Authors' Response to Reviewer Comments

The paper presents a newly developed WRF-CO2 4DVar assimilation system. The authors have documented the full system and provided a reasonable demonstration of the system. The system is well-documented and based on known optimization methods. The adjoint model has been carefully developed and implemented for tracer applications, and therefore provide an important step toward a full 4DVar system for regional inversions of fluxes. The main concerns and technical comments listed hereafter need to be addressed before final publication.

The reviewer has provided very insightful and constructive comments about our manuscript and the model code. We greatly appreciate the reviewer's time and effort for helping us improve our work. We have carefully considered the reviewer's comments and conducted further code development, simulation experiments, and manuscript revision to address these comments. This revision has substantially enhanced our understanding of the broad topic of atmospheric tracer data assimilation, in addition to improving the quality of this manuscript and our model code. For this, we express our sincere gratitude toward the reviewer. Detailed point-by-point responses are given following each comment below.

Note: all the references used in this response can be found in the reference of the revised manuscript.

General comments:

[1]- Missing definition of error covariances (both observations and prior flux errors): Here the system has been described and tested without any details on the construction of error covariances. The inverse framework needs to address that problem. The authors should include a paragraph about it.

In addition, one of the tests used an incorrect observation error covariance matrix R, inconsistent between the fully correlated errors (bias of 50% everywhere) and the R matrix (diagonal terms only - no correlation). The test should reconcile or address this problem.

The reviewer's comment on error matrices improved our understanding of this critical issue, especially with regards to real-data applications. The inverse modeling tests we used in this paper follow the same approach of Henze et al. (2007) regarding the error matrix specification and error-free synthetic data. But we failed to provide a clear justification for using the idealized error specification and a discussion of the need for more vigorous error matrix treatment in real-data applications. To address this, we have completely rewritten the inverse modeling section to: (1) provide a clear explanation of the experiment designs, with details about the **B** and **R** specification in light of the error-free synthetic data used in our inverse modeling tests. (2) A review section of error specification for real-data applications from the past studies. Please see Section 3.4.1 (Page 16 line 5- Page 17 line 26) of the revised manuscript for the detailed changes.

Regarding the reviewer's concern of the incorrect error covariance matrix R used in the Case 1 inverse modeling test (case 1: prior flux scaling factor set to 1.5 at all surface grid points),

we believe it was caused by our original manuscript's lack of clear description of the inverse modeling test setup. To address this issue, we provide a detailed explanation below and in the revised manuscript (Section 3.4.1). First, the synthetic data inversions used in our paper use a highly idealized configuration for the sole purpose of testing our model code accuracy. This is the same approach used in Henze et al. (2007) for testing the GEOS-Chem adjoint model. In this aspect, our inverse modeling tests are different from the synthetic data inversions typically used in observation system simulation experiments (OSSE). In OSSE type inversions, synthetic data are generated by a transport model driven by a true flux, then errors are added to the model simulated atmospheric CO_2 before being used to constrain the inversion. This process of generating synthetic data for OSSE inversion is described in detail in Chevallier et al. (2007). Since the synthetic observations contain errors, the observation error matrix must properly represent the error (both the variance and covariance) in this type of synthetic data inversions.

In comparison, the inverse modeling tests used in our paper are designed to test the adjoint model and optimization code accuracy, instead of an OSSE type synthetic inversion. This leads to a key difference in our inverse modeling test setup: *no errors were added to the forward model simulated CO_2.* Because the synthetic-observations contain no error, the observation error matrix **R** has no error correlation. The idea is to start with any prior flux, constrained by the error-free synthetic observations, the 4DVar system should converge to the true flux, thus confirming our assimilation system has been correctly implemented. In addition to the error-free observations, another key feature is that the background error matrix **B** is set to infinity (by setting $\mathbf{B}^{-1}=0$) and the observation error **R** is set to the identity matrix. The combination of the three features (error-free observations , $\mathbf{B}^{-1}=0$, and $\mathbf{R}=\mathbf{I}$), means that: (1) the background cost function is always zero: deviations of the analysis from the prior fluxes will incur no cost because $\mathbf{B}^{-1}=0$. (2) All the synthetic observations are given the same weight in the observation cost function because ($\mathbf{R}=\mathbf{I}$). (3) The 4DVar system is driven by the synthetic observations only, and the system iteratively decreases the cost function by moving the analysis from a given first guess flux to the true flux. Because deviations from the prior flux incur no cost and the observations contain no error, the system will eventually converge to the true flux. This is highly idealized setting, but is an effective way to examine assimilation system code accuracy, as done in Henze et al. (2007).

Given the above descriptions, we consider it is reasonable to use the same error configuration ($\mathbf{B}^{-1}=0$ and $\mathbf{R}=\mathbf{I}$) for the two cases of inversions (Case 1 and 2). Because both inversions were constrained by the same set of error-free synthetic-observations, $\mathbf{R}=\mathbf{I}$ is appropriate to assign all observations the same weight in cost function calculations. As the reviewer correctly pointed out that for Case 1, the first guess fluxes is perfectly correlated because the flux scaling factor is set to 1.5 for all surface grid points, thus all the off-diagonal elements of the background error matrix **B** (not the observation error matrix **R**) should be set to 1.0. However, this detail in **B** becomes irrelevant because **B** is set to infinity so that analysis deviation will incur no cost (so that the analysis can converge to the true flux). We hope this will provide a satisfactory response to the reviewer's concern. Please also see the rewritten Section 3.4.1 (Page 16 line 5- Page 16 line 27) of the revised manuscript for more details.

[2]- Boundary conditions: The inversion system needs to address this problem which is not trivial and could increase significantly the cost of the 4DVar. In addition, the optimization becomes more complicated as the two unknowns (boundary inflow and surface fluxes) could be optimized to correct for concentrations. The authors need to address that problem, and if not solving for, should describe a path toward implementing this part of the code. At this point, the WRF-CO2 4DVar is not usable for real-data inversions because of this gap. This is a major weakness in the current study.

We thank the reviewer for pointing out the importance of the tracer lateral boundary condition for our regional inversion system. To address this issue, we further developed our current 4DVar system to include the capacity of tracer lateral boundary condition optimization. The new code development includes: (a) the adjoint and tangent linear code of the tracer lateral boundary condition, and (b) expanding the two variational optimization schemes to accommodate the lateral boundary conditions in the state vector. The tracer lateral boundary condition adjoint and tangent linear code development are rather straightforward because they share the same code for dynamic (advection/diffusion), physics (ACM2 PBL) and chemistry (convective tracer transport in the chemistry module) with the surface flux code already developed in WRF-CO2 4DVar. Please note that WRF-Chem system does not use the relaxed boundary zone for chemistry variables (as compared to other scalar variables), and we followed the same treatment in the adjoint code development.

We have confirmed the accuracy of the tangent linear and adjoint code of the lateral boundary condition. However, we have not used the new code to test the combined boundary condition and flux optimization. As the reviewer pointed out, including lateral boundary conditions in the inversion is not trivial because it greatly increases the size of the state vector. Based on past regional inversion studies, we recognize that it may not be practical to optimize the lateral boundary conditions at the model grid resolution, and thus aggregation (both spatial and temporal) of the boundary condition scaling factor is necessary. Given this consideration, we implemented a mechanism in WRF-CO2 4DVar to allow a flexible mapping from the 3d lateral boundaries into the 1d state vector. This mapping gives the user the flexibility to determine whether and how to aggregate the lateral boundary condition scaling factors. While this new code has not been tested for the inverse modeling experiments in this manuscript, we do plan to test it in a follow-up study.

The description of the tracer lateral boundary condition code development is in Section 4 of the revised manuscript (Page 18 line 15-Page 20 line 2). We have reposted the WRF-CO2 4DVar code and it now includes the update with the tracer lateral boundary conditions.

[3]- Adjoint evaluation: Instead of a qualitative comparison of adjoint sensitivity and back-trajectories, the authors can compute the actual HYSPLIT footprints, and combine them with prior fluxes to compare to the adjoint sensitivity results. This analysis is fast considering the short simulation period and that the tools to compute tower footprints are publicly available. This evaluation would reinforce the confidence in the adjoint model.

We appreciate the reviewer's suggestion regarding the comparison. Through conducting further simulations, footprint calculations, and comparisons, our understanding of Eulerian vs.

Lagrangian tracer transport has been substantially enhanced, in addition to the improvement of the present manuscript.

In calculating the HYSPLIT footprint, we released 30,000 particles from each receptor location. Our sensitivity tests found that further increasing the particle number has negligible impact on the resulting footprint calculation (results not shown). All the backward trajectories used in calculating the footprint (Fig. 1 below and Fig. 11 of the revised text) were simulated with HYSPLIT using the PBL heights provide by the WRF generated meteorological data.

We summarized the new simulations and compared the resulting HYSPLIT footprints with that calculated by WRF-CO2 adjoint model (Fig. 2 below and Fig. 11 in the revised text). Note the adjoint sensitivities have been redrawn in a raster format (from the contours used in the original manuscript) to facilitate comparison with the HYSPLIT footprints. Since our adjoint model accuracy has been confirmed in this paper and the HYSPLIT model has been used in numerous studies, we provide a discussion for attributing the differences between the two sets of footprints. We believe that these differences are at least partially caused by: (1) the finite difference advection of WRF and Lagrangian advection of HYSPLIT. (2) The different treatment of vertical mixing in the boundary layer.

Through working through this part of the revision suggested by the reviewer, we realized that the detailed comparison between an Eulerian adjoint model and a Lagrangial model is of vital importance for real-data applications, but such comparisons are not been well documented in the literature. We acknowledge that our current simulations and analysis could be further improved regarding the impact of PBL and cumulus schemes, and advections. We plan to carry out further investigation along this direction to help us better understand and treat transport model errors for CO2 inversions in a follow-up work.

The above described HYSPLIT footprint calculations and comparison can be found at Page 15 Line 13 – Page 16 Line 2 of the revised manuscript.

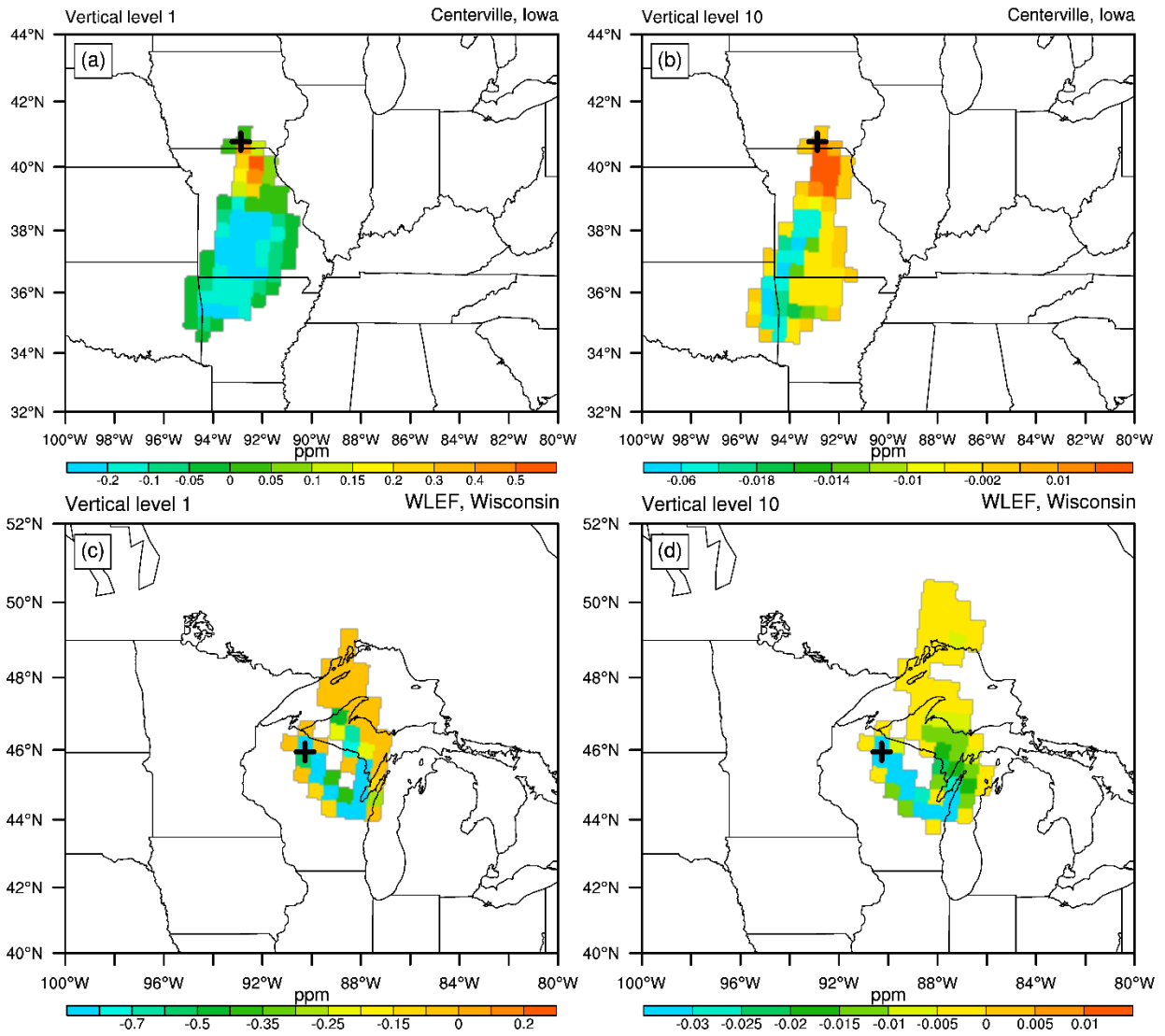

Figure 1. Footprints calculated using HYSPLIT backward trajectories for the four receptors (two vertical levels at each of the two tower sites). The black cross marks the tower location. These HYSPLIT footprints are calculated and plotted on the same grid and color scale as their WRF-CO2 adjoint sensitivity counterparts (Fig 2 next page).
Note: this figure is Fig. 11 of the revised manuscript.

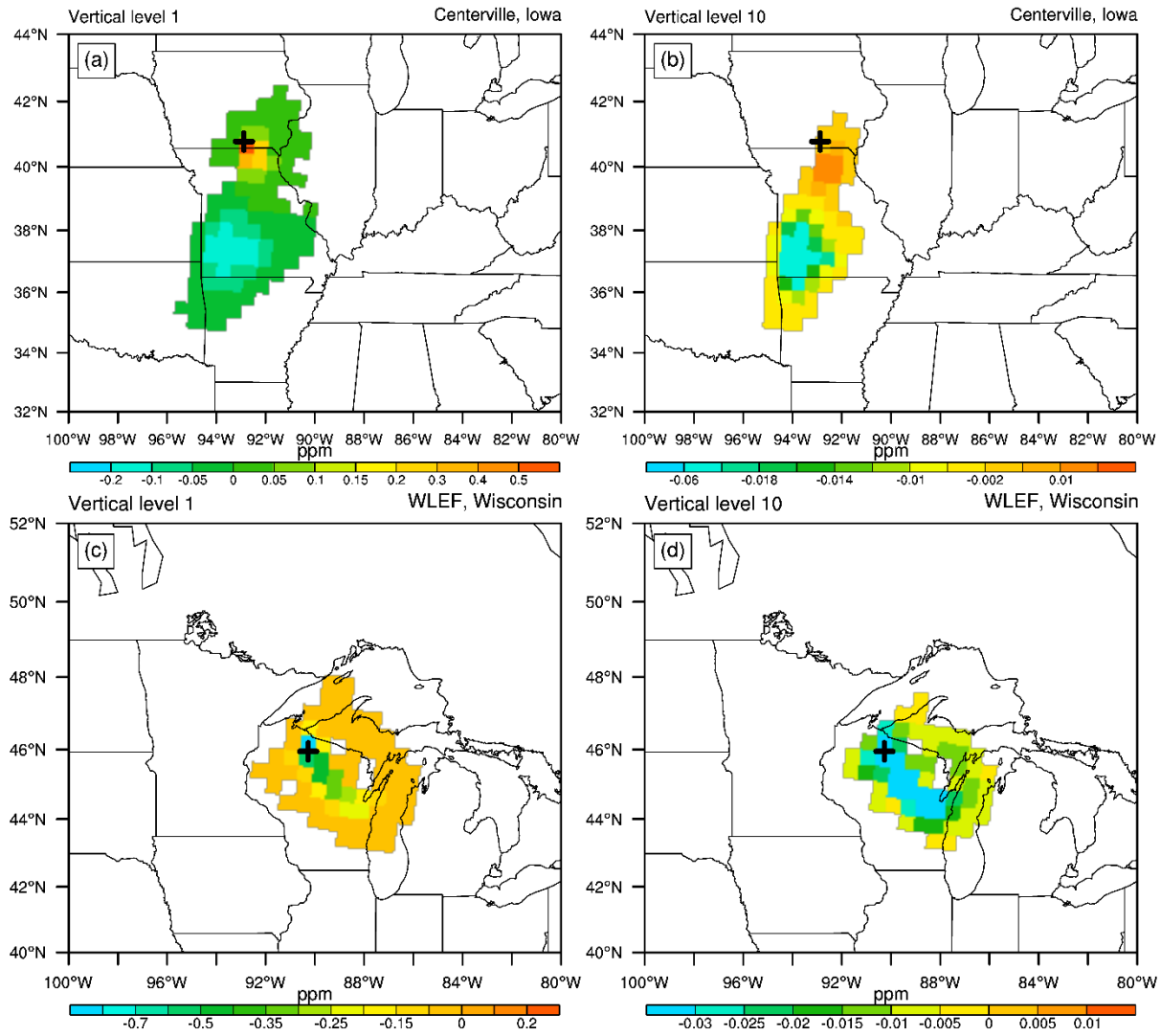

Figure 2. Footprint calculated by the WRF-CO2 adjoint model at four receptor locations (two vertical levels at each of the two tower locations). The black cross in each figure marks the location of the tower. Each footprint is plotted using same color scale with its HYSPLIT counterpart in Fig. 1 (last page). Note: this figure is Fig. 10 of the revised manuscript.

Technical comments:

[1] P1-L16-21: This paragraph needs to include more references. Only two papers are cited, somehow arbitrarily. Please add the references corresponding to the different statements. Chevallier et al., 2005 is a good reference but some of the major papers should also be cited here.

This paragraph has been completely rewritten to provide an organized review of the atmospheric CO2 inversion studies. Please see page 1 lines 16-24 of the revised text for details.

[2] P1-L22: An ensemble approach would also need a CTM. Correct the statement.

This sentence has been corrected in the revised manuscript (Page 2 line 3-4).

[3] P1-L23: What about the boundary conditions? You are developing a regional system within a bounded domain. The state vector also includes prior information related to the boundary inflow. Revise the sentence.

The sentence has been revised to include the CO2 lateral boundary condition in the state vector. Please see the page 2 line 1-2 of the revised manuscript. Regarding tracer lateral boundary conditions, please also refer to our response to general comment [1].

[4] P2-L1: Same comment related to boundary conditions. Some bibliography on this problem is needed here. Background conditions are critical to limited-domain inversions.

Code development for the tracer lateral boundary conditions and related literature reviews has been added in a new Section 4 (Page 18 line 15-Page 20 line 2 of the revised manuscript).

[5] P2-L2-3: Why do you bring the dimensions of the Jacobian in the introduction while not discussing it further in this paragraph?

The sentence about Jacobian matrix dimensions has been deleted.

[6] P2-L5-6: THis is only true for single tracer simulations in Eulerian mode. One can run multiple tracers in one simulation, or run millions of particles at once in a Lagrangian framework. Please revise this statement.

This sentence has been corrected. Please see page 2 line 10-14 of the revised text for details.

[7] P2-L6-11: The idea here is to explain the advantage of using a variational approach. While the arguments are correct, the justification lacks some clarity and a more rigorous description of the variational approach (minimzation using the gradients, adjoint model,...). There are also some disadvantages to the variational approach that should be stated here (no explicit posterior uncertainty, problem of convergence,...). This is an important paragraph that highlights the value of your work. You should revise this text to explain the major technical features in variational methods, and link it with the previous examples of 3d/4d-var studies in the following paragraphs.

