Response to Review 1

The impact of meteorological analysis uncertainties on the spatial scales resolvable in CO2 model simulations" by Saroja M. Polavarapu et al.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-346, 2016.

Original comments are in black text. Our responses are in blue text.

Unrequested modification: In addition to changes prompted by reviewers' questions, we decided to change the predictability metric from the square root of the global mean of zonal variance to a more intuitive one: global mean of zonal standard deviation. In doing so, a coding error in the original diagnostic was found that affected only vorticity and divergence: abnormally large normalized predictability error was produced in the stratosphere and mesosphere. Since the square root and global mean operators do not commute, the change in Figures 9-10 resulting from the change in diagnostic did not have to be small, but were in fact minimal. Thus the changes seen in the revised Figs. 9-10 are mainly from correcting the coding error.

This paper presents the adaptation of the GEMS-MACH model to simulate atmospheric CO2 within an NWP framework. An evaluation of the simulations using different types of observations is performed. The question of predictability associated with different aspects of the model (growth of transport errors, imperfect initial conditions, and impact of convection) is also addressed. The paper is very well written and the predictability study shows very interesting results that are relevant to the atmospheric composition and carbon cycle community. So I would recommend the paper to be accepted subject to minor corrections. A list of general and more specific comments can be found below.

Response: We thank the Reviewer for the encouraging remarks and for the thorough reading of the paper. We believe that the revised manuscript has greatly benefitted from the comments of both Reviewers.

GENERAL COMMENTS
- The potential role of the CO2 fluxes in the predictability of atmospheric CO2 is not clearly presented. Although this is not the main focus of this study, the generally important role of the fluxes in the predictability of CO2 should be emphasized. The relevance of the fluxes would become more evident if the predictability diagnostic would include a measure of the evolution of mean error with forecast lead time. In the paper two different surface fluxes are used but their differences are not explained. A description of the main differences (not only in formulation but in actual flux difference) would help the interpretation of the results in terms of flux errors. Can their difference be used to test the impact of flux errors on the predictability?
  Response: As the reviewer notes, the role of CO_2 fluxes in the predictability of atmospheric CO_2 is not a main focus here and this is partly because the only fluxes we had available at the time were two sets of retrieved fluxes from CarbonTracker. As implied by the Reviewer, the use of two essentially equivalent retrieved fluxes does not shed light on the impact of flux errors on predictability. More can be learned by comparing prior and posterior fluxes but we did not have access to this. As noted in the discussion section of the original manuscript (p.24, lines 27-35), some interesting extensions and applications to different observing systems can be done and are now in progress. The only reason that these two sets of fluxes were used was to illustrate this potential. However, it is clear that this potential is not being well illustrated as hoped. Moreover, it seems to be detracting from the main message. Therefore, we have removed all references to the two sets of fluxes and used only CT2013B which is the preferred product (according to Andy Jacobson of NOAA). In addition, we have tried to emphasize the important role of fluxes in

predictability results throughout the text of the revised manuscript. As for bias, we have computed the global mean of predictability error normalized by the standard deviation (as in Fig. 9 of the revised manuscript) and included it below. For CO_2, the bias is generally negligible compared to the standard deviation in the troposphere. This is likely because in our experimental design, the flux of the reference and perturbed cycle is the same. Moreover, model errors and initial CO_2 state errors are also eliminated because we focus on meteorological state errors. Bias is quite prevalent however in model forecast to measurement comparisons where flux, initial state, meteorology, model formulation, observation and representativeness errors are present. A paragraph discussing the important of forecast bias was added to the discussion section of the revised manuscript.
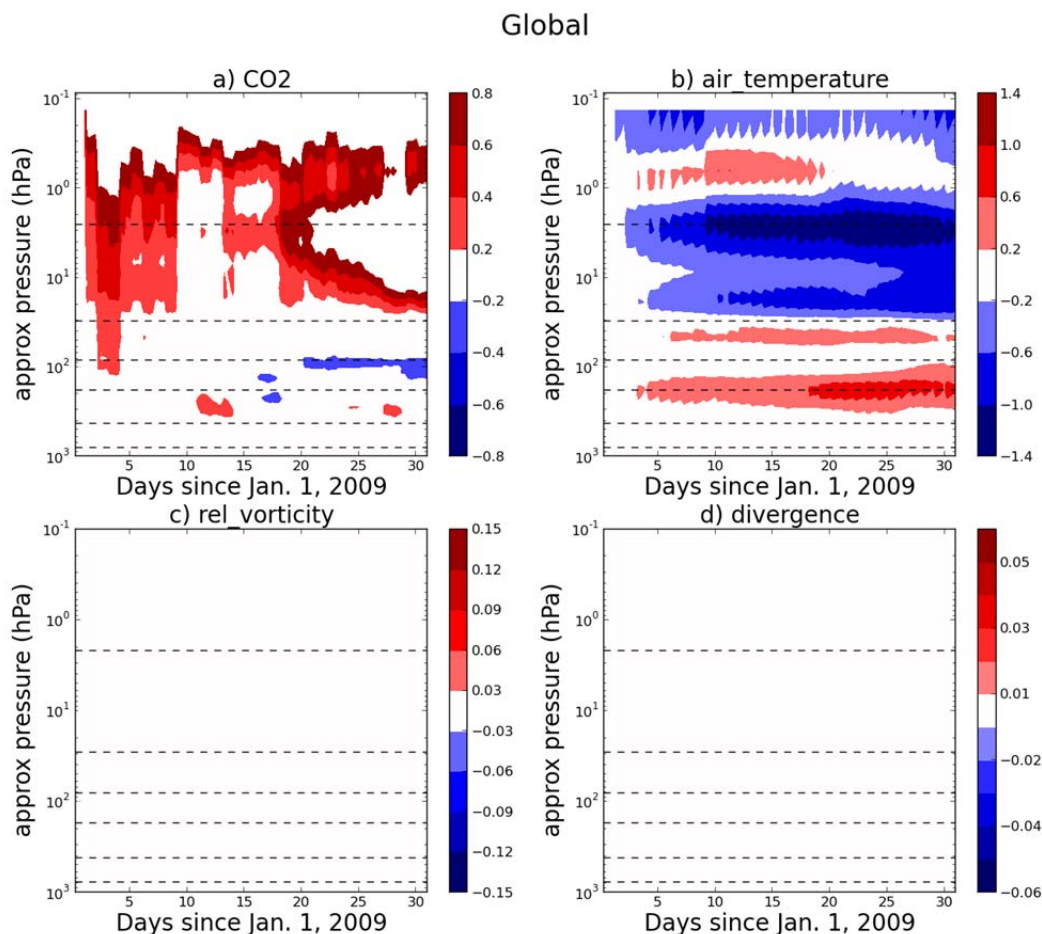

Figure R1-1: Global mean difference between the reference and climate-mode cycles during January 2009. Values are normalized by the global mean zonal standard deviation of the initial state. Mean departures are negligible for the wind field (relative vorticity and divergence fields), are important for the temperature near the tropopause and negligible for CO_2 in the troposphere.

- The characterisation of analysis errors using a 6-hour shift seems an unrealistically large estimate of the analysis error. If that is the case, the results might not be indicative of the real impact of using imperfect analyses in CO2 models (e.g. in flux inversions or in forward simulations), but will only provide an upper limit for scales that can be predicted even with the use of inaccurate initial conditions for the meteorology. I think it would be useful to compare this specification of uncertainty with the uncertainties provided by your analysis system.

Response: The reviewer is correct that a 6h shift could be large compared to a 6h analysis error. But as explained in the manuscript, this was done because we did not have access to analysis errors in 2009-2010. While an ensemble Kalman Filter has been operational at ECCC for some time, results were not archived. Archives begin in 2011 but are not reliable. Looking at EnKF ensemble spread estimates for January 2012 (a different time period), we see that analysis errors are indeed smaller than 6h difference errors but there is no simple pattern to the scaling factor—it depends on time, location, height and variable. Not surprisingly, no proxy can replace a true analysis error even though the analysis error is itself an imperfect estimate. Of course, as noted on p25, lines 12-15, we are in the process of developing an EnKF and will someday have access to actual analysis errors from our own system. When that system is available we will be able to quantify the limiting spatial scales due to analysis uncertainties. For the purpose of this work, we agree that it is important to emphasize the fact that the 6h shift is an overestimate of analysis errors and this is done in the revised manuscript. But the main conclusions regarding uncertain analyses, namely, that (1) some scales in CO_2 cannot be resolved due to the presence of analysis uncertainties, and (2) that for small enough spatial scales, the error spectrum due to analysis uncertainties asymptotes to the predictability spectrum hold regardless of the size of the error.

- The comparison of the errors associated with imperfect analysis, the predictability and omission of convective transport as well as flux errors are done for 1 to 2 month-long simulations. It is not clear whether the same results would hold for shorter simulations (e.g. 5-day or 10-day simulations) when the predictability error is not saturated.
  Response: Actually the simulations were done for one or two years but the spectra were averaged over a month in order to get more robust spectra. On short time scales like 5-10 days one would need an ensemble of short forecasts that can be averaged and we do not have that capability yet. This is routinely done at operational centres and when our assimilation cycle (rather than just forward model simulations) based on the ensemble Kalman filter is developed, this could be done. Here our focus is on the saturated error levels where robust statistics can be obtained by averaging spectra. Animations of daily spectra reveal that spectra are generally slowly varying in time and height but small noise exists so that averaging in time and height is appropriate. Our spectra are computed from roughly 31x12=372 samples. The question of short time scales for model and other components of transport error is, however, an interesting one that is left for future work with our ensemble Kalman filter once it is developed.

- Finally, it would be useful to have a clearer message of the implications of the findings regarding the impact of the different error sources for the carbon cycle community.
  Response: We agree that implications of our results for the carbon cycle community need to be much better presented. Therefore, we have completely rewritten the abstract and the discussion section, and revised the introduction. We have also now added a new section, just after the Introduction, which clarifies the relevance of predictability for flux inversions. In this section the terminology used in atmospheric dynamics and data assimilation is connected to that used in flux inversions. Prediction or forecast error is related to transport error, and we define the components of transport error, namely, flux error, initial CO_2 state error, meteorological state errors, and model error. With these definitions, the results and their implications are easier to explain.

SPECIFIC COMMENTS

- The CAMS CO2 analysis and forecasting system is using the Bermejo and Conde (2002) mass fixer since 2015 with the introduction of the high resolution CO2 forecast. A paper has been recently submitted to GMD to document it:
  Agusti-Panareda, M. Diamantakis, V. Bayona, F. Klappenbach, and A. Butz (2016): Improving the inter-hemispheric gradient of total column atmospheric CO2 and CH4 in simulations with the ECMWF semi-Lagrangian atmospheric global model (submitted to GMD)
  Response: Thanks for the update. The sentence on page 7 lines 1-2 has been changed to:

"For greenhouse gas transport, the ECMWF forecast model also uses the Bermejo-Conde scheme (Agusti-Panareda et al., 2016). "

- Page 11, line 15: Why should the the mass mixing ratio of CO2 be adjusted to account for changes in the atmospheric pressure associated with analysis increments? Although they will affect the conservation of CO2 mass in the model, the associated error is small and it does not grow with time. So why can't this error be considered part of the small error in mass conservation associated with model errors in atmospheric pressure?
  Response: The reviewer is correct that it is not necessary to account for this error since it is small and does not grow with time. We chose to account for it because it removes a noisy component of the mass budget plots and makes it easier to spot problems.

- Page 13, line 14: How frequently are the retrieved fluxes from Carbon Tracker updated?
  Response: Fluxes are provided every 3h and inserted every model time step (15 minutes). This information was added to the revised section 4.2.

- Page 13, line 11: It is not clear what is the purpose of using two versions of the Carbon Tracker fluxes. What is the difference between CT2010 and CT2013B apart from the fact that the latter extends further ahead in time? A figure showing the difference in the seasonal cycle and the annual mean distribution might help to understand how large the differences are and where/when they are more pronounced.
  Response: The two products are largely comparable. CT2013B is a later product and is thus viewed as "better" by NOAA. The only reason that both products were used is to see the impact of flux differences on spatial scales resolved. But because the two products are effectively equivalent, this point is not being well illustrated. Moreover, it is clear that the use of two equivalent flux products is detracting from the main conclusions regarding predictability. Therefore, we have removed the experiment using CT2010 from the manuscript (Figs. 4,5,6,8,11,12) in order to remove this distraction.

- Page 14, lines 23-24: This line gives the impression that improving the diurnal cycle of CO2 in the boundary layer is only a matter of improving the parameterisation of turbulent diffusion. Whereas the atmospheric CO2 diurnal cycle near the surface also depends on the diurnal cycle of the CO2 fluxes. So improving the turbulence mixing parameterisation will not necessarily improve the CO2 diurnal cycle, in particular if there are compensating errors in the retrieved fluxes. Also, since the retrieved fluxes are not using observations of the CO2 diurnal cycle, their diurnal cycle will be mainly reflecting the prior fluxes. This means that the error in the atmospheric CO2 diurnal cycle will have an even larger component from the CO2 flux error.
  Response: We agree with this point as it was not our intention to oversimplify the issue of boundary layer modelling. Therefore we have removed the last 3 sentences of this paragraph.

- Page 14, lines 30-31: "The GEM-MACH-GHG simulations with different posterior fluxes are closer to each other than to CarbonTracker in the tropical Atlantic". I see large differences with CarbonTracker in the Arctic and Tropical Pacific too. Again, it would be useful to know the difference in fluxes between CT2010 and CT2013. If their difference is very small, then it is not surprising that the two simulations are so similar and we mainly see the differences between the CT transport and the GEMS-MACH transport.
  Response: As noted above, it is clear that the use of two essentially equivalent sets of retrieved fluxes is not helpful, so the simulation with CT2010 was removed from this article.

- Page 15, lines 1-2: I do not understand how the finding that the fluxes are more important for explaining the tropical structure can be linked to the finding that posterior fluxes are sensitive to prior fluxes in the tropics.
  Response: With the removal of the GEM simulation with CT2010 fluxes, this sentence becomes irrelevant and was removed.

- Page 15, lines 24-25: The main reason why the biases here are lower than those from Massart et al. (2016) is because of the use of retrieved fluxes. If the retrieved fluxes from flux inversion systems are well constrained by the observed annual growth rate (which is similar in all background stations) then they should produce annual biases close to zero with forward models. The GOSAT data assimilated by Massart et al. (2016) is too spare to be able to constrain the global growth rate by just adjusting the atmospheric concentrations using a 12-hour window.
  Response: This is a good point. See response to the next point.

- Page 15, lines 25-26: I don't agree with this hypothesis. The main difference between the results here and those by Massart et al. (2016) is the fact that the retrieved fluxes are well constrained by the observed annual global atmospheric growth rate from surface observations, which should be similar to the growth rate observed in most TCCON stations. The system used by Massart et al (2016) is not adjusting the fluxes using long assimilation windows and therefore it is having a harder time constraining the annual global growth rate.
  Response: This argument makes sense. The sentence: "Since the analyses in Massart were based on GOSAT whereas CT2013B used surface data, this may indicate a better agreement of surface observations with TCCON." was changed to: "This is because the surface observations assimilated by CarbonTracker with long assimilation windows are able to constrain the global atmospheric growth rate whereas the system used by Massart et al. (2016) does not use long assimilation windows and thus does not constrain the global growth rate as effectively."

- Page 15, line 30: excluding Eureka, Karlsruhe and Izana.
  Response: Like Eureka, Karlsruhe and Izana also have large RMS, but this statement specifically discusses bias.

- Page 16, line 5: The statement "has GEM-MACH-GHG exceptionally good agreement" is not appropriate unless it is specified that the agreement is better than CT in the free troposphere but all the models have similar error magnitude in the boundary layer.
  Response: Agreed. The statement was modified to: "In the annual average, GEM-MACH-GHG has good agreement with these independent measurements while CT2013B has a very slight positive bias."

- Page 16, line 8: "excellent annual results *in the free troposphere* because . . . "
  Response: Agreed. The statement was modified as suggested.

- Page 16, lines 12-13: In autumn the gradient seems to be "too large". In the paper it says it is "too small". Moreover, the explanation that the vertical mixing just above the boundary layer is too weak would not make sense if the gradient was too small.
  Response: Thanks for catching this error. The statement was corrected.

- Page 16, line 15: The departure of the vertical gradient from those observed can only be attributed solely to the model formulation if the two posterior fluxes used are significantly different. In the paper there is no evidence that this is the case for the profiles used over Canada.

Response: This point is valid. Additionally, for reasons already mentioned, all references to GEM-MACH-GHG simulations with the second posterior flux were removed. So, this statement has been removed.

- Page 16, lines 17-18: On a season by season basis, I don't think it is possible to say that "Overall, compare to Carbon Tracker, vertical gradients in GEM-MACHGHG agree better with independent measurements in the mid troposphere but less well in the lower troposphere." This is only the case for the annual mean, but as mentioned in the paper the annual mean does not reflect the errors associated with the transport, but it reflects that fact that the errors have opposite sign in different seasons. So I would remove this statement.
  Response: Done.

- Page 16, line 23: what about the errors associated with the unresolved (parameterised) transport?
  Response: Transport here refers to the model's determinist transport including parameterized subgrid scale processes which are represented deterministically.

- Page 16, lines 25-26: The assumption of the fluxes being "perfectly known" is unrealistic.
  Response: We changed the sentence: "In a data assimilation or flux inversion system, when the fluxes are well constrained by observations, the ability to estimate the CO2 fields using observations will ultimately be limited by the loss of meteorological predictability, just as the quality of weather forecasting products are."

- Page 17, line 5: The synoptic variability of the biogenic CO2 fluxes can also have an impact on the synoptic variability of atmospheric CO2 (Chan et al. 2004 in Tellus B, Agusti-Panareda et al. 2014 in ACP).
  Response: Good point. The statement was modified to "The predictability problem on weather timescales is related to forecast sensitivity to initial conditions, but the atmospheric variability of CO_2 on diurnal (Law et al., 2008), synoptic (Chan et al. 2004, Agusti-Panareda et al. 2014) and seasonal and interannual (Gurney et al., 2002, 2004; Baker et al. 2006a, LeQuéré et al., 2015) time scales is largely governed by the terrestrial biospheric fluxes and hence is determined by sensitivity to boundary rather than initial conditions."

- Page 17, lines 8-10: It is important to also mention that the predicatibility error should include the random error as well as the bias. This paper addresses mainly the limits of predictability in terms of variance but often the more challenging problem for models using NWP assimilation windows is how to deal with the large-scale growing biases in the background air.
  Response: The bias is most often revealed when comparing forecasts to observations. In this case, forecast or transport model error is affected by flux errors convolved with meteorological state errors, initial CO_2 state errors and model formulation errors. Additionally, observation and representativeness errors are present in this comparison, but the most likely source of the bias is the convolution of flux and atmospheric transport. In our work, with a reference simulation, model formulation errors are absent. Flux and initial CO_2 state errors are also absent when we focus on the contribution of meteorological state errors to forecast (transport model) errors. Thus as shown in Figure R1-1, bias is not important in our study, but the Reviewer is correct that in the usual transport error assessments, it is an important concern. This point is made in a new paragraph in the revised discussion section.

- Page 18: The number of days where the forecast is deemed to have skill is just an empirical measure that is useful in the context of comparing different parameters and different factors (as mentioned in the discussion section). I would emphasize that the numbers of 2-3 days and 5 days

should not be taken as a theoretical limit (see line 27 in the paper). The reason why the predictability seems to be longer in the boundary layer could explained by the normalization of the predictability score, as well as the stronger influence from the prescibed fluxes. The normalisation of the random error will imply that the layers with larger zonal variability (e.g. near surface or in upper troposphere/lower stratosphere with well-defined zonally propagating waves) will have lower values of the predictability diagnostic. It is also worth noting that the specific predictability diagnostic used in this paper focuses on the ability of the model to simulate the expected zonal variability. However, there could be many other measures focusing on other aspects of the predicatibility (eg.anomaly correlations as done by Massart et al, 2014, or mean error growth) that are not shown in this study. These should be made clarified in the paper.
Response: Agreed. We tried to caution against viewing the numbers of 2-3 or 5 days as the result given their dependence on some arbitrary parameter choices made, in general. But we agree that the statement (p. 19, line 27) should be qualified. We also agree that the normalization may be playing a role in the relative predictability between layers, and this is now mentioned in the revised manuscript. In addition, the anomaly correlation results of Massart et al. (2015) are now discussed in the revised discussion section.

- Page 20, line 3: Which surface fluxes?
  Response: Those from the terrestrial biosphere. The text has been corrected.

- Page 20, lines 20-21: It is not clear how the spectra shown in Fig 11 is computed, specially the specification of the total and the zonal wave numbers shown in Fig 12 and 13.
  Response: As our model grid is defined on Gaussian latitudes, fields can be decomposed level by level into spherical harmonics following Boer (1983). Then using a triangular truncation and summing over either total or zonal wavenumber, the spectra are obtained as function of the remaining wavenumbers.

- Page 20, lines 27-28: Does this mean that the spatial scales of the predictability errors are only assessed for a 1 to 2 month simulation?
  Response: No, the simulation was for 1 year, the statistics were averaged over 1 month. By averaging the statistics, some of the noisiness of the spectra (seen in animations of daily spectra) is removed. One month was found sufficient to smooth the spectra yet allow for some seasonal variation to be seen.

- Page 21: Although the emphasis is on the predictability of the meterological parameters affecting the predictability of CO2, the influence of the CO2 fluxes in a real forecast setting (i.e. where the fluxes cannot be retrieved) should be emphasized. It seems as if they just play a secondary role as they are just mentioned as a remaining factor after all the others have been considered. A lot of emphasis is given to the role of the ocean and land surface conditions, but from Fig S7 I would say that there is still predictabtility at large scales even when those surface conditions are shifted by 3 months. I think this proves that at large-scales the fluxes are really the dominant factor for the predictability.
  Response: We agree. There was no intention to downplay the important role of CO_2 fluxes in predictability at large scales. The fact that the land and ocean surface play any role at all is rather interesting and novel (we believe). This is what we were trying to point out. However, we have modified the text (sections 5.2 and 6) to emphasize that we also demonstrate the dominant role of CO_2 fluxes on predictability at large scales.

- Page 22, lines 1-2: I do not understand why the results are consistent with the transport biases acting at large-scales. Isn't the limit of predictability associated with imperfect analysis most relevant for small scales?
  Response: Yes. It is the small scales that are the least predictable. But these unpredictable scales get surprisingly large as altitude increases. Chevallier et al. (2010) shows that small transport errors can produce flux biases of large spatial scale for GOSAT OSSEs. However, the connection is not that strong, so this statement was dropped.

- Page 22, lines 13-16: I would say that the larger influence of the flux differences near the surface is not because the retrieved fluxes assimilated observerations near the surface, but because any surface flux will have a much larger influence on the CO_2 near the surface than at upper levels. The diminishing impact of flux differences with height has nothing to do with whether the observations assimilated were near the surface, in the mid troposphere or for the total column, but on the fact that the influence of any surface flux will always diminish with height because of the transport and mixing away from the source/sink region.
  Response: Quite true. However, this paragraph was removed in order to remove the distraction of having used two different but comparable sets of retrieved fluxes.

- Page 22, lines 28-29: This sentence is not clear.
  Response: The sentence was modified to: "*However, since predictability was already close to lost even with the correct land and ocean surface (power in predictability error is same order of magnitude as that for reference state), this conclusion may not be warranted the worsened predictability in July to September with incorrect land and ocean surface fields is not a clear indication of influence of their influence in the tropics.*"

- Page 23, line 9: Can you associate the zonal spectra to the tropics exclusively? What about the Rossby wave in mid-latitudes?
  Response: No, the zonal spectra are not exclusively related to the tropics, but the tropics tend to be better represented in zonal spectra. These diagnostics can only be taken as indications of tropical predictability, since the measure is not specific to the tropics. However, the statement concerned the experiment with two posterior fluxes so it was removed.

- Page 23, lines 24-25: I think it would be useful to add the range of errors found on the seasonal time-scale when you say that the model compares well with observations(e.g. within +-2ppm based on Fig 8).
  Response: Good idea. The statement was modified.

- Page 23, line 27: The statement that the gradient in the model is excellent can be misleading as it only applies to the annual mean. Since the gradient for specific seasons is what really reflects the transport uncertainties, I would say that the gradient is slightly overestimated in the free troposphere probably due to a lack of mixing in the model.
  Response: "excellent" was changed to "slightly overestimated"

- Page 23, line 31: This is not the first time when the predictability of CO_2 has been investigated (see Massart et al. 2016, ACP). The results from Massart et al. (2016) show the forecast of column-averaged CO_2 has skill up to day 5, so the 2-3 day is not a theoretical limit, but it is dependend on the diagnostic used. This should be clarified somehow.
  Response: Agreed. A discussion of the results obtained by Massart et al. (2016) was added to the completely revised new discussion section.

- Page 24, line 5: The climate predictability is important, but so is the predictability of the CO2 fluxes.
  Response: Agreed. This whole section was rewritten. In doing so, the importance of the CO_2 fluxes on predictability results is better emphasized.

- Page 24, line 6: If I understood well, the impact of imperfect analysis is tested by shifting the analysis field by 6 hours. This is not equivalent to "a 6 h forecast error" as mentioned in the paper.
  Response: True. We have taken the 6h shift in analyses as a proxy for 6 h forecast error, but it should not be identified as such. This statement was clarified to avoid making the equivalence. Additionally, we have reviewed the entire manuscript and eliminated other instances where 6h analysis error is mentioned when referring to the proxy.

MINOR COMMENTS
- Page 14, lines 1-2: It is difficult to see the blue line associated with the model in the middle panel.
  Response: We agree. We had tried to reorder the colours but one of the curves will always be hard to see. As it is, this panel shows that the seasonal variation of the diurnal cycle amplitude is captured well. Where the blue curve is difficult to see is in the summer, and a better way to assess this is by computing and comparing the diurnal cycle amplitude to observations. This is done in Figure 4. We did however modify this plot to reduce the obscuring of the curves by the legend in the bottom panel.

- Figure 1: I think the sentences "..there should be 4 times as many blue boxes as shown. Some of these were omitted for clarify of presentation" could be misleading if the meteo analysis in the CO2 model is only used at 00 UTC. If not, this should be clarified in the paper.
  Response: The meteorological analyses were only used at 00 UTC although the CMC produces them every 6 h. These sentences were removed from the caption.

- Figures 3: It is difficult to see the blue line in the middle panel.
  Response: This is the same point mentioned two bullets ago.

- Page 15, line 1: replace "dynamics" by "transport".
  Response: Done.

Response to Review 2

The impact of meteorological analysis uncertainties on the spatial scales resolvable in CO2 model simulations" by Saroja M. Polavarapu et al.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-346, 2016.

Original comments are in black text. Our responses are in blue text.

Unrequested modification: In addition to changes prompted by reviewers' questions, we decided to change the predictability metric from the square root of the global mean of zonal variance to a more intuitive one: global mean of zonal standard deviation. In doing so, a coding error in the original diagnostic was found that affected only vorticity and divergence: abnormally large normalized predictability error was produced in the stratosphere and mesosphere. Since the square root and global mean operators do not commute, the change in Figures 9-10 resulting from the change in diagnostic did not have to be small, but were in fact minimal. Thus the changes seen in the revised Figs. 9-10 are mainly from correcting the coding error.

We thank the Reviewer for the comments. We believe that the revised manuscript has greatly benefitted from the comments of both Reviewers.

The paper includes two parts. It first describes a new coupled meteorological and tracer transport model based on Environment and Climate Change Canada's operational weather and prediction model, and then discusses the predictability of CO_2 due to initial state sensitivity and land and ocean surface states. While the paper devotes most of the space to describe the steps adapting the GEM-MACC to do tracer transport, both the title and abstract do not reflect such effort. I would recommend dividing the current paper into two papers: one is on model development that is more suitable for Geophysical Model Development (GMD), and the other is on CO_2 predictability, which may be suitable for ACP. At the current stage, the paper reads more like a model development paper.

Response: Before starting this manuscript we had considered writing a separate paper just documenting the model. However, we felt that although it was an arduous, painstaking task to adapt GEM-MACH to simulate CO_2, this work alone would not justify publication. This is because of the intrinsic important of fluxes to CO_2 simulations. Since we do not yet estimate CO_2 fluxes, only the transport component of simulations can be assessed, and only to a limited degree since using retrieved fluxes from another model means that model's transport errors are convolved with ours. This is an important point that is demonstrated here. It makes sense to document a system which is semi-operational, near-real time, or which estimates fluxes. On the other hand, using a coupled meteorology and transport model has some advantages that we explore here, namely, the spatial scales on which meteorological uncertainties impact transport error. These are highly original and significant results relevant to the understanding of CO_2 assimilation results. This is what is described in the main portion of the paper. However, this being the first instance of the use of this model for greenhouse gas simulation, it was necessary to document the model components and show a few results to assure the reader that predictability results, which are necessarily model dependent, and shown in section 4 are generally relevant. However, we agree with the Reviewer that this was not well indicated by the title and abstract of the original manuscript. Accordingly, we have modified the title in the revised manuscript to: *Greenhouse gas simulations with a coupled meteorological and transport model: The predictability of carbon dioxide*. Moreover, the abstract, Introduction, and discussion have been revised to obtain a more coherent picture focused on the coupling of CO_2 and weather and climate. A new section was also added to connect the ideas of predictability and transport model error. Since the aspect of transport model error that is difficult to

address with an offline model is the error due to uncertain meteorology, the use of a coupled meteorological and transport model is critical for this work.

The following are my detailed comments:

1) The term of mass-conservation is loosely used. In CO_2 transport, what needs to be conserved should be the total number of molecules in the atmosphere, not mixing ratio.

   Response: The mass budget of CO_2 has been a primary objective in the development of our modelling system. We spent a significant amount of time to make changes to the NWP model to ensure the conservation of the total amount of CO_2 molecules during the full length of the simulation. In section 2 we describe changes made to the model to express CO_2 in terms of dry mixing ratio and the effort made to conserve dry air in the model. Nevertheless we have revised the text to avoid any confusion. In section 2.2 the text has been modified to explicitly mention how we compute global mass from mixing ratio. In section 2.3, the line mentioning the conservation of mixing ratio has been removed. That statement had existed only to explain why we do not use horizontal diffusion, but it was removed to avoid any possible confusion.

2) The reason of using 24-hour tracer-transport forecast cycle is not well illustrated in the text. Though spurious gravity wave could be generated during forecast, NWP has been using filtering technique to reduce spurious gravity waves (Nezlin et al., 2009). In the off-line tracer transport model, the meteorology analysis fields are read in every 6 hours, while this paper uses 24-hour. The paper only compares the 24-hour forecast to other reanalysis products. The accuracy of 24-hour meteorology forecast compared to the analysis fields read in by GEM-MACC needs to be assessed. Also, the sensitivity of CO_2 transport to the length of tracer-transport forecast-cycle needs to be quantified, which can be addressed by changing the forecast cycle to 6 hours. Nezlin, Y., S. Polavarapu, and Y. J. Rochon (2009), A new method of assessing filtering schemes in data assimilation systems, Q. J. R. Meteorol. Soc., 135, 1059–1070.

   Response: Our experiments have shown that the quality of the CO_2 simulations does not necessarily improve with more frequent cycling. Figure R-2.1 below (not shown in the manuscript) indicates indeed that the spatial distribution of CO_2 is not sensitive to the cycling frequency. By doing 24 hr cycling we reduce significantly the amount of CPU without sacrificing anything in terms of quality of results. The actual impact on the CO_2 evolution was not large in terms of XCO_2 (Fig. R2-1). The differences seem random in both space and time (from animations of figures like Fig. R2-1). However, differences do accumulate in the stratosphere and mesosphere presumably due to different wave generation characteristics which impact the slow overturning Brewer-Dobson circulation. The 24h update yielded better (slower) transport in the middle atmosphere (for reasons not yet known). Although the mass of CO_2 in the stratosphere and mesosphere is not important, this difference in circulation meant that the 24h update cycle was preferred. As for filtering techniques, this was a concern when developing our system's configuration. However, we found that it had very little impact with a 24h cycle and was more expensive to run.
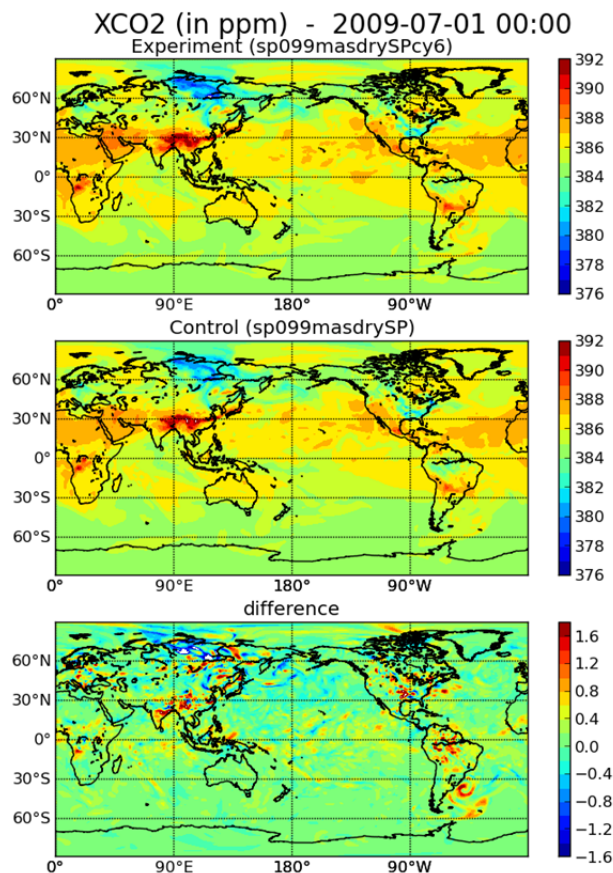
XCO2 (in ppm) - 2009-07-01 00:00

Fig. R2-1: The CO_2 state on July 1, 2009 00Z from 2 model simulations which are identical except for a 6h update cycle (top) and a 24h update cycle (middle) and differences between the top and middle panels (bottom). Difference patterns change daily but remain on fine spatial scales.

