Response to anonymous Referee #2

We thank the reviewer for their thorough review of our paper and helpful suggestions. Their comments led to substantial clarifications of our arguments and the implications of our work, as highlighted in the responses to the following comments. The reviewer comments are shown in italic font, and our response is shown below each comment, in plain text.

The manuscript presents a number of interesting new ideas and the analysis is very original. Several results seem to have important implications for published, and ongoing work in this field. My biggest concern with the manuscript in its current form is that even after reading it carefully several times, I am neither sure what these implications are nor whether they are fully justified. This is because these implications are often not clearly labeled and discussed. When they are, it is unclear whether the set of assumptions needed to come to the result is still valid in the actual situation they bear on. I fear that the averagely interested reader will therefore not pick out the potentially important message contained in this analysis, which would be a missed opportunity given the amount of work that has obviously gone into this manuscript. I therefore recommend that the manuscript is improved and clarified on a number of key points, which I will outline below.

Point 1: The manuscript currently tries to address many issues that are relevant to current forward and inverse modeling studies of CO2. Although commendable, it makes the paper quite long, and sometimes unfocused. After deriving and explaining the concept of relaxation time (up to section 4.2 and Figure 6) the reader is taken through a large number of applications and ways that the quasi-equilibrium approach can be used. I suggest that the authors shorten the paper by dropping at least section 4.5 and figure 11, and perhaps even section 4.4.

We significantly shortened the revised paper and edited to improve readability. We shortened the revised paper by moving figures 1 and 4 and their discussions to supplemental materials. These sections motivated the equilibrium approximation in different ways, whereas the revised paper adds references to provide similar motivation for our work (as noted in the reply to comment 11 below). The revised manuscript focuses more on the results of applying the equilibrium approximation. We also simplified and shortened the derivation of the autocorrelation function and moved that section to the end of the manuscript (see our reply to comment 21).

Point 2: I am not ready to believe yet that the new mixing diagnostic is able to discriminate a poor transport model from a good one. The main reason is that I haven't seen a model fail yet. Your analysis suggests that only at one location (HFM winter) there is a serious discrepancy but whether this is a model deficiency or not remains unclear. Until I can see your diagnostic as a number or metric that discriminates clearly, it remains to me a quite academic measure.

Our metric was defined by a relaxation time that characterizes the exponential decay rate of concentration autocorrelations. We estimated this decay rate graphically in our discussion paper, but in the revised manuscript we quantify this rate by fitting an exponential function to observed and modeled autocorrelations. This analysis provides a single number for comparing models, observations, and theory, but does not change our conclusions about vertical mixing at our study site (SGP). See our response to comment (22) for further details.

Our diagnostic can be derived from the conservation law under a set of known assumptions that can be tested in the field, which makes this a highly practical and scientific metric, and not simply an academic measure.

Since the submission of our manuscript, NOAA released comparisons between CarbonTracker concentrations and aircraft flask data at the SGP site, which show no significant differences in summer or winter:

http://www.esrl.noaa.gov/gmd/webdata/ccgg/CT2010/profiles/profiles_SGP_01P2.pdf
http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/profiles.php?site=SGP_01P2#imagetable

These results are consistent with our diagnostic, which did not indicate vertical mixing errors at SGP in summer or winter.

The reviewer misunderstands our results at HFM. We did not apply our diagnostic to concentration measurements at HFM or LEF, and so cannot conclude that there are model errors at those sites. The CT/TM5 vertical concentration gradient is inconsistent with ECMWF reanalysis subsidence rates at HFM, but those subsidence rates are model-derived and so may also be in error. In light of the possible misinterpretation of these results, we moved the HFM and LEF results from Fig. 6 to supplemental materials, and focus instead on comparing CT/TM5 to our observations at SGP in a new figure that combines Fig. 5 and Fig. 6a,d (see our response to comment 22).

The revised manuscript adds the statement that we cannot directly compare our results to previous work (Stephens et al., 2007) because that study used observations from sites other than SGP (see our response to comment 31).

