few other examples in the supplementary material. Yes, we agree that some single observation experiments would be nice to show as well. We might not show those types of experiments in this paper since it is already quite long, but we definitely will in subsequent papers.



Analysis increments (posterior-prior AOT) and posterior AOT fields from the 2DVar NAVDAS-AOD and the DART-EAKF for a dust event on August 2, 2013 (18Z).

3. Page 6 Line 22. How is the adaptive inflation estimated? Is it based on first guess departures? I know that the reader can look up the references, but just a sentence to explain briefly what the estimation is based on would be welcome.

   Response: We added some more information on the adaptive inflation in the methods section. Thank you for the suggestion.

   Manuscript change: An alternative method to a uniform multiplicative inflation is adaptive covariance inflation (Anderson 2009) which produces 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.**

4. Page 7 Line 12. 25% seems like a large perturbation, although later you say that it might be small for certain emissions (for example fires). How is this value assigned? I am surprised that location-dependent perturbations did not help with the ensemble performance, as you later mention that for localized sources the ensemble had the problem of over-correlating them. Perhaps the perturbations should be a function of the source spatial extension and intensity. I really do not know, just wondering.

   Response: In general, we have seemed to get the opposite reaction, that the source perturbation is small. We selected 25% as a pretty conservative estimate of the source perturbation with the expectation that we would evaluate the system performance and see what adjustments needed to be made. In general, we found that the system did pretty well in representing uncertainty (spread ~ RMSE) with this perturbations with the exception of fire-impacted regions. For these regions, we weren't getting enough spread, especially for high AOT events. This tells us that we probably need to increase the source perturbation for fires. Fire emissions are also highly uncertain, so needing perturbations larger than 25% for these emissions is not unexpected.

   The perturbations to the aerosol sources aren't location-dependent (we will work to make this clearer in the methods). We initially tried random perturbations that were drawn for each grid, however, we ruled this method out (page 28079, lines 20-21). The method that we did use in this work was to apply a randomly drawn perturbation for each aerosol source and for each ensemble member. This essentially creates large correlations between all emissions of aerosol of a given source-type (dust as an example), only limited by the localization. So for emissions in which the correlation lengthscale is smaller than the localization lengthscale (such as pollution sources), we identified issues. For events in which the correlation lengthscale is greater than the localization lengthscale, this method worked well (ie. large smoke and dust plumes). We have plans to reassess the source perturbations in future work to better deal with emissions for pollution events and small fires.

5. Page 11 Line 9. Please do explain briefly the methodology behind AI

   Response: We agree. We added some additional information in the methods section that we think will clarify the AI discussion in the rest of the paper.

6. Line 19. That points to model shortcomings which are not likely to be corrected with DA

   Response: Yes, we agree. We think that this is related to issues in the smoke emissions. The adaptive inflation is trying to correct for inconsistencies in the model prior and the observations by inflating. The smoke emissions are persistent during this time period and are likely contributing to the discrepancy between the model distribution and observations which leads to over-inflation by the AI and eventually, a crashing of the model. This is why we think we need to do some tuning to the smoke emissions and hope that this will alleviate the problem. We hope that once we tune the smoke emissions, the AI can be run without any measures for preventing inflation blow up.

   Manuscript change: 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.**

7. Page 12 Line 11. Well phrased. This is another one of the issues related to the fact that the aerosol problem is under-constrained.

   Response: Thanks!

