Geosci. Model Dev. Discuss., 8, C801–C819, 2015
www.geosci-model-dev-discuss.net/8/C801/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions

Interactive comment on "Development and application of the WRFPLUS-Chem online chemistry adjoint and WRFDA-Chem assimilation system" *by* J. J. Guerrette and D. K. Henze

J. J. Guerrette and D. K. Henze

jonathan.guerrette@colorado.edu

Received and published: 19 May 2015

We are thankful for all of the constructive comments and questions by M. Krol. The corresponding responses and manuscript changes are given below.

1 Responses to specific comments

1. **In the introduction, no reference is made to the pioneering work of Elbern et al. with the EURAD model, who worked on 4D-VAR chemical data assim-**

ilation for more than two decades.

The introduction to this paper focused on literature describing data assimilation methods that either utilized an online model or were for aerosols. Bocquet at al. (2014) review existing online and offline chemical data assimilation capabilities in more detail. As such, we have modified page 2315, line 25-26 as follows:

> To address this, chemical data assimilation can be used to improve short-term forecasts. Bocquet at al. (2014) review existing methods and previous applications of chemical data assimilation in CTMs and NWP-chemistry models.

2. **In general, it would be interesting to compare the approach described here to other approaches. For instance, some 4DVAR approaches (e.g. Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., et al. (2009). Inverse modeling of global and regional CH 4emissions using SCIAMACHY satellite retrievals. Journal of Geophysical Research, 114(D22), D22301. doi:10.1029/2009JD012287) use a two-step inversion. Observations that are not fitted within 3σ after the first optimization are left out with an argument that the model is not able to reproduce these observations. In the current study, some of the high aircraft observations may be due to specific layered outflow from a specific convection event, which is not (and might never be) adequately resolved by the model. Nevertheless, the advanced estimation of model error with the different settings in WRF is impressive. Without a true inversion, however, it is not possible to assess how well the observations finally will be matched. My main point here is that a discussion of this work in the context of existing techniques would be of added value.**

We agree that the online data assimilation method will need to be compared against existing CTM 4D-Var systems, especially to evaluate the benefits of in-

cluding online physics. Still, the purpose of this paper is to describe development of a chemical adjoint with online meteorology, which is the first step toward enabling the online 4D-Var system in WRFDA-Chem. The WRFDA-Chem platform is not at a state where a comparison might be made between online and offline approaches.

Additionally, the weighting scheme presented here cannot be compared to some other observation filtering method outside the context of a 4D-Var inversion. As stated on pp. 2339-2340, we introduce the weighting scheme, but do not exhaustively test it. You are correct that future 4D-Var studies will require distinction between residual error due to emission inventory and physical parameterization errors. As you mention, Bergamaschi et al. (2009) remove observations outside 3σ after an initial 4D-Var optimization that includes four outer loop iterations. Analyzing residual errors after an inversion is very useful to determine where model descriptions are weak. As such, we have amended the paragraph starting on page 2337, line 13 as follows:

> Observations with significant model bias would require the largest perturbation in control variables to alleviate, and would seem to inform the inversion process the greatest. However, they must also have low total variance to contribute to an inversion. Figure 9 shows the surface and aircraft SD plotted versus residual error. Also plotted in that figure are one and two SD zones, as well as lines of constant $\lambda^*_{k,o}$ for all $w_k = 1$. Any residual falling outside the 2σ zone has a combined model and observation SD that is small enough to determine with 95 $\%$ confidence ($p < 0.05$) that the residual error deviates from zero (i.e., the model and observation disagree). These statistically significant model errors indicate that some kind of inversion is worthwhile. In their multicycle 4D-Var approach, Bergamaschi et al. (2009) eliminate observations outside three SD's after an initial 4D-Var cycle, with the thought

> that incorrect model physics prevent those residual errors from being fixed with 4D-Var. Thus, while statistically significant residuals are important to driving a 4D-Var inversion, that they remain afterward is a strong indication of errors in the model description that cannot be fixed through adjustments to emissions. Figure 9 shows that the relative contributions of observation and model variances is in general proportional to the relative magnitudes of observed and modeled concentration. Specifically, model (observation) variation contributes to a large fraction of uncertainty in positive (negative) residuals.