The description of 4DVar assimilation has been rewritten to include the reviewer's suggestions. The new paragraph is in page 2 line 21-28 of the revised text.

[8] P2-L19: CarbonTracker is using a lagged Ensemble Kalman Smoother, not a variational approach. Revise the statement.

Thanks for pointing this out. CarbonTracker has been removed from the list of 4DVar

assimilations. It is discussed as an example for ensemble Kalman smoother in the revised manuscript (page 1 line 22, and page 2 line 16-19).

[9] P2-L13-26: this paragraph focuses entirely on 4d-var systems. What about other variational approaches?

Regarding other variational approaches: although there are a number of studies using 3DVar assimilation to estimate atmospheric chemistry mixing ratios, we have not found any 3DVar inversion studies for CO2 surface fluxes. This is most likely because 3DVar does not include a CTM and thus can not link fluxes to atmospheric mixing ratio. To strengthen this section, we added discussion about ensemble Kalman filter smoother (see our response to technical comment [8]). Page 2 line 16-19 of the revised manuscript.

[10] P2-29: More examples are needed here. Gerbig et al. (2009) is not an inversion but an overview on regional inverse modeling strategies. Include previous regional inversion studies with their achievements.

This has been fixed. In the revised text, a number of references to regional inverse modeling are given in page 3 line 10-24.

[11] P2-L31: GEOS-Chem is a global system not optimla for regional applications. Multiple examples of regional inversions have been published over the years.

GEOS-Chem reference has been removed from the regional inversion application list.

[12] P2-L34: Include references for LPDM-based inversion studies.

References to Lagrangian Particle Dispersion Model based inversion are in Page 3 Line 10-24 of the revised manuscript.

[13] P2-L35: "assimilated meteorology". Do you mean "meteorological analyses"?

“assimilated meteorology” has been corrected as “meteorological analyses”

[14] P3-L1-10: These are the studies you need to describe up front. This paragraph should be merged with the previous one.

The description about the past global and regional inversion studies have been moved together. Page 2 line 30- Page 3 line 24 of the revised manuscript.

[15] P3-L12: This statement should come before describing inversions and methods.

This statement has been modified and moved to the beginning of the first paragraph (page 1 line 16-17 of the revised manuscript).

[16] P3-L13: The differences in CMS products is an illustration but other examples would be better suited here. Consider inter-comparison studies in particular (e.g. Peylin et al. (2013) for CO2).

The CMS reference has been replaced with inversion inter-comparison studies including (Gurney et al 2002 and Peylin et al 2013), and the sentence has moved to page 1 line 16-17 in the revised manuscript.

[17] P3-L14-15: "high priority". Confusing. What do you mean here? Please rephrase.

Considering that the need for developing a high resolution 4DVar system has been made in other statements, this sentence is removed in the revised manuscript.

[18] P3-L29: The problem of convective transport for tracers is not trivial, and most of the convective schemes in WRF do not even produce mass fluxes explicitly. Some of them use a simple parameterization with 1D variables while other schemes only provide mass fluxes at the cloud base/top which is insufficient for convective mass transport. Few of the schemes have a full 3D mass flux, and mass conservation is not even guaranteed. Please explain with more details here which schemes would allow for convective transport of tracers with "limited new code development".

We agree with the reviewer's opinion about the importance of the tracer convective transport. To address this issue, the convective transport in Section 2.4.3 has been strengthened into a separate paragraph, which now includes details as suggested by the reviewer. In the revised text, we point out the following facts about the WRF-Chem tracer convective transport: (1) the tracer convective transport in WRF-Chem is treated by a simplified Grell convective scheme in the chemistry module, instead of by the cumulus schemes in the physics package. (2) This simplified Grell scheme for tracer convective transport uses convective precipitation rain rate to scale the base mass flux. This is a rather crude presentation comparing to the stochastic ensemble closure used in Grell and Freitas (2014).

Please see page 10, line 16-25 of the revised manuscript for the detailed changes.

[19] P3-L34: Justify this statement with a quantity or a reference.

Two references have been added justify this statement.

Please see page 4, line 9-10 of the revised manuscript for the detailed change.

[20] P4-L1: Typo. Subscript for "2" missing in "CO2".

Fixed.

[21] P4-L1-4: Provide a simple quantity instead of several sentences. For example, calculate the

impact on the solar energy when considering higher CO2 concentrations at the local scale. Or cite a study describing it.

To provide a quantitative assessment of CO_2 impacts on meteorology for short term simulations, we conducted the following sensitivity study. Using the model set up as described in Section 3.1 of the manuscript, two 48-hours simulations were conducted. The CO_2 mixing ratio is set to 391 ppm in the first and 500 ppm in the second simulation. Because the two simulations are identical otherwise, differences in meteorology fields were attributed to the CO_2 difference. The results show that magnitude of difference in horizontal wind and temperature are very small: Mean/standard deviation of U, V, and T at the end of the 48 hours simulation are 0.0794/0.1408 m/s, 0.0791/0.1459 m/s, and 0.0366/0.0614 K.

The above sensitivity results are summarized in page 4 line 9 -14 of the revised text.

Note: Radiative schemes in WRF-Chem do not use the predicted CO_2 mixing ratio (from the chemistry module) for radiative transfer calculations, they instead use either fixed values or interpolated climatological values.

[22] P4-L10: Typo. "estimated"

Fixed.

[23] P4-L6-12: This is even more relevant for turbulence with eddy turn-over times in minutes.

Thanks for pointing this out. The sentence has been revised to include the importance of online chemistry transport regarding the turbulence mixing. Page 4 line 20 of the revised manuscript.

[24] P4-L22: Replace "observational data" by "atmospheric observations" or "observations"

Changed to “observations”

[25] P4-L22 (also L23): Add "the" to "flux", or replace "flux" by "fluxes".

Changed to “the CO2 flux”.

[26] P4-L25: Replace "emission" by "fluxes". "emissions" refer to positive fluxes only.

“emission” has been replaced by “fluxes” here and a number of other places in the text to avoid confusion.

[27] P5-Eq 2-4: Describe the variables used in the equation (x, x0, y, H, and M) and refer to Table 1 for the full list of symbols.

A description of variables used in the equations and reference to Table 1 have been added as suggested by the reviewer. Page 5 Line 19-23 of the revised manuscript.

[28] P5-L10: Why "essentially"? Is it something else?

The word “essentially” has been removed from the sentence.

[29] P5-L11: Replace "emission" by "flux"

Fixed.

[30] P5-L12: Is a lagged approach? Or independent flux estimates? Refer to later sections to describe the approach.

WRF-CO2 4DVar is implemented for independent flux estimates. Reference to later sections for detailed descriptions has been added.

[31] P7-L17: Replace "CO2 related processes" by "physical and dynamical processes involved in the atmospheric transport of CO2"

The phrase has been replaced as suggested. Please see page 8 line 1 of the revised manuscript.

[32] P7-L20: Typo. "to keep"

Fixed.

[33] P7-L22: To be clear for the readers who are not familiar with WRF, the chemistry module was added, which is more accurate than stating that WRF-Chem "replaced" WRF. WRF is still used in WRF-Chem.

The sentence has been rephrased as the reviewer suggested (Page 8 line 2-3 of the revised manuscript).

[34] P7-L24-25: The tracer here has no impact on the code. One can substitute CO2 by CO or CH4. Assuming there is no chemistry involved, the transport of an inert and mass-free tracer has no incidence on the selected gas. Revise the sentence.

The sentence has been revised as suggested by the reviewer. Page 8 Line 5-9 of the revised manuscript.

[35] P7-L27: It means that you removed the GHG option and used a passive tracer. It would be clear to the future users if you state that. WRF-CO2 assumes that VPRM is used for the biogenic component. Otherwise it becomes equivalent to the original passive tracer mode. Clarify.

The sentence has been revised to clarify that VPRM is not used and CO2 is treated as a passive tracer. Please see page 8 line 5-9 of the revised text.

[36] P8-L25: Typo. "was"

Fixed.

[37] Section 2.4.2: This section is clear and helpful for future code development. One comment

here: Additional thoughts on joint optimization of meteo and CO2 data would be useful. How could a combined assimilation be implemented using your 4DVar system?

Thanks. A paragraph has been added about possible extension to a joint optimization of meteorology and CO_2 fluxes. This new paragraph (page 13, line 1-19 of the revised manuscript) is placed after the WRF meteorology field validation and discussed in the context of addressing transport error and bias for CO_2 inversion. This new paragraph discusses the possibility of using a combined meteorology and CO_2 4DVar inversion to correct transport model errors. It also points out that if CO_2 feedback on meteorology is not considered (which is insignificant for short term simulation, please see our response to technical comment [21]), the primary code modification for expanding to such a joint 4DVar system will be in the optimization schemes, because the adjoint code for meteorology (in WRFPLUS) and CO_2 (WRF-CO2 4DVar) has already been implemented.

Please see Page 13 Line 1-19 of the revised manuscript for the detailed change.

[38] P9-L29: The argument to select the PBL scheme is understandable, but lacks some scientific background. Is ACM2 a good option for CO2 turbulent mixing in the PBL? Several schemes have been tested with known systematic errors. The selection process should be based on model performances rather than technical reasons. You should at least comment on the model performances of the ACM2 PBL scheme. Add some metrics or published studies to assess the ACM2 schemes.

The justification for choosing ACM2 for PBL parameterization has been strengthened and forms a separate paragraph in the revised manuscript. It includes a brief description of ACM2 and its performance assessment in the published literature.

Please see page 10 line 4-14 of the revised manuscript for the detailed changes.

[39] P9-L31: To clarify this sentence, you need to explain that this scheme is parameterized based on precipitation rate. It is not the actuall mass flux from the Grell scheme but rather a crude representation of the vertical convective transport. Clarify in the text.

This has been clarified as suggested by the reviewer. Please see page 10 line 19-21 of the revised manuscript.

[40] P10-L27: "biospheric"

Fixed.

[41] P10-L27: Biospheric fluxes vary diurnally from negative to positive values depending on the time of day. For this reason, inversion problems need to address separately the two components or by time of day (night versus day) or by component (respiration versus photosynthesis). The daily mean will be irrelevant when transported into the concentration space as the timing of the atmospheric mixing is coupled to the timing of the fluxes and therefore cannot be simply averaged over an entire day. The 4DVar could be used for 3-hourly fluxes,

which would be more accurate and avoid biases due to day/night components. This problem has been discussed in several papers (e.g. Gourdji et al., 2012 - https://www.biogeosciences.net/9/457/2012/bg-9-457-2012.html). Whereas this problem will not be critical in a pseudo-data study, it will be critical in real-data inversions.

We agree with the reviewer about separating the biospheric fluxes by time in inversion. The present implementation of WRF-CO2 4DVar allows the flexibility for choosing the temporal resolution of fluxes inversion: the fluxes can be optimized at the same temporal resolution as the flux data allow. For instance, as pointed out by the reviewer, the optimization can be carried out on 3-hourly fluxes when using the CarbonTracker fluxes as prior. We added texts to clarify this issue and to emphasize that for real-data application, time varying fluxes should be used.

Please see Page 11 Line 18-23 of the revised manuscript for the detailed change.

[42] P10-L31 to P11-L4: The definition of boundary conditions by a global model is highly uncertain and cannot be ignored in the optimization process. Several studies have discussed that problem (Trusilova et al., 2010; Schuh et al., 2010; Goeckede et al., 2011; Lauvaux et al., 2012). See the general comment. If the problem is not solved in this paper, it should be highlighted as a major shortcoming in this study.

We agree with the reviewer that it is imperative for a regional inversion to address the boundary condition. In response, we have developed this part of code. Please refer to our detailed response to major comment [2].

[43] Figure 5: Conventional contour lines to illustrate pressure systems and frontal structures would be much easier to catch for the readers.

Figure 5 has been replotted following the reviewer's suggestion. To avoid clustering the figure, the state boundaries are not included.

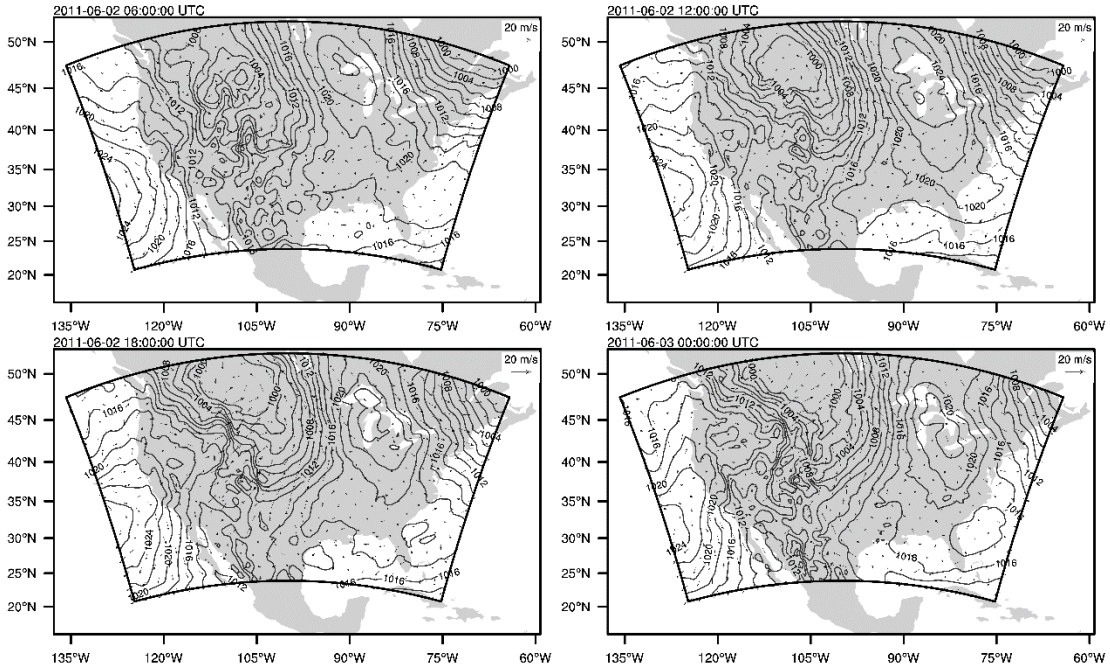

Figure. Sea level pressure (hPa) and horizontal wind (m s^{-1}) at models's lowest vertical level plotted during the 24-hour simulation starting at 2011-06-02 UTC

[44] P11-L22: Not only variances but covariances as well.

Thanks! Corrected.

[45] P11-L29 to P12-L7: This evaluation is limited to wind speed and direction but is not directly representative of CO2 concentration errors. Clarify here how these errors could be used to inform about transport model error variances and covariances.
From Figure 6, the wind speed is biased high, which is consistent with other studies. How would this problem be considered in the 4DVar framework?

Errors in meteorology fields will cause transport model error, which is a major part of the data-model mismatch error. For real observation applications, transport model errors are specified as part of the observation error covariance matrix R. Large transport error should be represented by higher variance/covariance in R to reduce the weight of the data-observation mismatch.

We think the wind bias can be addressed by nudging the wind field toward the observations. For instance, Gupta et al. (1997) found that nudging model simulated winds in the boundary layer to the radar wind profiles substantially improved estimates of plume dispersion. Another approach for addressing meteorology simulation bias will be a joint meteorology and CO2 assimilation. We lay out a brief pathway toward expanding our current system to such a joint meteorology/CO2 4DVar assimilation system in the future.

The above statements has been incorporated in the revised manuscript (Page 13 Line 1-19). Please also see our response to technical response [37].

[46] Section 3.2: The evaluation is convincing and seems to confirm that the adjoint has been correctly implemented. but please explain why the differences are considered "accurate". Cite other references for the threshold. For example: why is "10^-10" acceptable?

Citations to other adjoint code development accuracy assessments have been added for reference in the revised text. The added citations include: Henzen et al (2007) for GOES-Chem 4DVar system and Meirink et al. (2008) for TM5 4DVar system.

[47] Section 3.3: This section remains qualitative and not highly informative. The adjoint sensitivity seems to agree with the overall shape of the footprint. This comparison would be more convincing if the footprints were computed using HYSPLIT and combined with the prior fluxes. A short simulation (24 hours in your case) is really inexpensive for a Lagrangian model and would provide an independent evaluation of your adjoint transport.

This technical comment is the same as the major comment [3]. Please see our detailed response in the major comment section.

[48] P15-L1: The definition of R is inconsistent for case 1. If all pixels are multiplied by 1.5, the error correlations are equal to 1 over the entire domain. But you defined the R matrix with an independent error space. Clarify.

This technical comment is the same as the major comment [1]. Please see our detailed response in the major comment section.

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

Tao Zheng[1,4], Nancy H.F. French[2], and Martin Baxter[3]

[1]Department of Geography, Central Michigan University, Mount Pleasant, MI. USA
[2]Michigan Tech Research Institute, Michigan Technological University, Ann Arbor, MI. USA
[3]Department of Earth and Atmospheric Sciences, Central Michigan University, Mount Pleasant, MI. USA
[4]Institute of Great Lakes Research, 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. In this paper, an adjoint based four-dimensional variational (4DVar) assimilation system, WRF-CO2 4DVar, is developed to provide an alternative approach. This system is developed based on the Weather Research and Forecasting (WRF) modeling system, including WRF-Chem, WRFPLUS, and WRFDA, all in version 3.6. In WRF-CO2 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~~ flux 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. WRFPLUS forward, tangent linear, and adjoint models are modified to include ~~CO_2 related processes~~the physical and dynamical processes involved in the atmospheric transport of CO_2. 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 (Peters et al., 2007). Both inventory based bottom-up and atmospheric inversion based top-down approaches have been widely used to investigate carbon sources and sinks. Most~~ While rising atmospheric CO_2 has been well documented by observations, major uncertainties still exist in attributing it to specific processes (Gurney et al., 2002; Peylin et al., 2013). Atmospheric CO_2 inversions estimate surface carbon fluxes from atmospheric CO_2 measurements. Since the early study by Enting et al. (1995), a large amount of effort has been devoted to developing and applying atmospheric CO_2 inversion methods. Most of these inversions are based on ~~Bayes theorem, in which CO_2 flux is optimized~~ a Bayesian framework, and a wide range of different approaches have been used, including: synthesis inversion (Rayner et al., 1999; Bousquet et al., 1999; Peylin et al., 2002; Gurney et al., 2002), geostatisitical estimation (Michalak et al., 2004; Gourdji et al., 2012), Kalman smoother (Bruhwiler et al., 2005), Ensemble

Kalman smoother (Peters et al., 2005), and 4d variational inversion (Chevallier et al., 2005; Baker et al., 2010). All of these inversion approaches are optimization systems which yield an optimal estimate of CO_2 fluxes by minimizing a ~~quadratic form cost functionconsisting of background cost and observation cost. The minimization of the cost functioncan be achieved by analytical or variational approaches. Chevallier et al. (2005) provides a concise explanation of the differences between the two approaches .~~

~~Both analytical and variational inversions~~ Bayesian cost function. Within these optimization systems, the observation vector is formed by atmospheric CO_2 measurements, and the state vector is formed by CO_2 fluxes and lateral boundary conditions (only for regional inversion systems). The relationship between CO_2 fluxes and atmospheric CO_2 is described by the influence function, which is also called the footprint or adjoint sensitivity. Because all of the inversion approaches use a chemistry transport model (CTM) to relate CO_2 ~~flux~~ fluxes to atmospheric CO_2. ~~From the perspective of an optimization system, atmospheric CO_2 forms the observation vector, and CO_2 flux forms~~ , the influence function in theory can be calculated by the CTM using a finite difference method. However the practical limits imposed by computational costs often necessitate the aggregation of flux to reduce the state vector size, which leads to aggregation errors (Bocquet, 2009; Kaminski et al., 2001; Turner and Jacob, 2015). In practice, different inversion systems use different approaches 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,~~ determine the influnce function. Some inversion systems, including synthesis inverison, geostatistical estimation, and Kalman smoother, require the influence function to be explicitly constructed before the inversion. In comparison, Ensemble Kalman smoother and 4DVar inversion do not require precalculation of the influence function. Precalculation of the influence function is typically carried out using a finite difference approach with the CTM when the state vector is smaller than observation vector, or by the adjoint model of the CTM when the observation vector is smaller than the state vector. While most of the Lagrangian CTM models have their adjoint developed together (Uliasz, 1993; Lin et al., 2003; Stohl et al., 2005; Stein et al., 2015), separate and considerable efforts were often needed to develop and maintain the adjoint for Eulerian CTM models (Hourdin et al., 2006; Meirink et al., 2008). An ensemble Kalman smoother requires neither 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 (Bocquet, 2009; Kaminski et al., 2001; Turner and Jacob, 2015). In comparison, variational approaches~~ nor precalculation of the influence function. Instead it creates an ensemble of CO_2 flux fields and runs a CTM for each ensemble member. By sampling the ensemble flux fields and their corresponding atmospheric CO_2, the ensemble Kalman smoother calculates the Kalman gain matrix without explicitly constructing the influence function (Peters et al., 2005), and it provides posterior flux error estimates. The main disadvantage of the ensemble Kalman smoother is that use of a small number of ensemble members

is likely to lead to misrepresentation of the true prior error variance.