We also note a typo in the initially submitted draft to which the reviewer is referring, where the Fig. 6 caption had winter and summer reversed (but the labels at the top of the figure were correct). This error was corrected before publication of the online discussion paper, but the reviewer was not aware of this correction. This might have led to part of their confusion about our diagnostic, since previous studies indicated overestimated vertical mixing errors at HFM in summer, not winter (Stephens et al. 2007, supplemental materials).

Point 3: It will be important also to compare the findings with the ones of Betts et al. (2003) where they study the equilibrium boundary layer over land at larger time scales than diurnal. Their main finding is the importance of surface coupling with boundary layer process like cloud formation.

We added a paragraph to the introduction pointing out how our work follows from previous equilibrium boundary layer studies:

> … over longer periods of time boundary layer depth reflects a statistical equilibrium between surface heat fluxes acting to increase, and infrared radiative cooling acting to decrease boundary layer depth, with clouds coupled through their effect on radiative cooling (Betts, 2004). Radiative cooling in turn balances adiabatic warming in subsiding branches of the circulation, bringing boundary layer trace gas concentrations into equilibrium between surface fluxes and transport by the divergent circulation (Helliker et al., 2004; Betts et al., 2004).

Note that our study takes vertical concentration gradients as given (i.e. observed), and asks how one can infer the vertical mixing rate from these observations, which is an inverse problem. Previous studies (Betts 2004) addressed the forward problem of how to predict vertical mixing rates given the external forcing and coupling between processes such as clouds and radiation.

Further points:

4. Abstract, line 5-10: The phrasing here seems quite strong to me. The previous diagnostics that are mentioned as unreliable here refer to the PBL height and the vertical CO2 gradients, but your sections 4.5 and 4.4 describe that these two quantities play a role in vertical mixing, but do not give the complete picture (subsidence rates and surface flux need to be considered too). To simply call these unreliable diagnostics does not do justice to their contribution I think.

We changed the phrasing in the revised manuscript.

5. Page 3, line 57-60: I have not seen evidence in the rest of this paper that shows how your equilibrium PBL model can help reconcile global flux inversion results with inventories. Can you be more specific about what you meant to show and how it is demonstrated?

We clarified our objectives in the introduction of the revised manuscript.

> Observational constraints can help in reconciling atmospheric inversions that infer strong northern terrestrial sinks, with carbon inventories that estimate weaker northern but stronger tropical terrestrial sinks. However, different observations support different conclusions about transport model errors that would imply either stronger or weaker northern terrestrial sinks.
>
> … This paper uses boundary-layer equilibrium concepts to interpret discrepancies between modeled and observed concentration gradients in terms of transport model errors. Our results show that these discrepancies most likely result from overestimated as opposed to underestimated vertical transport and mixing, implying overestimated northern terrestrial carbon sinks in the inversions.

However we find that seasonal concentrations alone cannot distinguish transport model errors from errors in prior-specified surface fluxes, and we propose a new diagnostic to isolate the effects of transport and mixing.

6. Page 3, equation 1: This form of the mixed-layer equation has an explicit subsidence term, whereas often this effect is included already in the calculation of the mixed layer height evolution (dh/dt). How is dh/dt treated in the rest of this work and does it not also include the effect of subsidence already (more subsidence inhibits dh/dt)? In other words how is the effect of atmospheric dynamics at diurnal scales (determination of h and surface energy forcing) or at larger scales (subsidence and the role of horizontal advection) taken into account?

No, the advection of CO2 by the subsiding wind is not already accounted for by the term involving transient changes in mixed-layer depth. See Betts (1992), which we closely follow in our equation (1). A paragraph was added to the revised manuscript to clarify this issue (see section 2.1, second paragraph, of the revised manuscript). The effect of surface energy forcing on *h* is taken into account by using the time-dependent, modeled mixed-layer depth. Our study uses the CT/TM5 boundary layer depth, as predicted from the surface energy forcing in the ECMWF model underlying CT/TM5. We added this statement to the revised manuscript (Section 2.2, second paragraph).

7. Page 3, line 70: "Where Reynolds averaging is implied." This might be formally correct as the mixed-layer equations require one to average over turbulent time scales, but I do not think the meaning of this remark will be clear within the context. Could you add some references to more basic explanations of this equation and its applications (Betts et al, Tennekes 1973, . . .)?

We added a reference to Betts (1992).

