Anonymous Referee #1
General comments:

1) The introduction includes quite a bit of information related to offline and online meteorology. However, I have a hard time understanding why the online system was chosen if the feedback of CO2 to the meteorology was ignored for the convenience. Why not instead choose development on an offline regional model? Also, will the system be updated periodically to account for the regular updates in the WRF model family?

Regarding the rational for developing on-line chemistry transport based 4DVar system without considering the CO_2 feedback to meteorology.
***First**, a major benefit of using an online chemistry transport model is that it provides meteorology fields at much finer grid spacing and time interval. For high resolution regional inverse modelings, CO_2 vertical transport is primarily driven by resolved vertical wind velocity, instead of parameterized physical schemes. Offline models, driven by archived analysis or meteorology model output with much larger time interval, have difficult simulate chemistry transport fine grid spacing.*
***Second**, in WRF system, CO_2 can impact meteorology fields through the radiation schemes (longwave/shortwave). As a regional CO_2 inverse modeling system, WRF-CO2 4DVar is designed to run in short period time (hours to weeks) for constrain emission flux with observation data. For such time span are insignificant in most cases. For the applications WRF-CO2 4DVar is designed for, including CO2 impacts on meteorology will require a large amount of code development while offering limited performance improvement. This can be done in the future, but it is beyond the scope of the present paper.*

Regarding the future WRF system update:
We do plan to keep developing and updating the WRF-CO2 4DVar system. In addition to inclusion of observational operators for application with real observation data, we do plan to periodically update the system to keep up with the WRF system. We note that many updates of WRF system are not relevant to WRF-CO2 4DVar system, such as those for physical schemes and chemical mechanisms not used in CO_2 transport. Because WRF-CO2 uses a subset of WRF system, we only need to keep updating those relevant procedures, most of which are in the dynamical core (for advection and diffusion), convective chemistry transport (in chemistry module), and planet boundary layer schemes that treat chemistry transport (in physics module). The amount of work required for these updating are manageable for us.

2) As the authors noted, meteorology is critical to the quality of CO2 transport. Throughout the paper, I am surprised that there is no evaluation of the WRF meteorology, but this could be easily done considering the numerous observations available within the CONUS domain. In fact, I do have some concerns on the WRF setup as in specific comments 2)-3) below.

We agree with the referee that accurate meteorology simulation is of critical importance: error in meteorology will lead the inverse system to mistakenly assign transport error to fluxes sources. Thus it is imperative to ensure the quality of the meteorology field when applying the system to invert real observational data.

At the present stage, WRF-CO2 4DVar system does not include any observational operator (and their TL/AD counterparts), thus it is not ready for applying real observational data yet. The objectives of the present paper are to (1) develop and test the accuracy of the tangent linear and adjoint models, and (2) to implement the two iterative optimization schemes and test their effectiveness with synthetic data

(pseudo observation). In such pseudo observation based tests, both the observed and simulated CO2 are generated by the same meteorology but different CO_2 flux through scaling factor).This setup ensures that meteorology is error-free, and no transport error is present in the inverse system.

With this said, we completely agree with the referee that meteorology must be evaluated before the system is used with real observations. In response, we conducted comparison of the meteorology simulated by WRF-CO2 forward model against CFSv2. Since CFSv2 is an analysis which assimilated a large amount of quality controlled observations, it can be used in lieu of observational data here.
The inverse experiments described in the manuscript span the 24 hours starting at 2011-06-02 00:00 UTC. We interpolated CFSv2 to the WRF grid, and compare the two datasets at et 6-hour interval.

Figure 1 shows sea level pressure and horizontal wind at first vertical level from WRF (left column) and CFSv2 (right column).
Figure 2 shows horizontal wind and geopotential at the 30^{th} vertical level.

These comparisons indicate WRF simulated meteorology is close to the analysis and is valid for the purpose of the pseudo observation based tests used in the present paper.

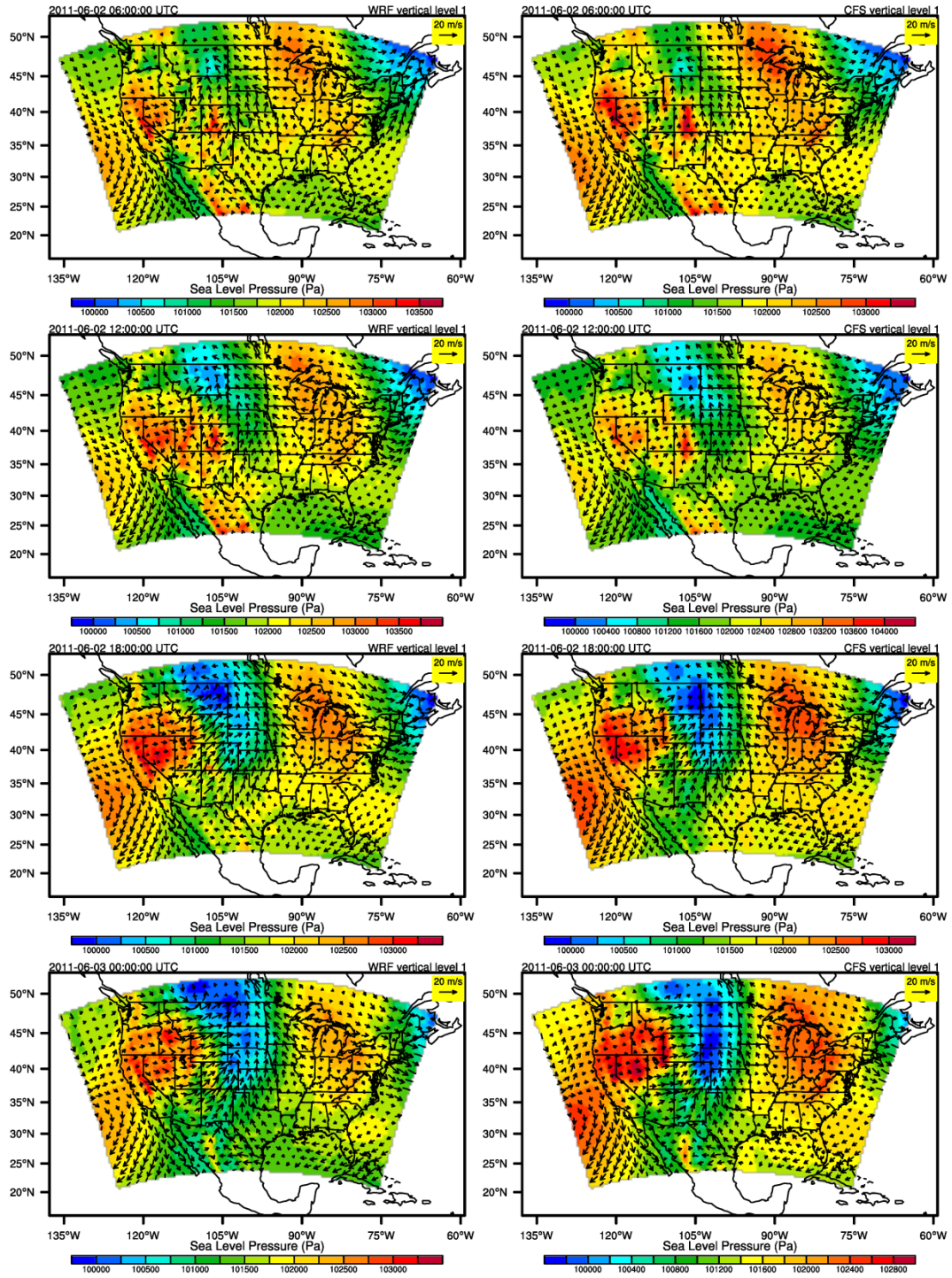
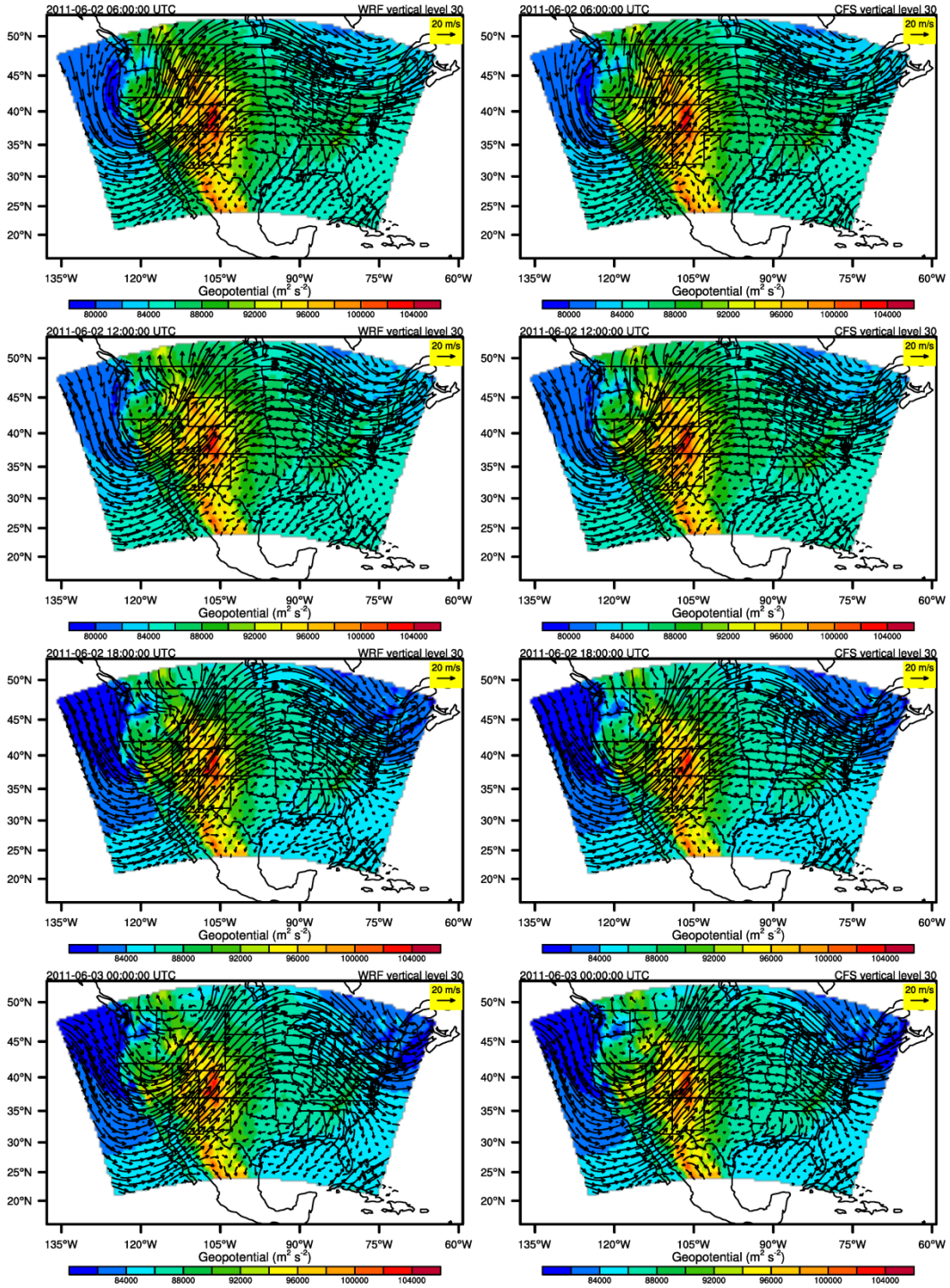


Figure 1. Sea level pressure (background) and horizontal wind at the first vertical level (arrows) as simulated by WRF (left column) and interpolated from CFSv2 (right columns). The four figures are plotted at 6-hour interval for the 24 hour simulation period (2011-06-02 00:00 UTC to 2011-06-03 00:00 UTC).



1)Figure 2. Geopotential and horizontal wind at the 30th vertical level (arrows) as simulated by WRF (left column) and interpolated from CFSv2 (right columns). The four figures are plotted at 6-hour interval for the 24 hour simulation period (2011-06-02 00:00 UTC to 2011-06-03 00:00 UTC).

3) The comparison of L-BFGS-B and the CG has been done before. It'd be good to relate the prior results to yours and highlight any unique findings from your work.
In Section 4of the revised manuscript, comparison between L-BFGS –B and Lanczos-CG regarding memory and computation cost are added, and related to previous research findings (Guerrette and Henze 2015).

4) Information on computation requirement and cost would be helpful.
Detailed information about memory requirement and walltime are added in Section 4 of the revised manuscript.

Additionally, here are some places where clarifications or corrections are needed:
1) The cost functions, etc, are not quite consistent with literature on the similar topic. Vectors should be in bold.
We accidently dropped the transpose operator equations 3,4, and 6. These are fixed. All vectors in equations and inline text are in bold now.

2) Page 9, Line 26: Was indirect soil nudging enabled when PX LSM was used? It is recommended to enable it in retrospective analysis because little testing has been done for running PX with the indirect soil nudging disabled. See the WRF users guide and literature.
The indirect soil nudging was not used in the experiments described in the original manuscript. We greatly appreciate the referee brought it to our attention. All the new simulations reported in the revised manuscript were conducted with the indirect soil nudging activated .

3) Page 10: Met IC/LBC from CFS on which resolution? Potential problems of down-scaling that to 48km should be discussed. Again, some model evaluation should be added.

CFSv2 analysis data of 1x1 degree horizontal grid are used to generate the meteorology initial and boundary condition. For model evaluation, please refer our response your general comment #2, and Figure 1 and 2 of this document.

Potential problem associated with downscaling should be added.
For purpose of model development and testing, the simulation matches well with analysis data (Figure 1 and 2) and the meteorology are valid for testing model accuracy and inverse modeling test. We do note that more care should be exercised when WRF-CO2 4DVar is used with actual observation data and potential transport error.

4) Page 10: any biomass burning emissions included? Does daily emission include any diurnal variability? Please include the emission amount in Section 3.1, to help understand the figure and results in Section 3.4.
Biomass burning emission:
In the simulations described in the original manuscript, biomass burning emissions were not included. In the revised manuscript, we switched from using EDGAR/CASA emission flux to CarbonTracker optimized fluxes. All four fluxes (fire, biosphere, fossil fuel, and ocean) are used. So, biomass burning

emission is included in the simulations reported in the revised manuscript.

Diurnal variability:
Because CarbonTracker fluxes are of 3-hour interval, diurnal variability is included. However we need to point out that in the inverse experiments, emission scaling factor is applied to the mean daily value of biosphere flux. This means no diurnal variability in the inverse experiments. Our inverse experiment is constrained by error-free pseudo observations, which is an ideal configuration to prove that WRF-CO2 inverse framework works. We acknowledge that diurnal flux variability probably should to be included in application with real observations, which will require an different setting for the scaling factor (such as separate scaling factors for photosynthesis and respiration). As these considerations are beyond the scope of the present paper, they are not addressed here.

Emission amount:
Figure 3 (below) shows the sign and magnitude of daily mean biosphere flux used in all the simulations reported in the revised text.

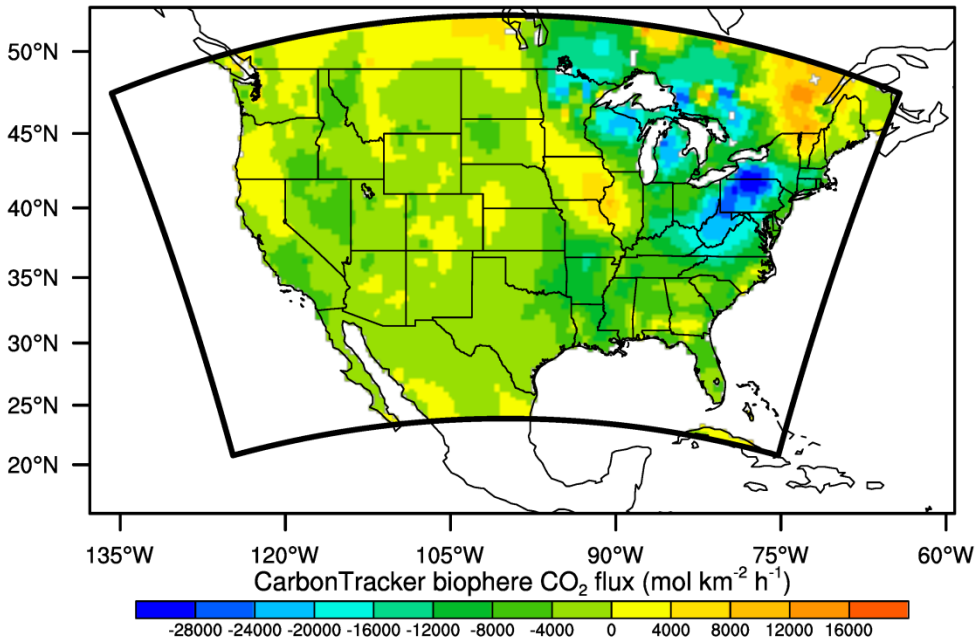


Figure 3. Daily mean CarbonTracker biophsere CO2 flux used in calculating sensitivities and inverse modeling experiments. The daily mean value is calculated as the arithmetic mean of the 3-hourly fluxes from 2011-06-02 00:00 UTC to 2011-06-03 00:00 UTC.

5) Page 10, line 7-8: More details on the global WRF-Chem simulation is needed. I assume it was done on a coarser resolution than 48 km. Would the sensitivity and other tests in the global domain differ much from the regional results? Some simple comparison can highlight the benefit of using a regional scale 4dvar system.

The global WRF-chem simulation was conducted with a 256 (east-west) by 128 (north-south) grid, which is about 156x156 km in horizontal. The WRF global model code includes polar filtering procedures for the converging latitude at high latitudes. These filtering procedures are located in multiple modules in the dynamical core. Because they are only needed for global model, we did not

develop their tangent linear and adjoint code for WRF-CO2 4DVar (which is designed for regional inverse only).This mean we can not run WRF-Chem for an global domain inversion to compare with WRF-CO2 4DVar in regional inversion.

*In the new experiments reported in the revised manuscript, chemistry initial and boundary conditions are **not** obtained by global WRF-Chem simulation anymore, but are from CarbonTracker CO2 mole fraction (global 3x2 degree, 3-hourly interval product, 2016 version).The interpolation from CarbonTracker grid to WRF is described in Section 3.1of the revised manuscript.*

6) Figure 1 and 2 need to be cited in the text. As red is already included in your emission color scheme, I suggest using a non-red color to show the locations of towers sites in Figure 3.
The optimization scheme diagram in Figure 1 and 2 are now cited in the text. In addition both diagrams have been improved for clarity in response to the second referee's comments.

Figure 3 are redrawn. In response to the second referee's comments, the sensitivity (TL/NL/AD)tests in the revised manuscript are done at the 20 towers (as receptors) and a different set of 35 locations (as sources). For clarity, we split the original Figure 3 to two separate figures: the Figure 3 marks the source and receptor locations and Figure 4. show the biosphere flux magnitude (please refer to our response to your comment #4

7) It'd be good to show Figure 5 and 6 along with trajectory calculations as in some prior works.

Our understanding is that trajectory calculations are carried out using Lagrangian particle dispersion models, such as LPDM, FlexPart, or Hysplit, in backward trajectory mode. As an adjoint based inverse system, WRF-CO2 4DVar does not calculate backward trajectory (it directly computes the product of the Jacobain with a forcing vector). So, we can not plot trajectory calculation. We did improve the footprint (adjoint sensitivity) plot by adding the results from receptors placed at the 10^{th} vertical level (in addition to the 1^{st} level).

Some typos and grammar errors:
1) Page 5, line 19: according -> according to
Fixed.

2) Page 12, line 28: facotr -> factor
Fixed.

3) Page 37, Table 5 caption: givne ->given
Fixed.

Response to Anonymous Referee #2

The manuscript, "Development of the WRF-CO2 4DVar assimilation system", describes exactly that. The authors have created linearized model versions of the WRF-Chem CO2 transport mechanisms, validated their performance against finite difference approximations, and demonstrated their utility in a simplified pseudo-data experiment. The introduction covers much of the relevant literature necessary to get to the same starting point as the authors, and we make a few additional suggestions below. Theadjoint and tangent linear model developments are described thoroughly, and would be helpful for any person working their way through the code at a later time. The adjoint model evaluation falls a little short, and we provide some suggestions for ways it could be improved. The pseudo-data inversion test, while quite simple and unrealistic, demonstrates that the inversion framework is working. It is a first step that undoubtedly took considerable effort, but needs some improvements in the application of the new tool. There is no discussion of the statistical nature of 4D-Var, which is paramount to that method's success with real data and its being labeled a Bayesian inversion technique. We have several specific comments as to how the discussion could be mademore precise and also miscellaneous technical corrections.

We thank the referee for the time they have taken to improve the paper with their insightful and detailed comments. Below is a summary of the major work conducted in response to the referee's comments. The point-to-point response is in the following sections.

- Model code debugging: as the referee point out, the adjoint model was not error free. To address the problem, we systematically debugged the model code. Errors were isolated, identified and corrected. The evaluation through sensitivity calculation confirms that the three model components (NL/TL/AD) match as expected.

- Optimization experiment: in the revised text, synthetic observation data are from 30 vertical levels from bottom up. They were from the bottom level only in the original text.

- In the footprint calculations, receptors are now placed at 1^{st}, 5^{th}, and 10^{th} . They were placed on the 1^{st} level only in the original text.

- TL/AD/FD sensitivity comparison. (1) tangent linear, adjoint, and finite difference sensitivities are calculated for source and receptors cells at different locations in both horizontal and vertical). (2) The receptor cells are placed at the 1^{st}, 5^{th}, and 10^{th} vertical levels at each tower site.