Like ensemble Kalman smoothers, 4DVar inversions do not require ~~the Jacobian matrix to be explicitly constructed, instead they directly compute the product of the Jacobian with a forcing vector~~precalculation of the influence function, but they do require the adjoint of a CTM to calculate the observation cost function. 4DVar inversions use a CTM and prior CO_2 fluxes to calculate the simulated CO_2, which is ~~the gradient vectorused for optimizing the state vector.~~ compared with observations to obtain the innovative vector. The adjoint of the CTM is then used to calculate the cost function gradient based on the innovative vector. Through iterative minimization of the cost function, 4DVar inversions estimate the optimal posterior fluxes. By avoiding calculation of the influence function, 4DVar inversions enable CO_2 fluxes to be estimated at much higher resolution provided that sufficent observations are available. The major disadvantages of 4DVar inversions are: they do not explicitly provide posterior flux error estimates (additional computation is required), and their convergence is not always guaranteed.

~~A number of four dimensional variational (~~4DVar ~~) assimilation~~ systems have been ~~developed and applied to global scale CO_2 inversions.~~ widely used for CO_2 inversion at global and regional scales. Examples of global 4DVar inversions include the following: The off-line transport model Parameterized Chemistry Tracer Model (PCTM) (Kawa et al., 2004) and its adjoint have been used for CO_2 inversions (Baker et al., 2010, 2006; Butler et al., 2010; Gurney et al., 2005). Chevallier et al. (2005) developed a 4DVar system based on the LMDZ model (Hourdin et al., 2006) to assimilate CO_2 observation data from the Television Infrared Observation ~~the~~ Satellite Operational Vertical Sounder (TOVS). This system has also been used to invert surface CO_2 ~~observation data (Chevallier, 2007; Chevallier et al., 2010)~~observations (Chevallier, 2007; Chevallier et al., 2010). The TM5 4DVar system (Meirink et al., 2008), based on the TM5 global two-way nested transport model (Krol et al., 2005), is ~~used in the CarbonTracker CO_2 data assimilation system (Peters et al., 2007) and is~~ included in the TransCom satellite intercomparison experiment (Saito et al., 2011). TM5 4DVar has also been used to investigate total column CO_2 seasonal amplitude (Basu et al., 2011) and to assimilate the Greenhouse Gases Observing Satellite (GOSAT) observations (Basu et al., 2013). Another widely used inversion system is the GEOS-Chem 4DVar (Henze et al., 2007; Kopacz et al., 2009) with its CO_2 module updated by Nassar et al. (2010). GEOS-Chem 4DVar has been used to estimate CO_2 fluxes from the Tropospheric Emission Spectrometer (TES) and the GOSAT CO_2 observations (Nassar et al., 2011; Deng et al., 2014), and it is also part of JPL's (Jet Propulsion Laboratory) Carbon Monitoring System (Liu et al., 2014)

~~CO_2 inversions at regional scale have become an active research front in recent years, driven~~ Regional 4DVar inversion studies have been driven in part by the need to resolve biosphere-atmosphere carbon exchange at smaller scales (Gerbig et al., 2009), and by the need to address policy-relevant objectives, such as assessing emission reduction effectiveness (Ciais et al., 2014) and the impact of regional scale sources like wildland fire (French et al., 2011). ~~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 (Wecht et al., 2014). The majority of regional inversions use analytical approaches and typically use a Lagrangian particle backward dispersion model (LPDM) to compute~~

the required influence function. For instance, Gerbig et al. (2003) used an analytical approach to minimize the cost function and the STILT (Lin et al., 2003) model driven by assimilated meteorology meteorological analyses to calculate the influence function. In a later study, STILT driven by ECMWF meteorology the European Center for Medium-Range Weather Forecasts (ECMWF) meteorological data is used to calculate the influence function to investigate the impacts of vertical mixing error (Gerbig et al., 2008). More recently, Lauvaux et al. (2012) also used an analytical solution for cost function minimization and LPDM (Uliasz, 1993) to compute the influence function. In another study, Pillai et al. (2012) used STILT driven by meteorology meteorological 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 (Nehrkorn et al., 2010). More recently, Alden et al. (2016) investigated biopheric CO_2 flux in the Amazon using an analytical inversion approach (Yadav and Michalak, 2013) with the influence function calculated by the STILT and Flexpart (Stohl et al., 2005) models. Also, Chan et al. (2016) applied a regional CO_2 inversion in Canada with both analytical and Markov chain Monte Carlo (MCMC) LPDM based approaches. Influence The influence function is also calculated with the Flexpart model in this study.

While 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 (Ott et al., 2015). 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 (Gerbig et al., 2009). Toward this endIn this paper, a regional CO_2 inversion system with online meteorology, WRF-CO2 4DVar, was developed by modifying modifying the WRFDA and WRFPLUS system (v3.6) in an approach similar to that used for black carbon emission inversion by Guerrette and Henze (2015, 2017) (hereafter GH15/17). WRFDA is a meteorology data assimilation system, which includes a 4DVar assimilation system (Barker et al., 2012; Huang et al., 2009) and related adjoint and tangent linear models (WRFPLUS) (Zhang et al., 2013). Designed to improve weather forecasts, WRFDA 4DVar optimizes meteorological initial and boundary conditions by assimilating a variety of observational data. WRFPLUS was modified to include CO_2 related transport processes and the cost function was configured so that the state vector consists of CO_2 flux fluxes instead of meteorological fields. In developing WRFDA-Chem for black carbon inversion, GH15/17 excluded radiation, cumulus, and microphysics parameterization schemes from the tangent linear model and adjoint model and adjoint models 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 inclusion of the full physics schemes in WRF-CO2 4DVar's tangent linear model and adjoint model and adjoint models with limited new code development (see Section 2.4.2.1). As transport model error is detrimental to 4DVar inversion accuracy (Fowler and Lawless, 2016; Gerbig et al., 2009), it is beneficial to use the full physics schemes in the tangent linear and adjoint models for WRF-CO2 4DVar. In addition, while GH15/17 excluded convective transport of chemistry species in WRFDA-Chem, the tangent linear and adjoint code for this process was developed for WRF-CO2 4DVar to reduce the vertical mixing error (see Section 2.4.4). Like GH15/17, .3). Two optimization schemes were developed for WRF-CO2 4DVar: an incremental optimization with Lanczos-CG was developed, as well as and an L-BFGS-B

based optimization.~~but we also implemented an L-BFGS-B based optimization.~~

In the WRF system, ~~CO2~~ CO_2 mixing ratio variations ~~can impact the~~ could impact meteorology fields through the radiation scheme~~. This feedback is ignored in WRF-CO2 4DVar and instead CO2 is modeled as a passive tracer. WRF-CO2 4DVar is designed for regional inversions with a~~ , which proivdes temperature tendency to the dynamical core (Iacono et al., 2008; Skamarock et al., 2
None of the radiation schemes (as of Version 3.6) use the simulated CO_2 from the chemistry module, but instead use climatological values. A sensitivity test was conducted to assess the short term impacts of CO_2 variation on meteorology. The results show that with CO_2 mixing ratio changed from 391 ppm to 500 ppm, the mean difference in horizontal wind (U,V) and air temperature at the end of the 48-hour simulations are 0.0794 ms^{-1}, 0.0791 ms^{-1}, and 0.0366 K, respectively. Although these differences grow with time, their magnitude are considerably smaller compared with the contribution from other factors for the short assimilation window (days to weeks) ~~. For such applications, the potential feedback of CO2 variation on the meteorology is insignificant for most cases. Including such impacts would require a large amount of new code development while offering very limited performance improvement. By ignoring the CO2 feedback,~~ that WRF-CO2 4DVar ~~can~~ is designed for. Based on the above analysis, the impact of CO_2 on meteorology is ignored in WRF-CO2 4DVar and CO_2 is modeled as a passive tracer. This simplification allows WRF-CO2 4DVar to use the full version of most WRF physics schemes in its tangent linear and adjoint models to minimize the linearization error (Tremolet, 2004).

Compared with offline regional inversion systems, WRF-CO2 4DVar has an advantage provided by the close one-way coupling between meteorological processes and chemistry transport. For example, adequately ~~resolving CO2 vertical transport~~ representing CO_2 vertical transport and eddy turbulent mixing in high resolution regional simulations is crucial~~and~~ , as vertical motions in the atmosphere exhibit significant temporal variability. Grell et al. (2004) shows that less than 40% of the total vertical velocity variability in a 3 km resolution simulation is captured by a 1-hour output interval. He ~~estimate~~ estimated that the meteorological output interval must be less than 10 minutes in order to capture more than 85% of the variability in cloud resolving simulations. In WRF-CO2 4DVar, ~~CO2~~ CO_2 transport runs at the same time step as the meteorology, avoiding the problems facing its offline counterparts.

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. Section 4 describes the treatment of CO_2 lateral boundary conditions in the WRF-CO2 4DVar system. Finally, a summary and outlook are presented in Section ~~4.~~ 5.

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 the CO_2 flux by assimilating CO_2 ~~observational data~~ observations into an atmospheric chemistry transport model. The CO_2 flux is optimized through use of a linear scaling factor:

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

Where $\widetilde{E}$ is the CO_2 ~~emission read from emission~~ flux read from flux files, k_{co2} is the ~~emission~~ flux scaling factor, and E is the effective CO_2 flux. It is the effective flux E that is used in WRF-Chem's emission driver to update CO_2 mixing ratio (q_{co2}). The ~~emission~~ flux 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 $J(\mathbf{x})$ 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 the background cost function $J_b(\mathbf{x})$ 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 $J_o(\mathbf{x})$ is defined as

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

In Eqs. (3-4), ~~the~~ $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, M is the transport model, and H is the observational operator, $\mathbf{y}_k$ is the observation vector, $\mathbf{x}^b$ is the prior state vector. The superscript n indicates that $\mathbf{x}^n$ is the optimized state vector at the n$^{\text{th}}$ iteration. For a full list of variables used in this paper, please refer to Table 1. Throughout the paper, bold face lower case characters represent vectors and bold face upper case characters represent matrices.

Like other data assimilation systems, WRF-CO2 4DVar is ~~essentially~~ an optimization scheme. Its state vector $\mathbf{x}$ consists of the ~~emission~~ flux scaling factors k_{co2} and lateral boundary condition scaling factors. The summation K in Eq. (4) indicates the entire assimilation time period is evenly split into K observation windows during which observational data are ingested

into the assimilation system. Details about how observations are ingested in the variational optimization schemes are given in Section 2.2 and 2.3

Two optimization schemes are implemented in WRF-CO2 4DVar to minimize the cost function. The first scheme uses a limited memory BFGS minimization algorithm (L-BFGS-B) (Byrd et al., 1995) and the second uses the Lanczos version of the conjugate gradient (Lanczos-CG) (Lanczos, 1950) 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 (Byrd et al., 1995) is a quasi-Newton method for nonlinear optimization with bound constraints. L-BFGS-B has been used in a number of atmospheric chemistry inverse modeling systems, including the GEOS-Chem adjoint model system (Henze et al., 2007) and the TM5 4DVar system (Meirink et al., 2008). 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 its 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 $H(M(\mathbf{x}^n))$ term in Eq. (4), which is then paired with the observation vector $\mathbf{y}_k$ to calculate the innovation vector $\mathbf{d}_k = H(M(\mathbf{x}^n)) - \mathbf{y}_k$. Next, the innovation vector and observation error covariance $\mathbf{R}$ are used to calculate the observation cost function $J_o(\mathbf{x})$ as expressed in Eq. (4). Finally, the background cost function $J_b(\mathbf{x})$ is calculated according to Eq. (3), and combined with the observation cost function $J_o(\mathbf{x})$ to form the total cost function $J(\mathbf{x})$ according to Eq. (2).

L-BFGS-B requires the values of the cost function $J(\mathbf{x})$ and its gradient $\nabla J(\mathbf{x})$ in searching for the optimized k_{co2}. The gradient is calculated using Eq. (5).

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

The first term on the right hand side of Eq. (5) is the observation gradient and the second is the background gradient. The observation 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 $\widetilde{H}^T\mathbf{R}^{-1}(H(M(\mathbf{x}^n)) - \mathbf{y}_k)$, which is the adjoint forcing. (2) The adjoint forcing is then ingested by the WRF-CO2 adjoint model during its backward (in time) integration, which yields the observation gradient. Supplied with the values of the cost function and 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 (Zhu et al., 1997). Version Lbfgsb.3.0 (Luis Morales and Nocedal, 2011) will be implemented in the next model update.

2.3 Incremental optimization

The second optimization scheme implemented for WRF-CO2 4DVar ~~was~~ is the incremental approach commonly used in NWP data assimilation systems, including ECWMF 4DVar (Rabier et al., 2000) and WRFDA (Barker et al., 2012). 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):

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

Compared to Eq. (4), Eq. (6) approximates the innovation vector by a sum of two parts. The first part, $H(M(\mathbf{x}^{n-1})) - \mathbf{y}_k$, is the innovation vector from the previous iteration. The second part, $\widetilde{H}(\widetilde{M}(\mathbf{x}^n - \mathbf{x}^{n-1}))$, is the state vector analysis increment $(\mathbf{x}^n - \mathbf{x}^{n-1})$ transformed by the tangent linear model $\widetilde{M}$ and tangent linear observation operator $\widetilde{H}$. With the linear approximation of the cost function the gradient is 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 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~~ optimized state vector in the inner loop. The outer loop first calls the forward model M and adjoint model $\widetilde{M}^T$ to calculate $\widetilde{M}^T\widetilde{H}^T R^{-1}(H(M(\mathbf{x}^{n-1}) - \mathbf{y}_k))$ and $B^{-1}(\mathbf{x}^{n-1} - \mathbf{x}^b)$, which remain unchanged during the subsequent inner loop calculation. The analysis increment $(\mathbf{x}^n - \mathbf{x}^{n-1})$ is optimized in the inner loop, which calls the tangent linear and adjoint models to calculate the third and fourth items of Eq. (7). Inner loop calculation is carried out by Lanczos-CG (Lanczos, 1950), which can optionally estimate eigenvalues of the cost function Hessian matrix ($\nabla^2 J(\mathbf{x})$). 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 (Barker et al., 2012; Huang et al., 2009). 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 the CO_2 flux, instead of the meteorological initial and boundary conditions. This difference means ~~CO_2 related processes~~ the physical and dynamical processes involved in the atmospheric CO_2 transport are needed in WRF-CO2 4DVar's model components. To include ~~the CO_2 related processes, WRF-Chem was used to replace WRF as the forward model. Then, a thorough variable dependence analysis was conducted 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~~

~~WRF-Chem replaced WRF as the forward model component of WRF-CO2 4DVar. As an atmospheric chemistry extension of WRF, WRF-Chem~~ these processes, the chemistry module was added to the forward model. The chemistry module includes chemistry, deposition, photolysis, advection, diffusion, and convective transport of chemistry species (Grell et al., 2005). These processes are included in different modules of WRF-Chem: ARW (Advanced Research WRF) dynamical core, physics driver, and chemistry driver. ~~The~~ GHG (Greenhouse Gas) tracer option ~~of WRF-Chem was used but with the CO and CH_4 removed , leaving only CO_2 related procedures~~was removed and CO_2 is treated as an inert tracer. In the emission driver, CarbonTracker 2016 version (Peters et al., 2007) replaces the online biogenic CO_2 model Vegetation Photosynthesis and Respiration Model (VPRM) (Mahadevan et al., 2008). This change is made because WRF-CO2 4DVar optimizes ~~for~~ the CO_2 flux instead of online emission model parameters.

2.4.1 Variable dependence analysis

The tangent linear and adjoint models of WRFPLUS ~~need~~ needed to be modified to include the ~~CO2 related processes~~ physical and dynamical processes involved in the atmospheric transport of CO_2, so that they will be consistent with the forward model. ~~The results of the~~ A thorough variable dependence analysis ~~is~~ was conducted and the results are 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, including gas and aqueous phase chemistry, dry and wet deposition, and photolysis. 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 addition, 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~~ and adjoint models. The difference 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, Tremolet (2004) 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~~and adjoint models. Since this second category of processes ~~are~~ is 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, the tangent linear code of the simplified versions of the cumulus, surface, and microphysics schemes, ~~were~~ was removed and replaced with ~~their corresponding~~ the code for the full schemes as used in the forward model. 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. The simplified versions of the cumulus, surface, and microphysics schemes used in the forward sweep of WRFPLUS's adjoint model ~~,~~ were removed and replaced with the full schemes used in the forward model. Since these processes do not directly modify CO_2 mixing ratio, their corresponding adjoint code ~~wes~~ was removed 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 the 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 made for advection and diffusion are described in Section 2.4~~.3~~.2, and those for emission, turbulent mixing in the PBL, and convective transport of CO_2 are detailed in Section 2.4~~.4~~.3.

2.4.2 Advection and diffusion of ~~CO2~~CO_2

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~~ for these processes has been implemented in WRFPLUS. It should be noted that the variables for these inert tracers are part of WRF, instead of WRF-Chem. WRF-Chem uses a separate array for its chemistry species. Since ~~WRF was replaced with~~ WRF-Chem is used as the forward model ~~in~~ of WRF-CO2 4DVar, the CO_2 mixing ratios are included in the chemistry array. In the GHG option of WRF-Chem used 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 inert tracers in WRFPLUS, subroutines solve_em_tl and solve_em_ad were modified to add the tangent linear and adjoint code for the advection and diffusion of the ~~chem~~ chemistry array. The modifications 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 (Zhang et al., 2013) 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.

2.4.3 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 (Fowler and Lawless, 2016). For instance, Stephens et al. (2007) 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 the A-SCOPE mission of 0.02 Pg C yr^{-1} per 10^6 km^2 (Houweling et al., 2010). In addition, Jiang et al. (2008) 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, 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~~ in the convective transport in the chemistry driver. The subgrid scale filter in the dynamical core treats both horizontal and vertical ~~diffusions~~diffusion, 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~~

For PBL parameterization, ACM2 ~~PBL scheme also treats chemistry species (?). The ACM2 scheme was chosen for in~~ (Pleim, 2007) was chosen for WRF-CO2 4DVar~~so that CO_2 vertical mixing is treated by the PBL parameterization.~~ . ACM2 is a hybrid local-nonlocal closure PBL scheme, and it updated the non-local scheme ACM1 (Pleim and Chang, 1992) by adding an eddy diffusion component. Because ACM2 explicitly defines local and nonlocal mass fluxes, it is particularly applicable for atmospheric chemistry simulations. In a one-dimensional model evaluation, ACM2 showed a very good agreement with large-eddy simulations for PBL heights with a very slight low bias (Pleim, 2007). In addtion to WRF, ACM2 has been implemented in the fifth-generation Pennsylvania State University-NCAR Mesoscale Model (MM5) and the Community Multiscale Air Quality (CMAQ) model. An evaluation using PBL heights derived from radar wind profiles showed that the MM5-ACM2 is capable of realistic simulation of PBL heights (Pleim, 2007). Hu et al. (2010) evaluated three WRF PBL schemes and found that ACM2 resulted in less bias than the local closure scheme Mellor-Yamada-Janjic (MYJ) when compared with surface and boundary layer observations. Furthermore, model evaluations also showed that ACM2 performed well with CMAQ for air pollution simulations (Nolte et al., 2015; Appel et al., 2017) .

