Author's response

Dear Dr. Williams,

please receive our revised manuscript titled "Hind- and forecasting of regional methane from coal mine emissions in the Upper Silesian Coal Basin using the online nested global regional chemistry climate model MECO(n)(MESSy v2.53)" by A. Nickl, M. Mertens, A. Roiger, A. Fix, A. Amediek, A. Fiehn, C. Gerbig, M. Galkowski, A. Kerkweg, T. Klausner, M. Eckl and P. Jöckel.

The replies to both referees, as well as a diff with the detailed textual changes in the revised manuscript are appended.

Best regards,
Anna-Leah Nickl

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 #1

Received and published: 10 December 2019

General Comments: This paper provides an evaluation of a modeling forecast system setup to forecast methane plumes emitted from coal mines in the Upper Silesian Coal Basin in Poland. The aim is to forecast the methane plumes in order to assist with the flight planning of several measurement campaigns. An evaluation of the skill of the model at two different resolutions (2.8 and 7 km) as compared to three different airborne observational datasets is presented.

The authors present a comprehensive overview of the biases in this model evaluation paper which is of interest to the scientific community and fits within the scope of the GMD journal.

Dear Referee, thank you very much for the positive view of the overall study.

However, in general the paper is lacking an in-depth interpretation of the results. More could be discussed in terms of the sources of uncertainty and error. While the authors provide a reasonable interpretation of the results in the Discussion and Conclusion sections, the paper would be more interesting to read if more interpretation and analysis was given throughout the paper, rather than just reporting the biases.

We revised the manuscript according to your specific comments below.

In terms of grammar, the paper is generally well written, but the author is inconsistent in using the past and present tense. The paper should be written consistently in either the past or present tense.

We now use consistently past tense in the revised version.

Specific Comments: Title and Abstract: I find the title and abstract misleading because most of the evaluation presented in this paper is focused on the analysis simulation rather than the forecast simulation. The title should be changed to something like "Modeling and forecasting of. . ." to reflect this.

Thank your for this hint. We changed the title, using the word "hindcasting".

In addition, the abstract and introduction should also reflect that most of the evaluation is focused on 1) assessing the impact of the model's spatial resolution on the simulation of methane plumes originating from ventilation shafts in the coal mines, 2) assessing the uncertainty in the model's methane concentrations using different air-borne measurements,

We are grateful to this specific comment and changed the abstract accordingly.

and 3) comparing the results of using two different emission inventories on peak methane concentrations over the coal mines.

As the purpose of this study was not comparing two different emission inventories we did not include such a statement in the abstract. We rather provide an improved explanation of using the EMPA/EDGAR inventory and the point source inventory in the revised manuscript and below (comment 2.2.1 Methane tracer).

1 Introduction: Page 2, Line 7: Explain isotope carbon-13 and how it can be used to infer sources of ch4 emission.

We now give an explanation on stable carbon isotopes in the introduction part.

2 Model and Forecast System: The authors mention updating the applied emissions inventory to EDGARv4.3.2 which could help in reducing the biases. They could also consider using the CAMS-GLOB-ANT anthropogenic global emissions which are currently used by the ECMWF-IFS models and are based on the EDGARv4.3.2 and CEDS inventories and extrapolated to the current year (https://eccad3.sedoo.fr/).

We are grateful for this hint for our future studies, and we mention this emission inventory in the outlook of the revised manuscript.

2.2.1 Methane Tracers: Please provide an explanation as to why you are evaluating two tracers. Is it to compare the different emission inventories? If so, it should be clearly stated. Otherwise, if it is just to get the other sources of methane emissions, and if the internal inventory of point sources is more accurate, why not replace the EDGAR emissions over the coals mines with these point sources?

The purpose of using these two tracers was not to evaluate the different emission inventories. The EMPA/EDGAR inventory (used for the CH4_FX tracer) was used to provide a methane distribution, which contains all known methane sources (within the nested regions including the “background” methane which is advected into the model domain). The point source emission inventory (CoMet ED v1) was used **in addition** for a second methane tracer tracing only the emissions from the coal mine ventilation shafts. In this way, we are able to trace methane enhancements (of the first tracer, CH4_FX, which is equivalent to what has been measured) back to the coal mine emissions. In other words, the “point source tracer” was used as additional diagnostic for tracing the individual emission solely from the point sources (which have been the focus of the campaign). In that context, it was not our purpose to evaluate the point source emission inventory (at least not in the present study), and therefore we did not replace the point sources in the EMPA/EDGAR inventory.

We improved the corresponding explanation in the revised manuscript.

The authors should also consider using the CAMS-REG-AP regional inventory for Europe which is developed by TNO in the Netherlands and can be downloaded from the ECCAD data repository

(https://eccad3.sedoo.fr/). These emissions are provided up to the year 2016 and are based on more detailed regional information than the global inventories.

Thank you for pointing on the emission inventories, which we will consider in future simulations.
One issue, however, with regional emission inventories in our nested global/regional setup is that the regional inventories need to cover at least our nested region, and that the inventories used for the global and the regional domains must be consistent (e.g. w.r.t. potential biases), because otherwise artefacts at the regional model boundaries can occur. This has to be checked carefully.

Page 7, Line 12: Explain in more detail what is meant by ". . .however, for the RCP8.5 scenario. . .". Why is a scenario used?

This sentence refers to the simulation we use for the initialization.
The initial conditions of CH4_FX are derived as monthly climatological average (2007-2016) of the simulation SC1SD-base-01, which is similar to the RC1SD-base-10 simulation (described in detail by Jöckel et al., 2016). Whereas for RC1SD-base-10 the prescribed emissions of the last available year (2011) have been used for later years as well, SC1SD-base-01 has been performed with the boundary conditions of the RCP8.5 emission scenario. We used this for initialization as our current best guess including transient boundary conditions after 2010.
Since this is confusing, we removed the sentence in the revised manuscript.

3 Evaluation of Analysis Simulation: Please clearly explain exactly what is meant by "analysis simulation" for readers who are not familiar with forecast systems. It should be stated that the analysis simulation is constrained by the meteorology.

We guess that the referee refers to section 2.3 here, because here the forecast system is explained. We rephrased accordingly:
"In order to achieve the best initial conditions of PCH4 and CH4_FX, the daily forecast simulations are branched from a continuous analysis simulation, which is essentially a hind-cast simulation until the start of the forecast day."

Furthermore, in section 3, we rephrase the first sentence:
"As the analysis simulation is nudged towards the ECMWF operational analysis data, we assume that this simulation reproduces the observed meteorology best."

3.1 Observational data: The flight pattern for J1 and J2 should also be provided. Regarding the flight pattern shown in S2 for P4 and P5, it is redundant to show altitude on the y-axis and in the color-scale. Instead latitudinal information would be more useful.