3. **I do not see why the summation is split in eq. 2a and does not simply run to n. Please explain.**

   The summation is split to distinguish the inner loop control variable increment, $\vec{\delta x}$. We have changed Eqs. (2a) and (2b) to the following:

   $$\mathsf{J_b} = \frac{1}{2}\left[\vec{\delta x} + \sum_{i=1}^{n-1}\left(\vec{x}^i - \vec{x}^{i-1}\right)\right]^\top \mathbf{B}^{-1} \left[\vec{\delta x} + \sum_{i=1}^{n-1}\left(\vec{x}^i - \vec{x}^{i-1}\right)\right] \text{ and } \mathsf{J_o} = \frac{1}{2}\sum_{k=1}^{K}\left\{H_k\left[M_k(\vec{x}^n)\right] - \vec{y}_k\right\}^\top \mathbf{R}_k^{-1} \left\{H_k\left[M_k(\vec{x}^n)\right] - \vec{y}_k\right\}$$
   $$\approx \frac{1}{2}\sum_{k=1}^{K}\left[\mathbf{H}_k\mathbf{M}_k\vec{\delta x} - \vec{d}_k\right]^\top \mathbf{R}_k^{-1} \left[\mathbf{H}_k\mathbf{M}_k\vec{\delta x} - \vec{d}_k\right].$$

4. **Page 2321, line 2: "earliest emission time": up to now "x" was a general variable, that is now linked suddenly to emissions. Please explain this better.**

   Indeed, this sentence should be kept more general. We will replace "the earliest emission time" with "earlier simulation times".

5. **Page 2321, line 17: "nonlinear to a quadratic form". I see what you mean,**

but strictly speaking quadratic is also nonlinear.

This sentence was incorrectly paraphrased. Page 2321, lines 16-17 will be corrected to: "The purpose of the two-level optimization is that approximating M with M simplifies the full problem to a quadratic problem, and guarantees a unique solution to the minimization..."

6. **Page 2325: "some cost function at location p and time step f with". Up to now, the cost function was introduced as a global variable (eq. 1). Defining it here as a space and time dependent variable is confusing.**

   In order to make the new definition less confusing, we have explicitly stated the new definition instead of only implying it. We have changed the text on page 2325 from lines 4 to 13 as follows:

   > ... We define a new cost function equal to a single predicted state variable, locally defined in grid cell p and at the end of time step f, $J = SV_{p,f}$. We use the TLM, ADM, and a centered finite difference approximation from the FWM to evaluate derivatives $\chi_{p,q} = \frac{\partial J}{\partial x_{q,0}}$, with respect to some CV at location q and the initial time, 0. The finite difference derivatives are calculated from $\chi^{NL}_{p,q} \approx \frac{J(x_{q,0}+\delta x)-J(x_{q,0}-\delta x)}{2\delta x}$, where each evaluation of J results from a FWM simulation with some perturbed value of $x_{q,0}$. δx varies between 0.1 and 10% of the value of $x_{q,0}$. The adjoint and tangent linear derivatives are found by forcing the model gradient fields, λ^* and λ, at J and x_q, respectively. ...

7. **Page 2325: Comparing the results of eqs. 5 and 6 is known as the gradient test (see e.g. ECMWF documentation). Normally, you take dx that approaches zero and the finite difference gradient will approach the true gradient until numerical rounding errors become important. To my experience, for double precision calculations, derivatives can be approximated**

within 10^{-7} before rounding errors kick in. In my applications there is convergence until dx is about 10^{-6}. I do not see why one has to fiddle around with different values of dx (0.1, 1, 10%, i.e. relatively large values) to see what value performs best.

We appreciate the referee pointing out that the gradient test is not specific to any modeling system, and have corrected the sentence beginning page 2325, line 3 to be: "Here we use an alternative verification approach based on Taylor series derivative approximations, and similar to that used by, e.g., Henze et al. (2007), to verify WRFPLUS-Chem."

The referee is right that continuous model equations should lead to finite difference approximations becoming more accurate as step size is decreased, which is a benefit to using them to approximate derivatives for nonlinear systems. In their Fig. 4, Henze et al. (2007) showed that this is not the case for discontinuous numerical algorithms, where larger perturbations may lead to smaller errors. This phenomena is described by Thuburn and Haine (2001), and it likely arises in WRFPLUS-Chem due to flux limiters in the 5th order, monotonic, horizontal tracer advection.