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 (Grell et al., 2004). This convective transport module includes a deep convective process and a shallow convective process. The deep convective transport process requires the convective precipitation rate calculated by the cumulus scheme (in the physics module of WRF): It calculates the base mass flux based on the convective precipitation rate. Compared to the ensemble stochastic approach used in the Grell-Freitas cumulus scheme (Grell and Freitas, 2014), this is a rather crude representation of the vertical convective transport. The shallow convective process requires three parameters passed in from the cumulus scheme in the physics module: updraft

originating level, cloud top level, and total base mass flux. Only two cumulus schemes in WRF provide these parameters: the Grell-Freitas (Grell and Freitas, 2014) and Grell 3D Ensemble (Grell and Devenyi, 2002).

Because the ACM2 PBL and chemistry convective transport are not included in WRFPLUS, their tangent linear and adjoint code were developed for WRF-CO2 4DVar. First the automatic differentiation tool TAPENADE (Hascoet and Pascual, 2013) was used 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. Then the TAPENADE generated code was manually modified to remove redundancy and unnecessary loops. It should be pointed out that these code developments were 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 the 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. First the simulation model setup is described, then the sensitivity tests and inverse modeling experiments are presented.

3.1 Model setup

WRF-CO2 4DVar is setup with a domain covering the continental United States with 48 km $\times$ 48 km grid spacing and 50 vertical levels (Fig~~3~~. 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 (Mlawer et al., 1997), Goddard shortwave radiation (Chou and Suarez, 1999), Pleim surface layer (Pleim, 2006), Pleim-Xiu land surface model (Pleim and Xiu, 2003), ACM2 PBL ~~(?)~~(Pleim, 2007), Grell-Freitas cumulus (Grell and Freitas, 2014), and Thompson microphysics (Thompson et al., 2008). Positive-definite transport is applied to the transport of scalars and CO_2.

CO_2 fluxes used for the simulations are from the CarbonTracker 2016 version (hereafter CT2016) (Peters et al., 2007). 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~~ biospheric 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~~ biospheric flux used for the 24 hour simulation is shown in Figure 4. The model configuration and emission data used are summarized in Table 3. Although the

daily mean biospheric flux was used for the forward and inverse modeling in this paper, the WRF-CO2 4DVar implementation allows flexibility in configuring the prior fluxes. For instance, fluxes from respiration and photosynthesis can be estimated independently, and at higher temporal resolution (Gourdji et al., 2012). When using these options with real observations, the balance between the degrees of freedom in the state vector and the constraints provided by the observations need to be carefully considered.

Model 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 (Saha et al., 2014). CO_2 initial and lateral boundary conditions are from the CT2016 global 3×2 degree CO_2 mole fraction. A method similar to PREP-CHEM-SRC (Freitas et al., 2010) was used to horizontally and vertically interpolate CT2016 mole fraction data to the WRF grid.

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 5 shows the instantaneous values of Sea Level Pressure (SLP) and horizontal wind at the model's lowest vertical level at ~~each~~ every 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.

In the model setup, the initial and boundary meteorological conditions are generated by downscaling the CFSv2 data. Downscaling coarse resolution global reanalysis data could lead to poor WRF performance. Although this potential problem is not a concern for the present ~~pseudo data~~ pseudo-observation based inversion experiments, it must be properly treated in ~~the future application with true observation data~~future applications with real observations. Error in the initial condition will lead to erroneous ~~emission source~~ flux attribution, especially for inversions with a short assimilation window.

In order to be useful for applications which employ real observational data, WRF-CO2 4DVar requires accurate ~~simulation~~ simulations of the meteorological fields by the forward model~~, in addition to accurate tangent linear and adjoint models~~. Because transport error can only be partially accounted for in the 4DVar system through the observation error ~~variance~~covariance matrix, it is imperative to minimize errors due to inaccurate simulation of meteorological processes as much as possible. Although the present paper uses pseudo-observation data (which have zero transport error by definition) in its inversion experiments, future applications with ~~true observational data~~ real observations will require vigorous evaluation of the model simulated meteorology and associated transport error. In the following, the forward model simulated horizontal winds at the

surface and 500 hPa constant pressure surface are evaluated using in-situ measurements from weather stations and radiosondes.

For the surface level, WRF simulated 10m winds are compared against surface weather station measurements archived in the NOAA ~~Integrate~~ Integrated Surface Dataset (Smith et al., 2011). Hourly surface wind measurements from more than 2,000 stations within the WRF domain are used for the evaluation. Comparisons of wind speed and wind direction are carried out at the top of each hour during the 24 hour simulation period starting at 00:00 UTC 02 June 2011. Excluding missing observations, this results in 31,745 valid data pairs, which are summarized in the histograms of Fig. 6. RMSE for the hourly wind speed is 2.16 m s^{-1} and the mean difference in the hourly wind direction is 29.4°.

For the upper level, WRF simulated 500 hPa horizontal winds were compared against radiosonde measurements from 90 stations obtained from the NOAA/ESRL radiosonde database (https://ruc.noaa.gov/raobs/). Since most stations release ~~balloon~~ balloons at 00:00 and 12:00 UTC, WRF winds were compared against the radiosonde measurements at a 12 hour interval during the 24 hour simulation period. The results are shown in Figure 7: RMSE of wind speed is 2.54, 4.0, and 5.11 m s^{-1}, at 2 June 00:00 UTC, 2 June 12:00 UTC, and 3 June 00:00 UTC, respectively. Wind direction difference between WRF and radiosonde is 11.5°, 16.4°, and 19.1° at the three times. Locations of the weather stations and radiosonde sites used in the evaluations can be found in the supplement document.

The ~~above-described~~ above described evaluations using in-situ measurements indicate that the meteorological simulation is of adequate accuracy for the pseudo ~~observation~~ observations based inverse modeling tests conducted in this paper. ~~Future applications with true observational data will need to quantify the simulation errorin the~~ When the 4DVar ~~system.~~ system is applied with real observations, the error and bias must be considered. In WRF-4DVar's cost function configuration, the observation error matrix $\mathbf{R}$ is a combination of three error sources: measurement error, aggregation error, and transport model error. The uncertainty of the CO_2 measurements is about 0.05%, while the transport and aggregation errors are typically an order of magnitude larger (Bruhwiler et al., 2005). For real observation applications, the variance and covariance in $\mathbf{R}$ need to represent the transport error. Furthermore, Fig. 6 shows that WRF simulated 10m wind speed is biased high, which is likely to result in bias in the simulated atmospheric CO_2 mixing ratio. Because Bayes inversion framework assumes unbiased observation error, it may be imperative to correct the error for inversions. One approach is to nudge the meteorology fields toward the observations. For instance, Gupta et al. (1997) found that nudging the model simulated winds in the boundary layer to the radar wind profile observations substantially improved estimates of plume dispersion. An alternative approach is to use a combined 4DVar inversion of meteorology and CO_2 fluxes. For instance, Bocquet et al. (2015) discussed data assimilation using coupled chemistry meteorology models (CCMM). If the CO_2 impact on meteorology is not considered, the current implementation of WRF-CO2 4DVar can be extended to a joint meteorology and CO_2 assimilation system. Since the adjoint code for meteorology has been developed and tested in WRFPLUS and WRFDA (Zhang et al., 2013; Barker et al., 2012), the major modification would be in the optimizaiton schemes where the combined state vector of meteorology and CO_2 is

optimized. It should be noted that in such a joint assimilation framework, optimization of meteorology is an initial condition problem, whereas the CO_2 flux optimization is a boundary condition problem (bottom and lateral boundaries).

3.2 Accuracy of tangent linear and adjoint sensitivities

Next, the accuracy of the newly developed tangent linear and adjoint models ~~whereby~~ was evaluated by comparing their sensitivity calculations against finite difference sensitivity calculated by the forward model. Grid cells involved in the sensitivity calculations 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 in 60 receptor cells.

A tangent linear model run for a grid cell will calculate the tangent linear sensitivity $\partial \mathbf{q_{co2}}/\partial k_{co2}$, which approximates a column vector of the forward model's Jacobian matrix and quantifies the influence of the cell's ~~emission~~ flux 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 $\partial q_{co2}/\partial \mathbf{k_{co2}}$, 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.

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 $\mathbf{g\_q_{co2}}$ are the tangent linear sensitivities $\partial \mathbf{q_{co2}}/\partial k_{co2}$. 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 $\mathbf{a\_k_{co2}}$ 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 of a receptor, such as shown in Fig~~4.~~ 4 of Gerbig et al. (2008).

The tangent linear sensitivity is first compared against the finite difference sensitivity. After confirming the accuracy of the tangent linear model, the adjoint sensitivity is compared against the tangent linear sensitivity.

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

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

The magnitude of Δx used in Eq. (8) is determined by comparing the result from a range of different values. The finite sensitivities were calculated at the 35 sites using Δx set to 0.01, 0.1, and 1.0, and the results show that the magnitude of all differences is less than 10^{-10} (results not shown) because WRF-CO~~2~~ 2 is largely linear. For all subsequent calculations, $\Delta x = 0.1$ is used for Eq. (8).

Sinceboth Since 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 8 shows the comparison between the finite difference and tangent linear sensitivities at 9 of the 35 source cells. The dark straight lines in the figures are the 1:1 line. The maximum and minimum of the difference between finite difference and tangent linear sensitivities are given for each source cell. Results at the rest of the sources are similar (not shown). All differences are less than 10^{-10}, confirming that the tangent linear model is accurate.

The adjoint model is next evaluated 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 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. 9. The minimum and maximum value of the difference between tangent linear and adjoint sensitivities in all three groups are no greater than 10^{-6}, which is comparable to the accuracy tests from other adjoint model developments (Meirink 2008; Henze 2007), indicating that the adjoint model is accuratehas been correctly implemented.

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 flux scaling factor of all surface cells. It is similar to the receptor footprint typically calculated using LPDM, such as Fig. 4 of Gerbig et al. (2008) and Fig. 1 of Alden et al. (2016). But $q_{co2}/\mathbf{k}_{co2}$ differs from footprint in that the former contains the combined impact of tracer transport and emission flux magnitude, while the latter is determined by tracer transport alone. The spatial patterns of the adjoint sensitivity was were examined to discern the impacts of tracer transport. Figure 10 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 the receptor placed at the 1st and 10th vertical levels are is 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 agrees with the fact that low level wind during the simulation period is predominantly southernlysoutherly, 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 of Fig. 10 shows that the highest magnitude of $q_{co2}/\mathbf{k}_{co2}$ is closest to the tower itself, indicating a large impact from local fluxfluxes. In comparison, when the receptor is placed at the 10th vertical level, the peak magnitude of its adjoint sensitivity is a much farther distance southward, and it features much wider spread, indicating transport of flux distant from that receptor is dominant. Results from WLEF located further south of the tower site. Results from the WLEF site 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 similar pattern of impacts of

transport and receptor placement height (not shown).

To provide a comparative view of the source-receptor relations, backward trajectories of particles released from the Centerville and WLEF sites were also calculated using the Lagrangian model HYSPLIT (Stein et al., 2015). WRF-CO2 forward model simulated meteorology saved at 1-hour intervals ~~is~~ was used to drive the HYSPLIT trajectory calculations. ~~Two~~ To compare with the adjoint model result, two sets of simulations were carried out for each of the two tower sites~~:~~. For each simulation, 30,000 particles were released from the ~~height approximates the first vertical level in WRF in the first set, and from the tenth level in the second. Each set of simulation consists of 300 particles, and their starting locations were uniformly distributed within a 48x48 km grid boxcorresponding with the WRF grid spacing~~location of the corresponding WRF grid box used in the adjoint sensitivity calculations. The starting locations of the particles were randomly distributed within the grid box. The resulting backward trajectories were combined with the biospheric CO_2 flux to calculate the footprint for the receptor locations. The HYSPLIT ~~backward trajectory simulation results wre plotted in~~ footprints were calculated on the same grid as used in the WRF-CO2 simulations to facilitate the comparisons between the two models.

Figure 11 shows the HYSPLIT calculated footprints for the Centerville and WLEF sites at the two different vertical levels. The four figures in Fig. 11 are the HYSPLIT counterparts of the adjoint sensitivity figures in Fig. 10. A comparison between Fig. 10 and Fig 11. ~~To avoid cluttering the figures, the trajectories are plotted using only 60 out the 300 particles released in each case. The backward trajectories demonstrate similar spatial patterns as the adjoint sensitivities~~ shows that the results from HYSPLIT and the WRF-CO2 adjoint model compare well spatially. For instance, for the receptor placed at the first vertical level at Centerville, Iowa (Fig. 10~~). On the other hand, there are discernible differences between the two, which can be attributed to that representation of some tracer transport processes in HYSPLIT that are different from WRF, especially diffusion and convective transport.~~ (a) and Fig. 11(a)), the footprints from both models are primarily located in Missouri and Northwestern Arkansas. Based on the horizontal wind fields at the first level, these areas were upwind of the receptor location during the simulation period. Overall, the WRF-CO2 adjoint sensitivities contain larger surface areas compared to their HYSPLIT footprint counterparts. This difference is likely caused by the more diffusive nature of tracer transport in WRF-CO2: its finite difference scheme for tracer advection contains numerical diffusion, and it also includes an explicit horizontal diffusion term in the tracer transport (Skamarock et al., 2008). A further comparison at individual grid points reveals magnitude differences between the footprints from HYSPLIT and the WRF-CO2 adjoint model. This is mainly caused by the different treatments of turbulent vertical mixing by the two models. In WRF-CO2, the PBL and convective schemes parameterize tracer vertical mixing (see Section 2.4.3). For vertical mixing, HYSPLIT either uses the PBL heights calculated by WRF or it calculates PBL heights independently by analyzing temperature profiles. The footprints shown in Fig. 11 were simulated by HYSPLIT using PBL heights from the WRF-CO2 ACM2 PBL scheme. In a separate set of HYSPLIT simulations with PBL heights calculated from the temperature profiles, only minor differences are observed in the resulting footprints (not shown).

3.4 ~~Inverse modeling test~~

3.4 Inverse modeling tests

~~After confirming the validity of the~~

3.4.1 Inverse modeling setup

The sensitivity tests in section 3.2 have confirmed that the tangent linear and adjoint models ~~, the effectiveness~~ of WRF-CO2 ~~4DVAR in inverse modeling experiments was tested. Pseudo-observation data generated by the forward model run were used in these inverse modeling experiments, which start with prescribed prior values for the emission scaling factors and seek to recover their true values. To generate pseudo-observation data, the forward model ran for 24 hourswith emission scaling factor set to unity at all surface grid points, saving CO_2 mixing ratio (q_{co2}) every 4~~ 4DVar are correctly implemented. In this section, inverse modeling tests are conducted to confirm that the two optimization schemes described in Section 2.2 and 2.3 are also correctly implemented. The inverse modeling tests here are designed following the approach used in Henze et al. (2007). To confirm that the GEOS-Chem 4DVar code was correctly developed, Henze et al. (2007) set $\mathbf{B^{-1}} = \mathbf{0}$ and $\mathbf{R} = \mathbf{I}$ (the identity matrix) and constrained the optimizations with error-free pseudo-observations. Because $\mathbf{B^{-1}} = \mathbf{0}$, analysis deviations from the first guess cause no increase in the cost function (see Eq. (3)). This means that if the 4DVar code is correctly implemented, the optimization will converge to the true solution used to generate the pseudo-observations. Such a configuration of $\mathbf{B}$ and $\mathbf{R}$, although highly ideal and unrealistic for real applications, is an effective way to test the code accuracy in isolation from external errors. If the code is correctly implemented, the optimization will converge to the true solution used to generate the pseudo-observations. Because the background error is set to infinity ($\mathbf{B^{-1}} = \mathbf{0}$), the optimization should converge to the true solution with any first guess. A different first guess will impact the process of the convergence, but not the result: the optimization should eventually converge to the true solution.

Following Henze et al. (2007), inverse modeling tests here involve the following steps:

1. Run the WRF-CO2 forward model for 24 hours, using the daily mean biospheric CO_2 flux (Fig. 4) as the true biospheric CO_2 fluxes.

2. Generate pseudo-observations by saving the model simulated atmospheric CO_2 at all grid points of the bottom 30 vertical levels every 4 hours. This ~~generated~~ creates a set of ~~six~~ 6 pseudo-observation files, ~~each of which include the instantaneous q_{co2} at the model's 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.~~

   ~~Inverse modeling experiments were conducted for two cases of prior k_{co2}. In the first case, the prior emission scaling factor overestimates the true values by 50%~~ (which contain no error with respect to the true biospheric CO_2 flux used in Step 1.

3. Generate a set of first guess biospheric CO_2 fluxes.

4. Set the background error to infinity ($\mathbf{B^{-1}} = \mathbf{0}$) and the observation error to the identity matrix ($\mathbf{R} = \mathbf{R^{-1}} = \mathbf{I}$).

5. Run the L-BFGS-B and incremental optimizations with the first guess biospheric CO_2 flux (Step 3), constrained by the pseudo-observations (Step 2), until the optimized biospheric flux converges to the true biopheric CO_2 flux (Step 1).

Steps 3-5 repeat twice for two different sets of first guess biospheric CO_2 fluxes:

- Case 1: set flux scaling factor $k_{co2} = 1.5$ at all ~~cells). In the second case, the prior emission scaling factor is~~ surface grid point.
- Case 2: set flux scaling factor k_{co2} randomly distributed between ~~0.5 and 1.5.~~ 0.5 and 1.5.

Figure 12 shows the two ~~experiment cases as scatter plots between the true biosphere CO_2 and its background value (the prior).~~

~~Both L-BFGS-B and incremental optimization (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 identity matrix). This configuration is equivalent of (~~sets of first guess biospheric CO_2 as compared with the true biospheric CO_2 fluxes. Each point in the figures represents a surface grid point. It should be noted that because the Case 1 ~~) the setting total cost function to the observation cost function, and (2) setting the gradient to the observation gradient~~first guess overestimates the true fluxes by 50% at all surface grid points, the background error are perfectly correlated, implying that all off-diagonal elements in $\mathbf{B}$ should be set to unity. However, since the inverse modeling tests are designed to be driven solely by the pseudo-observations (by setting $\mathbf{B^{-1}} = \mathbf{0}$), the detailed content of $\mathbf{B}$ becomes irrelevant. 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~~ also be noted that the same set of pseudo-observations ~~.~~

~~Because the pseudo-observation data~~ (Step 1) are used for both of the two cases of first guesses, and the pseudo-observations were not perturbed with errors, it is appropriate to set $\mathbf{R} = \mathbf{I}$ for both cases. This simply assigns all the observations equal weight in calculating the observation cost function using Eq. (4). In these inverse modeling tests, because the pseudo-observations are of q_{co2} at the forward model's grid points, the mapping between model space and observation space is trivial: the observation operator (H), tangent linear observation operator ($\widetilde{H}$), 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.~~ ($\widetilde{H}^T$) are all simply the indentity matrix. For application with real observations, however, each type of CO_2 observations will need its own set of H, $\widetilde{H}$, and $\widetilde{H}^T$ to map between the model space and observation space.