- Cumulus activity indication: extra variables are implemented in the model to track when/where the convective tracer transport is activated during the simulation. This information is plotted and used to ensure there are sources and receptors located within or near the cumulus activity. As the referee pointed out, this is necessary to evaluate the accuracy of the newly developed TL/AD code of the convective chemistry transport scheme (module_ctrans_grell).

Major Comments:

• Section 2.3: Incremental 4D-Var is used to optimize nonlinear systems. But the CO2 tracer simulation is inherently linear. Thus, I don't really see at this point what the benefit would be of an incremental formulation, nor how updating the inner loop with an outer loop integration would provide any additional information. Thus, the importance of including both of these methods and their comparisons for CO2 inversions needs further justification.

We thoroughly examined the model code for the linearity pointed out by the referee. We found that WRF-CO2 is linear except at certain situation when positive definite chemistry transport is used. Our examination shows that the ACMPBL and convective transport (ctrans_grell) are both linear with respect to CO2. With positive definite chemistry advection, nonlinearity can be introduced when the predicted minimum possible CO2 is negative at a grid point. This will trigger a renormalization procedure, which is nonlinear. We confirmed this nonlinearity through examining finite difference sensitivity around grid point where the above mentioned renormalization is artificially triggered. In order to trigger the renormalization, we created large horizontal CO2 gradients. We do not believe such large CO2 gradient is very common in nature, but may be possible for wild fire emission. In the 24-hour simulation used in this paper, the renormalization was not triggered in any grid cell, thus the system is linear.

The outer loop updating used in the original text was necessitated by the error in the adjoint model code, and we mistakenly attributed it to the loss of conjugacy. We greatly appreciate the referee's insightful comment. With the corrected adjoint model, Lanczos-CG based incremental optimization does not need the outer loop update: only one outer loop iteration is applied in all inverse modeling experiments.

Our inverse experiment results (with the corrected adjoint model, and observation at 30 vertical levels) show that Lanczos-CG converge substantially faster than L-BFGS-B. Although we are aware that this performance difference may be specific to our experiment setup, we consider it is necessary to include both optimization schemes in WRF-CO2 4DVar for future applications.

• Fig 8: As the authors surely know, adjoint sensitivities should agree with the tangent linear sensitivities to numerical precision. The differences between these sensitivities vs the finite difference sensitivities, given that the latter match the tangent linear sensitivities, is an indication that the adjoint model is not error free. For the cases tested here, the errors are manageable, yet there is no guarantee that the errors would not grow for longer simulations. The authors should thus continue to debug their adjoint code, possibly by performing this type of test around the tangent linear and adjoint code of individual physics components developed here (such as ACM2 PBL mixing). If they can not resolve the code bugs this way, they should at least perform additional tests using different receptor locations and simulations of increasing (and decreasing) length to examine how the numerical errors may be accumulating.

Yes, we are aware that adjoint model was not error free. We thank the referee for reminding us to correct it. Following the referee's suggestion, we debugged the individual processes in isolation. We also modified the code to test all three models in a single time step mode. Code errors were identified and corrections were made. We evaluated the updated code by comparing the adjoint sensitivity against the tangent linear and finite sensitivities. As suggested by the referee, these sensitivities are calculated with sources and receptors at different grid cells and the receptors cells are placed at multiple vertical levels. (See Fig. 2 next page for the source and receptor placement. We also ensured that there are sources and receptors placed within or near cumulus activities for testing the convective transport code. (See Fig. 3 for cumulus indicator).

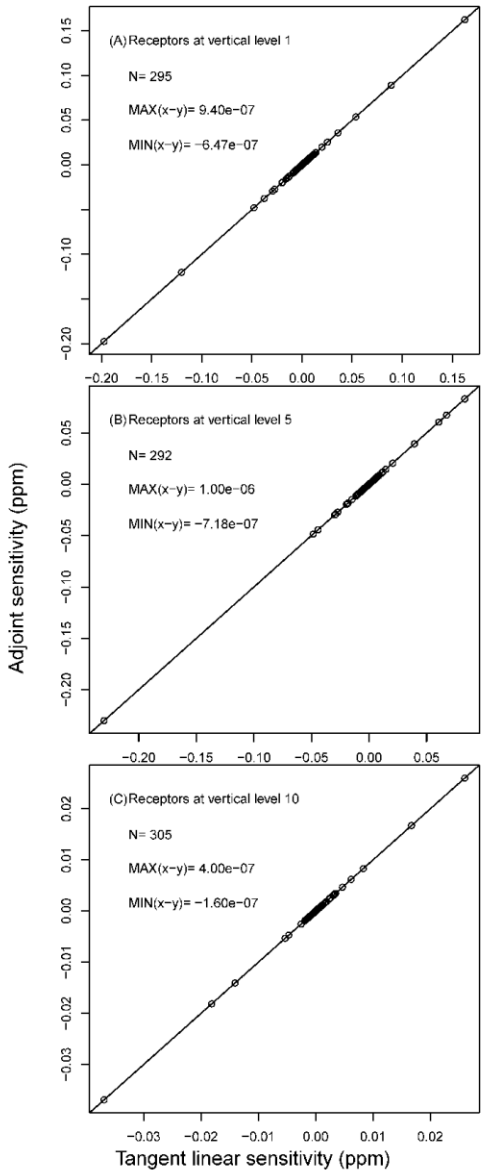


Figure 1. Comparison between tangent linear sensitivity (x-axis) and adjoint sensitivity (y-axis). The sensitivities are organized into three groups for receptors placed at the 1st, 5th, and 10th vertical levels. Because there 35 sources (red starts Fig. 2) and 20 tower sites (red triangles in Fig. 2), there are 700 (35x20) pairs of tangent linear and adjoint sensitivities for each group. Sensitivities with absolute values less than 10^{-10} are not included in comparison, resulting in 295, 292, and 305 pairs of comparisons. The solid lines in each figure are the 1:1 lines. Because the values are very close, we summarize each figure by the minimum and maximum difference between the tangent linear and adjoin sensitivities, instead of the slope and r-squared.

• p. 12, line 7: It is not required that a grid cell be both a receptor and a source to have non-zero sensitivity, as evidenced by Figures 5 and 6. The source and receptor must simply be significant (large source and large perturbation to concentration due to that particular source). Choosing them to be the same grid cell is likely to produce good agreement between the adjoint and finite difference methods even when the advection, PBL, or convective transport adjoint may be incorrect. *Did the authors choose identical grid cells for source and receptor?* If so, then additional tests are required to prove that the adjoint code works as described. Additionally, validating the convective transport adjoint and tangent linear codes requires demonstration during a period when the subgrid cumulus parameterization is active and for sources and receptors near that phenomenon. The authors should present some indicator of cumulus activity in an additional figure.

Placement of the source and receptor cells: In the original manuscript, the source and receptors were the same grid cells. They were the bottom level grid cells where the 20 towers are located.

We agree that code should be tested with sources and receptors placed at different grid cells as long as there are discernible impact during the simulation period. To address this, we conducted the sensitivity calculations with the updated model in a more systematic approach: (1) Receptors are still placed at the 20 tower sites in horizontal, but at each tower site, 3 receptors are placed at different vertical levels: level 1, 5, and 10. (2) Sources are at a different set of 35 grid cells placed around the receptors. The placement of sources and receptors used in the revised manuscript is in Figure 2 below.

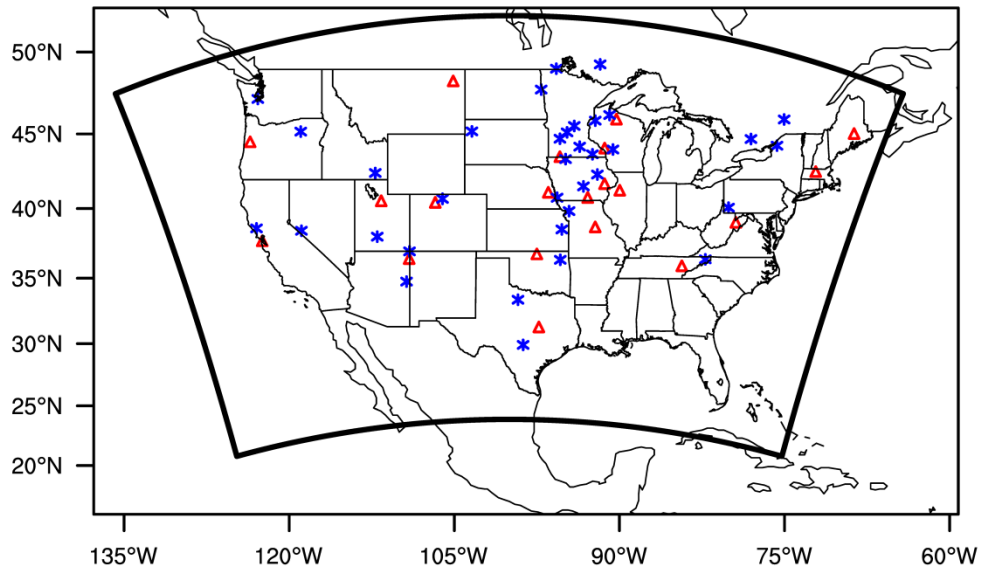
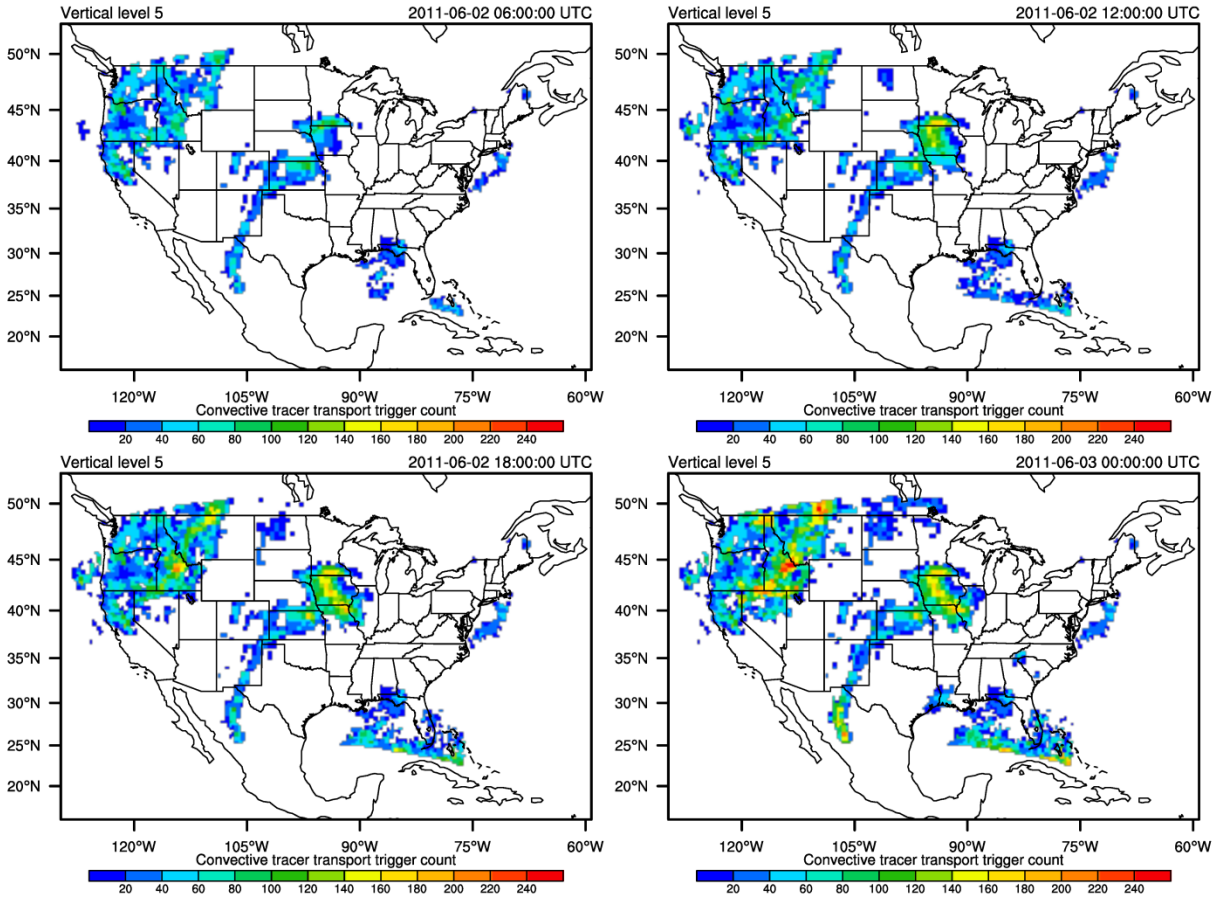


Figure 2. Placement of the sources (blue stars) and receptors (red triangles). Receptors are the placed at 1st, 5th, and 10th vertical levels of grid cell of the 20 towers (Table 4 of the manuscript). Being the surface flux source, all sources are place at 1st vertical level.

Indicator of cumulus activities:

To ensure some sources and receptors are placed within or near cumulus activities, we implemented an counting mechanism within the convective transport code (module_ctrans_grell.F in the chemistry directory). In WRF, whether chemistry species at each grid cell is vertically transported at a given time step by the cumulus process is determined by a number of tests and marked by a pair of flags (one for deep convection and another for shallow convection). These two flags are reset at each time step. We added two variables to track the two flags across time steps: each time the convective transport process is triggered at a grid cell, its count increases by one. We refer the new variables "convective tracer transport trigger count" and used it to examine the cumulus activities. Comparison of Figure 3 (below) and Figure 1 confirms there are sources and receptors placed within or near cumulus activities.



Fiure 3. Convective tracer transport trigger count plotted at 6-hour intervals at vertical level 5. These counts are vertical level specific as the convective transport activations are determined for each vertical level. The figures show that deep convections are trigger during a large portion of the 24-hour simulation at the Pacific northwest and Midwest (center round Iowa). A comparison with source/receptor cells placement in Figure 1 confirms that there are

sources/receptors placed around cumulus activities, and thus tangent linear and adjoint code of the convective transport are indeed tested.

• p. 12, lines 16-17: Similar to the comment above, choosing the observations to be in the lowest model layer reduces the importance of having the adjoint and tangent linear treatments of vertical mixing. The observations should be spread more thoroughly, vertically.

We agree with the referee.
To address this issue, new pseudo observation data are generated from the forward model run at the first 30 vertical levels (out of a total of 50 levels). This means that there are 30 observations at each horizontal grid. We emphasize this setting is for the sole purpose of testing the inverse system with ideal error-free synthetic data. The inverse experiment results are shown in Fig. 10-13 of the revised manuscript.

• From a software perspective, I'm a bit confused about the distinction between the present work and that of GH15/16, where the adjoint of the BC tracer is developed. So, essentially the update here is that BC has been changed to CO2, and convective transport has been added? On a broader note, would it be beneficial to the community to view these as two options within a single chemical 4D-Var system, rather than as two different models? I realize this likely results from development of these systems over time, in parallel, but thinking to the future I wonder if a model merge would be in order. To illustrate my point, imagine if rather than a consolidated WRF-Chem model, we had separate WRF-Chem-BC, WRF-Chem-CO2, WRF-Chem-CO, . . . etc. models. That would clearly hinder development of the tool as a whole, which shares many common elements across the different tracer simulations.

Yes, we developed WRFCO2 4dvar system as part of an effort proposing to NASA's carbon science program. Our development is in parallel with G15/16, and no collaboration has been involved yet.
We agree with the referee that coordinated code development will benefit the community. We will contact Dr. Henze's group for collaborating on future code consolidation/merging.

Further Comments:
1. p. 2, line 33: The reference to Streets 2013 is a bit odd, as that paper is specifically a review of remote sensing based constraints on emissions and focuses mainly on reactive trace gases and aerosols. As the present work doesn't seem geared towards remote sensing observations, some other references to literature on regional CO2 inverse modeling would be a better fit here.

This reference has been removed.

2. p. 2, line 9: Probably more correct to say "instead they directly compute the product of the Jacobian with a forcing vector, which is the gradient used for optimizing

the state vector."

This sentence has been corrected following the referee's suggestion.

3. p. 2, line 10: The notion that posterior error can not be calculated from a variational inversion is outdated. Posterior error can be calculated analytically using the Lanczos vectors from a CG minimization in the incremental 4D-Var framework following Fisher and Courtier (1995) - as currently done in operational weather forecast centers such as ECMWF and the UK Met Office - for minimal additional cost. Efficient posterior error estimate for non-incremental 4D-Var frameworks are described in Bousserez et al., QJRMS, 2015, including previous works on Monte Carlo (e.g., Chevallier et al., JGR, 2007) and stochastic (Rabier and Courtier, QJRMS, 1992) methods.
Fisher M, Courtier P. 1995. Estimating the Covariance Matrices of Analysis and Forecast Error in Variational Data Assimilation, Technical Memorandum 45. ECMWF: Reading, UK.

We thank the referee for correcting us. This statement about posterior error calculation has been corrected accordingly.

4. The introduction states that both bottom-up and top-down approaches are used, but does not say why that is the case. It is recommended to move the first paragraph of Section 4 to the introduction.

The related paragraphs have been rearranged following the suggestion.

5. p. 2., line15: The GEOS-Chem CO2 4D-Var system is also part of JPL's Carbon Monitoring System, e.g. Liu, Bowman, Lee, et al., Tellus B, 2014; Liu, Bowman, and Lee, JGR, 2016.

Thanks for pointing this out. The text has been modified accordingly.

6. p. 2, line 18: Also Chevallier et al., JGR, 2007.

Thanks. This has been fixed.

7. p. 2: 2: For regional CO2 inversions, the list isn't entirely complete, see also Alden, Miller, Gatti, et al., Global Change Biology, 2016; Chan, Chan, Ishizawa, et al., GMDD, 2016. There are others, but I think it suffices to say the literature review could be a bit more comprehensive (or, alternatively, scoped / phrased as to be more narrow).

Agree. The regional CO2 inversion review has been strengthened with the following the two additional literatures.

8. p. 2, line 33: Here and in several other places, the authors use the phrase "influence

function" without every having defined it.

Definition of the influence function has been added.

9. Equations 3,4: define superscript n in this context.
Definition for superscript n has been added.

10. Equations 6,7: I understand incremental 4D-Var, but I think still the authors should rigorously define the superscript n in this context for the sake of completeness, which I believe should differentiate between inner and outer loops. Also, incremental 4D-Var is usually employed with a square-root preconditioning, which I don't see here.

Text has been added to define the superscript n and how it changes differently within the inner and outer loop of the incremental optimization.
The square-root preconditioning is not applied in the pseudo-data based inversion experiments. As explained in Section 3.4 of the manuscript, background error matrix is set to infinity and observation error matrix is set to the identity matrix. This is realized in the code by setting the cost function equal to the observation cost function, and the cost function gradient to the observation cost function gradient. For this setup, we believe the preconditioning does not need to be applied.

11. Fig 1: The way that the observations fit into this diagram doesn't make sense, since currently it implies the arrow coming out of the right side of the Simulated box passes information both to the right and left. Some separate arrows from the Observation box seem to be needed.

Two separate arrows has been added out of the modeled and simulated box to avoid the possible confusion in data flow direction. (See Fig 1 in the updated manuscript.)

12. Fig 2: Despite the caption, this doesn't really show how the CG method is implemented to anyone already not familiar enough with incremental 4D-Var to know that it likes between the "no" and Tangent linear model and involves an updated estimate to the (preconditioned) increment.

We added to two addition boxes (for residual vector and updated gradient respectively) between the tangent linear and adjoint model. This helps explain the data flow between the the TL an AD models in the inner loop. Also the 'exit' box is changed to 'inner loop converged' box to emphasize the condition to exit the inner loop. (See Fig. 2 of the update manuscript.)

13. p. 3, lines 11-13: This sentence is grammatically incorrect, the phrasing is confusing, and the conclusion is drawn weakly. What is the "potential" that online transport based inversion systems have demonstrated? At a bare minimum, add a reference (e.g., Grell and Baklanov, 2011) or remove that statement.

We chose to have the statement removed.

14. p. 3, line 18, p .6, line 3, and p .6, line 23: The original reference for WRFPLUS is Xiao et al. (2008) [DOI: http://dx.doi.org/10.1175/2008MWR2235.1]. The version you use (v3.6) includes the work by Zhang et al. (2013) [mentioned elsewhere in your text], and should be included in these references. Huang et al. (2009) specifically used WRFPLUS for 4D-Var in WRFDA, but did not develop WRFPLUS. Barker et al. (AMS, 2005) originally developed WRFDA (for 3D-Var). These two latter references should be used as references for WRFDA. Barker et al. (2012) is an update on software development for WRFDA, and doesn't even mention WRFPLUS. The appropriate references need only be given at the first mention of these particular models, and do not need to be repeated throughout as references to the entire model. The exception is when discussing a particular aspect of those works.

We really appreciate the referee to clear this up for us. The references are fixed.

15. Throughout, the author should use bold characters for vector notation (i.e., x, y, q, kco2). These would be particularly illustrative on p. 10, lines 21, 24, and 31 to indicate whether the denominator or numerator of @qco2/@kco2 is a vector**.**
All vectors in the equations and inline text have been changed to bold face characters. At place when a symbol can be either vector or scalar, it is kept as non-bold character.

16. p. 5, line 7 and p. 14, line 23. You mention that L-BFGS-B can be used to calculate the posterior covariance, which is true although robustness of this approach with regards to the initial inverse Hessian estimate is an issue when using this algorithm, see Bousserez et al., QJRMS, 2015. As mentioned elsewhere in our review and this manuscript, Lanczos CG can also be used to estimate posterior error due to the eigen decomposition (well documented). Thus, the ability to calculate posterior error estimations is not a valid distinction between these two. Further, calculation of posterior error is not included for either method in this work. So while this could be mentioned in the introduction or discussion of future work, the methods section should only refer to methods that are actually used in this work or ones that provide reasoning for why you used a particular approach.