8. **In the paper (page 2326) it is written: "A range of finite difference perturbations dx is used for U, T , and Q_v control variables in order to find a value of χ_{NL} with the best compromise between truncation and roundoff error." Another problem might be that perturbations to U in the forward model perturb the physics (atmospheric flow is normally defined in vorticity and divergence), and that this violates some mass-conservation constraints. This might be the reason for the strange behavior presented later in figure 5, which look rather suspect in my opinion. The sensitivities for something linear as emissions (figure 5, first panel) look perfectly fine and what would be expected.**

   In response to your questioning perturbations in U, the standard WRFPLUS ver-

ification test described by Zhang et al. (2013) perturbs U, T, and Q_v, but does so across the entire domain simultaneously. Their approach results in a decrease of error as the domain-wide perturbation is reduced, however it likely averages out any discontinuities introduced at specific perturbation locations. To be clear, the T used in the model is actually the perturbation potential temperature. Thus we have changed T to $\delta\Theta$ in the paper.

We agree that the finite difference results in Fig. 5 reflect substantial nonlinear responses. As such, we repeated the test with new source code and post processing. Unfortunately, we were unable to determine the initial locations, q, that we used previously due to a data loss. However, we repeated these tests in locations likely to be equivalent; a revised version of Fig. 5 is shown below. We think that the resulting derivatives are within the tangent linear regime expected in an online model. Also note that the derivatives with respect to U and Q_v are evaluated very near the coastline, which could be a source of additional advective discontinuities (as discussed in response 7 above).

We have also modified the text starting on page 2329, line 3 until the end of that paragraph as follows:

> However, the TLM has inflection points at the same times as the finite difference approximations, including during fast transient periods, such as for $\frac{\partial U}{\partial U}$ and $\frac{\partial U}{\partial Q_v}$. The duration over which the tangent linear assumption is valid for chemical responses to U and Q_v depends on the size of the perturbation and on the local regime of meteorology. For instance, the test location shown is very near the California coast, but better agreement was found for an inland response location, p. Further testing of the coupled derivatives will be necessary to determine over what time period they are suitable for inverse modeling, and under what conditions the model nonlinearities cease to be a limiting factor. Future inversions with coupled physics and chemistry will need to verify that

> $\frac{\partial J}{\partial \alpha}$ has a near linear response over the time frame considered. The behaviors noted here are similar or improved across the other thirteen pairs of q and p.

9. **Page 2326: The adjoint test presented is figure 2 was compromised by some errors, as mentioned later in the paper (page 2328). So it seems logical to replace figure 2 by a corrected one.**

   The improved agreement between the ADM and TLM described on page 2328 resulted from changing single precision variable definitions to double precision in the TL version of the dry deposition velocity calculation. This changed the TLM derivatives in the 8th digit and beyond. There would be no perceptible change to Fig. 2, other than the Max Rel. Err. result. Because running these verification tests would take several weeks to repeat, and the ADM/TLM agreement is already within the bounds of a single precision calculation in Fig. 2, we believe it is unnecessary to repeat the verification.

10. **Page 2326, line 5: Q_v has not been introduced in the paper.**

    The sentence starting on page 2326, line 4 has been corrected to: "The CVs include initial conditions for BC_1, zonal wind (U), temperature (T), and water vapor mixing ratio (Q_v), and also BC emission scaling factors (α_{BC})."

11. **Page 2328, line 26: "BC concentrations respond linearly to a 1% perturbation of emissions for at least 48 h". Is there any reason that a non-linear response can be expected when coupling with BC and radiation is turned off?**

    Since the emission and transport processes are linear, we did not expect nonlinearities. However, it was not known whether discontinuities in the advection would introduce non-physical nonlinearities. We believe it is important to state this was not the case, at least for these tests. We modified the sentence beginning on

page 2328, line 25 as follows: “Most importantly for multi-day 4D-Var emissions inversions, and as would be expected, BC concentrations respond linearly to ...”

12. **Page 2331, line 28: reference missing**

    We have fixed the reference in the manuscript.