8. Page 3, line 74: The term "storage" is not well-known in the CO2 modeling community (I think it has its basis more in the flux measurement community). Perhaps add a sentence that explains this term as the change of CO2 within the mixed-layer domain.

Clarification was added to the revised paper.

9. Page 4, line 103: When introducing the relaxation time it might be good to explain directly which steady-state is reached within that time scale and what it means for the CO2 budget. Page 4, line 105: Connected to the previous point, which vertical velocity is implied by the term W? Is this a subsidence rate? Or a turbulent mixing velocity? Or an entrainment velocity?

We added a new paragraph at the beginning of Section 3:

> The relaxation time characterizes how long it takes for tracer concentrations to come into equilibrium with vertical transport by the subsiding flow. The

characteristic subsidence rate W reflects the aggregate of all processes affecting subsidence, including synoptic systems. Frontal passages and other mechanisms of ascent are coupled to subsidence through the conservation of mass and energy (e.g. Lawrence and Salzmann, 2008). Adiabatic warming approximately balances infrared radiative cooling in the subsiding branches of the circulation, which in turn balances latent heating by moist convection in the precipitating storm systems that comprise the ascending branches.

10. Page 4, line 111: What type of atmospheric circulations do you refer to? The ones that would enhance W? Are these synoptic systems with large subsidence? Or frontal passages with strong lofting of surface CO2?

Please see our response to the above comment. Also note that the mention of 'atmospheric circulations' on line 111 was removed as part of our effort to shorten the manuscript and clarify our arguments.

11. Page 5, line 136 to 146: I had trouble following the logic of this paragraph. This is one of the examples where I feel an important point is made about diurnal versus seasonal rectifier effects and how entrainment is or is not important in these processes, but I cannot reproduce your arguments. Which evidence from short-term observations do you refer to when you say they are extrapolated to explain the seasonal rectifier effect? Why would I expect the growth and decay of mixed-layer concentrations to increase when averaging over longer time-scales?

Section 3.4 makes the same point about diurnal versus seasonal rectifier effects, so we decided to remove the more confusing paragraph (lines 136 to 146) and focus on Section 3.4 (which became section 3.1 in the revised manuscript). This change also helps to shorten the paper, in response to the reviewer's point (1) about length.

Our point was that the storage and entrainment terms represent rates of change, as calculated by changes in concentrations or mixed-layer depths, divided by a time interval. Concentrations that fluctuate by 10 ppm from day to day would have to fluctuate by 90 ppm over the course of a 90-day season for the storage term to remain as large in seasonal averages as in daily averages. Such large fluctuations would show up in timeseries of concentrations, but they do not, as indicated by comparing the 90-day and 1-day running averages in Fig. 1a,b. In fact the seasonal changes are the same order of magnitude as daily changes, which means the storage term becomes an order of magnitude smaller at seasonal timescales than at daily timescales.

The same argument applies to the entrainment term, by examining the mixed-layer depth timeseries in Fig. 1c. These observations suggest that neglecting the vertical and horizontal advection terms will result in large errors at seasonal timescales, when storage and entrainment account for a smaller portion of the CO2 budget relative to advection.

We can motivate the same result in terms of length scales. Fluctuations in mixed layer depth can move a particle in the mixed-layer over a distance on the order of the mixed layer depth (~1000 m), but this distance cannot exceed the maximum mixed-layer depth. However advection can move a particle by a distance on the order of the wind speed multiplied by the time over which the wind blows. Therefore, over longer periods of time, advection becomes much more efficient than fluctuations in boundary layer depth at moving mass into and out of the mixed-layer.

12. Why would you even take 90-day running means of this data when the quantitative evaluation that follows in 3.4 uses non-overlapping segments? Please help me understand the argument.

The 90-day running averages (in Fig. 1) aid in comparing seasonal variability to shorter-term (e.g. daily) variability.

We use non-overlapping segments in section 3.4 to create statistically independent samples of each budget term as a function of timescale. There are several years of data, each having one 90-day "non-overlapping" segment in each season (e.g. from June 1 to August 30). We added the statement (Section 3.1, third paragraph of the revised manuscript): "Note that the 90-day non-overlapping segments resulted in one 90-day sample for each year, for a total of 5 and 7 samples for the SGP and CT/TM5 data, respectively, whereas the 45-day non-overlapping segments contain 14 and 10 samples, respectively, and so on."