Thank you for pointing this out.
We added Figure 9, showing the flight pattern for J1 (a) and J2 (b).
We additionally added Figure 11, showing the flight pattern for P4 (a) and P5 (b), and the two flight routes (c) and (d) to the corresponding section. The flight pattern show latitude, pressure and bias-corrected CH4_FX, the flight routes show longitude, latitude, pressure at flight level and the exact positions of the coal mines. We removed the Fig. S1 from the Supplement.

3.2 Comparison with Analysis results: It is shown here that the model has a slight systematic negative bias in background CH4. Possible sources of the bias, which is presumably inherited from the global model, should be discussed (i.e. emissions, representation of chemical processes, etc.). Specifically, what impact could the prescribed OH field have on the simulated CH4 concentrations in terms of a chemical sink?

Thank you for the comment. We now discuss possible sources of the bias at the beginning of the analysis.

Moreover, since the aim of the forecast system is to simulate methane plumes arising from the coal mines, I would suggest that the rest of the analysis be performed on the anomalies (with respect to background values) of the simulated and observed methane concentrations rather than the absolute values. This would remove the model's bias in the background methane concentrations and allow for a more straightforward comparison of the peaks related to emissions from the ventilation shafts of the coal mines.

We agree. We therefore define an average background value using the most frequently occurring difference between in-situ measurements and model results, subtract this bias and redo the statistical analyses. Unfortunately, we cannot apply the same bias to the integrated total column average mixing ratios (comparison to CHARM-F measurements). This is discussed in Section 3.2.

Page 11, Line 13: It can't be seen on the graph that small scale patterns and are better resolved in CM2.8. The sentence should be taken out or more proof provided.

Thank you for the hint. We removed this sentence.

3.2.2 Comparison with in-situ measurements: In discussing Figure 8(a) and (b), the authors state that "Despite the negative bias, peak mixing ratios of CM7 and CM2.8 reach values close to those of the observations. . .", however, it seems that since the simulated background methane has a negative bias (>10 umol/mol) then the model is actually overestimating the increase in the peak methane mixing ratios in the plumes compared to the observations, in particular for CM2.8. Could the authors comment or clarify this point? An evaluation of the anomalies would have been an effective way to remove the model's systematic bias from the background methane and evaluate the model's ability to reproduce the peaks observed in the methane plumes.

Thank you for this comment. Indeed, this sentence is misleading and is actually supposed to point on the overestimation of the model. We now changed the sentence accordingly. As already stated above, we also revised the analysis after bias correction.

Can any conclusions be drawn regarding the relationship between the stability of the boundary layer, spatial resolution and the model's performance in simulating the methane plumes? Accurately simulating the PBL is critical in forecasting the methane plume. Has their model's PBL scheme been evaluated elsewhere? If so, it should be referenced here and discussed.

To our knowledge, there is only one paper by Collaud et al. (2014) analyzing the COSMO simulated PBL height. However, they did not systematically study the impact of the model resolution. Moreover,

they come to a conclusion, which contradicts our finding (e.g. Fig. 12 in the revised manuscript, Fig. 9 in the original manuscript and corresponding discussion). We want to stress, however, that our simulations are short term and cannot provide a detailed analysis of the PBL height. Yet, these questions are highly interesting for further studies, but beyond the scope of our manuscript.

The revised manuscript includes a brief discussion on that in the discussion section.

3.2.3 Taylor Diagram: I don't think that this section adds much information that hasn't already been presented in the timeseries plots. I would suggest to either summarize the results in a meaningful way or to remove it. For example, can you draw any conclusions about the model's bias with regard to the different types of observations? Why are the biases with the J observations lower than the P observations, and why do Tables 3 and 4 show the contrary? The authors should either present a full analysis of the differences in the biases (i.e. instrument type, PBL height, time of day, concentration in the plume, location, windspeed and direction, etc), or simply report on the range of uncertainty that is found using these three datasets which is already quite useful information in terms of assessing the model's skill.

The Taylor diagram does not contain any information about biases, just correlations, normalized standard deviations, and centered/normalized root mean square errors. For this, we are hesitating to remove the Taylor diagram (see also comments by referee #2), because it allows to compare all different data sets regardless of their underlying biases.

However, as stated above, we revised our statistical analysis which now contains a bias corrections for the J and P datasets, but not for the C dataset, as we explain in the revised text.

We further guess that the referee here (in agreement with referee #2) refers to overall deviations (between bias-corrected model results and observations) and not the overall bias. Therefore, we added some more discussion and analysis to the revised text.

4.1 Theoretical Forecast Skill: I'm not convinced that the Taylor diagram brings any additional information that can't be deduced from Figure 12.

We agree, the Taylor diagram does not give any additional information. Neither does the respective explanation. We removed the whole sentence including the diagram.

4.2 Expected Skill Score: Please explain exactly what the expected skill score is because the fact that the model's skill does not decrease in the same manner as the theoretical skill score does not make sense to me. If we assume that the analysis is a "perfect simulation", and compared to the forecast simulation the theoretical forecast skill decreases to almost zero by day 6, how is it possible that the expected forecast skill in comparison to the observations is essentially the same on day 1 as day 6?

We came to the conclusion that the term "expected skill score" is badly chosen and highly misleading. This has also been pointed out by referee #2. Thus, we now use "actual skill score", which refers to the real observations. We give an explanation on different results of the two skills:
"Whereas the theoretical skill score is defined to measure the skill, averaged over the entire model domain, the actual skill score compares the model results to observational data. The latter measure the

downwind methane plumes, which are easier to forecast than the variability of the methane background in the overall model domain."

The authors present the model biases using different observations but more explanation or interpretation would be appreciated. For example, on page 19 line 1 it is stated that ". . .Sv is highest for J1 and J2. . ." but no explanation/speculation is offered as to why.

According to the different flight patterns between the J, C and the P observations, J "measures" the vertical gradient, P the small scale horizontal gradients, and C the larger scale horizontal gradients of the column integrated methane. Thus, the deviations of the model results seem to differ or to be inconsistent, which is however not the case. This is better explained in the revised text.

Again, I don't think that the Taylor diagram presented in Figure 16 adds any new information. It is clear from the plots in Figures 14 and 15, that the model's skill score for predicting the J observations is higher than for the other observations, especially the P observations. What would be interesting is for the authors to offer an explanation as to why this is the case. Why is there more variability in the skill score for the P observations than for the J or C observations? Unless the authors can draw some interesting conclusions such as this, I would suggest removing the Taylor diagrams and replacing them with the HALO and D-FDLR flight patterns.

We removed Figure 16 and the corresponding section.

Page 20, line 7: The authors state "All forecast days show a normalized standard deviation close to 1. . .meaning that all forecast days show similar amplitudes. . .". In theory, this can't be deduced from the standard deviation alone.

The referee is right! However, we removed this section in the revised manuscript, according to your comment above.

Technical Corrections: The author should go through the entire paper, especially (but not only) Section 3.1 and make sure they are consistent with either using the past or present tense.

We now use past tense in the revised manuscript. Section 3.1 still uses some past tense, as it refers to the sampling of the data in 2018.