A very simple error configuration ($\mathbf{B^{-1}} = \mathbf{0}$, and $\mathbf{R} = \mathbf{I}$) was used in the inverse modeling tests here, but such a setting is only appropriate to confirm code accuracy using error-free observations. For real data applications, an appropriate specification of background ($\mathbf{B}$) and observation error ($\mathbf{R}$) is a critical and challenging task. Ideally the variance and covariance in $\mathbf{B}$ should be specified based on comparisons between prior fluxes and accurate flux measurements (Chevallier et al., 2006; Gerbig et al., 2006). But available flux measurements are often of insufficient amount, thus necessitating assumptions regarding the form of the

background error matrix. For instance, prior flux errors were treated as uncorrelated in Gockede et al. (2010), and Rodenbeck et al. (2003) u
an exponential decaying spatial correlation for the prior flux error. In another study, Peylin et al. (2005) found significantly different flux estimation resulted from varying the prior flux error correlation scale from 500 km to 2000 km. For the observation error covariance matrix $\mathbf{R}$, the spatial and temporal error correlation were often negelected in earlier inversion studies (Gurney et al., 2002; I With more recent inversion studies using continuous observation at towers (Law et al., 2008), airborne observations (Lauvaux et al., 2008), and satellite observations (Chevallier et al., 2005), attempts have been made to represent the spatial and temporal correlation of observation errors. For instance, Kountouris et al. (2015) found the temporal autocorrelation time for observation data using the VPRM model is around 30 days.

3.4.2 Inverse modeling results

The results from inverse modeling experiments with Case 1 prior are shown in Fig. 13 and 14. Figure ~~10~~ 13 shows the iterative reduction of the cost function $J(\mathbf{x})$, gradient norm $\|\nabla J(\mathbf{x})\|$, and RMSE. The iteration number for Lanczos-CG is all from its inner loop, and only one outer loop is used. The figures show both L-BFGS-B and Lanczos-CG 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 after. In gradient norm reduction, both schemes feature periodic oscillations embedded in the large scale downward trend. By comparison, Lanczos-CG has a smaller magnitude oscillation and steeper downward trend than L-BFGS-B. It should be noted that while L-BFGS-B 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. ~~10~~ 13 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~~actual inversion applications. Figure 13(c) shows that both optimization schemes reduce RMSE of daily biosphere flux monotonically, and Lanczos-CG achieves better reduction after about the first 10 iterations. Figure ~~12~~ 14 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 ~~solution~~solution.

The results of inverse modeling experiments using Case 2 prior are shown in ~~Fig~~Figs. 15 and 16. The reductions of $J(\mathbf{x})$, $\|\nabla J(\mathbf{x})\|$, and RMSE are similar to Case 1 in that Lanczos-CG substantially outperforms L-BFGS-G after about the first 10 iterations. Table 5 summarizes the results from 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~~inversion with real observations.

4 Tracer lateral boundary condition

The lateral tracer boundary condition is necessary to connect regional tracer simulations to the global background tracer distribution (Gerbig et al., 2003). A number of regional inversion studies have explored the sensitivity of the estimated posterior flux to the lateral boundary condition. For instance, Schuh et al. (2010) found a 30% magnitude difference in the retrieved North America biospheric flux when boundary conditions from two different global models were used (CarbonTracker and PCTM). In an inversion study over the state of Oregon, Gockede et al. (2010) found the estimated biospheric CO_2 fluxes were highly sensitive to systematic changes in the advected background CO_2 through the lateral boundaries. To address the lateral boundary uncertainty, Lauvaux et al. (2008) used LPDM backward trajectories to calculate the atmospheric CO_2 sensitivity to the lateral boundary conditions, and optimized lateral boundary conditions along with surface fluxes in a synthesis inversion approach. An alternative is to use part of the observations to correct the lateral boundary error before the inversion, which then only includes surface fluxes in its state vector (Lauvaux et al., 2012). In the pseudo-observation based inverse modeling tests described in Section 3 of this work, CO_2 lateral boundary conditions do not contain error, and they were not included in the state vector for optimization. When WRF-CO2 4DVar is applied with real observations, uncertainties of lateral boundary conditions need to be appropriately treated. To use either approach used in Lauvaux et al. (2008) or in Lauvaux et al. (2012), the adjoint code for tracer lateral boundary conditions would need to be developed for the WRF-CO2 4DVar system.

In the WRF-Chem dynamical core, chemistry mixing ratios are updated at each time step by the advection and diffusion tendencies. Then chemistry mixing ratios at the lateral boundaries are updated with the chemistry boundary condition using the flow dependent method, which uses the horizontal wind direction to determine whether the chemistry mixing ratio at a boundary grid point should be updated by the lateral boundry. If the horizontal wind direction indicates tracer inflow at a boundary grid, Eq. (9) will be applied to the grid point.

$$q_{co2} = q_b + q_{b,t}\Delta t \tag{9}$$

Where q_{co2} represents CO_2 mixing ratio at a lateral boundary grid, q_b and $q_{b,t}$ are the CO_2 mixing ratio and tendency at the correponding lateral boundary.

To develop the lateral boundary related tangent linear and adjoint code, Eq. (9), is replaced by Eq. (10) in WRF-CO2 4DVar.

$$q_{co2} = k_{co2}(q_b + q_{b,t}\Delta t) \tag{10}$$

Where k_{co2} represents the CO_2 lateral boundary scaling factor. Please note that in Eqs. (9) and (10), the time dependence has been dropped for the sake of simplicity. The corresponding tangent linear and adjoint of Eq. (10) are given in Eqs. (11) and (12).

$$g\_q_{co2} = g\_k_{co2}(q_b + q_{b,t}\Delta t) \tag{11}$$

$$a\_k_{co2} = a\_k_{co2} + a\_q_{co2}(q_b + q_{b,t}\Delta t) \tag{12}$$

Where g_q_{co2} and a_q_{co2} are the tangent linear and adjoint variable of q_{co2}, and g_k_{co2} and a_k_{co2} are the tangent linear and adjoint variables of k_{co2}.

Most code development for tracer lateral boundary conditions are in the input_chem_data module of the chemistry directory, along with some additional code modification to enable the lateral boundary condition variables to be passed forward (k_{co2} and g_k_{co2}) and backward (a_k_{co2}) in time. The two optimization schemes of WRFCO2-4DVar have also been implemented to allow for flexibilities in state vector specification. The user can choose to include lateral boundary conditions in the state vector to be optimized, which is a similar approach as in Lauvaux et al. (2008) (but using a 4DVar optimization). Alternatively, the user can choose to correct the lateral boundary (using the adjoint model) before the inversion, and not to include lateral boundary in the state vector (Lauvaux et al., 2012).

When applied with real observations, whether and how to aggregate lateral boundary scaling factors is not trivial (Lauvaux et al., 2008, 20 On one hand, including lateral boundary scaling factors without spatial aggregation will greatly increase the state vector size, likely causing the inversion to be under-constrained. On the other hand, aggregating lateral boundary scaling factors may cause aggregation error (Kaminski et al., 2001). While the actual treatment of lateral boundary scaling factor aggregation is beyond the scope of this work, a mapping mechanism has been implemented in WRF-CO2 4DVar to facilitate the aggregation. In WRFCO2-4DVar, q_{co2}, g_q_{co2} and a_q_{co2} are defined on the model grid, but k_{co2}, g_k_{co2}, and a_k_{co2} are defined as 1d variables in the state vector. The mapping mechanism implemented in procedure da_cv_to_wrf and its adjoint counterpart allows for many-to-one mappings from the 3d grid variables to the 1d state vector. This mapping mechanism allow the user flexibility in determining whether and how to aggregate the lateral boundary condition.

5 Summary and outlook

WRF-CO2 4DVar was developed as a data assimilation system designed to constrain surface CO_2 ~~flux~~ fluxes by combining an online atmospheric chemistry transport model and observation data in a Bayesian framework. Two optimization schemes were implemented 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~~models, all developed on top of the WRFPLUS system. While WRFPLUS's forward model is WRF, WRF-Chem was used as WRF-CO2 4DVar's forward model to include CO_2 in the system, and the tangent linear and adjoint models were modified to keep their consistency with the forward model. Like most other CO_2 inverse modeling systems, WRF-4DVar ignores the possible impacts of atmospheric CO_2 variation on the meteorology. This simplification enables the use of the same full physical parameterizations in the forward, tangent linear, and adjoint ~~model~~models. This configuration reduces linearization error while allowing the WRF system's large number of physical parameterizations to be used in WRF-CO2 4DVar without requiring a large amount of new code development.

WRF-CO2 4DVar's tangent linear and adjoint models were tested by comparing their sensitivities' spatial patterns with the dominant wind patterns. The results make physical sense given the meteorological transport. The accuracy of tangent linear and adjoint models ~~were~~ was evaluated 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. Finally, the system was tested in inverse modeling with ~~pseudo-observation data~~pseudo-observations, and the results show that both optimization schemes successfully recovered the true values with reasonable accuracy and computation cost.

While Lanczos-CG performs better than L-BFGS-B in the inverse modeling tests, it must be pointed out that the tests are very limited. Although a comprehensive comparison between the two optimization schemes is beyond the scope of the present paper, it is important 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 calls the forward model. For a tracer transport system like WRF-CO2 4DVar, the tangent linear model can skip some of the costly physics parameterizations, such as the radiation scheme. This difference means that 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 processor core), it takes about 10 minutes walltime to complete one inner loop of Lanczos-CG. L-BFGS-B takes about 10% more walltime to complete one iteration.

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, this calculation is carried out on only the root core and broadcasted to the other process cores. 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 processor core. 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 core. 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 processor core, where m is the maximal inner loop iteration allowed. Although it is possible to reduce the per processor core a memory allocation from $m \times n$ to n by disactivating the modified Gram-Schmidt orthonormalization step, it is typically not recommened.

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, which is designed to run a longer simulation than the typical 6-hour run intended for WRFDA. GH15/17 implemented a second-order checkpoint mechanism to overcome the memory limit. This approach breaks the whole simulation period into sections, saves restart files at the end of each section by the forward model. This approach requires extra calls of the forward model to recalculate the trajectory for each section during backward integration (See Fig. 3 of GH15)

To overcome the memory limit posed by a long simulation. In WRF-CO2 4DVar, a different approach was implemented . In WRF-CO2 4DVar, to overcome this memory limit: the forward model saves the trajectory at each time step in memory, as WRFPLUS does. After a number of integration steps, the memory on each task processor core is dumped to an external file, and the memory is then reused. Each external file is marked with its starting timestamp and the processor core 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 processor core is only needed for 30 steps instead of 720 steps. This will results in 24 (720/30) trajectory files on each task processor core, and the total number of trajectory files depends on the number of processor core cores 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 processor core, rather than being shared among all processor cores. This means all model runs in an inverse experiment must use the same domain patch configuration, which is the most common practice.

In future development, we plan to implement observation operators for real observations, including those from towers, satellites, and airborne instruments. This is required for applying WRF-CO2 4DVar with real observation dataobservations. As a regional inverse system, the correct treatment of chemistry tracer lateral boundary conditions is important. We plan to include chemistry initial and boundary conditions in the state vector in the next updatetest the lateral boundary condition adjoint code (Section 4) in a follow-up study. 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 tests in the present paper.

In addition, we also plan to periodically update the WRF-CO2 4DVar system to keep up with WRF system updates. Such updates will mainly consist of replacing the forward model with the updated WRF code, and developing the tangent linear and adjoint code for the relevant updated procedures. As the variable dependence analysis (Section 2.4.2.1) indicates that the tangent linear and adjoint code are only needed for a portion of WRF procedures, the amount of work required for updating WRF-CO2 4DVar is manageable. In addition, future development of WRF-CO2 4DVar will also be dependent on updates to WRFPLUS, which has always been updated along with WRF.

6 Code availability

WRF-CO2 4DVar source code can be retrieved via https://doi.org/10.5281/zenodo.839260 .1184200

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 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 three

reviewers greatly improved both the model and the paper. This work was partially supported by a Central Michigan University CST research incentive fund.

References

Alden, C. B., Miller, J. B., Gatti, L. V., Gloor, M. M., Guan, K., Michalak, A. M., van der Laan-Luijkx, I. T., Touma, D., Andrews, A., Basso, L. S., Correia, C. S. C., Domingues, L. G., Joiner, J., Krol, M. C., Lyapustin, A. I., Peters, W., Shiga, Y. P., Thoning, K., van der Velde, I. R., van Leeuwen, T. T., Yadav, V., and Diffenbaugh, N. S.: Regional atmospheric CO2 inversion reveals seasonal and geographic differences in Amazon net biome exchange, Global Change Biol., 22, 3427–3443, 2016.

Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O. T., Hogrefe, C., Luecken, D. J., Bash, J. O., Roselle, S. J., Pleim, J. E., Foroutan, H., Hutzell, W. T., Pouliot, G. A., Sarwar, G., Fahey, K. M., Gantt, B., Gilliam, R. C., Heath, N. K., Kang, D., Mathur, R., Schwede, D. B., Spero, T. L., Wong, D. C., and Young, J. O.: Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1, Geosci. Model Dev., 10, 1703–1732, 2017.

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO2, Tellus B, 58, 359–365, 2006.

Baker, D. F., Boesch, H., Doney, S. C., O'Brien, D., and Schimel, D. S.: Carbon source/sink information provided by column CO2 measurements from the Orbiting Carbon Observatory, Atmos. Chem. Phys., 10, 4145–4165, 2010.

Barker, D., Huang, X.-Y., Liu, Z., Auligne, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X.: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System WRFDA, Bull. Amer. Meteor. Soc., 93, 831–843, 2012.

Basu, S., Houweling, S., Peters, W., Sweeney, C., Machida, T., Maksyutov, S., Patra, P. K., Saito, R., Chevallier, F., Niwa, Y., Matsueda, H., and Sawa, Y.: The seasonal cycle amplitude of total column CO2: Factors behind the model-observation mismatch, J. Geophys. Res, 116, d23306, 2011.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO2 fluxes estimated from GOSAT retrievals of total column CO2, Atmos. Chem. Phys., 13, 8695–8717, 2013.

Bocquet, M.: Toward Optimal Choices of Control Space Representation for Geophysical Data Assimilation, Mon. Weather Rev., 137, 2331–2348, 2009.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Zabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Perez Camano, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmos. Chem. Phys., 15, 5325–5358, 2015.

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P.: Inverse modeling of annual atmospheric CO2 sources and sinks 1. Method and control inversion, J. Geophys. Res, 104, 26 161–26 178, 1999.

Bruhwiler, L., Michalak, A., Peters, W., Baker, D., and Tans, P.: An improved Kalman Smoother for atmospheric inversions, Atmos. Chem. Phys., 5, 2691–2702, 2005.

Butler, M. P., Davis, K. J., Denning, A. S., and Kawa, S. R.: Using continental observations in global atmospheric inversions of CO2: North American carbon sources and sinks, Tellus B, 62, 550–572, 2010.

Byrd, R. H., Lu, P., and Nocedal, J.: A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Stat. Comp., 16, 1190–1208, 1995.

Chan, E., Chan, D., Ishizawa, M., Vogel, F., Brioude, J., Delcloo, A., Wu, Y., and Jin, B.: Description and evaluation of REFIST v1.0: a regional greenhouse gas flux inversion system in Canada, Geosci. Model Dev. Discussion, 2016.

Chevallier, F.: Impact of correlated observation errors on inverted CO2 surface fluxes from OCO measurements, Geophys. Res. Lett., 34, 2007.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Breon, F. M., Chedin, A., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res., 110, d24309, 2005.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Froehlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgui, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, d21307, 2010.

Chevallier, F., Viovy, N., Reichstein, M., and Ciais, P.: On the assignment of prior errors in Bayesian inversions of CO2 surface fluxes, Geophys. Res. Lett., 33, 2006.

Chou, M. D. and Suarez, M.: A solar radiation parameterization for atmospheric studies, Tech. Rep. NASA/TM-1999-10460, vol. 15, 38 pp, NASA, 1999.

Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregon, A., Rayner, P. J., Miller, C., Gobron, N., Kinderman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R. J., Balsamo, G., Bopp, L., Breon, F. M., Broquet, G., Dargaville, R., Battin, T. J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J. G., Cook, R. B., DeFries, R., Engelen, R., Gurney, K. R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luyssaert, S., Miller, J., Moriyama, T., Moulin, C., Myneni, R. B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J. D., Piao, S. L., Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., van der Werf, G., Wickland, D., Williams, M., and Zehner, C.: Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system, Biogeosciences, 11, 3547–3602, 2014.

Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO2 from GOSAT XCO2 data, Atmos. Chem. Phys., 14, 3703–3727, 2014.

Enting, I. G., Trudinger, C. M., and Francey, R. J.: A Synthesis Inversion of the Concentration and Delta-C-13 of Atmospheric Co2, Tellus B, 47, 35–52, 1995.

Fowler, A. M. and Lawless, A. S.: An Idealized Study of Coupled Atmosphere–Ocean 4D-Var in the presence of model error, Mon. Weather Rev., 144, 4007–4029, 2016.

Freitas, S. R., Longo, K. M., Alonso, M. F., Pirre, M., Marecal, V., Grell, G., Stockler, R., Mello, R. F., and Sanchez Gacita, M.: PREP-CHEM-SRC-1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models, Geosci. Model Dev., 4, 419–433, 2010.

French, N. H. F., de Groot, W. J., Jenkins, L. K., Rogers, B. M., Alvarado, E., Amiro, B., de Jong, B., Goetz, S., Hoy, E., Hyer, E., Keane, R., Law, B. E., McKenzie, D., McNulty, S. G., Ottmar, R., Perez-Salicrup, D. R., Randerson, J., Robertson, K. M., and Turetsky, M.: Model comparisons for estimating carbon emissions from North American wildland fire, J. Geophys. Res., 116, 2011.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res, 108, 4756, 2003.

Gerbig, C., Lin, J. C., Munger, J. W., and Wofsy, S. C.: What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, Atmos. Chem. Phys., 6, 539–554, 2006.

Gerbig, C., Korner, S., and Lin, J. C.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, Atmos. Chem. Phys., 8, 591–602, 2008.

Gerbig, C., Dolman, A. J., and Heimann, M.: On observational and modelling strategies targeted at regional carbon exchange over continents, Biogeosciences, 6, 1949–1959, 2009.

Gockede, M., Turner, D. P., Michalak, A. M., Vickers, D., and Law, B. E.: Sensitivity of a subregional scale atmospheric inverse CO2 modeling framework to boundary conditions, J. Geophys. Res., 115, 2010.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, 2012.

Grell, G. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, 2002.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmos. Chem. Phys., 14, 5233–5250, 2014.

Grell, G. A., Knoche, R., Peckham, S. E., and McKeen, S. A.: Online versus offline air quality modeling on cloud-resolving scales, Geophys. Res. Lett., 31, l116117, 2004.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled online chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, 2005.

Guerrette, J. J. and Henze, D. K.: Development and application of the WRFPLUS-Chem online chemistry adjoint and WRFDA-Chem assimilation system, Geosci. Model Dev., 8, 1857–1876, 2015.

Guerrette, J. J. and Henze, D. K.: Four dimensional variation of black carbon emissions during ARACTAS-CARB with WRFDA-Chem, Atmos. Chem. Phys., 17, 7605–7633, 2017.

Gupta, S., McNider, R., Trainer, M., Zamora, R., Knupp, K., and Singh, M.: Nocturnal wind structure and plume growth rates due to inertial oscillations, J. Appl. Meteorol. Climatol., 36, 1050–1063, 1997.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002.

Gurney, K. R., Chen, Y. H., Maki, T., Kawa, S. R., Andrews, A., and Zhu, Z. X.: Sensitivity of atmospheric CO2 inversions to seasonal and interannual variations in fossil fuel emissions, J. Geophys. Res., 110, d10308, 2005.

Hascoet, L. and Pascual, V.: The Tapenade Automatic Differentiation Tool: Principles, Model, and Specification, ACM Trans. Math. Software, 39, 20, 2013.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413–2433, 2007.

Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J. L., Fairhead, L., Filiberti, M. A., Friedlingstein, P., Grandpeix, J. Y., Krinner, G., Levan, P., Li, Z. X., and Lott, F.: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, Climate Dyn., 27, 787–813, 2006.