13. Page 6, line 185: In this part of your analysis you exclude mountain sites, but you later on use HFM to show that mixed-layer equilibrium times can be used to identify erroneous vertical mixing in the model possibly due to the effect of the Appalachian mountains. Also, you use HFM as an example to show that mixed-layer heights and mixing can be anti-correlated. This seems contradictory to me.

We respect the reviewer's concerns about the generality of our findings at HFM and have added a sentence to the revised manuscript to describe similar results along the east coast and eastern Atlantic, as shown in a new supplemental figure described below and attached at the end of this response.

First, recall that we used the ECMWF interim reanalysis for part of our work because the CT/TM5 data did not include vertical velocities. We excluded Estevan Point, British Columbia (ESP), which has a surface topography of 343 m (above sea level) in ECMWF reanalysis, but a measurement site elevation of 7 m. Such differences in elevation could result in errors in subsidence magnitudes of about 25%, according to Fig. 11. The only other North American site included in Stephens et al. (2007) that we excluded here was Briggsdale, Colorado (CAR), but inverse model errors were much smaller there than at HFM or LEF, according to their study.

Of course we could always include more sites, but this will not change our conclusions about the existence of anti-correlated mixed-layer heights and mixing. To indicate the generality of our results at HFM, we included a figure (attached at the bottom of this response) showing the difference between winter and summer subsidence rates in a longitude-height cross-section at the latitude of HFM. This figure shows that many North American locations experience weaker subsidence in summer than winter, particularly over the east coast and extending well into the Atlantic. Much of the variability in subsidence occurs on local to regional scales, indicating the influence of topography and perhaps land-sea contrasts. However, we also see a larger-scale pattern with enhanced winter subsidence over the Atlantic and east coast, and diminished winter subsidence over the Midwest and central plains, possibly reflecting the influences of stationary wave patterns. These results support our conclusion that seasonal rectifier effects depend on the tropospheric circulation, in addition to variations in boundary layer height.

14. Page 7, section 3.5: Although this is a very nice analysis of the terms in equation 1, I kept wondering while reading this where the horizontal advective tendency has gone? Does it simply average out since you include many synoptic events when using large windows? Or is the budget you present not actually closed and are we missing a term that is of equal size as the vertical advective term or the surface flux term? Since the surface + horizontal flux is not included in Figure 4 it is hard to see whether an equilibrium is reached between the vertical advective term and surface flux, or whether these fluxes are both simply dwarfed by horizontal advection.

We are specifically referring to how the storage and entrainment terms become smaller relative to vertical advection as the timescale increases. We included horizontal advection in subsequent analyses of the CO2 budget and vertical concentration gradients. We added a statement to the revised manuscript to clarify when horizontal advection should be considered part of the equilibrium approximation:

> More generally, the equilibrium approximation to equation (1) should include horizontal advection if this term is of similar magnitude to vertical advection and surface fluxes. We included horizontal advection in our study of the seasonal cycle, shown in the following section.

We also rearranged the revised manuscript, so readers can immediately see that horizontal advection does not change our argument about the relative magnitude of storage and entrainment compared to vertical advection. We moved Figure 4 to supplemental materials because it contains the same information as Figure 3 (that storage and entrainment become smaller relative vertical advection at seasonal timescales). We moved the section on seasonal budgets (Section 4.3, including Figs. 7,8,9, in the old manuscript) to right after the section on scaling (after Section 3.5 in the old manuscript).

15. Page 8, line 245: What do you mean by "non-linear vertical tracer advection"?

We mean the term defined on line 340, of Section 4, which accounts for the correlation of perturbations in vertical velocity with perturbations in vertical concentration gradients. We thank the reviewer for pointing out the need to define this term as it appears. We rearranged the revised paper so that this term is defined when it is first used (Section 4.3 was moved to Section 3).

16. Page 8, line 249: Is this the same assumption as line 240 where horizontal wind divergence varies slowly in time compared to perturbations in h?

