Authors' Response: 'High-resolution simulations of atmospheric CO2 over complex terrain- representing the Ochsenkopf mountain tall tower' by Pillai et al.

We would like to thank anonymous referee for his/her valuable comments. Authors' responses to these comments are as follows. The referee's comments are in black color fonts and the authors' responses are in blue color.

Whereas the conclusion, that meso-scale models are better capable to reproduce high resolution spatial and temporal variability, may not be stunning, this paper has several merits: 1) it combines available data (meteorological and tracer data, ground, aircraft based, and model data) in a logical way within one framework, 2) it addresses several rather different atmospheric phenomena. To my knowledge a meso-scale model evaluation of atmospheric CO2 transport has not been performed earlier on this scale. At the same time, such model systems are highly needed to explain and attribute the variability of the high accuracy observations taken at tall towers and flasks collected in the last decade(s). Unfortunately, the large extent of the study also puts limits to the depth at which the underlying processes can be studied. As a result, the study does not relate model-data mismatches to specific processes or model configurations, nor does it result in suggestions on how to improve the models. I consider this as a drawback of the integral method chosen, although I do suggest that the authors include a paragraph in section 5 to discuss potential improvements.

A more serious concern with the paper is the lack of a discussion of how good is good enough. On quite a number of occasions, the authors state that 'WRF or STILT captures XX "relatively well"', even though the difference between model and observations is sometimes up to a few degrees C / g/kg/ ppm. Prior to assessing whether the model does good or not, the paper needs to discuss what the model is supposed to do best for its particular purpose, and how that can be tested. In this case, the model is supposed to do CO2 transport good at high spatial and temporal resolution, so it needs to do vertical mixing, meso-scale circulations, pbl height and surface fluxes well. What are acceptable margins (e.g. compare to TM3 and measurement accuracy)? How important is a bias compared to amplitude? Only with a definition of the model requirements can statements about performance be made.

We agree with the referee's comment about the definition of the acceptable transport model bias in terms of CO_2 fluxes.

So we modified Sec.1 and included the following statement

"The target is to be able to retrieve fluxes from inverse modelling with an accuracy of 10% on monthly time scales; hence this criterion is used to assess the model performance. As the simulated CO_2 concentrations are based on a combination of transport and flux modeling and both of which are associated with uncertainties, the method will yield an upper bound for the transport error compatible with the 10% flux uncertainty."

Also modified Sect. 3.2

"As mentioned in the Sect. 1, we target at a monthly averaged flux uncertainty (in $\mu.moles/(m^2s^{-1})$ *) with the upper boundary of 10% . This criterion affords a monthly averaged transport model bias of 0.4 to 0.6 ppm which is calculated by transporting simulated flux (CO_2 uptake (Gross Ecosystem Exchange) or release (ecosystem respiration) fluxes) of 10 % uncertainty during summer period."*

Also modified Sect. 3.2.1

"The high-resolution models slightly overestimated the CO_2 concentrations with biases ranging from 0.3 to 0.6 ppm (table 2), which is in the targeted range (10 %) of the monthly averaged flux uncertainty."

Specific comments: Page 6880, line 4-9: this sentence is too long
-modified to

"In addition, a Lagrangian particle dispersion model, Stochastic Time-Inverted Lagrangian Transport Model (STILT) (Lin et al., 2003), driven with high-resolution assimilated meteorological fields, is used to simulate the upstream influence on the observation point (i.e. the footprints). These footprints are then multiplied by VPRM fluxes (NEE) as well as fluxes from fossil fuel emissions in order to simulate CO_2 concentrations at the observation location."

Page 6880, last paragraph: it remains unclear, at this point, what method you will apply to address the objectives. It may help to better explain the methods here or just before.
-included
"In order to meet these goals, the simulations from the high-resolution modeling framework are compared with observations from various data streams as well as with simulations from a coarse resolution model."

Page 6881, section 2: information about the region (topography, land use, industrial activity, etc.) is missing, as well as the location of the tower, and the distance between the tower and the wind profiler.
-modified Sect. 2.1

"The tower at Ochsenkopf is 163 m tall and is located in the second highest peak of Fichtelgebirge mountain range (1022 m a.s.l.; 50° 1'48" N, 11°48'30" E) in Germany. The surrounded area of the tower is covered predominantly with conifer forest and has low population as well as industrial activity."

Page 6883, line 8: The WRF domain of 500 x 500 km seems somewhat small, resulting in a large part of the model domain being influenced by the boundary conditions. Please comment on this.

The domain is large enough to avoid the boundary influence at the measurement locations - i.e. the location of the tower and the aircraft track are much away from the domain boundary and hence there will not be any boundary effects.

