Interactive comment on "Forecasting of regional methane from coal mine emissions in the Upper Silesian Coal Basin using the on-line nested global regional chemistry climate model MECO(n) (MESSy v2.53)" by Anna-Leah Nickl et al.

In black we repeat the referees comments, in red are our replies.

Anonymous Referee #2

Received and published: 6 January 2020

This paper describes the setup and application of a nested atmospheric transport modeling system to support an aircraft measurement campaign with daily model forecasts of methane over a coal-mining region in Poland. The performance of the forecasts is assessed both in terms of theoretical skill (comparing 1-6 day CH4 forecasts with analyses) and in terms of "expected skill" (or actual skill) by comparison with aircraft observations of total column and in situ CH4. The model is shown to be capable of simulating the structures and amplitudes of the CH4 observations well, although this is more a qualitative than a quantitative statement since there is a large uncertainty in the underlying emission inventories and since there is no comparison with other modeling systems.

Dear Referee, thank you very much for the appreciation of our work.
Indeed there is not yet a comparison with other modeling systems, but one further study with the WRF-Stilt model is in preparation.

The paper is well written, clearly structured, and the analyses are detailed (sometimes too detailed) and sound. MECO(n) is an impressively flexible model system capable of online nesting multiple instances of a regional model (COSMO) in a global model (ECHAM). The paper presents a relevant application of the model, which takes full advantage of its nesting and online processing capabilities (e.g. sampling the model fields at each time step along aircraft trajectories).

We are very grateful for this positive valuation of the MECO(n) model system and its presented application.

Supporting aircraft measurement campaigns requires models with sufficient resolution (much better than the horizontal and vertical distance travelled during the flights), but it is not clear a priori, what resolution is really needed and whether very high resolution brings sufficient added value to justify the additional computational cost. By comparing the results of two different model instances with a resolution of 7 km and 2.8 km, respectively, the paper shows that the results (CH4 sampled along the aircraft tracks) are very comparable and that the higher resolution does not bring a great benefit, though some small-scale details were better resolved. An interesting but also surprising finding is that the model skill (evaluated against observations) did not depend clearly on forecast lead-time, i.e. a 3- or 4-day forecast performed equally well as a 2-day forecast. Unfortunately, there is little discussion of this result.

Thank you for that comment, we now discuss actual forecast skill in more detail, regarding also your comments below and the comment of referee #1.
Overall, I consider the publication acceptable with minor revisions, but I have a few main points and a number of small corrections/suggestions.

Main points:
- There are two methane tracers, PCH4 and CH4_FX, the first one representing emissions only from coal mining and the second all emissions (anthropogenic+natural) plus background CH4. The anthropogenic emissions in CH4_FX are based on EDGAR v4.2FT2010. The authors need to check how large emissions in the USCB region are in EDGAR in comparison with the total emissions of the COMET ED v1 inventory used for PCH4. To my understanding, fugitive emissions from solid fuels belong to category 1B1 (see IPCC 1996 reporting guide-lines), which is available as separate category in the EDGAR inventory. Without a comparison of these numbers, it is difficult to understand the results presented in Figure 8, which suggest and overestimation of the amplitude of CH4 enhancements for the tracer CH4_FX but an underestimation for the tracer PCH4. Furthermore, how do coal-mining emissions compare with other emissions e.g. from agriculture in this region (according to EDGAR?).

We moved this analysis from the discussion section into Sect. 3.2.2 and expanded it further, discussing CH4_FX and PCH4.
We compared total EDGAR v4.2FT2010 to CoMet ED v1. Therefore, we took every grid cell, for which we had information in CoMet ED v1, summed the emissions and compared them to the point source emissions in the respective grid cell. We further compared the 1B1 sector of EDGAR v4.2FT2010, which makes up more than 96% of the emissions. Agriculture plays a minor role, here. Only 11.18 kt/a of methane arise from EDGAR v4.2FT2010 sector 4 in the region (longitude: 18.3 °E to 19.4 °E, latitude: 49.9 °N to 50.4 °N).

- The simulation results are biased low because of a too low background. This should not be surprising considering that the simulation was initialized from a monthly climatological average of a period, when atmospheric CH4 was lower than in 2018. This bias is thus arbitrary and not of interest for the study (we are much more interested in the excursions from the background), but it dominates much of the statistics discussed and presented in the tables. I therefore suggest computing an overall bias (e.g. mean difference averaged over all flight sections measuring background) and subtract this constant offset from all simulation data, at least when computing the RMSE and NMBE statistics. In the current tables, the RMSE is of the order of 0.1 umol/mol, which is of a similar magnitude as the amplitude of the observed CH4 peaks, which would actually suggest a very poor model skill.

We revised our analysis and included a bias correction for the P and J observations. As explained in the text, for the C observations, this correction is, however, not applicable.

- The discussion on forecast skill is rather lengthy, especially the discussion of the Taylor diagrams. I found it useful to summarize the results of all model-observation comparisons in a Taylor diagram as shown in Figure 11, but I am much less convinced of the use of Figures 13 and 16 summarizing the 1- to 6-day forecast skills at a single location (Fig. 13) and for the aircraft measurements (Fig. 16). Much of the information is already conveyed by the other figures. The discussion of Figures 13 and 16 is lengthy and not providing much additional insight.

Thank you for this comment; we agree with that point and removed figures 13 and 16, as well as the corresponding text passage.

Furthermore, one should be very careful in the interpretation of the results presented in Fig. 11, since much of the findings are simply a consequence of the different flight patterns. The high correlations in the HALO in situ measurements (J1, J2), for example, are primarily due to the large altitude changes on these flights probing a large vertical gradient in CH4. But also horizontal flight patterns may critically affect the results, depending on the complexity of the pattern, the overall distance flown, the time spent in sampling background versus polluted air, etc.