We thank the referee for correct us on this issue. The text related to posterior covariance calculation between the two optimization schemes have been removed from the manuscript.

17. p. 5, the term "cost function" is used 13 times on this page alone. It is suggested to reduce "cost function gradient" to "gradient".
"Cost function gradient" has been change to "gradient" through the text except where full term is needed to avoid ambiguity.

18. A comparison of Lanczos CG and LBFGS-B based solely on cost function reduction and RMSE is not sufficient. The authors should be more instructive and explicit as to the tradeoffs between them. In regards to p. 6, line 20, and p. 14,

line 21: How much less memory does Lanczos CG require for your particular application, as a percentage? Is that a good reason for choosing it over L-BFGS-B in this case? Are there ways to reduce the memory requirements of each? The most accurate Lanczos CG algorithm requires storing all basis vectors and performing full reorthogonalization after each iteration. Do you include that step? If not, why? *This is a salient topic*, since you discuss the loss of conjugacy later in the manuscript. Also, how can Lanczos CG be adapted for parallel computation in a way that differs from L-BFGS-B? The name of an algorithm or a reference should be included. What are the respective wall-clock times of the two methods? Lastly, It's also not clear why one is more amenable to parallel programming than the other (p. 14, line 21), as both are sequential techniques, unless that is strictly a consequence of the aforementioned memory requirements.

(1) Regarding the comparison between L-BFGS-B and Lanczos CG, we added discussion about the memory requirement, and parallel implementation related issue in Section 4.
(2) Regarding the reorthogonalization. Yes the reorthogonaliation is implemented in the code used in the original optimization experiment. The referee is correct that the degradation of the Lanczos CG with increased inner loop iteration was not caused by the loss of conjugacy, but the error in the adjoint model (and thus the calculated gradient vector). After we corrected the adjoint model, the need for the second outer loop does not exist anymore.
(3) Walltime used in our experiment with the two schemes are documented and added in the text.
(4) References for the L-BFGS-B Fortran code (Algorithm 788) compiled in WRF-CO2 4DVar is added.

19. p. 6, line 20: Lanczos CG provides approximations of both the leading eigenvalues and eigenvectors (eigenmodes), not only the former.

Thanks. This statement has been corrected.

20. p. 7, lines 8-9: Does VPRM calculate fluxes at the grid-scale in every time step? You can scale fluxes whether they are provided online or offline.

We agree with the referee that the fluxes can be scaled whether they are from offline data files or calculated by online model (VPRM). Because running the VPRM model requires additional datasets (satellite derived vegetation indexes and land cover classification maps) and some parameter tuning, we choose to use offline CarbonTracker CO2 fluxes instead. Both methods are valid, but using the offline files allows us to focus on the core code development by avoiding some extra input data preparation.

21. p. 7, line 15: So a tagging scheme for source specific CO2 has been implemented as well? This might present an interesting feature for testing the adjoint sensitivities and 4D-Var system, or performing low-dimensional analytic inversions.

The present model code does not include tagging scheme for specific CO2 sources. But as the chemistry 4d variable in all three models (NL/TL/AD) includes separate variables for each individual fluxes, tagging scheme can be implemented with some minor code modification in the future development.

22. p. 7, line 32: Well, to be more precise about the wording, the adjoint is just the backward sweep, although for nonlinear systems it would need information from the forward sweep. I'm not sure what that information would be though, for a linear CO2 simulation.

The statement about the adjoint and forward/backward sweep has been revised.
Regarding the referee's question about what information from the forward sweep would be needed by the adjoint (the backward sweep): (1) Meteorology state variables: In the dynamical core, advection and diffusion of chemistry species are carried in each of the three sub-steps of the Runge-Kuta loop. At a given time step, at the start of the backward sweep, only the meteorology state variables at the last sub-step is available while all three all needed. This requires the forward sweep to save (push to local stack) the meteorology at each sub-step to be used in the backward sweep. (2) CO2 mixing ratio: as explained in our response to the referee's major comment #1, nonlinearity can be present in WRF-CO2 when the positive definite advection predicts the minimum CO2 mixing at a given grid cell to be negative and trigger the renormalization. When such nonlinearity occurs, cost function gradient depends on not only perturbation but also background value of CO2.

23. p. 10, line 29: The emission-normalized sensitivities can also be found by dividing the full sensitivities by the emissions. Two separate simulations are not required. You might say that you calculated them this way, but to suggest "this is done" in general by a specific approach is misleading.

We agree and the text has been revised to avoid misleading the readers.

24. p. 11, line 5: I realize this is more of a numerical demonstration than scientific result, but it is strange to define the adjoint forcing for tower observations to be at the surface rather than tower height, as in practice these types of measurements would have a greater emissions footprint (hence the rational for using a tower . . .).
We agree. We conducted new experiment with the improved code, and set the adjoint forcing at the 1st and 10th vertical level of the WRF model grid. Footprint figures (Figs 8 of the revised manuscript) are redrawn with the new simulation results. We also noted the difference of footprints between the adjoint forcing at the 1st and 10th levels, as the referee pointed out in his comment.

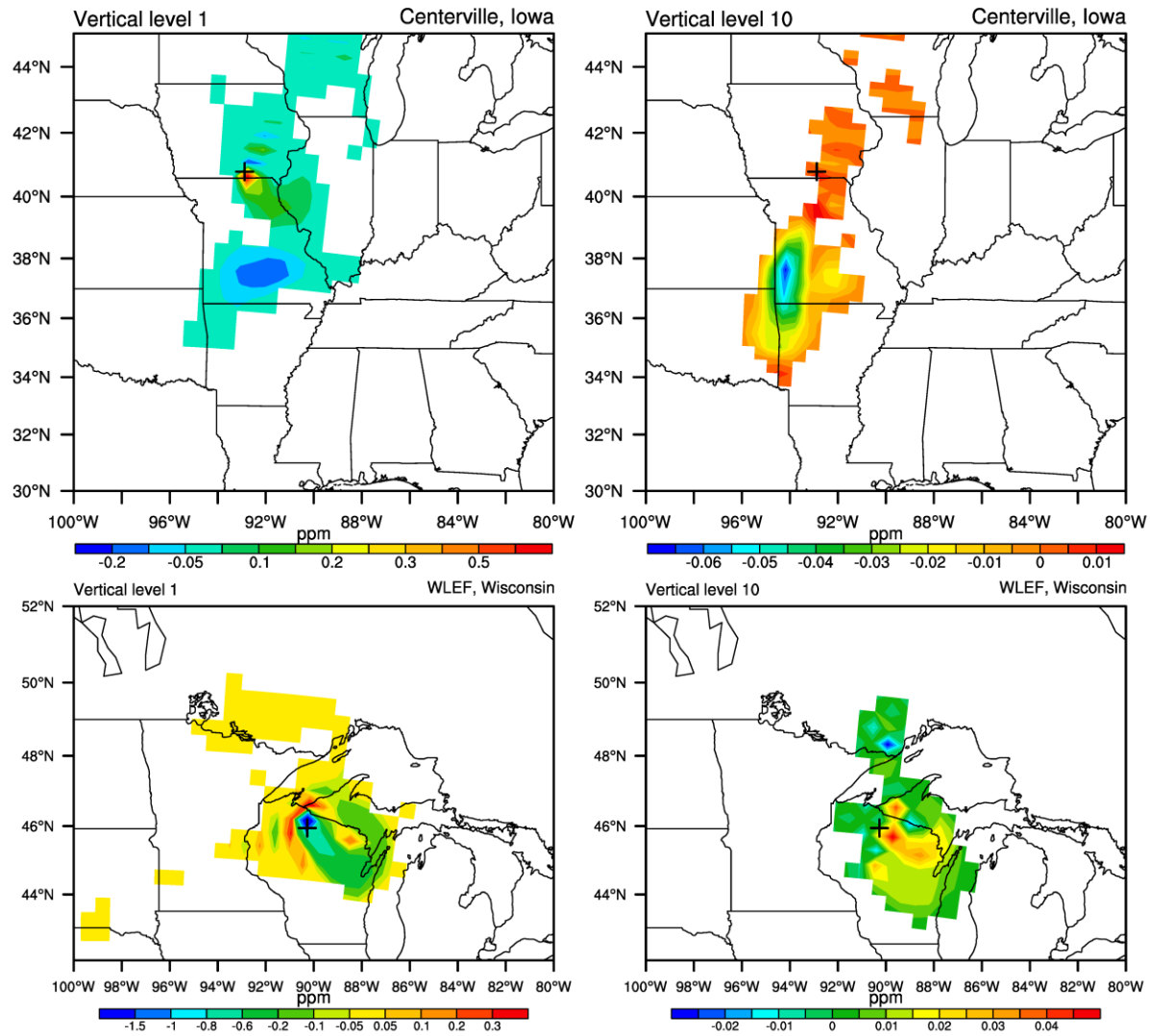


Figure 4. The adjoint sensitivities (footprint) of tower sites at Centerville, Iowa (top panels) and WLEF, Wisconsin (lower panels). At each site, the adjoint sensitivities are calculated twice: one with receptor placed a the 1st vertical level, and another at the 10th level.

25. Section 3.3: It seems like accuracy should be evaluated first, before presenting the sensitivity results in section 3.2.

The order has been switched so that the model accuracy evaluation is presented before the sensitivity spatial pattern.

26. Equation 8: What value used for delta_x? It can sometimes be difficult to find a perturbation value that balances truncation and roundoff error when using this equation to verify adjoint sensitivities.

The delta_x used in final calculation is 0.1. Prior to the final calculation for finite difference sensitivity, we conducted test using delta_x ranging from 0.01 to 1.0 to assess the impact of

magnitude of delta_x. The results indicate there is virtually no difference between 0.01, 0.1, and 1.0 regarding the finite difference value. We attribute this to the fact that CO2 tracer transport is linear. The impact of delta_x on the finite difference sensitivity is documented in Figure 5 below. This figure is included in the supplement document, but not in the manuscript.

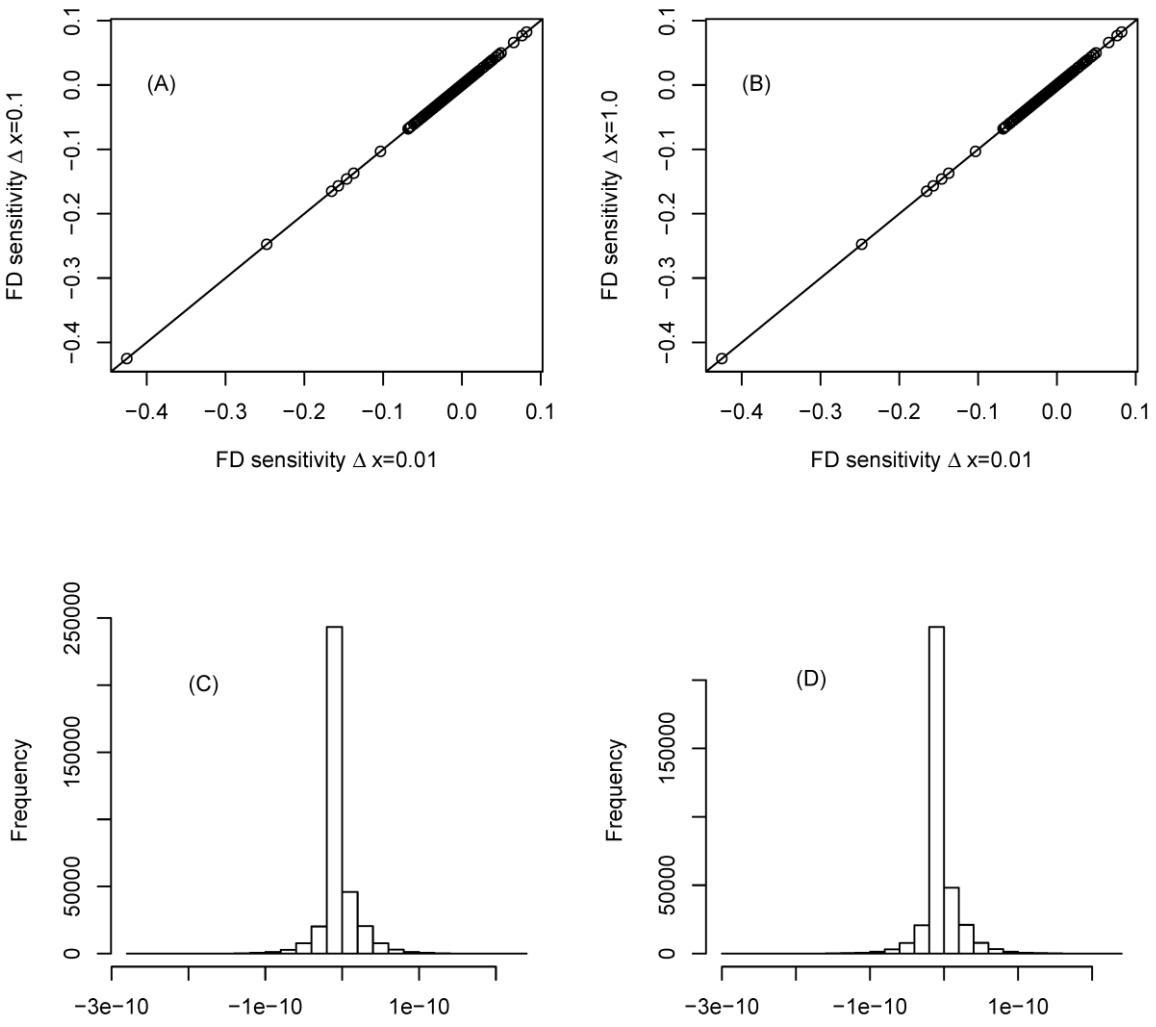


Figure 5. Impacts of the delta_x on finite difference sensitivity calculation. Finite difference of a source is calculated three times, with delta_x set to 1.0, 0.1, and 0.01 respectively. The results are compared grid cell to grid cell. Figure (c) is the histogram of difference in finite difference sensitivity calculated with delta_x=0.01 and delta_x=0.1. Figure (d) is the histogram of difference between detal_x=0.01 and 1.0. Both the scatterplots and histograms show the difference in finite difference caused by delta_x is negligible.

27. p. 12, line 23: By assuming B−1 = 0 and R = I, the pseudo-data case ignores how uncertainties will affect the convergence of Lanczos-CG and LBFGS-B. How would the performance of these two approaches differ with imperfect observations? With an unbiased prior? Determining the correct treatment of B and R is an active research area, which the authors do not address. Do the authors plan to explore more realistic covariance definitions in the future? At a minimum, this should be discussed in Section 4.

We agree with the referee that these are very important issues when applying the system to actual observation data. As the referee pointed out, treatment of background and observation error covariance is an active research area. Our objective of the present paper is to develop and test the TL/AD model and the optimization scheme. In-depth discussion on how the two optimization schemes perform with actual data is beyond the present paper's cope. But we did add a statement to remind the readers the nature of the pseudo-data based inverse experiments.

28. p. 12, line 32-24: Did the authors confirm a loss of conjugacy in the Lanczos basis vectors? Also, did the authors make any attempt to force conjugacy through full re-orthonormalization (e.g., Modified Gram Schmidt)? That mechanism is built in to release version 3.6 of WRFDA. While re-orthonormalization uses extra memory, that resource requirement is often very small relative to that of the model integrations. The authors should justify a decision that adds iterations to the optimization. After including full re-orthonormalization, the number of iterations for Lanczos CG to converge in each outer iteration should be proportional to the degrees of freedom (DOF) constrained by the chosen observations (see, e.g., Rodgers, 2000), entirely independent of the conjugacy issue. At that point, the necessity of multiple outer iterations would be caused by a nonlinearity in the forward model, possibly the PBL treatment or convective transport. The authors make no attempt to characterize such a nonlinearity that would necessitate using a nonlinear optimization strategy.

In the original text, we did not examine conjugacy. The loss of conjugacy was a mere guess, and turned out to be a wrong one. We really appreciate the referee pointing it out.
As we explained in debugging the adjoint model, the degradation of the incremental inner loop was caused by the inaccuracy of adjoint model as opposed to the loss of conjugacy.

Yes, re-orthonormalization is implemented WRF-CO2 following the WRFDA code. This means that loss of conjugacy was not possible, but we did not realize this while writing the original manuscript. This has been corrected in the revision.

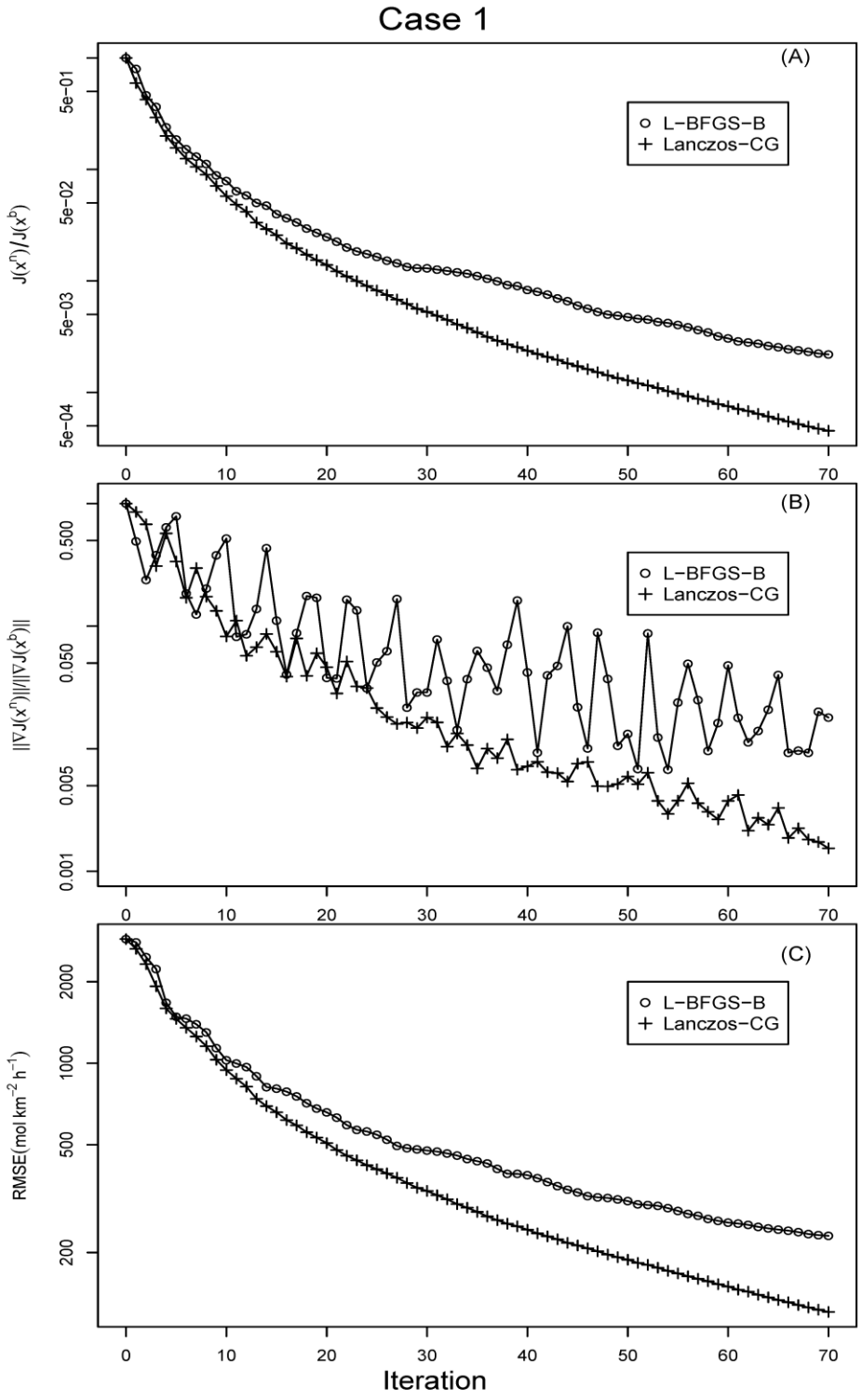


Figure 6. The results of the inverse experiment (Case 1). Reduction of the cost function and gradient norm are expresses as ratio to their respective starting value. Only one outer loop is used in the Lanczos-CG optimization.

29. Section 3.4: While interesting and valuable, numerically, there should be some statement with regards to the unphysical nature of the test setup, to emphasize that this is strictly a numerical test and not the expected level of performance (in terms of cost function reduction or RMSE) that would be obtained in a real inversion.

Thanks for pointing this out. We added description about the unrealistic nature of the experiment setup and that real observation data application will need more careful treatment of the errors.

30. In Section 4, the third paragraph needs a topical sentence. Also, the two sentences "We evaluated ... sensitivity." should be combined into one and made more concise.
A topic sentence has been added and the two sentences have been combined.

31. p. 15, lines 1-11: While mentioning the ability to use different aggregation techniques may illuminate a budding area of research to the reader, the details given should be both accurate and concise. The authors' discussion of smoothing and aggregation error (i.e., from Turner et al., 2015) are based in the assumption that no correlation is used in B. Taken out of context, this could be very confusing to the reader. Full non-ambiguous coverage of that topic would require more than a paragraph, but such a description is not appropriate for this section. Indeed, large portions of Section 4 (paragraphs 4, 5, and 6) ought to be rewritten or removed. Many of the references are out of date, and do not represent the state of the science.

The discussion about aggregation and errors has been removed in the revised manuscript, as they are not central to the present paper's objectives. In its place, we added detailed discussion about memory and computation requirement of the two optimization schemes.