Abstract: Line 4: change "measuring" to "measurement" Line 8: Change the sentence to read "In order to help with the flight planning during the campaigns. . ."

We changed both in the revised manuscript.

1 Introduction: page 2, Line 20: change "climate change strategies" to "climate change mitigation strategies"

We added the term "mitigation" to the sentence.

2 Evaluation of Analysis Simulation: Page 12, line 10: change sentence to ". . .observed peaks in the afternoon flight are lower those of the morning flight."

Page 13, line 4: change “very precisely” to “more precisely”
Page 13, line 6: change “constant offset” to “systematic bias”
Page 13, line 12: change “. . .simulated boundary layer. . .” to “. . .simulated boundary layer height. . .”

We corrected all points in the revised manuscript.

5 Discussion: page 22, line 5: There is something wrong with the sentence “This the intended result given the fact, . . .”. Perhaps a word is missing.

We changed the sentence to: “This is the intended result considering that . . .”.

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 CH4 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 CH4 for this flight to demonstrate that the model captures the vertical gradient of CH4 quite accurately.

We add Fig S1 to the supplement showing observations and model results as CH4 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.

~~Forecasting~~ Hind- and 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)

Anna-Leah Nickl[1], Mariano Mertens[1], Anke Roiger[1], Andreas Fix[1], Axel Amediek[1], Alina Fiehn[1], Christoph Gerbig[2], Michal Galkowski[2,3], Astrid Kerkweg[4,*], Theresa Klausner[1], Maximilian Eckl[1], and Patrick Jöckel[1]

[1]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Max Planck Institute for Biogeochemistry, Jena, Germany
[3]AGH University of Science and Technology, Krakow, Poland
[4]Institute of Geosciences and Meteorology, University of Bonn, Germany
[*]now at: Research Center Juelich, Institute of Energy and Climate Research, Juelich, Germany

Correspondence: Anna-Leah Nickl (anna-leah.nickl@dlr.de)

Abstract. Methane is the second most important greenhouse gas in terms of anthropogenic radiative forcing. Since pre-industrial times, the globally averaged dry mole fraction of methane in the atmosphere has increased considerably. Emissions from coal mining are one of the primary anthropogenic methane sources. However, our knowledge about different sources and sinks of methane is still subject to great uncertainties. Comprehensive ~~measuring~~ measurement campaigns, as well as reliable chemistry climate models, are required to fully understand the global methane budget and to further develop future climate mitigation strategies. The CoMet 1.0 campaign (May to June 2018) combined airborne in-situ, as well as passive and active remote sensing measurements to quantify the emissions from coal mining in the Upper Silesian Coal Basin (USCB, Poland). Roughly 502 kt of methane are emitted from the ventilation shafts per year. In order to help ~~the campaigns flight planning~~ with the flight planning during the campaigns, we performed 6-day forecasts using the on-line coupled, three times nested global and regional chemistry climate model MECO(n). We applied three nested COSMO/MESSy instances going down to a spatial resolution of 2.8 km over the USCB. The nested global/regional model system allows for the separation of local emission contributions from fluctuations in the background methane. Here, we introduce the forecast setup and assess the ~~model skill~~ impact of the model's spatial resolution on the simulation of methane plumes from the ventilation shafts. Uncertainties in simulated methane mixing ratios are estimated by comparing different ~~observations with the individual forecast~~ air-born measurements to the simulations. Results show that MECO(3) is able to simulate the observed methane plumes and the large scale patterns (including vertically integrated values) reasonably well. Furthermore we receive reasonable forecast results up to forecast day four.

1 Introduction

In terms of radiative forcing methane is the second most important anthropogenically altered greenhouse gas (Myhre et al., 2013). The globally averaged dry mole fraction of methane has increased rapidly since 2007 (Nisbet et al., 2014, 2016), and its growth has even accelerated in 2014 (Nisbet et al., 2019; Fletcher and Schaefer, 2019), where the annual rise was 12.7 ± 0.5 ppb. (Nisbet et al., 2019). The reason for the rapid methane growth in the atmosphere is currently under debate and discussed in several studies (Schaefer et al., 2016; Nisbet et al., 2016, 2019; Saunois et al., 2017; Thompson et al., 2018). The largest increase of methane is observed in the tropics and midlatitudes (Nisbet et al., 2019). Differences in isotopic methane source signatures (δ^{13}C and δD) can further help to constrain different source contributions (e.g. of thermogenic or biogenic origin) to the global methane budget. A depletion in global δ^{13}C indicates a shift from fossil fuel emissions towards more microbial sources (Schaefer et al., 2016; Nisbet et al., 2016, 2019). Nisbet et al. (2016) suggest natural emissions from wetlands as a result of positive climate feedback are the primary source of the methane enhancement. On the contrary, Schaefer et al. (2016) propose that the increase in atmospheric methane since 2007 mainly originates from enhanced agricultural activity. Additionally, a change of the atmospheric oxidation capacity, i.e. a reduction of the OH sink, could play a role and may explain the shift in isotopic signature (Rigby et al., 2017). Increasing fossil fuel emissions could also explain the rise in atmospheric methane (Thompson et al., 2018). Shale gas is more depleted in δ^{13}C relative to conventional gas and could be associated with the observed global ~~depeltion~~ depletion in δ^{13}C, too (Howarth, 2019). And Schwietzke et al. (2016) pointed out that the fossil fuel emissions are 20 % to 60 % higher than previously thought. However, we still do not fully understand all factors that affect the sources and sinks of methane (Saunois et al., 2016). Furthermore, a reduction of anthropogenic emissions is attractive and inexpensive and due to its relatively short lifetime ($\sim$ 9 years), it could rapidly cause a change in the global methane budget (Dlugokencky et al., 2011). Comprehensive measurements and the use of chemistry climate models can therefore help to improve further climate change projections, and to develop potential climate change mitigation strategies.

