Anonymous Referee #1
General comments:

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

Regarding the rational for developing on-line chemistry transport based 4DVar system without considering the CO_2 feedback to meteorology.

***First**, a major benefit of using an online chemistry transport model is that it provides meteorology fields at much finer grid spacing and time interval. For high resolution regional inverse modelings, CO_2 vertical transport is primarily driven by resolved vertical wind velocity, instead of parameterized physical schemes. Offline models, driven by archived analysis or meteorology model output with much larger time interval, have difficult simulate chemistry transport fine grid spacing.*

***Second**, in WRF system, CO_2 can impact meteorology fields through the radiation schemes (longwave/shortwave). As a regional CO_2 inverse modeling system, WRF-CO2 4DVar is designed to run in short period time (hours to weeks) for constrain emission flux with observation data. For such time span are insignificant in most cases. For the applications WRF-CO2 4DVar is designed for, including CO2 impacts on meteorology will require a large amount of code development while offering limited performance improvement. This can be done in the future, but it is beyond the scope of the present paper.*

Regarding the future WRF system update:

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

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

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

At the present stage, WRF-CO2 4DVar system does not include any observational operator (and their TL/AD counterparts), thus it is not ready for applying real observational data yet. The objectives of the present paper are to (1) develop and test the accuracy of the tangent linear and adjoint models, and (2) to implement the two iterative optimization schemes and test their effectiveness with synthetic data *(pseudo observation). In such pseudo observation based tests, both the observed and simulated CO2 are generated by the same meteorology but different CO_2 flux through scaling factor).This setup ensures that meteorology is error-free, and no transport error is present in the inverse system.*