32. p. 15, line 18-21: Other areas to improve upon would be more accurate treatment of data and model (i.e. transport and representational) errors in R, and error correlations in B, and posterior error estimation.

Thanks! These suggestions have been added in the revised manuscript.

33. Section 5: I'm not sure this meets the requirements of GMD, and may delay the publication of this work until the code is publicly available.

Accept. The source code has been submitted to zenodo at https://doi.org/10.5281/zenodo.839260
Future development will be made available to the public access in the same fashion too.

Technical Corrections:
1. Add appropriate punctuation to Eqs. 1, 2, 3, 4, 6, and 7.

Fixed.

2. p. 1, line 21: Remove “inversion” at the end of the sentence, as it is implied in the first half of this statement.
Fixed.

3. p. 2, line 16: “LDMZ” should be changed to “LMDZ”
It is corrected.

4. p. 2, line 17: Change “inverse” (noun) to “invert” (verb).
It is fixed.

5. p. 3, line 3: “LPDM” is undefined. Possibly define and change “Lagangian particle backward trajectory model” to “Lagangian particle dispersion model (LPDM)” on p. 2, line 33.
Definition for LPDM is added, and the Lagrangian particle backward trajectory model is changed to LPDM.

6. p. 4, line 12: “Where” to “where”
Fixed.

7. The first term in parentheses in Eqs. 3, 4, and 6 need transpose operators. Additionally, it would be less confusing if brackets and braces are used in addition to parentheses where warranted.
Thanks for pointing out the missing transpose operators. They are added. The two equation presentation has been improved by using brackets and braces.

8. p. 6, line 9: Correct “innoviation” to “innovation”
Fixed.

9. p. 6, line 11: Remove “Eq. 7”, since you are referencing the very next line of the text.
Fixed.

10. p. 6, line 20: Correct “lead” to “leading”
Fixed

11. p. 7, lines 12, 16: “inner” to “inert”
Fixed

12. p. 7, lines 14-16. Combine the two sentences that both state this category does not apply to CO2.
Fixed

13. p. 7, line 20: “(Zhang et al.,” to “Zhang el al., (”

Fixed

14. p. 7, line 30: Correct “simplied” to “simplified”
Fixed

15. p. 8, lines 12 and 16: Correct “inner” to “inert”
Fixed

16. p. 8, line 12: Correct “use” to “uses”
Fixed

17. p. 9, line 1: Change “chemistry vertical mixing” to “vertical mixing of chemical species”
Fixed

18. p. 9, line 3: “dynamical” to “the dynamical”
Fixed

19. p. 9, line 24: “set up” to “setup”
Fixed

20. p. 10, line 5: “simulation spans” to “simulations span”
Fixed

21. p. 10, line 7: “condition” to “conditions”
Fixed

22. p. 11, line 3: The reference should be to Gerbig et al. (2008). Also, use the correct parenthetical format for inline references.
The Reference is fixed.

23. p. 11, line 3: Change “footprint at a receptor” to “footprint of a receptor”
Fixed

24. p. 11, lines 9, 11, 12, 15: The figure references are off by 1.
Fixed

25. p. 11, line 18: Correct “no shown” to “not shown”
Fixed

26. p. 12, line 23: Correct “identify” to “identity”
Fixed

27. p. 12, line 28: Correct “facotr” to “factor”
Fixed

28. p. 12, line 30-31: Lanczos-CG is repeated twice. Also, use either “Lanczos CG” or “Lanczos-CG” throughout the document.
Fixed.

29. p. 13, lines 2 and 21: cost function needs an article, such as “the”
Fixed.

30. p. 13, lines 2-4: Add commas before and after “by the 30th iteration”.
Fixed

31. p. 13, lines 16-17: The opening to this sentence, “Starting at 2336.5 mol km-2 h-1,” is confusing or out of place.
The sentence has been rephrased.

32. p. 13, lines 24: change “the Lanczos” to “Lanczos” for consistency
Fixed.

33. p. 13, line 26: extra “the”
Fixed.

34. p. 14, line 8: Change “system” to “systems”
Fixed

35. p. 14, line 10: Modify, “Such configuration”, which is grammatically incorrect.
Text is modified to correct the grammatical error.

36. p. 14, line 11: Change “incurring” to “requiring”
Fixed

A list of major changes made in the revised manuscript

- Model code debugging: as the referee point out, the adjoint model was not error free. To address the problem, we systematically debugged the model code. Errors were isolated, identified and corrected. The evaluation through sensitivity calculation confirms that the three model components (NL/TL/AD) match as expected.

- Optimization experiment: in the revised text, synthetic observation data are from 30 vertical levels from bottom up. They were from the bottom level only in the original text.

- In the footprint calculations, receptors are now placed at 1^{st}, 5^{th}, and 10^{th} . They were placed on the 1^{st} level only in the original text.

- TL/AD/FD sensitivity comparison. (1) tangent linear, adjoint, and finite difference sensitivities are calculated for source and receptors cells at different locations in both horizontal and vertical). (2) The receptor cells are placed at the 1^{st}, 5^{th}, and 10^{th} vertical levels at each tower site.

- Cumulus activity indication: extra variables are implemented in the model to track when/where the convective tracer transport is activated during the simulation. This information is plotted and used to ensure there are sources and receptors located within or near the cumulus activity. As the referee pointed out, this is necessary to evaluate the accuracy of the newly developed TL/AD code of the convective chemistry transport scheme (module_ctrans_grell).

- Chemistry initial and boundary conditions used in the simulations have been changed to CarbonTrack2016 CO2 mol fraction.

- The four CO2 fluxes (fossil fuel, fire, biosphere, and ocean) have been changed to CarbonTracker2016 fluxes.

- Comparison has been conducted between WRF simulated meteorology and that interpolated from CFSv2.

Development of the WRF-CO2 4DVar assimilation system v1.0

Tao Zheng[1], Nancy French[2], and Martin Baxter[3]

[1]Department of Geography, Central Michigan University, Mount Pleasant, MI. USA
[2]Michigan Technological Research Institute, Michigan Technological University, Ann Arbor, MI. USA
[3]Department of Earth and Atmospheric Sciences, Central Michigan University, Mount Pleasant, MI. USA

Correspondence to: Tao Zheng (zheng1t@cmich.edu)

Abstract. Regional atmospheric CO_2 inversions commonly use Lagrangian particle trajectory model simulations to calculate the required influence function, which quantifies the sensitivity of a receptor to flux sources. To provide an alternative, we developed an adjoint based four-dimensional variational (4DVar) assimilation system, WRF-CO2 4DVar. This system is developed based on the Weather Research and Forecasting (WRF) ~~model~~ modeling system, including WRF-Chem, WRFPLUS, and WRFDA. In WRF-CO2 ~~4DVR~~4DVar, CO_2 is modeled as a tracer and its feedback to meteorology is ignored. This configuration allows most WRF physical parameterizations to be used in the assimilation system without incurring a large amount of code development. WRF-CO2 4DVar solves for the optimized CO_2 emission scaling factors in a Bayesian framework. Two variational optimization schemes are implemented for the system: the first uses the L-BFGS-B and the second uses the Lanczos conjugate gradient (CG) in an incremental approach. We modified WRFPLUS forward, tangent linear, and adjoint models to include CO_2 related processes. The system is tested by simulations over a domain covering the continental United States at 48 km $\times$ 48 km grid spacing. The accuracy of the tangent linear and adjoint models are assessed by comparing against finite difference sensitivity. The system's effectiveness for CO_2 inverse modeling is tested using pseudo-observation data. The results of the sensitivity and inverse modeling tests demonstrate the potential usefulness of WRF-CO2 4DVar for regional CO_2 inversions.

1 Introduction

Quantification of surface-atmospheric carbon exchange is important for understanding the global carbon cycle (**?**). Both inventory based bottom-up and atmospheric inversion based top-down approaches have been widely used to investigate carbon sources and sinks. Most atmospheric CO_2 inversion methods are based on Bayes theorem, in which CO_2 flux is optimized by minimizing a quadratic form cost function consisting of background cost and observation cost. The minimization of the cost function can be achieved by analytical or variational approaches. **?** provides a concise explanation of the differences between the two approaches.~~**?** provides a comprehensive review of atmospheric chemistry inversions using satellite observations, many of which have been applied to CO_2 inversion.~~

Both analytical and variational inversions use a chemistry transport model (CTM) to relate CO_2 flux to atmospheric CO_2. From the perspective of an optimization system, atmospheric CO_2 forms the observation vector, and CO_2 flux forms the state

vector to be optimized. Central to all CO_2 inversion approaches is the Jacobian matrix which relates changes in flux to change in model-simulated atmospheric CO_2. For an inversion system with a $n \times 1$ state vector and a $m \times 1$ observation vector, its Jacobian matrix is a $m \times n$ matrix. Analytical inversions require the explicit construction of the Jacobian matrix, which can be carried out by either CTM (as the forward model) or its adjoint model. While a forward model calculates the Jacobian matrix by columns, an adjoint model calculates it by rows. The size of the state vector or observation vector determines the number of forward or adjoint model runs needed for constructing the Jacobian matrix. The practical limit imposed by the computational cost of the Jacobian matrix construction and the memory demand of matrix inversion often necessitate the aggregation of flux to reduce state vector size in analytical inversions, which leads to aggregation error (**???**). In comparison, variational approaches do not require the Jacobian matrix to be explicitly constructed, instead they ~~propagate the overall adjoint forcing backward in time in searching for the optimized~~ directly compute the product of the Jacobian with a forcing vector, which is the gradient vector used for optimizing the state vector.~~On the other hand, analytical inversions yield an estimation of the posterior error along with the optimized flux. Variational inversions require significant extra computation to estimate the posterior error.~~

A number of four dimensional variational (4DVar) assimilation systems have been developed and applied to global scale CO_2 inversions. The off-line transport model Parameterized Chemistry Tracer Model (PCTM) (**?**) and its adjoint have been used for CO_2 inversions (**????**). **?** developed a 4DVar system based on the ~~LDMZ~~ LMDZ model (**?**) to assimilate CO_2 observation data from Television Infrared Observation the Satellite Operational Vertical Sounder (TOVS). This system has also been used to ~~inverse~~ invert surface CO_2 observation data ~~(?).~~ (**??**) The TM5 4DVar system (**?**), based on the TM5 global two-way nested transport model (**?**), is used in the CarbonTracker CO_2 data assimilation system (**?**) and is included in the TransCom satellite intercomparison experiment (**?**). TM5 4DVar has also been used to investigate total column CO_2 seasonal amplitude (**?**) and to assimilate the Greenhouse Gases Observing Satellite (GOSAT) observations (**?**). Another widely used inversion system is the GEOS-Chem 4DVar (**??**) with its CO_2 module updated by **?**. GEOS-Chem 4DVar has been used to estimate CO_2 fluxes from the Tropospheric Emission Spectrometer (TES) and the GOSAT CO_2 observations (**??**)~~.~~, and it is also part of JPL's (Jet Propulsion Laboratory) Carbon Monitoring System (**?**)

CO_2 inversions at regional scale have become an active research front in recent years, driven by the need to resolve biosphere-atmosphere carbon exchange at smaller scales (**?**), and by the need to address policy-relevant objectives, such as assessing emission reduction effectiveness (**?**) and the impact of regional scale sources like wildland fire (**?**). ~~Compared with global inversions, there are fewer model choices for regional inversions.~~ A number of regional inversion systems have been developed and applied. For instance, GEOS-Chem 4DVar's nested simulation ability provides a means for regional inversions, such as its application for CH_4 inversion over North America (**?**). The majority of regional inversions use analytical approaches and typically use a Lagrangian particle backward ~~trajectory model~~ dispersion model (LPDM) to compute the required influence function. For instance, **?** used an analytical approach to minimize ~~for~~ the cost function and the STILT (**?**) model driven by assimilated meteorology to calculate the influence function. In a later study, STILT driven by ECMWF meteorology is used to calculate the influence function to investigate the impacts of vertical mixing error (**?**). More recently, **?** also used an analytical

solution for cost function minimization and LPDM (**?**) to compute the influence function. In another study, **?** used STILT driven by meteorology data from WRF to calculate the influence function for comparing Lagrangian and Eulerian models for regional CO_2 inversions. To improve accuracy, STILT has been coupled to WRF, in which the latter provides online meteorology to STILT to avoid interpolation error (**?**). More recently, **?** investigated biopheric CO_2 flux in Amazon using an analytical inversion approach (**?**) with influence function calculated by the STILT and Flexpart (**?**) models. Also, **?** applied regional CO_2 inversion in Canada with both analytical and Markov chain Monte Carlo (MCMC) LPDM based approach. Influence function is also calculated with the Flexpart model in this study.

~~Most of these global and regional inversion systems transport CO_2 using offline meteorology. An exception is the Monitoring Atmospheric Composition and Climate Interim Implementation (MACC-II) greenhouse gases system based on ECMWF Integrated Forecasting System (IFS) (**?**). In this system, CO~~While the rising atmospheric CO_2 ~~transport is computed online using the IFS model . Because meteorological fields are crucial determining the quality of transport (**??**), and model transport uncertainty is key to the inversion quality (**?**), online transport based inversion systems have the potential to investigate and mitigate the transport error.~~

~~In this paper~~has been well documented by observational data, major uncertainties still exist in attributing it to specific processes. For instance, the two sets of terrestrial biosphere CO_2 flux databases in NASA's carbon monitoring system flux pilot project differ substantially (**?**). In order to better resolve the terrestrial biosphere's response to the rising CO_2, inverse modeling at the regional scale is a high research priority (**?**). Toward this end, we developed WRF-CO2 4DVar, a regional CO_2 inversion system with online meteorology. This system is developed by modifying the WRFDA and WRFPLUS system (v3.6) in a similar approach to that used by **??** (GH15/16 afterward) for black carbon emission inversion. WRFDA is a meteorology data assimilation system, which includes a 4DVar assimilation system (**??**) and related adjoint and tangent linear models (WRFPLUS) ~~(**??**)~~(**?**). Designed to improve weather forecasts, WRFDA 4DVar optimizes meteorological initial and boundary conditions by assimilating a variety of observational data. We modified WRFPLUS to include CO_2 related processes and we configure the cost function so that the state vector consists of CO_2 flux instead of meteorological fields. In developing WRFDA-Chem for black carbon inversion, GH15/16 excluded radiation, cumulus, and microphysics parameterization schemes from the tangent linear model and adjoint model because developing these procedures for black carbon would incur a large amount of new code development. In WRF-CO2 4DVar, CO_2 is a tracer, meaning its impacts on meteorology are ignored. This configuration allows us to include full physics schemes in WRF-CO2 4DVar's tangent linear model and adjoint model with limited new code development (see Section 2.4.2). As transport model error is detrimental to 4DVar inversion accuracy (**??**), we deem it important to use the full physics schemes in the tangent linear and adjoint models for WRF-CO2 4DVar. In addition, while GH15/16 excluded convective transport of chemistry species in WRFDA-Chem, we developed the tangent linear and adjoint code for this process in WRF-CO2 4DVar to reduce the vertical mixing error (see Section 2.4.4). Like GH15/16, we implemented an incremental optimization with ~~the Lanczos version of conjugate gradient~~Lanczos-CG, but we also implemented ~~the~~ a L-BFGS-B based optimization~~and compared the performance of the two approaches (see Section 2.2 and 3.4)~~.

The remainder of this paper is organized as follows: Section 2 details the implementation of the two variational optimization schemes for cost function minimization, and the modification to the tangent linear and adjoint models. Section 3 examines the accuracy of sensitivity calculated by the tangent linear and adjoint models, and the system's effectiveness in inverse modeling. Finally, a summary and outlook are presented in Section 4.

2 Method

This section describes the WRF-CO2 4DVar cost function configuration and the associated minimization schemes, followed by a description of the forward, tangent linear, and adjoint models.

2.1 Cost function configuration

WRF-CO2 4DVar is designed to optimize CO_2 flux by assimilating CO_2 observational data into an atmospheric chemistry transport model. CO_2 flux is optimized through use of a linear scaling factor:

$$E = k_{co2} \times \widetilde{E} \tag{1}$$

Where $\widetilde{E}$ is the CO_2 emission read from emission files, k_{co2} is the emission scaling factor, and E is the effective CO_2 flux. It is the effective flux that is used in WRF-Chem's emission driver to update CO_2 mixing ratio (q_{co2}). The emission scaling factor k_{co2}, its tangent linear variable g_k_{co2}, and its adjoint variable a_k_{co2} are used in calculating model sensitivity and minimizing the cost function defined in Eq. (2). The readers can find a list of the notations used in this article in Table 1. Throughout the paper, bold face lower case characters represent vectors and bold face upper case characters represent matrices.

The cost function ~~$J(x)$~~ $J(\mathbf{x})$ of WRF-CO2 4DVar follows the Bayes framework widely used in atmospheric chemistry and numerical weather prediction (NWP) data assimilations:

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) \tag{2}$$

~~Where~~ where the background cost function ~~$J_b(x)$~~ $J_b(\mathbf{x})$ is defined as

$$J_b(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^n - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x}^n - \mathbf{x}^b) \tag{3}$$

and the observation cost function ~~$J_o(x)$~~ $J_o(\mathbf{x})$ is defined as

$$J_o(\mathbf{x}) = \frac{1}{2} \sum_{k=1}^{K} (\{H([M(\mathbf{x}^n)) - y_k)] - \mathbf{y}\}^T \mathbf{R}^{-1} (\{H([M(\mathbf{x}^n)) - y] - \mathbf{y}_k)\} \tag{4}$$

In Eqs. (3-4), the superscript n indicates that $\mathbf{x}^n$ is the optimized state vector at the nth iteration.

Like other data assimilation systems, WRF-CO2 4DVar is essentially an optimization scheme. Its state vector ~~x~~ $\mathbf{x}$ consists of the emission scaling factors k_{co2}. The subscript k in Eq. (4) indicates the entire assimilation time period is evenly split into

~~K~~ k observation windows during which observational data are ingested into the assimilation system.

In WRF-CO2 4DVar, we implemented two optimization schemes to minimize the cost function. The first scheme uses a limited memory BFGS minimization algorithm (L-BFGS-B) (**?**) and the second uses the Lanczos version of conjugate gradient (~~Lanczos-CG~~Lanczos-CG) (**?**) minimization algorithm. Both schemes are iterative processes, and they call on WRF-CO2 4DVar model components (the forward, tangent linear, and adjoint models) to calculate the model sensitivity $\partial q_{co2}/\partial k_{co2}$ between the iterations. The two optimization schemes are described in Section 2.2 and 2.3, respectively, and the three model components are described in Section 2.4.

2.2 L-BFGS-B optimization

L-BFGS-B (**?**) is a quasi-Newton method for nonlinear optimization with bound constraints. ~~It utilizes the cost function gradient to approximate the Hessian matrix, which provides an estimation of posterior error.~~ L-BFGS-B has been used in a number of atmospheric chemistry inverse ~~model~~ modeling systems, including the GEOS-Chem adjoint model system (**?**) and the TM5 4DVar system (**?**). The diagram in Fig. 1 demonstrates the steps involved in the L-BFGS-B based optimization scheme. The scheme is an iterative process which searches for the optimized k_{co2} by minimizing the cost function defined in Eq. (2-4). Between its iterations, the minimization algorithm L-BFGS-B requires the values of the cost function and ~~cost function~~ it's gradient, which are supplied by the forward model and the adjoint model as indicated in Fig. 1.

The calculation of the cost function is carried out based on Eq. (2-4). Starting with the prior estimate of k_{co2}, the forward model run generates the CO_2 mixing ratio q_{co2}, which is transformed from the WRF model space to the observation space by the forward observation operator H. This results in the ~~$H(M(x^n))$~~ $H(M(\mathbf{x}^n))$ term in Eq. (4), which is then paired with the observation vector ~~y_k~~ $\mathbf{y}_k$ to calculate the innovation vector ~~$d_k = H(M(x^n)) - y_k$~~$\mathbf{d}_k = H(M(\mathbf{x}^n)) - \mathbf{y}_k$. Next, the innovation vector and observation error covariance $\mathbf{R}$ are used to calculate the observation cost function ~~$J_o(x)$~~ $J_o(\mathbf{x})$ as expressed in Eq. (4). Finally, the background cost function ~~$J_b(x)$~~ $J_b(\mathbf{x})$ is calculated according to Eq. (3), and combined with the observation cost function ~~$J_o(x)$~~ $J_o(\mathbf{x})$ to form the total cost function ~~$J(x)$~~ $J(\mathbf{x})$ according to Eq. (2).

~~In addition to the cost function,~~ L-BFGS-B ~~also requires the cost function gradient $\nabla J(x)$~~ requires the values of the cost function $J(\mathbf{x})$ and it's gradient $\nabla J(\mathbf{x})$ in searching for the optimized k_{co2}. The ~~cost function~~ gradient is calculated using Eq. (5).

$$\nabla J(\mathbf{x}) = \sum_{k=1}^{K} \widetilde{M}^T \widetilde{H}^T R^{-1}(\{H([M(\mathbf{x}^n) - y - \mathbf{y}_k)]\} + B^{-1}(\mathbf{x}^n - x - \mathbf{x}^b) \quad (5)$$

The first term on the right hand side of Eq. (5) is the observation ~~cost function~~ gradient and the second is the background ~~cost function~~ gradient. The observation ~~cost function~~ gradient is calculated in two steps: (1) The innovation vector is scaled by $\mathbf{R}^{-1}$ and transformed to the WRF model space by the adjoint observation operator, resulting in ~~$\widetilde{H}^T\mathbf{R}^{-1}(H(M(x^n)) - y_k)$~~ $\widetilde{H}^T\mathbf{R}^{-1}(H(M(\mathbf{x}^n))$ which is the adjoint forcing. (2) The adjoint forcing is ingested by the WRF-CO2 adjoint model during its backward (in time)

integration, which yields the observation ~~cost function~~ gradient. Supplied with the values of the cost function and ~~cost function~~ gradient, the L-BFGS-B algorithm finds a new value of k_{co2}, which is used for the next iteration. The iterative optimization process continues until a given convergence criterion is met.