The reviewer's comments 16-20 all pertain to the continuity equation. We did not explicitly show this equation in our original paper, which understandably has led to some confusion. We addressed this issue by adding the continuity equation to our revised manuscript (equation 5 of the revised manuscript), and thank the reviewer for pointing out the need to clarify the import role of mass continuity in our derivations.

Horizontal CO2 advection and horizontal divergence are two separate processes, and our diagnostic will most accurately reflect vertical mixing when both processes either vary slowly (e.g. seasonally) or represent temporally uncorrelated white noise. The reviewer's comment raises the interesting point that there may be a way to relate these two assumptions to an assumption on the statistics of horizontal wind fluctuations, but we have not yet derived such a relationship.

17. Page 8, line 254: Is this a redefinition of the relaxation time? You previously used a typical time-scale of H/W for the relaxation time whereas now it is related to the horizontal wind divergence only and no longer to H.

There is no inconsistency between these definitions, but we clarified this point in the revised manuscript and thank the reviewer for pointing out the possible confusion. H/W is the same as the vertically integrated horizontal wind divergence, after using the continuity equation. It helps to keep in mind that divergence has units of inverse time, and so does H/W, so they are dimensionally consistent, too.

18. Is this short time scale now the same assumption as "long time limit" (line 249), and as "slowly varying horizontal divergence" (line 240)? You can see how the many assumptions that you describe about times for averaging and times for variations are confusing me here, and hence I have a hard time understanding the implications of equation 4.

We used the term “long-time limit” in our derivation of the autocorrelation function. We removed mention of this term in the revised manuscript, by simplifying our derivation of the autocorrelation function (see our response to comment 21). The solution (equation 3 of the old manuscript) decays to 1/e of its initial value over the relaxation time. In the “long time limit”, i.e. after many relaxation times have passed, the exponential term in equation (3) vanishes and we obtain equation (4), after using the continuity equation to rewrite the divergence in terms of the subsidence rate.

“Slowly varying divergence” does not mean “slow divergence” and therefore does not conflict with the relaxation time being the short timescale. By “slowly varying” we mean approximately constant in time. We know that the divergence rate is not constant in reality, but as long as it varies slowly (seasonally), we can characterize it with a single number like in any other exponential decay problem involving rate constants.

19. Page 8/9, line 255-266: I believe it is important to address the statement in line 246 that says that the flux term F can include the slowly varying component of the horizontal flux. To what extent is this needed?

Horizontal advection is an important part of the CO_2 budget at some sites (see Fig. 8c), so an ideal vertical mixing diagnostic should include horizontal advection. We presented an approximate solution to this problem in the discussion paper, by noting that horizontal advection consists largely of a component that varies slowly in time compared to the relaxation time. We then used the fact the surface flux is mathematically equivalent to an external forcing on the vertical CO_2 gradient (in equation 2), so that there is a mathematical similarity between the solution to the continuity equation under slowly varying surface fluxes (equation 4) and the solution that includes the slowly varying part of horizontal advection together with surface fluxes. We clarified this similarity in the revised paper by formally incorporating the seasonal component of horizontal advection together with surface fluxes, as part of a generalized forcing function (which we call G, as opposed to F). This is a standard procedure in solving differential equations, to collect terms that can represent external forcing on one side of the equation and to make reasonable assumptions about the nature of the forcing that allow for realistic solutions.

20. The rest of your manuscript refers to F as surface flux and you make the argument that vertical gradients should be combined with the surface flux to assess vertical mixing in a model. But if horizontal advection is an important part of the surface forcing F, then it seems to me that horizontal motions affect two sides of the equation 4: both the vertical wind (as calculated from horizontal divergence) and the F. Does your diagnostic then still assess a model's ability to capture vertical exchange, or is it simply a test of horizontal winds?

Our model still assesses the vertical exchange. Horizontal divergence is related to vertical exchange through the continuity equation, which we added to the revised manuscript. Horizontal CO_2 advection and horizontal divergence are two separate processes. Please see our response to comment 16.