13. **Page 2332: I am a bit worried that you use two different measurement techniques for BC. BC is particularly tricky to measure and LAC and TOR might have different biases. For sure, BC and EC cannot be compared directly, because they are defined differently. Using ARCTAS and IMPROVE data in the same inversion might have to deal with a bias of one method to the other. Maybe it is important to highlight how comparable the data are. There is a wealth of literature available on bias correction of particular data streams (e.g. satellite data).**

    This comment is important, because it calls upon the relatively few comparisons that have been made between the many BC/EC measurement techniques. This is an important point to consider for any model-observation comparison, especially when utilizing multiple in-situ measurement approaches. Yelverton et al. (2014) compared SP2 BC and IMPROVE EC from TOR, in addition to eight other BC or EC measurement protocols, for a single concentration of carbonaceous particles. Those authors found that time-averaged mass concentrations of SP2 BC and IMPROVE EC agree within 7%, with the EC values being larger, and the two averages being within 2σ. Although we have not applied any correction to either observation set, the experiment performed here is meant as a demonstration. A 7% change in concentrations from either device would not change our qualitative conclusions. For future work, we will be sure to apply the correction. Additionally, refractory coatings in aged aerosol or near biomass burning sources may introduce additional measurement bias. In light of this discussion, we added the following text to the paragraph ending on page 2331:

> While we use IMPROVE elemental carbon (EC) and SP2 absorbing carbon as equivalents herein, Yelverton et al. (2014) found that the former is approximately 7% higher than the latter, but that their error bars overlap. For the qualitative analysis performed in this demonstration, bias correction would not change any of the final conclusions.

14. **Page 2335, eq. 16: Sure you divide by L^2?. Anyhow, it would be better to have the non-summed part before the summation sign for clarity. Also for eqs. 17, 20, 24.**

    The variance of the mean observation is $\text{Var}(\bar{y}_k) = Var\left(\frac{1}{L_k}\sum_{l=1}^{L_k} y_k\right)$ $= \frac{1}{L_k^2} Var\left(\sum_{l=1}^{L_k} y_{l_k}\right)$. The variance of the sum of any two variables, X and Y is (e.g., http://en.wikipedia.org/wiki/Variance#Basic_properties) $\text{Var}(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$. Since in our case, the observations are independent, the variance of the mean is then the sum of the variances, divided by the number of observations squared: $\text{Var}(\bar{y}_k) = \frac{1}{L_k^2}\sum_{l=1}^{L_k} Var(y_{l_k})$ $= \frac{1}{L_k^2}\sum_{l=1}^{L_k} \sigma_{l_k}^2$.

    In order to improve clarity, we moved the non-summed denominators before the summation symbol for Eqs. 14, 16, 17, 19, 20, and 24.

15. **Page 2338. In the discussion of the model-data mismatches and associated adjoint forcings, I miss a discussion of the role of the adjoint model. The H^T operator projects the mismatches to $\frac{dJ}{dc_k}$ and indicates how sensitive a particular observation is for a particular emission change. By only discussion the adjoint forcings and their magnitudes, I cannot see how you can defend the need for a w_k scaling in the covariance matrix. In my opinion, the need of this scaling only appears after a full inversion, and I would like to hear the authors' opinion about this.**

We agree that the paragraph on page 2338 is unclear as to the justification for applying w_k. We devised the weighting scheme based on past experience with 4D-Var, and from considering how the $\mathbf{H}^\top$ operator informs the optimization. When the adjoint forcing, $\lambda^*_{k,o}$, increases in magnitude, there is greater potential for the adjoint operator, $\mathbf{H}^\top$, to generate large sensitivities (positive or negative). Consider Fig. 9, where we show that for positive residuals, the model uncertainty dominates, while for negative residuals, the observation uncertainty dominates. This is a result of each uncertainty being proportional to the respective concentration.

The rest of the justification is given below, and replaces the paragraph on page 2338 starting with "There are several outlier negative...":

> When both the observed and modeled concentrations are small, the total variance decreases to the minimum possible value, governed by the MML and MDL. This generally happens in remote regions, where small concentrations result from some combination of small nearby sources and transport from many distant sources. If the total variance is small enough relative to the residual error, $\lambda^*_{k,o}$ will be very large, often larger than in cases with larger residual errors (see Fig. 10a). The adjoint model propagates a relatively large forcing from a small residual backward, resulting in large sensitivities to emission scaling factors. These sensitivities then translate to large emission perturbations in the optimization process.
>
> The residual errors in remote locations are likely within combined model and observation uncertainty, but the model variance at these locations is unrealistically small. The ensemble will underestimate variance at observations near low-biased prior sources due to the absence of tracer mass. The opposite may be true for a high-biased prior. The challenge then is to define the concentration uncertainty introduced by