For this work, it is important to demonstrate that our 24h forecasts are within the uncertainty of reanalyses. As the base model is an operational model, its forecasts are routinely verified against measurements and compared with other operational centres (as noted in the original manuscript). But what is more relevant to the carbon cycle community, and to this paper, is to demonstrate that stitching together a sequence of 24h forecasts has relevance to a time series of reanalyses since the latter can be used to constrain flux inversions. Therefore we could compare our analysis errors to that of reanalyses. However, we went much further than that and demonstrated that the entire forecast (up to 24h) is as close to a reanalysis as are any two reanalyses datasets. The point is that reanalyses are not perfect and the difference in reanalysis datasets is a measure of their uncertainty (and is commonly used as such in the climate science community). Thus our full time series of disjoint 24h forecasts has no more uncertainty than the time series of reanalyses. This point is now explained in the last paragraph of section 4.1 of the revised manuscript.

3) Figure 5 and 6 qualitatively compare the GEM CO_2 fields to CarbonTracker. A figure showing the difference between GEM CO_2 and CarbonTracker will be more quantitative. Also, it would be helpful to include a column CO_2 north-south gradient comparison. Figure 5 indicates that CarbonTrakcer and GEM may have quite different N-S gradient.

`

Response: Figures 5 and 6 have been replaced so that now one panel depicts the difference between GEM and CarbonTracker. The text has been correspondingly adjusted.

4) The explanations for the disagreement between obs and aircraft and some surface insitu observations are not convincing. Figure 3 a and figure 8 b and c show that the summer draw down at lower levels simulated by the GEM model is much stronger than observations. The authors attribute this to the accuracy of underlying fluxes (P15, L33), since the model simulated CO_2 agrees better with the observations when using the posterior fluxes constrained by GOSAT. While the accuracy of underlying fluxes could be the reason, I think it is more likely due to the accuracy of convective transport because the aircraft observations are over NA where surface flask observations are dense. On the contrary, both figures seem to indicate that the vertical mixing at lower levels is too weak during both summer and winter. I suggest the authors using MACC III that is constrained by surface flask observations to do one more CO_2 forward simulations, and then compare GEM CO_2 to MACC III CO_2 fields.

Response: In Fig. 3a the disagreement between the model and Alert observations was attributed to fluxes because different retrieved fluxes from GEOS-Chem (using GOSAT) observations did not have this mismatch. Thus, we can conclude a mismatch in long range transport between CarbonTracker and GEM-MACH-GHG. However, as noted by the Reviewer, the problem with the GEOS-Chem flux integration is that the observing system also changed along with the source of fluxes. While using MACC III posterior fluxes constrained by surface observations would help settle the issue, we also now have access to a similar dataset, namely, GEOS-Chem posterior fluxes constrained by only surface observations, as well as other new information. This additional information supports our original point of view. Firstly, a GEM-MACH-GHG simulation with GEOS-Chem fluxes constrained only by surface observations also agrees better with observations than when CarbonTracker fluxes are used (Fig. R2-2). Secondly, GEM-MACH-GHG constrained with CarbonTracker-CH4 fluxes also agrees well with measurements at Alert. Since the transport is identical for both CO_2 and CH_4 (both were carried in the same model) yet only CO_2 has a disagreement with surface observations at high latitudes (Fig. R2-3), the CarbonTracker CO_2 fluxes are clearly implicated. That is, the CarbonTracker transport errors do not match GEM-MACH-GHG transport errors in high latitudes in the autumn. With this additional information, we can now safely attribute the mismatch at high latitudes to the fluxes and hence to the mismatch of GEM and CarbonTracker transport at high latitudes. Figure S6 of the supplemental material was modified to include Fig. R2-2 and text was correspondingly modified. However, we did not add Fig. R2-3 to the supplemental material because our CH_4 simulations were not described in this article. This will be the topic of a forthcoming article (not yet in preparation). As for convective transport, our updraft velocities compare well to observed values, but we do not yet have tracer transport through the shallow convection scheme, and this could explain the overestimated vertical gradient seen in Figure 8. The shallow convection scheme is now described in section 3.5 and this point is now mentioned in connection with Fig. 8.
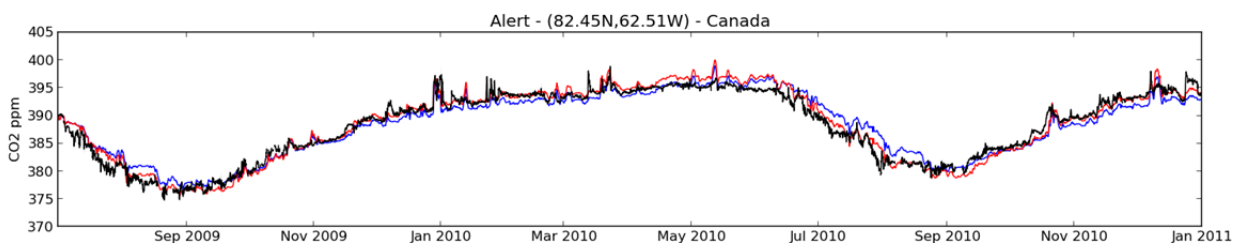

Figure R2-2: Time series of CO_2 at Alert. GEM-MACH-GHG simulations with GEOS-Chem posterior fluxes obtained with GOSAT (red) and with surface observations only (blue) are compared to

measurements (black) for July 2009-Dec 2010. Note the agreement with observations in autumn for both model simulations.
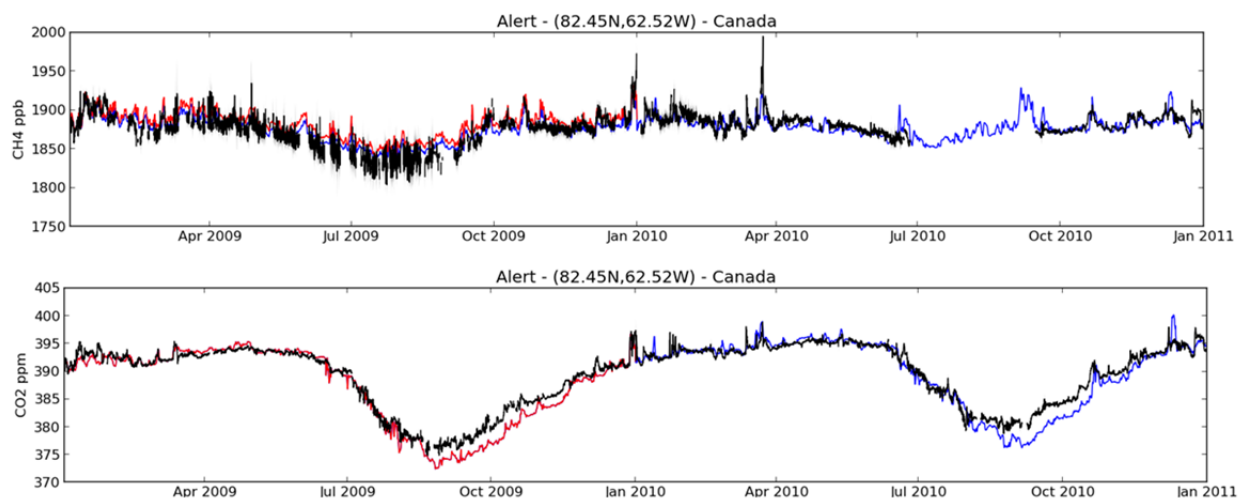

Figure R2-3: Time series of CH_4 (top) and CO_2 (bottom) at Alert. A recent GEM-MACH-GHG run with both CO_2 and CH_4 using posterior fluxes from CT2013B (for CO_2) and CarbonTracker-CH4 (for CH_4) (blue) is compared to measurements (black). Note that there is no disagreement in autumn for CH_4 as there is for CO_2 with the same model transport.

5) In adapting the model to do CO_2 transport, the authors do not include the horizontal diffusion (section 2.3). What is the impact of horizontal diffusion on CO_2 fields? The authors at least can run an experiment including horizontal diffusion for few day and then compare to the control run. Within a few days, the mass conservation is not that critical.

Response: During the model development (a couple of years ago), we did in fact look at the impact of horizontal diffusion. Not surprisingly, the CO_2 field is slightly smoother (the diffusion uses a del-6 operator) but there is an impact on global mass conservation on the time scale of one year. As noted in the original manuscript (p.8, lines 13-14), given our desire for exact mass conservation, horizontal diffusion cannot be included. Nevertheless, for interest sake, we show here a 2-year-old plot of spectra from a model run with the standard high order (del-6 operator) (red), a comparable run with no horizontal diffusion (black) and another with greatly enhanced diffusion (del-4 operator) (blue). The impact of horizontal diffusion is on wavenumbers 100 and higher. Similar results are seen for other vertical levels and other dates.
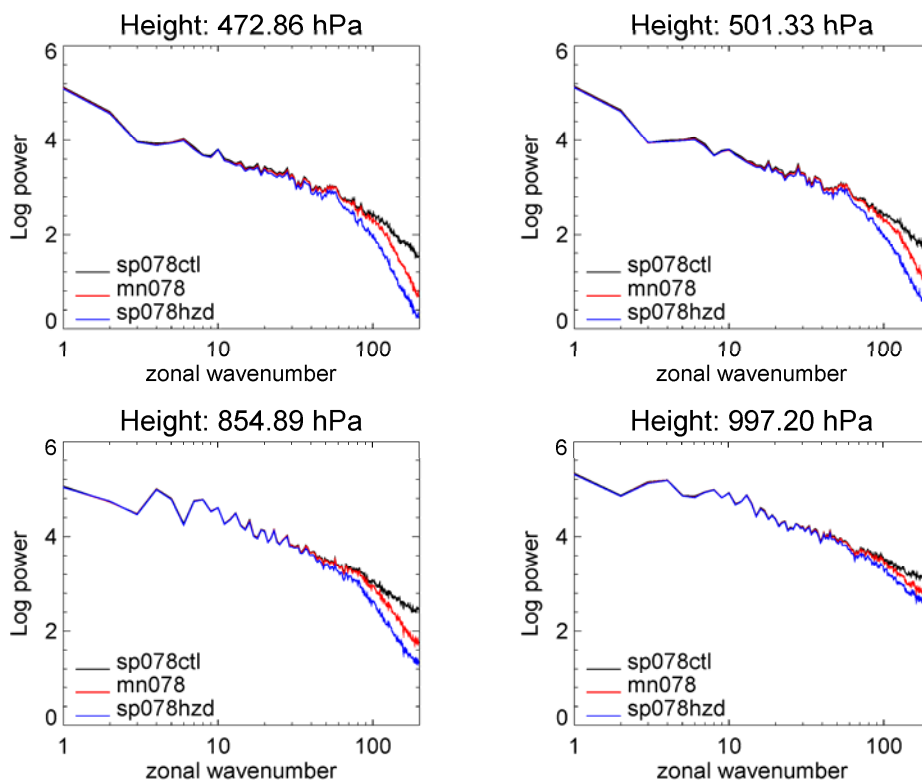
Figure R2-4: Spectra of CO_2 as a function of zonal wavenumber for various model levels whose approximate pressure is given above each panel. The black curves correspond to the control cycle while the red curves add horizontal diffusion with high order (del-6) filtering. The blue curves use an even stronger, less scale-selective filtering (del-4).

6) In the second part of the paper, the authors try to quantify the CO_2 predictability in weather time scales and seasonal time scales. The discussions are lack of physical interpretation of the results and the implication of the results for flux inversion. Unlike meteorology fields and other air quality variables, such as O3, CO_2 transport and CO_2 observations are mainly used to quantify surface fluxes, not for prediction; the CO2 prediction itself does not have any applications. Under this context, it is important to illustrate the connection between CO_2 predictability discussed here and the CO_2 flux inversions.

Response: We agree with the reviewer that it is important/necessary to identify the physical mechanisms which determine the CO_2 predictability. Indeed, a focus of this work is on the processes connecting meteorological and CO_2 predictability. For this reason there is an attempt to address it in section 4 where we discuss the impact of convective mixing on the error spectra. Aside from the consideration of spatial scales, atmospheric processes from stratospheric sudden warmings, to tropical modes of predictability, to atmospheric waves have been invoked in the discussion of results. It is exactly this type of analysis which is new to carbon cycle science and which is one of the novel features of this work.

A relevant matter for the flux inversion community is the use of transport model predictions (i.e. forecasts) in the computation of model-data mismatches. Transport model predictions are also performed with the fluxes retrieved from such an inversion to obtain CO_2 state estimates. The issue of transport model error (hence predictability) is of considerable concern to the flux inversion community. However, there may be a mismatch in terminology used by the weather prediction data assimilation community versus that used by the flux inversion community. To address this, a new section (section 2) was added to the revised manuscript to connect the concepts of predictability and transport error and to identify the components of transport error that are discussed in the article.

We recognize that the predictability of the CO_2 state has important implications for flux inversions. The new section connecting the concepts of predictability with transport error combined with the completely rewritten discussion section (the new section 6) address this issue. The abstract was also rewritten with this point in mind.

7) The authors simulate the analysis errors by shifting the met analysis by 6 hours. This would create large errors due to inaccurate description of diurnal cycle, which is especially significant over extratropics. Since ECCC has a hybrid approach to simulate background error covariance (page 10, line 24), the analysis error can be approximated by the spread of ensemble forecasts used in the hybrid scheme, which should be more realistic.

Response: The reviewer is correct that errors in the diurnal cycle would be created by shifting analyses by 6 h. We had considered this issue but decided it was not critical because the shifting is done everywhere at a given universal time, rather than local time. Thus the shift will be at different parts of the diurnal cycle for different longitudes, so the bias introduced by the shift averages out for the meteorological fields. It was assumed that even with the covariation of the wind fields with the CO_2 flux over 24 h, the impact when averaged over longitudes would again average out given that we only considered global diagnostics. However, the only way to know for sure if this is true, is to remove the diurnal cycle from the 6h analysis differences and rerun the perturbed analysis cycle. Therefore, we have done this. It is straightforward (though time consuming) to do this by simply removing the monthly mean of each hour from the perturbation. Figure R2-5 shows the impact of removing the diurnal cycle from 6h analysis differences on spatial spectra. The Figure corresponds to Figures 11-12 of the original manuscript with the red curves depicting the CO_2 spectra for differences in CO_2 due to 6h analysis differences, whereas the cyan curve depicts the impact from 6h analysis difference with the diurnal cycle removed. The month chosen is July 2009 because this is the month for which the impact of the diurnal cycle is the largest. In Fig. R2-5 (top row) there is an impact on the largest scales, with much reduced analysis uncertainty when the diurnal cycle is removed. The impact is largest near the surface where the boundary layer diurnal variation is important and in the stratosphere and above where tidal signals are important. However, the conclusion regarding the lack of predictability for smaller spatial scales (where the red or cyan curves cross the reference spectra (black or blue)) remains the same. Additionally, the conclusion regarding the lack of information at small spatial scales (the asymptoting of the red or cyan curves to the black one) remains the same. Because there is an impact from removing the diurnal cycle on predictable spatial scales, Figures 11-13 have been updated in the revised manuscript by replacing the red curve with the one corresponding to the new cycle with analysis perturbations having the diurnal cycle removed.
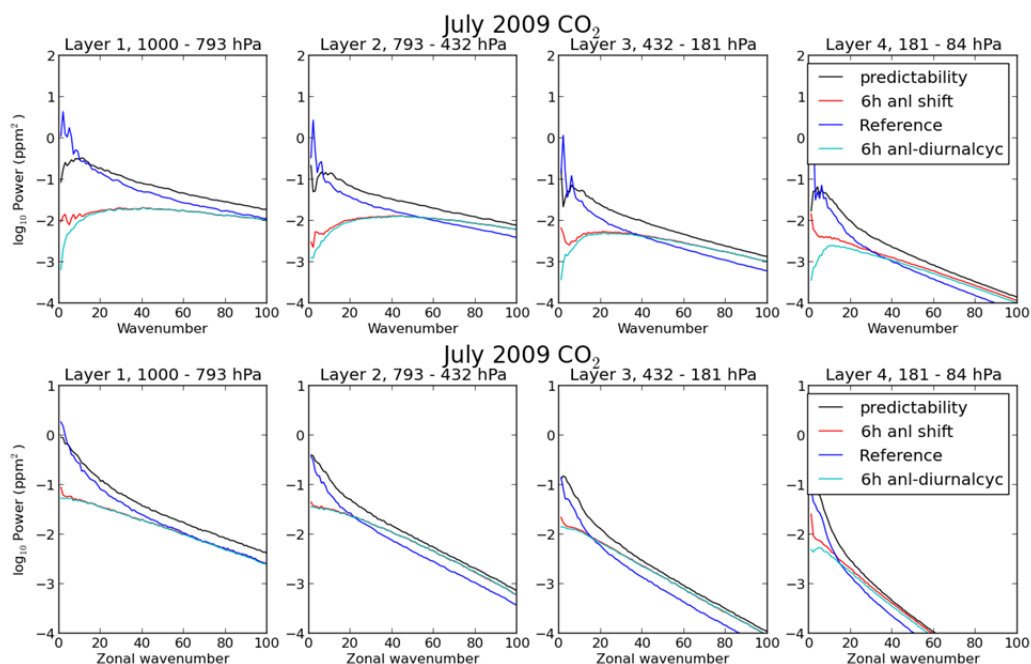
Figure R2-5: Spectra of CO_2 as a function of total (top row) or zonal (bottom row) wavenumber for July 2009 various model levels whose approximate pressure is given above each panel. The black curves represent predictability error while the red curves represent error due to a 6h shift in analysis. The red curves should be compared to the cyan curves which result when the diurnal cycle is removed from the 6h shift.

As stated in the original manuscript (p20, lines 9-11 and p25, lines 8-10), the ideal perturbation is a realization of analysis error, but dealing with operational systems poses significant logistic challenges. It is true that ECCC has a hybrid approach to obtaining background error covariances as of Nov. 18, 2014 but the hybrid system relies on a separate ensemble Kalman Filter system to provide realizations of forecast errors. It does not use nor does it compute analysis errors. However, ECCC also has a separate EnKF which is used for ensemble prediction along with the operational deterministic 4D-Var (i.e hybrid) analysis system. We had naturally considered the possibility of getting realizations of analysis error from the EnKF (though it uses lower resolution, a different model version and a lower lid from the deterministic system which produced the meterological analyses used here) but these were not archived by CMC in 2009-2010. The archives only start around 2011, and are not reliable so a full year of members could not be extracted. As noted in the original manuscript (p25, lines 11-16), we are devoting significant time right now to developing an ensemble Kalman Filter for greenhouse gas which, when ready, will give us exactly what we need. In the meantime, our proxy for analysis error (the 6h analysis difference minus the diurnal cycle) is useful (even though this proxy will be larger than analysis errors) since none of our main conclusions are affected by this choice, namely, that (1) some scales in CO_2 cannot be resolved due to the presence of analysis uncertainties, and (2) that for small enough spatial scales, the error spectrum due to analysis uncertainties asymptotes to the predictability spectrum hold regardless of the size of the error. Additionally, in the revised manuscript we stress that the proxy should over estimate analysis error.

~~The impact of meteorological analysis uncertainties on the spatial scales resolvable in CO$_2$ model simulations~~<u>Greenhouse gas simulations with a coupled meteorological and transport model: The predictability of CO$_2$</u>

5 Saroja M. Polavarapu[1], Michael Neish[1], Monique Tanguay[2], Claude Girard[2], Jean de Grandpré[3], Kirill Semeniuk[3], Sylvie Gravel[3], Shuzhan Ren[4], Sébastien Roche[5], Douglas Chan[1] and Kimberly Strong[5]

[1]Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, M3H 5T4, Canada
[2]Meteorological Research Division, Environment and Climate Change Canada, Dorval, Québec, H9P 1J3, Canada
[3]Air Quality Research Division, Environment and Climate Change Canada, Dorval, Québec, H9P 1J3, Canada
10 [4]Air Quality Research Division, Environment and Climate Change Canada, Toronto, Ontario, M3H 5T4, Canada
[5]Dept. of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

Correspondence to: Saroja Polavarapu (Saroja.polavarapu@canada.ca)

Abstract. A new model for greenhouse gas transport has been developed based on Environment and Climate Change
15 Canada's operational weather and environmental prediction models. When provided with realistic posterior fluxes for CO$_2$, the CO$_2$ simulations compare well to NOAA's CarbonTracker fields, and to near surface continuous measurements, columns from the Total Carbon Column Observing Network (TCCON), and NOAA aircraft profiles. This coupled meteorological and tracer transport model is used to study the ~~atmospheric modulation~~ <u>predictability</u> of CO$_2$ ~~transport~~. ~~The predictability of CO$_2$ due to initial state sensitivity is shorter than that for the temperature field but is consistent with the predictability of the
20 wind fields. However, when broken down into spatial scales, CO$_2$ has predictability at the very largest scales due to long time scale memory in surface CO$_2$ fluxes as well as in land and ocean surface forcing of meteorological fields. The predictability due to the land and ocean surface is most evident in boreal summer when biospheric uptake produces large spatial gradients in the CO$_2$ field. Predictability errors provide an upper limit for errors arising solely from the use of uncertain meteorological analyses. When considering meteorological analysis errors, CO$_2$ can be defined only on large
25 scales. Thus, there is a spatial scale below which information cannot be obtained simply due to the fact that meteorological analyses are imperfect. Compared to the spatial scales resolvable in the context of imperfect atmospheric analyses, the differences between two sets of posterior fluxes are resolvable only for very large scales. Similarly, the impact of convective tracer transport exceeds that due to atmospheric analysis errors for only the largest spatial scales.~~ <u>Predictability concerns the quantification of model forecast errors and thus of transport model errors. CO$_2$ predictions are used to compute model-data
30 mismatches when solving flux inversion problems and the quality of such predictions is a major concern. Here, the loss of meteorological predictability due to uncertain meteorological initial conditions is shown to impact CO$_2$ predictability. The predictability of CO$_2$ is shorter than that of the temperature field and increases near the surface and in the lower stratosphere.</u>

When broken down into spatial scales, CO_2 predictability at the very largest scales is mainly due to surface fluxes but there is also some sensitivity to the land and ocean surface forcing of meteorological fields. The predictability due to the land and ocean surface is most evident in boreal summer when biospheric uptake produces large spatial gradients in the CO_2 field. This is a newly identified source of uncertainty in CO_2 predictions but it is expected to be much smaller than uncertainties in fluxes. However, it serves as an upper limit for the more important source of transport error and loss of predictability which is due to uncertain meteorological analyses. By isolating this component of transport error, it is demonstrated that CO_2 can only be defined on large spatial scales due to the presence of meteorological uncertainty. Thus, there is a spatial scale below which information cannot be obtained simply due to the fact that meteorological analyses are imperfect. These unresolved spatial scales correspond to small scales near the surface but increase with altitude. By isolating other components of transport error, the largest or limiting error can be identified. For example, a model error due to the lack of convective tracer transport was found to impact transport error on the very largest (wavenumbers less than 5) spatial scales. Thus for wavenumbers greater than 5, transport model error due to meteorological analysis uncertainty is more important than the lack of convective tracer transport, for our model.

1 Introduction

Atmospheric observations of CO_2 are important for understanding the global carbon cycle and its response to perturbations from anthropogenic emissions into the atmosphere. The global carbon budget is routinely updated using atmospheric and ocean measurements in conjunction with a careful accounting of anthropogenic emissions (LeQuéré et al., 2015). Since the greatest uncertainty in the quantified exchanges of carbon between the atmosphere, ocean and land reservoirs is associated with the terrestrial biosphere flux (Friedlingstein et al., 2014), this component is determined as the residual between anthropogenic emissions and atmospheric and ocean changes (which are relatively well constrained by measurements). The growth in global atmospheric CO_2 is attributed to anthropogenic emissions but superimposed on this trend are seasonal and interannual variations which are largely attributed to the terrestrial biosphere (Ciais et al., 2013). Specifically, observations of the 13C isotope (which is fractionated during photosynthesis by land plants so its relative concentration is indicative of the terrestrial biosphere) reveal that the interannual and seasonal variations of the global atmospheric CO_2 budget are due to the terrestrial biosphere (Keeling, 2005). At the same time, biospheric uptake is influenced by climate variations (Nemani et al., 2003; Friedlingstein et al., 2006; Zhao et al., 2011). Specifically, the El Niño Southern Oscillation (ENSO) signal and volcanic eruptions can explain 75% of the variations in the terrestrial biospheric uptake (Raupach et al., 2008). Thus, atmospheric observations are important for determining the global terrestrial biospheric flux, while temporal biospheric flux variations explain atmospheric variability of CO_2 concentrations on seasonal and interannual timescales.

Beyond the global budget, an important challenge is to understand how temporal variations in atmospheric sources and sinks reflect the interplay between natural processes and anthropogenic perturbations and also the feedback between the

2

carbon cycle and climate variations. By combining atmospheric observations with atmospheric models, a spatial distribution of fluxes of CO_2 between the Earth's surface and the atmosphere can be obtained through inverse modelling (e.g. Rödenbeck et al., 2003; Patra et al., 2005; Baker et al., 2006a; Peylin et al., 2013; Chevallier et al., 2014). Such "flux inversions" performed on the global domain and incorporating only around 100 or so CO_2 observation stations near the surface are able to constrain the global atmospheric carbon budget, capture interannual and seasonal variations and attribute these to the terrestrial biosphere (Rödenbeck et al., 2003; Peylin et al., 2013). In fact, only a few observation sites representative of background values (far from CO_2 sources and sinks) are needed to constrain the global CO_2 budget because variations in the vicinity of source/sink regions are smoothed out by the time these locations are reached (Bruhwiler et al., 2011; Keeling, 2005). However, with more surface observations, the retrieved flux uncertainties can be reduced and the ability to retrieve smaller spatial scales is improved (Bruhwiler et al., 2011). Moreover, with the desire to understand the interplay between the natural and anthropogenically perturbed processes, observations near source and sink regions become important. Such observations will be influenced by atmospheric variations on diurnal, synoptic, seasonal and interannual time scales. On the diurnal timescale, the variation of CO_2 due to uptake by plants through photosynthesis in sunlit hours is strongly modulated by turbulent transport through the planetary boundary layer (PBL) which also evolves throughout the day. This is the so-called rectifier effect which helps to explain the annual mean north-south gradient of CO2 (Denning, 1995). Specifically, the uptake of CO_2 by plants during the spring growing season occurs when the PBL is generally unstable and deeper while in winter CO_2 increases due to biospheric respiration can build up when the boundary layer is stable and shallow. In addition, poleward heat transport by baroclinic disturbances is stronger in winter preferentially transporting high CO_2 values north relative to the summer, in the northern extratropics (Chan et al., 2008). Synoptic scale systems also influence CO_2 evolution, particularly in northern midlatitudes where advection can explain up to 70% of day-to-day variability (Parazoo et al., 2008). On the interannual time scale, variations in biospheric uptake can be partially attributed to climate variations (Patra et al., 2005) and flux inversion systems are able to attribute interannual variability of the global CO_2 budget to the tropical biosphere (Baker et al., 2006a). Because atmospheric observations of CO_2 contain the signals of both surface fluxes and atmospheric variations, and they are needed for data assimilation in state estimations or flux inversions, it is important to be able to accurately characterize and model how the atmosphere modulates CO_2 evolution.

In a flux inversion system, observations of CO_2 concentration are used to solve for surface fluxes. To relate the surface flux to an atmospheric concentration, an offline atmospheric transport model is typically used. In doing so, the mismatch between observed and modelled concentrations can be inverted to estimate a flux increment, but the atmospheric model's transport is assumed to be perfect (Baker et al., 2006b). The fact that it is not, means that one source of uncertainty in the inverse problem is not accounted for. For this reason, the need to characterize "transport errors" has been well recognized and has led to the formation in 1993 of an international group called TransCom (http://transcom.project.asu.edu/) focused on understanding and quantifying the contribution of transport errors to flux estimates from inverse models. Because different models are used by different flux inversions, retrieved fluxes based on multiple inversions may be more reliable than those from a given system since the unknown transport error (if it is random) may be averaged out. Moreover,

3

the sensitivity of flux inversion results to transport error can be identified since different models and hence different transport errors are used. Transport error generally refers to the deviation of CO_2 prediction from an (unknowable) true value due to the use of an imperfect transport model and thus includes errors due to model formulation (e.g. convective or boundary layer parameterizations or advection schemes), the use of imperfect meteorological analyses, and uncertain CO_2

5   initial conditions. Transport errors have been found to be an important source of errors in flux inversions (e.g. Chevallier et al., 2014; Chevallier et al., 2010; Houweling et al., 2010; Law et al., 1996). The sources of transport error arising from model formulation include errors in the representation of: mixing in the planetary boundary layer (Denning et al., 1995), vertical mixing in the free atmosphere (Stephens et al., 2007, Yang et al., 2007), synoptic scale and frontal motions (Parazoo et al., 2008) and convective transport (Ott et al., 2011; Parazoo et al., 2008).

10   The uncertainty in meteorological analyses is also an important source of transport error (Liu et al., 2011). In an effort to address this type of uncertainty which cannot be accounted for in flux inversion systems, NOAA's CarbonTracker attempted to perform an ensemble of inversions using different atmospheric analyses as well as different prior flux sets[1]. With a coupled meteorological and tracer forecast model (as used by Liu et al., 2011), the impact of meteorological uncertainties on tracer transport are more easily identified. Recently, coupled meteorological/tracer forecast models have

15   been developed at operational centers such as the European Centre for Medium Range Weather Forecasting (ECMWF) (Agusti-Panareda et al., 2014) and NASA Goddard's Global Modelling and Assimilation Office (GMAO) (Ott et al., 2015). In these operational systems, short-term predictions of greenhouse gases are produced and satellite observations are assimilated at ECMWF (Massart et al., 2016). Such products may provide useful background or a priori information for satellite retrievals as well as providing initial and boundary conditions for regional flux inversions. Coupled meteorological

20   and CO_2 data assimilation systems also provide useful information on the reliability of correlations between meteorological and CO_2 transport errors (Kang et al., 2011) and on the temporal propagation of the observed signal in the context of transport errors (Kang et al., 2012).

         The goal of this work is to better understand how atmospheric transport is modulating the CO₂ mixing ratio and how it is impacted by the component of transport error that is due to uncertain meteorological statesanalyses. While Liu et

25   al. (2011) have shown that meteorological forecast errors in the presence of CO_2 gradients produce CO_2 transport errors, here we consider , for the first time, the predictability of CO₂ and the spatial scales identifiable in the context of imperfect meteorological analyses. Predictability concerns the study of forecast uncertainty and thus corresponds to the study of transport model errors. While transport errors comprise flux errors, model formulation errors, initial state errors and meteorological state errors, it is the latter that is the focus of this work. Using a coupled meteorological and tracer transport

30   model, we first study the loss of CO_2 predictability due to uncertain meteorological initial conditions on weather (2 weeks) and seasonal time scales in order to obtain an upper limit to forecast errors arising from initial state meteorological analysis

---

[1] This approach was later abandoned because only one of the two sources of atmospheric analyses provided the necessary convective mass fluxes needed for accurate tracer transport. See http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2013B_doc.php#tth_sEc7.2

errors. Then the spatial scales of errors in CO_2 arising from the use of imperfect atmospheric analyses are determined. It is shown that there is a spatial scale below which CO_2 cannot be retrieved, simply due to the presence of meteorological analysis ~~errors~~uncertainty.   The advantage of isolating and comparing different components of transport error for a given model is then demonstrated. The ~~spatial scales resolved from using different posterior flux estimates and the~~ spatial scales of

5   model errors (such as the lack of convective transport of tracers) are ~~then~~ compared to the scales retrievable in the context of imperfect atmospheric analyses.

The article is organized as follows. In section 2, we discuss the concept of predictability for the tracer transport problem and identify the components of transport error while in section 3, we describe a new model for greenhouse gas transport based on an operational weather and environmental prediction model. In section 4~~3~~, this model is assessed in terms

10   of its meteorology (24 h forecasts are compared to reanalyses used in flux inversions) and its CO_2 transport.   The ~~atmospheric modulation of~~ component of CO_2 transport model error due to meteorological initial conditions is ~~present~~studied first in section 5~~4~~, considering both ~~starting with the predictability problem on~~ weather and seasonal timescales ~~and proceeding to the issue of predictability on seasonal timescales~~. This is followed by a study of the spatial scales of transport model error arising from uncertain meteorological analyses. ~~The focus of the analysis is on the spatial~~

15   ~~scales of CO_2 transport errors due to uncertain initial conditions and uncertain meteorological analyses.~~ Results are discussed and summarized in s~~S~~ection 6~~5~~.

2 Predictability and Transport Error

Atmospheric predictability concerns the quantification of forecast uncertainty with a weather or climate prediction model. The limited ability to forecast weather is related to the underlying chaotic dynamics of the atmosphere and is evident in the

20   sensitivity of forecasts to initial conditions. Yet even though the predictability of weather does not extend beyond 2 weeks, components of the climate can be predicted on seasonal and interannual timescales. The sources of predictability on climate timescales are related to atmospheric boundary conditions such as sea surface temperature, soil moisture, snow cover, vegetation and sea ice (Shukla and Kinter 2006). Because CO_2 is transported by the atmosphere, the predictability of the atmosphere is directly relevant to the predictability of CO_2.   However, atmospheric CO_2 concentrations are also strongly

25   influenced by surface sources and sinks. In fact, when concentrations are globally averaged to compute a global CO_2 budget, it is only these fluxes that are relevant since atmospheric transport does not change global mass. But when considering the spatial distribution of CO_2, both surface fluxes and atmosphere transport contribute to the quality of predictions of CO_2. While the forecast of CO_2 is not of public interest as a weather forecast is, the ability to predict CO_2 is needed for solving flux inversion problems. Specifically, a CO_2 prediction is compared to atmospheric observations to retrieve fluxes. Thus

30   knowledge of the accuracy of the CO_2 prediction is needed in the formulation of the model-data mismatch error statistics needed for solving the flux inversion problem (see Baker et al. 2006b). In addition, once the fluxes have been inverted, the

transport model is integrated and compared to measurements to assess the quality of the flux inversion. Thus the quality of the CO_2 prediction is assessed.