Houweling, S., Aben, I., Breon, F. M., Chevallier, F., Deutscher, N., Engelen, R., Gerbig, C., Griffith, D., Hungershoefer, K., Macatangay, R., Marshall, J., Notholt, J., Peters, W., and Serrar, S.: The importance of transport model uncertainties for the estimation of CO2 sources and sinks using satellite measurements, Atmos. Chem. Phys., 10, 9981–9992, 2010.

Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model, J. Appl. Meteorol. Climatol., 49, 1831–1844, 2010.

Huang, X.-Y., Xiao, Q., Barker, D. M., Zhang, X., Michalakes, J., Huang, W., Henderson, T., Bray, J., Chen, Y., Ma, Z., Dudhia, J., Guo, Y., Zhang, X., Won, D.-J., Lin, H.-C., and Kuo, Y.-H.: Four-Dimensional Variational Data Assimilation for WRF: Formulation and Preliminary Results, Mon. Weather Rev., 137, 299–314, 2009.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113, 2008.

Jiang, X., Li, Q. B., Liang, M. C., Shia, R. L., Chahine, M. T., Olsen, E. T., Chen, L. L., and Yung, Y. L.: Simulation of upper tropospheric CO2 from chemistry and transport models, Global Biogeochem. Cycles, 22, gB4025, 2008.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res., 106, 4703–4715, 2001.

Kawa, S. R., Erickson, D. J., Pawson, S., and Zhu, Z.: Global CO2 transport simulations using meteorological data from the NASA data assimilation system, J. Geophys. Res., 109, d18312, 2004.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns, J. Geophys. Res., 114, d04305, 2009.

Kountouris, P., Gerbig, C., Totsche, K. U., Dolman, A. J., Meesters, A. G. C. A., Broquet, G., Maignan, F., Gioli, B., Montagnani, L., and Helfter, C.: An objective prior error quantification for regional atmospheric inverse applications, Biogeosciences, 12, 7403–7421, 2015.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmos. Chem. Phys., 5, 417–432, 2005.

Lanczos, C.: An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators, J. Res. Nat. Bur. Stand., 45, 255–282, 1950.

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337–354, 2012.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459–3471, 2008.

Law, R. M., Peters, W., Roedenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S. J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Patra, P. K., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Takigawa, M., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO2: Experimental overview and diurnal cycle results for 2002, Global Biogeochem. Cycles, 22, gB3009, 2008.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res., 108, 4493, 2003.

Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., Collatz, G. J., Menemenlis, D., Ott, L., Pawson, S., Jones, D., and Nassar, R.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT X-CO2 sampling on the inference of terrestrial biospheric sources and sinks, Tellus B, 66, 2014.

Luis Morales, J. and Nocedal, J.: Remark on "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization", ACM Trans. Math. Software, 38, 2011.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X. M., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochem. Cycles, 22, gB2005, 2008.

Meirink, J. F., Bergamaschi, P., Frankenberg, C., d'Amelio, M. T. S., Dlugokencky, E. J., Gatti, L. V., Houweling, S., Miller, J. B., Roeckmann, T., Villani, M. G., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modeling of atmospheric methane emissions: Analysis of SCIAMACHY observations, J. Geophys. Res., 113, d17301, 2008.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res., 109, d14109, 2004.

Mlawer, E., Taubman, S., Brown, P., Iacono, M., and Clough, S.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102, 16 663–16 682, 1997.

Nassar, R., Jones, D. B. A., Suntharalingam, P., Chen, J. M., Andres, R. J., Wecht, K. J., Yantosca, R. M., Kulawik, S. S., Bowman, K. W., Worden, J. R., Machida, T., and Matsueda, H.: Modeling global atmospheric CO2 with improved emission inventories and CO2 production from the oxidation of other carbon species, Geosci. Model Dev., 3, 689–716, 2010.

Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W., Andres, R. J., Suntharalingam, P., Chen, J. M., Brenninkmeijer, C. A. M., Schuck, T. J., Conway, T. J., and Worthy, D. E.: Inverse modeling of CO2 sources and sinks using satellite observations of CO2 from TES and surface flask measurements, Atmos. Chem. Phys., 11, 6029–6047, 2011.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and forecasting-stochastic time-inverted lagrangian transport (WRF-STILT) model, Meteorol. Atmos. Phys., 107, 51–64, 2010.

Nolte, C. G., Appel, K. W., Kelly, J. T., Bhave, P. V., Fahey, K. M., Collett, Jr., J. L., Zhang, L., and Young, J. O.: Evaluation of the Community Multiscale Air Quality (CMAQ) model v5.0 against size-resolved measurements of inorganic particle composition across sites in North America, Geosci. Model Dev., 8, 2877–2892, 2015.

Ott, L., Pawson, S., Collatz, G., Gregg, W. W., Menemenlis, D., Brix, H., Rosseaux, C. S., Bowman, K. W., Liu, J., Eldering, A., Gunson, M. R., and Kawa, S. R.: Assessing the magnitude of CO2 flux uncertainty in atmospheric CO2 records using products from NASA's Carbon Monitoring Flux Pilot Project, J. Geophys. Res.: Atmosphere, 120, doi:10.1002/2014JD022 411, 2015.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proc. Natl. Acad. Sci. U.S.A., 104, 18 925–18 930, 2007.

Peters, W., Miller, J., Whitaker, J., Denning, A., Hirsch, A., Krol, M., Zupanski, D., Bruhwiler, L., and Tans, P.: An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res., 110, 2005.

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO2 data, J. Geophys. Res., 107, 4385, 2002.

Peylin, P., Rayner, P., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., and contributors, A.: Daily CO2 flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, Atmos. Chem. Phys., 5, 3173–3186, 2005.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Roedenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO2 inversions, Biogeosciences, 10, 6699–6720, 2013.

Pillai, D., Gerbig, C., Kretschmer, R., Beck, V., Karstens, U., Neininger, B., and Heimann, M.: Comparing Lagrangian and Eulerian models for CO2 transport - a step towards Bayesian inverse modeling using WRF/STILT-VPRM, Atmos. Chem. Phys., 12, 8979–8991, 2012.

Pillai, D., Buchwitz, M., Gerbig, C., Koch, T., Reuter, M., Bovensmann, H., Marshall, J., and Burrows, J. P.: Tracking city CO2 emissions from space using a high-resolution inverse modelling approach: a case study for Berlin, Germany, Atmos. Chem. Phys., 16, 9591–9610, 2016.

Pleim, J. and Chang, J.: A nonlocal closure model for vertical mixing in the convective boundary layer, Atmos. Environ., 26, 965–981, 1992.

Pleim, J. E.: A simple, efficient solution of flux-profile relationships in the atmospheric surface layer, J. Appl. Meteorol. Climatol., 45, 341–347, 2006.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing, J. Appl. Meteorol. Climatol., 46, 1383–1395, 2007.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model, J. Appl. Meteorol. Climatol., 46, 1396–1409, 2007.

Pleim, J. E. and Xiu, A. J.: Development of a land surface model. Part II: Data assimilation, J. Appl. Meteorol. Climatol., 42, 1811–1822, 2003.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J. F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, Quart. J. Roy. Meteor. Soc., 126, 1143–1170, a, 2000.

Rayner, P., Enting, I., Francey, R., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO2, delta C-13 and O-2/N-2 observations, Tellus B, 51, 213–232, 1999.

Rodenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO(2) flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, 2003.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, J. Climate, 27, 2185–2208, 2014.

Saito, R., Houweling, S., Patra, P. K., Belikov, D., Lokupitiya, R., Niwa, Y., Chevallier, F., Saeki, T., and Maksyutov, S.: TransCom satellite intercomparison experiment: Construction of a bias corrected atmospheric CO2 climatology, J. Geophys. Res., 116, d21120, 2011.

Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E., and Worthy, D. E. J.: A regional high-resolution carbon flux inversion of North America for 2004, Biogeosciencs, 7, 1625–1644, 2010.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W., and Powers, J.: A description of the Advanced Research WRF version 3, NCAR Tech Note NCAR/TN-475+STR, 2008.

Smith, A., Lott, N., and Vose, R.: The Integrated Surface Database Recent Developments and Partnerships, Bull. Amer. Meteor. Soc., 92, 704–708, 2011.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA'S HYSPLIT atmospheric transport and dispersion modeling system, Bull. Amer. Meteor. Soc., 96, 2059–2077, 2015.

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey,

R. J., and Denning, A. S.: Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO(2), Science, 316, 1732–1735, 2007.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 2005.

Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: implementation of a new snow parameterization, Mon. Weather Rev., 136, 5095–5115, 2008.

Tremolet, Y.: Diagnostics of linear and incremental approximations in 4D-Var, Quart. J. Roy. Meteor. Soc., 130, 2233–2251, b, 2004.

Turner, A. J. and Jacob, D. J.: Balancing aggregation and smoothing errors in inverse models, Atmos. Chem. Phys., 15, 7039–7048, 2015.

Uliasz, M.: The Atmospheric Mesoscale Dispersion Modeling System, J. Appl. Meteorol., 32, 139–149, 1993.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, J. Geophys. Res.-Atmospheres, 119, 7741–7756, times Cited: 13, 2014.

Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear Inverse Prob.: an example from carbon dioxide flux estimation, Geosci. Model Dev., 6, 583–590, 2013.

Zhang, X., Huang, X.-Y., and Pan, N.: Development of the Upgraded Tangent Linear and Adjoint of the Weather Research and Forecasting (WRF) Model, J. Atmos. Oceanic Technol., 30, 1180–1188, 2013.

Zhu, C., Byrd, R., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Software, 23, 550–560, 1997.

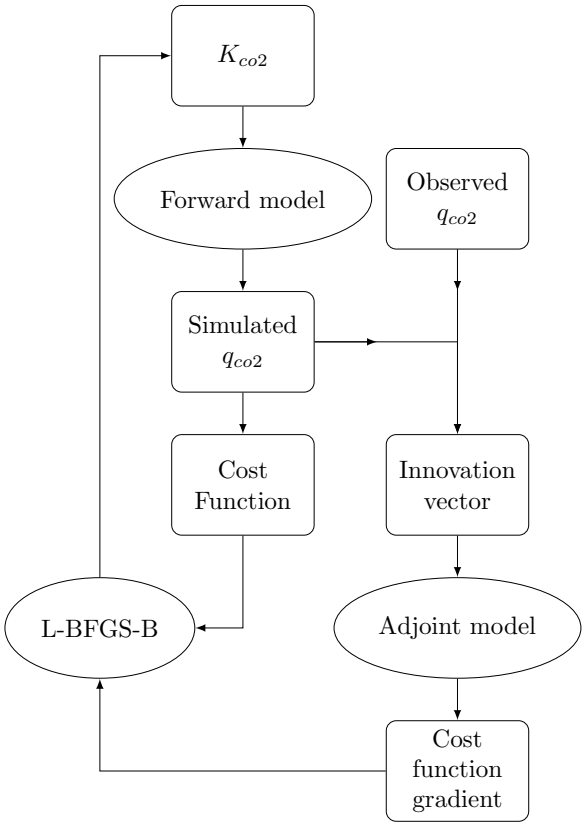

Figure 1. Diagram of L-BFGS-B based optimization implemented for WRF-CO2 4DVar.

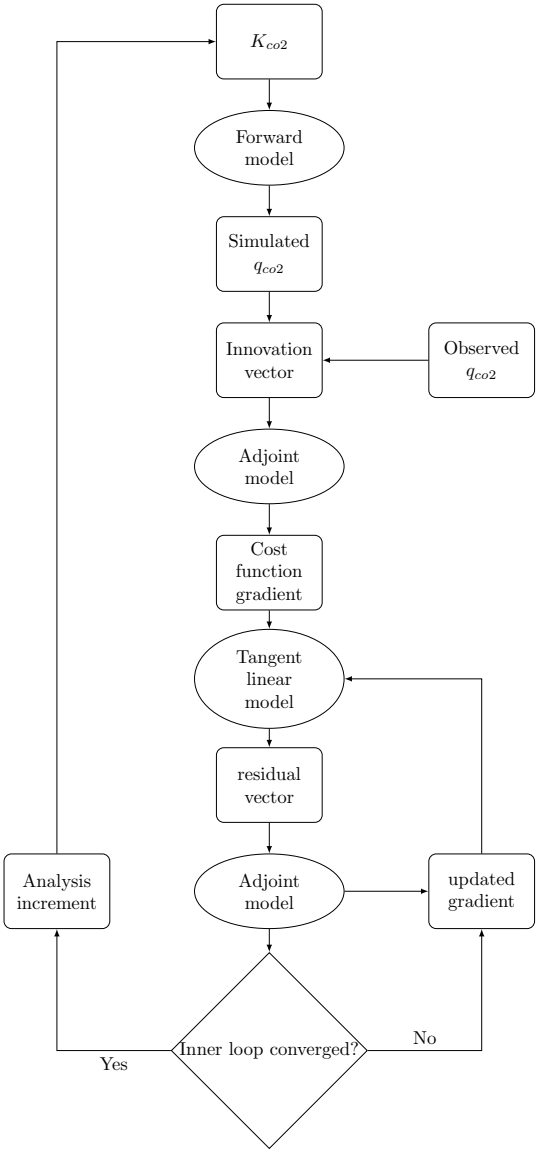

Figure 2. Diagram of Lanczos-CG based incremental optimization implemented for WRF-CO2 4DVar.

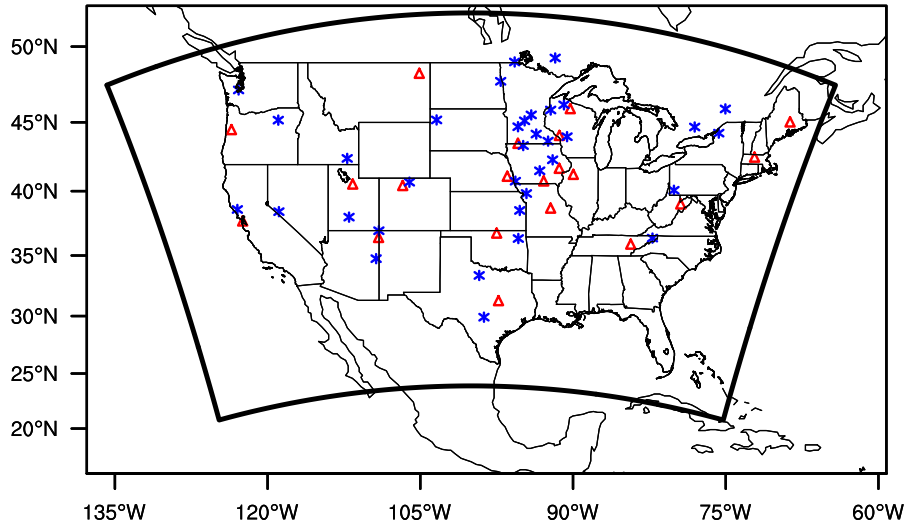

Figure 3. WRF-4DVar simulation domain covering the continental United State with 48 km×48 km grid spacing. The domain boundary is marked by the bold dark outline. Grid cells used for evaluating sensitivities are marked: red triangles are the 20 CO_2 tower sites used as receptor ~~location~~ locations; blue ~~starts~~ stars are source locations. While receptors are placed at the 1st, 5th, and 10th vertical level at each site, all sources are at the 1st level only.

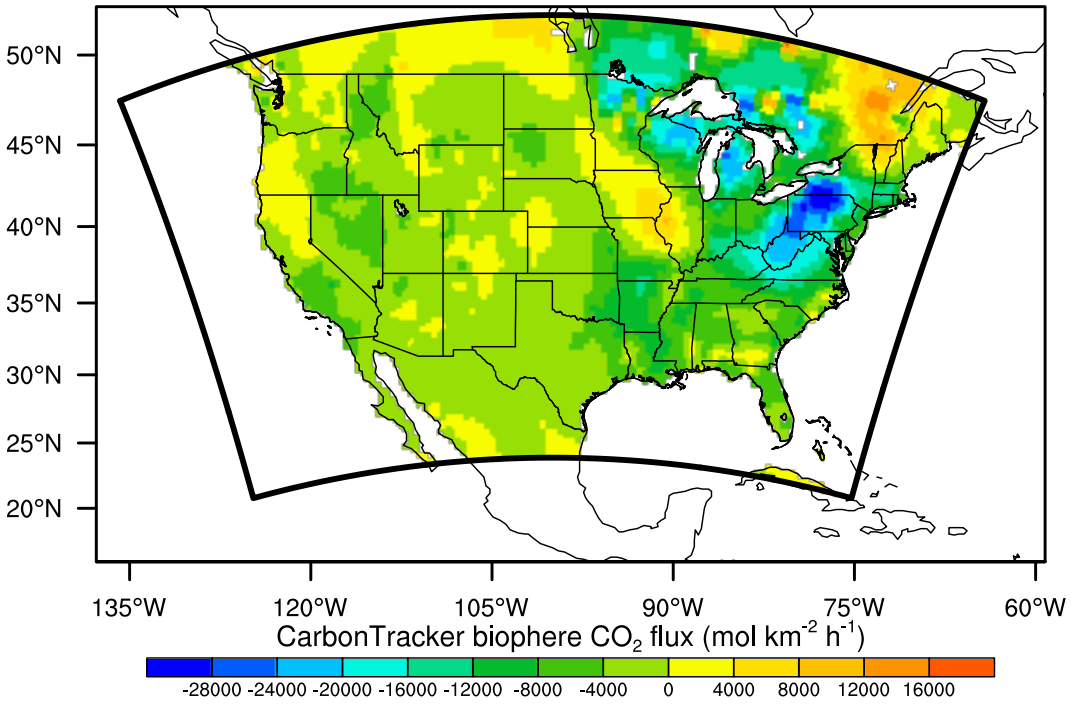

Figure 4. Daily mean CarbonTracker biosphere CO_2 flux, calculated as the arithmetic mean of the 3-hourly flux between 2011-06-02 00:00:00 UTC to 2011-06-03 00:00:00 UTC.

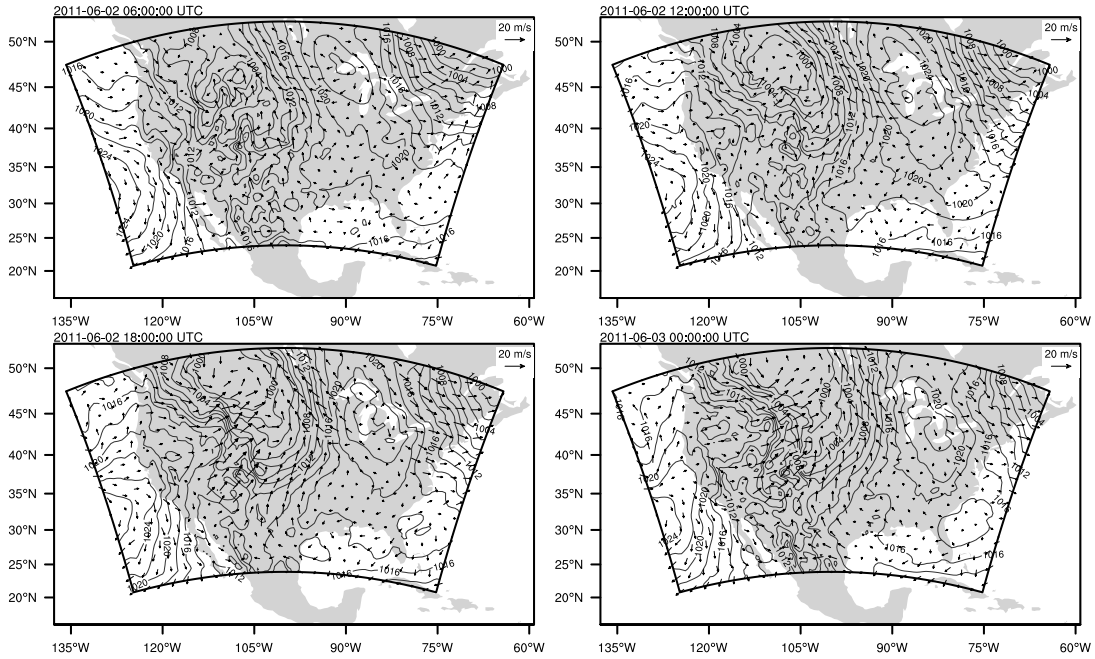

Figure 5. Sea Level Pressure (~~P~~hPa) and horizontal wind (m s^{-1}) at model's lowest vertical level plotted at 6-hour interval during the 24-hour simulation starting at 2011-06-02 00:00 UTC.