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

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

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

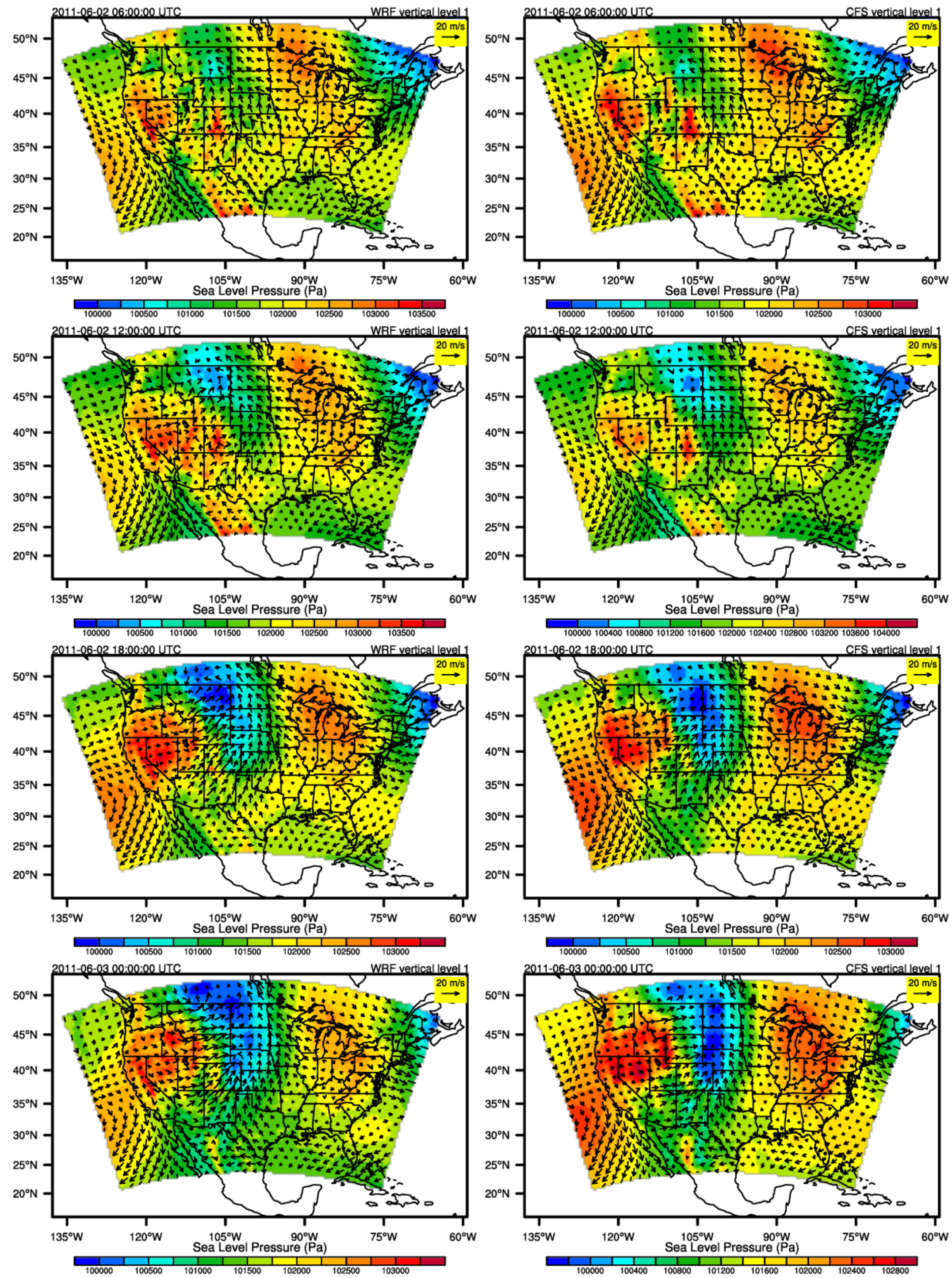
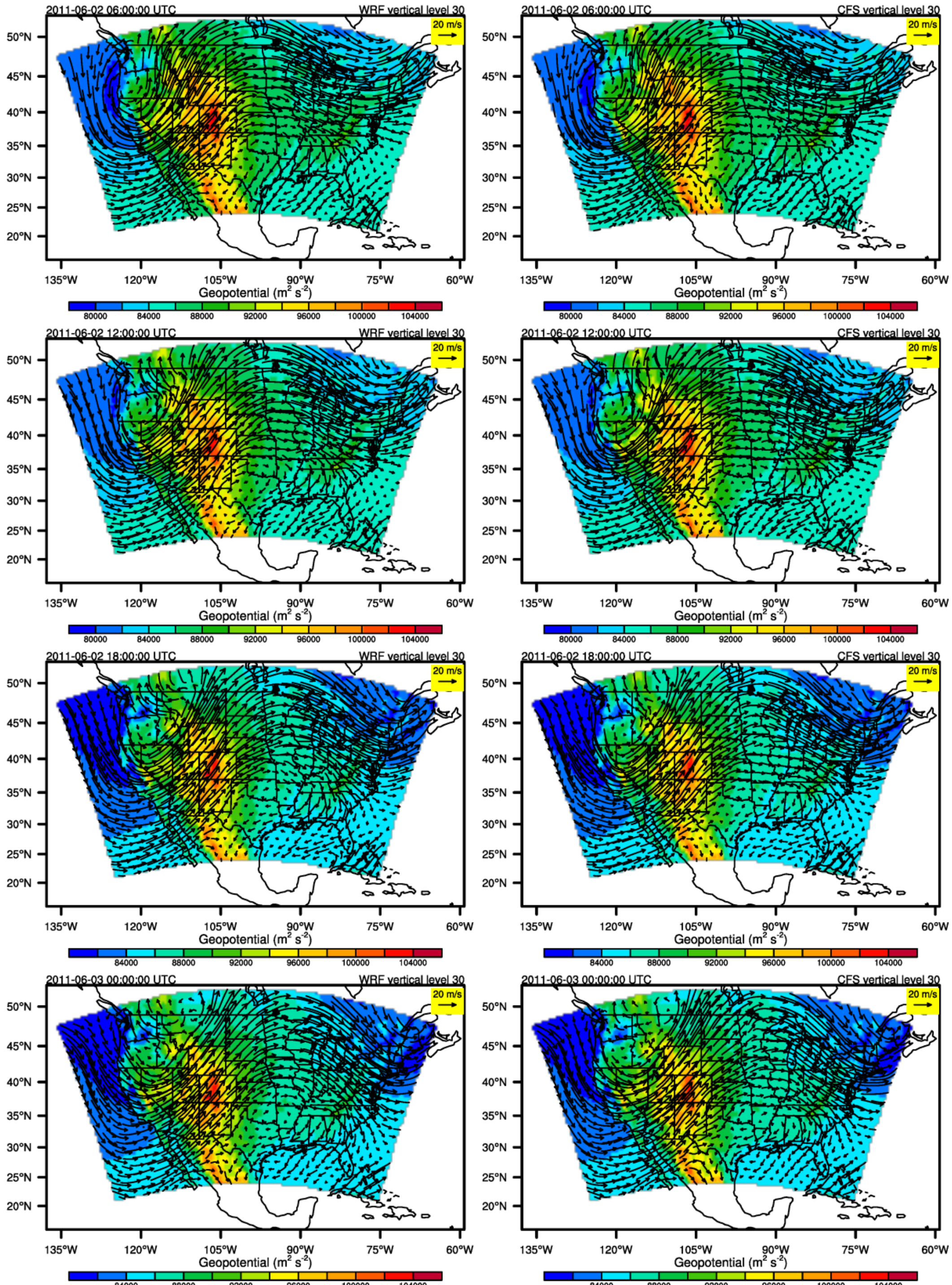