Since CO_2 prediction is frequently performed with a transport model, particularly in the context of flux inversions, the prediction or forecast error is called "transport error". There is considerable concern over the nature and quantification of various sources of transport error (see http://transcom.project.asu.edu/). As noted in the Introduction, atmospheric variability on diurnal, seasonal and interannual time scales impact CO_2 concentrations. Thus atmospheric predictability on all of these times scales impact the predictability of CO_2. Indeed, transport error on diurnal time scales due to mixing in the boundary layer (Denning et al., 1995), synoptic time scales (Chan et al. 2004, Parazoo et al. 2008), seasonal time scales (Denning 1995, Chan et al. 2008) and interannual time scales (Baker et al. 2006a) have all been identified as contributing to flux estimation errors.

The error of a CO_2 transport model can be defined as the departure of a CO_2 prediction from the true (but not known) value. Mathematically, this is expressed as $\varepsilon_k^{transport} = \mathbf{c}_{k+1} - \mathbf{c}_{k+1}^{true}$ where the updated concentration at time step k+1 is given by a transport model:

$$\mathbf{c}_{k+1} = \mathbf{T}_{k,k+1}\left(\mathbf{x}_k, \mathbf{c}_k, \mathbf{s}_k\right). \tag{1}$$

where $\mathbf{T}_{k,k+1}$ is the transport model which evolves the meteorological state, $\mathbf{x}_k$, the concentrations, $\mathbf{c}_k$, and the flux, $\mathbf{s}_k$ at time step k to time step $k+1$. This transport error has contributions from various sources which can be identified by considering the hypothetical case of a forecast obtained with perfect knowledge of fluxes, meteorology and initial CO_2 state. Such a forecast would still be imperfect because the transport model is imperfect:

$$\mathbf{c}_{k+1}^{true} = \mathbf{T}_{k,k+1}\left(\mathbf{x}_k^{true}, \mathbf{c}_k^{true}, \mathbf{s}_k^{true}\right) + \varepsilon_k^{mdl}. \tag{2}$$

The last term is the model error which arises due to errors in the model formulation. Various authors have investigated the role of model errors due to boundary layer mixing (Denning et al. 1995), vertical mixing in the free atmosphere (Stephens et al., 2007, Yang et al. 2007), synoptic and frontal motions (Parazoo et al. 2008), convective transport (Ott et al., 2011; Parazoo et al., 2008) and mass conservation schemes (Agusti-Panareda et al. 2016). Note that all but the last of these model errors also impact meteorological forecast uncertainty in the case of a coupled meteorological and tracer transport model. Of

6

course, model error is only one component of transport error. The additional error sources are due to errors in the meteorological state $\left(\boldsymbol{\varepsilon}^{met} = \mathbf{x}_k - \mathbf{x}_k^{true}\right)$, errors in the initial concentration $\left(\boldsymbol{\varepsilon}^c = \mathbf{c}_k - \mathbf{c}_k^{true}\right)$, and flux errors, $\left(\boldsymbol{\varepsilon}^s = \mathbf{s}_k - \mathbf{s}_k^{true}\right)$. Substituting (1) and (2) into the expression for transport error, and using these additional error definitions reveals that transport error is the convolution of all of these error sources. In other words,

$$\boldsymbol{\varepsilon}_k^{transport} = \mathbf{M}_k^{met}\boldsymbol{\varepsilon}_k^{met} + \mathbf{M}_k^c\boldsymbol{\varepsilon}_k^c + \mathbf{M}_k^s\boldsymbol{\varepsilon}_k^s - \boldsymbol{\varepsilon}_k^{mdl} + \text{higher order terms} \quad (3)$$

where

$$\mathbf{M}_k^{met} = \frac{\partial \mathbf{T}_{k,k+1}}{\partial \mathbf{x}}\left(\mathbf{x}_k^{true}, \mathbf{c}_k^{true}, \mathbf{s}_k^{true}\right), \quad \mathbf{M}_k^c = \frac{\partial \mathbf{T}_{k,k+1}}{\partial \mathbf{c}}\left(\mathbf{x}_k^{true}, \mathbf{c}_k^{true}, \mathbf{s}_k^{true}\right), \quad \mathbf{M}_k^s = \frac{\partial \mathbf{T}_{k,k+1}}{\partial \mathbf{s}}\left(\mathbf{x}_k^{true}, \mathbf{c}_k^{true}, \mathbf{s}_k^{true}\right).$$

Thus, the various error subcomponents interact with each other to produce transport error. When an offline transport model is used in a flux inversion, the meteorological state is assumed to be known so the first and last terms in (3) are absent. Since the meteorological state is obtained from a meteorological analysis (usually from a foreign assimilation or reanalysis system) the meteorological state error evolves from an analysis error. It is difficult to account for meteorological state errors on CO_2 transport errors in the flux inversion problem aside from producing an ensemble of inverse results with different meteorological forcings (e.g. from different reanalyses). However, with a coupled meteorological and transport model, Liu et al. (2011) have demonstrated that meteorological uncertainty is an important component of transport error (1.2-3.5 ppm at the surface and 0.8-1.8 ppm in a column mean). Since the goal of flux inversion is to attribute the mismatch of CO_2 predictions and observations to a flux adjustment, a good understanding of the nature and size of all of the components of transport error is needed because only then can a successful attribution of the mismatch to flux be made. While progress has been made on identifying the role of model and flux errors on transport errors, there is less known about the impact of meteorological state errors on CO_2 predictability. With a coupled meteorological and transport model, we are able to consider the impact of meteorological state errors on CO_2 predictability and thus on CO_2 transport error.

~~2~~3 Model Description

In order to simulate greenhouse gas evolution, we can take advantage of the comprehensive forecast models already developed for operational environmental prediction at Environment and Climate Change Canada (ECCC). However, it was

7

necessary to adapt these models for the purpose of tracer transport on multi-annual time scales. Specifically, it was necessary to implement a mass conservation scheme, redesign the tracer variables, modify the vertical mixing in the boundary layer, and add convective tracer transport. The basic modelling tools are described in the next subsection while the adaptations needed for this work are presented in subsection 3~~2~~.2 (mass conservation and tracer variable definitions), 3~~2~~.3 (horizontal diffusion), ~~2~~3.4 (convective transport) and 3~~2~~.5 (boundary layer mixing). The coupled meteorology and tracer transport forecasting system is described in subsection ~~2~~3.6.

3~~2~~.1 The Canadian operational environmental prediction models

For many decades, the Canadian Meteorological Centre (CMC) has been producing operational weather forecasts for public dissemination. Since 24 February 1997, these forecasts have utilized the Global Environmental Multiscale (GEM) model (Côté et al., 1998a,b; Girard et al., 2014). GEM is a gridpoint model which solves the hydrostatic (global domain) or nonhydrostatic (Yeh et al., 2002) (regional domain) primitive equations using a hybrid terrain-following vertical coordinate (Girard et al., 2014). As of Feb. 2013, the grid spacing of the global model is roughly 25 km, originally using a regular lat-lon grid, and since Dec. 2015, a Yin-Yang grid (Qaddouri and Lee, 2011). There are 80 vertical levels spanning the surface to 0.1 hPa. The usual physical processes of radiation (Li and Barker, 2005), boundary layer mixing (Bélair et al., 1999), shallow (Belair et al., 2005) and deep convection (Kain and Fritsch, 1990; Kain, 2004), orographic gravity wave drag (McFarlane, 1987), and nonorographic gravity wave drag (Hines, 1997a,b) are included in all model configurations. The land surface model and assimilation scheme are described in Bélair et al. (2003a,b). More details of the physics package are found in Mailhot et al. (1998).

Operational air quality forecasts have been produced by CMC since 2001 in order to provide real time forecasts of the air quality health index on a limited area domain covering most of North America. As of 18 November 2009, these forecasts have utilized GEM-MACH (Modelling Air quality and CHemistry) (Moran et al., 2010; Robichaud and Ménard, 2014; Makar et al., 2015). GEM-MACH is a version of GEM in which complete tropospheric chemistry (involving over 100 chemical reactions) is modelled online, where "online" refers to the fact that the chemistry module is fully integrated into the meteorological model time step. The operational products involve an analysis of ground-level ozone, fine and coarse particulate matter (PM2.5 and PM10), NO_2 and SO_2 over a limited area domain covering North America (Robichaud and Ménard, 2014). The grid spacing currently used is 10 km horizontally, with 80 vertical levels from the surface to 0.1 hPa. The operational forecasts are driven by time evolving meteorological boundary conditions from the operational regional deterministic prediction system (Fillion et al., 2010, Caron et al., 2015), while the chemical boundary conditions are defined using predetermined seasonally-averaged states. A global version of GEM-MACH is also in development for the purpose of providing boundary conditions for the regional model and a parameterized stratospheric chemistry model (McLinden et al., 2000) is used for UV index forecasting.

Our primary interest is in global greenhouse gas distributions, thus we chose to use the global GEM-MACH configuration. In the future, extensions of our system for regional greenhouse gas simulations could be based on the

operational regional GEM-MACH configuration. By definition, an operational forecasting system is constantly changing. Since our period of interest commences in 2009 with the launch of GOSAT on 23 January and since the upper boundary of GEM was raised from 10 to 0.1 hPa on 22 June 2009 (Charron et al., 2012), this stratospheric model configuration was chosen. However, for computational expediency (especially during the model development phase) the grid spacing was coarsened to 0.9° (400×200 grid points) which is roughly twice the grid spacing used by CMC's global deterministic prediction system (800×600 grid points) in 2009. The model time step is 15 minutes.

Our primary focus in this work is global carbon dioxide simulation, but extensions for other greenhouse gases (methane and CO) are also being developed. For that we employ a simple parameterized climate chemistry involving a single OH reaction for the methane and CO simulations. This chemical module is activated from the GEM-MACH chemical interface which is also used to handle emission inputs. This simplified model version which only includes the treatment of greenhouse gas (chemistry and transport) will be called GEM-MACH-GHG hereafter.

~~3~~2.2 Mass conservation and tracer variable definitions

The dry air mass of the atmosphere is known to be constant, since changes in trace gases are very small (Trenberth 1981, Trenberth and Smith, 2005). However, as with many weather and climate prediction models, GEM does not conserve dry air mass. Specifically, GEM loses 0.1 hPa in global mean surface pressure during a 10-day forecast, which is precisely the same rate as that seen in the ECMWF model (Diamantakis and Flemming, 2014).   This loss is only 0.01% of the global mass and thus there is a negligible impact on medium-range weather forecasts. However, for longer simulations relevant for climate timescales, this error can accumulate. Thus GEM (as does the ECMWF and other models) has a parameter to allow the global mean surface pressure to be conserved by adding a spatially uniform adjustment to each grid cell. The constant is determined by the constraint that the air mass at the end of the dynamics step equal that at the start of the step. For our simulations of tracer transport, it is necessary to use this switch to enforce the conservation of dry air mass.

The advection scheme in GEM uses a semi-Lagrangian approach as it affords longer time steps and the computational efficiency desirable in an operational context. However, semi-Lagrangian schemes are well known to be non-conservative (Williamson, 1990; Staniforth and Côté, 1991). GEM is a grid point model, and the semi-Lagrangian approach involves first determining the upstream positions of the grid cells (using a 3-dimensional trajectory in our case) and then interpolating the value of the advected field at these locations. During this first step, gradients in the wind field will limit the accuracy of advection and during the second step, errors associated with the interpolation of the advected field can deteriorate species mass conservation. By combining a finite volume approach with the semi-Lagrangian scheme, it is possible to devise inherently mass-conserving schemes such as SLICE (Semi-Lagrangian Inherently Conserving and Efficient) (Zerroukat and Allen, 2012). Indeed, this scheme is being investigated for implementation with GEM in the future. In the meantime, an interim solution is needed. The approach taken here, as in Diamantakis and Flemming (2014), is to adopt a global mass fixer. To be clear, the global mass is obtained by summing the individual masses of carbon in each grid box. The mass of carbon in a grid box is determined from the mixing ratio (mass of carbon divided by the mass of dry

9

air) times the mass of dry air in the same box. A global mass fixer scheme computes the global mass of tracer at the beginning and end of the advection step and then distributes the change in global mass among various grid cells. There is no unique way of distributing the mass spatially, so different choices result in different mass fixer schemes. While a few approaches were tried and tested with a few different chemical species (ozone, CH_4 and CO_2), the one which produced the most physically desirable results was that due to Bermejo and Conde (2002). In this scheme, the global mass is distributed according to the smoothness of the field. That is, mass is preferentially adjusted where gradients are larger (and interpolation error is known to be larger). The exact scheme used in GEM is precisely that selected by ECMWF and it is described in Diamantakis and Flemming (2014, section 3.1). Note that for For greenhouse gas transport, the ECMWF forecast model also uses the Bermejo-Conde scheme (Agusti-Panareda et al., 2016). uses a proportional mass fixer instead of the Bermejo-Conde scheme (Agusti-Panareda et al., 2014).

For tracers, another issue with the interpolation step of the semi-Lagrangian advection scheme is the potential creation of spurious subgrid scale structure due to Gibbs effects. To avoid this problem, GEM uses a quasi-monotonic interpolation scheme (Bermejo and Staniforth, 1992) which combines high-order and low-order interpolation schemes to prevent overshoots and undershoots, thereby preventing the formation of spurious extrema. However, just as Flemming and Huijnen (2011) noted, enforcing monotonicity was found to worsen the problem of mass non-conservation because it will tend to diffuse gradients. Therefore, monotonicity was replaced by an iterative locally mass conserving (ILMC) scheme (Sorensen et al., 2013). The idea behind the ILMC scheme is to locally preserve the shape of the field and its gradients by distributing the excess (or deficit of) mass due to spurious extrema in the cubic interpolated field to ever increasing shells around the upstream departure point but in such a way that spurious extrema are avoided. By design, the use of ILMC does not impact our lack of global mass conservation, so it is used for preserving positive definiteness of tracer fields in place of a quasi-monotonic interpolation scheme. The ILMC is performed first, and then the Bermejo and Conde (2002) scheme is applied to the ILMC-corrected field. Details on the implementation of monotonic and mass-conservation schemes in GEM and their impact on species transport are found in de Grandpré et al. (2016).

The primitive equations solved by GEM are naturally written in terms of a moist density or pressure variable. However, when tracers are defined as mixing ratios with respect to moist air, tracers become coupled to the water vapour evolution. Thus one of two options must be taken. Either the tracer variable must be redefined every time the water vapour field is modified in the model, or the tracers must be defined as mixing ratios with respect to dry air (e.g. Neale et al., 2010). The latter option has the advantage that measurements of CO_2 are frequently made in terms of dry mole fraction and are thus more easily related to a mixing ratio with respect to dry air. Since the nominal tracer equation within GEM assumed a moist mixing ratio (MMR), for this work, GEM was modified to permit tracers to be defined as mixing ratios with respect to dry air (DMR). This involved: (1) modifying the global mean surface pressure adjustment to ensure global dry air mass conservation, (2) converting the mass fixer scheme (i.e. the Bermejo-Conde scheme) and the ILMC to deal with DMR, (3) modifying the vertical diffusion equation to handle DMR, and (4) ensuring emissions are correctly inserted into GEM's bottom model layer. Once all of these changes had been made, the tracer mass change in a given time step was found to still

display a coupling to the water vapour change in that same time step. The reason was found to be due to the fact that the continuity equation in GEM does not account for the change in surface pressure due to the change in water vapour which occurs as a result of mass flux at the Earth's surface (i.e. because of precipitation or evaporation). By adding a new adjustment to surface pressure field after the physics step to account for the change in global water vapour mass, and

5   correspondingly redefining the tracer variable using a vertical regridding approach similar to   Jöckel et al. (2001) except for the use of dry pressure coordinates, global mass conservation of CO_2 was obtained during a model forecast. This deficiency in GEM's continuity equation is not unusual as it was also present in NASA GMAO's GEOS5 model (Takacs et al., 2015). The adjustment to the model dynamics done here to account for the modified continuity equation is rather similar to that described in section 2a of Takacs et al. (2015) although our work was done independently. The main difference is that only

10   water vapour mass is considered here whereas cloud liquid water and cloud ice mass are also considered in Takacs et al. (2015) (although the authors note that additional masses due to liquid and ice phases are negligible). As these authors demonstrate, without the correction to the model's continuity equation, dry air mass conservation is not assured. They also note that in the context of an assimilation cycle, further constraints on dry air mass are needed during the analysis and initialization steps. Furthermore, they show that the impact of the errors is evident in the mismatch of global water vapour

15   tendency based on the water vapour field itself versus that based on precipitation minus evaporation changes for reanalysis products such as MERRA (Rienecker et al., 2011) and ERA-Interim (Dee et al., 2011).

3~~2~~.3 Horizontal diffusion

It is standard to apply an explicit high order (typically ∇^6) diffusion operator to the meteorological fields in applications involving the global GEM model. It is arguably consistent therefore to apply the same level of diffusion to the tracer fields.

20   However, because the constituent variables are defined as mixing ratios, the operator is not mass conservative. ~~Rather it will conserve mixing ratio.~~ Because of the effort involved in obtaining a careful accounting of tracer mass (described in section 3~~2~~.2), it was decided to not apply any horizontal diffusion to the tracers.

3~~2~~.4 Convective Transport

The parameterization of deep convection most frequently used by GEM and GEM-MACH is due to Kain and Fritsch (1990).

25   The Kain and Fritsch (KF) scheme is based on a single column bulk mass flux approach. Entrainment of ambient air into the cloud environment associated with updrafts and downdrafts is proportional to the corresponding mass flux and inversely proportional to the cloud radius. So narrower convective towers experience more entrainment of lower-buoyancy ambient air and consequently have less intensity and lower cloud tops. The KF scheme is used in several forecast models (e.g. Japanese Meteorological Agency model, Saito, 2012; Bologna Limited-Area Model (BOLAM), Lagouvardos et al., 2003;

30   High-Resolution Limited-Area Model (HIRLAM), Eerola, 2013; Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF-NMM), Gallus Jr. and Bresch, 2006). With the Kain-Fritsch scheme, GEM has a good representation of convectively coupled waves and is able to capture the Madden-Julian oscillation (Lin et al., 2008).

11

The version of the Kain-Fritsch scheme used for global deterministic prediction, which was our starting point, had to be modified for the purpose of greenhouse gas transport. The original parameters resulted in too frequent penetration of convection into the stratosphere in the tropics. This is because the updraft core radius had been set to 1500 km globally. This is a reasonable value for extratropical convection over land, and since the target region of ECCC's forecasting systems is Canada, it is a valid choice. However, observations (Lucas et al., 1994 and references therein) show that the updraft core radius varies with latitude and from land to ocean. Thus over the oceans, a value of 900 km is used in the tropics with 1000 km in the extratropics. Over tropical land grid points, a value of 1200 km is used. With these settings, tropical convection overshoots fall in the correct range (below 80 hPa). However, the model climatology is altered. The parameter changes impact land-sea contrasts which influence stationary Rossby wave generation and the spin-up of the Brewer-Dobson circulation in the stratosphere. They also affect flow over topography which impacts orographic wave drag and its associated transport circulations. An animation (**Fig. S1**) shows a comparison of the impact of changing Kain-Fritsch parameters versus that of adding convective tracer transport on CO_2 evolution. As expected, the change in parameters impacts the stratospheric distribution after a few months of simulation, whereas the introduction of tracer transport through deep convection impacts the tropical tropospheric CO_2 distribution at all times as well as the northern hemisphere in summer. The zonal mean values are small but so too are zonal standard deviations (both are less than 0.3 ppm). Another animation which compares the impact of changing KF parameters versus that of adding convective tracer transport on column mean or XCO_2 (**Fig. S2**) reveals that the magnitude of the impact is smaller for the change in parameters (maxima of 0.3 ppm) than for the introduction of tracer transport through deep convection (maxima of 0.8 ppm).

3.5 Vertical mixing

Turbulence in the Planetary Boundary Layer (PBL) is important for the transport of heat, momentum, moisture and constituent fluxes from the surface to the atmosphere. The PBL scheme in GEM is described in Bélair et al. (1999), Benoit et al. (1989) and Mailhot and Benoit (1982). Since a summary of GEM's PBL scheme was recently presented in McTaggart-Cowan and Zadra (2015) and Aliabadi et al. (2016), we note here only the main differences between GEM's PBL parameterization and those of other models used for greenhouse transport or flux inversion.

Shallow convection also exerts a significant impact on the tracer distribution and vertical mixing in the lower troposphere (e.g. Dacre et al., 2007). The shallow convection scheme in GEM is a modified Kuo parameterization (Kuo, 1974) developed at EC (Belair et al., 2005). As with the original Kuo scheme it does not address convective tracer transport. However, unlike the Kain-Fritsch scheme it is not based on a mass flux formulation that can be easily adapted to handle tracers. For this reason the version of GEM used for this study does not include tracer transport by shallow convection.

The vertical transport of greenhouse gas fluxes follows that of other constituents and is described by:

$$\frac{\partial C}{\partial t} = -\frac{1}{\rho}\frac{\partial}{\partial z}\left(\rho\overline{w'C}\right) = \frac{1}{\rho}\frac{\partial}{\partial z}\left[\rho K\left(\frac{\partial C}{\partial z} - \gamma\right)\right] \tag{4}$$

where C is the constituent mixing ratio on resolved scales, ρ is air density, w' is subgrid scale vertical velocity and the product with an overbar is the vertical flux of constituent due to subgrid scale turbulence. The first equality denotes the impact of subgrid scale motions on the resolved constituent distribution. The subgrid scale flux is parameterized through the second equality, and is a function of the vertical gradient of the constituent. γ is a counter gradient flux relevant for an
5   unstable PBL. K is the thermal eddy diffusivity and it depends on properties of the flow. Most models define K diagnostically in terms of an eddy length scale and local gradients of wind and virtual potential temperature. Such schemes are said to invoke a first-order closure. However, GEM's scheme specifies K through a prognostic equation for turbulent kinetic energy and is thus deemed a 1.5 order closure scheme (Holtslag, 2015). When K is determined diagnostically through local flow properties, the case of an unstable or strongly convective PBL is not well represented. Thus, an extension
10 for non-local mixing due to large-scale eddies is often invoked in which the counter-gradient term represents large eddy flux and K permits eddy length scales comparable to the PBL height in the case of unstable boundary layers. CarbonTracker (Peters et al., 2004) and GEOS-Chem (Lin and McElroy, 2010) both use such a nonlocal scheme based on Holtslag and Boville (1993). ECMWF also uses a diagnostic nonlocal scheme (Köhler et al., 2011). In GEM's 1.5 order closure scheme, K can also represent nonlocal effects due to large eddies in an unstable PBL through a careful definition of eddy length scale
15 (Bélair et al., 1999). One notable difference between GEM's prognostic nonlocal scheme and diagnostic nonlocal schemes is that the latter require PBL height as an input for determining K, whereas in GEM's scheme, the PBL height is diagnosed from the turbulent kinetic energy profile (so K is not a direct function of PBL height).

     A minimum value for eddy diffusivity of 0.1 m^2 s^{-1} is imposed in the PBL in GEM-MACH for air quality applications. This value was empirically chosen to balance the impacts on the various tropospheric reactive species (Paul
20 Makar, 2012, personal communication). In this work, the minimum value was raised to 10 m^2 s^{-1} because a value of 1 or 0.1 was found to occasionally lead to too little mixing. This was most evident in spuriously low CO_2 concentrations in the daytime during summer when biospheric fluxes are large and negative. (Note that the eddy diffusivity is kept fully varying in space-- it is only the minimum value that is slightly altered. This is in contrast to earlier versions of GEOS-Chem in which instantaneous full mixing of emissions and mixing ratios occurred in the PBL. See Lin and McElroy, 2010 for example.) A
25 negative consequence of raising the minimum value to 10 is a reduced-amplitude of the diurnal cycle of CO_2 (as will be seen in section 43). However, many models have difficulty in capturing the amplitude of the diurnal cycle (Law et al., 2008, Patra et al., 2008) so this consequence was considered tolerable at present. Moreover, the comparison of model results to measurements on sub diurnal time scales is very difficult, leading Law et al. (2008) to recommend that "comparisons with observations should only be made for daily or longer time averages, and possibly for only part of the diurnal cycle". Thus,
30 many inversion systems assimilate only afternoon mean observations and do not attempt to capture sub diurnal time scales (Peters et al., 2010).

32.6 The coupled meteorology and tracer transport forecast cycle

To transport greenhouse gases, a simulation cycle is used. Meteorological analyses archived at the CMC (Canadian Meteorological Centre) are inserted periodically to constrain the transport to reality. While operational assimilation cycles use an update frequency of 6 h, model forecasts during such a short period will be contaminated by spurious gravity wave generation, as evident in surface pressure time series (as well as other fields) (see Daley, 1991, ch. 6). However, after 24 hours, spurious gravity waves have generally dispersed so an update frequency of 24 h was chosen, just as in Agusti-Panareda et al. (2014). (Note that a 6 h update cycle had also been tested and the resulting CO_2 evolution was very close to that obtained with a 24 h update cycle, but because the stratospheric circulation was poorer, the 24 h update cycle is preferred.) A schematic diagram of the transport cycle is depicted in **Figure 1**Figure 1. The upper half of the figure depicts the operational system which collects meteorological observations over a 6 h window and uses these in a 4-Dimensional Variational Assimilation or 4D-Var (Gauthier et al., 2007) to generate an estimate of the meteorological state for a deterministic prediction. Although the operational deterministic prediction system now uses a hybrid approach to background error covariance estimation (Buehner et al., 2015), the analyses used here (from 2009-2010) were generated using the previously operational 4D-Var system (Charron et al., 2012). On the cycle's start date at 0 UTC, the meteorological analysis is combined with an initial state for greenhouse gases (here CO_2 only) and a 24 h coupled meteorology and tracer forecast is produced. The 24 h tracer forecast is subsequently combined with the meteorological analysis for the next day at 0 UTC (blue boxes) to produce a new coupled initial state for the second day's forecast (red filled circles). Note that no additional "initialization" or filtering (see Daley 1991, ch. 6, 9, 10) scheme such as a digital filter or incremental analysis updates (Bloom et al., 1996) is used, as the impact on the CO_2 field was found negligible. Also note that in our simulation cycles, no assimilation of greenhouse gases is performed. An assimilation system is currently in development as discussed in section 65.

In Figure 1Figure 1, every time a new analysis is inserted, a new surface pressure field informed by atmospheric observations is introduced. This surface pressure analysis will differ from the 24 h forecast of surface pressure because the model is not perfect, so the model forecast cycle will experience an abrupt shift in global mean surface pressure and hence, global air mass. The change in global air mass will then impact the global tracer mass. To maintain global tracer mass conservation across this discontinuity, it is necessary to redefine the tracer mixing ratio for the change in air mass. The scheme adopted follows that used by GEM for the global surface pressure adjustment. Specifically, a spatially uniform increment in tracer mixing ratio is determined based on the constraint that the global tracer mass be preserved (see Appendix A for details). The adjustment so obtained is small, with a mean value of -8×10^{-5} ppm and a standard deviation of 0.015 ppm over a one year simulation. Since the surface pressure analyses are stored with only 16-bit precision (because the analysis uncertainty does not warrant more precision than this), this ultimately limits our knowledge of the tracer mixing ratio and the local tracer mass. With our global adjustment scheme, the adjustments are much smaller than 16-bit precision will allow, and thus the scheme is justifiable.

14

4 3 Model evaluation

We will use GEM-MACH-GHG to study the influence of uncertain atmospheric transport on the spatial scales recoverable in CO_2 fields, but since this model has not yet been used for greenhouse gas transport, it is necessary to first document its ability to transport CO_2.

4 3.1 Evaluation of meteorological fields

ECCC has been delivering operational weather forecasts for over 40 years and the products have been based on the GEM model for 19 years. As with many national weather forecast centres, ECCC participates in regular intercomparisons of operational forecasts following WMO standards. For the global configuration, forecasts on the medium (up to 10 days) range are compared every month and are available at

http://web-cmoi.cmc.ec.gc.ca/verification/monthly/observations/obs_monthly_e.html. Thus the quality of GEM transport has been documented. However, since changes were made to the model that affect weather prediction (specifically, the change to the continuity equation, the implementation of conservation of the dry air contribution to the global mean surface pressure, and the convective transport scheme parameters) it is necessary to demonstrate the quality of the meteorological fields for the purpose of greenhouse gas transport. Since offline transport models often use reanalyses such as ERA-Interim or MERRA to transport constituents, we compare our 24 h forecasts against these products. In general, analyses more closely match observations than do forecasts because models are imperfect, and reanalyses should be superior in quality to operational analyses because the former use a superset of observations and a single, recent model version. Thus, our 24 h forecasts cannot be expected to be superior to any of ERAI, MERRA or JRA-55 (Kobayashi et al., 2015) reanalyses. Nevertheless, because of the usage of a 24 h forecast cycle (Figure 1Figure 1), it is important to verify that transport on this forecast range is reasonable, and reanalyses serve as high quality reference fields.

The initial meteorological fields used at the start of every cycle (Figure 1Figure 1) come from archived operational products interpolated to our lower-resolution grid and topography. Thus our transport can only drift from reality for 24 h before it is corrected so that, even with our modifications to GEM, significant transport errors are not expected. Nevertheless, since changes of several hPa in 3-day forecasts of surface pressure do arise from the adjustment to the continuity equation and these could impact synoptic-scale forecasts, it is worth verifying the quality of our modified transport. Figure 2Figure 2 (left column) shows the monthly mean difference of temperature, zonal and meridional wind differences between GEM-MACH-GHG and ERAI fields for July 2009 based on 6 hourly difference fields. ERAI fields were obtained from http://reanalysis.org on pressure levels in GRIB format and interpolated to GEM's grid at 1.5° resolution. GEM fields were also output on the same pressure levels and resolution. The mean differences can be compared to those between other reanalyses: MERRA and ERAI (middle column) and JRA-55 and ERAI (right column). The latter difference fields were obtained from http://reanalysis.org using the Web-based Reanalysis Intercomparison Tools (WRIT, Smith et al., 2014) with the values saved in NetCDF format for plotting. It is evident that for all fields, GEM-MACH-GHG

15

is as similar to ERAI as other reanalysis products although our products are 24 h forecasts (not analyses nor reanalyses). For temperature (top row), the largest differences appear in the stratosphere, for all models. In the stratosphere where models and observations are biased, assimilation can be challenging (Polavarapu and Pulido, 2016) so this result is not surprising. The extratropical zonal mean stratosphere is dominated by the slow Brewer-Dobson circulation which is forced by waves propagating upward from the troposphere (Andrews et al., 1987; Vallis, 2006). The spectrum of waves from a given model depends on parameterizations such as convection and gravity wave drag which generate high frequency subgrid scale waves, as well as resolved waves. So different models can be expected to have rather different spectra, and thus different forcing of the Brewer-Dobson circulation. For wind fields, largest differences occur in the tropics. Zonal wind differences are greatest in the tropical stratosphere (middle row). The zonal mean tropical stratosphere is dominated by the quasi-biennial oscillation (QBO) which is driven by vertically propagating waves. Climate models have difficulty capturing the QBO as it depends on vertical resolution, gravity wave drag parameterizations and the spectrum of waves generated by tropical convection schemes (Baldwin et al., 2001; Campbell and Shepherd, 2005). Analyses can capture the QBO by assimilating the few radiosonde wind observations available in the tropics. However, the amplitude and phase of the QBO captured may depend on the underlying model and assimilation system characteristics. In general, since few direct measurements of winds are available to constrain analyses, and there are no simple dynamical balances available to infer winds from observations related to mass fields, model biases in the tropics are difficult to correct with data assimilation (Polavarapu and Pulido, 2016). In Figure 2Figure 2, the meridional winds (bottom row) differ most in the tropical troposphere where these issues of lack of measurements, and simple dynamical balances will prevail.

In summary, the meteorological fields used to transport CO_2 in our system are similar in quality to those from reanalyses which have been used with offline transport models. Reanalyses are not perfect and their uncertainty can be quantified through the degree of disagreement with other reanalysis products. Thus, when offline transport models use reanalyses, this uncertainty impacts their transport quality. While we use a 24 h update cycle, this subsection has demonstrated that throughout the 24 h forecast, the uncertainty is comparable to that of an individual reanalysis product. Thus results on the sensitivity of transport error to meteorological uncertainty obtained with our sequence of 24 h forecasts has direct relevance to transport errors obtained with a 6 h insertion of reanalysis products. While statistics for July 2009 were shown in Figure 2Figure 2, those for Dec. 2009, July 2010 and Dec. 2010 are found in the Supplemental Material (**Figs. S3-S5**).

43.2 Evaluation of CO$_2$ fields

Having established that the meteorological fields that will transport constituents are sufficiently accurate, the CO_2 evolution can be assessed. Since the goal is to assess GEM-MACH-GHG as a transport model, realistic surface-to-atmosphere fluxes are required. Without realistic sources and sinks of CO_2, discrepancies with observations could be equally attributed to erroneous fluxes as to erroneous transport. Even with a good source of retrieved atmospheric fluxes, the evaluation of GEM-MACH-GHG transport requires care (as discussed below). For this purpose, posterior fluxes from CarbonTracker 2013B

(hereafter referred to as CT2013B, Peters et al., 2007) ~~and CarbonTracker 2010 (hereafter CT2010)~~ were obtained and regridded in a mass conservative way to GEM's grid. CarbonTracker was chosen because it is generally recognized as a good product, is regularly monitored and updated, and is readily available from http://carbontracker.noaa.gov. These fluxes were also used for a similar purpose in Houweling et al. (2010). CT2013B fluxes are available every 3 h, with fluxes inserted at every model timestep. With ~~fluxes from CT2010 and~~ an initial condition from ~~CT2010~~ CarbonTracker for 1 January 2009 0 UTC, ~~GEM-MACH-GHG was run for 2009. With~~ and the CT2013B fluxes, GEM-MACH-GHG was run for all of 2009-2010. The resulting CO_2 simulations are compared against observations assimilated by ~~CT2010 and~~ CT2013B (continuous observations), observations not assimilated by ~~CT2010 and~~ CT2013B (columns from Total Carbon Column Observing Network (TCCON) and aircraft profiles), and gridded 3-dimensional concentration fields from ~~CT2010 and~~ CT2013B.