The AIRSPACE project (Aircraft remote sensing of greenhouse gases with combined passive and active instruments) aims for a better understanding of sources and sinks of the two most important anthropogenic greenhouse gases: carbon dioxide and methane. Several measurement campaigns within the project e.g. CoMet (Carbon Dioxide and Methane Mission), are carried out to increase the number of airborne and ground-based (Luther et al., 2019) measurements of CO_2 and CH_4. CoMet 0.5 in August 2017 combined ground based in-situ and passive remote sensing measurements in the Upper Silesian Coal Basin (USCB) in Poland, where large amounts of methane are emitted due to hard coal mining (roughly 502 kt CH_4/yr, CoMet internal CH_4 and CO_2 emissions over Silesia, version 2 (2018-11), further denoted as CoMet ED v2). CoMet 1.0, that took place in May and June 2018, additionally included airborne in-situ as well as passive and active remote sensing measurements in ~~the~~ Upper Silesia and Central Europe. In order to localize the methane plumes and to obtain the best measurement strategies for the campaigns, it is helpful to have reliable forecasts of the methane distribution in the atmosphere. We performed model based forecasts over the entire period of the campaigns using a coupled global and regional chemistry climate model. While local features are often not resolved in global climate models, it is important for the CoMet forecasts to resolve the local methane emissions from the coal mining ventilation shafts in the USCB. Therefore a smaller scale atmospheric chemistry

model is required, which is provided by the on-line coupled model system "MESSyfied ECHAM and COSMO models nested n times" (MECO(n), Kerkweg and Jöckel, 2012b; Mertens et al., 2016). To increase the resolution of our forecasts, we apply a nesting approach with three simultaneously running COSMO/MESSy instances down to a spatial resolution of 2.8 km. Section 2.2 ~~describes~~ presents the model setup and the implementation of two different methane tracers. We describe the details of the new forecast system (Sect. 2.3) and discuss its evaluation. We ~~evaluated~~ evaluate the model performance by comparing the methane mixing ratios simulated by the two finest resolved COSMO/MESSy instances with airborne observational data. In Sect. 3 we show the comparisons with data that were sampled using three different measuring methods during the CoMet 1.0 campaign. Moreover, we assess the forecast performance firstly by internal comparison of the individual forecast days with the analysis ~~simulations of CoMet 0.5 and CoMet~~ simulation of CoMet 1.0 (Sect. 4.1) and secondly by comparison of the forecast results with the observations of CoMet 1.0 (Sect. 4.2).

2 Model and Forecast System

2.1 Model Description

The numerical global chemistry climate model ECHAM/MESSy (EMAC, Jöckel et al., 2010) consists of the Modular Earth Submodel System (MESSy) that is coupled to the general circulation model ECHAM5 (Roeckner et al., 2006). EMAC comprises various submodels that describe different tropospheric and middle atmospheric processes. It is operated with a 90 layer vertical resolution up to about 80 km altitude, a T42 spectral resolution (T42L90MA) and a time step length of 720 s. For our purpose, EMAC is nudged by Newtonian relaxation of temperature, vorticity, divergence and the logarithm of surface pressure towards the ECMWF operational forecast or analysis data. Sea surface temperature (SST) and sea ice coverage (SIC), that are also derived from the ECMWF data sets, are prescribed as boundary conditions. The EMAC model is used as a global driver model for the coarsest COSMO/MESSy instance.

The model COSMO/MESSy consists of a Modular Earth Submodel System (MESSy, Jöckel et al., 2005) which is connected to the regional weather prediction and climate model of the Consortium for Small Scale Modelling (COSMO-CLM further denoted as COSMO, Rockel et al., 2008). The COSMO-CLM is the community model of the German regional climate research community jointly further developed by the CLM-Community. Details on how the MESSy infrastructure is connected to the COSMO model are given in the first part of four MECO(n) publications (Kerkweg and Jöckel, 2012a). Several COSMO/MESSy instances can be nested on-line into each other in order to reach a regional refinement. For chemistry-climate applications the exchange between the driving model and the respective COSMO/MESSy instances at its boundaries must occur with high frequency. This is important to achieve consistency between the meteorological situation and the tracer distribution. Furthermore, the chemical processes should be as consistent as possible. In MECO(n) the model instances are coupled on-line to the respective coarser COSMO/MESSy instance. The coarsest COSMO/MESSy instance is then on-line coupled to EMAC. ~~Contrary~~ In contrast to the off-line coupling, the boundary and initial conditions are provided by direct exchange via computer memory using the Multi-Model-Driver (MMD) library. This coupling technique is described in detail in Part 2 of the MECO(n) documentation series (Kerkweg and Jöckel, 2012b). The chemical processes are described in submodels,

which are part of MESSy. These submodels do not depend on spatial resolution and can be used similarly in EMAC and all COSMO/MESSy instances. A detailed evaluation of MECO(n) with respect to tropospheric chemistry is given in the fourth part of the MECO(n) publication series (Mertens et al., 2016). In the present study we use MECO(3) based on MESSy version 2.53.

The MESSy submodel S4D (Jöckel et al., 2010) on-line samples the model results along a specific track of a moving object, such as air planes or ships. The simulation data is horizontally (and optionally also vertically) interpolated to the track and sampled at every time step of the model. This guarantees the highest possible output frequency (each model time step) of respective vertical curtains along the track. The submodel SCOUT (Jöckel et al. 2010) on-line samples the model results as a vertical column at a fixed horizontal position. The high frequency model output is useful for comparison with stationary observations, such as ground-based spectroscopy or lidar measurements.

2.2 Model Setup

To resolve the local emissions from the ventilation shafts in the USCB, we ~~operated~~ operate MECO(n) with three nested instances, MECO(3), see Figure 1. The first COSMO/MESSy instance (hereafter called CM50) covers the European area and is operated at a resolution of 0.44° (~50 km) and with a time step length of 240 s. CM50 is on-line coupled to EMAC, resulting in a direct exchange of boundary conditions between the global model and the regional COSMO/MESSy model.
The second COSMO/MESSy instance (hereafter called CM7) covers the area over Central Europe and is operated with a resolution of 0.0625° (~7 km) and a time step length of 60 s. The smallest instance (hereafter called CM2.8) covers the Upper Silesia in Poland and thus also the target region. CM2.8 has a resolution of 0.025° (~2.8 km) and a time step length of 30 s. The individual finer COSMO/MESSy instances (CM7 and CM2.8) are on-line driven from the respectively coarser model domain (CM50 and CM7). In doing so, the respective coarser domain provides the boundary data for the smaller domain at each of its model time steps (EMAC: 720 s → CM50: 240s → CM7: 60s → CM2.8: 30s). Figure 2 shows an overview of the initial and boundary data exchange between the different domains. CM50 and CM7 are operated with 40 vertical layers, and the smallest domain CM2.8 is operated with 50 vertical layers, that cover the atmosphere from the surface up to an altitude of 22 km. A sponge zone begins at 11 km which reaches the models top and nudges the models prognostic variables with increasing weights towards the driving model.