21. Page 9, lines 278-294: Once again the logic behind the derivation of the observed relaxation time is difficult to follow. Different parts of previously derived results are put together, and assumptions are made that relate fluctuations of terms in the equations to the autocorrelation of observed CO_2 gradients. Assumptions are made on the fluctuations of surface fluxes but stated to be not quite true, but despite that the proposed metric will work. As reader, I can only decide to believe your math and see what comes next, but there remains a feeling that everything which follows should be regarded with some skepticism. Of course, the results in figures 5 and 6 suggest that indeed the autocorrelation method yields the same metric as the full derivation of the relaxation time scale from the terms in the equations. So perhaps it would be better to use figure 6 (or a subset thereof) to first convince the reader that the relaxation time scale can be derived from the observed gradients only, and then apply it as a metric.

We greatly simplified the derivation of the relaxation time by drawing an analogy between the mixed-layer tracer conservation equation (equation 1) and an autoregressive process (which we define in equation 3 of the revised manuscript). We use this analogy as a guide to the derivation and the assumptions involved. We think the revised derivation is much easier to follow and the assumptions behind our method have been made clearer.

The reviewer remarks that the assumptions behind our diagnostic cause some skepticism. Not knowing if or when a method fails should not increase ones confidence in applying it to the real world. One of the advantages of our diagnostic is that it can be derived from the conservation law under a set of known assumptions that can be tested in the field. These assumptions state under what conditions observations of vertical concentration profiles can be used to constrain transport and mixing, which is a valuable contribution to this research area.

22. Page 10, line 314: This is the result that drives my point #2. Although it is interesting to see that there is no evidence for a vertical transport bias in CT/TM5 at SGP, I cannot help but think that perhaps the metric you show is not discriminating enough. To my eye, it looks like any of the gray curves and symbols in figure 5a+b and 6a-e could have been swapped without obvious differences jumping out at me. Could you visualize or quantify the different relaxation times in some way that makes them stand out more? Or could you construct a case where the simulated relaxation times are proven wrong by observations, for instance by using an Alert, Canada modeled time series of CO2 to estimate the relaxation time at SGP?

Our diagnostic is consistent with a recent study that does not find significant errors in CT/TM5 vertical concentration gradients at SGP in summer or winter (see our response to Point 2 above). We addressed the reviewer comments by combining the SGP results shown in Figs. 5 and 6, to directly compare our observations to CT/TM5. We performed an additional analysis to ensure that any differences between CT/TM5 and SGP observations stand out in our diagnostic. This analysis is described in the revised paper as follows:

> We estimated divergence rates of 6.23×10^{-6} and 1.04×10^{-5} s^{-1} for summer and winter, respectively, using a non-linear least-squares fit of equation (7) [equation 6 in the old manuscript] to the CT/TM5 autocorrelations. We obtained similar divergence rates from the observed autocorrelations (6.10×10^{-6} and 1.13x10-5 s^{-1} for summer and winter, respectively). Differences in sample means between the CT/TM5 and observationally-derived divergence rates were not statistically significant, and furthermore would amount to vertical concentration gradient errors on the order of 0.1 ppm (using $\Delta c \sim -F/w$), much smaller than the errors reported in a previous study at other sites (Stephens et al., 2007).
>
> Our proposed diagnostic has the capability of determining if large errors seen in transport model inversions result from errors in the modeled divergent wind field. Transport model inversions predict vertical concentration gradients almost twice as large as observations at some sites (Stephens et al., 2007). The equilibrium

boundary layer approximation implies that these errors would appear as similarly large errors in mixed-layer divergence rates, or equivalently in subsidence rates (i.e. twice as fast in transport models as in observations, according to $\Delta c \sim -F/w$). Such large errors are on the order of the difference between summer and winter divergence rates at SGP (6.23×10^{-6} and 1.04×10^{-5} s^{-1}, respectively) which is statistically significant in our analysis, and therefore should be detectable using the methods presented here.

23. Page 11, line 335-338: Your Reynolds decomposition of non-linear vertical advection suggests to me that the non-linear part alluded to previously was the covariation between vertical motions and vertical gradients?

Yes.

24. Page 11, line 350: Is the CT/TM5 surface flux shown in the figure the actual diagnosed surface flux from the biosphere in gridbox SGP? Or is it the sum of the calculated contributions in Figure 8? In the latter case it would include horizontal advection but in the former it would not.