Figure 3~~Figure 3~~ compares CO_2 time series from GEM-MACH-GHG with CT2013B fluxes to surface observations from ECCC's greenhouse gas measurement network (Worthy et al., 2005) at Alert, East Trout Lake and Sable Island. Alert is a remote Arctic site far from CO_2 sources so it can be used to assess long-range transport. East Trout Lake is close to sources and will reflect diurnal variations in CO_2 fluxes convolved with boundary layer mixing. Sable Island is downstream of sources on the eastern edge of the continent and effectively reveals synoptic-scale variability. The comparison at Alert (top panel) reveals general agreement between the model and measurements throughout the 2-year period, with the exception of low values during the autumn of both years. This departure can be explained by a discrepancy between CarbonTracker's transport and GEM-MACH-GHG's transport. CT2013B fluxes were obtained by minimizing the difference between observations and CarbonTracker's forecasts and thus reflect the amount of flux needed to bring CarbonTracker in line with measurements. Since CarbonTracker's transport is not perfect, the retrieved fluxes retain a signature of CarbonTracker's transport errors which may or may not match GEM-MACH-GHG's transport errors. To demonstrate this point, GEM-MACH-GHG was run with another set of retrieved fluxes, this time from GEOS-Chem (Deng et al. 2016) ~~using GOSAT data~~, for July 2009-Dec 2010 and no such autumnal drift is seen (**Fig. S6**). Thus GEM-MACH-GHG is potentially able to capture long and seasonal time scales of CO_2 transport. The time series at East Trout Lake (middle panel) shows that the smaller amplitude of the diurnal cycle in winter relative to summer is captured by the model. However, CarbonTracker's diurnal cycle amplitude in summer at this site appears too large. The amplitude of the diurnal cycle is further discussed in the context of **Figure 4**~~Figure 4~~. Finally, the bottom panel compares observed and modelled time series at Sable Island where synoptic-scale variations are more evident. It is clear that the model captures synoptic scales well, as expected from a weather forecast model. In particular, variations on the one month timescale are seen in August and September of 2009 and the model follows these variations well. The ability of GEM-MACH-GHG to capture synoptic scales is important, since the ground-based measurement network can resolve the global carbon budget and very large (continental) spatial scales (e.g. Peylin et al., 2013). Then in a data assimilation or flux inversion system, the model can supplement the large spatial scales observable with this network with realistic synoptic scales (e.g. Agusti-Panareda et al., 2014).

17

The mean diurnal cycle at individual stations can reveal more clearly the realism of a model's boundary layer variation. Here we choose two sites close to sources and sinks because the amplitude of the diurnal cycle at such sites should vary through the year. **Figure 4**~~Figure 4~~ shows that the model's mean diurnal cycle at East Trout Lake with GEM-MACH-GHG with CT2013B (blue) ~~or CT2010 (red)~~ fluxes compare well to measurements (black) whereas CT2013B (green) has a too-large amplitude in April and July 2009 (consistent with Figure 3~~Figure 3~~). However, the model's behaviour varies with location and time. At Fraserdale, GEM-MACH-GHG's diurnal cycle amplitude is clearly too low in all months whereas CarbonTracker fares better in July and October 2009. The fact that the diurnal variability from GEM-MACH-GHG is lower than that observed stems from the choice made for the minimum value of eddy diffusivity (section ~~2~~3.5). Lowering this value can increase the amplitude of the diurnal cycle, but with the increased risk of occasional spuriously low values of CO_2 during summer daytime. As noted earlier, it is difficult to compare models to measurements on sub-diurnal time scales (Law et al., 2008; Patra et al., 2008) and most models have difficulty in capturing boundary layer evolution so flux inversions typically use only afternoon mean measurements. ~~Nevertheless, in the future, this minimum value may be revised as GEM MACH GHG's boundary layer is routinely assessed. The fact that the diurnal cycle amplitude is not poor at all locations may make finding a single best parameter value rather difficult. Indeed, Aliabadi et al. (2016) show that the realism of PBL height estimates in stable conditions compared to observations depends on location as well as the choice of parameterization.~~

GEM-MACH-GHG transport is directly compared to CarbonTracker's transport in **Figs.** 5 and 6. **Figure 5**~~Figure 5~~ presents the column mean CO_2 weighted by air mass on 1 July 2009 0 UTC and 31 Dec. 2009 0 UTC from GEM-MACH-GHG with CT2013B fluxes (top panel), ~~GEM MACH GHG with CT2010 fluxes~~ CT2013B (middle panel) and ~~CT2010 mole fractions~~ the difference between these two (bottom panel). ~~The largest differences between the simulations occur in the summer, possibly due to the different convection schemes used in the two models. For example, the GEM MACH GHG simulations with different posterior fluxes are closer to each other than to CarbonTracker in the tropical Atlantic.~~ The higher resolution of GEM-MACH-GHG is evident particularly in July where the largest differences occur at small spatial scales. CT2013B has more CO_2 than GEM-MACH-GHG in both the tropics and the northern hemisphere. However, in winter (right column), CT2013B has more CO_2 mainly in the tropics. The tropical differences in both seasons are likely related to differences in convection schemes in the two models. The increased CO_2 in the northern midlatitudes in boreal summer may be due to differing meridional transport or to differing rates of vertical mixing or a combination of both. Zonal mean fields also reveal that the greatest differences between the two GEM-MACH-GHG simulations and CT201~~0~~3B are in the ~~tropics in~~ summer (**Figure 6**~~Figure 6~~, top panel). Throughout the troposphere and near the surface, CT2013B has more CO_2. In winter, CT2013B has a slight deficit near the surface compared with GEM-MACH-GHG. ~~In winter (**Figure 6**, bottom panel), GEM MACH GHG with CT2010 fluxes matches CT2010 in the tropics better than GEM MACH GHG with CT2013B fluxes, indicating that the fluxes are more important for explaining the tropical structure than differing model dynamics at this time. This is consistent with the finding that posterior fluxes are sensitive to prior fluxes particularly in data sparse regions such as the tropics (Peters et al., 2007, 2010). However the differences between both GEM simulations and~~

18

~~CT2010 are very small compared to the difference between CT2010 and CT2013B in the tropics (not shown) (which is largely attributed to the convective mass flux fixes in the latter see http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2015_doc.php).~~ Differences in the stratosphere, ~~particularly on 31 December 2009~~ between the two models are also evident in **Figure 6**~~Figure 6~~. Since GEM-MACH-GHG has better vertical resolution compared to CarbonTracker (80 versus 25~~34~~ levels) and GEM is designed to have a realistic stratosphere (Charron et al., 2012), differences in stratospheric and mesospheric flow are to be expected. ~~Differences at these levels are even greater when compared to CT2013B where vertical resolution was further reduced to 25 levels (not shown).~~ However, the mass of CO_2 in the stratosphere and mesosphere is very small so that column mean or surface values would be insensitive to such differences. ~~In summary, overall, GEM-MACH-GHG with CarbonTracker's retrieved fluxes reproduces CarbonTracker's fields rather well.~~ In summary, when GEM-MACH-GHG is run with CT2013B fluxes, differences in transport errors between CarbonTracker and GEM-MACH-GHG are evident. Differences are within 3 ppm in column mean and 4 ppm in zonal mean. Neither (nor any) model can be expected to have perfect transport so the acceptability of transport is generally gauged through comparisons of model predictions to measurements.

To assess seasonal time scales, it is useful to compare to the Total Carbon Column Observing Network (TCCON) measurements (Wunch et al., 2011). The data used for this study are from the GGG2014 release, available on the network's website http://tccon-wiki.caltech.edu. All sites with measurements in 2009 and 2010 are selected, as listed in **Table 1**~~Table 1~~. The details of how the model profiles were converted to column-averaged dry mole fraction (XCO_2) and smoothed following Wunch et al. (2010) is provided in Appendix B along with the precise definitions of the statistics discussed here (bias, root mean square or RMS, and scatter). The statistics for 2009-2010 are shown in **Table 2**~~Table 2~~. The bias is below 1 ppm for every station except Eureka which only had 49 hours of measurements in 2010. No selection is applied when considering which sites are included in the ALL, MEAN or standard deviation (SD) statistics. At Eureka, the -2.66 ppm bias significantly impacts the station-to-station SD of the bias, it is 0.3 ppm without Eureka. (Note that an error in surface pressure was recently discovered at Eureka and preliminary results for a correction suggest a 0.5 ppm difference in XCO2, which would reduce the Eureka EC-CAS bias by a corresponding amount.) Except for Eureka and Park Falls, the station bias is positive and the overall standard deviation is 0.96 ppm with a high correlation coefficient of 0.95. Although we did not perform a data assimilation, the posterior fluxes from CarbonTracker contain information from the observations they used, so we can compare our **Table 2**~~Table 2~~ to Massart et al. (2016, Table 2). The biases of individual stations are mostly lower here, as is the overall averaged bias. ~~Since the analyses in Massart were based on GOSAT whereas CT2013B used surface data, this may indicate a better agreement of surface observations with TCCON.~~ This is because the surface observations assimilated by CarbonTracker with a long assimilation window are able to constrain the global atmospheric growth rate whereas the system used by Massart et al. (2016) does not use long assimilation windows and thus does not constrain the global growth rate as effectively. The full time series for Park Falls (USA) and Wollongong (Australia) are shown in **Figure 7**~~Figure 7~~ and the seasonal bias as well as the seasonal statistics using data from all sites combined are shown in **Table 3**~~Table 3~~. Seasons are defined as DJF (December, January, and February), MAM (March, April, and May), JJA (June, July,

19

and August) and SON (September, October, and November). The model is able to reproduce seasonal variations of XCO_2 with biases ranging between -1 and +1 ppm (excluding Eureka) and scatter values consistently below 1 ppm. The large negative bias at Eureka in autumn is consistent with the time series shown in Figure 3~~Figure 3~~. Other northern stations (Sodankylä and Bialystok) have similar but smaller biases. As discussed earlier, the discrepancy of GEM-MACH-GHG transport with that of CarbonTracker (as imprinted in the posterior fluxes) may explain~~s~~ this behaviour since other posterior fluxes do not have this particular issue (**Fig. S6**).

The vertical structure of model CO_2 is compared to NOAA aircraft profiles (Sweeney et al. 2015) over Canada and the U.S. in **Figure 8**~~Figure 8~~. Following Agusti-Panareda et al. (2014), mean model profiles at the nearest model gridpoint and timestep to the profile location were averaged over all profiles for a season. Both observed and model values were binned into 1 km layers. The observations are from ObsPack2013 (Masarie et al., 2014) and include only profiles from Canada or the contiguous U.S. The annually averaged model profiles are shown in panel a. In the annual average, GEM-MACH-GHG has ~~exceptionally~~ good agreement with these independent measurements while CT2013B has a very slight positive bias. Both models are quite good compared to the ensemble of models shown in Stephens et al. (2007, Fig. 2b). However, the other panels reveal that GEM-MACH-GHG's excellent annual result in the free troposphere is because of compensating errors in different seasons. The boreal winter season (Dec-Jan-Feb) is not shown because Dec 2008 was not simulated. However, the behaviour of the model in boreal winter 2009 (based on Jan-Feb) and 2010 is qualitatively similar to its behaviour in spring (panel b). Panels b-d reveal that GEM-MACH-GHG agrees quite well with observations from 3-6 km in all seasons. However, from 1-3 km, vertical gradients are too sharp but of opposing directions in boreal spring (panel b) and summer (panel c). In autumn (panel d) the gradient is slightly too ~~small~~large while CT2013B is almost perfect. Thus, the vertical mixing just above the boundary layer is too weak, as with most models (see Yang et al., 2007). A possible reason for the overestimation of vertical gradient with GEM-MACH-GHG is the lack of tracer transport through shallow convection. ~~Because the GEM-MACH-GHG profiles obtained with the 2 different posterior fluxes have very similar profiles in all seasons, the departure of vertical gradients from those observed is attributed to the model's formulation.~~ These biases in vertical gradients will be relevant for regional flux inversions (see Stephens et al. 2007) that may use GEM-MACH-GHG results as boundary conditions. ~~Overall, however, compared to CarbonTracker, vertical gradients in GEM-MACH-GHG agree better with independent measurements in the mid troposphere but less well in the lower troposphere.~~

54 The predictability of CO_2 in the context of uncertain atmospheric analyses

Having established that GEM-MACH-GHG can simulate CO_2 reasonably well, we turn our attention to the question of how atmospheric transport modulates the CO_2 distribution. The evolution of CO_2 can be described by the species transport equation and thus may be considered to be perfectly predictable. However, this is only true if the advecting fields are perfectly known, and this is never the case. With a coupled meteorology and forecast model, the impact of the uncertainty of

20

meteorological fields on CO_2 transport can be explored. In a data assimilation or flux inversion system, ~~if the fluxes are perfectly known~~when the fluxes are well constrained by observations, the ability to estimate the CO_2 fields using observations will ultimately be limited by the loss of meteorological predictability, just as the quality of weather forecasting products are. Thus it is useful to identify these limits on the spatial scales that can be retrieved in analyses of CO_2 because these predictability limits can then be compared to the spatial scales of CO_2 retrievable in the presence of meteorological analysis ~~errors~~uncertainty. In this section, we use GEM-MACH-GHG to determine the predictability of CO_2 on weather and climate time scales with the latter referring to subseasonal to seasonal scales. In other words, we isolate and study the component of transport error due to meteorological state (initial condition or analysis) errors. In section ~~5~~4.1, the classic weather predictability problem of uncertainty in meteorological initial conditions is considered whereas longer time scales and the transport errors due to uncertain meteorolog~~ical analysis errors~~y, and model errors ~~and flux errors~~ are considered in section ~~5~~4.2.

~~5~~4.1 Weather time scales

Although flux inversion systems focus on retrieving relatively long time scale signals (than 2 weeks), it is useful to first consider CO_2 predictability on weather time scales before considering errors on longer time scales (next subsection). Specifically, we isolate the impact of the loss of meteorological predictability on CO_2 predictability. The meteorological predictability problem on weather timescales is related to forecast sensitivity to initial conditions, but the atmospheric variability of CO_2 on diurnal (Law et al., 2008), synoptic (Chan et al. 2004, Agusti-Panareda et al. 2014) and seasonal and interannual (Gurney et al., 2002, 2004; Baker et al. 2006a, LeQuéré et al., 2015) time scales is largely governed by the terrestrial biospheric fluxes and hence is determined by sensitivity to boundary rather than initial conditions. Nevertheless, the CO_2 predictability error arising from loss of meteorological predictability on weather time scales has not to our knowledge been identified, and it can be used to identify an upper limit on forecast errors. This may be relevant for operational data assimilation or forecasting systems such as those at ECMWF (Agusti-Panareda et al. 2014) and NASA Goddard (Ott et al., 2015) which use update cycles of 12 or 24 h and also examine the quality of short term forecasts. It will also serve as an upper limit for transport error arising from~~in~~ the presence of uncertain meteorological analyses in section ~~5~~4.2.

Predictability of weather normally refers to the sensitivity of forecast errors to initial conditions such that any infinitesimal perturbation will lead to diverging forecasts in a finite length of time. This is the so-called butterfly effect and it occurs because of the underlying nonlinear chaotic dynamics of the governing equations (Palmer, 2006). To compute predictability error of meteorological variables on weather time scales, one can simply start with a reference simulation and perturb the initial conditions. Eventually, the forecasts will diverge, but the error will saturate at climatological levels. Once saturation has been reached, the statistics of this predictability error can be determined. However, with a transported tracer such as CO_2, the model forecast requires regular insertion of wind fields. During our forecast cycle, the meteorology is constrained by analyses every 24 h and departure from reality will represent at most a 24 h forecast error. Thus for

21

transported constituents, the definition of the predictability experiment is slightly different. The reference simulation will be taken as the GEM-MACH-GHG two year run with CT2013B fluxes. Then a comparable "climate cycle" is run in which the model, initial conditions, CO_2 fluxes and surface forcing are identical to those used in the reference cycle. However, with the second and all subsequent cycles, the meteorological fields are not replaced by analyses but are instead copied from the

5 24 h forecast fields. Thus the meteorology fields used to transport the CO_2 field in this "climate cycle" are never updated with observations (analyses) and will thus depart from those used in the reference cycle in the first two weeks. The divergence of the CO_2 field in the "climate cycle" from that in the reference cycle, once the error has saturated, defines the CO_2 predictability error arising from the loss of meteorological predictability.

Starting from 1 January 2009 0 UTC, the reference and climate cycles are run for one month with fields saved every

10 6 h. The differences between the corresponding fields from the two cycles are computed for temperature, zonal and meridional wind components, CO_2 and surface pressure. Vorticity and divergence fields are computed from the wind difference fields. Then the ~~square root of the~~ global mean of the zonal ~~variance~~ standard deviation of each difference field is computed after first subtracting the zonal mean. The resulting global mean values are called predictability errors and they have the same units as the corresponding forecast field. In order to get a comparable scale for all variables, the errors were

15 normalized by the global mean ~~square root~~ of the zonal ~~mean variance~~standard deviation for a reference state. The choice of this state is arbitrary but has implications on the maximum values attained (as will be discussed below). The reference field was taken as the initial state used to launch both the control and climate cycles and corresponds to 1 January 2009 0 UTC. **Figure 9**~~Figure 9~~ shows the time and height variation of the predictability error normalized by the variability of the reference state for all four variables. When the predictability error approaches the variability of the reference state, values

20 approach 1. Thus predictability is expected only when the relative error is much less than 1. In **Figure 9**~~Figure 9~~, we see that temperature loses predictability within 10 days, as expected (i.e. the normalized error reaches 0.8). However, CO_2 loses predictability in the troposphere within 2-3 days except very near the surface where it rea~~approa~~ches 5 days. The predictability of CO_2 more closely resembles that of the wind (vorticity and divergence) fields which also lose predictability in fewer than 5 days in the troposphere. This makes sense because the wind fields are used to transport the CO_2 field. The

25 difference in evolution of the CO_2 field in the reference and climate cycles (figures not shown) reveals largest values to be associated with gradients created by large fluxes (whether natural or anthropogenic). This ability of the uncertainty in wind analyses to act on CO_2 gradients and to spread the uncertainty downstream was previously illustrated by Liu et al. (2011) using an ensemble of wind fields. CO_2 is more predictable in the stratosphere, with the loss of predictability occurring after 5 days in the lower stratosphere. The extended predictability of CO_2 in the lower stratosphere is similar to that seen in the

30 vorticity field (bottom left panel). The reason that the vorticity field is more predictable than the divergence field (with a loss of predictability occurring after 3-4 days in the troposphere) is because the vorticity field is associated with slower rotational modes whereas the divergence field is often associated with higher frequency waves. The atmospheric kinetic energy spectrum is dominated by rotational motions in the troposphere (Koshyk et al., 1999; Skamarock et al., 2014). In the stratosphere, the zonal mean flow in winter is driven by very large-scale vertically propagating planetary waves (Andrews et

al, 1987; Vallis, 2006) so large-scale rotational modes dominate the energy spectrum and extended predictability in vorticity and CO_2 results.

~~Stripes of 24 h in frequency are seen in the vorticity and divergence plots. This occurs because the forecasts in the reference cycle are abruptly returned to reality every 24 h with the insertion of a new analysis. When the predictability error of the 24 h forecast error is large compared to that of the analysis valid at the same time, we see stripes at the 24 h period. Thus we conclude that the normalized 24 h forecast error of the wind field is much larger than that of the temperature field. This makes sense because the global observing system is dominated by information about the mass field with relatively sparse direct observations of the wind field (Baker et al., 2014). In addition, the mass field (which is reflected in the temperature field) is a much smoother field and is thus more easily observable with a given network relative to fields which are dominated by smaller spatial scales (such as vorticity or divergence).~~

While predictability is considered lost when its normalized error approaches 1, sometimes the relative error is much greater than 1. This occurs because of the arbitrary choice taken for the normalization. In our case, the reference state corresponds to the initial state which corresponds to a relatively quiescent synoptic situation since January 2009 marks the strongest and most prolonged stratospheric major warming on record (Manney et al., 2009). The criteria for a stratospheric sudden warming (SSW) were met on 24 January ~~-~~2009 when zonal mean easterlies replaced the climatologically normal westerlies at 10 hPa. However, easterlies were noted in the mesosphere prior to this date (Manney et al., 2009). This pattern is consistent with the appearance of anomalously large predictability error in <u>relative</u> vorticity ~~and divergence~~ in the mesosphere prior to 24 January and the appearance of anomalously large errors in the ~~lower~~ stratosphere <u>(layer 6) after 15 January</u> ~~on this date~~. Although global mean values are shown, it is zonal ~~variance~~<u>standard deviation</u> that is computed, and the departure from a zonal mean will be large during a wave 2 vortex splitting event such as occurred in 2009. Thus zonal <u>standard deviations</u>~~variances~~ are anomalously large throughout much of the northern hemisphere because the climate cycle does not capture this event whereas the reference cycle does. The extent of the disturbance in the northern hemisphere is large enough to influence the global mean values. Anomalously large CO_2 predictability error appears in the mid stratosphere around 20 January 2009. The disturbance of CO_2 due to the SSW is to be expected since the disturbance of CO in the mesosphere, N_2O in the mid stratosphere and H_2O in the lower stratosphere as the vortex deformed and split was evident in MLS (Microwave Limb Sounder) observations (Manney et al., 2009).

Figure 10~~Figure 10~~ compares the layer mean normalized predictability error for different variables for the layers indicated by the dashed lines in **Figure 9**~~Figure 9~~ and labelled in white text at the left edge of each panel. Near the surface, the normalized CO_2 predictability error closely follows that of the specific humidity field (for about <u>3</u>~~7~~ <u>8</u> days). Both moisture and CO_2 fields are advected by the wind fields and are similarly affected by the predictability of the wind fields. In this bottom layer, CO_2 and moisture are more predictable than the wind field presumably because of their dependence on surface fluxes <u>a</u>~~-~~<u>lthough the normalization which is layer dependent cannot be ruled out as a contributing factor.</u> However, they are both less predictable than the temperature field (Fig. 10a). In the troposphere ~~and stratosphere~~, the loss of CO_2 predictability is similar to the loss of predictability in vorticity (panels b,<u>c</u>~~d,e,f~~) <u>for the first few days</u>. After that, the loss of

23

predictability for CO_2 is faster than that for the vorticity and divergence fields. While the normalization may play a role in this result, it is also worth noting that the predictability of vorticity and divergence fields will differ from that of wind components because very smooth wind field errors generate little error in vorticity and divergence fields. Throughout the atmosphere, temperature loses predictability at a slower rate than does CO_2. While both moisture and CO_2 are transported by wind fields, moisture is also a dynamic variable thus the loss of predictability for specific humidity is not the same as that for CO_2 in the lower to mid troposphere (panels b,c) beyond day 5. Predictability error increases from day 1 to reach saturation levels in 10-15 days for all levels except the upper stratosphere (Fig. 10f) which is being affected by the stratospheric sudden warming. Thus in the next section, climatological levels of predictability error will be discussed for periods longer than 1 month in the troposphere.

Waves of 24 h in period are seen, particularly, in the vorticity and divergence plots in Fig. 10. This occurs because the forecasts in the reference cycle are abruptly corrected every 24 h with the insertion of a new analysis. When the predictability error of the 24 h forecast error is large compared to that of the analysis valid at the same time, we see stripes at the 24 h period. Thus we conclude that the normalized 24 h forecast error of the wind field is much larger than that of the temperature field. This makes sense because the global observing system is dominated by information about the mass field with relatively sparse direct observations of the wind field (Baker et al., 2014). In addition, the mass field (which is reflected in the temperature field) is a much smoother field and is thus more easily observable with a given network relative to fields which are dominated by smaller spatial scales (such as vorticity or divergence).

In summary, the global predictability of CO_2 due to uncertain meteorological initial conditions on weather time scales in the free troposphere is very short (less than 2 days with our model)in the free troposphere and is associated with the predictability of wind fields. This predictability limit refers only to sensitivity to meteorological initial conditions and it will be counter balanced by the predictability coming from biospheric fluxes on diurnal and synoptic scales. However, because atmospheric flow modulates CO_2 evolution on these time scales (Parazoo et al., 2008, Law et al., 2008), it is useful to identify these theoretical limits. It is important to note (and this is discussed more fully in section 6) that the predictability diagnostic used here is dependent on an arbitrary normalization and an arbitrary threshold to define predictability so absolute results are not expected. Rather the relative predictability between variables or between atmospheric layers is expected to be more reliable. To improve this type of predictability, more observations will be needed where the wind fields have finer spatial scales and where convection is occurring. The current global meteorological measurement network is relatively sparse in the tropics where convection is important, but new observations from space-borne lidars may be able to remedy this problem (Baker et al., 2014).

54.2 Seasonal time scales

As noted earlier, CO_2 observations contain information on seasonal to interannual time scales. Specifically, the global surface network used in flux inversions is able to constrain the global CO_2 budget and capture seasonal and interannual variability of the global fluxes (e.g. Baker et al., 2006a; Peylin et al., 2013). The source of predictability on subseasonal to

seasonal and longer time scales partially derives from climate predictability on those scales but also from long time scale information contained in ~~surface~~terrestrial biospheric fluxes which are, in turn, influenced by climate variability (Patra et al., 2005). In this section, we explore the predictability of CO_2 on longer time scales and compare these to CO_2 simulation errors due to the use of uncertain meteorological analyses.

5   The predictability experiment represents an extreme case in which no information from observations is present in the wind fields after the initial time. In reality, in our CO_2 transport cycle (Figure 1~~Figure 1~~), the wind fields are constrained to observations by the insertion of a meteorological analysis every 24 h. However, the analyses are not perfect and have a certain level of uncertainty. To simulate this uncertainty, we could perturb the analyses every 24 h with an analysis error. In a variational data assimilation system such as that used at ECCC, it is possible to estimate the analysis error covariance

10   matrix but it is expensive to do so, and such estimates are not routinely made. On the other hand, a simple perturbation such as random spatially uncorrelated errors will not be useful as they will primarily generate unbalanced motions. What is more relevant is a perturbation of the size and shape of the 6 h analysis error, given the use of a 6 h forecast cycle in operations. (Reanalyses are also available at 6 h intervals and these are sometimes used to constrain flux inversions.) Thus, in order to simulate a coherent 6 h analysis error, we simply insert the analysis state valid 6 h prior to the actual analysis time (i.e. the

15   one from 18 UTC of the day before, instead of the correct one from 0 UTC), relabelling the date and time to the correct ones. In addition, because the shift in the diurnal cycle for the meteorology would impact the CO_2 predictability, the diurnal cycle is removed from the perturbations by subtracting the monthly mean of each synoptic hour. Then, the deviation of the CO_2 field from this perturbed analysis cycle from the reference cycle defines the error due to the use of uncertain meteorological analyses. This error should be much smaller than the predictability error arising from uncertain meteorological initial

20   conditions. However, it should be larger than an actual analysis error because of the additional component corresponding to the evolution of the true state in 6 h. Thus it is an over-estimate of analysis error.

Figure 11~~Figure 11~~ shows the monthly mean spatial spectra of various difference fields averaged over several model levels for July (top row) and December (bottom row) of 2009. The spectra refer to the spherical harmonics (Boer 1983) of a scalar field multiplied by its complex conjugate and summed over zonal wavenumbers. The x-axis then defines a

25   total wavenumber. While spectra were computed for each day at 0 UTC, these were averaged over the month to filter some noise and identify a robust signal. In addition, they were averaged over 12 model levels to get representative spectra for a few atmospheric layers, namely, the bottom 4 layers shown in **Figure 9~~Figure 9~~**. The blue curves in **Figure 11~~Figure 11~~** depict mean spectra of the CO_2 state from the reference cycle similarly averaged in time and in the vertical dimension. The black curves represent the predictability error arising from uncertain meteorological initial conditions. This error is very

30   small at the end of the first cycle (on 2 January 2009) but rapidly increases during the first 2 weeks to saturate at climatological levels (section 4~~5~~.1). Since we are interested in this saturated level of error, we do not consider the first month of errors. The two months chosen in **Figure 11~~Figure 11~~** represent the variation seen in various months of the year. The predictability error is seen to be lower than the reference state itself for very large spatial scales but quickly equals (around wavenumber 10) then surpasses the power in the reference state. The reason that the power in the predictability

25

error can be larger than that in the state itself is because it involves the difference of two fields. In the limit where two fields become uncorrelated, the variance of the difference equals the sum of the variances. If the two fields have the same climatological variance, the variance of the difference is twice the climatological variance. Thus it is not surprising that the power in the predictability error should surpass that in the reference state for small spatial scales. What is more intriguing is

5   that some information is still retained in the largest scales (wavenumbers less than 5) even after 6 or 12 months of simulation. The source of this predictability at very large scales is mainly due to surface fluxes of CO_2 which are the same for both the reference and climate cycles. However, it might also be partially from the surface forcing of meteorological fields. Subseasonal to seasonal predictability is manifested in modes of variability such as the Madden-Julian oscillation (MJO), the Pacific North American (PNA) pattern, midlatitude blocking events and the North Atlantic oscillation (NAO)

10   (Waliser, 2006) and their predictability derives from atmospheric boundary conditions, namely, sea surface temperature, soil moisture, snow cover, vegetation and sea ice (Shukla and Kinter, 2006). These ocean and land surface conditions influence fluxes of moisture, and sensible and latent heat into the atmosphere which may change low level atmospheric convergence and lead to atmospheric heating anomalies which influence the large-scale flow. To see if long time scales in the meteorological analyses explainplay a role in the large-scale predictability seen in **Figure 11**Figure 11, an experiment was

15   run in which the predictability experiment was repeated, but this time incorrect surface fields (from 3 months later) were used. With no information from atmospheric observations as well as a seasonally shifted error in the surface forcing, the predictability error is worsened at these largest scales, particularly in the summer (June, July and August) near the surface (**Fig. S7**), and in the lower troposphere (not shown). The differences in CO_2 evolution in the two predictability experiments during boreal summer are largest in the northern extratropics (not shown). This confirms that the ocean and land surface are

20   playing a role in predictability of the system at the largest scales in the lower troposphere in boreal summer. In the mid and upper troposphere, smaller impacts are seen but the impact is largest in the spring (not shown). In **Fig. S7**, the remaining predictability seen at wavenumbers below 10 for all months is then attributed to the common CO_2 fluxes used by the reference and predictability experiments. Thus Fig. S7 confirms that the dominant source of predictability at large scales seen in Figure 11 is due to surface fluxes.

25   In summary, a direct impact of climate predictability on CO_2 predictions through the ocean and land surface is seen through the worsened predictability in boreal summer months when biospheric CO_2 fluxes are largest. Since CO_2 fluxes were specified, and were the same in the reference and predictability experiments, this climate signal is retained and explains most of the CO_2 predictability at large scales in **Figure 11** and **S7**. Finally, it is worth noting that a shift of one month in surface fields resulted in no real deterioration of the predictability error since there is still significant correlation among

30   surface fields due to a one month lag.

    The red curves in **Figure 11**Figure 11 depict the 6 h analysis error spectra due to analysis uncertainty. As expected, these spectra are reduced compared to predictability errors because meteorological analyses are used in this cycle, although they are 6 h out-of-date. In particular, significant error reduction at large scales is seen. However, the red curves also intersect the reference state spectra at increasingly smaller wavenumbers as height increases. Thus, if analysis

26

~~uncertainty is~~errors are considered, there is a gain of information over climatological error levels defined by the predictability error but only for the larger spatial scales. The analysis perturbations used here overestimate analysis error so we can expect that the spatial scales resolved are a conservative estimate. But whatever the size of the error, there will be a limit to the spatial scales resolved because of meteorological analysis imperfections. At the very least, the meteorological observing system has a given spatial resolution and moreover, meteorological assimilation systems (like ECCC's) may compute analysis increments at a coarser resolution than that of the model. Beyond the point where the red and blue curves intersect, there is no useful information in the CO_2 field at these scales due to meteorological analysis ~~errors~~uncertainty because the power in the CO_2 prediction error is larger than that in the CO_2 state. In fact, the CO_2 predictions with ~~6 h~~ analysis ~~errors~~uncertainty (red curves) asymptote to the predictability error spectra for large wavenumbers. For these spatial scales, using the updated meteorological analysis is no better than having no updates at all. Near the surface, there is the greatest gain of information, but in the upper troposphere, the spectra of CO_2 errors due to analysis uncertainty ~~error~~ are less than that of the reference state only for wavenumbers lower than 3~~2~~0. Thus the fact that the meteorological analysis has information on only certain spatial scales places limits on the spatial scales that can be retrieved in a transported field such as CO_2. ~~This result is consistent with the observation that transport biases act at large scales (Chevallier et al. 2010).~~

~~The final set of curves (cyan lines) in **Figure 11** depict the spectra of differences in an experiment which is identical to the reference but which uses a different set of posterior fluxes (CT2010 instead of CT2013B). When information differs in the two sets of retrieved fluxes the spectra of differences are large. It is evident that the differences are greatest at the largest spatial scales and nearer the surface. Since the observing system is broadly similar in the two systems (CT2010 and CT2013B) and both are based on the surface network which defines large spatial scales (e.g. Bruhwiler et al., 2011; Peylin et al., 2013), differences in the two sets of posterior fluxes appear on the largest scales. Where the cyan curves crosses the red curves mark the points where this difference is surpassed by CO_2 differences due to analysis error. For example, in the upper troposphere, the differences produced by using the two different posterior fluxes are always smaller than the uncertainty in CO_2 resulting from a 6 h analysis error (layer 4). But near the surface (layer 1), particularly in July, there are large scale differences resulting from the use of the two different posteriors that exceed the CO_2 errors due to the use of imperfect atmospheric analyses. Thus changes to the CarbonTracker system that produced the 2010 and 2013B posterior fluxes are understandably most evident near the surface where the assimilated observations are located and differences are largest in boreal summer. The difference in CO_2 produced by the two products diminishes with height because both CT2010 and CT2013B assimilated surface observations only.~~

Another view of the spectra can be obtained by summing over total wavenumber and plotting with respect to zonal wavenumber (**Figure 12**~~Figure 12~~) because zonal wavenumber spectra are more indicative of the tropical signal. In this figure, it is evident that there is almost no information retained in the predictability error (i.e. the black curve lies above the blue one). In other words, predictability error spectra surpass that of the reference state at all levels except near the surface for the first few zonal wavenumbers. When a 3 month shift in surface fields is included in the predictability experiment, all predictability is lost in July and August since the power in the predictability error (red curves) exceeds that of the reference

27

state (black curves) for all wavenumbers (**Fig. S8**). Predictability is mostly lost in September as well. Thus using the correct land and ocean surface fields may be relevant for capturing predictability in July, August and September in the tropics. The dominant mode of tropical subseasonal variability is the MJO which is characterized by very large scales (zonal wavenumbers 1 and 2 in wind and rainfall fields) in the tropics (Waliser, 2006) and GEM can capture predictability

5   associated with this mode (Lin et al. 2008) supporting the notion that GEM has predictive skill in the tropics on seasonal scales. However, since predictability was already close to lost even with the correct land and ocean surface (power in predictability error is same order of magnitude as that for reference state), ~~this conclusion may not be warranted~~ the worsened predictability in July to September with incorrect land and ocean surface fields is not a clear indication of influence of their influence in the tropics. Indeed, outside of these three months, shifting the land and ocean surface fields by

10   3 months has little impact (red and black curves are similar in **Fig. S8**).   This suggests that most of the information retained at the largest scales (seen in **Figure 11**~~Figure 11~~) in the predictability experiment is coming from the extratropics or signals with latitudinal structure for these months. Indeed, the CO_2 evolution (not shown) in the reference and predictability experiments differs the most in northern hemisphere extratropics. Thus the large-scale predictability seen in the northern extratropics is due to large CO_2 variability during October to May associated with biospheric fluxes but from June to

15   September, ~~the~~ the ocean and land forcing of GEM's climate is also important. When ~~6 h~~ analysis ~~errors are~~ uncertainty is simulated (red curves), there is a gain of information over predictability error but fewer than 20 zonal waves are resolved in the mid troposphere (**Figure 12**~~Figure 12~~, layers 2 and 3). In the lowest layer, about 40 waves are resolved. Compared to **Figure 11**~~Figure 11~~, there are significantly fewer waves being resolved. Thus structure in the meridional direction is better resolved in analyses than is structure in the zonal direction. ~~Finally, the differences in the two posterior fluxes (CT2010 and~~

20   ~~CT2013B) are smaller than CO_2 differences due to meteorological analysis errors everywhere, in December. In July, only near the surface and only for the smallest few wavenumbers are differences larger than analysis errors (i.e. cyan curve is above the red curve). Thus, we conclude that, except for July near the surface, the two sets of posterior fluxes have retrieved almost identical information in the tropics.~~

      It is also useful to examine other sorts of model errors in terms of their impact on CO_2 predictions. **Figure**

25   **13**~~Figure 13~~ shows the spectra of errors due to the inclus~~addit~~ion (or not) of convective tracer transport (cyan curves). The impact of adding convective tracer transport is primarily at large scales and exceeds CO_2 errors due to imperfect wind analyses (red curves) only in the mid troposphere for wavenumbers less than 5. Spectra for April 2009 were shown in **Figure 13**~~Figure 13~~ because the impact of adding convective tracer transport was largest in spring months. The impact is always less than that due to imperfect wind analyses if zonal wavenumber spectra are considered (bottom row). Thus, for

30   our model, this type of model error is exceeded by transport error due to uncertain meteorological analyses for most spatial scales. This type of analysis in which components of transport error are compared may thus be useful in diagnosing the dominant sources of transport error for a given transport model.