> the model physics, independent of the magnitude of emissions, which we attempt to do with a weighting scheme. The weights are used only to inflate variance, which when very low is thought to misinform the adjoint about concentration errors. Variance reduction may be necessary for observations near high-biased sources. Also, while we apply the weights to the total variance, they could be applied to only the model portion. Here we are developing a philosophy for scaling the variances, of which the following description is but one example.

To clarify the impact of adding the weights, we also will add the following sentences after line 13 on page 2340: "This is a considerable change from the unity weights where $|\lambda^*_{k,o}|$ was as large as $200\,\mu g^{-1}m^3$ in the region between the 1σ and 2σ zones."

16. **Page 2341, discussion figure 11. Figure is unclear to me. Results are shown for anthropogenic (red?) and biomass burning (blue?) BC emissions. I understand that the black markers highlight boxes with biomass burning, but how do I see where the anthropogenic emissions are and how this relates to the blue and red colors?**

    We see where there might be some confusion. We have changed the legend to make it the figure more clear (see below). Hopefully the changes are sufficient, and you are able to understand the following. The color bar on the right indicates the magnitude of the sensitivities, whereas the shapes indicate the type of emissions. The squares apply to anthropogenic emissions, and the circles are for biomass burning sources. Black of either shape indicates a sensitivity near zero, as shown in the color bar. Because the grid cells with the largest anthropogenic and fire sources are not collocated, showing the largest magnitude normalized sensitivity in each surface grid cell does not obscure any critical information. Please indicate if any of these points are unclear.

17. **Page 2343, line 21. The authors write: "The increased burning sensitivity magnitude indicate the weighting scheme is successful at generating a cost function that is more robustly sensitive to emission perturbations." I think this is not a valid reasoning. It is not surprising that the inversion is somewhat sensitive to the settings of the physical model parameters, simply because the boundary layer scheme determines how emissions are transported in the atmosphere and how the simulated observations look like. By using a different weighting these sensitivities become more alike, but this is not a proof that the new weighting scheme is better or worse: it simply gives different results because outliers receive more weight compared to better simulated observations. Like stated before: it is unclear why the authors felt the need to deviate from Bayesian statistics. Only after a true inversion and calculation of the associated statistics (e.g. χ^2 values) one might conclude that the weighting scheme gives more favorable results.**

The inclusion of a weighting term does not deviate from Bayesian statistics, but instead potentially introduces a non-Gaussian distribution of observation error. Alternatively, and as described in the response to comment 13 above, the statistics may still be Gaussian if we are inflating variance where it is not fully captured by the model ensemble. Still, we have reworded Page 2339, beginning on line 17, as follows:

> However, care must be taken when selecting γ, β, and $\tilde{r}_k$ to ensure convergence in 4D-Var. Use of these weights may imply that residual errors do not fit a Gaussian distribution.

Page 2343, line 21 does not state the results are more favorable and we agree that such an assessment cannot be made until after a true inversion. As robust may be a strong word choice, we have made the following changes:

Abstract, page 2314, starting on line 11: "A cost function weighting scheme was devised to increase adjoint sensitivity consistency in future inverse modeling studies."

Page 2343, starting on line 21: "The increased burning sensitivity magnitudes indicate the weighting scheme is successful at generating a cost function that is more consistently sensitive to emission perturbations across multiple model configurations."

Page 2345, starting on line 10: "Results indicate that the weighting scheme is effective at generating consistent sensitivities of the cost function to emissions across multiple model configurations."

We believe adjoint sensitivity consistency is demonstrated. In the forward model simulation, both the SLAB and PX LSM schemes result in a net negative model bias. Because the anthropogenic sensitivities are net positive in both cases, and we assume the emissions are the only source of bias in the demonstration, then the burning emissions must be under-predicted. Without the weighting scheme, the two configurations disagree about the sign of burning emission perturbations that would reduce model bias. With the weights, the sensitivities between SLAB and PX LSM schemes agree that burning emissions should be increased. An inversion using the weighting scheme would not necessarily result in more correct emissions, but such an assessment could be made after respective 4D-Var inversions.