2.2.1 Methane tracers

CoMet aims to quantify the methane emissions in the USCB region, which actually arise from coal mining. In order to separate these emissions within our model, we ~~defined~~ define two different methane tracers. One tracer takes into account all methane emission fluxes (hereafter called CH4_FX) and includes the background methane, which is advected into the model domain. The second tracer (hereafter called PCH4) only considers the point source emissions of the ventilation shafts~~(hereafter called PCH4) and the other tracer takes into account all methane emission fluxes (hereafter called~~ . In this way, we are able to trace

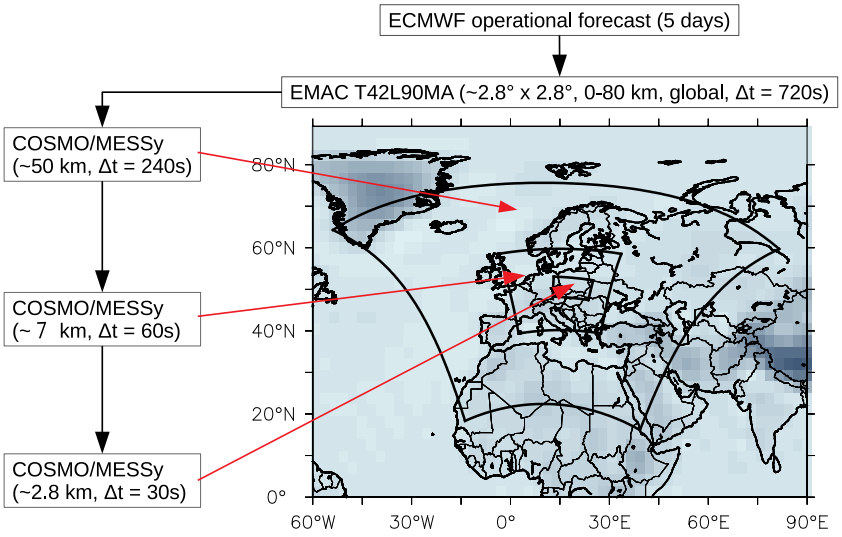

Figure 1. Overview of all three COSMO/MESSy domains over Europe (CM50), over Central Europe (CM7) and over the USCB in Poland (CM2.8), as well as the corresponding temporal and spatial resolution. The black arrows indicate the data exchange between the different models. The driving model EMAC is nudged towards divergency, vorticity, temperature and the logarithm of surface pressure from ECMWF. SST and SIC are prescribed as boundary conditions.

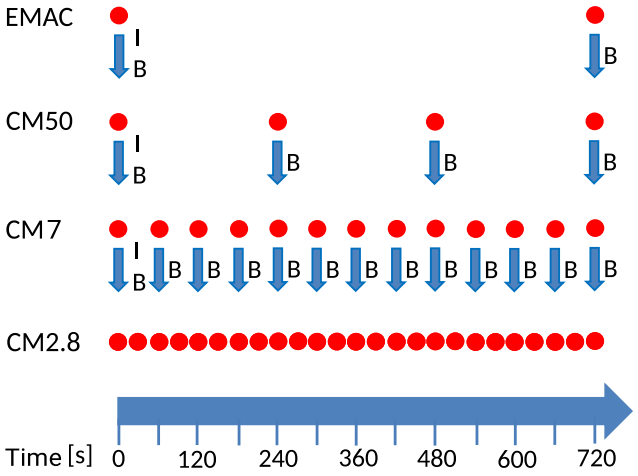

Figure 2. The illustration shows the initial and boundary data exchange between EMAC and the different COSMO/MESSy instances. The blue arrows symbolize the data exchange between the different model instances. B stands for boundary data and I for initial data. The red circles visualize the specific time steps of data exchange.

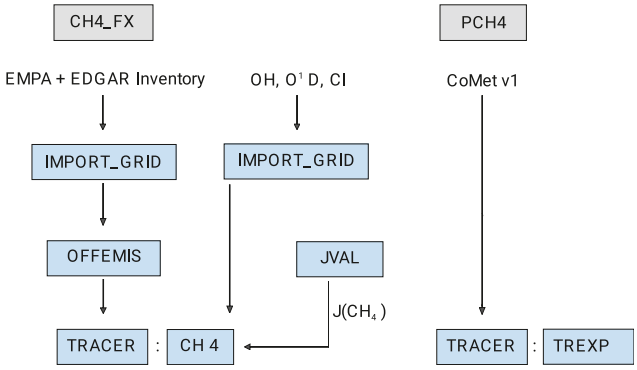

Figure 3. Illustration of submodels which are used for the different methane tracers. CH4_FX tracer (left side): Methane emission data and the oxidation reaction partners OH, O^1D and Cl are read from the netcdf files and transformed to the computational grid by the submodel IMPORT_GRID. OFFEMIS converts the emission fluxes into tracer tendencies, and the CH4 submodel simulates the chemical loss of methane using the predefined fields of the oxidation partners and the calculated photolysis rate from the submodel JVAL. PCH4 (right side): submodel TREXP is used for the point source emissions and tracer definition.

back the methane enhancements of the first tracer CH4_FX ~~)~~(equivalent to what has been measured) to the coal mine emissions. Figure 3 shows an overview of both tracers, the involved submodels and the corresponding emission inventories. We ~~initialized~~ initialize these two independent tracers for EMAC and for all three COSMO/MESSy instances equally. The initial conditions for the forecast simulations are derived from a continuous analysis simulation, which is described in detail in Sect. 2.3.

CH_4 Point sources (PCH4)

The PCH4 tracer considers only point source emissions, that are emitted by the ventilation shafts of the various coal mines in the USCB. In Fig. 4 the entire territory (49.90°N - 50.40°N latitude and 18.30°E - 19.40°E longitude) together with the location of the ventilation shafts are shown. To prescribe the emissions coming from the different shafts, we ~~used~~ use CoMet ED v1, an ~~internal~~ inventory mainly based on the European Pollutant Release and Transfer Register (E-PRTR 2014, Retrieved from http://prtr.eea.europa.eu, Feb 08, 2017), but also on data from Wyzszy Urzad Gorniczy 2014 (Retrieved from http://www.wug.gov.pl/download/5710.pdf, Feb 08, 2017). Further details on the names and exact position of the different mines can be found in the Supplement. The total point source methane emissions in this area are estimated to be 465 kt/a (CoMet ED v1 inventory). Emissions of single coal mines are split equally between the corresponding ventilation shafts. For the definition of point sources, we ~~applied~~ apply the MESSy submodel TREXP that is described in detail by Jöckel et al. (2010).

Gridded methane emissions (CH4_FX)

The second tracer is called CH4_FX and includes all methane emission fluxes, anthropogenic and natural. We ~~used~~ use an inventory which consists of two different parts, both monthly averaged: the year 2012 of the EMPA inventory (Frank, 2018)

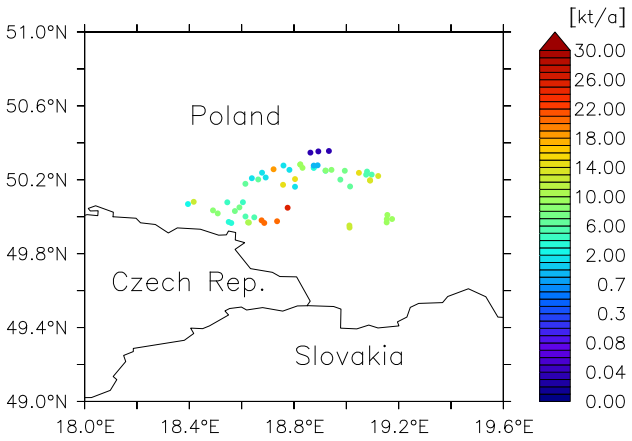

Figure 4. The map shows the locations and the emissions of methane in tons per year of the ventilation shafts in the USCB (CoMet ED v1). All ventilation shafts are gathered in the south-west of Poland close to the polish city Katowice and the Czech border.