Fig. 11 is Fig. 8 in the revised manuscript. We completely agree with the referee and now explain the relation to the flight patterns (and measurement technique – vertical column vs. in-situ) in the revised manuscript.

Minor points:
- Page 2, Line 28: Please explain what you mean by "internal"

“internal” simply refers to the fact that this inventory was compiled in preparation for this measurement campaign and it was not publicly available. We removed the “internal” from the revised text.

- P4, L4: You mention that the simulation data can optionally be interpolated vertically. Was such vertical interpolation applied, or were the simulated fields only taken from the closest vertical layer?

Here, we sampled the vertical “curtain” along the horizontal flight tracks on-line on the original vertical model grids for direct output. This “curtain” was then further sub-sampled onto the flight altitude by linear interpolation (for P and J data) and by integration up to the flight altitude (for C data).
We mention this in the revised manuscript on P10, L 1-2.

- P6: Figure 3 could be improved. The black text in the blue boxes is a bit difficult to read.

We chose a lighter color for the revision.

- P7, L4: Were the O1D and Cl fields obtained from a full chemistry simulation?

Yes. We added this information to the revised text.

- P7, Lines 13-18: I didn’t really understand these sentences: Why do you need an "interpolation in time"? Why are data of time steps at 06:00 UTC and 12:00 UTC needed if the nudging requires two time steps AHEAD of the simulated time, which starts at 00:00 UTC?

This is solely due to technical constraints. We use the original nudging routines of the ECHAM5 base model. The nudging is applied in every time step, and the nudging fields (here the 6-hourly forecast data) are linearly interpolated in time. In addition, SST/SIC are consistently prescribed and also linearly interpolated in time in every model time step, however, in the standard configuration only every 12 hours. For this linear interpolation, the data are required 12 hours ahead.
We added this information to the revised text.

- P8, L9: I thing HPC stands for High Performance Computing (not "Performing").
Thanks. That was probably the spell checker ...

- P9, Figure 5: Why are the grey and green dashed lines with arrows going from right to left? This seems to suggest that e.g. the forecast starting at 12:00 is branched from an analysis, which has seen nudging data between 12:00 and 24:00 on that same day. Is this really true?

No. The issue is explained by the linear interpolation (see above) of the nudging data in time. A forecast starting at 12:00 UTC from the analysis simulation, requires the analysis simulation to be advanced until 12:00 UTC, which in turn (due to the time interpolation) required analysis nudging data until 24:00 UTC of that day. That is why we can branch off FC simulations with a lag of 12 hours, only. This is indicated by the dashed arrows and we actually do not have a better idea on how to visualize this.

- Figures 7 and 8: The flight patterns or at least the altitude profiles should also be shown in the main body of the paper, not just in the supplement, because this is essential information. It is important to know, for example, whether the individual peaks correspond to different plumes or whether the same plume was sampled back and forth multiple times. It is also important to know whether changes in CH_4 mole fractions are due to changes in flight altitude rather than due to transecting a plume.

We now include figure 11, showing flight pattern and flight routes.

- Tables 2, 3 and 4: I suggest adding the correlation coefficients (or R-square as a measure of the variance explained).

We now added the correlation coefficient to all tables.

- P13, L21: I think it would be useful to show a vertical profile of CH_4 for this flight to demonstrate that the model captures the vertical gradient of CH_4 quite accurately.

We add Fig S1 to the supplement showing observations and model results as CH_4 versus pressure altitude.

- P17, equations of skill scores: Does one of these skill scores correspond to the dashed line in Figure 11? Is so, please mention.

No. The dashed line refers to the centered normalized root mean square error (NRMSE).

- "Expected skill" doesn’t sound right to me. What about "Actual skill", or "True skill"?

Thank you for this comment. Indeed, “expected skill” is not the most appropriate term. We changed it to “actual skill”.

- P18, L1: For which period (how many days) did you compare the forecasts with the analysis simulation?

6 forecast days, starting each day between June 1 and June 22, 2018. Changed in text.

Corrections:

- Page 2, Line 28: "in the Upper Silesia" -> "in Upper Silesia"

- P4, L12: I suggest using "time step" instead of "time step length" here and in the following sentences.

We prefer to keep “time step length”, which is the length of one time step in the model, whereas “time step” refers to the actual time step, e.g. the 1st, 2nd, 3rd, etc.

- P13, L4: "very precisely". I would rather say "quite precisely"

- P13, L5: "at June" -> "on June"

- P13, L12: Change to "below the top of the boundary layer"

- P13, L24: "correlate well" has a positive connotation. "correlate closely" sounds better to me in this case.

- P14, L3: Change "Contrary," to either "In contrrast," or to "On the contrary," here and at other places.

- P14, L4: "expect that the model is able" -> "expect the model to be able" C5

- P15, L8: "suit well" -> "fit well"

Thank you for all the corrections, we changed them all (except one on P4, L12, see above) according to your specific suggestions.

- P15, L9: Isn’t the NRMSE high rather than low?

You are right, this sentence is wrong. The correlation coefficient is low, but NRMSE is rather high. The sentence is changed now in the revised manuscript.

- P16, L1: "spacial" -> "spatial"

- P20, L14: "amplitude height" -> "amplitude"

- P22, L5: "This the intended result given" doesn’t sound right.

- P24, L6: "the boundary layer is too low" -> "the boundary layer is too shallow". The top of the boundary layer can be too low, but not the boundary layer itself.

- P24, L13: "PCH4 correlates well with the observed methane emissions". There was no observation of emissions but only of concentrations.

- P24, L23: "might probably" -> "might"

- P24, L22: "forecast day" -> "forecast days"

All corrected.