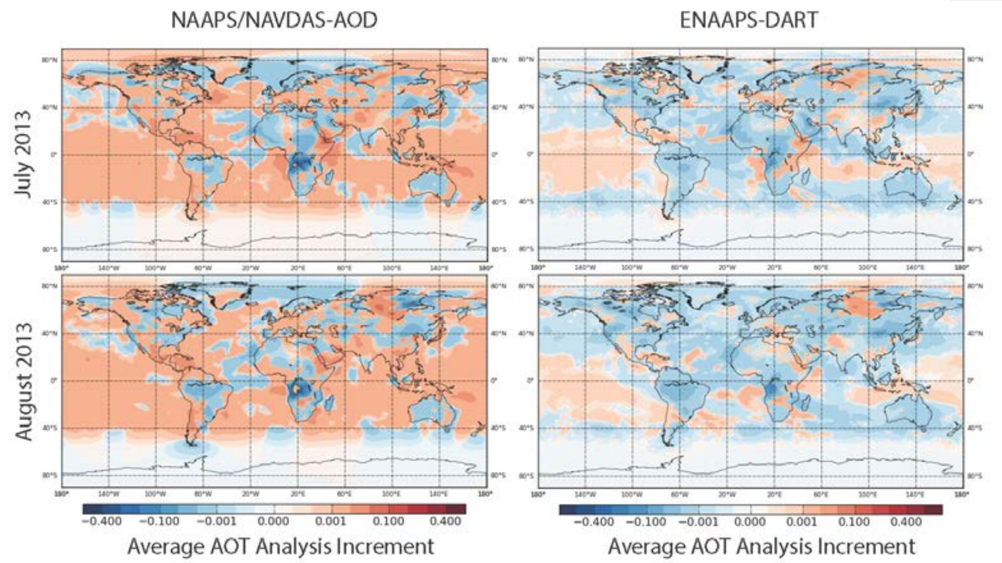
8. Page 15 Line 35. An interesting conclusion about the observation errors being too large for small AOTs. Perhaps the methodology of Desroziers et al (2005) could be applied to ascertain so in a more mathematical way. [Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385–3396. doi:10.1256/qj.05.108]

   Response: We should clarify that we mean that the observational error is too large relative to the ensemble spread. This is probably more likely due to aerosol being a positive-definite, therefore, it is hard to get enough spread near-zero. We added some discussion on this point. However, this still has an important implication for data assimilation in that the observations won't have much impact. We have been discussing different ways to deal with this issue, including doing a data transform on the observations before assimilation.

9. Page 19 Line 12. The fact that the RMSE values of the two analysis are not statistically different might also mean that the system is driven more by the observations than the background, and perhaps the observations errors are too small. This may seem to contradict what said on page 15 line 35, but the two things may co-exist as the balance is to be obtained between the background errors and the observation errors and it is possible that the analysis draws too much to the observations (i.e. the background errors are large with respect to the observation errors). Again, perhaps an analysis of the departures of both the variational and ensemble analyses could offer some insight on this particular aspect.

   Response: Yes, we agree, the observations are pulling the priors in the two systems to similar values (although there are differences such as over ocean where we aren't verifying with AERONET). The ensemble system tends to produce larger AOT values (positive bias) and the observations in general pull the AOT lower. The deterministic system tends to produce smaller AOT values (negative bias) and the observations tend to pull the AOT higher (see figure below).

However, the analysis increments tend to be smaller for the ensemble system than the deterministic system. This means that the forecast error as specified for the 2DVar system is quite large and puts a lot of weight toward the obs. The forecast error determined by the ensemble is smaller and as a result has a smaller analysis increment (ie. less obs impact), this would indicate that the ensemble system is doing a better job in the short-term forecast (prior) at least at AERONET sites



10. Page 19 Line 15. Please use another verb other than "produced", like "displayed".

    Response: Thank you, we updated this and similar wording throughout the text.

11. Page 19, Line 39 As already mentioned, it would be good to see a plot of the background error covariance matrices for the variational and the ensemble system (single observation experiment increments would also do the job). Figure 15 shows some of this, but it would be good to have a dedicated single observation experiment.

    Response: We may reserve single observations experiments for subsequent papers since this paper is already quite lengthy.

12. Page 20 Line 16. To be fair to the variational system, it is definitely not tuned at all to capture sharp gradients. I presume the 2D-Var background error covariance matrix is spatially homogeneous, constant and with fixed correlation length. It seems to be asking too much of the system.

    Response: Yes, we agree that the 2DVar system won't be able to capture gradients based on how the error covariance matrix is defined. This is meant to demonstrate why an ensemble approach might be the chosen approach moving forward or at least should be part of the operational runs, especially as we begin to incorporate spatially-limited observations.

13. Page 24 Line 32. Have you looked what happens at longer forecast ranges than 24h?

    Response: No, not at this point in time. However, we have plans to implement this system semi-operationally and will begin to evaluate forecasts out to a few days.