2 References

- P. Bergamaschi, C. Frankenberg, J. F. Meirink, M. Krol, M. G. Villani, S. Houweling, F. Dentener, E. J. Dlugokencky, J. B. Miller, L. V. Gatti, A. Engel, and I. Levin, "Inverse modeling of global and regional CH4 emissions using SCIAMACHY

satellite retrievals," Journal of Geophysical Research: Atmospheres, vol. 114, no. D22, p. n/a–n/a, 2009.

- M. Bocquet, H. Elbern, H. Eskes, M. Hirtl, R. ?abkar, G. R. Carmichael, J. Flemming, A. Inness, M. Pagowski, J. L. Pérez Camaño, P. E. Saide, R. San Jose, M. Sofiev, J. Vira, A. Baklanov, C. Carnevale, G. Grell, and C. Seigneur, "Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models," Atmospheric Chemistry and Physics Discussions, vol. 14, no. 23, pp. 32233–32323, Dec. 2014.
- D. K. Henze, A. Hakami, J. H. Seinfeld, and others, "Development of the adjoint of GEOS-Chem," Atmospheric Chemistry and Physics, vol. 7, no. 9, pp. 2413–2433, 2007.
- J. Thuburn and T. W. N. Haine, "Adjoints of Nonoscillatory Advection Schemes," Journal of Computational Physics, vol. 171, no. 2, pp. 616 – 631, 2001.
- T. L. B. Yelverton, M. D. Hays, B. K. Gullett, and W. P. Linak, "Black Carbon Measurements of Flame-Generated Soot as Determined by Optical, Thermal-Optical, Direct Absorption, and Laser Incandescence Methods," Environmental Engineering Science, vol. 31, no. 4, pp. 209–215, Apr. 2014.
- X. Zhang, X.-Y. Huang, and N. Pan, "Development of the Upgraded Tangent Linear and Adjoint of the Weather Research and Forecasting (WRF) Model," Journal of Atmospheric and Oceanic Technology, vol. 30, no. 6, pp. 1180–1188, Feb. 2013.

Interactive comment on Geosci. Model Dev. Discuss., 8, 2313, 2015.

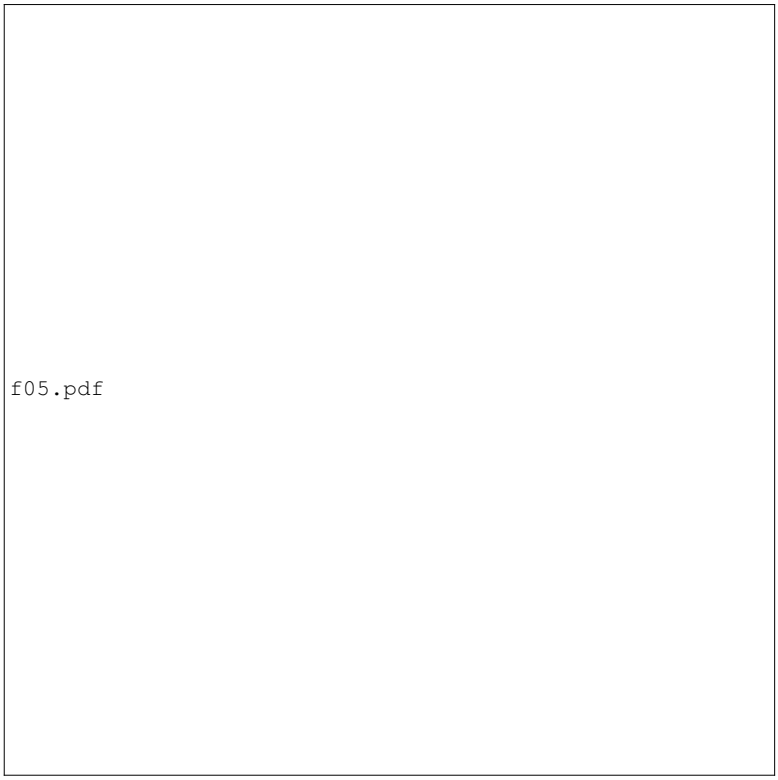

Fig. 1. Fully normalized time variant sensitivities calculated with the TLM with second order checkpointing and with multiple finite difference perturbation sizes. Each plot is for a single pair of source and receptor locations, q and p.

f11.pdf

Fig. 2. Normalized sensitivities ($\frac{\partial lnJ}{\partial lnE_{i,j,d}}$) of the 4D-Var cost function (for surface and aircraft observations) with respect to anthropogenic and burning emission scaling factors overlaid on MODIS Aqua true color images for six days during the simulation. Anthropogenic sensitivities with magnitudes less than 1 % of the maximum anthropogenic sensitivity magnitude are removed. There is a marker for all grid cells with non-zero burning emissions.