with a 1.0° × 1.0° grid resolution and the EDGAR v4.2FT2010 (Retrieved from http://edgar.jrc.ec.europa.eu, May 30, 2017) inventory with a finer grid resolution of 0.1° × 0.1°. All anthropogenic (including rice cultivation) emissions are used from the EDGAR v4.2FT2010. Natural emissions and emissions caused by biomass burning are used from the EMPA inventory. The emission data are imported and transformed to the computational grid (IMPORT_GRID, Kerkweg and Jöckel, 2015). The emission fluxes are then converted into tendencies of the tracer CH4_FX (OFFEMIS, Kerkweg and Jöckel, 2012b, therein described as OFFLEM). Processes that are related to the methane chemistry in the model are described in the MESSy submodel CH4 (Frank, 2018). The submodel simulates the chemical loss of methane including the depletion by photolysis rate calculated by the submodel JVAL (Sander et al., 2014). The CH4 submodel uses predefined fields of the oxidation reaction partners OH, O^1D and Cl which, for our setup, ~~were~~ are derived as monthly averages (2007-2016) from a previous interactive chemistry simulation and read by IMPORT_GRID.

2.3 The Forecast System

In order to achieve the best initial conditions of PCH4 and CH4_FX, the daily forecast simulations are branched from a continuous analysis simulation, which is essentially a hind-cast simulation until the start of the forecast day. In the analysis simulation EMAC is nudged by Newtonian relaxation of temperature, vorticity, divergence and the logarithm of surface pressure towards the 6-hourly ECMWF operational analysis data. SST and SIC, derived from the same data set, are prescribed as boundary conditions for EMAC. The initial conditions of CH4_FX ~~have been~~ are derived as monthly climatological average (2007-2016) of the simulation SC1SD-base-01, which is similar to the RC1SD-base-10 simulation ~~(Jöckel et al., 2016), however for the RCP8.5 emission scenario~~(described in detail by Jöckel et al. (2016)). PCH4 is initialized with zero. The starting ~~dates~~ date

of the analysis simulations ~~are August 1st, 2017 for CoMet 0.5 and~~ is April 1st, 2018~~for CoMet 1.0. This~~ , which results in a spin up time of ~~8 days and~~ 45 days~~, respectively. For the~~ . Nudging is applied in every model time step. The nudging fields (6-hourly data) and the prescribed SST and SIC (12-hourly data) are linearly interpolated in time. For this interpolation in time, starting and continuing the analysis simulation requires two nudging time steps ahead of the simulated time. An analysis simulation which should start at 00:00 UTC, hence requires the nudging data of the time steps 06:00 UTC and 12:00 UTC. Once the respective time period is simulated and the corresponding restart file is written, a new forecast simulation is triggered. The forecast branches as a restart from the analysis simulation and simulates a time period of six days by using the 6-hourly ECMWF operational forecast data for the EMAC nudging. PCH4 and CH4_FX are automatically initialized from the restart files. Throughout this process the analysis simulation continues. The forecast system is visualized schematically in Fig. 5. As soon as the pre-processed nudging files become available, the analysis simulation runs for about 50 min. Each forecast simulation takes about 8 hours and the post processing takes another 1.5 to 2 hours. The 8 hours are for 144 message passing interface (MPI) tasks on an Intel Xeon E5-2680v3 based Linux Cluster (6 nodes à 12 dual cores), whereby 6, 18, 56, and 64 tasks were used for the model instances EMAC CM50, CM7, and CM2.8, respectively. In our example, a forecast that simulates a time period starting at forecast day one at 00:00 UTC, is readily post-processed on forecast day two at around 04:30 UTC (after approximately 28.5 hours). Throughout both campaigns, forecasts were delivered every 12 hours and made available online on a web page. In order to guarantee a continuous and uninterrupted supply of forecasts, we run the simulations alternately on two independent HPC (High ~~Performing~~ Performance Computing) Clusters. An example of a forecast web product, which shows the forecast starting on June 07, 2019 at 00:00 UTC can be found here: https://doi.org/10.5281/zenodo.3518926 (Jöckel et al., 2019). The post-processing ~~included~~ include the vertical integration of PCH4 and CH4_FX into a total column dry air average mixing ratio, called $XPCH_4$ and XCH_4 for PCH4 and CH4_FX, respectively. It is calculated as follows:

$$XCH_4 = \frac{\sum(\chi_{CH_4} \cdot m_{dry})}{\sum m_{dry}} \quad (1)$$

where χ_{CH_4} is the methane mixing ratio, m_{dry} stands for the mass of dry air in a grid box and summation is carried out over all vertical levels. Figure 6 shows the design of $XPCH_4$ and XCH_4 which appeared on the forecast website. It is an example of a snapshot during CoMet 1.0 simulated with CM2.8.

3 Evaluation of Analysis Simulation

3.1 Observational Data

During CoMet 1.0, methane was measured by active remote sensing. The instrument is an "integrated path differential absorption" (IPDA) Lidar called CHARM-F (Amediek et al., 2017), which was installed on board of the German Research Aircraft HALO (High Altitude and LOng Range). CHARM-F ~~is~~ was operated by the German Aerospace Center (DLR) in Oberpfaffenhofen and ~~measures~~ measured the weighted atmospheric columns of the methane dry-air mixing ratio from the surface to