The L-BFGS-B based optimization in WRF-CO2 4DVar is implemented based on the Fortran code of Algorithm 788 version Lbfgsb.2.1 (**?**). We plan to change it to version Lbfgsb.3.0 (**?**) in the next model update.

2.3 Incremental optimization

The second optimization scheme we implemented for WRF-CO2 4DVar is the incremental approach commonly used in NWP data assimilation systems, including ECWMF 4DVar (**?**) and WRFDA (**?**). A major difference between the L-BFGS-B based optimization and the incremental optimization is that the former optimizes for the state vector while the latter optimizes for the state vector analysis increment. The incremental assimilation scheme uses a linear approximation to transform the observation cost function from what is defined in Eq. (4) to Eq. (6):

$$J_o(\mathbf{x}) = \frac{1}{2}\sum_{k=1}^{K}(\{H([M(\mathbf{x}^{n-1})) - y] - \mathbf{y}_k + ((x^n - x^{n-1})))\widetilde{H}[\widetilde{M}(\mathbf{x}^n - \mathbf{x}^{n-1})]\}^T \mathbf{R}^{-1}(\{H([M(\mathbf{x}^{n-1})) - y] - \mathbf{y}_k + ((x^n - x^{n-1})))\widetilde{H}[\tag{6}$$

Compared to Eq. (4), Eq. (6) approximates the innovation vector by a sum of two parts. The first part, ~~$H(M(x^{n-1})) - y_k$~~$H(M(\mathbf{x}^{n-1})) -$ is the ~~inoviation~~ innovation vector from the previous iteration. The second part, ~~$\widetilde{H}(\widetilde{M}(x^n - x^{n-1}))$~~$\widetilde{H}(\widetilde{M}(\mathbf{x}^n - \mathbf{x}^{n-1}))$, is the state vector analysis increment ~~$(x^n - x^{n-1})$~~ $(\mathbf{x}^n - \mathbf{x}^{n-1})$ transformed by the tangent linear model $\widetilde{M}$ and tangent linear observation operator $\widetilde{H}$. With the linear approximation of the cost function ~~, the cost function gradient is calculated by Eq. (7)~~ and gradient are calculated by

$$\begin{aligned}\nabla J(\mathbf{x}) = &\sum_{k=1}^{K}\widetilde{M}^T\widetilde{H}^T\mathbf{R}^{-1}\{H[M(\mathbf{x}^{n-1}) - \mathbf{y}_k]\} + \mathbf{B}^{-1}(\mathbf{x}^{n-1} - \mathbf{x}^b) + \\ &\sum_{k=1}^{K}\widetilde{M}^T\widetilde{H}^T\mathbf{R}^{-1}\{\widetilde{H}[\widetilde{M}(\mathbf{x}^n - \mathbf{x}^{n-1})]\} + \mathbf{B}^{-1}(\mathbf{x}^n - \mathbf{x}^{n-1})\end{aligned} \tag{7}$$

In WRF-CO2 4DVar, the incremental optimization is implemented as a double loop in which the outer loop calculates the first and second items on the right hand side of Eq. (7), while the inner loop calculates the third and fourth items. The ~~outer loop~~ superscript $n-1$ indicates that $\mathbf{x}^{n-1}$ is the optimized state vector in the last outer loop, and superscript n indicates that $\mathbf{x}^n$ is the optimzed state vector in the inner loop. The outer loop first calls the forward model M and adjoint model $\widetilde{M}^T$ to calculate ~~$\widetilde{M}^T\widetilde{H}^T R^{-1}(H(M(x^{n-1}) - y_k))$ and $B^{-1}(x^{n-1} - x^b)$~~$\widetilde{M}^T\widetilde{H}^T R^{-1}(H(M(\mathbf{x}^{n-1}) - \mathbf{y}_k))$ and $B^{-1}(\mathbf{x}^{n-1} - \mathbf{x}^b)$, which remain unchanged during the subsequent inner loop calculation. The analysis increment ~~$(x^n - x^{n-1})$~~ $(\mathbf{x}^n - \mathbf{x}^{n-1})$ is optimized in the inner loop, which calls the tangent linear and adjoint models to calculate the third and fourth items of Eq. (7). ~~For the inner loop calculation , we use the Lanczos CG~~ Inner loop calculation is carried out by Lanczos-CG (**?**), which ~~requires~~

~~less memory than L-BFGS-B and can be readily adapted to distributed memory parallel computation. In addition, Lanczos CG can optionally estimate the lead~~ can optionally estimate eigenvalues of the cost function Hessian matrix ($\nabla^2 J(\mathbf{x})$). ~~$\nabla^2 J(x)$).~~ The diagram in Fig. 2 shows the structure of the Lanczos-CG based incremental optimization implemented in WRF-CO2 4DVar.

2.4 Forward, tangent linear, and adjoint models

WRFPLUS consists of three model components: the WRF model, its tangent linear model, and its adjoint model (**??**). The three models are used by WRFDA to optimize the initial meteorological condition in order to improve numerical weather prediction. Unlike WRFDA, WRF-CO2 4DVar is designed to optimize CO_2 flux, instead of the meteorological initial and boundary conditions. This difference means CO_2 related processes are needed in WRF-CO2 4DVar's model components. To include the CO_2 related processes, we first use WRF-Chem to replace WRF as the forward model. Then, we conducted a thorough variable dependence analysis to determine how to modify the tangent linear and adjoint model in order keep them consistent with WRF-Chem (the forward model).

2.4.1 Forward model

We replaced WRF with WRF-Chem as the forward model component of WRF-CO2 4DVar. As an atmospheric chemistry extension of WRF, WRF-Chem includes chemistry, deposition, photolysis, advection, diffusion, and convective transport of chemistry species (**?**). These processes are included in different modules of WRF-Chem: ARW (Advanced Research WRF) dynamical core, physics driver, and chemistry driver. We use the GHG (Greenhouse Gas) tracer option of WRF-Chem but have the CO and CH_4 removed, leaving only CO_2 related procedures. In the emission driver, we use the CASA-GFED v4 biosphere flux (**?**) to replace the online biogenic CO_2 model Vegetation Photosynthesis and Respiration Model (VPRM) (**?**). This change is made because WRF-CO2 4DVar optimizes for CO_2 flux instead of online emission model parameters.

2.4.2 Variable dependence analysis

The tangent linear and adjoint models of WRFPLUS need to be modified to include the CO2 related processes so that they will be consistent with the forward model. The results of the variable dependence analysis is summarized in Table 2, which groups WRF-Chem processes into three categories regarding CO_2 tracer transport. The first category includes the chemistry processes that do not apply to CO_2~~. This category contains~~ , including gas and aqueous phase chemistry, dry and wet deposition, and photolysis. ~~Because they are not applied to CO_2, these~~ These processes are simply excluded from the forward, tangent linear, and adjoint models in WRF-CO2 4DVar.

The second category is comprised of the physical parameterizations that do not provide CO_2 tendency, but provide meteorological tendency. This category includes radiation, surface, cumulus, and microphysics parameterizations. While the full

physics schemes of surface, cumulus, planetary boundary layer (PBL), and microphysics are used in the forward model of WRFPLUS, simplified versions of these schemes are used in its tangent linear and adjoint models. In addtion, WRFPLUS uses full radiation schemes (longwave and shortwave) in its forward model, but it excludes radiation schemes from its tangent linear model and adjoint model. The differences in the physical parameterizations between the forward model and tangent linear/adjoint models in a 4DVar system is a source of linearization error. For instance, **?** found linearization error in ECMWF 4DVar larger than expected and recommended more accurate linear physics for higher resolution 4DVar systems. Because WRF-CO2 4DVar ignores the impacts of CO_2 mixing ratio variation on the meteorological fields, no tangent linear and adjoint variables for meteorological fields are needed in its tangent linear model and adjoint model. Since this second category of processes are not directly involved in CO_2 transport, there is no need for their tangent linear and adjoint procedures in WRF-CO2 4DVar. In WRFPLUS's tangent linear model, we removed the tangent linear code of the ~~simplied~~ simplified versions of the cumulus, surface, and microphysics schemes, and replaced them with ~~the forward code of~~ their corresponding full schemes as used in the forward model. ~~An adjoint model conducts a forward sweep and a backward sweep.~~ In WRFPLUS's adjoint model, the forward sweep updates the state variables and local variables just as in the forward model, but it also stores these variables' values for the subsequent backward sweep, which updates the adjoint variables of the state variables. We removed the simplified versions of the cumulus, surface, and microphysics schemes used in the forward sweep of WRFPLUS's adjoint model, and replaced them with the full schemes used in the forward model. Since these processes do not directly modify CO_2 mixing ratio, we simply removed their corresponding adjoint code from the backward sweep of the adjoint model, as indicted by the 'X' in Table 2.

The third category includes advection, diffusion, emission, and turbulence mixing in PBL, along with convective transport of CO_2. Because these processes directly modify CO_2 mixing ratio, their tangent linear code and adjoint code are needed for WRF-CO2 4DVar. The modifications we made for advection and diffusion are described in Section 2.4.3, and those for emission, turbulent mixing in PBL, and convective transport of CO_2 are detailed in Section 2.4.4.

2.4.3 Advection and diffusion of CO2

WRF includes the advection and diffusion of inert tracers along with other scalars in its ARW dynamical core. The tangent linear and adjoint code of these processes has been implemented in WRFPLUS. It should be noted that the variables for these ~~inner~~ inert tracers are part of WRF, instead of WRF-Chem. WRF-Chem ~~use~~ uses a seperate array for its chemistry species. Since we replaced WRF with WRF-Chem as the forward model in WRF-CO2 4DVar, CO_2 mixing ratio are included in the chemistry array. In the GHG option of WRF-Chem we use for WRF-CO2 4DVar, CO_2 from different sources (anthropogenic, biogenic, biomass burning, and oceanic) are represented by separate variables in the chemistry array. Following the treatment for the ~~inner~~ inert tracers in WRFPLUS, we modified subroutines solve_em_tl and solve_em_ad to add the tangent linear and adjoint code for the advection and diffusion of the chem array. The modifications we made include adding calls to the procedures that calculate advection and diffusion tendencies, updating the chemistry array with the tendencies and boundary conditions, and addressing the Message Passing Interface (MPI) communications. The new upgrade to WRFPLUS described in (**?**) greatly expedited this part of development for WRF-CO2 4DVar. The 'Add' in Table 2 for advection and diffusion

emphasizes that their tangent linear and adjoint code are added to WRF-CO2 4DVar based on the existing WRFPLUS code without substantial new code development.

2.4.4 Vertical mixing of CO_2 in PBL and convective transport

An accurate representation of vertical mixing is important for inversion accuracy, because misrepresentation causes transport error, which manifests itself in the innovation vector and causes error in posterior estimation (**?**). For instance, **?** pointed out that global chemistry transport model error in vertical mixing and boundary layer thickness could cause significant overestimation of northern terrestrial carbon uptake. A comparison of four global models found that model transport uncertainty exceeds the target requirement for A-SCOPE mission of 0.02 Pg C yr^{-1} per 10^6 km_{2}^{-2} (**?**). In addition, **?** reported that convective flux is likely underestimated in boreal winter and spring based on simulated upper tropospheric CO_2 from 2000 to 2004 using three chemistry transport models.

In WRF-Chem, ~~chemistry vertical mixing~~ vertical mixing of chemical species is treated in three separate parts: in the vertical diffusion (subgrid scale filter) in the dynamical core, in the PBL scheme in the physics driver, and within convective transport in the chemistry driver. The subgrid scale filter in the dynamical core treats both horizontal and vertical diffusions, but vertical diffusion is turned off if a PBL scheme is used. While all PBL schemes implemented in WRF-Chem treat the vertical turbulent mixing of temperature and moisture, only the ACM2 PBL scheme also treats chemistry species (**?**). We choose to use the ACM2 scheme in WRF-CO2 4DVar so that CO_2 vertical mixing is treated by the PBL parameterization. Convective transport of chemistry species in WRF-Chem is not treated by the cumulus scheme in the physics driver, but by a separate convective transport module (module_ctrans_grell) in the chemistry driver (**?**).

Because the ACM2 PBL and chemistry convective transport are not included in WRFPLUS, we developed their tangent linear and adjoint code for WRF-CO2 4DVar. We first used the automatic differentiation tool TAPENADE (**?**) to generate the tangent linear and adjoint code based on the forward code: module_bl_acm for the ACM2 PBL and module_ctrans_grell for the chemistry convective transport. We then manually modified the TAPENADE generated code to remove redundancy and unnecessary loops. It should be pointed out that these code developments are made significantly simpler because the meteorological state variables are merely passive variables in the tangent linear and adjoint code. For instance, to calculate the moist static energy and environmental values on cloud levels, the chemistry convective transport code (module_ctrans_grell) in the chemistry driver calls a number of subroutines in the cumulus parameterization code in the physics driver. Because these subroutines in cumulus parameterization only involve meteorology state variables and not the chemistry array, no tangent linear or adjoint code is needed for them in WRF-CO2 4DVar.

3 Results

This section presents an accuracy assessment of the newly developed WRF-CO2 4DVar system. We first describe the simulation model setup, then the sensitivity tests and inverse modeling experiments.

3.1 Model setup

WRF-CO2 4DVar is ~~set up~~ setup with a domain covering the continental United States with 48 km × 48 km grid spacing and 50 vertical levels ~~.~~ (Fig 3.). The domain dimension is 110 points in east-west and 66 points in north-south direction. Model configuration includes: Rapid Radiative Transfer Model (RRTM) longwave radiation (**?**), Goddard shortwave radiation (**?**), Pleim surface layer (**?**), Pleim-Xiu land surface model (**?**), ACM2 PBL (**?**), Grell-Freitas cumulus (**?**), and Thompson microphysics (**?**). Positive-definite transport is applied to the transport of scalars and CO_2. ~~Emissions inventories~~

CO_2 fluxes used for the simulations ~~include: anthropogenic flux from the Emission Database for Global Atmospheric Research (EDGAR; http://edgar. jrc. ec.europa.eu/index.php) version 4.2, biosphereflux from CASA-GFED v4 (?), and ocean flux from Estimating the Circulation and Climate of the Ocean Phase II (ECCO2)-Darwin(?). A common emission scaling factor is applied to the combined EDGAR and CASA-GFED CO_2 flux, both of which are disaggregated to daily values before being ingested by WRF-CO2 4DVar. ECCO2-Darwin ocean flux is used in the simulations, but not included as the assimilation system's control variables.~~ are from the CarbonTracker 2016 version (CT2016 afterward) (**?**). These fluxes are the optimized surface fluxes at a 3-hour interval and at 1×1 degree spatial resolution. The four individual CO_2 fluxes (biosphere, fossil fuel, fire, and ocean) are spatially interpolated to the WRF grid, and saved in chemistry input files. In the following sensivitity tests and inverse experiments, the emission scaling facotr k_{co2} is applied only to the biosphere flux. Daily mean biosphere fluxes are calculated as the arithmatic mean of the 3-hourly CT2016 fluxes at each surface grid cell, and the scaling factor k_{co2} is applied as in Eq. (1). The daily mean biosphere flux used for the 24 hour simulation is shown in Figure 4. The model configuration and emission ~~inventories~~ data used are summarized in Table 3.

Model ~~simulation spans~~ simulations span 24 hours from 00 UTC 02 June to 00 UTC 03 June, 2011. Meteorological initial and lateral boundary conditions are prepared using the NCEP Climate Forecast System Version 2 (CFSv2) 1×1 degree 6-hourly products (**?**). CO_2 initial ~~condition~~ and lateral boundary conditions are ~~generated by running a WRF-Chem global domain simulation for the 120 hours prior to and during the WRF-CO2 4DVar simulations period. Figure 3 shows the model domain (bold outline) with its terrestrial region colored by the CASA-GFED biosphere CO_2 flux. The red triangles in the figure mark the locations of 20 tower sites at which model sensitivities are assessed in Section 3.3.~~ from the CT2016 global 3×2 degree CO_2 mole fraction. We used a method similar to PREP-CHEM-SRC (**?**) to horizontally and vertically interpolate CT2016 mole fraction data to the WRF grid.

3.2 ~~Spatial patterns of sensitivities~~

First, the forward model (WRF-Chem) was run for 24 hours with the CO_2 emission as described in the last section. Trajectory files that contain model state variables including both meteorology and CO_2 mixing ratio are saved at model dynamical time step intervals (120 seconds). These files are required for the subsequent tangent linear and adjoint model runs. Figure 4 shows the instantaneous values of Sea Level Pressure (SLP) and horizontal wind at the model's lowest vertical level at ~~06 UTC, 12 UTC, 18 UTC, and 23 UTC of 2 June 2011.~~ each 6 hours. The figure shows that a high pressure system was located off the west coast, causing a northerly surface wind off southern California, and a westerly wind for most of the Pacific Northwest. A low pressure system intensified over Montana and North Dakota during the 24 hours, causing a strong southerly wind over the Midwest. In the northeast, as a low pressure system moved eastward out of the domain, the surface wind shifted from southwesterly to westerly. We conducted comparison between the WRF simulated meteorology and that interpolated from the CFSv2 at 6-hour intervals. The comparison results (not shown) indicate that the WRF simulated meteology closely matches the CFSv2 in terms of SLP and winds at multiple vertical levels.

3.2 Accuracy of tangent linear and adjoint sensitivities

We next examined the accuracy of the newly developed tangent linear and adjoint models by comparing their sensitivity calculations against finite difference sensitivity calculated by the forward model. Grid cells involved in sensitivity calculation are shown in Fig. 3, in which the 35 blue stars are the source cells, and the 20 red triangles are 20 tower sites where the receptors are placed. All the 35 sources are placed at the grid's bottom vertical level. Receptors are placed at the 1st, 5th, and 10th vertical level at each of the 20 tower sites, resulting to 60 receptor cells.

A tangent linear model run for a grid cell will calculate the tangent linear sensitivity ~~$\partial q_{co2}/\partial k_{co2}$~~$\partial \mathbf{q_{co2}}/\partial k_{co2}$, which approximates a column vector of the forward model's Jacobian matrix and quantifies the influence of the cell's emission change on CO_2 mixing ratio of its receptor cells downwind. In comparison, an adjoint model run for a grid cell will calculate adjoint sensitivity ~~$\partial q_{co2}/\partial k_{co2}$~~$\partial q_{co2}/\partial \mathbf{k_{co2}}$, which approximates a row vector of the forward model's Jacobian matrix and quantifies the influence on the cell's CO_2 mixing ratio by its source cells upwind. Because k_{co2} multiplies emission in Eq. (1), the magnitude of the sensitivity is determined by both the magnitude of emission and meteorological transport.

~~We first examined the influence of the meteorological transport on sensitivity by excluding the influence of emission magnitude. This is done by setting emission at all grid cells to unity (1 mol km^{-2} h^{-1}) for a given tangent linear or adjoint model run.~~ To calculate tangent linear sensitivity at a grid cell, g_k_{co2} is set to unity at the cell and zero at all other cells at the start of a tangent linear model run. Upon completion, the values of ~~g_q_{co2}~~ $\mathbf{g\_q_{co2}}$ are the tangent linear sensitivities ~~$\partial q_{co2}/\partial k_{co2}$. Because emissions are set to unity at all model cells, the unit of such obtained sensitivity is ppm/(mol km^{-2} h^{-1}).~~ $\partial \mathbf{q_{co2}}/\partial k_{co2}$. To calculate adjoint sensitivity at a cell, an adjoint model run starts with a_q_{co2} set to unity at the cell and zero at

all others, and the values of ~~a_k_{co2}~~ a_k_{co2} at the end of the simulation are the adjoint sensitivities. The adjoint model running in this mode is analogous to using a Lagrangian particle transport model in backward trajectory mode to compute the footprint ~~at~~ of a receptor, such as shown in Fig 4. of ~~(?)~~?.

~~Table 4 lists the locations of tower sites marked on Fig 3. We calculated tangent linear and adjoint sensitivities at each of the grid cells where these tower sites are located. Figure 5 shows the results for Ozarks, Missouri (top row) and WLEF, Wisconsin (bottom row). Tangent linear sensitivity at Ozarks shows that the areas where surface level CO_2 are most influenced by Ozarks are located north and northwest of it. In comparison, adjoint sensitivityat Ozarks shows its surface level CO_2 was most influenced by the emission of areas located to the south of it. The spatial patterns of both the tangent linear and adjoint sensitivities at Ozarks indicate the influence of meteorological transport of CO_2 (Fig. 3), as surface wind in this area was largely southerly and southeasterly during the simulation period. In comparison, the tangent linear and adjoint sensitivities at WLEF (lower row of Fig. 4) show that its receptor cells are largely located to its west and its source cells located to its southeast. Surface wind over Wisconsin and Minnesota was easterly and southeasterly during the simulation period (Fig. 3), confirming the sensitivities' spatial patterns. The results from Sutro, California (top row) and Hidden Peak, Utah (bottom row) are shown in Fig 6. While the adjoint sensitivities (right column) in Fig. 6 show that both sites are mostly influenced by areas west of them owing to the steady westerly wind (Fig. 3), the tangent linear sensitivities (left column) show marked differences in their influence downwind: The northerly wind along the Californian coast advects Sutro's emission southward,~~ We first compared the tangent linear sensitivity against the finite difference sensitivity. After confirming the accuracy of the ~~southwesterly wind over Utah, Colorado, and Wyoming advects Hidden Peak's emission toward northeast. The tangent linear and adjoint sensitivities patterns correlate similarly with the meteorology at the other 16 sites (no shown).~~

3.3 ~~Accuracy of tangent linear and adjoint sensitivities~~

~~We next examined the accuracy of the tangent linear and adjoint models by comparing their sensitivitycalculations against the finite difference~~ tangent linear model, we compared the adjoint sensitivity against the tangent linear sensitivity.