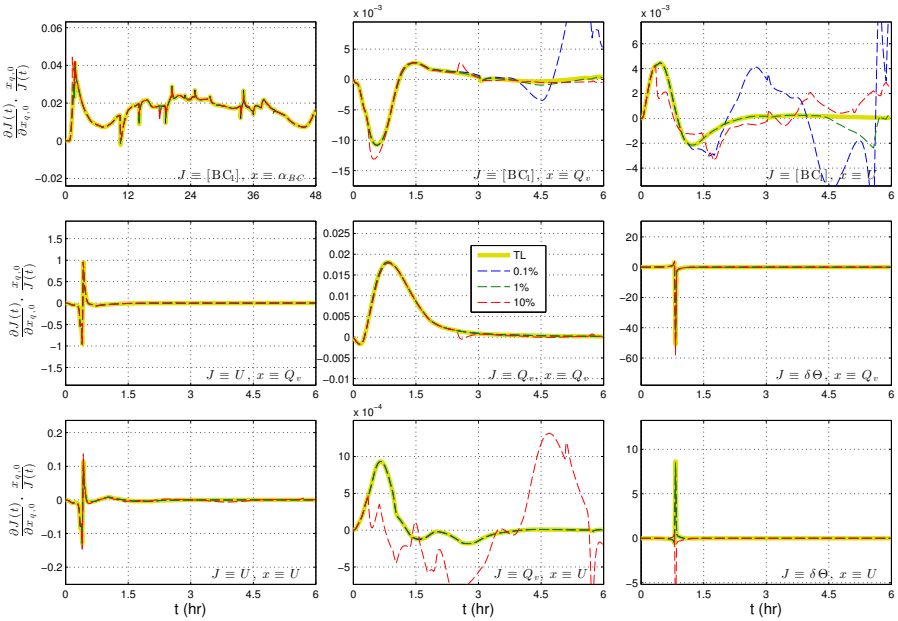

Fig. 3. Fully normalized time variant sensitivities calculated with the TLM with second order checkpointing and with multiple finite difference perturbation sizes. Each plot is for a single pair of source an

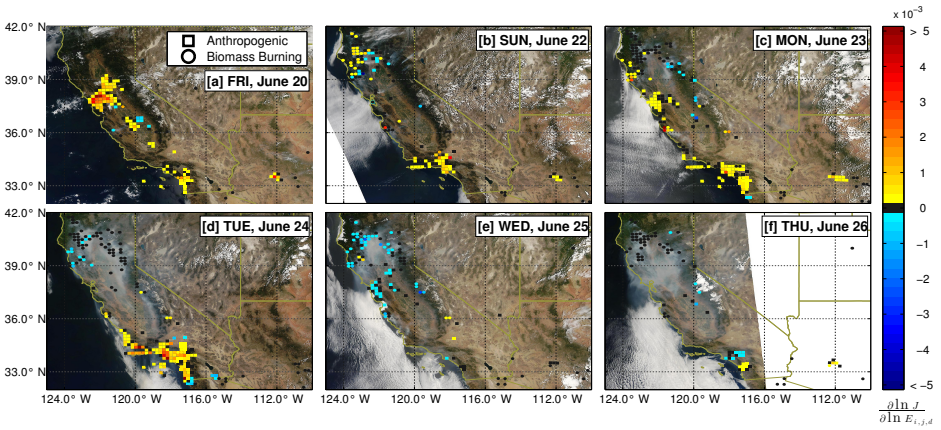

Fig. 4. Normalized sensitivities ($\frac{\partial \text{ln }J}{\partial \text{ln }E_{i,j,d}}$) of the 4D-Var cost function (for surface and aircraft observations) with respect to anthropogenic and burning emi