6 Summary and Discussion

A new capability for simulating CO_2 using ECCC's operational weather and environmental prediction models has been developed. The adaptations required for greenhouse gas simulation include the implementation of a global mass fixer for the semi-Lagrangian tracer transport scheme, the implementation of a mixing ratio defined with respect to dry air for tracer variables, the addition of convective tracer transport and slight modification of a parameter in the boundary layer scheme. A sequence of 24 h meteorological forecasts is used to transport CO_2 fields in a forecast cycle involving a coupled meteorological and tracer transport model. Using prescribed posterior fluxes from NOAA's CarbonTracker, the transport of the model has been assessed. The 24 h meteorological forecasts are as similar to the ERA Interim reanalyses as are other reanalyses (MERRA and JRA-55). The CO_2 fields compare well to observations assimilated in the posterior fluxes (surface hourly measurements) as well as to independent observations (TCCON and NOAA aircraft profiles) and to CarbonTracker mole fractions. Synoptic and seasonal timescales are well captured but, as with most transport models, the diurnal cycle amplitude is too low in summer. The vertical gradient in the mid troposphere is excellent but the gradient from 1-3 km does not agree as well with observations and the error in the gradient changes with season.

Using this coupled meteorological and tracer forecast model (called GEM-MACH-GHG), the influence of atmospheric transport and the uncertainty of meteorological analyses on CO_2 predictions was explored. The predictability of CO_2 has been described for the first time. Predictability on weather time scales for CO_2 is 2-3 days in the troposphere, reaching 5 days near the surface and in the lower stratosphere for GEM-MACH-GHG. Reduced predictability in the stratosphere was seen during the prolonged and strong stratospheric sudden warming of January 2009. Predictability of CO_2 is shorter than that for temperature and is comparable to the predictability of the wind fields in the free atmosphere. Near the surface, the predictability of CO_2 is similar to that of the specific humidity field. Predictability on seasonal timescales was seen at the largest scales and is primarily attributed to long timescales in CO_2 fluxes which are in turn partly due to climate predictability. On top of this, three-month errors in land and ocean surface fields are seen to directly impact CO_2 predictability in boreal summer in the northern extratropics. By simulating a 6 h forecast error, meteorological analyses are shown to resolve large scales in CO_2 transport, but there is a spatial scale below which there is no information beyond predictability error. By comparing spectra in terms of total and zonal wavenumbers, it is seen that more information is retained in the CO_2 field in north-south direction as opposed to the zonal direction. The impact of convective tracer transport on CO_2 fields are largest in boreal spring and exceeds errors due to the use of imperfect atmospheric analyses only in the mid troposphere and only for wavenumbers below 5.

By definition, predictability experiments use a reference simulation against which perturbed simulations are compared. Thus the errors obtained here are system dependent but if the system is representative, then results are likely to be representative of other systems. Indeed, being an operational weather forecast model, GEM is routinely evaluated and we have shown that 24 h forecasts with our modified version of GEM are comparable to reanalysis products. On the other hand, seasonal predictability is model dependent and is most likely related to model parameterization of subgrid scale processes,

29

particularly, convection (Shukla and Kinter, 2006). Subseasonal predictability is limited by the fact that most models do not capture the MJO (Waliser, 2006). Even if they do capture it, and the initial state has a strong MJO signal, forecast skill may still not be improved because of the complex interplay between MJO and other modes of variability (Lin et al., 2008). Thus, it may be useful for individual models to be able to characterize CO_2 predictability error, particularly on longer time scales, as well as the spatial scales definable in the presence of imperfect meteorological analyses.

The length of time attributed to predictability on weather time scales (e.g. **Figure** 9) is based on a normalized predictability error and an arbitrary threshold value of 0.8. Because of the arbitrariness of the choice of reference state used for normalization, and the threshold value, the length of time for which a variable remains predictable should not be taken as fundamental. Rather, the important result is the relative time of predictability for the variables. Specifically, the fact that the loss predictability of CO_2 follows that of the wind field rather than the temperature field is expected to be a robust result.

Using our experimental design, we have also compared the spatial scales of the impact of using different posterior fluxes (from CT2010 and CT2013B) to those arising from the impact of imperfect meteorological analyses. The information of observations is present in the posterior fluxes but since both sets of posteriors used very similar observing systems, little difference was seen in the resulting CO_2 simulations. However, if prior fluxes are used to define a reference simulation then various posterior fluxes could be used to define flux analysis increments. Specifically, the spatial scales resolved by various observing systems (e.g. GOSAT or OCO-2 versus the surface network) could be computed and compared to the scales resolvable in the context of imperfect meteorological analyses. Indeed, such experiments are in progress and will be described in a subsequent article.

The impact of uncertain meteorological analyses on the spatial scales resolved in CO_2 simulations was presented here in the context of a coupled meteorological and tracer forecast model. With an offline transport model, the impact of uncertain atmospheric analyses could also be considered. For example, when integrating posterior fluxes to get CO_2 concentrations, two simulations could be performed: one with the correct analyses and another with analyses shifted by 6 h. In addition, the fluxes retrieved using the correct atmospheric analyses could be compared to those retrieved using a series of temporally shifted atmospheric analyses. The difference in the retrieved fluxes in these experiments might then shed light on the magnitude of the impact of transport error due to atmospheric analyses on retrievals.

The impact of uncertain meteorological analyses on CO_2 simulations was assessed using a 6 h shift in analysis states as a proxy for 6 h analysis errors. This was done because such estimates of meteorological analysis errors are not available but a perturbation of the size and shape of a 6 h analysis error was desired. While other proxies for 6 h analysis errors could be devised, none would be any more valid. On the other hand, it is possible to directly obtain analysis and forecast errors by implementing an ensemble Kalman filter for CO_2 state estimation. Indeed, a greenhouse gas data assimilation system based on an augmented state (meteorology, constituent, and fluxes) ensemble Kalman filter is now under development. This new system is called EC-CAS (ECCC Carbon Assimilation System) and also uses existing tools developed at ECCC, namely, the operational global ensemble prediction system (Houtekamer et al., 2014). With EC-CAS, the atmospheric modulation of CO_2 forecast uncertainty would be directly simulated and the impact on flux estimate uncertainties could then be determined.

A new capability for simulating CO_2 using ECCC's operational weather and environmental prediction models has been developed. The adaptations required for greenhouse gas simulation include the implementation of a global mass fixer for the semi-Lagrangian tracer transport scheme, the implementation of a mixing ratio defined with respect to dry air for tracer variables, the addition of convective tracer transport and slight modification of a parameter in the boundary layer scheme. A sequence of 24 h meteorological forecasts is used to transport CO_2 fields in a forecast cycle involving a coupled meteorological and tracer transport model. The 24 h meteorological forecasts are as similar to the ERA Interim reanalyses as are other reanalyses (MERRA and JRA-55). That means that throughout the sequence of 24 h forecasts, the meteorological uncertainty is comparable to that of a reanalysis dataset which could be used to constrain an offline transport model. Using prescribed posterior fluxes from NOAA's CarbonTracker (CT2013B), the transport of the model has been assessed. The CO_2 fields compare well to observations assimilated in the posterior fluxes (surface hourly measurements) as well as to independent observations (within 2 ppm for TCCON and NOAA aircraft profiles) and to CarbonTracker mole fractions (within 3 ppm in column means). Synoptic and seasonal timescales are well captured but, as with most transport models, the diurnal cycle amplitude is too low in summer. The vertical gradient in the mid troposphere is slightly overestimated but the gradient from 1-3 km does not agree as well with observations and the error in the gradient changes with season.

The predictability of CO_2 concerns forecast uncertainty. CO_2 forecasts or predictions are implicitly used in flux inversions when computing the model-data mismatch. The forecast or prediction error of a transport model is often referred to as "transport error" and transport error is a significant component of posterior flux errors from flux inversions. Transport error is comprised of meteorological state, CO_2 state, flux and model formulation errors. While considerable research has been devoted to better understanding the nature of model formulation and flux errors on CO_2 predictability (or CO_2 transport error), much less is known about the nature of meteorological state errors and their impact on CO_2 transport errors. Liu et al. (2011) have shown that transport errors due to meteorological uncertainty are 1.2-3.5 ppm near the surface and 0.8-1.8 ppm in the column mean. Here we examine the atmospheric processes and scales implicated in such impacts. Of course, flux and model formulation errors are a significant component of transport error and the various contributions to transport error do not act in isolation. However, isolating this impact provides new insight into transport error processes. Moreover, it permits the identification of the spatial scales of CO_2 that can be trusted in the context of meteorological analysis uncertainties.

The fact that weather becomes unpredictable beyond 2 weeks has implications for CO_2 transport. With an experimental design developed to isolate the impact of weather predictability due to imperfect initial conditions on CO_2, we have demonstrated that CO_2 also has limited predictability. That is, uncertainty in meteorological initial conditions is evident in CO_2 simulations when other sources of transport error are removed. This is a newly identified source of uncertainty which is primarily of pedagogical interest because transport models always have updated information from meteorological analyses. However, isolating this error process leads to insights into the processes which couple CO_2 and meteorological errors. Moreover, this type of error forms the upper limit of transport errors due to uncertain meteorological analyses. The predictability of CO_2 due to uncertain meteorological initial conditions is found to be short (just a few days) and is comparable to that of the wind field during this time. After a few days, CO_2 loses predictability faster than the wind

31

field. Predictability is greatest near the surface and in the lower stratosphere. Reduced predictability in the stratosphere was seen during the prolonged and strong stratospheric sudden warming of January 2009. Predictability of CO_2 is shorter than that for temperature at all levels but is similar to that of the specific humidity field for the first few days, after which CO_2 becomes less predictable than moisture. The predictability measure used is based on zonal variability. In addition, the measure is normalized by the variability of the initial state so that loss of predictability by different variables could be compared. Since there is an arbitrariness in the choice of the normalization state, and in choosing a threshold to define predictable states, absolute times for predictability cannot be obtained. Rather, it is the relative predictability that is expected to be more reliable. Specifically, the greater loss of predictability for CO_2 than the temperature, humidity or wind fields (the latter after a few days) and the increased predictability of CO_2 near the surface and lower stratosphere compared to that in the free troposphere are the new results. While these results are model specific (as are predictability results in general) and metric specific, there are reasons to expect more generality of the results. Increased predictability of CO_2 near the surface can be explained by the constraint of fluxes near the bottom boundary. Increased predictability of CO_2 in the lower stratosphere makes sense because of dominance of slow rotation motions in that layer of the atmosphere. Nevertheless, these results need to be confirmed by other models and metrics. Massart et al. (2016) have computed the predictability of XCO2 forecasts obtained from analyses which assimilated GOSAT observations and found predictability up to 5 days globally. Results were not produced for CO_2 at various altitudes nor compared with those for meteorological variables. However the predictability of XCO2 was computed for different regions. Predictability was greater in the tropics than in the extratropics. The measure used was an anomaly correlation wherein forecasts anomalies with respect to a climatology are correlated with analysis anomalies. This measure of predictability is useful for identifying predictability of a given forecast system at specific dates and forecast ranges. However, as with all metrics there are known deficiencies. In particular, the anomaly correlation can be optimistic where observation density is poor because forecasts and analysis anomalies will be perfectly correlated in the extreme case of no observations. Thus, the predictability results of Massart et al. (2016) are not directly comparable with those obtained here.

Even with no information from meteorological analyses, the predictability of CO_2 was seen on seasonal timescales for very large spatial scales and was found to be primarily due to surface fluxes. However, worsened predictability of CO_2 was seen on these scales in the northern extratopics in boreal summer if the surface atmospheric forcings were incorrectly specified. In other words, the land and ocean surface also play a role in CO_2 predictability on seasonal timescales. The proposed mechanism is that the land and ocean surface perturb atmospheric circulations which transport CO_2 and the impact is greatest where CO_2 gradients are largest: in the northern midlatitudes in boreal summer.

Our predictability error is defined by the error due only to imperfect meteorological initial conditions. Thus compared to a reference cycle, the perturbed cycle has no updates of meteorology with analyses. The more realistic situation of imperfect meteorological information was addressed by perturbing the meteorological analyses used in a reference cycle with errors. Ideally, meteorological analysis errors should be used but as these are not available for our system during the time period of interest, a proxy was used. The proxy is based on the closest available analysis (i.e. one available 6 h prior to

that used in the reference cycle) so perturbations involved a difference of analyses. The diurnal cycle was removed from the perturbations by first removing the monthly mean of each synoptic hour from the analyses. Since the resulting perturbations include the time evolution of the meteorology during 6 h as well as analysis errors, such a perturbation is expected to be larger than that due to an actual analysis error. By decomposing the departure of the CO_2 evolution in the perturbed cycle from that in the reference cycle into spherical harmonics, the spatial scales of the differences were computed. It is seen that the spatial spectra of CO_2 differences due to imperfect meteorological analyses exceeds that due to the CO_2 state itself for some scales. Thus, for spatial scales smaller than this crossover point, CO_2 is not predictable, simply due to the presence of meteorological uncertainties. The spatial scales so identified are specific to our model's domain and resolution and the monthly time scale used for averaging spectra. The spectra of the difference due to imperfect meteorology are seen to asymptote to the predictability error spectra, meaning that for small enough spatial scales, the CO_2 field is seeing no updated information from the meteorology at all. Thus, transport error is impacted by meteorological state errors with the error being greatest for the smallest spatial scales. By comparing spectra in terms of total and zonal wavenumbers, it is seen that more information is retained in the CO_2 field in north-south direction as opposed to the zonal direction. Note that the spatial scales identified as predictable depend on the choice of fluxes used in both reference and perturbed simulations.

These experiments demonstrate that the predictability of CO_2 is limited by the presence of meteorological analysis errors as well as flux and model errors. The CO_2 observing system determines the spatial scales that can be resolved in CO_2 state estimate and in CO_2 flux estimates. However, the meteorological observing system also imposes limits on the spatial scales that can be resolved. Here these limits are found to be well below those imposed by the CO_2 observing system. However, the unresolved spatial scales are shown to increase with altitude. As the CO_2 observation density increases (for example with satellite data from GOSAT, OCO-2 and future planned missions) knowing these limits will become increasingly important.

Other components of transport error can also be isolated and compared. Considerable work has been done on identifying the contribution of model formulation errors to transport errors. Looking at the CO_2 spatial scales resolved by different model configurations provides a new tool for identifying the most important model errors. As an example, a model error was introduced by removing the transport of tracers through deep convection in a perturbed CO_2 simulation. The impact of convective tracer transport on CO_2 fields is seen to be largest in boreal spring but exceeds errors due to the use of imperfect atmospheric analyses only in the troposphere and only for wavenumbers below 5. Thus for all but wavenumbers below 5, transport errors due to imperfect meteorology are larger than those due to the lack of convective tracer transport. In other words, by comparing components of transport error, the limiting source of error for a given spatial scale can be determined. An important application of this diagnostic is as follows. With realistic flux perturbations, transport errors due to flux uncertainties could be compared to transport error due to meteorological state uncertainties. In doing so, the spatial scales of transport error dominated by flux errors and those dominated by meteorological state errors could be identified. Alternatively, if prior fluxes are used to define a reference simulation then various posterior fluxes could be used to define flux analysis increments. Specifically, the spatial scales resolved by various observing systems (e.g. GOSAT or OCO-2

33

versus the surface network) could be computed and compared to the scales resolvable in the context of imperfect meteorological analyses. Indeed, such experiments are in progress and will be described in a subsequent article.

By definition, predictability experiments use a reference simulation against which perturbed simulations are compared. Thus the errors obtained here are system dependent but if the system is representative, then results are likely to be representative of other systems. Indeed, being an operational weather forecast model, GEM is routinely evaluated and we have shown that 24 h forecasts with our modified version of GEM are comparable to reanalysis products. On the other hand, seasonal predictability is model dependent and is most likely related to model parameterization of subgrid scale processes, particularly, convection (Shukla and Kinter, 2006). Subseasonal predictability is limited by the fact that most models do not capture the MJO (Waliser, 2006). Even if they do capture it, and the initial state has a strong MJO signal, forecast skill may still not be improved because of the complex interplay between MJO and other modes of variability (Lin et al., 2008). Thus, it may be useful for individual models to be able to characterize CO_2 predictability error, particularly on longer time scales, as well as the spatial scales definable in the presence of imperfect meteorological analyses.

In this work, the focus of predictability (transport) error assessment was on the random component. This is because the bias of the transport error was found to be small compared to the random component for all timescales considered here. This occurs because of the use of a control simulation when defining errors. That is, model formulation and representativeness errors are absent from such error calculations. In addition, we isolate various contributions to transport error by holding fixed other sources of error such as flux error. In reality, when comparing CO_2 model predictions to measurements all of these error sources are present and the mean error (or bias) is an important concern. Thus our methodology of isolating components of transport error provides insight into the component that is largest for a given spatial scale, may not shed light on biases that develop when all of these errors interact.

A limitation of this study is that the impact of uncertain meteorological analyses on CO_2 simulations was assessed using a 6 h shift in analysis states as a proxy for 6 h analysis errors. This was done because such estimates of meteorological analysis errors are not available but a perturbation of the approximate size and shape of a 6 h analysis error was desired. While other proxies for 6 h analysis errors could be devised, none would be any more valid. On the other hand, it is possible to directly obtain analysis and forecast errors by implementing an ensemble Kalman filter for CO_2 state estimation. Indeed, a greenhouse gas data assimilation system based on an augmented state (meteorology, constituent, and fluxes) ensemble Kalman filter is now under development. This new system is called EC-CAS (ECCC Carbon Assimilation System) and also uses existing tools developed at ECCC, namely, the operational global ensemble prediction system (Houtekamer et al., 2014). With EC-CAS, the atmospheric modulation of CO_2 forecast uncertainty would be directly simulated and the impact on flux estimate uncertainties could then be determined.

35

Every 24 h when a new surface pressure analysis becomes available, there will be a sudden change in local and global mean surface pressure when the 24 h forecast is replaced by the analysis. Here we derive the scheme used to ensure that the tracer mass is not affected by this change in surface pressure. First we define some global mass quantities. The global dry air mass is

$$M^a_{dry-air} = \frac{1}{g} \sum_{i,j,k} \Delta p^a_{i,j,k} \left(1 - q^a_{i,j,k}\right) \Delta A_{i,j} \quad , \tag{A1}$$

the global mass of tracer in the forecast is

$$M^f_c = \frac{1}{g} \sum_{i,j,k} c^f_{i,j,k} \Delta p^f_{i,j,k} \left(1 - q^f_{i,j,k}\right) \Delta A_{i,j} \tag{A2}$$

and the global mass of tracer after adjustment is given by

$$M^a_c = \frac{1}{g} \sum_{i,j,k} c^a_{i,j,k} \Delta p^a_{i,j,k} \left(1 - q^a_{i,j,k}\right) \Delta A_{i,j} \quad . \tag{A3}$$

Here i,j,k are longitude, latitude and vertical grid indices, $\Delta A_{i,j}$ is the area of grid box (i,j), $\Delta p_{i,j,k}$ is the vertical pressure difference across the grid box at level k, $c_{i,j,k}$ is the mass mixing ratio with respect to dry air at the center of grid box (i,j,k), q is specific humidity and g is the gravitational constant. The superscripts a and f refer to the analysis and forecast, respectively. We seek a spatially invariant adjustment (ε) to the tracer mixing ratio:

$$c^a_{i,j,k} = c^f_{i,j,k} + \varepsilon \quad . \tag{A4}$$

The adjustment parameter (ε) is determined from the constraint that the global adjusted tracer mass equal that of the forecast tracer. (A4) has the nice property of exactly conserving spatial gradients. For a tracer with large background value (like CO_2) this adjustment is much smaller than field itself, and is neglegible when comparing to analysis errors. However, this additive adjustment scheme may have undesirable effects for tracers with a large dynamic range. Substituting (A4) into (A3) yields

$$M^a_c = M^*_c + \varepsilon M^a_{dry-air} \tag{A5}$$

where

$$M^*_c = \frac{1}{g} \sum_{i,j,k} c^f_{i,j,k} \Delta p^a_{i,j,k} \left(1 - q^a_{i,j,k}\right) \Delta A_{i,j} \quad . \tag{A6}$$

Solving for ε with the constraint that $M^a_c = M^f_c$ yields

$$\varepsilon = \frac{M^f_c - M^*_c}{M^a_{dry-air}} \quad . \tag{A7}$$

36

For a fair comparison to TCCON observations, GEM-MACH-GHG XCO_2 simulations are smoothed using the TCCON a priori profiles and averaging kernels to account for the sensitivity of the measurements. The vertical column (VC) of CO_2 is defined as:

$$VC_{CO_2} = \sum_{i=1}^{N} \frac{f_{CO_2,i}}{g\, m_{air,i}} \Delta p_i \qquad (B1)$$

with N being the number of layers in the profile, $f_{CO2,i}$ the mean mole fraction in the ith layer of pressure thickness Δp_i and mean molecular weight of air $m_{air,i}$. g is the Earth acceleration due to gravity and is kept constant (g=9.80616 m.s^{-2}). GEM-MACH-GHG produces profiles of CO_2 dry mole fraction $f_{CO_2}^{dry} = \frac{f_{CO_2}}{1-f_{H_2O}}$ as well as specific humidity $q = \frac{[kg_{H_2O}]}{[kg_{air}]}$. Considering $\frac{1-f_{H_2O}}{m_{air}} = \frac{1-q}{m_{air}^{dry}}$ we can express (B1) as:

$$VC_{CO_2} = \sum_{i=1}^{N} \frac{(1-f_{H_2O,i})f_{CO_2,i}^{dry}}{g\, m_{air,i}} \Delta p_i = \sum_{i=1}^{N} \frac{(1-q_i)f_{CO_2,i}^{dry}}{g\, m_{air,i}^{dry}} \Delta p_i \qquad (B2)$$

and the vertical column weighted by the TCCON column averaging kernels is :

$$VC_{CO_2,a} = \sum_{i=1}^{N} \frac{(1-q_i)f_{CO_2,i}^{dry}\, a_i}{g\, m_{air,i}^{dry}} \Delta p_i \; . \qquad (B3)$$

The column-averaged dry mole fraction XCO_2 is the ratio of the vertical column of CO_2 and the vertical column of dry air. Finally the smoothed GEM-MACH-GHG XCO_2 is obtained as:

$$X_{CO_2}^{model} = X_{CO_2}^{a\,priori} + \frac{VC_{CO_2,a}^{model} - VC_{CO_2,a}^{a\,priori}}{VC_{air}^{dry}} \; . \qquad (B4)$$

GEM-MACH-GHG CO_2 profile simulations extend from 2009 to 2010 with an output frequency of 15 min. Average hourly XCO_2 is considered for the comparisons and all TCCON sites with observations in 2009-2010 are used. There are both 120HR and 125HR measurements at Lauder in 2009-2010, with the 125HR dataset starting 2 February 2010; only the 125HR data is used. Several statistical parameters are derived: N is the number of pairs (hours) for which there are TCCON measurements. The Bias is the average difference between the model and TCCON:

$$Bias = \frac{\sum_{i=1}^{N}\left(X_{CO_2}^{model} - X_{CO_2}^{TCCON}\right)_{(i)}}{N} \; . \qquad (B5)$$

The root-mean-square (RMS) of the differences is :

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}\left(X_{CO_2}^{model}-X_{CO_2}^{TCCON}\right)^2_{(i)}}{N}} \quad . \tag{B6}$$

The Scatter is the standard deviation of the differences :

$$Scatter = \sqrt{\frac{\sum_{i=1}^{N}[\left(X_{CO_2}^{model}-X_{CO_2}^{TCCON}\right)_{(i)}-Bias]^2}{N-1}} \quad . \tag{B7}$$

The mean bias is, with N_s the number of stations :

$$Bias_{mean} = \frac{1}{N_S}\sum_{i=1}^{N_S} Bias_{(i)} \quad . \tag{B8}$$

The standard deviation of a station 's bias is useful for estimating the variability of the bias from station to station :

$$SD = \sqrt{\frac{\sum_{i=1}^{N_S}(Bias_{(i)}-Bias_{mean})^2}{N_S-1}} \quad . \tag{B9}$$

Finally, R is the Pearson's correlation coefficient :

$$R = \frac{\sum_{i=1}^{N}(X_{CO_2\ (i)}^{model}-X_{CO_2\ mean}^{model})(X_{CO_2\ (i)}^{TCCON}-X_{CO_2\ mean}^{TCCON})}{\sqrt{\sum_{i=1}^{N}(X_{CO_2\ (i)}^{model}-X_{CO_2\ mean}^{model})^2}\sqrt{\sum_{i=1}^{N}(X_{CO_2\ (i)}^{TCCON}-X_{CO_2\ mean}^{TCCON})^2}} \quad . \tag{B10}$$

References

Agustí-Panareda, A., Massart, S., Chevallier, F., Boussetta, S., Balsamo, G., Beljaars, A., Ciais, P., Deutscher, N. M., Engelen, R., Jones, L., Kivi, R., Paris, J.-D., Peuch, V.-H., Sherlock, V., Vermeulen, A. T., Wennberg, P. O., and Wunch, D.: Forecasting global atmospheric CO2, Atmos. Chem. Phys., 14, 11959-11983, doi:10.5194/acp-14-11959-2014, 2014.

Agusti-Panareda, A., Diamantakis, M., Bayona, V., Klappenbach, F., and Butz, A.: Improving the inter-hemispheric gradient of total column atmospheric CO_2 and CH_4 in simulations with the ECMWF semi-Lagrangian atmospheric global model, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-143, in review, 2016.

Aliabadi, A.A, Staebler, R.M, de Grandpré, J., Zadra, A.,Vaillancourt P.A.: Comparison of Estimated Atmospheric Boundary Layer Mixing Height in the Arctic and Southern Great Plains under Statically Stable Conditions: Experimental and Numerical Aspects, Atmosphere-Ocean, 54:1, 60-74, 2016.

Andrews, D. G., Holton, J. R. and Leovy, C. B.: Middle Atmosphere Dynamics, Academic Press, 1987.

Baker, D. F., et al.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988 – 2003, Global Biogeochem. Cycles, 20, GB1002,doi:10.1029/2004GB002439, 2006a.

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

5   Baker, W. E., Atlas, R.,Cardinali, C., Clement, A., Emmitt, G. D., Gentry, B. M., Hardesty, R. M., Källén, E., Kavaya, M.J., Langland, R., Ma, Z., Masutani, M., McCarty, W., Pierce, R.B., Pu, Z., Riishojgaard, L.P., Ryan, J., Tucker, S., Weissmann, M., and Yoe, J. G.: Lidar-Measured Wind Profiles: The Missing Link in the Global Observing System. Bull. Amer. Meteor. Soc., 95, 543–564, doi: http://dx.doi.org/10.1175/BAMS-D-12-00164.1, 2014.

Baldwin, M.P., Gray, L.J., Dunkerton, T.J., Hamilton, K., Haynes, P.H., Randel, W.J., Holton, J. R ., Alexander, M.J.,

10 Hirota, I., Horinouchi, T., Jones, D.B.A., Kinnersley, J.S., Marquardt, C., Sato, K., Takahashi, M.: The quasi-biennial oscillation, Rev. Geophys., 39(2), 179–229, doi:10.1029/1999RG000073, 2001.

Bélair, S., Mailhot, J., Strapp, J.W., and MacPherson, J.I.: An Examination of Local versus Nonlocal Aspects of a TKE-Based Boundary Layer Scheme in Clear Convective Conditions, J. Appl. Meteor., 38, 1499–1518, doi: http://dx.doi.org/10.1175/1520-0450(1999)038<1499:AEOLVN>2.0.CO;2 , 1999.

15 Bélair, S., Louis-Philippe, C., Mailhot, J., Bilodeau, B., and Delage, Y.: Operational Implementation of the ISBA Land Surface Scheme in the Canadian Regional Weather Forecast Model. Part I: Warm Season Results, J. Hydrometeor, 4, 352–370. doi: http://dx.doi.org/10.1175/1525-7541(2003)4<352:OIOTIL>2.0.CO;2 , 2003a.

Bélair, S., Brown, R., Mailhot, J., Bilodeau, B., and Crevier, L.-P.: Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results, J. Hydrometeor., 4, 371–386,

20 doi:10.1175/1525-7541(2003)4,371:OIOTIL.2.0.CO;2, 2003b.

Bélair, S., Maihot, J., Girard, C. and Vaillancourt, P.: Boundary-layer and shallow cumulus clouds in a medium-range forecast of a large-scale weather system. Mon. Weather Rev. 133: 1938–1960, 2005.

Benoit, R., Côté, J. and Mailhot, J.: Inclusion of a TKE Boundary Layer Parameterization in the Canadian Regional Finite-Element Model, Mon. Wea. Rev., 117, 1726–1750. doi: http://dx.doi.org/10.1175/1520-

25 0493(1989)117<1726:IOATBL>2.0.CO;2 , 1989.

Bermejo, R. and Conde, J.: A conservative quasi-monotone semi-Lagrangian scheme, Mon. Weather Rev. 130, 423–430, 2002.Bermejo, R. and Staniforth, A.: The Conversion of Semi-Lagrangian Advection Schemes to Quasi-Monotone Schemes, Mon. Wea. Rev., 120, 2622-2632, 1992.

Bloom, S., Takacs, L., DaSilva, A., and Ledvina, D.: Data assimilation using incremental analysis updates, Mon. Wea.

30 Rev., 124: 1256-1271, 1996.

Blumenstock, T., Hase, F., Schneider, M., García, O. E., and Sepúlveda, E.: TCCON data from Izana, Tenerife, Spain, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.izana01.R0/1149295, 2014.

Boer, G.J.: Homogeneous and Isotropic Turbulence on the Sphere, J. Atmos. Sci., 40, 154-163, 1983.

Bruhwiler, L. M. P., Michalak, A. M., and Tans, P. P.: Spatial and temporal resolution of carbon flux estimates for 1983–2002, Biogeosciences, 8, 1309-1331, doi:10.5194/bg-8-1309-2011, 2011.

Buehner, M., McTaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., Lapalme, E., Laroche, S., Macpherson, S.R., Morneau, J., and Zadra, A.: Implementation of Deterministic Weather Forecasting Systems based on Ensemble-Variational Data Assimilation at Environment Canada. Part I: The Global System, Mon. Wea. Rev., 143, 2532-2559, doi: http://dx.doi.org/10.1175/MWR-D-14-00354.1, 2015.

Campbell, L. J. and Shepherd, T.G.: Constraints on Wave Drag Parameterization Schemes for Simulating the Quasi-Biennial Oscillation. Part I: Gravity Wave Forcing. J. Atmos. Sci., 62, 4178–4195. doi: http://dx.doi.org/10.1175/JAS3616.1, 2005.

Caron, J.-F., Milewski, T., Buehner, M., Fillion, L., Reszka, M., Macpherson, S., and St-James, J.: Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part II: The regional system. Mon. Wea. Rev., 143, 2560–2580, doi:10.1175/MWR-D-14-00353.1, 2015.

Chan, D., Ishizawa, M., Higuchi, K., Maksyutov, S. and Chen, J.: Seasonal CO2 rectifier effect and large-scale extratropical atmospheric transport, J. Geophys. Res., 113, D17309, doi:10.1029/2007JD009443, 2008.