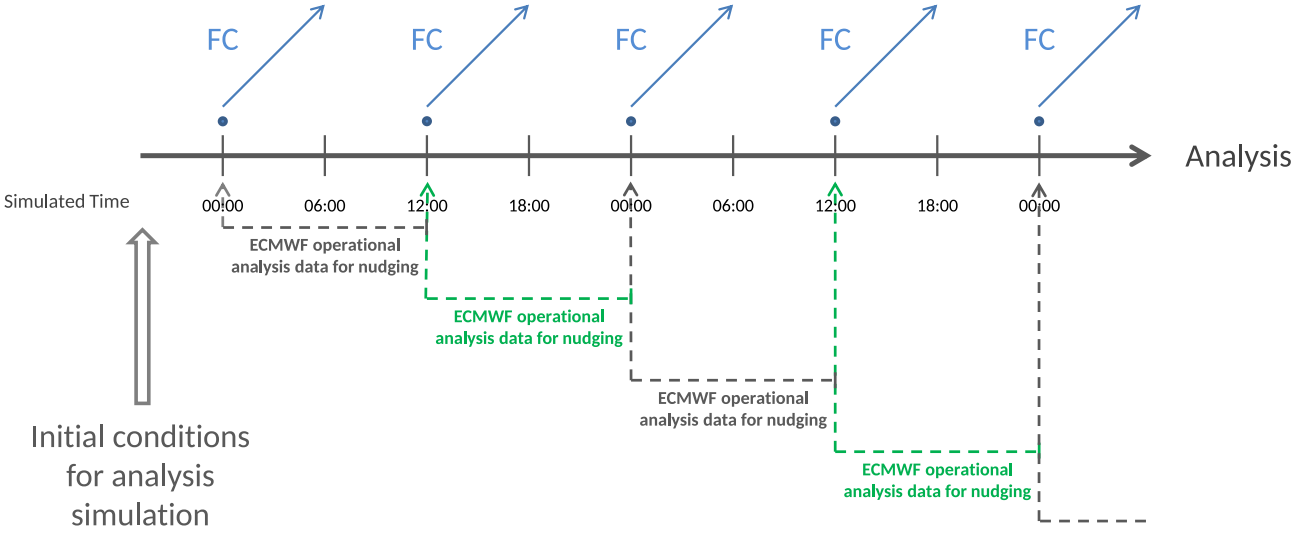

Figure 5. Chronology of the analysis simulation (dark grey) and the branching of the forecasts (FC, blue). The analysis simulation continues as soon as the ECMWF operational analysis data (two time steps ahead) are available for nudging. The required nudging time steps are indicated by the dotted lines in grey and green. A forecast simulation is branched (blue dots) every 12 hours from the analysis at 00:00 or 12:00 and simulates a time period of 6 days. The initial conditions are provided by restart files of the analysis simulation.

the flight altitude of the research aircraft. We compare our model results to the observations of the HALO D-ADLR flights on June 6th, 2018 and on June 7th, 2018. For simplicity, both data sets are hereafter called C1 and C2. See also Table 1, which lists all flights considered in this study and their abbreviations. Both data sets have a temporal resolution of 1 s and are already smoothed horizontally with a box window corresponding to 2 km flight distance.

Additionally, methane was sampled in-situ by Cavity Ring Down Spectroscopy (CRDS). A JIG instrument (Jena Instrument for Greenhouse Gases, Filges et al., 2015), that ~~also measures~~ measured the methane mixing ratio in-situ by CRDS, was installed on board of HALO and operated by the Max-Plank Institute for Biogeochemistry in Jena. Our model results are compared to the observations of the HALO flights on June 6th, 2018 and on June 7th, 2018. Data sets are abbreviated with J1 or J2 (see Table 1). Both data sets have a temporal resolution of 1 s. A Picarro CRDS G1301-m instrument was installed on board of the DLR research aircraft Cessna 208B (D-FDLR) and operated by the DLR in Oberpfaffenhofen. We compare seven flight observations to our model. Data sets are named accordingly P1 - P7 (see Table 1) and have a temporal resolution of 1 s.

Upon completion of CoMet 1.0, we conducted the analysis and forecast simulations again and used the specific geographical flight track coordinates (in degrees), pressure altitudes (in hPa) and time steps (in UTC) of all flights for the S4D submodel.

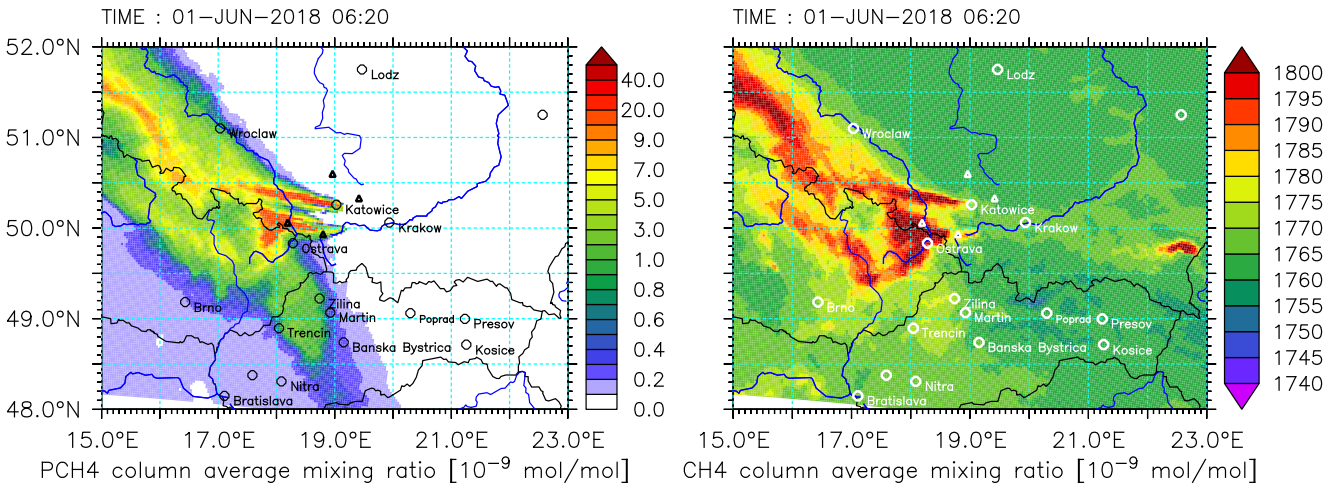

Figure 6. Snapshot of the methane forecasts during CoMet 1.0 simulated with the finest resolved COSMO/MESSy instance CM2.8. The total column dry air average mixing ratio in mol/mol is calculated for PCH4 (left) and CH4_FX (right). The area encompasses the USCB and shows the evolution of methane plumes in the atmosphere. Note that the color bar on the left is pseudo logarithmic for better visualization.

Table 1. Overview of the abbreviations for all observational methane data sets.

Abbreviation	Flight	Instrument	Type of observation
C1	HALO, 6th of June 2018	CHARM-F	XCH_4
C2	HALO, 7th of June 2018	CHARM-F	XCH_4
J1	HALO, 6th of June 2018	JIG (with Picarro CRDS G2401-m)	in-situ
J2	HALO, 7th of June 2018	JIG (with Picarro CRDS G2401-m)	in-situ
P1	D-FDLR, 29th of May 2018	Picarro CRDS G1301-m	in-situ
P2	D-FDLR, 1st of June 2018	Picarro CRDS G1301-m	in-situ
P3	D-FDLR, 5th of June 2018	Picarro CRDS G1301-m	in-situ
P4	D-FDLR, 6th of June 2018, morning	Picarro CRDS G1301-m	in-situ
P5	D-FDLR, 6th of June 2018, afternoon	Picarro CRDS G1301-m	in-situ
P6	D-FDLR, 7th of June 2018	Picarro CRDS G1301-m	in-situ
P7	D-FDLR, 11th of June 2018	Picarro CRDS G1301-m	in-situ

The simulated data were then sampled as track-following curtains at each model time step; i.e. every 720 s, 240 s, 60 s and 30 s for EMAC, CM50, CM7 and CM2.8, respectively. However, our evaluation in this study only considers the two finest COSMO/MESSy instances CM7 and CM2.8. For the comparisons with the in-situ observation, the curtain is further sub-sampled onto the flight altitude by linear interpolation. As the observed data ~~has~~ have a finer temporal resolution than the

model output, they ~~were~~ are averaged over 60 s for CM7 and over 30 s for CM2.8. In order to compare our model results with those of the CHARM-F measurements, we ~~calculated~~ calculate the dry air mixing ratio between surface and aircraft (in the following referred to as $X_{fl}CH4$) using the S4D submodel output.