Fig. (9a) shows that the surface flux recreated from the sum of all budget terms (as shown in Fig. 8) approximately agrees with the surface flux predicted from the CarbonTracker data assimilation system (Fig. 9b), as a consistency-check. We clarified this point in the revised paper. Fig. 9a and Fig. 9b both reflect horizontal advection, because the CO2 budget relates the two processes.

25. Page 12, section 4.5; although this analysis is nicely done, I feel like the amount of information that has been handed is too much at this point. The main message from this section as summarized in the last paragraph could go into the discussion or conclusions, and one could shorten the paper by a few more paragraphs and one figure without losing anything of your story. Please consider this.

The revised paper is much shorter.

26. Page 13, line 430-434: This discussion is much more nuanced than the statement in your abstract.

We changed the wording of the abstract in the revised paper.

27. Page 14, lines 447-460: This simple explanation of your main points is very helpful to non-expert readers, well done. Perhaps consider more of these inserts along the way.

Thanks. We edited the revised paper to make it easier to read.

28. Page 14, line 470: I am not sure I understand this argument. If one starts from a biased prior flux and subsequently adjusts it because observed gradients do not match modeled ones, then what brings the risk of overestimating the land uptake? Somewhere *in this argument the model transport has to appear. The way I read this now I see your "compensating stronger summer uptake" simply as a correct adjustment of an overestimated prior flux? This does not yet explain NH land sinks that are too large in my opinion.*

We respect the reviewer's opinion, but await further evidence before making conclusions about the NH land sinks. We are not the first to suggest that errors in inverse model concentration fields could also reflect errors in prior-specified surface fluxes (Kaminski et al., 2001; Gurney et al., 2005, Peters et al. 2007; Yang et al. 2007). Underestimated summer fluxes could also explain underestimated vertical concentration gradients, according to our results (equation 7 of the discussion paper). We would like to see the sensitivity of inversions to the prior flux seasonal amplitude and timing of the seasonal cycle before accepting the reviewer's viewpoint.

29. Page 15, line 478: I might misunderstand your argument again, but increased vertical mixing to me seems to lead to smaller vertical gradients, not larger.
Page 15, line 486: "desirable" = "plausible"?

Yes, that was a typo, and plausible is indeed a better word, thank you.

30. Page 16, top paragraph: This argument seems very speculative and distracts a bit from the rest of the discussion. Please consider removing it.

We removed it.

31. Page 16, line 519-524: To what extent do your results suggest that the Stephens et al (2007) results are wrong? Have you tried to look at his supplementary material to formulate a different theory on what the problems are with the transport models used or the fluxes derived? After all, Stephens et al. could not find a single model that reproduced obscured vertical gradients in both seasons as well as in the annual mean. Your figure 12 might give a guideline to come to a new hypothesis.

We claim that previous efforts to infer vertical mixing rates from concentration gradients will work only if there are no errors in prior-specified surface fluxes. Therefore we would like to see the sensitivity of inversions to the prior flux seasonal amplitude and timing of the seasonal cycle before accepting their viewpoint, and that of the reviewer. The reviewer has not commented on other studies supporting our suggestion that vertical concentration gradient errors could also reflect errors in prior fluxes (Peters et al., 2007; Yang et al., 2007), although we cited those studies in our discussion paper. Other possibilities include errors from seasonally varying fossil fuel emissions (Gurney et al.. 2005), or aggregation errors that result from specifying the prior fluxes over large geographical regions (Kaminski et al., 2001). The latter possibility was noted in the supplemental materials of Stephens et al. (2007).

We hope our work will encourage other measurement programs to use our diagnostic or other metrics designed to separate the effects of surface fluxes and vertical mixing on concentrations.

On the issue of the existence of systematic transport model errors, we added a paragraph to the revised manuscript (Section 5, fourth paragraph) clarifying that our study does not include the full set of sites and models used in previous studies:

> Our results do not necessarily conflict with previous work suggesting large errors in vertical transport and mixing (Stephens et al., 2007), particularly because we have not applied our diagnostic to sites showing large discrepancies between observations and transport models inversions. Furthermore, CarbonTracker uses a higher resolution (two-way nested) version of the TM5 model than what inverse studies have used, and its treatment of surface fluxes differs from those studies by assimilating observed and model-derived surface fluxes every 3-hours and including effects of inter-annual variability and vegetation fires. CT/TM5 comes into closer agreement with carbon inventories and has overall smaller concentration profile errors compared to the TransCom3 model inversions (Peters et al., 2007).