Chan, D., Yuen, C. W., Higuchi, K., Shashkov, A., Liu, J., Chen, J., Worthy, D.: On the CO2 exchange between the atmosphere and the biosphere: the role of synoptic and mesoscale processes. Tellus B, 56. doi:http://dx.doi.org/10.3402/tellusb.v56i3.16424, 2004.

Charron, M., Polavarapu, S., Buehner, M., Vaillancourt, P.A., Charette, C., Roch, M., Morneau, J., Garand, L., Aparicio, J. M., MacPherson, S., Pellerin, S., St-James, J. and Heilliette, S.: The Stratospheric Extension of the Canadian Global Deterministic Medium-Range Weather Forecast System and Its Impact on Tropospheric Forecasts. Mon. Weather Rev. 140, 1924–1944, 2012.

Chevallier, F., Feng, L., Bösch, H., Palmer, P.I. and Rayner: On the impact of transport model errors for the estimation of CO2 surface fluxes from GOSAT observations, Geophys. Res. Lett., 37, L21803, doi:10.1029/2010GL044652, 2010.

Chevallier, F., Palmer, P.I., Feng, L., Boesch, H., O'Dell, C.W. and Bousquet, P.: Toward robust and consistent regional CO2 flux estimates from in situ and spaceborne measurements of atmospheric CO2, Geophys. Res. Lett., 41, 1065–1070, doi:10.1002/2013GL058772, 2014.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R.B., Piao S., and Thornton, P.,: Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation, Mon. Wea. Rev., 126, 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2., 1998a.

40

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M. and Staniforth, A.: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results, Mon. Wea. Rev., 126, 1397–1418, doi:10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2, 1998b.

Dacre, H. F., Gray, S. L. and Belcher, S. E.: A case study of boundary layer ventilation by convection and coastal processes, J. Geophys. Res., 112, D17106, doi:10.1029/2006JD007984, 2007.

Daley, R. Atmospheric Data Analysis, Cambridge University Press, 1991.

Dee, D. P., and 35 co-authors: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quart. J. R. Meteorol. Soc., 137, 553-597, DOI: 10.1002/qj.828, 2011.

de Grandpré, J., Tanguay, M., Qaddouri, A., Zerroukat, M., and McLinden, C.A.: Semi-Lagrangian Advection of Stratospheric Ozone on a Yin-Yang Grid System, Mon.Wea.Rev., 144, 1035-1050, 2016.

Deng, F., Jones, D. B. A., O'Dell, C., Nassar, R. and Parazoo, N. C.: Combining GOSAT XCO2 observations over land and ocean to improve regional CO2 flux estimates, J. Geophys. Res. Atmos., 121, doi:10.1002/2015JD024157, 2016.

Denning, A. S., Fung, I. Y. and Randall, A. D.: Latitudinal gradient of atmospheric CO2 due to seasonal exchange with land biota. Nature, 376, 240–243, 1995.

Deutscher, N., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., Grupe, P., and Katrynski, K.: TCCON data from Bialystok, Poland, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.bialystok01.R0/1149277, 2014.

Diamantakis, M. and Flemming, J.: Global mass fixer algorithms for conservative tracer transport in the ECMWF model, Geosci. Model Dev., 7, 965-979, doi:10.5194/gmd-7-965-2014, 2014.

Eerola, K., Twenty-one years of verification from the HIRLAM NWP system, Wea. Forecast., 28, 270-285, doi:10.1175/WAF-D-12-00068.1, 2013.

Fillion, L., Tanguay, M., Lapalme, E., Denis, B., Desgagné, M., Lee, V., Ek, N., Liu, Z., Lajoie, M., Caron, J.-F., Pagé, C.: The Canadian regional data assimilation and forecasting system. Weather Forecast. 25, 1645-1669, 2010.

Flemming, J. and Huijnen, V.: IFS Tracer Transport Study, MACC (Monitoring Atmospheric Composition and Climate) Deliverable G-RG WP 4.2. Available from https://www.google.ca/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0CBwQFjAAahUKEwjY2eCY5PzIAhXCox4KHbePBbo&url=https%3A%2F%2Fwww.gmes-atmosphere.eu%2Fdocuments%2Fdeliverables%2Fg-rg%2Fifs_transport_study.pdf&usg=AFQjCNFS_Zd6GP5s5oJgtkJFGGaOrM4AHw&bvm=bv.106923889,d.dmo , 2011.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, J. Clim., 19, 14, 3337-3353, DOI: http://dx.doi.org/10.1175/JCLI3800.1, 2006.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K. and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, J. Clim., 27, 511-526, DOI: http://dx.doi.org/10.1175/JCLI-D-12-00579.1, 2014.

Gallus Jr., W. A., Bresch, J. F., Comparison of impacts of WRF dynamic core, physics package, and initial conditions on

5   warm season rainfall forecasts, Mon. Wea. Rev., 134, 2632-2641, doi: 10.1175/MWR3198.1, 2006.

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., and Morneau, J.: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. Mon. Wea. Rev., 135, 2339–2354, 2007.

Girard, C., Plante, A., Desgagné, M., McTaggart-Cowan, R., Côté, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P. A., Zadra, A.: Staggered Vertical Discretization of the

10   Canadian Environmental Multiscale (GEM) Model Using a Coordinate of the Log-Hydrostatic-Pressure Type. Mon. Wea. Rev., 142, 113, DOI: 10.1175/MWR-D-13-00255.1, 2014.

Griffith, D. W. T., Deutscher, N., Velazco, V. A., Wennberg, P. O., Yavin, Y., Aleks, G. K., Washenfelder, R., Toon, G. C., Blavier, J.-F., Murphy, C., Jones, N., Kettlewell, G., Connor, B., Macatangay, R., Roehl, C., Ryczek, M., Glowacki, J., Culgan, T., and Bryant, G.: TCCON data from Darwin, Australia, Release GGG2014R0, TCCON data archive, hosted by the

15   Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.darwin01.R0/1149290, 2014a.

Griffith, D. W. T., Velazco, V. A., Deutscher, N., Murphy, C., Jones, N., Wilson, S., Macatangay, R., Kettlewell, G., Buchholz, R. R., and Riggenbach, M.: TCCON data from Wollongong, Australia, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee,

20   USA, doi:10.14291/tccon.ggg2014.wollongong01.R0/1149291, 2014b.

Gurney, K. R., et al.: Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature, 415, 626– 630, 2002.

Gurney, K. R., et al. : TransCom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, Global Biogeochem. Cycles, 18, GB1010, doi:10.1029/2003GB002111, 2004.

25   Hase, F., Blumenstock, T., Dohe, S., Groß, J., and Kiel, M.: TCCON data from Karlsruhe, Germany, Release GGG2014R1, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.karlsruhe01.R1/1182416, 2014

Hines, C. O.: Doppler-spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 1: Basic formulation, J. Atmos. Sol.-Terr. Phys., 59, 371–386, 1997a.

30   Hines, C. O.: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 2: Broad and quasi monochromatic spectra, and implementation. J. Atmos. Sol.-Terr. Phys., 59, 387–400, 1997b.

Holtslag, A. A. M. and Boville, B., A.: Local versus Non-Local Boundary Layer Diffusion in a Global Climate Model, J. Climate, 6:1825-1842, 1993.

Holtslag, A. A. M.: Atmospheric Boundary Layer and Air Pollution, Modeling and Parameterization, in Encyclopedia of Atmospheric Sciences, Vol. 2, Academic Press, Elsevier, http://dx.doi.org/10.1016/B978-0-12-382225-3.00087-6 , 265-273, 2015.

Houtekamer, P.L., Deng, X., Mitchell, H.L., Baek, S.-J. and Gagnon, N.: Higher Resolution in an Operational Ensemble Kalman Filter, Mon. Wea. Rev., 2014, 142, 1143-1162, 2014.

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 CO_2 sources and sinks using satellite measurements, Atmos. Chem. Phys., 10, 9981-9992, doi:10.5194/acp-10-9981-2010, 2010.

Jöckel, P., von Kuhlmann, R., Lawrence, M. G., Steil, B., Brenninkmeijer, C.A.M., Crutzen, P. J., Rasch, P. J. and Eaton, B.: On a fundamental problem in implementing flux-form advection schemes for tracer transport in 3-dimensional general circulation and chemistry transport models, Q. J. R. Meteorol. Soc., 127, 1035-1052, 2001.

Kain, J.S., and Fritsch, J.M.: A one-dimensional entraining/detraining plume model and its application in convective parameterizations, J. Atmos. Sci. 47, 2784-2802, 1990.

Kain, J.S.: The Kain-Fritsch convective parameterization: an update, J. Appl. Meteorol. 43, 170-181, 2004.

Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T. and Ide ,K.: "Variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation, J. Geophys. Res., 116, D09110, doi:10.1029/2010JD014673, 2011.

Kang, J.-S., Kalnay, E., Miyoshi, T., Liu, J. and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, J. Geophys. Res., 117, D24101, doi:10.1029/2012JD018259, 2012.

Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., and Meijer, H.A.: Atmospheric CO_2 and $13CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications. In: A History of Atmospheric CO_2 and Its Effects on Plants, Animals, and Ecosystems [J. R. Ehleringer, T. E. Cerling and M. D. Dearing (eds.)]. Springer Science+Business Media, New York, NY, USA, and Heidelberg, Germany, pp. 83–113, 2005.

Kivi, R., Heikkinen, P., and Kyro, E.: TCCON data from Sodankyla, Finland, Release GGG2014R0., TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.sodankyla01.R0/1149280, 2014.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Japan, 93, 5-48, doi: 10.2151/jmsj.2015-001, 2015.

Köhler M, Ahlgrimm M, Beljaars A: Unified treatment of dry convective and stratocumulus topped boundary layers in the ECMWF model, Q. J. R. Meteorol. Soc., 137, 43–57, DOI:10.1002/qj.713, 2011.

Koshyk, J.N., Boville, B.A., Hamilton, K., Manzini, E., and Shibata, K.: Kinetic energy spectrum of horizontal motions in middle atmosphere models, J. Geophys. Res., 104, D22, 27,177-27,190, 1999.

43

Kuo, H.L., 1974: Further studies on the parameterization of the influence of cumulus convection on large-scale flow. J. Atmos. Sci., 31, 1232-1240.

Lagouvardos, K., Kotroni, V., Koussis, A., and Feidas, F., The meteorological model BOLAM at the National Observatory of Athens: Assessment of two-year operational use, J. Appl. Meteor., 42, 1667-1678, doi: 10.1175/1520-0450(2003)042<1667:TMMBAT>2.0.CO;2, 2003.

Law, R. M., Rayner, P. J., Denning, A. S., Erickson, D., Fung, I. Y., Heimann, M., Piper, S. C., Romonet, M., Taguchi, S., Taylor, J. A., Trudinger, C. M., and Watterson, I. G.: Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO2 inversions, Global Biogeochem. Cy., 10(4), 783–796, 1996.

Law, R. M., et al.: TransCom model simulations of hourly atmospheric CO2: Experimental overview and diurnal cycle results for 2002, Global Biogeochem. Cycles, 22, GB3009, doi:10.1029/2007GB003050, 2008.

Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N.: Global Carbon Budget 2015, Earth Syst. Sci. Data, 7, 349-396, doi:10.5194/essd-7-349-2015, 2015.

Li, J. and Barker, H.W.:A radiation algorithm with correlated k-distribution. Part I: local thermal equilibrium, J. Atmos. Sci., 62, 286-309, 2005.

Lin, H., Brunet, B. and Derome, J.: Forecast Skill of the Madden–Julian Oscillation in Two Canadian Atmospheric Models. Mon. Wea. Rev., 136, 4130–4149, doi: http://dx.doi.org/10.1175/2008MWR2459.1, 2008.

Lin, J.-T., and McElroy, M.: Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing, Atmospheric Environment, 44, 1726-1739, doi:10.1016/j.atmosenv.2010.02.009, 2010.

Liu, J., I. Fung, Kalnay, E., and Kang, J.-S.: CO2 transport uncertainties from the uncertainties in meteorological fields, Geophys. Res. Lett., 38,L12808, doi:10.1029/2011GL047213, 2011.

Lucas, C., Zipser, E. J. and Lemone, M. A.: Vertical velocity in oceanic convection off tropical Australia, J. Atmos. Sci., 51, 3183-3193, 1994.

Mailhot, J. and Benoit, R.: A Finite-Element Model of the Atmospheric Boundary Layer Suitable for Use with Numerical Weather Prediction Models, J. Atmos. Sci., 39, 2249–2266. doi: http://dx.doi.org/10.1175/1520-0469(1982)039<2249:AFEMOT>2.0.CO;2 , 1982.

44

Mailhot, J. and Co-authors: Scientific description of RPN physics library—Version 3.6. Atmospheric Environment Service Tech. Rep., 188 pp. [Available online at http://collaboration.cmc.ec.gc.ca/science/rpn/physics/physic98.pdf .], 1998.

Makar, P. A., Gong, W., Milbrandt, J., Hogrefe., C., Zhang, Y., Curci, G., Žabkar, R., Im, U., Balzarini, A., Baró, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, M., Honzak, L., Hou, A., Jiménez-Guerrero, P., Langer, M., Moran, M.D. , Pabla, B., Pérez, J.L., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., Zhang, J., Galmarini, S.: Feedbacks between air pollution and weather, Part 1: Effects on weather, Atmospheric Environment, 115, 442-469, http://dx.doi.org/10.1016/j.atmosenv.2014.12.003, 2015.

Manney, G. L., Schwartz, M.J., Krüger, K., Santee, M.L., Pawson, S., Lee, J. N., Daffer, W. H., Fuller, R. A., and Livesay, N. J.: Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming, Geophys. Res. Lett., 36, L12815, doi:10.1029/2009GL038586, 2009.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth Syst. Sci. Data, 6, 375-384, doi:10.5194/essd-6-375-2014, 2014.

Massart, S., Agustí-Panareda, A., Heymann, J., Buchwitz, M., Chevallier, F., Reuter, M., Hilker, M., Burrows, J. P., Hase, F., Desmet, F., Feist, D. G., and Kivi, R.: Ability of the 4-D-Var analysis of the GOSAT BESD XCO2 retrievals to characterize atmospheric CO2 at large and synoptic scales, Atmos. Chem. Phys. , 16, 1653-1671, doi:10.5194/acp-16-1653-2016, 2016.

McFarlane, N. A.: The effect of orographically excited gravity-wave drag on the circulation of the lower stratosphere and troposphere. J. Atmos. Sci., 44, 1775–1800, doi:10.1175/1520-0469(1987)044,1775:TEOOEG.2.0.CO;2, 1987.

McLinden, C. A., Olsen, S .C., Hannegan, B . , Wild, O., and Prather, M. J.: Stratospheric ozone in 3-D models' A simple chemistry and the cross-tropopause flux. J. Geophys. Res., 105 (D11): 14,653-14,666, 2000.

McTaggart-Cowan, R., and Zadra, A.: Representing Richardson Number Hysteresis in the NWP Boundary Layer, Monthly Weather Review, 143, 1232-1258, 2015.

Moran, M.D., Ménard, S., Talbot, D., Huang, P., Makar, P.A., Gong, W., Landry, H., Gravel, S., Gong, S., Crevier, L.-P., Kallaur, A., Sassi, M.: Particulate-matter forecasting with GEM-MACH15, a new Canadian air-quality forecast model, in: Steyn, D.G., Rao, S.T. (Eds.), Air Pollution Modelling and Its Application XX. Springer, Dordrecht, 289-292, 2010.

Neale, R.B., Chen, C.C., Gettelman, A., Lauritzen, P.H., Park, S., Williamson, D.L., Conley, A.J., Garcia, R., Kinnison, D., Lamarque, J.F. and Marsh, D.: Description of the NCAR community atmosphere model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+ STR. 2010 Available from http://www.cesm.ucar.edu/models/cesm1.2/cam/docs/description/cam5_desc.pdf

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999, Science, 300 (5625), 1560-1563, 2003.

Notholt, J., Petri, C., Warneke, T., Deutscher, N., Buschmann, M., Weinzierl, C., Macatangay, R., Grupe. P.: TCCON data from Bremen, Germany, Release GGG2014R0. TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.bremen01.R0/1149275, 2014.

5 ObsPack2013, Cooperative Global Atmospheric Data Integration Project, updated annually. Multi-laboratory compilation of atmospheric carbon dioxide data for the period 2000-2012; (2013): obspack_co2_1_PROTOTYPE_v1.0.4b_2014-02-13; National Oceanographic and Atmospheric Administration, Global Monitoring Division (NOAA GMD). http://dx.doi.org/10.3334/OBSPACK/1001, 2013.

Ott, L., Pawson, S. and Bacmeister, J.: An analysis of the impact of convective parameter sensitivity on simulated global
10 atmospheric CO distributions, J. Geophys. Res., 116, D21310, doi:10.1029/2011JD016077, 2011.

Ott, L. E., Pawson, S., Collatz, G. J., Gregg, W.W., Menemenlis, D., Brix, H., Rousseaux, 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 CarbonMonitoring Flux Pilot Project, J. Geophys. Res. Atmos., 120, doi:10.1002/2014JD022411, 2015.

15 Palmer, T.N.: Predictability of Weather and Climate: from theory to practice, in: Predictability of Weather and Climate, Cambridge University Press, Cambridge, U.K., 306-341, 2006.

Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S., and Baker, I. T.: Mechanisms for synoptic variations of atmospheric CO2 in North America, South America and Europe, Atmos. Chem. Phys., 8, 7239-7254, doi:10.5194/acp-8-7239-2008, 2008.

20 Patra, P. K., Maksyutov, S. and Nakazawa, T.: Analysis of atmospheric CO2 growth rates at Mauna Loa using CO2 fluxes derived from an inverse model. Tellus B, 57: 357–365. doi: 10.1111/j.1600-0889.2005.00159.x, 2005.

Patra, P. K., et al.: TransCom model simulations of hourly atmospheric CO2: Analysis of synoptic-scale variations for the period 2002–2003, Global Biogeochem. Cycles, 22, GB4013, doi:10.1029/2007GB003081, 2008.

Peters, W., Krol, M. C., Dlugokencky, E.J., Dentener, F. J., Bergamaschi, P., Dutton, G., Velthoven, P. v., Miller, J. B.,
25 Bruhwiler, L. and Tans, P. P.: Toward regional-scale modeling using the two-way nested global model TM5: Characterization of transport using SF6, J. Geophys. Res., 109, D19314, doi:10.1029/2004JD005020, 2004.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, 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. Nat. Acad. Sci., 104,
30 18925–18930, 2007.

Peters, W., Krol, M. C., Van Der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., DiSarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Van der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning,

A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M. and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations. Global Change Biology, 16: 1317–1337. doi:10.1111/j.1365-2486.2009.02078.x, 2010.

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., Rödenbeck, 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, doi:10.5194/bg-10-6699-2013, 2013.

Polavarapu, S. and Pulido, M.: Stratospheric and Mesospheric Data Assimilation: The role of middle atmospheric dynamics, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Vol. 3, Seon K. Park and Liang Xu (Editors), Springer, In Press. 2016.

Qaddouri, A., and Lee, V.: The Canadian Global Environmental Multiscale model on the Yin-Yang grid system, Q.J.R. Meteorol. Soc., 137, 1913-1926, doi: 10.1002/qj.873, 2011.

Raupach, M. R., Canadell, J. G., and Le Quéré, C.: Anthropogenic and biophysical contributions to increasing atmospheric CO2 growth rate and airborne fraction, Biogeosciences, 5, 1601-1613, doi:10.5194/bg-5-1601-2008, 2008

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624-3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Robichaud, A. and Ménard, R.: Multi-year objective analyses of warm season ground-level ozone and PM2.5 over North America using real-time observations and Canadian operational air quality models Atmos. Chem. Phys., 14, 1-31, 2014.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, 2003.

Saha, S., and Coauthors: The NCEP Climate Forecast System Reanalysis, Bull. Amer. Meteor. Soc., 91, 1015–1057, doi: 10.1175/2010BAMS3001.1 , 2010.

Saito, K., The JMA Nonhydrostatic Model and its Applications to Operation and Research, in Atmospheric Model Applications, Ismail Yucel (Ed.), InTech, ISBN: 978-953-51-0488-9, doi: 10.5772/2012, 2012.

Sherlock, V. B., Connor, Robinson, J., Shiona, H., Smale, D., and Pollard, D.: TCCON data from Lauder, New Zealand, 125HR, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.lauder02.R0/1149298, 2014.

Shukla, J. and Kinter, J.L. III: Predictability of seasonal climate variations: a pedagogical review, in: Predictability of Weather and Climate, Cambridge University Press, Cambridge, U.K., 306-341, 2006.

Skamarock, W.C., Park, S.-H., Klemp, J.B. and Snyder, C.: Atmospheric Kinetic Energy Spectra from Global High Resolution Nonhydrostatic Simulations, J. Atmos. Sci., 71, 4369-4380, DOI:10.1175/JAS-D-14-0114.1, 2014.

Smith, C.A., Compo, G.P. and Hooper, D.K.: Web-Based Reanalysis Intercomparison Tools (WRIT) for Analysis and Comparison of Reanalyses and Other Datasets. Bull. Amer. Meteor. Soc., 95, 1671–1678, doi: http://dx.doi.org/10.1175/BAMS-D-13-00192.1, 2014.

47

Staniforth, A., and Côté, J.: Semi-Lagrangian Integration Schemes for Atmospheric Models-A Review, Mon.Wea.Rev., 119, 2206-2223, 1991.

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.,

5 Langenfelds, R.L., Steele, L.P., Francey, R.J., Denning, A.S.: Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO2, Science, 316, DOI:10.1126/science.1137004, 1732-1735, 2007.

Strong, K., J. Mendonca, D. Weaver, P. Fogal, J.R. Drummond, R. Batchelor, R. Lindenmaier. 2014. TCCON data from Eureka, Canada, Release GGG2014R0. TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A, doi:10.14291/tccon.ggg2014.eureka01.R0/1149271, 2014.

10 Sørensen, B., Kaas, E., and Korsholm, U.: A mass-conserving and multi-tracer efficient transport scheme in the online integrated enviro-HIRLAM model, Geosci. Model Dev. 6, 1029–1042, 2013.

Sussmann, R. and Rettinger, M.: TCCON data from Garmisch, Germany, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.garmisch01.R0/1149299, 2014.

15 Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., Andrews, A. E., Lang, P. M., Neff, D., Dlugokencky, E., Miller, J. B., Montzka, S. A., Miller, B. R., Masarie, K. A., Biraud, S. C., Novelli, P. C., Crotwell, M., Crotwell, A. M., Thoning, K. and Tans, P. P..: Seasonal climatology of CO_2 across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Refernce Network, J. Geophys. Res. Atmos., 120, doi:10.1002/2014JD022591, 2015.

20 Takacs, L. L, Suarez, M. J. and Todling, R.: Maintaining Atmospheric Mass and Water Balance in Reanalyses. Q. J. R. Meteorol. Soc., In Press. 2015.

Trenberth, K. E., Seasonal Variations in Global Sea Level Pressure and the Total Mass of the Atmosphere: J. Geophys. Res., 86, 5238-5246, 1981.

Trenberth, K. E. and Smith, L.: The Mass of the Atmosphere: A Constraint on Global Analyses, J. Climate, 18, 864–875,

25 doi: http://dx.doi.org/10.1175/JCLI-3299.1, 2005.

Vallis, G. K.: Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, 745 pp, 2006.

Waliser, D. E.: Predictability of tropical intraseasonal variability, in: Predictability of Weather and Climate, Cambridge University Press, Cambridge, U.K., 275-305, 2006.

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N., Petri, C., Grupe, P., Vuillemin, C., Truong, F.,

30 Schmidt, M., Ramonet, M., and Parmentier, E.: TCCON data from Orleans, France, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.orleans01.R0/1149276, 2014.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., Keppel-Aleks, G., Allen, N., and Ayers, J.: TCCON data from Park Falls, Wisconsin, USA, Release GGG2014R0, TCCON data archive, hosted by the

Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.parkfalls01.R0/1149161, 2014a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin, J.: TCCON data from Lamont, Oklahoma, USA, Release GGG2014R0, TCCON data archive, hosted by the Carbon Dioxide

5 Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.14291/tccon.ggg2014.lamont01.R0/1149159, 2014b.

Williamson, D. L.: Semi-Lagrangian moisture transport in the NMC spectral model. Tellus A, 42: 413–428. doi: 10.1034/j.1600-0870.1990.t01-3-00002.x, 1990.

Worthy, D E., Platt, J. A., Kessler, R., Ernst, M., Audette , C., and Racki, S.: An update on the Canadian GHG measurement

10 program, In: Report of the 12th WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques, Toronto, Canada, September 2003, D. Worthy and L. Huang (editors), World Meteorological Organization Global Atmosphere Watch, Report 162, 220-231, 2005.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The total carbon column observing network, Philos. T. Roy. Soc. A, 369, 2087–2112,

15 doi:10.1098/rsta.2010.0240, 2011.

Wunch D., Toon G. C., Wennberg P. O., Wofsy S. C.,Stephens B. B., Fischer M. L., Uchino O., Abshire J. B., Bernath P., Biraud S. C., Blavier J.-F. L., Boone C., Bowman K. P., Browell E. V., Campos T., Connor B. J., Daube B. C., Deutscher N. M., Diao M., Elkins J. W., Gerbig C., Gottlieb E., Griffith D. W. T., Hurst D. F., Jim´enez R., Keppel-Aleks G., Kort E.A., Macatangay R., Machida T., Matsueda H., Moore F., Morino I., Park S., Robinson J., Roeh C. M., Sawa Y., Sherlock V.,

20 Sweeney C., Tanaka T., and Zondlo M. A. : Calibration of the Total Carbon Column Observing Network using aircraft profile data, Atmos. Meas. Tech., 3, 1351-1362, 2010.

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.

25 Yeh,K.-S., Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M. and Staniforth, A.: The CMC–MRB Global Environmental Multiscale (GEM) Model. Part III: Nonhydrostatic Formulation, Mon. Wea. Rev., 130, 339–356, doi: http://dx.doi.org/10.1175/1520-0493(2002)130<0339:TCMGEM>2.0.CO;2, 2002.

Zerroukat, M., and Allen, T.: A three-dimensional monotone and conservative semi-Lagrangian (SLICE-3D) for transport problems, Q.J.R.Meteorol.Soc., 138, 1640–1651, 2012.

30 Zhao, X., Tan, K., Zhao, S., Fang, J.: Changing climate affects vegetation growth in the arid region of the northwestern China, J. Arid Environments, 75(10), 946-952, ISSN 0140-1963, http://dx.doi.org/10.1016/j.jaridenv.2011.05.007, 2011.

Table 1: TCCON stations used, their location and data reference.

49

Site	Latitude	Longitude	Reference
Eureka	80.05	-86.42	Strong et al. (2014)
Sodankylä	67.37	26.63	Kivi et al. (2014)
Bialystok	53.23	23.02	Deutscher et al. (2014)
Bremen	53.10	8.85	Notholt et al. (2014)
Karlsruhe	49.10	8.44	Hase et al. (2014)
Orléans	47.97	2.11	Warneke et al. (2014)
Garmisch	47.48	11.06	Sussmann and Rettinger (2014)
Park Falls	45.94	-90.27	Wennberg et al. (2014a)
Lamont	36.60	-97.49	Wennberg et al. (2014b)
Izaña	28.30	-16.48	Blumenstock et al. (2014)
Darwin	-12.43	130.89	Griffith et al. (2014a)
Wollongong	-34.41	150.88	Griffith et al. (2014b)
Lauder	-45.05	169.68	Sherlock et al. (2014)

5

Table 2: Statistics for the average hourly XCO2 comparison between TCCON measurements and GEM-MACH-GHG simulations: RMS (ppm), Bias (ppm), Scatter (ppm), and correlation coefficient R. The « ALL » line shows the mean of each parameter using data from all sites combined. MEAN is the average of each parameter and SD is the standard deviation of the station 's bias. N is the number of data pairs (or sites for MEAN and SD) used in the computation of the statistics.

Site	Latitude	N	RMS (ppm)	Bias (ppm)	Scatter (ppm)	R
Eureka	80.05	49	2.86	-2.66	1.08	0.77
Sodankylä	67.37	1384	1.23	0.19	1.22	0.98
Bialystok	53.23	1279	1.16	0.55	1.03	0.95
Bremen	53.10	455	1	0.48	0.88	0.95
Karlsruhe	49.10	274	1.42	0.95	1.05	0.91
Orléans	47.97	910	0.76	0.27	0.71	0.98
Garmisch	47.48	1194	1.1	0.22	1.08	0.93
Park Falls	45.94	2427	0.97	-0.2	0.95	0.97
Lamont	36.60	4490	0.87	0.01	0.87	0.94
Izaña	28.30	221	1.13	0.45	1.04	0.84
Darwin	-12.43	1704	0.67	0.28	0.61	0.88
Wollongong	-34.41	1451	0.89	0.38	0.8	0.8
Lauder	-45.05	826	0.58	0.35	0.47	0.88
ALL		16615	0.98	0.14	0.96	0.95
MEAN		14	1.13	0.1	0.9	0.91
SD		14		0.84		

5

Table 3: Seasonal bias (ppm) for the hourly averaged XCO2 comparison between TCCON measurements and GEM-MACH-GHG simulations. The mean (using data from all sites) bias (ppm), scatter (ppm) and correlation coefficient (R) are also shown under « ALL ». Seasons with less than 10 pairs are not included in the « ALL » calculations.