Figure 1. Sea level pressure (background) and horizontal wind at the first vertical level (arrows) as simulated by WRF (left column) and interpolated from CFSv2 (right columns). The four figures are plotted at 6-hour interval for the 24 hour simulation period (2011-06-02 00:00 UTC to 2011-06-03 00:00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



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

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

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

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

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

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

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

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

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

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

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

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

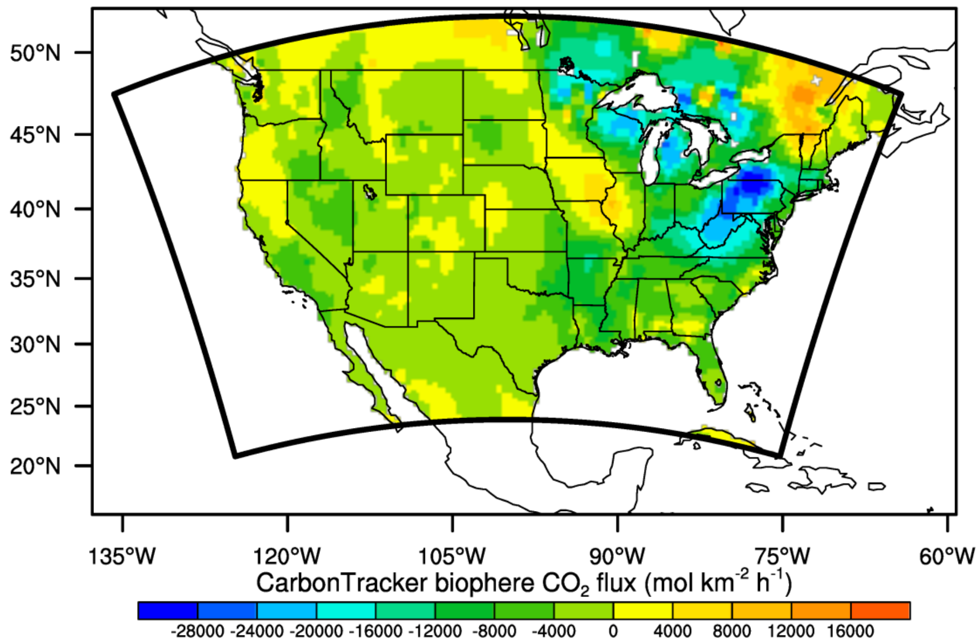


Figure 3. Daily mean CarbonTracker biophsere CO2 flux used in calculating sensitivities and inverse modeling experiments. The daily mean value is calculated as the arithmetic mean of the 3-hourly fluxes from 2011-06-02 00:00 UTC to 2011-06-03 00:00 UTC.

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

The global WRF-chem simulation was conducted with a 256 (east-west) by 128 (north-south) grid, which is about 156x156 km in horizontal. The WRF global model code includes polar filtering procedures for the converging latitude at high latitudes. These filtering procedures are located in multiple modules in the dynamical core. Because they are only needed for global model, we did not *develop their tangent linear and adjoint code for WRF-CO2 4DVar (which is designed for regional inverse only).This mean we can not run WRF-Chem for an global domain inversion to compare with WRF-CO2 4DVar in regional inversion.*

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

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

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

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

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

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

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

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