Finite difference sensitivities are calculated using the two-sided formula (Eq. (8)).

$$\frac{\partial f}{\partial x} = \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} \tag{8}$$

~~All results described in this section are obtained by model simulations with full emissions as described in Section 3.1. The resulting sensitivities $\partial q_{co2}/\partial k_{co2}$ include the influence of both meteorological transport and the magnitude of emission, and their units are ppm. This is different from the results represented in Section 3.2, where the sensitivities do not include the influence of the magnitude of emission.~~

~~Using~~ The magnitude of Δx used in Eq. (8) is determined by comparing the result from a range of different values. At a number of the 35 sites, we calculated the finite sensitivities using Δx set to 0.01, 0.1, and 1.0, and the results show that the maginitude of all differences is less than 10^{-10} (results not shown). This is due to the fact that the WRF-CO_2 is largely linear. For all subsequent calculaions, $\Delta x = 0.1$ is used for Eq. (8)~~, two forward model runs were used to calculate the finite~~

~~sensitivity at a given grid cell, which forms a column vector of the model's Jacobian matrix.~~ .

Because both finite difference and tangent linear sensitivities form columns of the Jacobian matrix, their values can be compared cell by cell for all receptor cells for a given site. Figure ~~7~~ 6 shows the comparison between the finite difference and tangent linear sensitivities at ~~nine sites~~9 of the 35 source cells. The dark straight lines in the figures are the 1:1 line. ~~At two of the sites, WLEF and Cannan Valley, the sensitivity $\partial q_{co2}/\partial k_{co2}$ shows negative values, due to the fact that the combined EDGAR and CASA-GFED emission is negative (carbon sink). The slope and R^2 of the linear fit show that the tangent linear sensitivity agrees very well with the finite difference sensitivity~~The maximum and minimum of the difference between finite difference and tangent linear sensitivities are given for each source cell. Results at the ~~other 11 sites~~ rest of the sources are similar (not shown). All differences are less than 10^{-10}, confirming that the tangent linear model is accurate.

We next ~~examined the accuracy of the adjoint model~~ evaluate the adjoint model by comparing adjoint sensitivities against the tangent linear sensitivities. Because finite difference sensitivities form columns of the Jacobian matrix while adjoint sensitivities form rows of the Jacobian matrix, they can only be compared at the intersections of the rows and columns ~~for the 20 sites. For the 24-hour simulation, sensitivity at most of these intersections are zero except for those where a cell is both a receptor and source. We compared adjoint and finite difference sensitivities at the 20 sites and the result is shown in Fig. 8. The result from Sutro, California is shown in an inset due to its much larger magnitude as the site is close to the large anthropogenic emission from San Francisco. The two sets of sensitivities compare well, with the linear fit slope equal to 0.9994 and R^2 equal to 0.9997~~of the Jacobian matrix, meaning there are 2160 (35×60) pairs of comparison. We organized these 2160 pairs into three groups based on the vertical levels a receptor is placed at and the result is shown in Fig. 7. The minimum and maximum value of the difference between tangent linear and adjoint sensitivities in all three groups are no greater than 10^{-6}, indicating that the adjoint model is accurate.

3.3 Spatial patterns of adjoint sensitivities

Adjoint sensitivity $q_{co2}/\mathbf{k}_{co2}$ quantifies how q_{co2} of a given receptor is impacted by the emission scaling factor of all surface cells. It is similar to the receptor footprint typically calculated using LPDM, such as Fig. 4 of **?** and Fig. 1 of **?**. But $q_{co2}/\mathbf{k}_{co2}$ differs from footprint in that the former contains the combined impact of tracer transport and emission magnitude, while the the latter is determined by tracer transport alone. We examined the spatial patterns of the adjoint sensitivity to discern the impacts of tracer transport. Figure 5 shows $q_{co2}/\mathbf{k}_{co2}$ of Centerville, Iowa (top row) and WLEF, Wisconsin (bottom row). At each tower site, $q_{co2}/\mathbf{k}_{co2}$ of receptor placed at the 1st and 10th vertical levels are plotted.

The adjoint sensitivities of the Centerville tower site indicate its q_{co2} results primarily from surface flux located immediately south of the site. This pattern matches the fact that low level wind during the simulation period is predominantly souhernly, tranporting tracers northward. There is also a marked difference in the adjoint sensitivity of the same tower site when the receptor is placed at a different height. The figure in the top left panel shows that the highest magnitude of $q_{co2}/\mathbf{k}_{co2}$ is closest

to the tower itself, indicating a large impact from local flux. In comparison, when the receptor is placed at the 10th vertical level, the peak magnitude of its adjoint sensitivity is in much farther distance southward, and it features much wider spread, indicating transport of flux distant from that receptor is dominant. Results from WLEF shows the adjoint sensitivity are located to the southeast of the site, matching the southeasterly wind patterns around Wisconsin during the simulation period. There are also clear difference between the receptors at the different vertical levels. Results from other sites all show simiar pattern of impacts of transport and receptor placement height (not shown).

3.4 Inverse modeling test

After confirming the validity of the tangent linear and adjoint models, we tested the effectiveness of WRF-CO2 4DVAR in inverse modeling experiments. Pseudo-observation data generated by the forward model run are used in these inverse modeling experiments, which start with prescribed prior values ~~of~~ for the emission scaling factors and seek to recover their true values. To generate pseudo-observation data, the forward model ran for ~~24 hours with EDGAR and CASA-GFED emission (disaggregated to daily values)~~24 hours with emission scaling factor set to unity at all surface grid points, saving CO_2 mixing ratio (q_{co2}) every ~~4~~ 4 hours. This generated a set of six pseudo-observation files, each of which include the instantaneous q_{co2} at the model's ~~lowest layer at all grid cells.~~ first 30 vertical levels starting from the bottom level at each grid point. Given the simulation domain dimension, each individual observation file contains $110 \times 60 \times 30$ data entries.

We conducted inverse modeling experiments for two cases of prior k_{co2}. In the first case, the prior emission scaling factor ~~underestimates~~ overestimates the true values by 50% (~~$k_{co2} = 0.5$~~ $k_{co2} = 1.5$ at all cells). In the second case, the prior emission scaling factor ~~overestimates the true values by 50% ($k_{co2} = 1.5$ at all cells).~~ is randomly distributed between 0.5 and 1.5. Figure 9 shows the two experiment cases as scatter plotS between the true biophsere CO_2 and its background value (the prior).

Both L-BFGS-B and incremental optimization (~~Lanczos CG~~Lanczos-CG) are applied to the two cases, giving four inverse modeling experiments in total. In all four experiments, background error covariance is set to infinity ($\mathbf{B^{-1}} = \mathbf{0}$) and equal weights are assigned to all observations ($\mathbf{R}$ set to ~~identify~~ identity matrix). This configuration is equivalent of (1) the setting total cost function to the observation cost function, and (2) setting the gradient to the observation gradient. It should be pointed out that this is an unrealistically simplified treatment of $\mathbf{B}$ and $\mathbf{R}$, used here for the sole purpose of testing the WRF-CO2 4DVar system with error-free pseudo-observations.

Because the pseudo-observation data are of q_{co2} ~~from~~ at the forward model's ~~lowest layer~~grid points, the mapping between model space and observation space is trivial: the observation operator, tangent linear observation operator, and adjoint observation operator are all set to the identity matrix. Again, it should be noted that application of real observation data will require development of observation operators and their tagent linear and adjoint counterparts.

The results from inverse modeling experiments with ~~prior emission scaling facotr k_{co2}=0.5~~ Case 1 prior are shown in Fig ~~9 and 10. Figure 9~~ 10 and 11. Figure 10 shows the iterative reduction of the cost function ~~$J(x)$ (top) and cost function gradient norm $\|\nabla J(x)\|$ (bottom). By the 30^{th} iteration, the cost function is reduced to 0.42×10^{-5} and 1.54×10^{-5} of its starting value by the Lanczos-CG~~ $J(\mathrm{x})$, gradient norm $\|\nabla J(\mathrm{x})\|$, and ~~L-BFGS-B and~~ RMSE. The iteration number for Lanczos-CG ~~optimization, respectively. The figure shows that Lanczos-CG reduces the cost function faster than L-BFGS-B in the first 10 iterations, but its reduction rate slows down and stagnates at round the 15^{th} iteration , caused by the gradual loss of conjugacy. Lanczos-CG exits~~ is all from its inner loop~~at end of 16^{th} iteration and starts a second outer loop in which the forward model is called to run with the last optimized k_{co2}, and the adjoint model is called to update the adjoint forcing. This results in a renewed conjugacy and increased cost function reduction rate. In comparison, L-BFGS-B has a more consistent cost function reduction rate through the iterations. The lower panel of Fig. 9 shows that by the 30^{th} iteration cost function gradient norm $\|\nabla J(x)\|$ is reduced to 0.6×10^{-3} and 1.35×10^{-3} of its starting value by the L-BFGS-B and Lanczos-CG optimization , respectively. Whereas the~~ , and only one outer loop is used. The figures show both L-BFGS-B and Lanczos-CG ~~results in a largely monotonic decreasing $\|\nabla J(x)\|$,~~ reduce the cost function monotonically. In about the first 10 iterations, the cost function reduction is more or less similar for the two optimization schemes, but Lanczos-CG starts to gradually outperform L-BFGS-B ~~features oscillations but results in a overall~~ after. In gradient norm reduction, both schemes feature periodic oscillations embedded in the large scale downward trend.

~~Figure 10 shows the progress of the optimized CO_2 emission, obtained by multiplying the combined EDGAR/CASA-GFED emission with the optimized k_{co2}. The figure shows the comparisonbetween the true emission value and optimized values after the 1^{st}, 5^{th}~~ By comparison, ~~10^{th}, 20^{th}, and 30^{th} iterations by both optimization approaches. The straight line in each figure is the 1:1 line, and Root Mean Squared Error (RMSE) is given in each figure. The figures show that after the first iteration, the~~ Lanczos-CG has a smaller magnitude oscilation and steeper downward trend than L-BFGS-B. It should be noted while L-BFGS-B ~~optimized emission barely departed from the prior value, reflecting the very small reduction in cost function by L-BFGS-B in its first several iterations as~~ calculates cost function and its gradient in each iteration, Lanzcos-CG only approximates these values in its inner loop. The cost function and gradient norm from Lanczos-CG shown in Fig. ~~9. In comparison, the first iterationby Lanczos-CG substantially moved the optimized emission toward the true values. The figure also~~ 10 are calculated by extra calls to the forward and adjoint models in each inner iteration, which doubles the computation cost and is not needed in practice. Figure 10(c) shows that both optimization schemes ~~sometimes push underestimated values to overestimates. After the 5^{th} iteration, the patterns are very similar between the two optimization schemes, although RMSE still indicates substantial differences. The L-BFGS-B optimization achieves a smaller RMSE than Lanczos-CG by the 20^{th} iteration, mirroring the cost function and cost function gradient norm reduction trend shown in Fig. 9. Starting at 2336.5 mol km^{-2} h^{-1}, RMSE is reduced to 17.0~~ reduce RMSE of daily biosphere flux monotonically, and ~~42.6 mol km^{-2} h^{-1} after 30 iterationsby the L-BFGS-B and~~ Lanczos-CG ~~, respectively.~~ achieves better reduction after about the first 10 iterations. Figure 12 shows the snapshots of the optimized daily mean biosphere flux (obtained as the product of the prior flux and the optimized scaling factor) at a selected set of iterations. These figures depict the iterative process of priors converging to the true soluction.

The results of inverse modeling experiments ~~with prior emission scaling factor $k_{co2} = 1.5$~~ using Case 2 prior are shown in Fig. ~~11 and 12.~~ 12 and 13. The reductions of ~~$J(x)$ and $\|\nabla J(x)\|$~~ $J(\mathbf{x})$, $\|\nabla J(\mathbf{x})\|$, and RMSE are similar to ~~the first case. Figure 11 shows that in 29 iterations, L-BFGS-B reduces cost function to 0.38×10^{-5} of its starting value, out-performing Lanczos CG (0.87×10^{-5}). The cost function gradient norm $\|\nabla J(x)\|$ is reduced to 0.45×10^{-3} and 0.91×10^{-3} of its starting value by L-BFGS-B and~~ Case 1 in that Lanczos-CG ~~, respectively. Figure 12 shows the iterative optimization of CO_2 emission by the two optimization approaches. Similar to the first case, L-BFGS-B results in a slower reduction of RMSE in about the~~ substantially outperforms L-BFGS-G after about first 10 iterations~~, but catches up with the Lanczos-CG afterward. After 29 iterations, the RMSE has been reduced from 2336.5 mol km^{-2} h^{-1} to 20.2 and 22.19 mol km^{-2} h^{-1}by the L-BFGS-B and Lanczos-CG, respectively.~~ . Table 5 summarizes the results from ~~the~~ all four inverse modeling experiments described above. It must be pointed out that these inverse modeling results are obtained from a highly unphysical setup, and they are not the expected level of performance (in terms of cost function and RMSE reduction) that would be obtained in a real inversion.

4 Summary and outlook

~~While the rising atmospheric CO_2 has been well documented by observational data, major uncertainties still exist in attributing it to specific processes. For instance, the two sets of terrestrial biosphere CO_2 flux databases in NASA's carbon monitoring system flux pilot project differ substantially (?). In order to better resolve the terrestrial biosphere's response to the rising CO_2, inverse modeling at the regional scale is a high research priority (?).~~

~~Toward this end, we~~ We developed the WRF-CO2 4DVar, a data assimilation system designed to constrain surface CO_2 flux by combining an online atmospheric chemistry transport model and observation data in a Bayesian framework. We implemented two optimization schemes for cost function minimization. The first is based on L-BFGS-B and the second is an incremental optimization using Lanczos-CG. The cost function and its gradient required by the optimization schemes are calculated by WRF-CO2 4DVar's three component models: forward, tangent linear, and adjoint model, all developed on top of the WRFPLUS system. While WRFPLUS's forward model is WRF, we use WRF-Chem as WRF-CO2 4DVar's forward model to include CO_2 in the system, and we modified the tangent linear and adjoint models to keep their consistency with the forward model. Like most other CO_2 inverse modeling ~~system~~systems, WRF-4DVar ignores the possible impacts of atmospheric CO_2 variation on the meteorology. This simplification enables us to use the same full physical parameterizations in the forward, tangent linear, and adjoint model. ~~Such~~ This configuration reduces linearization error while allowing the WRF system's large number of physical parameterizations to be used in WRF-CO2 4DVar without ~~incurring~~ requiring a large amount of new code development.

We tested WRF-CO2 4DVar's tangent linear and adjoint models by comparing their sensitivities' spatial patterns with the dominant wind patterns. The results make physical sense given the meteorological transport. We evaluated the accuracy of tangent linear and adjoint models by comparing their sensitivity against finite difference sensitivity calculated by the forward model. The results show that both tangent linear and adjoint sensitivities agree well with finite difference sensitivity. At last,

we tested the system in inverse modeling with pseudo-observation data, and the results show that both optimization schemes successfully recovered the true values with reasonable accuracy and computation cost.

~~The~~ While Lanczos-CG performs better than L-BFGS-B in our inverse modeling tests, it must be pointed out that our tests are very limited. Although a comprehensive comparison between the two optimization schemes ~~have their respective advantages and disadvantages. Lanczos CG based incremental optimization requires less memory and is more amenable to parallel programming, both of which are desirable for large dimension optimization problems typical in atmospheric chemistry inverse modeling. As a quasi-Newton approach,~~ is beyond the scope of the present paper, we wish to point out some of their differences as implemented in WRF-CO2 4DVar. First, the Lanczos-CG calls the tangent linear model in each inner loop iteration, while L-BFGS-B ~~may converge faster than Lanczos CG, and it can be extended to provide an approximation of the cost function's inverse Hessian, which is also the posterior error covariance. However,~~ calls the forward model. For a tracer transport system like WRF-CO2 4DVar, the tangent linear model can skip some of the costly physics parameteriztions, such as the radiation scheme. This difference means typically the tangent linear model is faster than the forward model, and as a result Lanczos-CG runs faster than L-BFGS-B. In our inversion modeling experiments (24-hour simulation with $\Delta t = 120$ seconds, 30 processes), it takes about 10 minutes walltime to complete one inner loop of Lanczos-CG. L-BFGS-B ~~requires much more memory, especially if one chooses to update the inverse Hessian.~~ takes slightly longer walltime to complete one iteration.

~~A potential application for~~ Second, provided with the cost function and its gradient, each iteration of L-BFGS-B calculates an updated state vector from its previous iteration. In WRF-CO2 4DVar~~will be using satellite CO_2 data to estimate surface flux. Technological advancement in satellite remote sensing has led to increasingly refined footprint size of CO_2 sensors. While the smaller footprint has the advantage of allowing CO_2 measurement during scattered cloud conditions,~~ , this calculation is carried out on only root process and broadcasted to the ~~mismatch between the satellite retrieval footprint size and the larger atmospheric chemistry transport modelgrid causes representation error (?). Because scale mismatch caused representation errors will lead to systematic bias in CO_2 flux estimates, it is desirable to use a transport model with finer grid spacing. In this regard, the WRF-CO2 4DVar system provides the flexibility of using any model grid spacing.~~ other processes. In comparison, Lanczos-CG calculates the state vector increment based on the cost function gradient alone (without the need for $J(\mathbf{x})$). The calculation is carried out on each process. The above difference has implications for memory requirements: The main memory allocation for L-BFGS-B is its workspace array, which is about $(2 \times k + 4) \times n$, where n is the size of the state vector (x), and k is the number of corrections used in the limited memory matrix. This memory allocation is only needed on the root process. The value of k is set by the user and the recommended value is between 3 and 20. In comparison, Lanczos-CG requires memory size of about $m \times n$ on each process, where m is the maximal inner loop iteration allowed. Although it is possible to reduce the per process a memory allocation from $m \times n$ to n by disactivating the modified Gram-Schmidt orthonormalization step, it is typically not recommened.

~~Although using a high resolution transport model could reduce scale mismatch and the related representation error, its actual application needs to address the balance of aggregation and smoothing error (?). For instance, if~~ Another consideration for memory requirements is related to I/O time cost. WRFPLUS saves its entire trajectory in memory to avoid expensive I/O operations. This is not a practical solution for WRF-CO2 ~~4DVar's model grid spacing decreases from 48 km × 48 km to 12 km × 12 km to better match the footprint size of GOSAT measurement, its state vector dimension will increase~~ 4DVar, which is designed to run a longer simulation than the typical 6-hour run intended for WRFDA. GH15/16 ~~times. The much larger state vector is not likely to be sufficiently constrained by observations, leading to smoothing error (??). On the other hand, decreasing model resolution means a smaller state vector and less smoothing error, but the spatial patterns of emission in the large grids are imposed by prior knowledge and not allowedto be optimized, leading to aggregation error (??). A solution toward an optimal balance of the representation, aggregation, and smoothing errors is to use a high resolution transport model but aggregate the transport model's native grid to reduce the state vector dimension. A number of studies have been conducted to search for the optimal aggregation approaches (????). A mapping mechanism between WRF-4DVar's native grid and state vector can be easily implemented to test such aggregation approaches.~~ implementd a second-order checkpoint mechanism to overcome the memory limit. This approach breaks the whole simulation period into sections, saves restart files at end of each section by the forward model. This approach requires extra call of forward model to recalculate trajecotry for each section during backward integration (See Fig. 3 of GH15/16)

~~Another potential application of~~ We implemented a different approach to overcome the memory limit posed by a long simulation. In WRF-CO2 4DVar~~is to investigate transport model uncertainty. For instance, ? analyzed atmospheric tracer sensitivity to convective transport parameterization, and ? investigated CO_2 inversion uncertainty caused by the transport model, including advection, PBL mixing, convection, and mesoscale processes . With its online meteorology and flexibility in configuring physical parameterizations and dynamical processes, WRF-CO2 4DVar provides a versatile platform for investigating transport model related error in CO_2 inversions.~~ , the forward model saves trajectory at each time step in memory, as WRFPLUS does. After a number of integration steps, the memory on each task process is dumped to an external file, and the memory is then reused. Each external file is marked with its starting timestamp and the process it belongs to. For instance, a 24-hour simulation with 120-second time step will have a total of 720 steps. If the system saves its trajectory to external files each 30 time steps, memory allocation on each task process is only needed for 30 steps instead of 720 steps. This will results in 24 (720/30) trajectory files on each task process, and the total number of trajectory files depends on the number of processes used. These trajectory files are read by both tangent linear and adjoint models in a similar way as standard WRF auxiliary files. In the above example, they are read in at each 30 time steps, substantially reducing I/O time compared with reading in at each step. These trajectory files are different from standard WRF auxillary files in that each file belongs to an individual process, rather than being shared among all processes. This means all model runs in an inverse experiment must use the same domain patch configuration, which is the most common practice.

~~Regarding its applications in the above discussed areas, we will address the following issues in the future development of WRF-CO2 4DVar: (1) developing~~ In future development, we plan to implement observation operators for ~~satellite retrieved column-averaged dry air mole fraction of CO2 (XCO_2); (2) implementing the option to include the initial and lateral boundary conditions of atmospheric CO_2~~ real observations, including those from towers, satellites, and airborne. This is required for applying WRF-CO2 4DVar with real observation data. As a regional inverse system, correct treatment of chemistry lateral boundary conditions is important. We plan to include chemistry initial and boundary conditions in the state vector ~~.~~in the next update. In addition, future applications of WRF-CO2 4DVar with real observations must use proper treatment of observation and background error covariance, which was not tackled in the pseudo-observation test used in the present paper.