Site	Latitude	N	Bias (ppm)	N	Bias (ppm)	N	Bias (ppm)	N	Bias (ppm)
			DJF		MAM		JJA		SON
Eureka	80.05	0	-	0	-	35	-2.18	14	-3.86
Sodankylä	67.37	56	0.83	450	1.04	610	0.06	262	-1.04
Bialystok	53.23	92	0.38	557	0.82	537	0.49	93	-0.61
Bremen	53.1	34	0.72	237	0.72	123	0.02	61	0.36
Karlsruhe	49.1	5	1.32	59	1.09	113	0.96	97	0.85
Orléans	47.97	73	-0.27	310	0.38	236	0.66	291	-0.03
Garmisch	47.48	107	0.2	254	0.4	439	0.29	394	0.03
Park Falls	45.94	311	-0.22	668	-0.13	676	-0.42	772	-0.06
Lamont	36.6	850	-0.03	983	-0.05	1602	-0.16	1055	0.34
Izaña	28.3	25	-0.65	44	-0.56	103	0.76	49	1.28
Darwin	-12.43	310	-0.04	391	0.1	126	0.5	877	0.44
Wollongong	-34.41	445	0.11	185	0.42	332	0.96	489	0.23
Lauder	-45.05	168	0.01	142	0.39	192	0.44	324	0.45
ALL									
N			2471		4280		5124		4778
Bias (ppm)			0.02		0.33		0.12		0.15
Scatter (ppm)			0.78		0.88		1.11		0.86
R			0.93		0.92		0.94		0.93

52
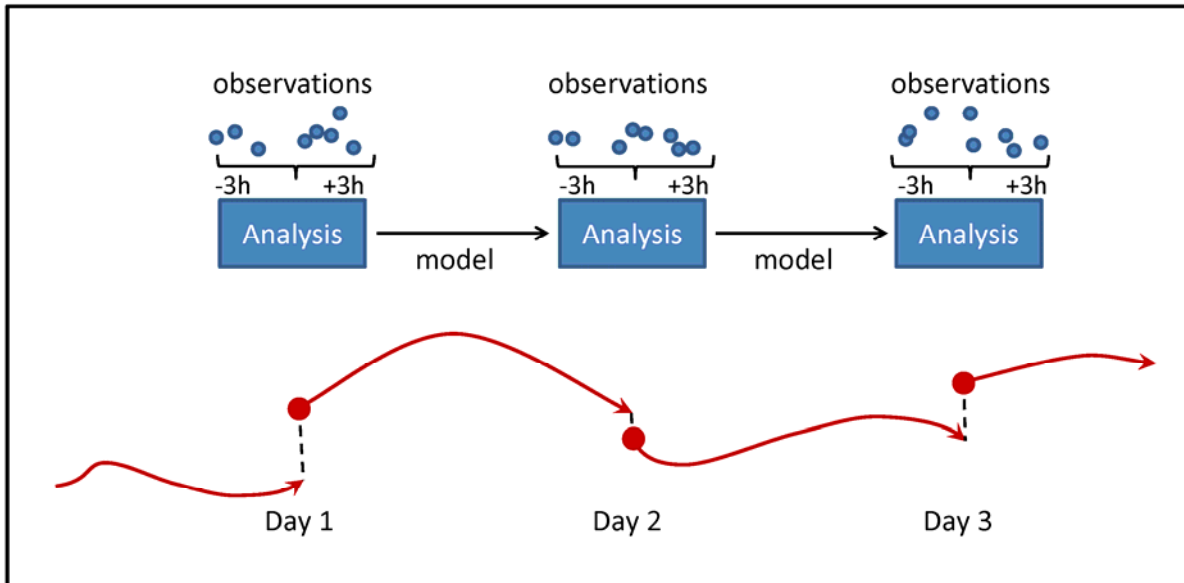
Figure 1: Schematic diagram of EC-CAS forward model cycles. Meteorological analyses were precomputed by CMC's operational global deterministic prediction system using all observations collected in a 6 h window centred on the analysis time. These analyses are represented by blue boxes. CO_2 tracers are added to these analyses to form an initial condition (solid red circles) for launching a 24 h forecast (red arrows) using the coupled meteorological/tracer model starting at 0 UTC of each day. Note that the meteorological update cycle is 6 h so there should be 4 times as many blue boxes as shown. Some of these were omitted for clarity of presentation.
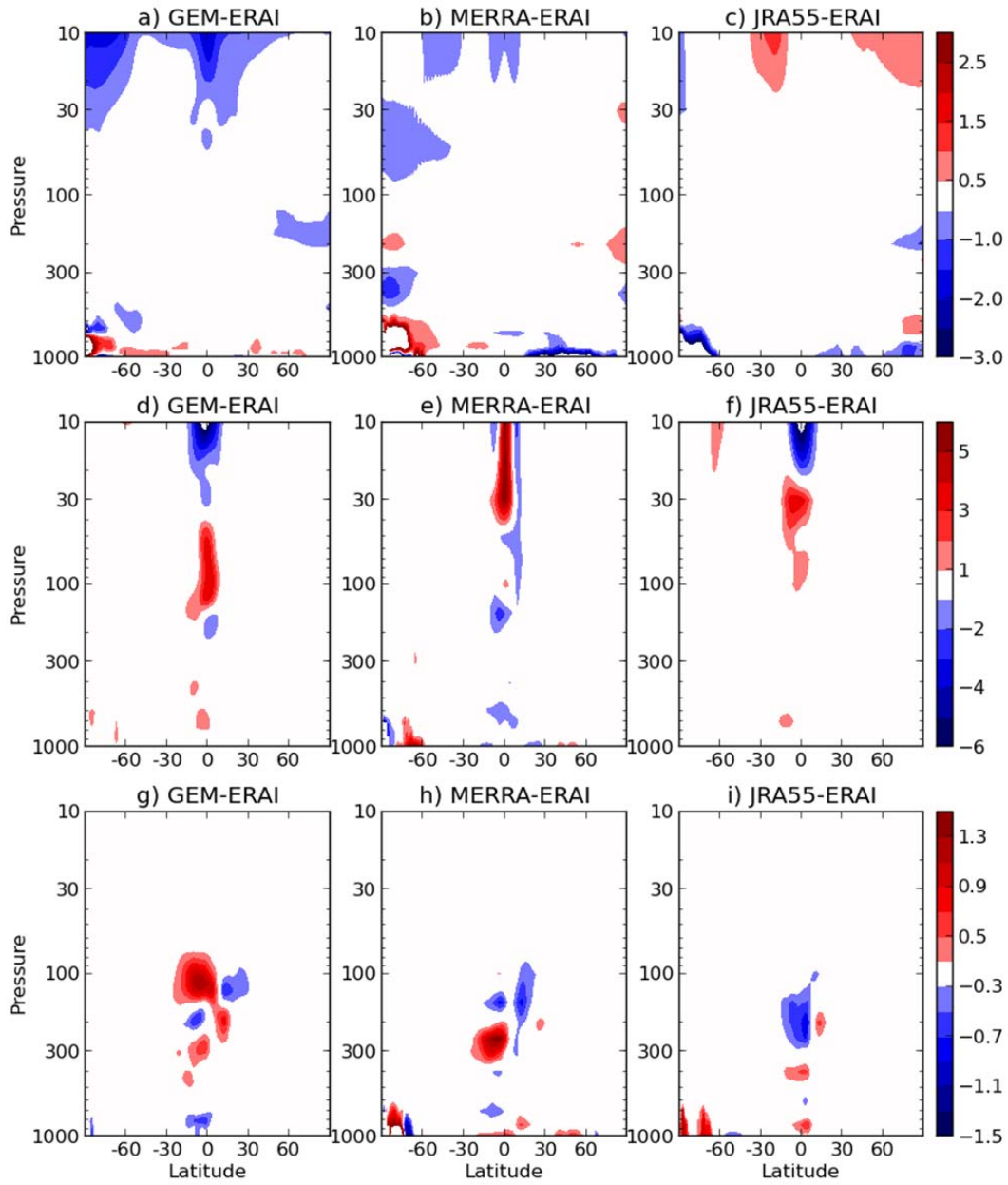
Figure 2: Comparison of GEM-MACH-GHG meteorological analyses with other reanalyses for July 2009. Monthly and zonal means of differences with respect to ERA-Interim fields of GEM-MACH-GHG (left column), MERRA (middle column) and JRA55 (right column) are shown for temperature in K (top row), zonal wind in m/s (middle row) and meridional wind in m/s (bottom row).
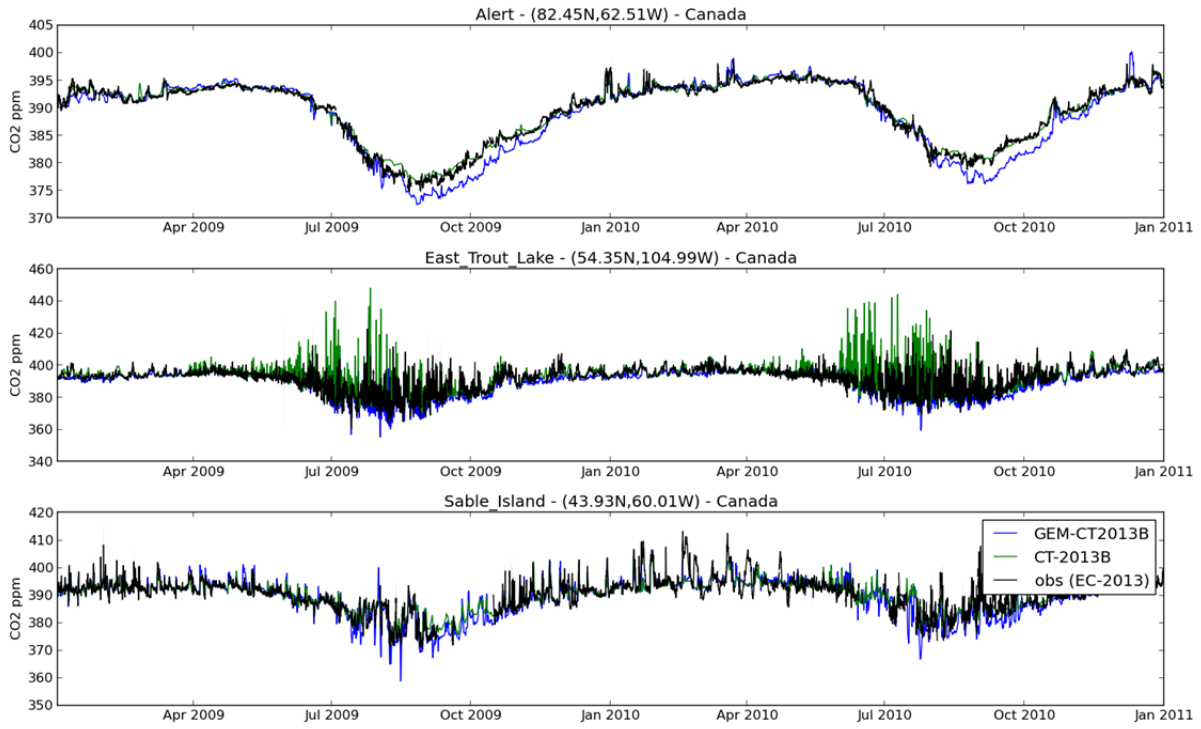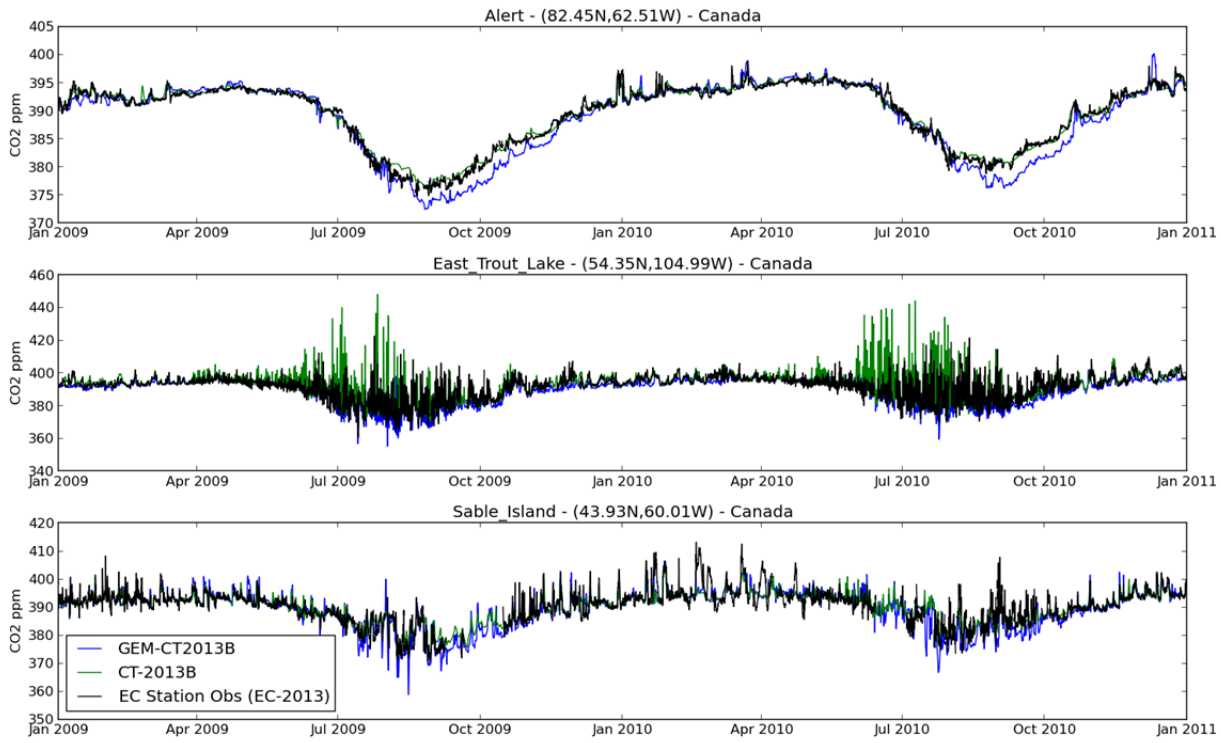
Figure 3: Comparison of GEM-MACH-GHG with surface observations at Alert (top), East Trout Lake (middle) and Sable Island (bottom). ECCC observations (black), GEM-MACH-GHG with CT2013B fluxes (blue) and CarbonTracker-2013B (green) time series are shown for each location.

East Trout Lake 2009
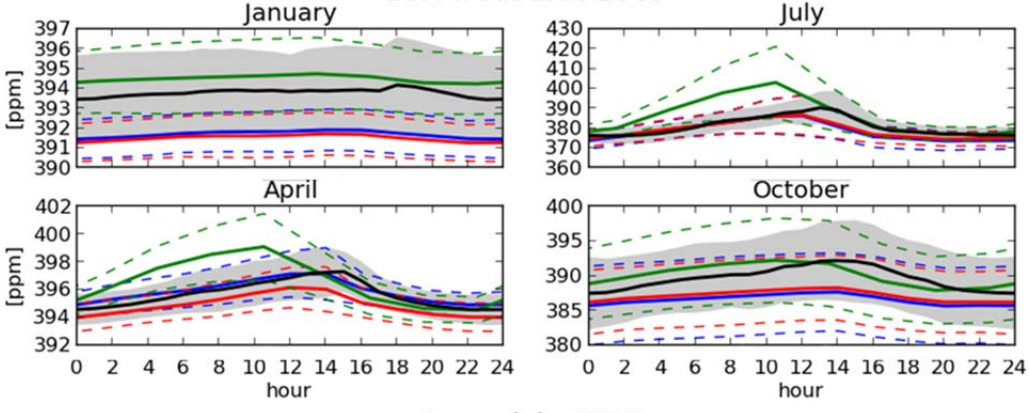
Fraserdale 2009
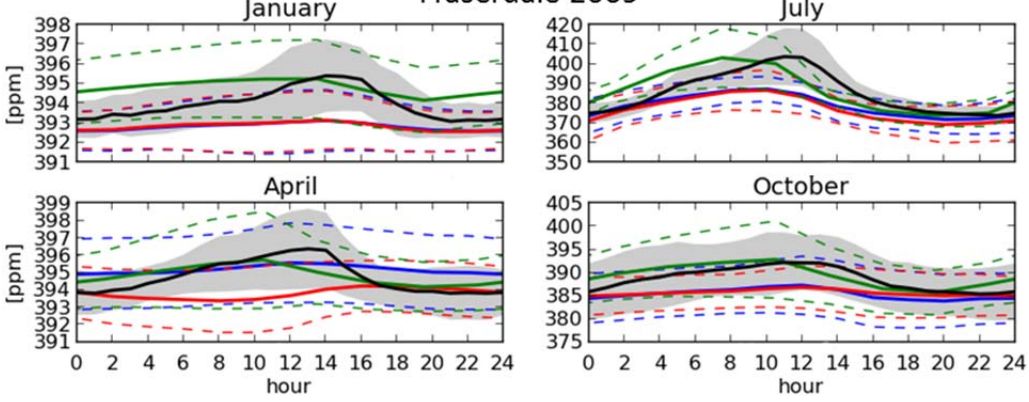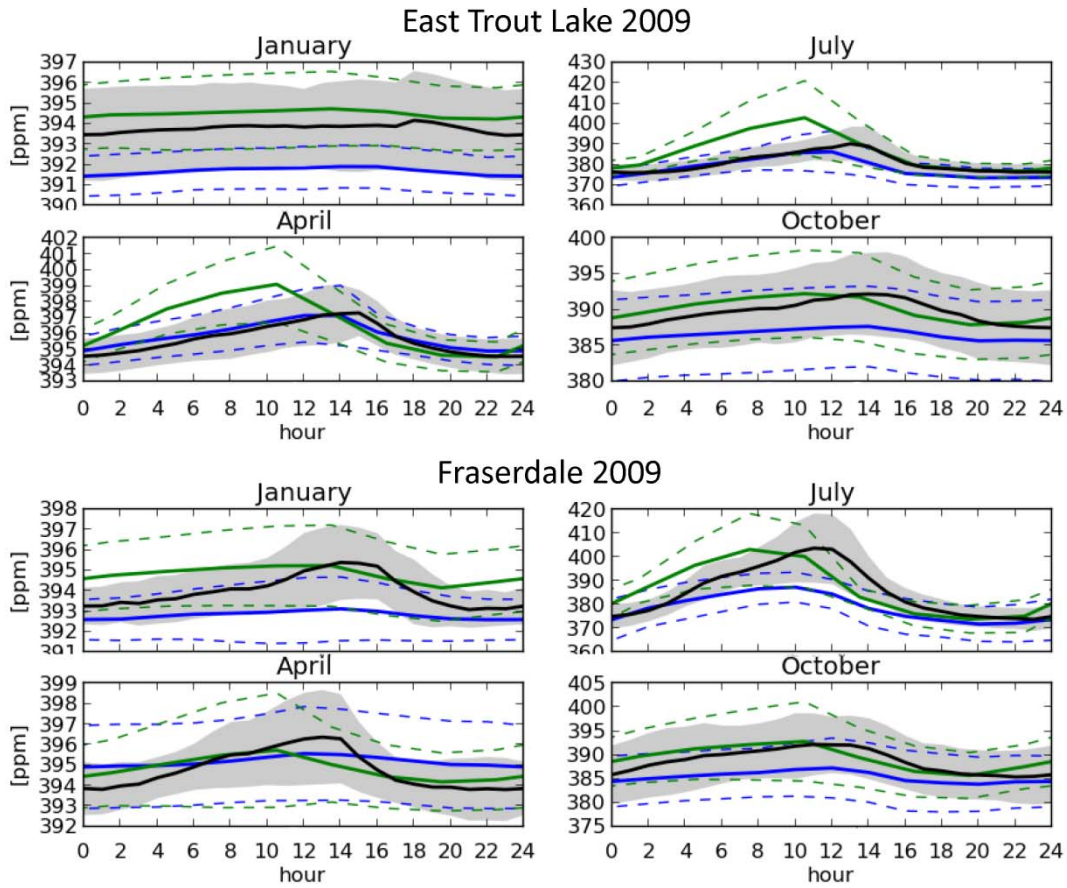
Figure 4: Mean diurnal cycle at East Trout Lake, Saskatchewan (top 4 panels) and Fraserdale, Ontario (bottom 4 panels) in 2009. Each panel shows the observed mean cycle from continuous measurements (black), CT2013B (green)~~,~~ and GEM-MACH-GHG with CT2013B fluxes (blue) ~~and GEM-MACH-GHG with CT2010 fluxes (red)~~. Time is given in UTC. The 4 panels correspond to the months of January, July, April and October, as labelled above each panel. The grey shaded region indicates one standard deviation above and below observed values while the dashed lines indicate the same for the model run with the corresponding colour.

58
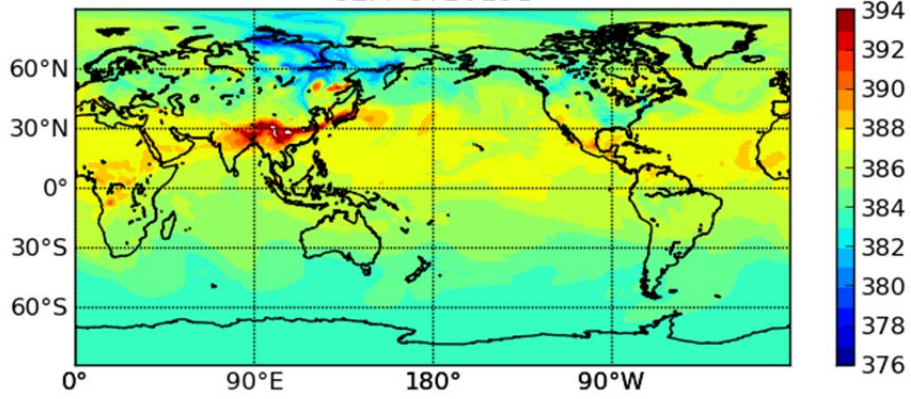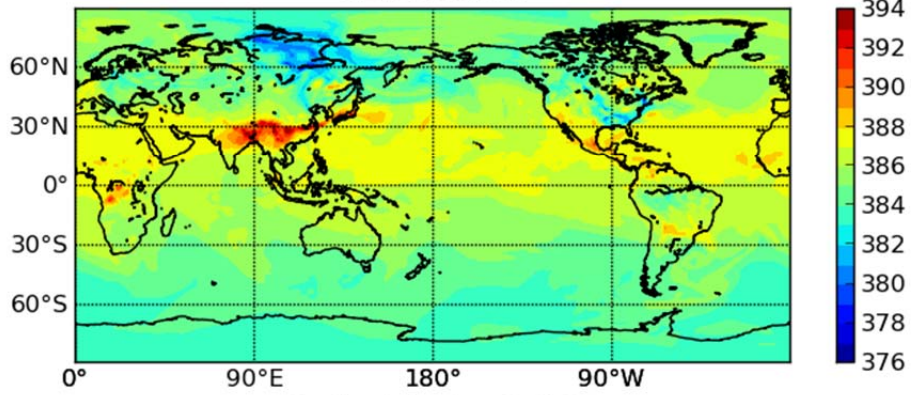
XCO2 (in ppm) - 2009-07-01 00:00
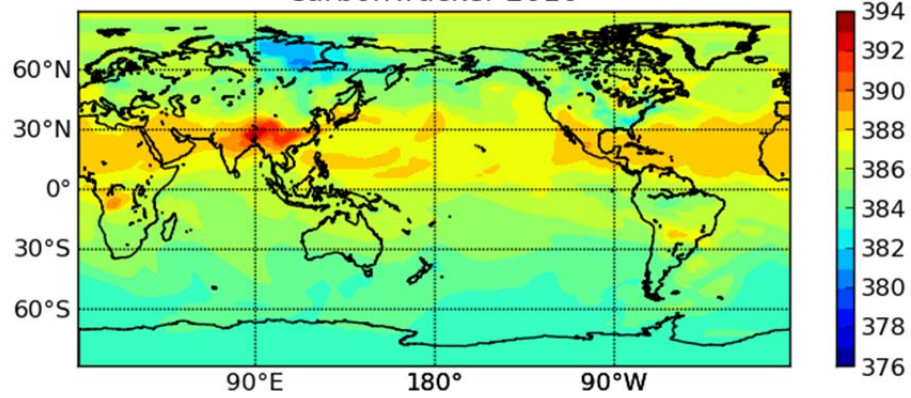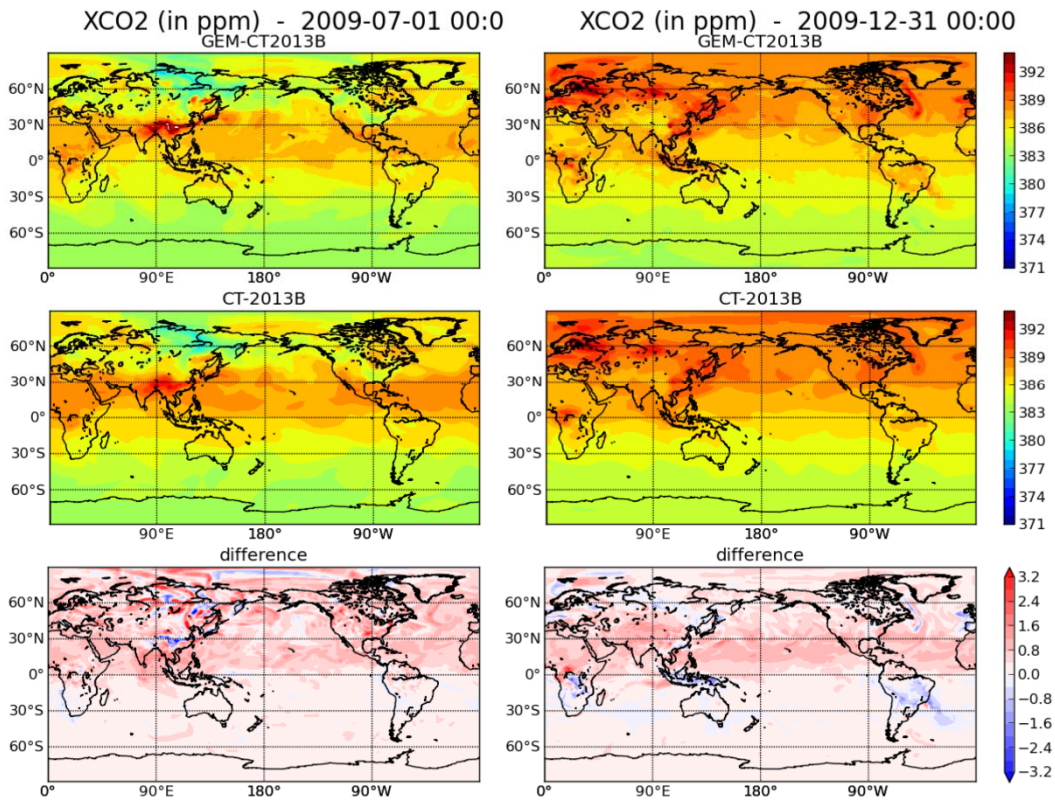
Figure 5: Column mean CO$_2$ ~~on 1 July 2009~~ for GEM-MACH-GHG with CT2013B fluxes (top **row**), ~~GEM-MACH-GHG with~~ CT201~~03~~**3B** ~~fluxes~~ (middle **row**) and CT201**3B minus GEM-MACH-GHG fields** ~~0~~ (bottom **row**) **for 1 July 2009 (left column) and 31 Dec. 2009 (right column)**.

Zonal mean CO2 (in ppm) - 2009-07-01 00:00
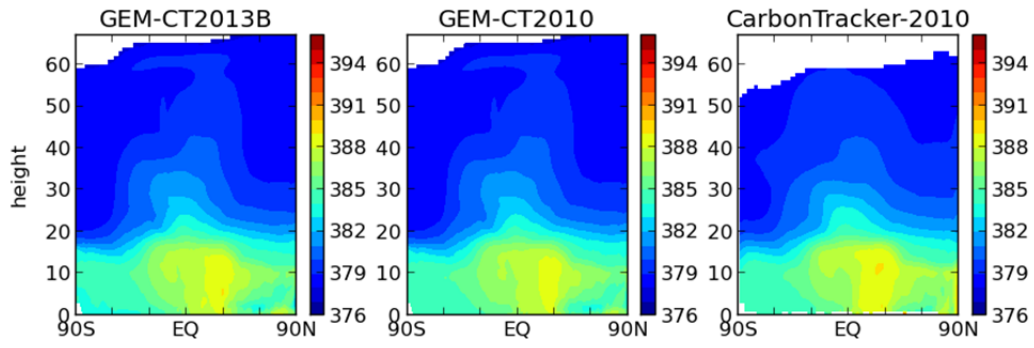
Zonal mean CO2 (in ppm) - 2009-12-31 00:00
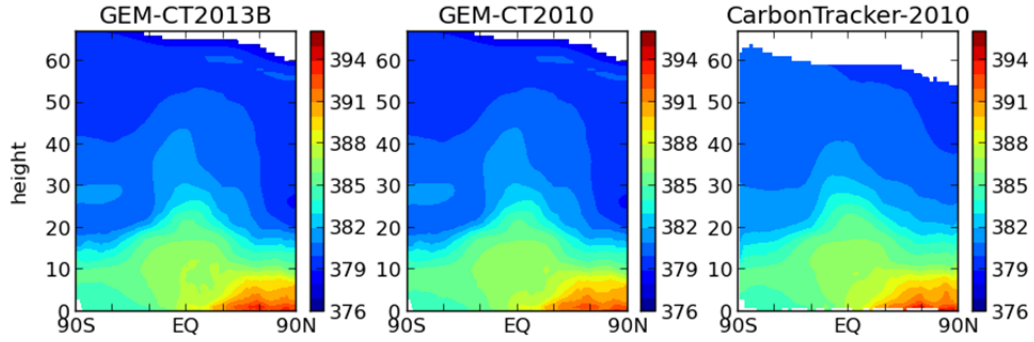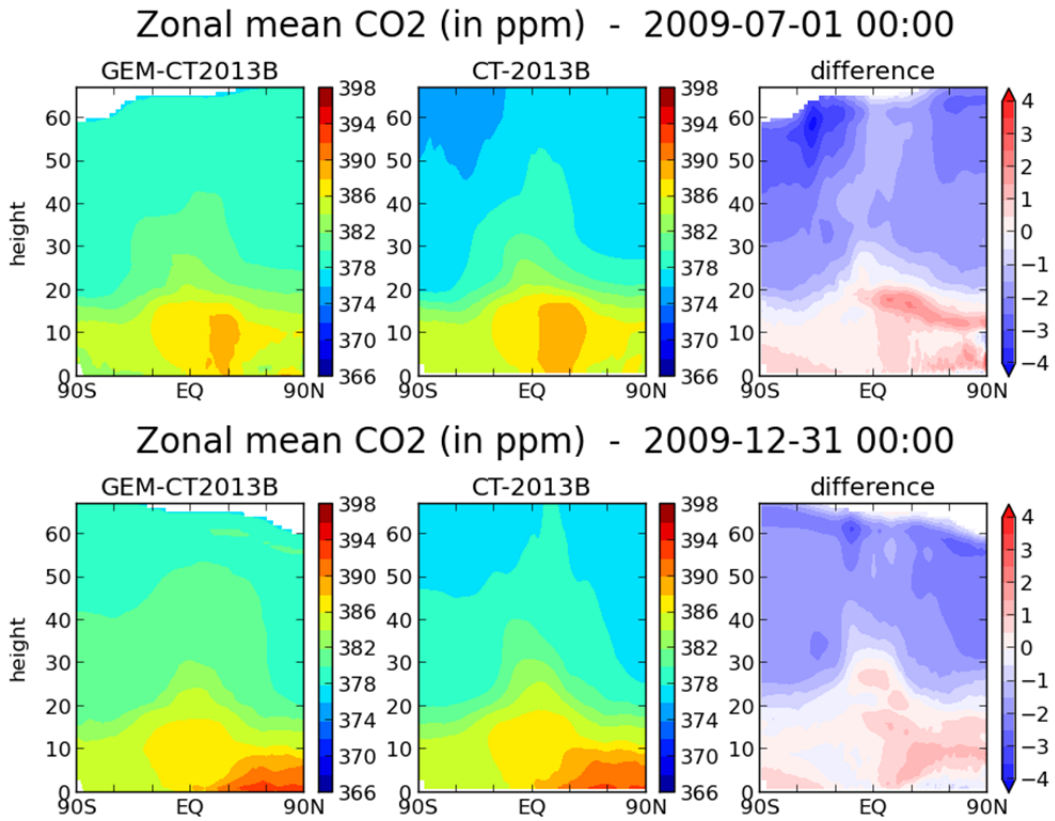
Figure 6: Zonal mean CO₂ on 1 July 2009 (top row) and 31 Dec. 2009 (bottom row) for GEM-MACH-GHG with CT2013B fluxes (left column), ~~GEM-MACH-GHG with CT2010 fluxes~~ CT2013B (middle column), and CT201~~3B minus GEM-MACH-GHG~~0 fields (right column).
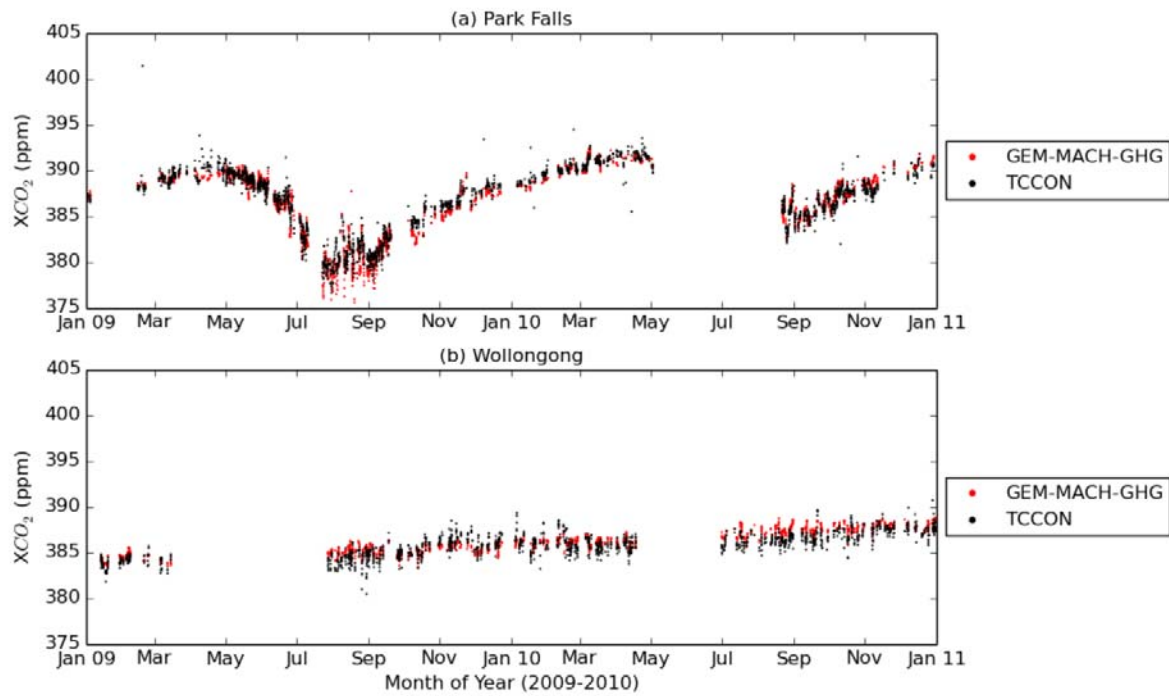
Figure 7: Time series of XCO2 (ppm) at (a) Park Falls, USA and (b) Wollongong, Australia, between January 1, 2009 and January 1, 2011. The dots are hourly averaged XCO2 for TCCON (black) and smoothed GEM-MACH-GHG simulations obtained using CT2013B posterior fluxes (red).