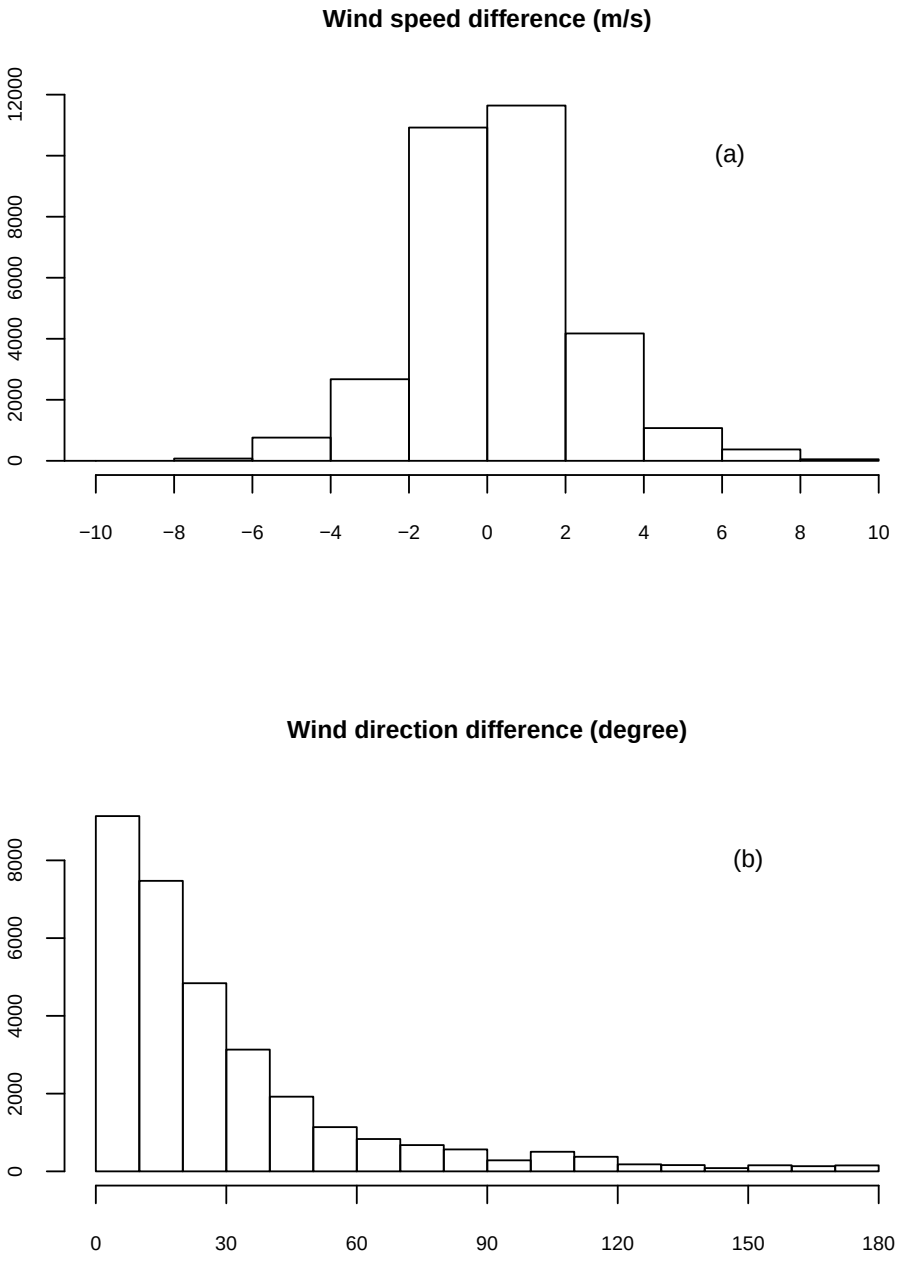

Figure 6. Histograms of the 10m wind speed difference (a) and wind direction difference (b) between WRF simulation and surface meteorological station measurements.

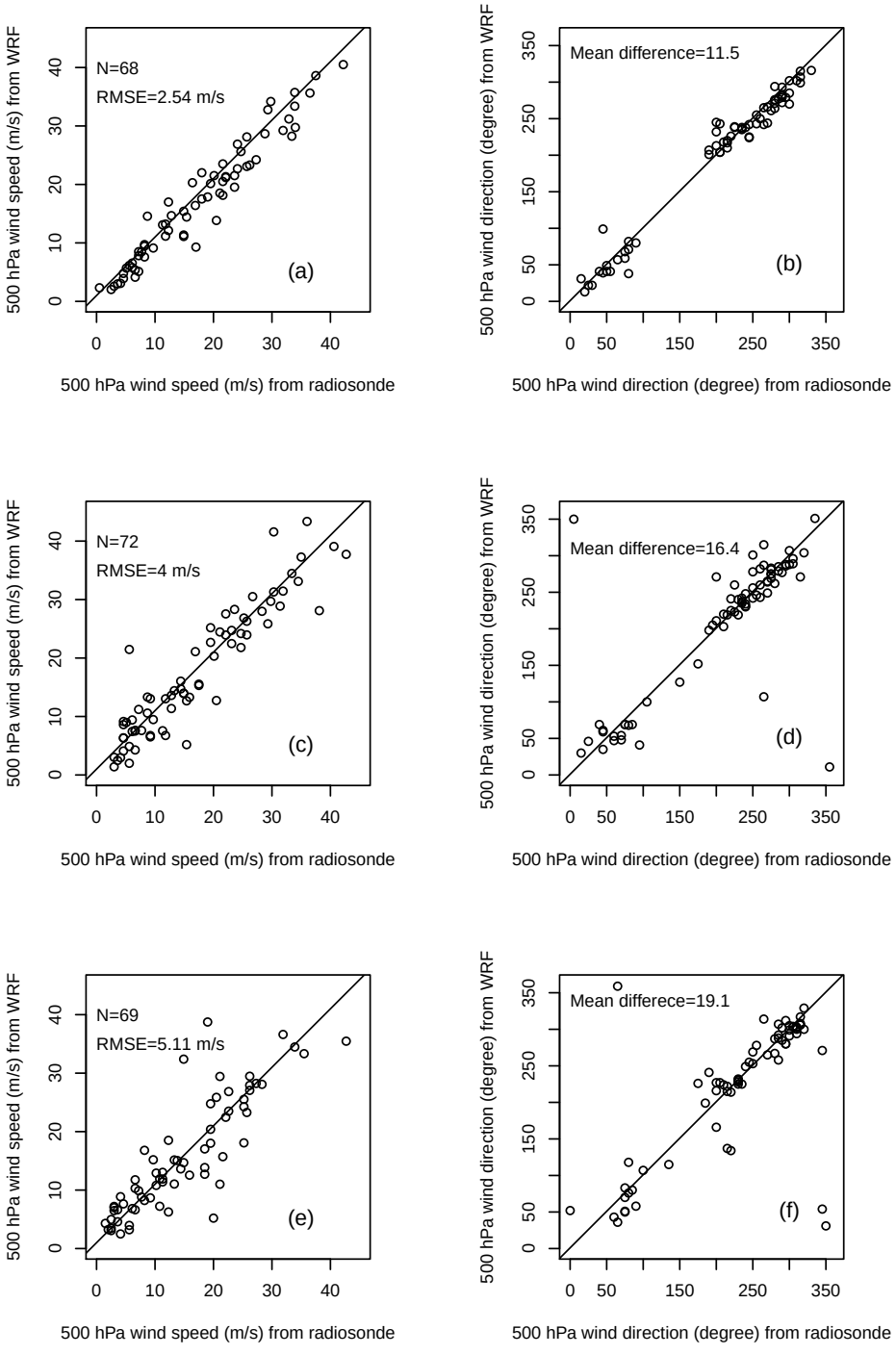

Figure 7. Comparison of ~~500hPa~~ 500 hPa wind speed and wind direction between WRF simulation and radiosonde measurements. Figures (a) ~~anb~~ and (b) are the comparison at 2011-06-02 00:00 UTC; Figures (c) and (d) are at 2011-06-02 12:00 UTC; and Figures (e) and (f) are at 2011-06-03 00:00 UTC. RMSE and relative error (RE) for wind speed and meann difference in wind direction are shown in each figure.

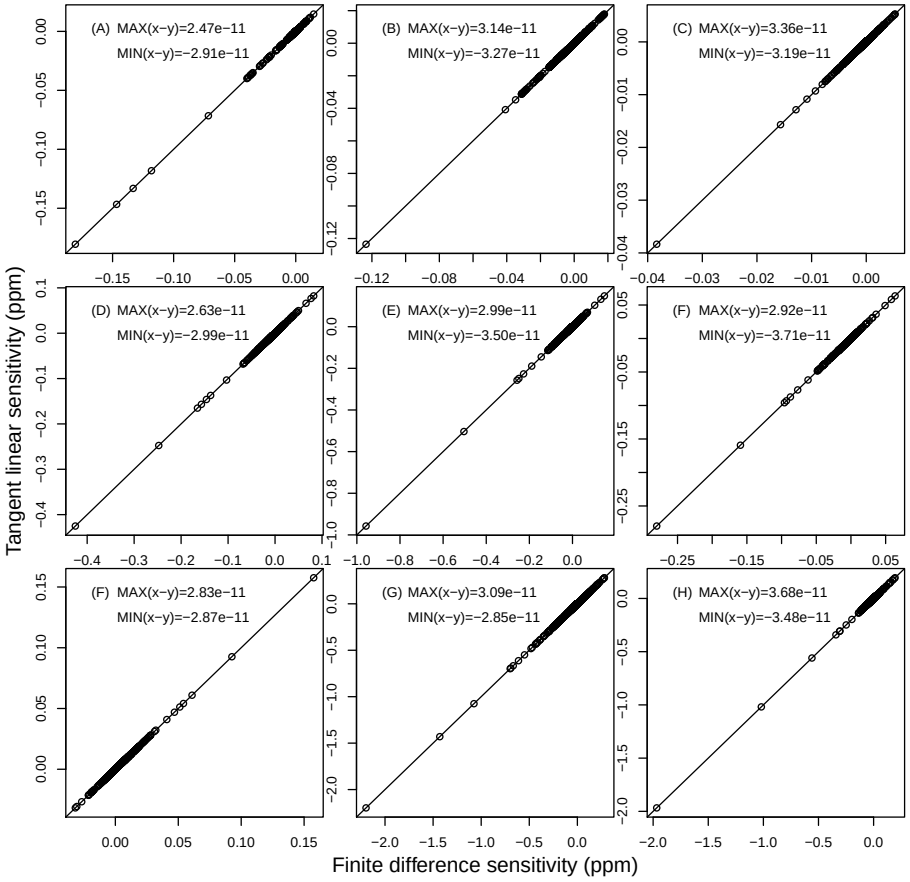

Figure 8. Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by finite difference (x axis) and tangent linear model (y axis) for nine ~~sources~~ source cell locations (see Fig. 3 for source locations).

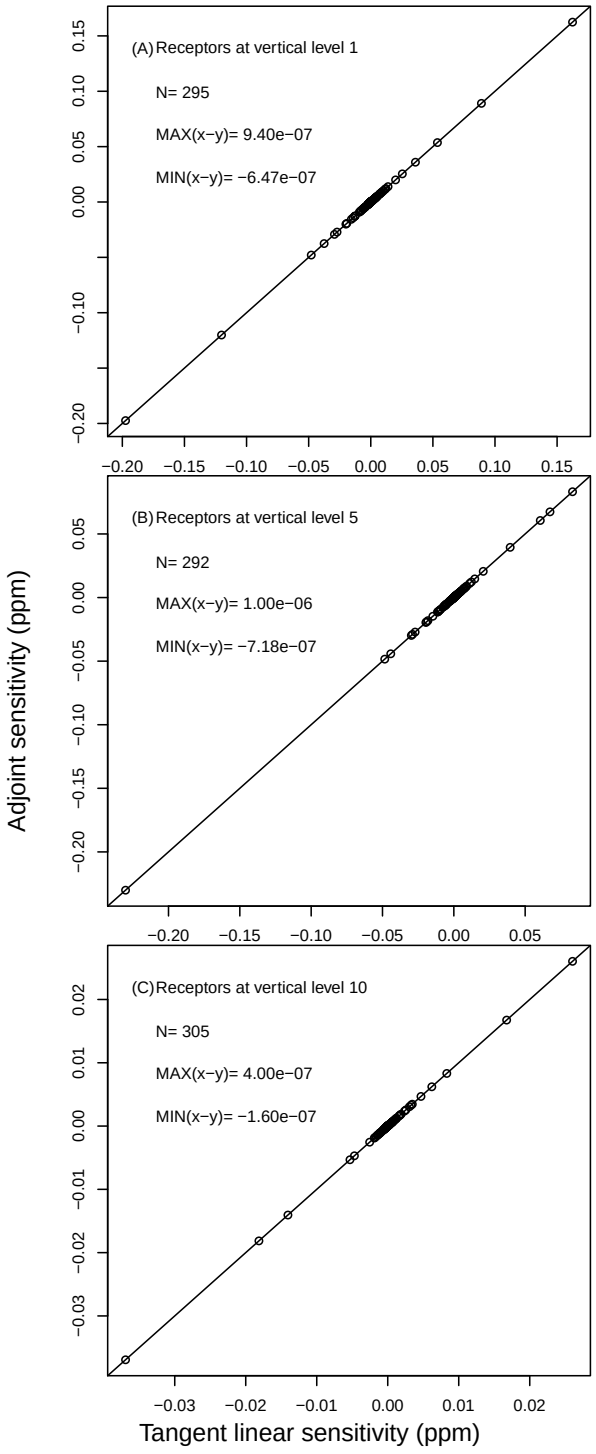

Figure 9. Comparison between $\partial q_{co2}/\partial k_{co2}$ calculated by the tangent linear (x axis) and adjoint model (y axis).

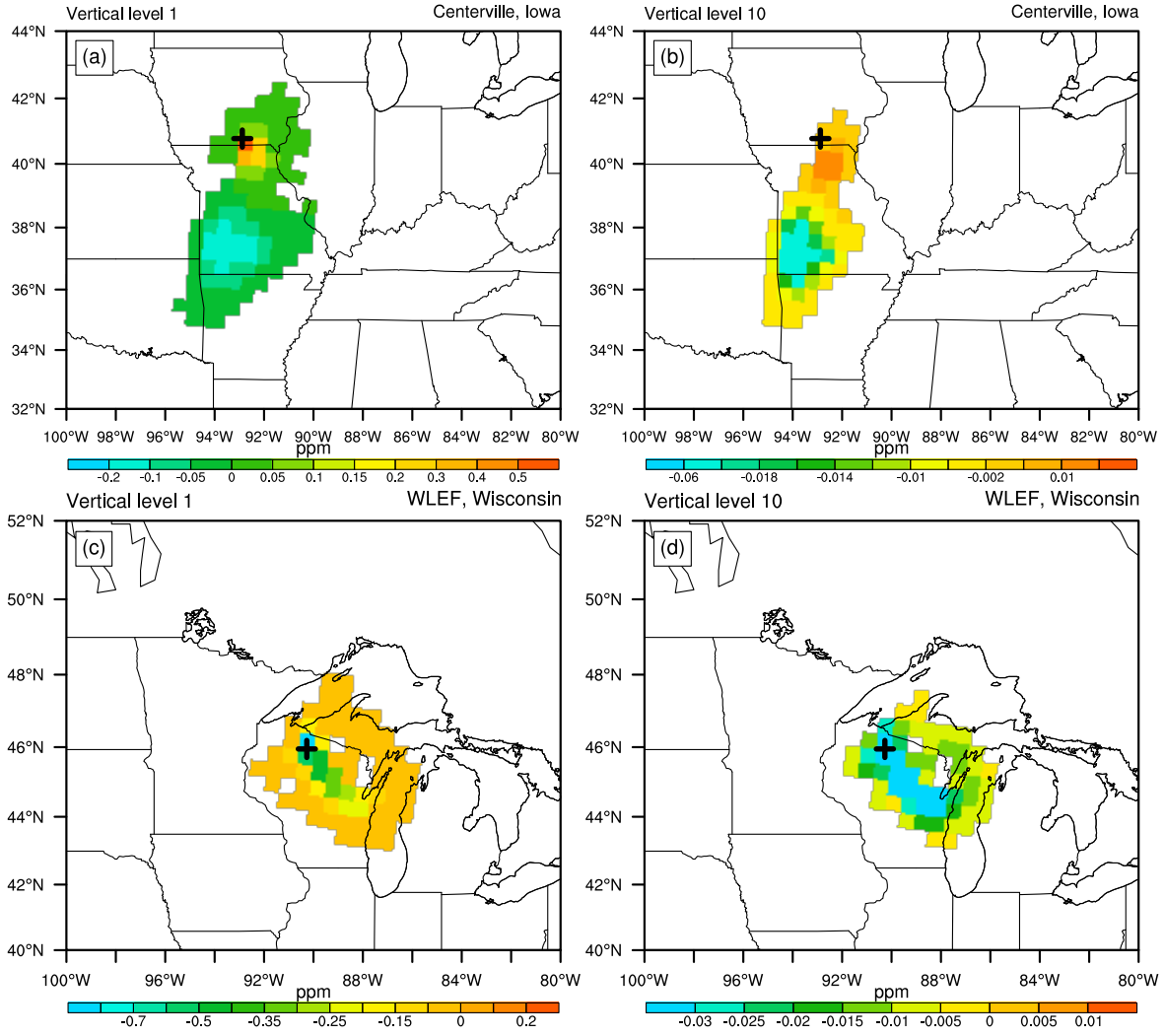

Figure 10. Adjoint sensensitivities calculated by the WRF-CO2 adjoint model. The top panel shows adjoint sensitivity of receptors placed at the 1st (a), and 10th (d) vertical level at Centerville, Iowa. The bottom panel shows adjoint sensitivity of receptors placed at the 1st (c), and 10th (d) vertical level at WLEF, Wisconsin. The black cross in each figure marks the corresponding tower site.