5 Code availability

~~The source code and compilation instruction of the~~ WRF-CO2 4DVar ~~assimilation system can be obtained by contacting the lead author: T. Zheng (zheng1t@emich.edu). A full release of the source code for the public use is being prepared. The system is under continuous development and major updates in the future will be released too.~~source code can be retrieved via https://doi.org/10.5281/zenodo.839260

Acknowledgements. The authors express their appreciation for the WRF/WRF-Chem/WRFDA/WRFPLUS development teams for making their code available in the public domain. Discussion with Joel ~~BeLanc~~ LeBlanc of Michigan Technological Research Institute (MTRI) improved the optimization schemes implementation and presentation in this paper. The insightful and detailed comments from the two referees greatly improved the model and the paper.

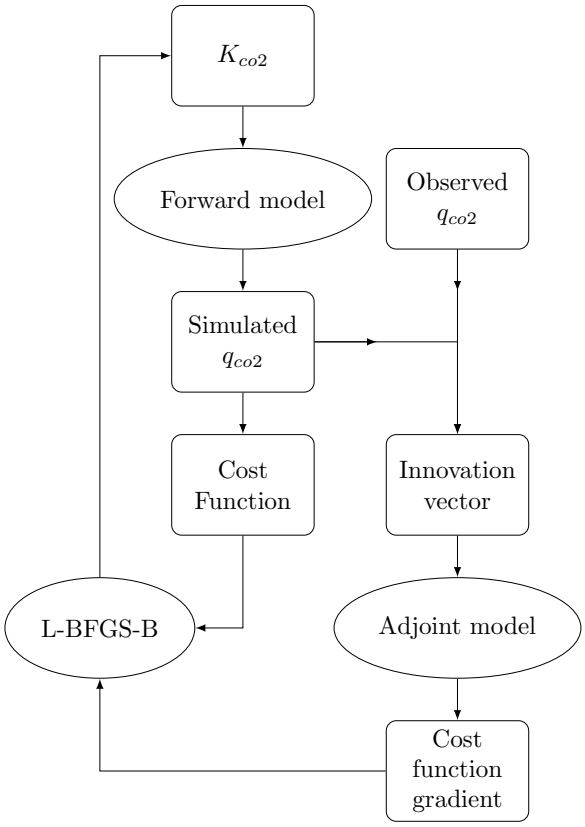


Figure 1. Diagram of L-BFGS-B based optimization implemented for WRF-CO2 4DVar.

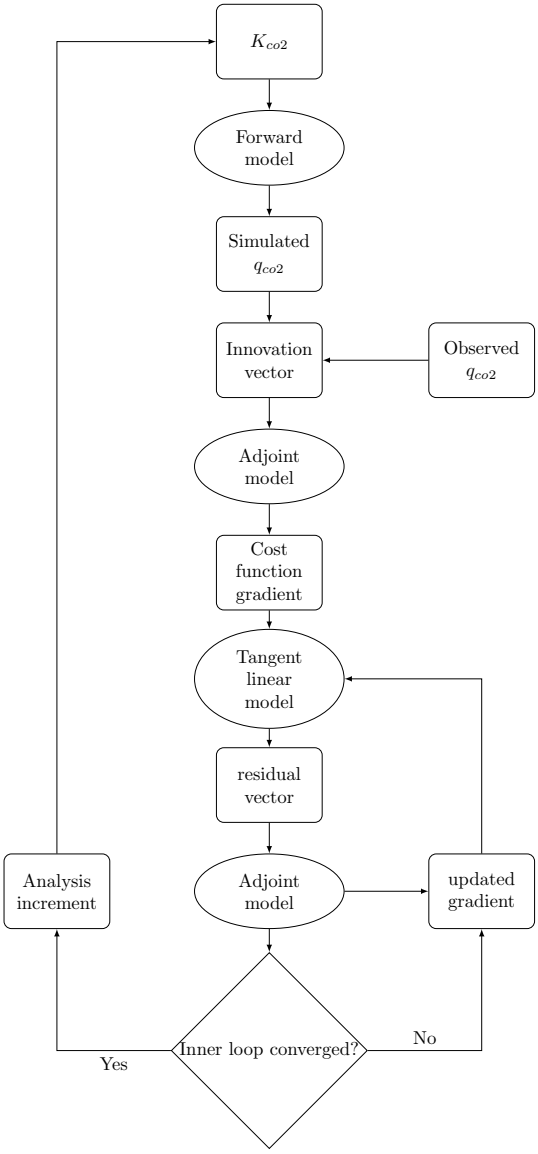


Figure 2. Diagram of ~~Lanczos conjugate gradient (CG)~~ Lanczos-CG based incremental optimization implemented for WRF-CO2 4DVar

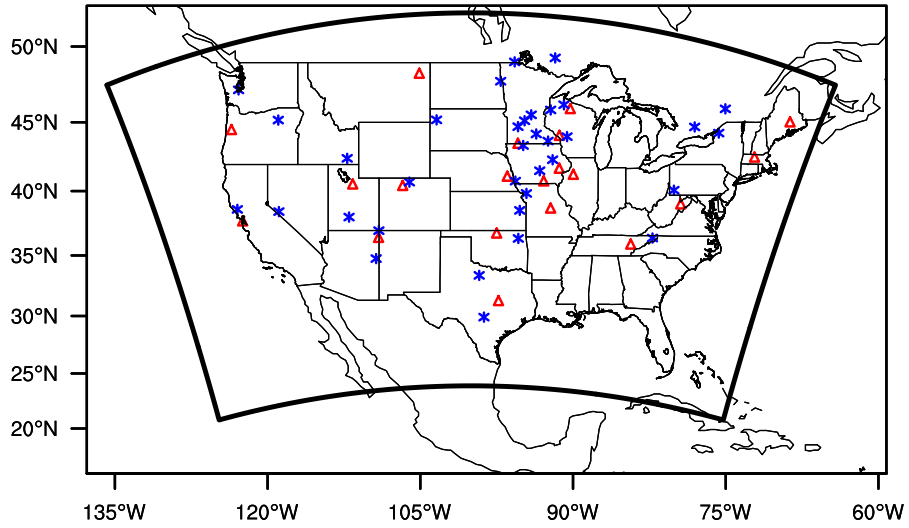


Figure 3. WRF-4DVar simulation domain covering the continental United State with 48 km×48 km grid spacing. The domain boundary is marked by the bold dark outline. ~~Land area is colored by CASA-GFED v4 biosphere CO_2 flux (mol km^{-2} h^{-1}). Locations of~~ Grid cells used for evaluating sensitivities are marked: red triangles are the 20 CO_2 tower sites used as receptor location ; blue starts are ~~marked by red triangles~~ source locations. While receptors are placed at the 1^{st}, 5^{th}, and 10^{th} vertical level at each site, all sources are at the 1^{st} level only.

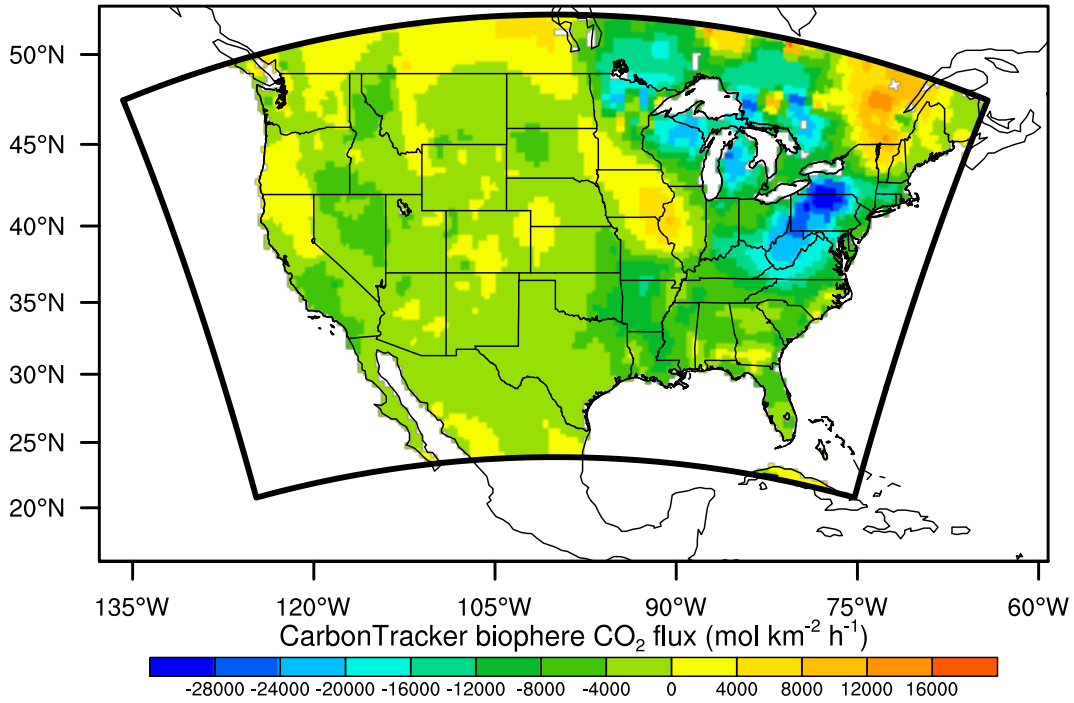
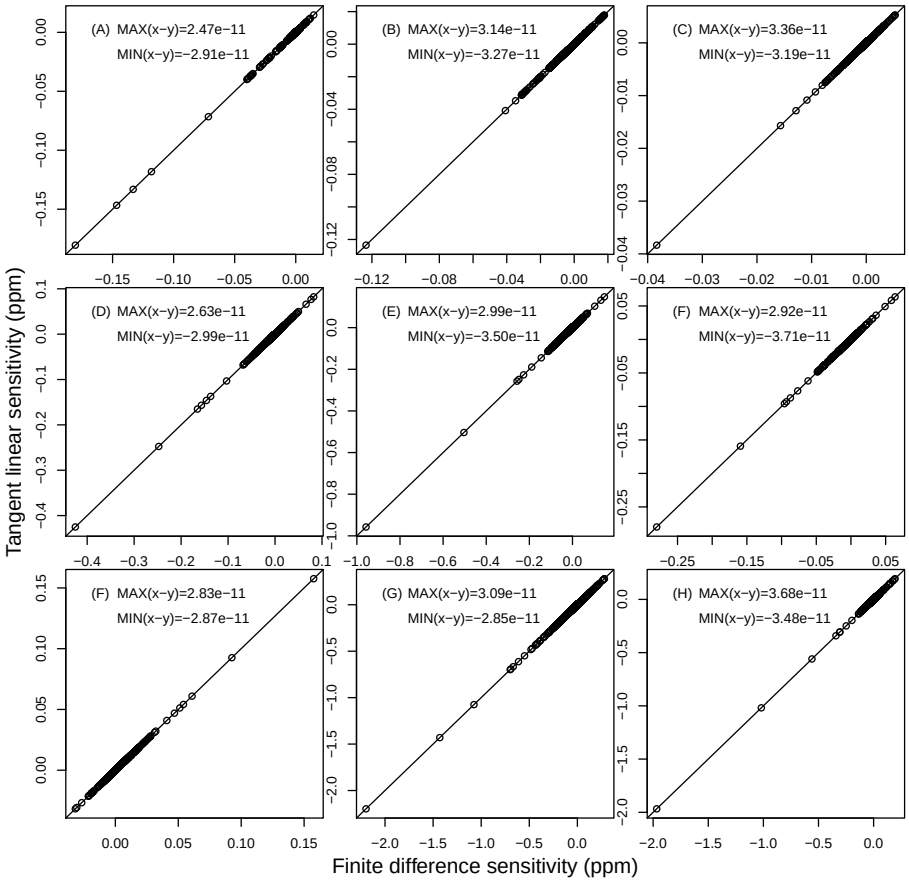


Figure 4. Daily mean CarbonTracker biosphere CO_2 flux, calculated as the arithmetic mean of the 3-hourly flux between 2011-06-02 00:00:00 UTC to 2011-06-03 00:00:00 UTC.



Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by finite difference (x axis) and tangent linear model (y axis) ~~at~~ for nine ~~flux CO_2 tower sites~~sources (see Fig. 3 for source locations).

Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by finite difference (x axis) and tangent linear model (y axis) ~~at~~ for nine ~~flux CO_2 tower sites~~sources (see Fig. 3 for source locations).

Figure 5. ~~Same as Figure 5 but for Sutro, California (top) and Hidden Peak, Utah (bottom).~~

Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by finite difference (x axis) and tangent linear model (y axis) ~~at~~ for nine ~~flux CO_2 tower sites~~sources (see Fig. 3 for source locations).

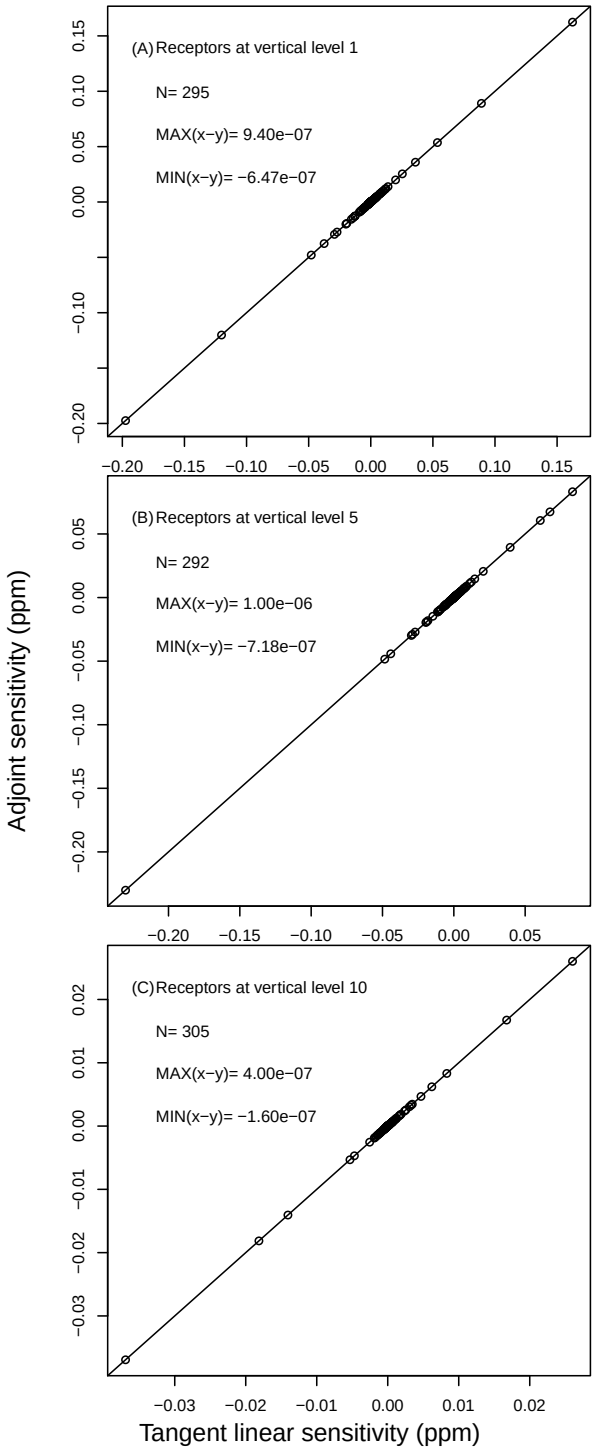


Figure 6. Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by ~~finite difference~~ the tangent linear (x axis) and adjoint model (y axis)~~at the model grid of the 20 CO_2 tower sites listed in Table 4. Result at Sutro, California, is shown in an inset due to its much larger magnitude~~.

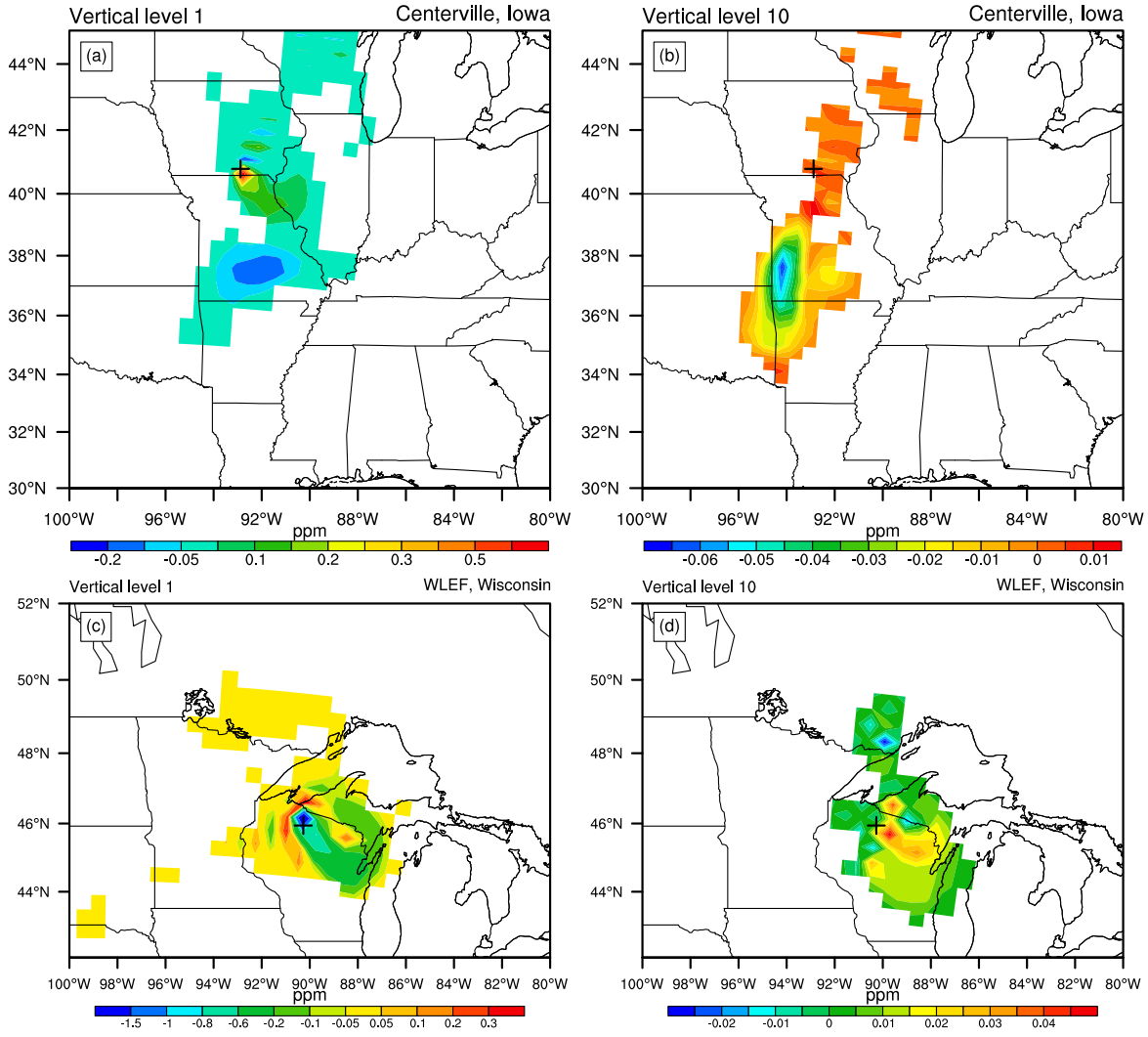


Figure 7. ~~Results~~ The top panel shows adjoint sensitivity of ~~inverse modeling case~~ receptors placed at the 1st (~~prior k_{co2} underestimates the true value by 50%~~a), and 10th (d) vertical level at Centerville, Iowa. The ~~top figure~~ bottom panel shows ~~the reduction~~ adjoint sensitivity of receptors placed at the ~~cost function~~1st (c), ~~represented by $J(x^n)/J(x^b)$,~~ and ~~the bottom~~ 10th (d) vertical level at WLEF, Wisconsin. The black cross in each figure ~~shows~~ marks the ~~reduction of the cost function gradient norm, represented by $\|\nabla J(x^n)\|/\|\nabla J(x^b)\|$~~corresponding tower site.