63
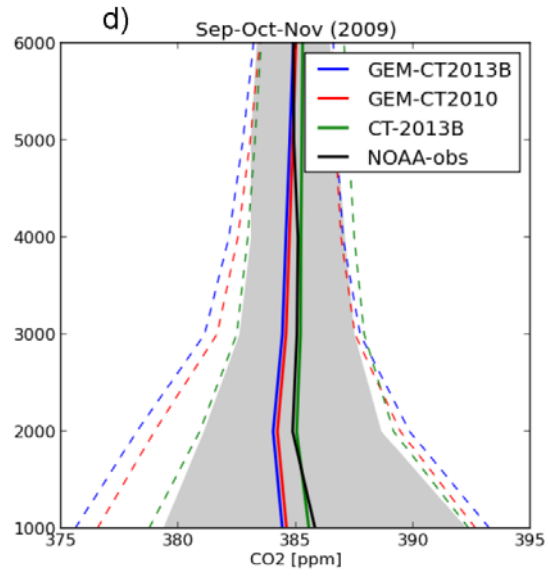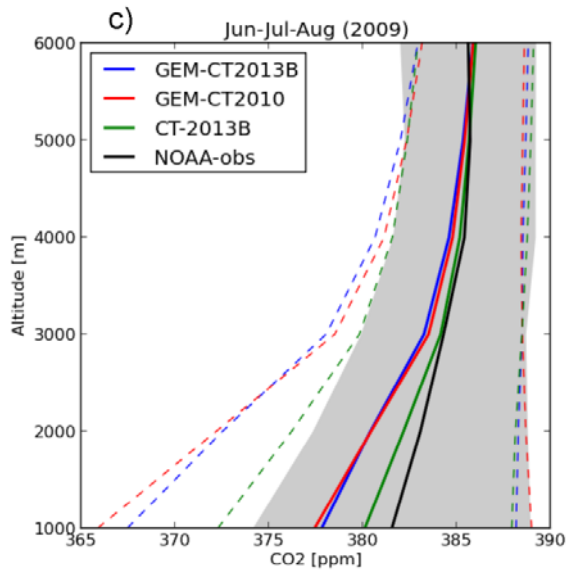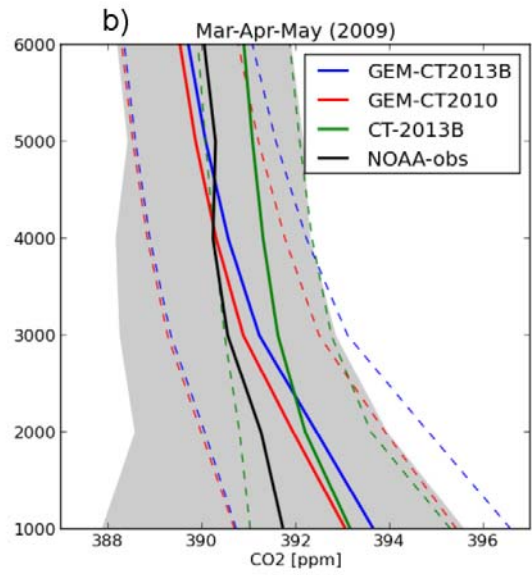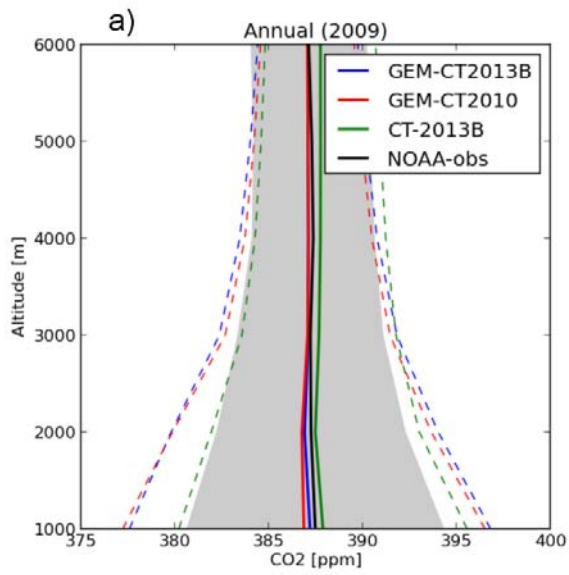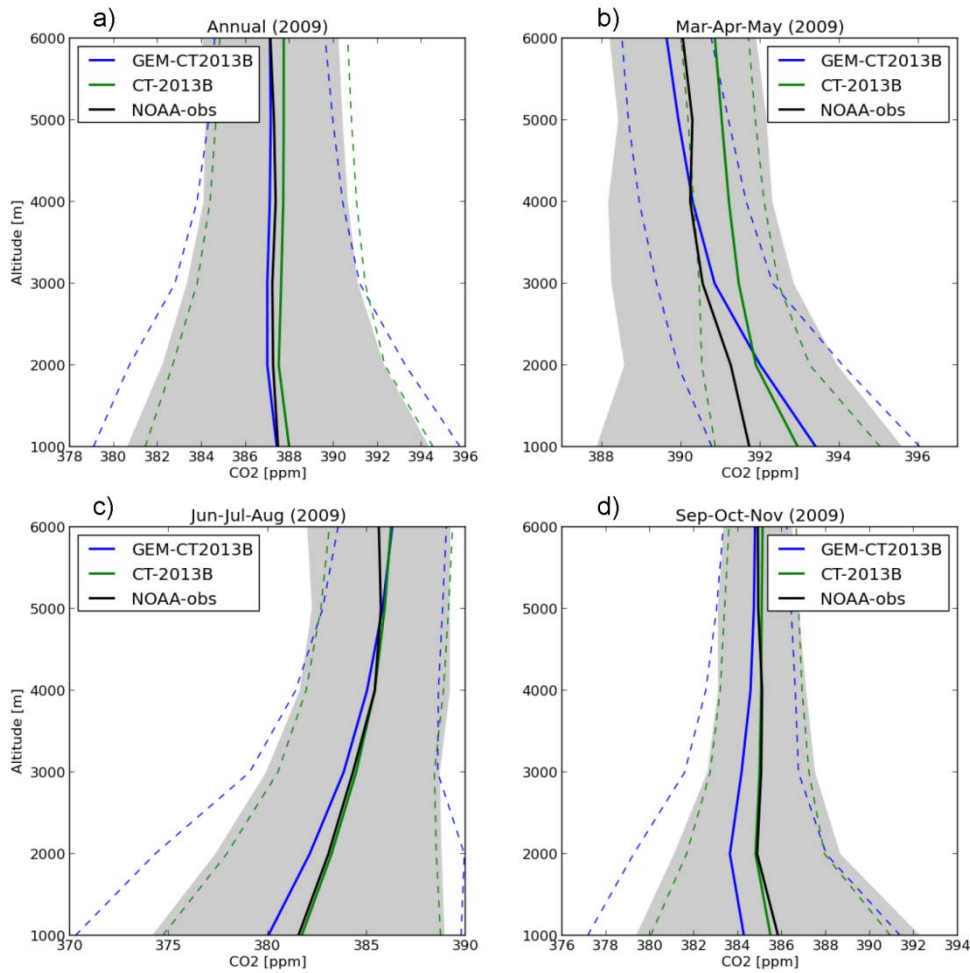
Figure 8: Comparison of model mean profiles to NOAA aircraft observations. Observations (black curves) are from obspack_co2_1_PROTOTYPE_v1.0.4_2013-11-25 for locations over continental U.S. and Canada, only. Observed and modelled profiles are binned over (a) 2009, (b) March to May 2009, (c) June to August 2009, (d) Sept. to Nov. 2009. CarbonTracker 2013B molefractions (green) and, GEM-MACH-GHG with CT2013B posterior fluxes (blue curves) and GEM-MACH-GHG with CT2010 posterior fluxes (red curves) are shown in all panels. The shaded grey regions indicate plus or minus one standard deviation for the observations while the dashed coloured lines indicate the same quantities but for the different model runs. Note that the x-axis range differs in each panel and that ticks are every 5 ppm except for panel b where they are every 2 ppm. Sites used are: Beaver Crossing, Nebraska; Bradgate, Iowa; Briggsdale, Colorado; Cape May, New Jersey; Charleston, South Carolina; Dahlen, North Dakota; East Trout Lake, Saskatchewan; Estevan Point, British Columbia; Fairchild, Wisconsin; Harvard Forest, Massachusetts; Homer, Illinois; Oglesby, Illinois; Park Falls, Wisconsin; Poker Flat, Alaska; Sinton, Texas; Southern Great Plains, Oklahoma; Trinidad Head, California; West Branch, Iowa; Worcester, Massachusetts.

65

a) CO2   b) Temperature   c) Vorticity   d) Divergence
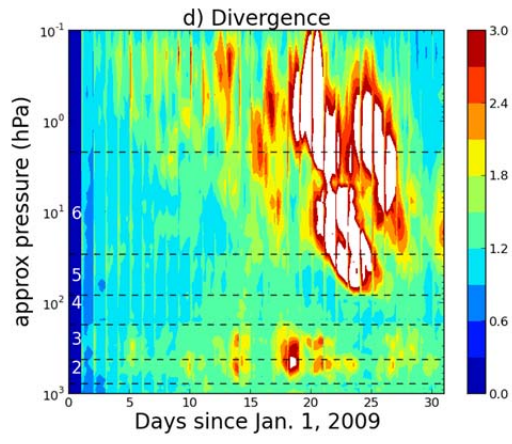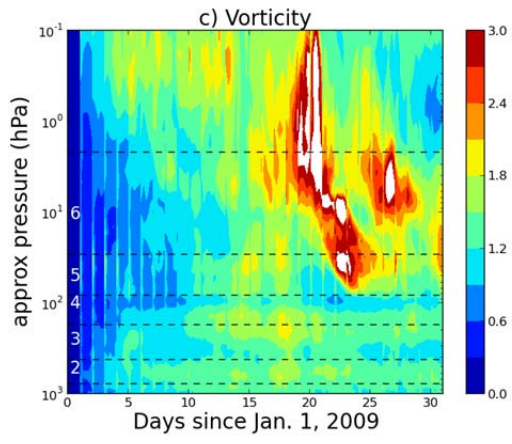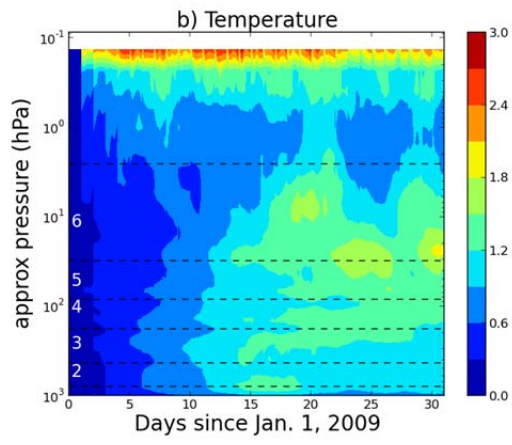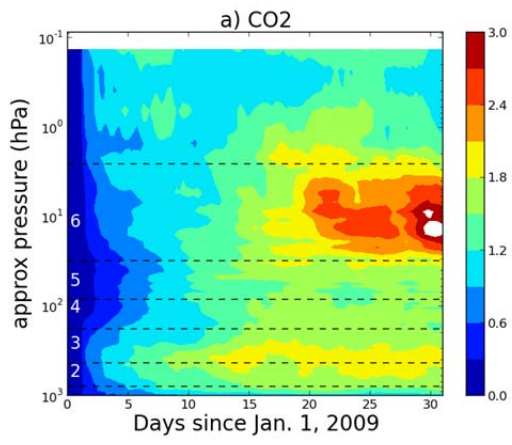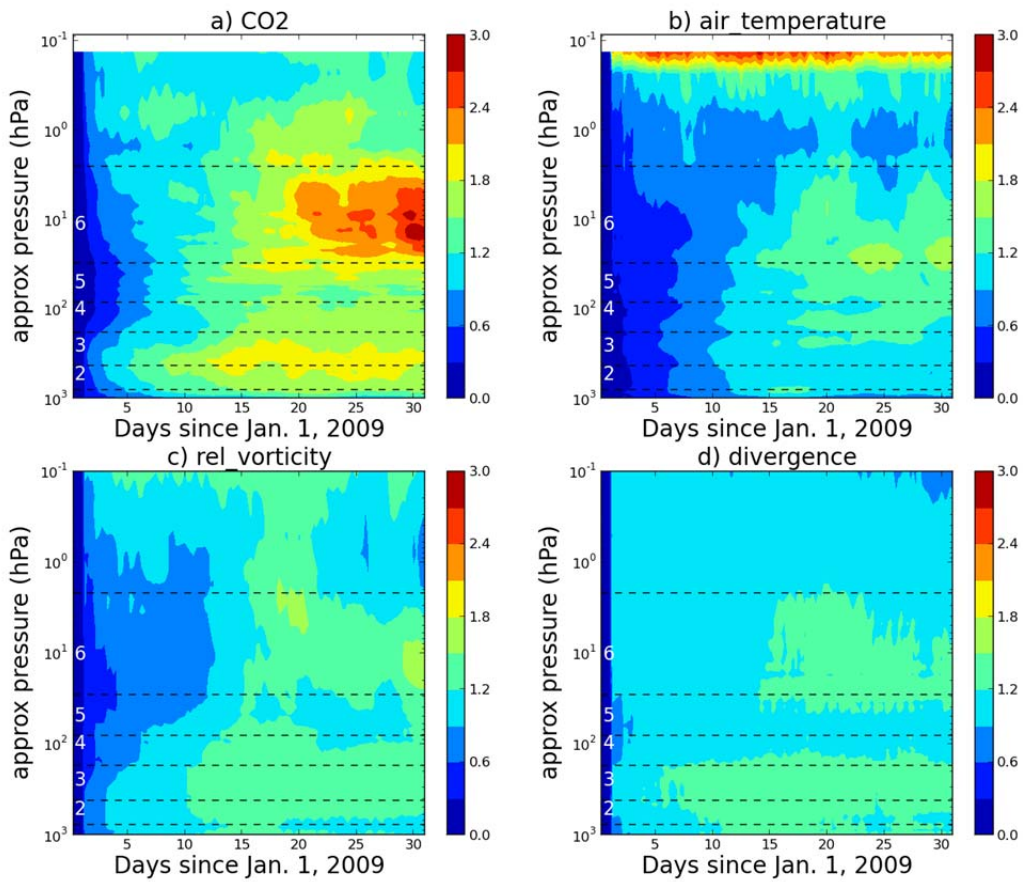
Figure 9: Predictability on weather time scales during January 2009. Predictability error is defined here as the ~~square root of~~ the global mean ~~variance~~zonal standard deviation of the difference in evolution of a control cycle from a climate cycle. This value is normalized by the global mean ~~variance~~zonal standard deviation of the corresponding field in the initial condition. Predictable regimes are ones for which this ratio is much less than 1. The normalized predictability error fields for a) CO_2, b) temperature, c) vorticity and d) divergence are plotted as a function of model vertical level (converted to approximate pressure with a reference surface pressure of 1000 hPa) and time in days since 1 January 2009. The layers labelled ~~~~with white text in each panel will be used for computing the layer mean averages shown in Figure 10.

67
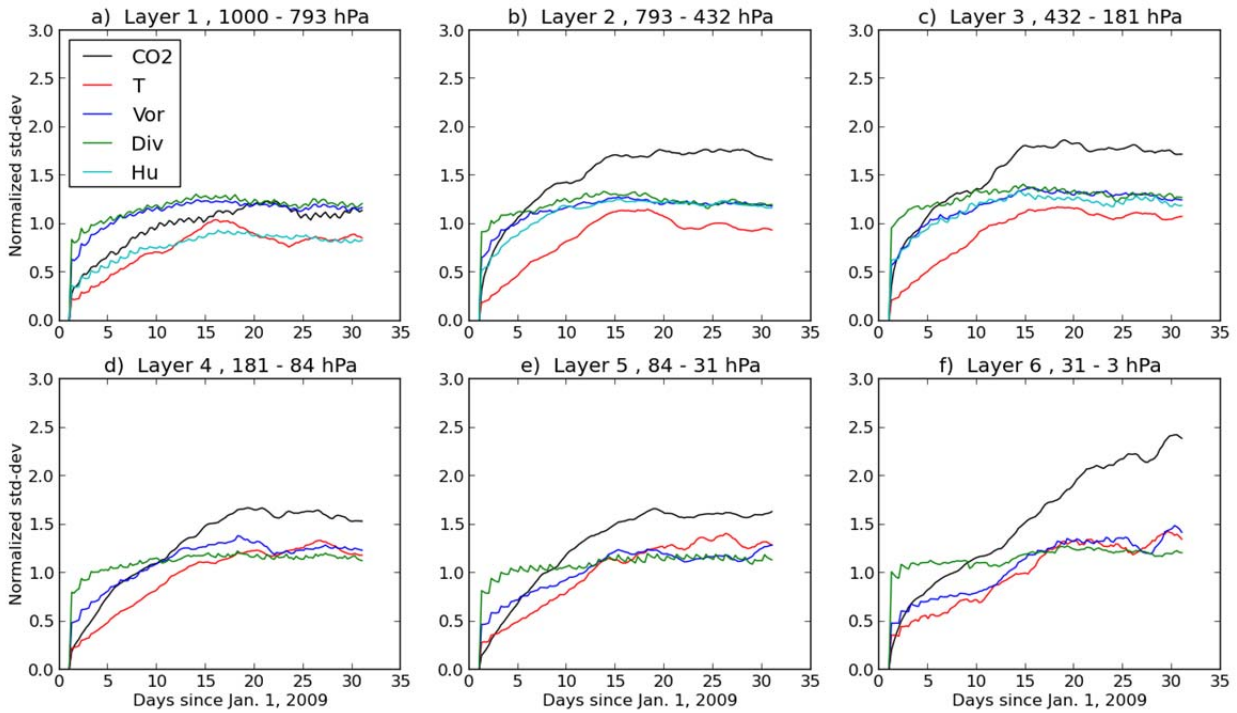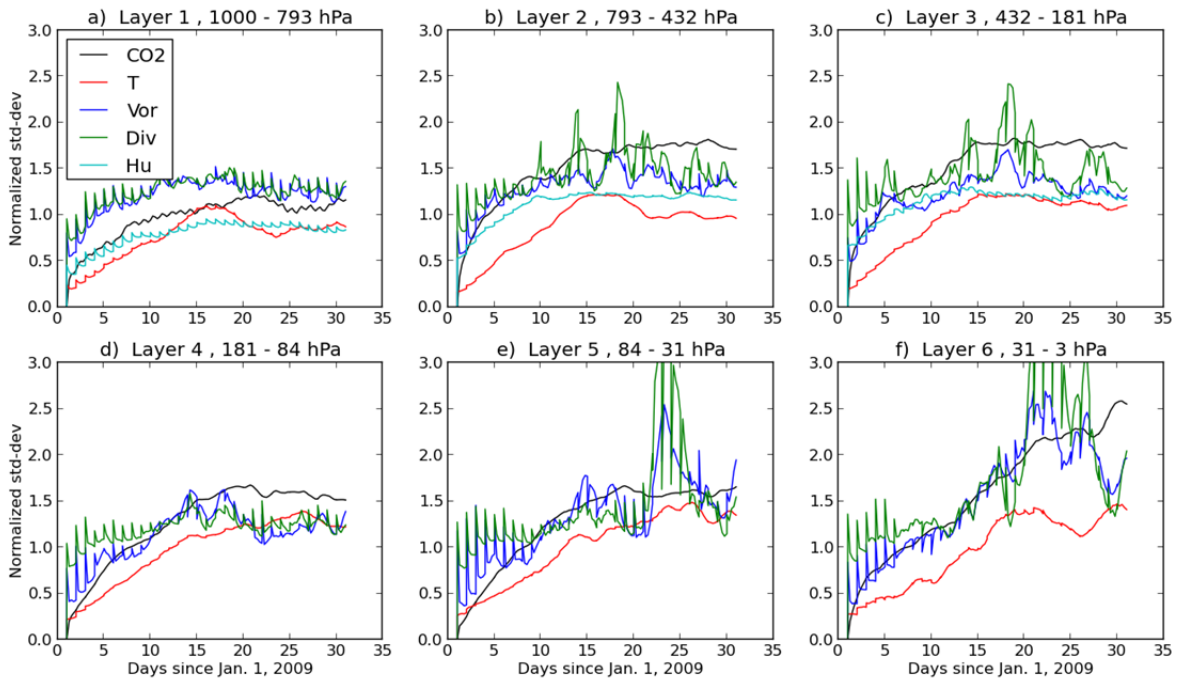
Figure 10: Time series of layer mean normalized predictability error during January 2009 for CO_2 (black curves), temperature (red curves), vorticity (blue curves), divergence (green curves) and specific humidity (cyan curves). For clarity, no curves for specific humidity are plotted in the bottom row of panels. The normalized errors are averaged over 12 model levels, the top and bottom levels used in the average are given in approximate pressure above each panel. The layer numbers are associated with the layers defined by dashed horizontal lines in Figure 9 and correspond to the near surface (layer 1, panel a), the lower troposphere (layer 2, panel b), the mid troposphere (layer 3, panel c), the upper troposphere (layer 4, panel d), the lower stratosphere (layer 5, panel e) and the upper stratosphere (layer 6, panel f).

5

69
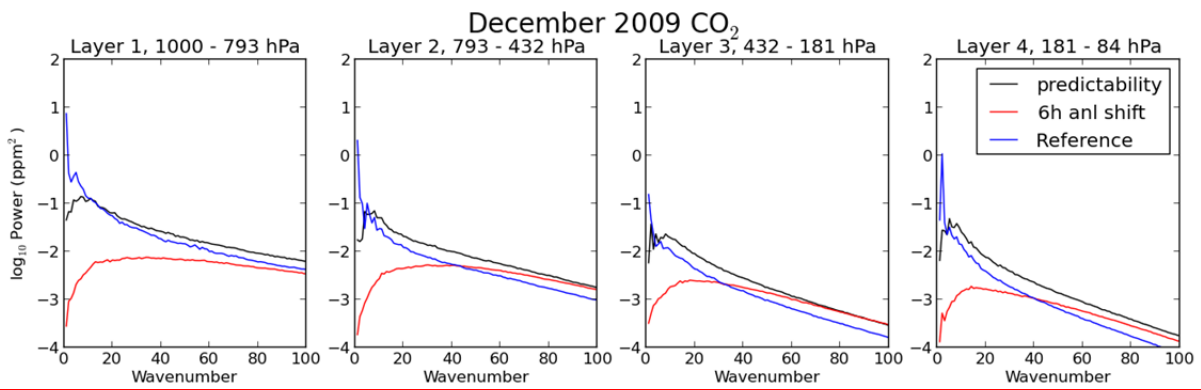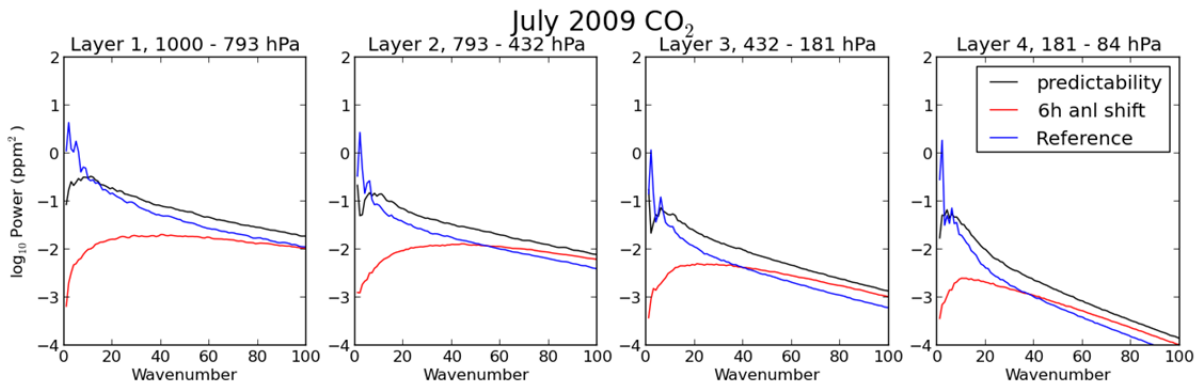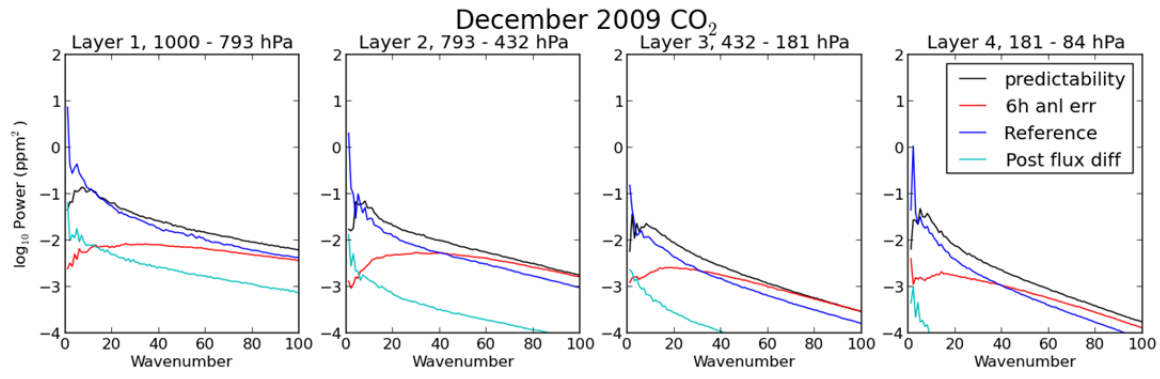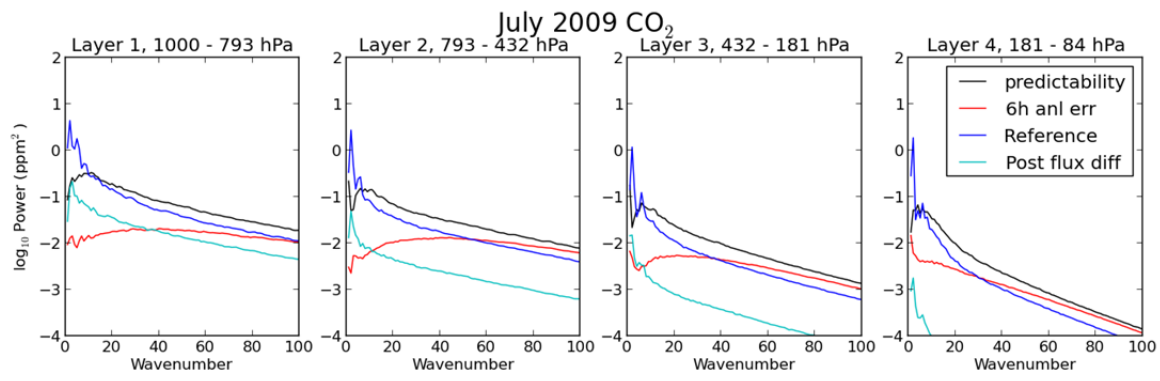
Figure 11: Spectra of various fields as a function of total wavenumber. Spectra are averaged over one month for July 2009 (top row) and December 2009 (bottom row) and over 12 model levels. The lower and upper model levels averaged are indicated above each frame. Approximate pressure is obtained from model level by multiplying by 1000 (which corresponds to assuming a reference surface pressure of 1000 hPa). The CO_2 reference state spectra (blue curves), predictability error (black curves) ~~and,~~ <u>and</u> error due to a 6 h shift in analysis fields (red curves) ~~and differences due to the use of different posterior fluxes (CT2010 or CT2013B) (cyan curves)~~ are shown.
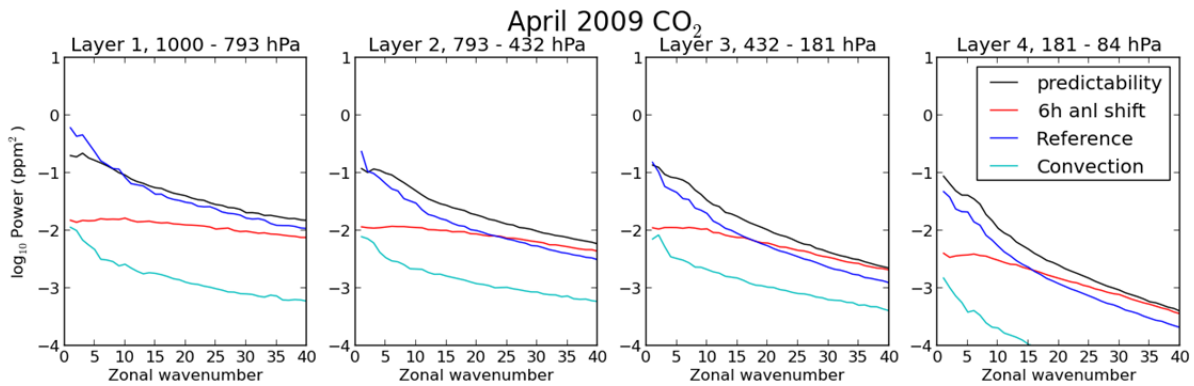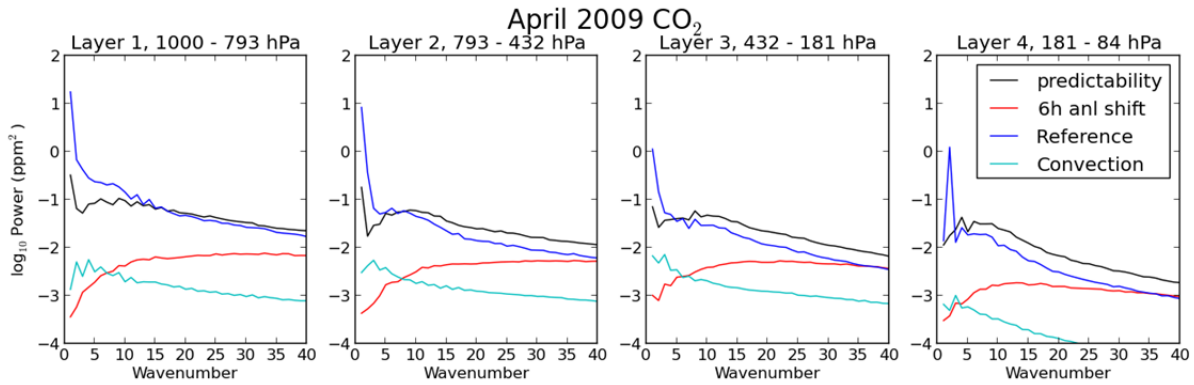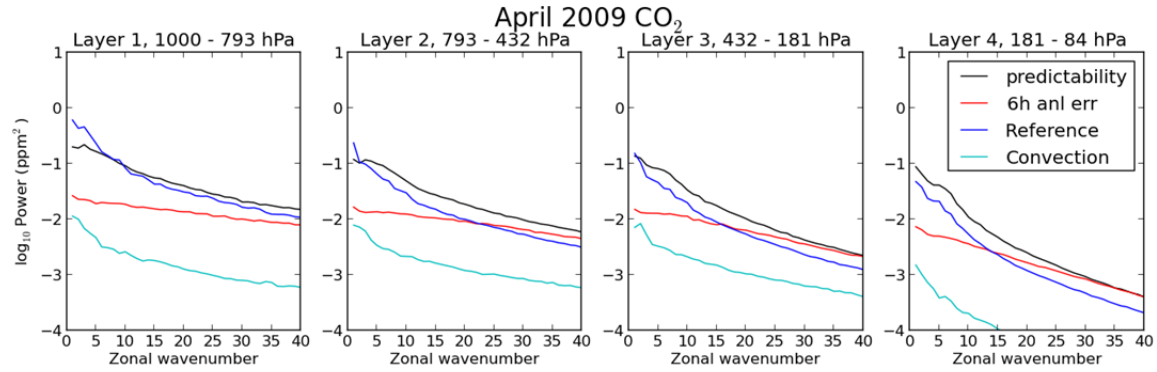
71
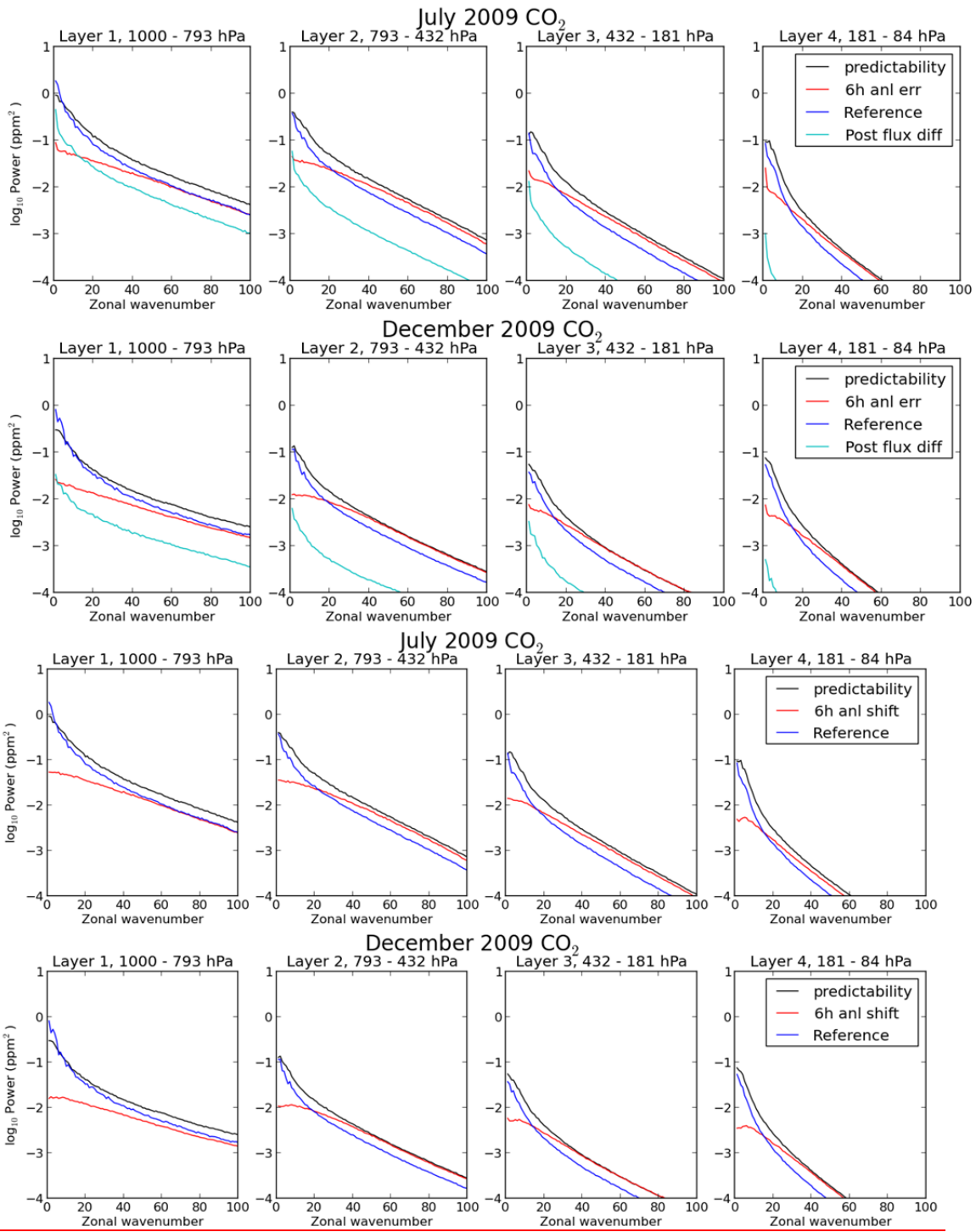
Figure 12: As in Figure 11 but for spectra as a function of zonal wavenumber.
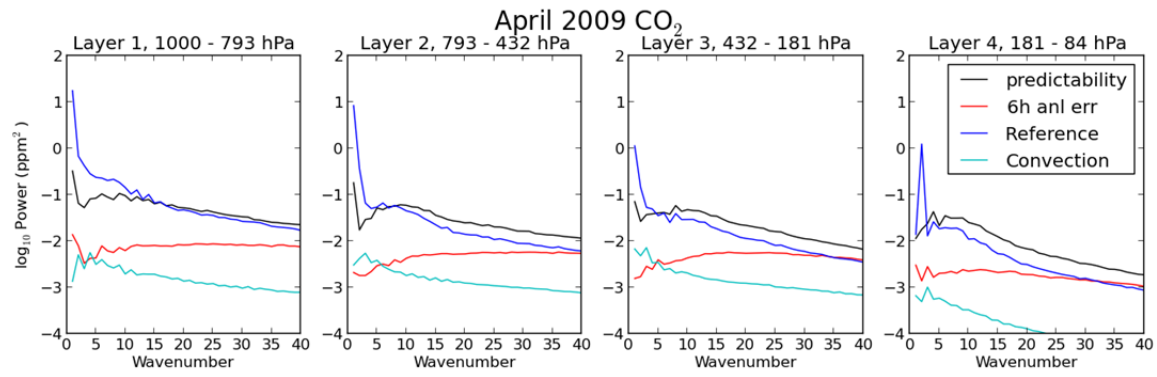
April 2009 CO$_2$

Figure 13: As in Figure 11 but for spectra in April 2009 and now errors due to the ~~addition~~ removal of convective tracer transport are shown in cyan curves. Spectra are shown as a function of total wavenumber (top row) or zonal wavenumber (bottom row). Note that only wavenumbers up to 40 are shown.