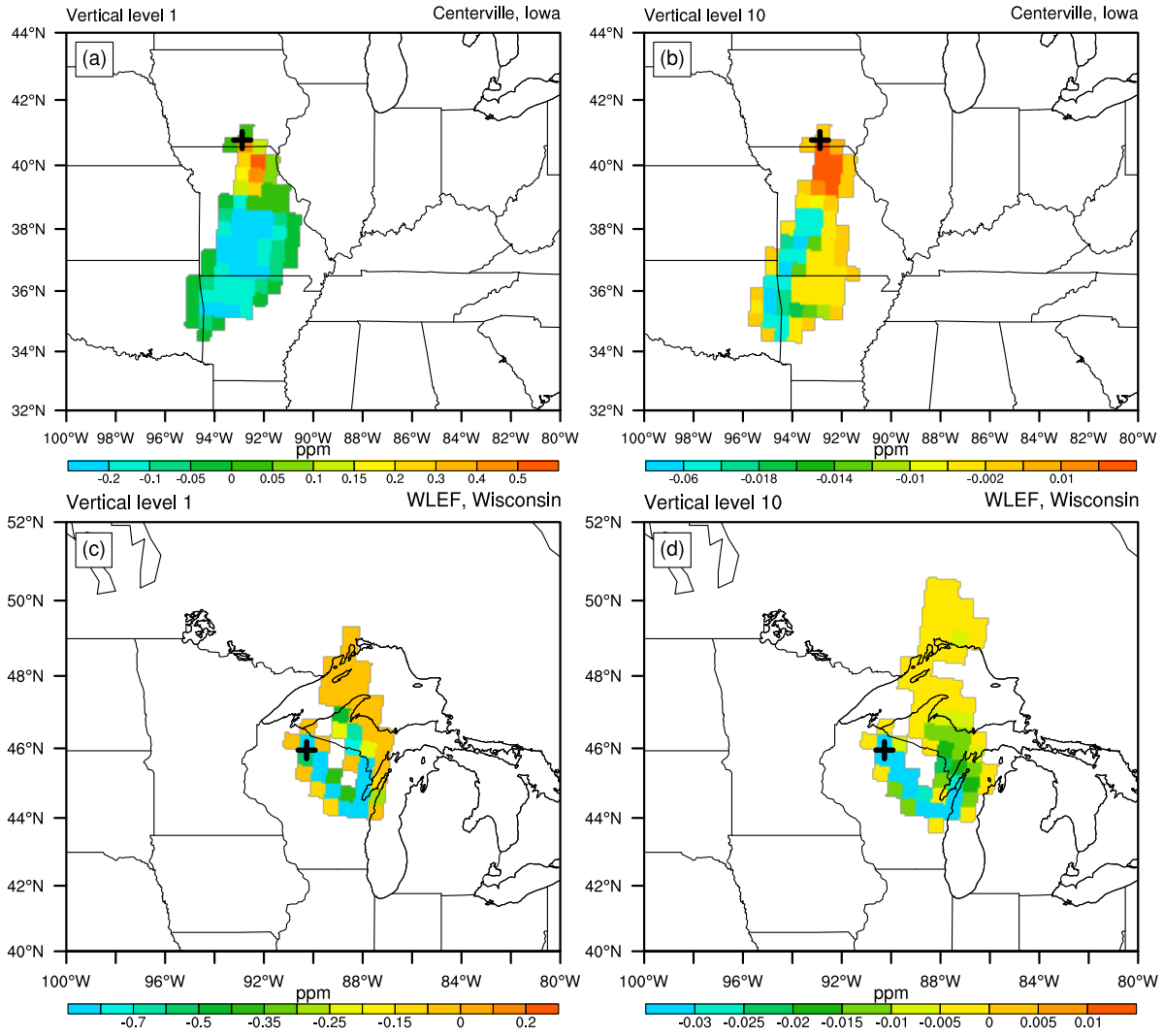

Figure 11. ~~Backward~~ Footprints calculated using HYSPLIT backward trajectories ~~of particles released from~~ and CarbonTracker biospheric fluxes for the tower sites at Centerville, Iowa and WLEF, Wisconsin. The ~~tower sites~~ receptor locations are ~~marked with red crosses. All particles are released at 2011-06-03 00:00 UTC and transported backward~~ the same as in ~~time for 24 hours~~Fig. ~~Figures (a) and (c) are~~ 10. Each HYSPLIT footprint is plotted in the ~~trajectories of particles released from the height match WRF's 1st vertical level, and Figures (b) and (d) are of the 10th level~~same color scale as its counterpart in Fig. 10 for comparision

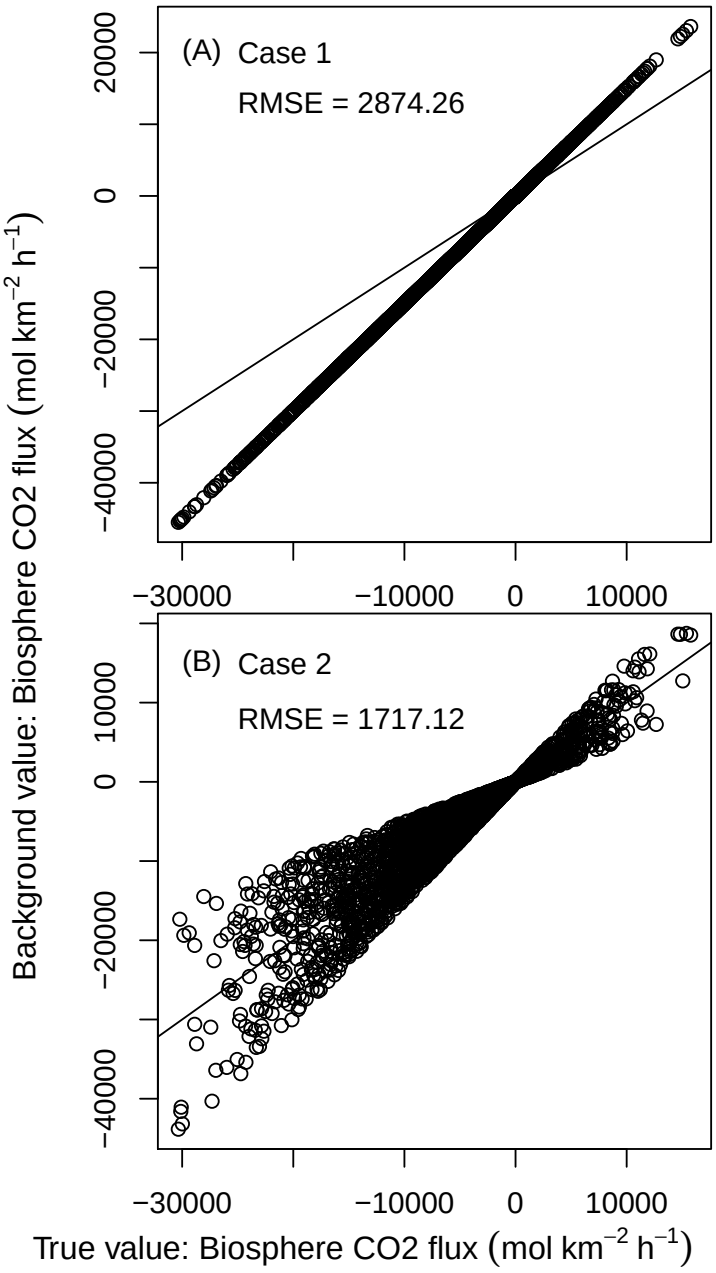

Figure 12. The first guess ~~of~~ biosphere CO_2 fluxes used in the two inverse modeling experiments. The x-axis is true daily mean CarbonTracker biosphere CO_2 value (as shown in Fig 4), and y-axis is the first guess (background value). The solid line in each figure is the 1:1 line.

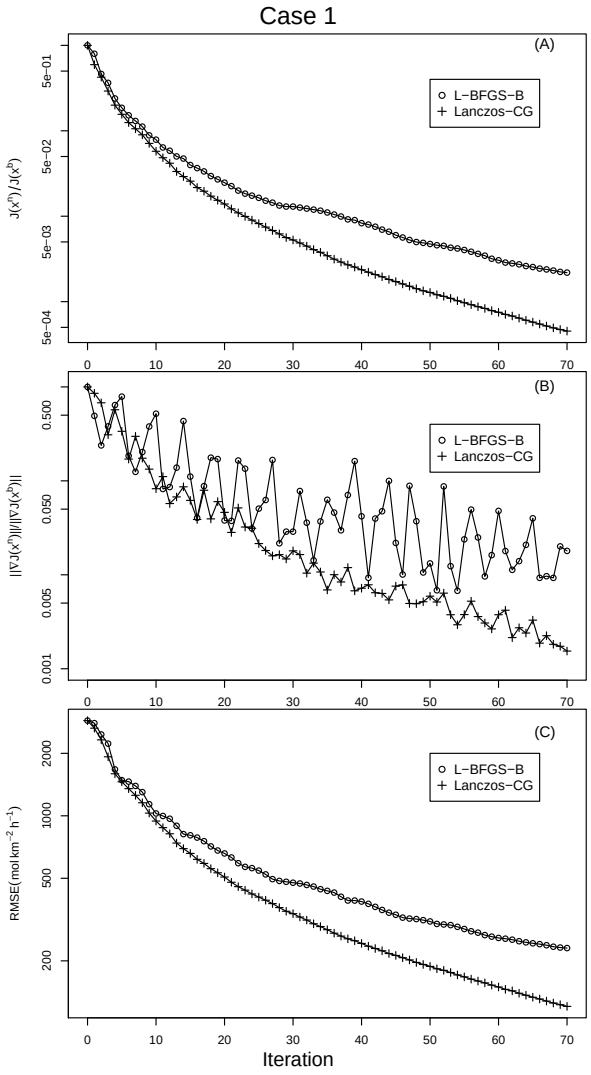

Figure 13. Results of inverse modeling experiment Case 1. Figure (a) shows the reduction of the cost function, represented by $J(x^n)/J(x^b)$. Figure (b) shows the reduction of the gradient norm, represented by $\|\nabla J(x^n)\| / \left\|\nabla J(x^b)\right\|$. Figure (c) shows the reduction of ~~biosphere~~ biospheric CO_2 flux RMSE.

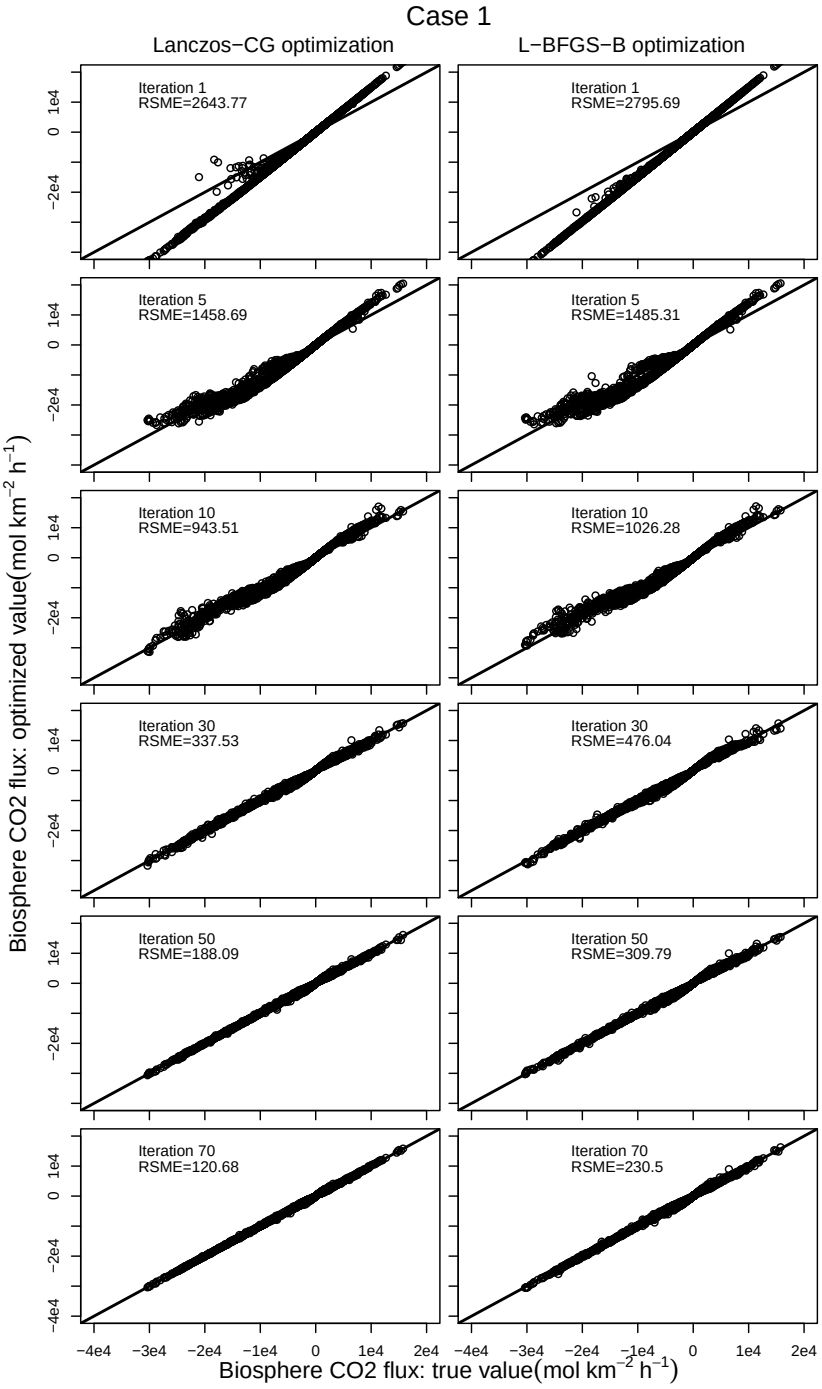

Figure 14. Comparison between the true and optimized CO_2 flux by Lanczos-CG (left column) and L-BFGS-B (right column) in inverse modeling experiment Case 1. The comparison and RMSE after the 1st, 5th, 10th, 30th, 50th, 70th iteration are shown in the figure. All iterations of Lanczos-CG ~~is~~ are from one outer loop.

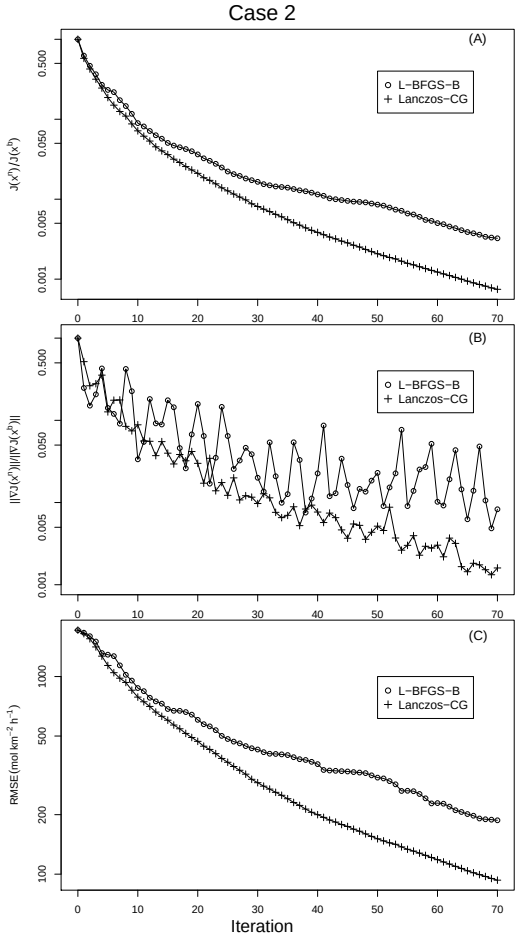

Figure 15. Same as Fig. ~~11~~13, but for inverse modeling experiment Case 2.

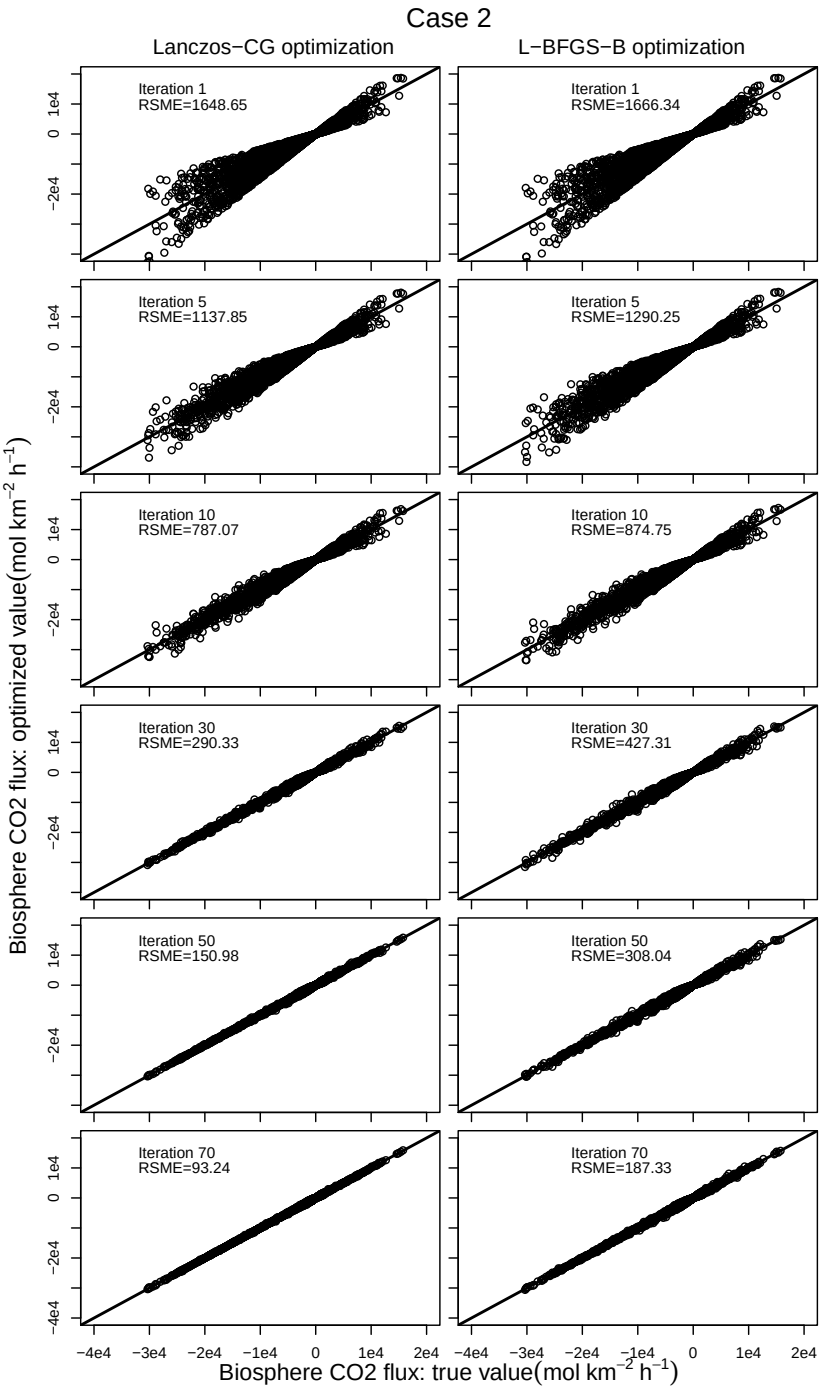

Figure 16. Same as Fig. ~~10~~14, but for inverse modeling experiment Case 2.

Table 1. A list of symbols used in this article

$J(\mathbf{x})$	Cost function
$J_b(\mathbf{x})$	Background cost function
$J_o(\mathbf{x})$	Observation cost function
$\nabla J(\mathbf{x})$	Cost function gradient
$\|\nabla J(\mathbf{x})\|$	Cost function gradient norm
$\nabla^2 J(\mathbf{x})$	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
$\mathbf{k}_{co2}$	CO_2 emission scaling factor
$\mathbf{q}_{co2}$	CO_2 mixing ratio (dry air)
$g\_\mathbf{k}_{co2}$	Tangent linear variable for CO_2 emission scaling factor
$a\_\mathbf{k}_{co2}$	Adjoint variable for CO_2 emission scaling factor
$g\_\mathbf{q}_{co2}$	Tangent linear variable for CO_2 mixing ratio (dry air)
$a\_\mathbf{q}_{co2}$	Adjoint variable for CO_2 mixing ratio (dry air)
$\mathbf{x}^b$	Prior estimate of CO_2 emission scaling factor
$\mathbf{x}x^n$	Analysis of CO_2 emission scaling factor
$\hat{\mathbf{x}}$	Analysis increment of CO_2 emission scaling factor
$\mathbf{y}_k$	Observation at the k^{th} assimilation window
$\mathbf{d}_k$	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 model forward sweep	Adjoint model backward sweep
Chemistry	X	X	X	X
Photolysis	X	X	X	X
Dry deposition	X	X	X	X
Wet deposition	X	X	X	X
Radiation	F	F	F	X
Surface	F	F	F	X
Cumulus	F	F	F	X
Microphysics	F	F	F	X
Advection	F	Add	F	Add
Diffusion	F	Add	F	Add
Emission	F	Dev	F	Dev
PBL	F	Dev	F	Dev
Convective transport	F	Dev	F	Dev

Table 3. WRF-CO2 4DVar model configuration and 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 CO_2 flux	CarbonTracker 2016
ocean CO_2 flux	CarbonTracker 2016
fire CO_2 flux	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 $J(x)$, gradient norm $\|\nabla J(x)\|$, and RMSE are given as the ratio to their respective starting values. Results of the two experiment cases are the values after 70 iterations.

Case 1

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

Case 2

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