32. Figures 5+6: What causes the gray line to be a band rather than a line? Why is the width of the band different for different lag?

The width indicates the range of exponential decays predicted by the ECMWF reanalysis divergent wind over the 7 summers and winters between 2001 and 2008. The revised manuscript uses the standard deviation for the width, as explained in the revised figure caption.

33. All figures: Colors for the lines would really help a lot. And color figures are free in this journal, so take advantage!

We combined figures 5 and 6 to provide a clearer comparison of observations and CT/TM5. We also improved the lines for clarity, and added color (see Fig. 8 in the revised manuscript). We will upload color versions of the other figures as soon as possible.

References:

Betts, A. K.: FIFE atmospheric boundary-layer budget methods, Journal of Geophysical Research-Atmospheres, 97, 18523-18531, 1992.

Betts, A. K., B. Helliker, et al.: Coupling between CO2, water vapor, temperature, and radon and their fluxes in an idealized equilibrium boundary layer over land, J. Geophys. Res.-Atmos. 109, 2004.

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.-Atmos., 110, D10308, doi:10.1029/2004jd005373, 2005.

Helliker, B. R., Berry, J. A., Betts, A. K., Bakwin, P. S., Davis, K. J., Denning, A. S., Ehleringer, J. R., Miller, J. B., Butler, M. P., and Ricciuto, D. M.: Estimates of net CO2 flux by application of equilibrium boundary layer concepts to CO2 and water vapor measurements from a tall tower, J. Geophys. Res.-Atmos., 109, D20106, doi:10.1029/2004JD004532, 2004.

Kaminski, T., P. J. Rayner, et al.: On aggregation errors in atmospheric transport inversions, Journal of Geophysical Research-Atmospheres, 106, 4703-4715, 2001.

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, P. Natl. Acad. Sci. USA, 104, 18925–18930, 2007.

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 CO2 , Science, 316, 1732–1735, 2007.

Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., Sweeney, C., and Wennberg, P. O.: New constraints on Northern Hemisphere growing season net flux, Geophys. Res. Lett., 34, L12807, doi:10.1029/2007GL029742, 2007.

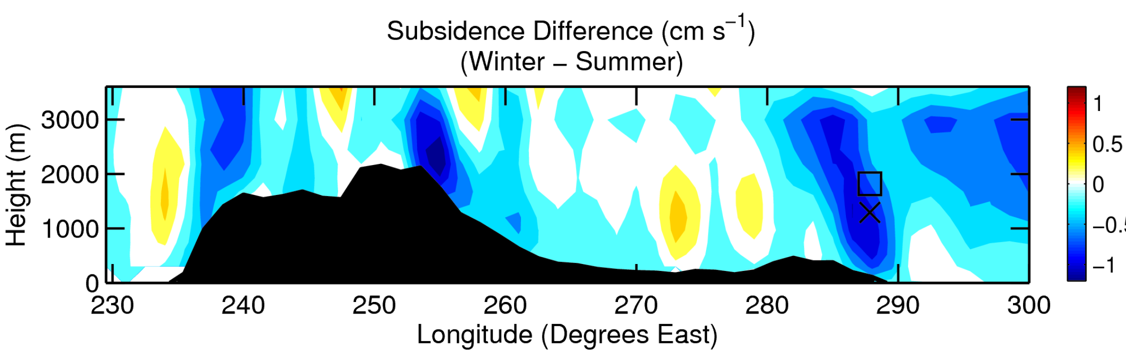

Figure: Longitude-height cross-section showing the difference between winter and summer subsidence velocities at the latitude of HFM. Negative values indicate greater subsidence (stronger descent) in winter than in summer. Subsidence is calculated by averaging the negative values of vertical velocity over each season (90-days). Black shading indicates the height of the surface topography. Symbols indicate summer (square) and winter (x) mixed-layer depth at the longitude of HFM.