Page 6883, line 24: The Pillai et al, 2011 is not submitted yet, so please do not refer to it.
???
-modified accordingly

Page 6884, section 3: Why don't you evaluate the model with respect to boundary layer height and/or turbulence characteristics (u*, tke, sigma_u, sigma_w, : : :)? And, in the whisker plots, why don't you make a distinction between data taken below and above the top of the boundary layer, because humidity, temperature, and CO2 often change considerably across the bl top. This may explain the large variability in specific humidity/relative humidity. You may even consider comparing average boundary layer measurements and model data when the bl is well mixed.
- included the boundary layer info. in the plot.

Page 6884, first paragraph, you discuss wind direction, but you show wind speed (Fig2)
- horizontal direction of wind is already indicated with arrowheads (see Fig. 2); however we see problem with poor resolution of the Fig and this will be improved.

Figure needs higher resolution, arrowheads are barely visible

Page 6885, line 6: so why did you choose to compare this time slot?
- included the sentence in Sec. 3.1.1
"The time-slot was chosen as a representative of the model performance when the boundary layer is well-mixed."

Page 6885, lines 9-15: it is more conventional to display model – observations, because then an overestimation becomes a positive difference.
-modified accordingly

Page 6885, lines 19 and 21: change 'parameters' to 'variables'
-modified accordingly

Page 6886, line 3: why do you compare moisture in terms of RH and not in terms of q, because RH depends also on T
-modified accordingly

Page 6887, line 5: 'a slight underestimation': the difference looks quite large to me

The figure is re-plotted; there was a colour scale difference in the plot of specific humidity. (see Fig. 5)

Page 6887, first paragraph: it is interesting that WRF and STILT capture the daytime minimum CO2 much better than the nighttime maximum, and that WRF and STILT both tend to underestimate the nighttime maximum. Can you explain this?
This is an indication of issues with the models in representing nocturnal vertical mixing of the tracers.

Page 6888, line 11: 'remarkably well': 1) what sort of temporal variability do you mean, 2) bias and standard deviation are indeed smaller than in TM3 and R2 larger, but they are still an order of magnitude larger than the measurement precision. Therefore I do not think the label 'remarkably well' is suitable.

-modified Sect. 3.2.1
"The summary statistics clearly indicate that high-resolution models are able to predict <u>reasonably</u> well the temporal patterns of CO_2 (measured at three different vertical levels on the tower) for different seasons <u>when compared to the coarse resolution model."</u>

Page 6888, line 23: 3rd reason: q has sinks (precipitation) in the atmosphere, where CO2 has not.
- added the third reason

Page 6889, line 23: 'a decrease' in spec. hum, not an increase?
-modified

Page 6890: how are the observed/modelled peaks defined in terms of spatial and temporal averages? Do differences in averaging explain part of the mismatch?

No. We used 3 hourly aggregated data from observations and measurements (i.e. same temporal resolution). The spatial scale was 2 km x 2 km - the transport in 3 hours covers more than 2 km x 2km

Page 6892, section 4.2.2 (Mountain wave activity). I do not see any waves in fig. 11, except in 11d, where w seems to correspond to U x dz_sfc/dx. I am not an expert in the field of gravity waves, therefore please explain where you see evidence of waves. To me Fig. 11 looks more like katabatic flow than like gravity waves.

We hope that you might have referred to Fig. 12 instead of Fig. 11. The strong gradient in potential temperature, relatively high vertical velocity and high wind speed indicate the maximum likelihood of mountain wave. Also you may see clear wave-like structure in the vertical velocity plot (Fig. 11 d).

We realize that the caption of the Fig. 12 needs to be modified. So now it is changed to "*…The overlaid arrows indicate horizontal wind direction at different altitude."*

Page 6894, line 9 ‘unrealistic gradient’: I would say that WRF’s gradient looks more like the observed gradient than STILT’s.

You may see the gradient between the levels in WRF-VPRM, when you look at the time slot between *0 to 6 UTC during 16–17 October 2006.*

-the text is modified to make it clear
“This was reproduced well in STILT-VPRM, while WRF-VPRM showed the decreasing tendency of CO_2 *concentrations on this period but with an unrealistic gradient between the layers [0 to 6 UTC during 16–17 October 2006], which might be due to the underestimated mixing process.”*

Table 2: sd is short for standard deviation? Or standard error?
It is standard deviation of the differences- the caption for Table 2 is modified

Figure 1: it is somewhat unclear whether the blue rectangles represent nested grids in STILT or WRF or both. Please re-formulate the caption
-modified

Figure 5: it is unclear what ‘dimona’ means (in the colorbar text)
-modified

Figures 10 and 12: the arrows show the wind direction. Which components do they show? Certainly not U and W?
It is U and V components. The caption is modified (Please see the comment above)

Figure 13: Why do you show only 20 hours, and not 24?
It is monthly averaged diurnal cycle of 3 hourly data (showing 0-21 hours since 24 hrs = 0 hrs)