

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 CO₂ fluxes in the predictability of atmospheric CO₂ 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 CO₂, 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 CO₂ 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.

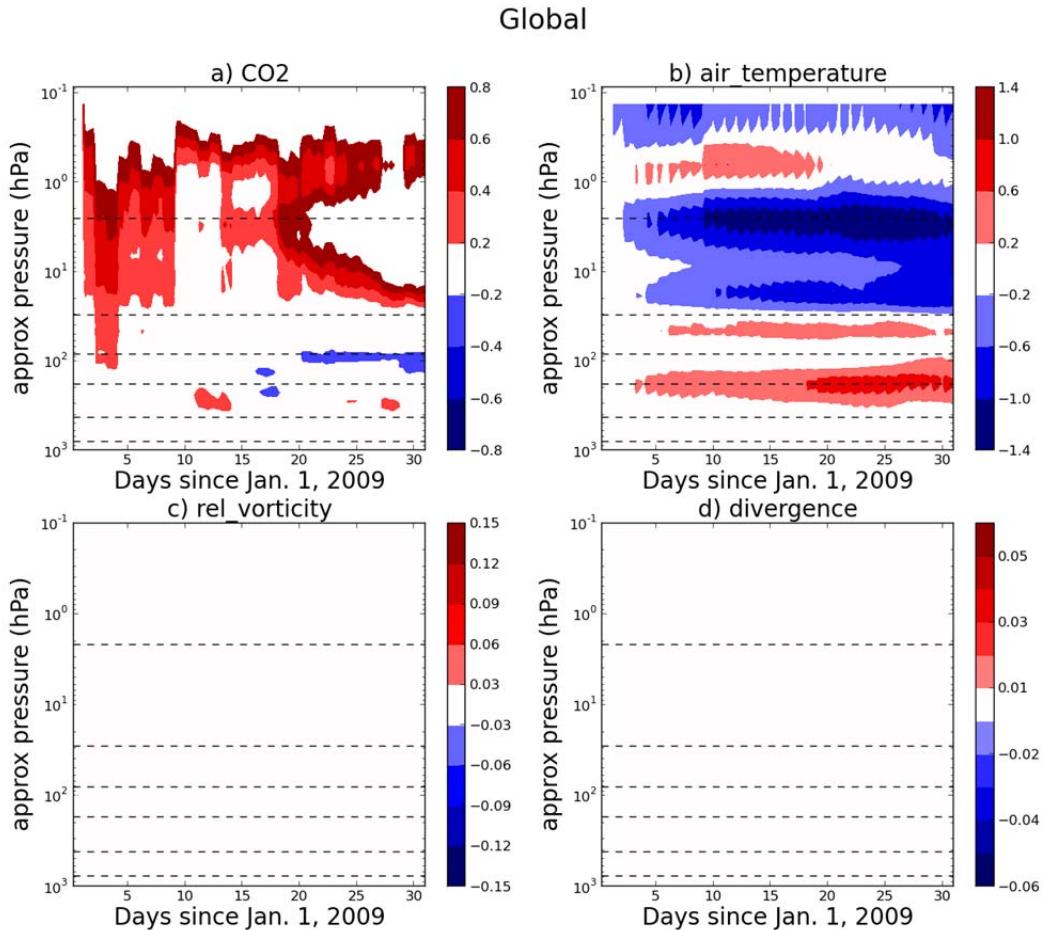


Figure R1-1: Global mean difference between the reference and climate-mode cycles during January 2009. Values are normalized by the global mean zonal standard deviation of the initial state. Mean departures are negligible for the wind field (relative vorticity and divergence fields), are important for the temperature near the tropopause and negligible for CO₂ 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 CO₂ 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 CO₂ 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 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 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 CO₂ 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 predictability 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 CO₂ 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 CO₂ 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 be explained by the normalization of the predictability score, as well as the stronger influence from the prescribed fluxes. The normalization 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 predictability (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 meteorological 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 predictability 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 CO₂ 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 CO₂ 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 observations near the surface, but because any surface flux will have a much larger influence on the CO₂ 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 CO₂ has been investigated (see Massart et al. 2016, ACP). The results from Massart et al. (2016) show the forecast of column-averaged CO₂ 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 CO₂ 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 clarity 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.