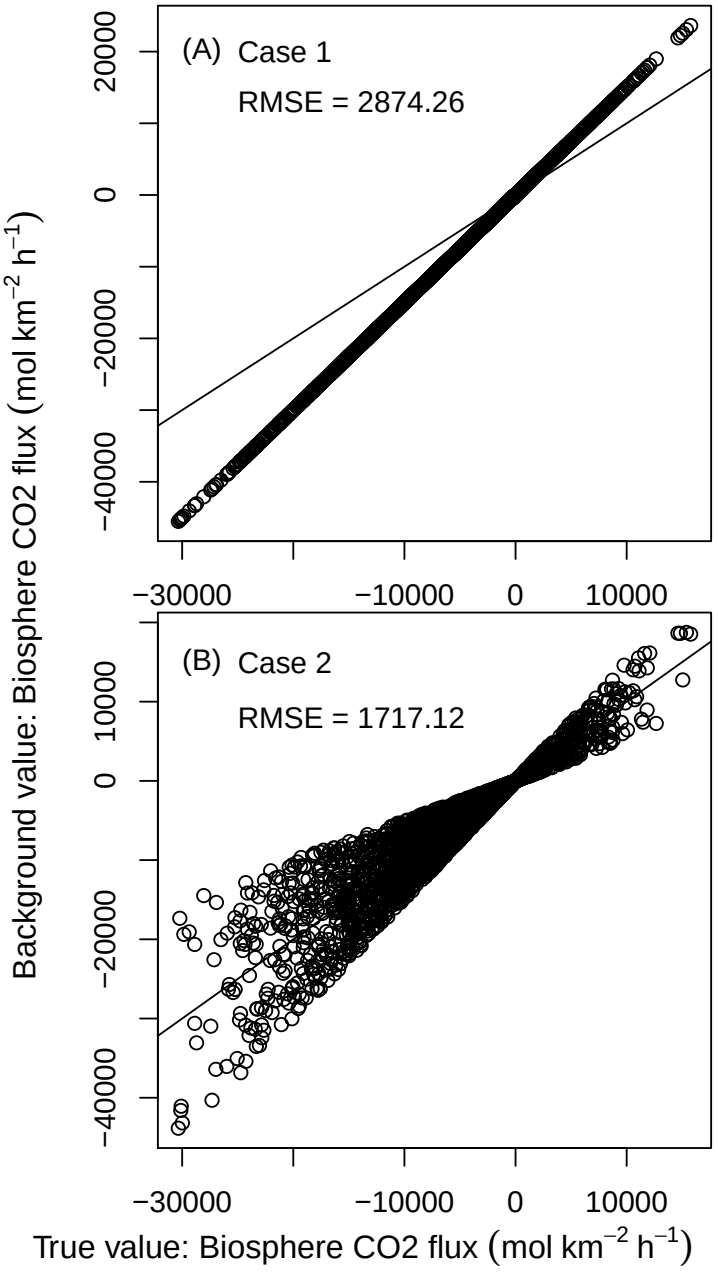


Figure 8. ~~Comparison between~~ The first guess of biosphere CO_2 fluxes used in the two inverse experiments. The x-axis is true ~~and optimized~~ daily mean CarbonTracker biosphere CO_2 ~~flux by L-BFGS-B~~ value (~~left column~~as shown in Fig 4), and ~~Lanczos CG (right column) in inverse modeling case 1 (prior k_{co2} underestimates~~ y-axis is the ~~true~~ first guess (background value~~by 50%~~). ~~Optimized flux after the 1^{st}, 5^{th}, 10^{th}, 20^{th}, and 30^{th} iterations are shown~~ The solid line in ~~the~~ each figure is the 1:1 line.

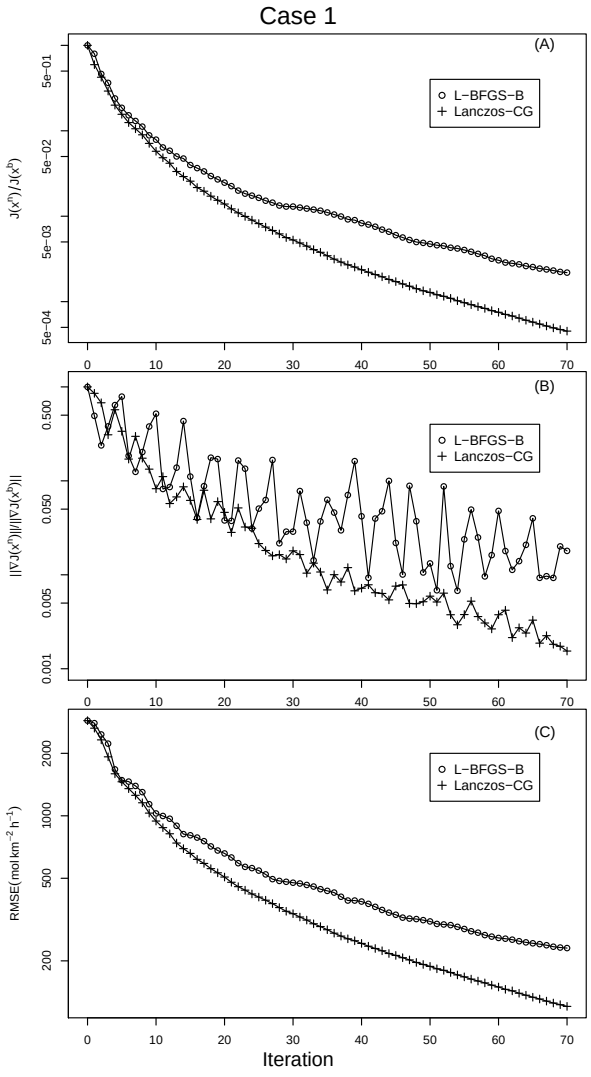


Figure 9. Results of inverse modeling ~~case 2~~ experiment Case 1. Figure (~~prior k_{co2} overestimates the true value by 50%~~a) ~~. The top figure~~ shows the reduction of the cost function, represented by $J(x^n)/J(x^b)$~~, and the bottom figure~~ . Figure (b) shows the reduction of ~~cost function~~ the gradient norm, represented by $\|\nabla J(x^n)\| / \left\|\nabla J(x^b)\right\|$. Figure (c) shows the reduction of biosphere CO_2 flux RMSE.

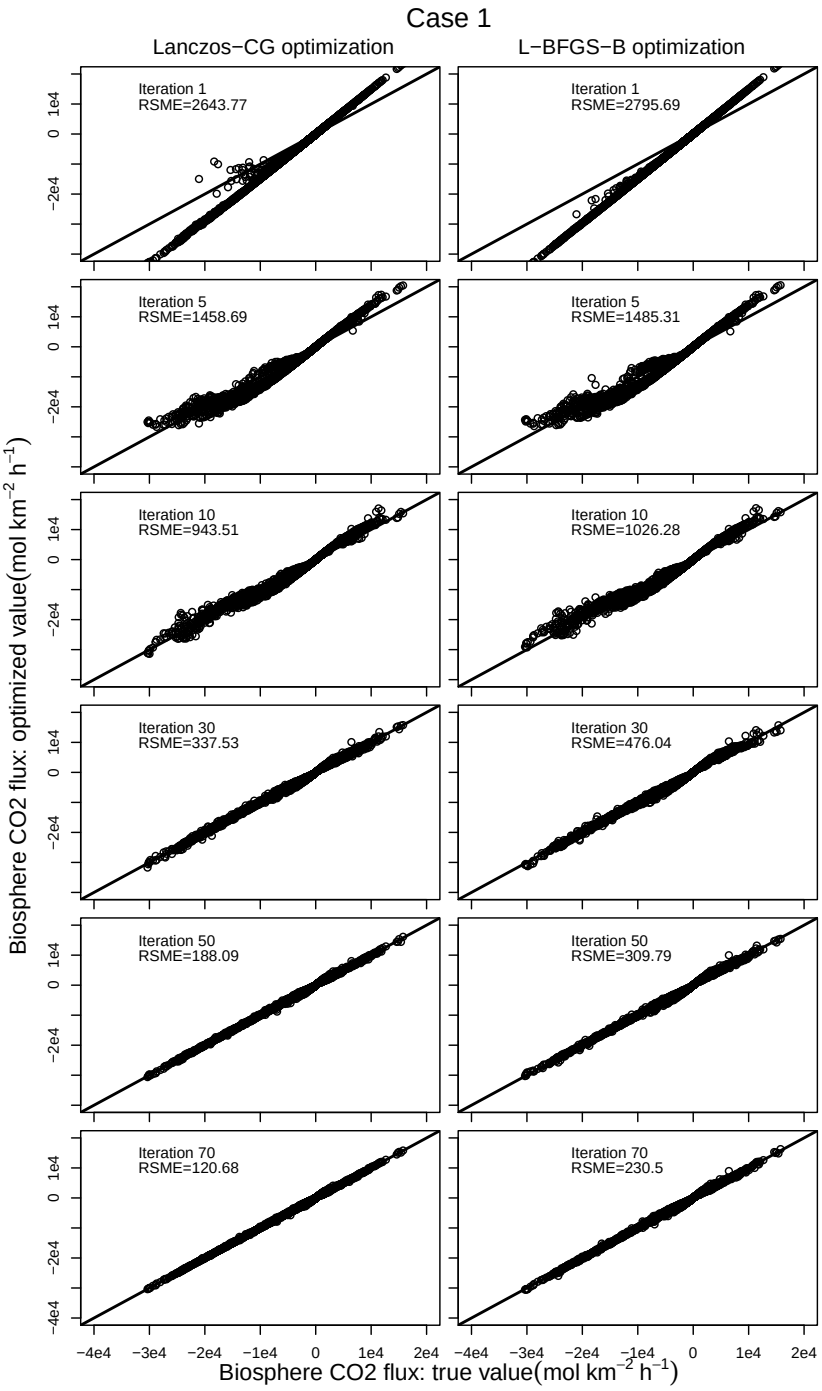


Figure 10. Comparison between the true and optimized CO_2 flux by ~~L-BFGS-B~~ Lanczos-CG (left column) and ~~Lanczos CG~~ L-BFGS-B (right column) in inverse modeling ~~case 2 (prior k_{co2} overestimates the true value by 50%). Optimized flux~~ experiment Case 1. The comparison and RMSE after the ~~1^{st}~~ 1^{st} 5^{th}, ~~5^{th}~~ 10^{th}, ~~10^{th}~~ 30^{th}, ~~20^{th}~~ 50^{th}, ~~and 29^{th} iterations~~ 70^{th} iteration are shown in the figure. All iterations of Lanczos-CG is from one outer loop.

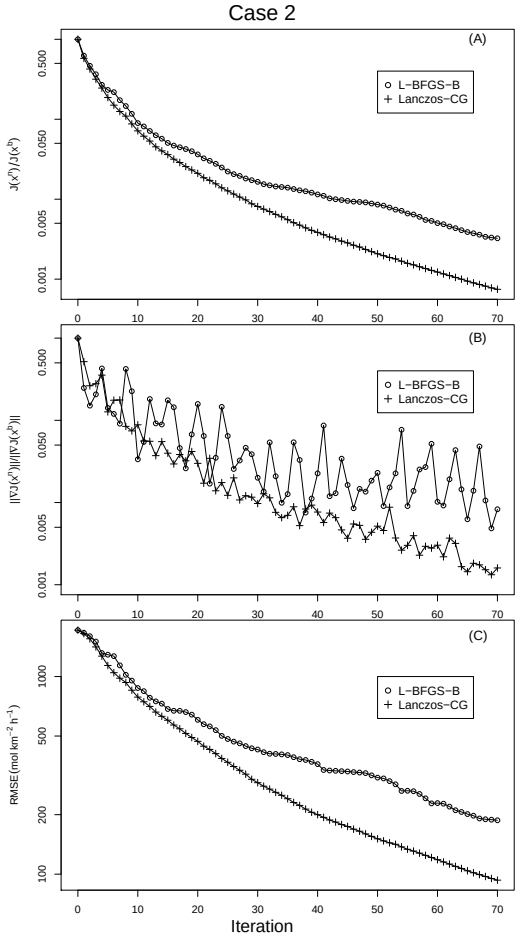


Figure 11. Same as Fig. 11, but for inverse modeling experiment Case 2.

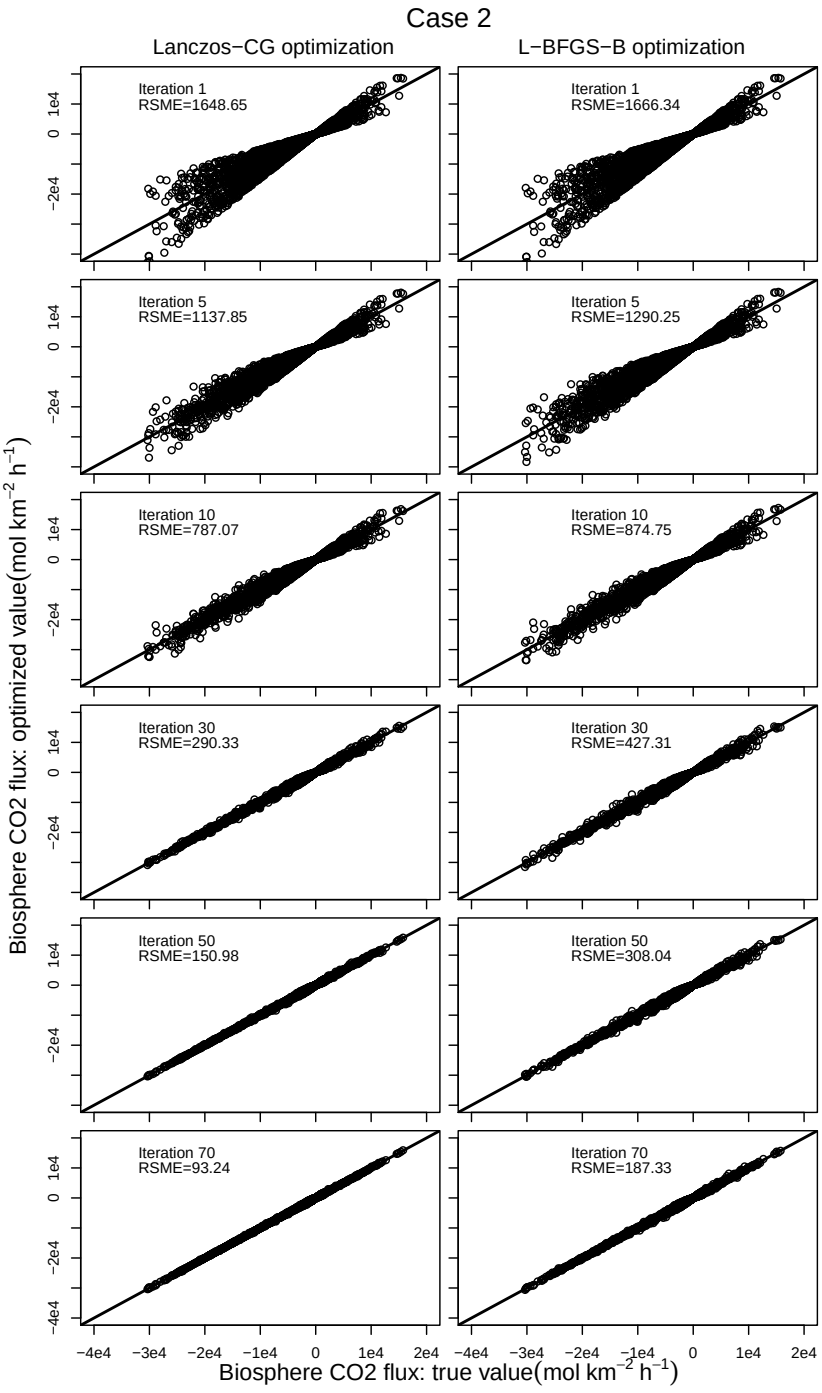


Figure 12. Same as Fig. 10, but for inverse modeling experiment Case 2.

Table 1. A list of symbols used in this article

~~$J(x)$~~ $J(\mathbf{x})$	Cost function
~~$J_b(x)$~~ $J_b(\mathbf{x})$	Background cost function
~~$J_o(x)$~~ $J_o(\mathbf{x})$	Observation cost function
~~$\nabla J(x)$~~ $\nabla J(\mathbf{x})$	Cost function gradient
~~$\|\nabla J(x)\|$~~ $\|\nabla J(\mathbf{x})\|$	Cost function gradient norm
~~$\nabla^2 J(x)$~~ $\nabla^2 J(\mathbf{x})$	Cost function Hessian
$\mathbf{B}$	Background error covariance
$\mathbf{R}$	Observation error covariance
M	WRF-CO2 forward model
$\widetilde{M}$	WRF-CO2 tangent linear model
$\widetilde{M}^T$	WRF-CO2 adjoint model
H	Observation operator
$\widetilde{H}$	Tangent linear observation operator
$\widetilde{H}^T$	Adjoint observation operator
~~k_{co2}~~ $\mathbf{k}_{co2}$	CO_2 emission scaling factor
~~q_{co2}~~ $\mathbf{q}_{co2}$	CO_2 mixing ratio (dry air)
~~g_k_{co2}~~ $g\_\mathbf{k}_{co2}$	Tangent linear variable for CO_2 emission scaling factor
~~a_k_{co2}~~ $a\_\mathbf{k}_{co2}$	Adjoint variable for CO_2 emission scaling factor
~~g_q_{co2}~~ $g\_\mathbf{q}_{co2}$	Tangent linear variable for CO_2 mixing ratio (dry air)
~~a_q_{co2}~~ $a\_\mathbf{q}_{co2}$	Adjoint variable for CO_2 mixing ratio (dry air)
~~x^b~~ $\mathbf{x}^b$	Prior estimate of CO_2 emission scaling factor
~~x^n~~ $\mathbf{x}x^n$	Analysis of CO_2 emission scaling factor
~~$\hat{x}$~~ $\hat{\mathbf{x}}$	Analysis increment of CO_2 emission scaling factor
~~y_k~~ $\mathbf{y}_k$	Observation at the k^{th} assimilation window
~~d_k~~ $\mathbf{d}_k$	Innovation vector at the k^{th} assimilation window

Table 2. Summary of variable dependence analysis for developing WRF-CO2 4DVar component models on top of WRFPLUS. In the table, an 'F' means a full physics scheme is used in the forward model, tangent linear model, or the forward sweep of the adjoint model. An 'X' means a process is not needed for CO_2 treatment. A 'Dev' means a process does not exist in WRFPLUS and has been developed for WRF-CO2 4DVar. An 'Add' means a process for CO_2 is simply added using the existing WRFPLUS code for other tracers.

Process	Forward model	Tangent linear model	Adjoint mode forward swee
Chemistry	X	X	X
Photolysis	X	X	X
Dry deposition	X	X	X
Wet deposition	X	X	X
Radiation	F	F	F
Surface	F	F	F
Cumulus	F	F	F
Microphysics	F	F	F
Advection	F	Add	F
Diffusion	F	Add	F
Emission	F	Dev	F
PBL	F	Dev	F
Convective transport	F	Dev	F

~~Chemistry X X X X Photolysis X X X X Dry deposition X X X X Wet deposition X X X X~~ height

Table 3. WRF-CO2 4DVar model configuration and ~~emission inventories~~ CO_2 flux used in sensitivity and inverse modeling tests.

Longwave radiation	Rapid Radiative Transfer Model (RRTM)
Shortwave radiation	Goddard shortwave
Microphysics	Thompson
Surface layer	Pleim-Xiu
Land surface	Pleim-Xiu
Planetary boundary layer	ACM2 PBL
Cumulus	Grell-Freitas
CO_2 advection	Positive-definite advection
~~Biosphere~~ biosphere CO_2 flux	~~CASA-GFED v4~~ CarbonTracker 2016
~~Anthropogenic CO_2 emission~~ ocean CO_2 flux	~~EDGAR v4.2~~ CarbonTracker 2016
~~Ocean CO_2 exchange~~ fire CO_2 flux	~~ECCO2-Darwin~~ CarbonTracker 2016
fossil fuel CO_2 flux	CarbonTracker 2016

Table 4. Summary of CO_2 tower sites. Sensitivity $\partial q_{co2}/\partial k_{co2}$ as calculated by WRF-CO2 4DVar's tangent linear and adjoint models is compared against finite difference sensitivity at these sites.

Site Name	Symbol	Latitude	Longitude
Kewanee	RKW	41.28^oN	89.77^oW
Centerville	RCE	40.79^oN	92.88^oW
Mead	RMM	41.14^oN	96.46^oW
Round Lake	RRL	43.53^oN	95.41^oW
Galesville	RGV	44.09^oN	91.34^oW
Ozarks	AMO	38.75^oN	92.2^oW
WLEF	LEF	45.95^oN	9.27^oW
West Branch	WBI	41.73^oN	91.35^oW
Canaan Valley	ACV	39.06^oN	72.94^oW
Chestnut Ridge	ACR	35.93^oN	84.33^oW
Fort Peck	AFP	48.31^oN	105.10^oW
Roof Butte	AFC_RBA	36.46^oN	109.09^oW
Storm Peak Lab	SPL	40.45^oN	106.73^oW
Argle	AMT	45.03^oN	68.68^oW
Harvard Forest	HFM	42.54^oN	72.17^oW
Southern Great Plains	SGP	36.80^oN	97.50^oW
Sutro	STR	37.75^oN	122.45^oW
Hidden Peak	HDP	40.56^oN	111.64^oW
Mary's Peak	ARC_MPK	44.50^oN	123.55^oW
KWKT	KWT	31.31^oN	97.32^oW

Table 5. Summary of inverse modeling experiment results. The reductions of cost function $J(x)$, ~~cost function~~ gradient norm $\|\nabla J(x)\|$, and RMSE are ~~givne~~ given as the ratio to their respective starting values. Results of the two experiment cases are values after 70 iterations.

Case 1

Reduction in	L-BFGS-B	Lanczos-CG
~~$J(x)$~~ $J(\mathbf{x})$	~~0.42×10^{-5}~~ 2.23×10^{-3}	~~1.54×10^{-5}~~ 4.72×10^{-4}
~~$\|\nabla J(x)\|$~~ $\|\nabla J(\mathbf{x})\|$	~~0.6×10^{-3}~~ 2.0×10^{-2}	~~1.35×10^{-3}~~ 1.7×10^{-3}
RMSE	~~0.72×10^{-2}~~ 8.01×10^{-2}	~~1.8×10^{-2}~~ 4.19×10^{-2}

Case 2

Reduction in	L-BFGS-B	Lanczos-CG
~~$J(x)$~~ $J(\mathbf{x})$	~~0.38×10^{-5}~~ 3.31×10^{-3}	~~0.87×10^{-5}~~ 7.76×10^{-4}
~~$\|\nabla J(x)\|$~~ $\|\nabla J(\mathbf{x})\|$	~~0.45×10^{-3}~~ 4.84×10^{-3}	~~0.91×10^{-3}~~ 1.32×10^{-3}
RMSE	~~0.86×10^{-2}~~ 1.09×10^{-1}	~~0.98×10^{-2}~~ 5.43×10^{-2}