3.2 Comparison with Analysis results

As the analysis simulation is ~~already~~ nudged towards the ECMWF operational analysis data, we assume ~~to reproduce the best possible meteorology~~ that this simulation reproduces the observed meteorology best. Thus, in order to find the best estimate of our model performance, the observations are compared to the analysis simulation results first. The model performance is analyzed with respect to pattern similarity and amplitude, i.e. root mean square error (RMSE), standard deviation, correlation coefficient and normalized mean bias error (NMBE):

$$NMBE = \frac{\sum(\chi_{sim} - \chi_{obs})}{n \cdot \overline{\chi}_{obs}} \cdot 100 \quad (2)$$

where ~~is~~ χ_{sim} is the simulated methane mixing ratio, χ_{obs} stands for the observed methane mixing ratio and the summation is over all n time steps. ~~The results are~~ Compared to the observations, all CH4_FX model results are equally biased towards lower methane mixing ratios (see Fig. S2-S5 in the Supplement). The systematic bias is due to the lower background methane in the model. This is likely caused by the initialization from a monthly climatological average of a period between 2007 and 2016 (SC1SD-base-01, see 2.3), where global methane mixing ratios were lower than in 2018 (Nisbet et al., 2019). As OH is initialized from the same simulation (also as monthly climatological mean) as methane, the OH field might play a role, too. Yet, due to the short simulation period, this should not have a significant influence in our MECO(n) simulations. In order to evaluate the anomalies resulting from coal mine emissions, rather than the discrepancies in background methane, we apply a bias correction to all model results involved in statistical comparisons with the in-situ observations. For this purpose, we define an average bias of 0.108 μmol/mol using the most frequently occurring difference between all D-FDLR in-situ observations and the model results of instance CM2.8. Biases between CHARM-F and XCH4_FX are lower than the average offset of 0.108 μmol/mol and it is difficult to determine a definite offset. Integration over a varying number of model levels or changes of topography, which are not resolved by the model, could be an explanation for this. The bias correction is therefore not applied to the vertically integrated values, which are compared to CHARM-F observations. The results are presented in Sect. 3.2.1 (CHARM-F) and 3.2.2 (D-FDLR and HALO in-situ). In Sect. 3.2.3 we discuss all statistical results graphically.

3.2.1 Comparison with CHARM-F observations

Figure 7 shows the observed XCH_4 values of C1 and C2 as black lines. Red and blue dots display the simulated $X_{fl}CH_4$ of CM7 and CM2.8, respectively and all methane mixing ratios are given in μmol/mol. On both days the observed patterns agree well with the simulated patterns. Peaks in the observed methane mixing ratios are represented in CM7 as well as in CM2.8. From a visual point, amplitudes also appear to be similar. Mismatches can be seen on 6 June, around 09:30 UTC (see Fig. 7 (a)) where model results are slightly shifted in time. Observed XCH_4 values follow a negative trend until 10:10 UTC, which is not simulated by the models. On June, 7 the observed amplitudes are larger than those of the model results. On both days CM7 and

Table 2. Summary of the results of the statistical analysis of C1 and C2 compared to the the simulated $X_{fl}CH4$. Listed are the root mean square error (RMSE) in μmol/mol~~and~~ , the normalized mean bias error (NMBE) in % ~~,~~ and the correlation coefficient (R) for the model domains CM7 and CM2.8.

Flight	RMSE (CM7)	RMSE (CM2.8)	NMBE (CM7)	NMBE (CM2.8)	R (CM7)	R (CM2.8)
C1	0.08	0.08	-4.1	-4.2	0.47	0.47
C2	0.10	0.10	-5.3	-5.3	0.75	0.75

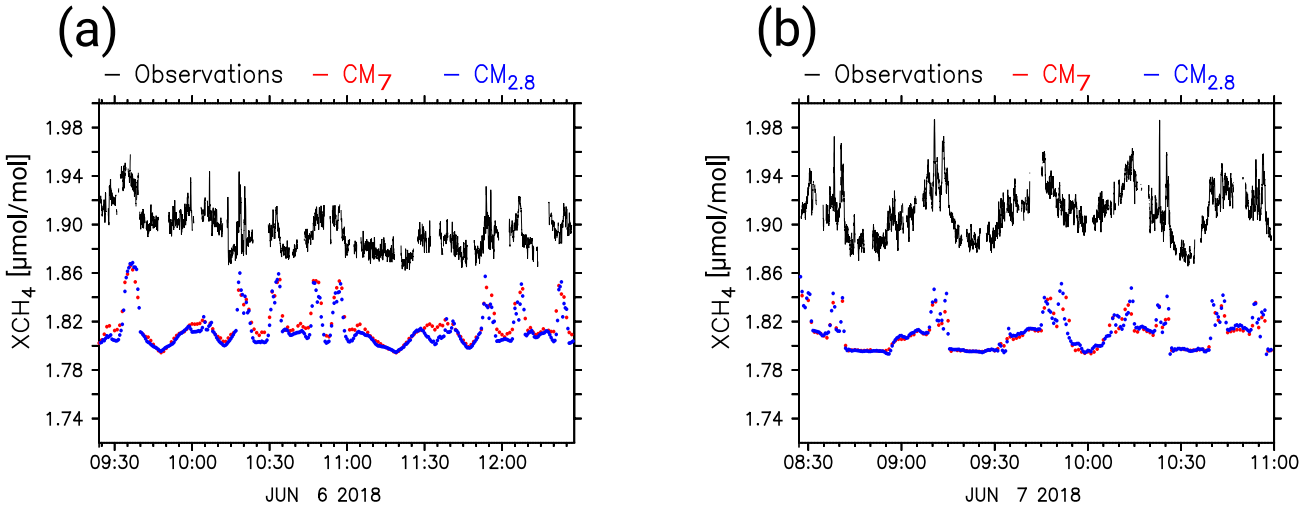

Figure 7. Results of the CHARM-F measurements and the S4D submodel sampling which was vertically integrated to yield $X_{fl}CH4$. XCH_4 of C1 (a) and C2 (b) are displayed in black and the simulated $X_{fl}CH_4$of CM7 and CM2.8 are shown as red and blue dots, respectively. Mixing ratios are shown in μmol/mol. The time axes display the time of the specific flight in UTC.

CM2.8 do not differ significantly from each other. ~~However, small scale patterns are better resolved in CM2.8 and amplitudes slightly exceed those of CM7.~~ Furthermore, the comparisons reveal a continuous and constant bias. Simulated $X_{fl}CH_4$values are shifted towards smaller values compared to C1 and C2. Table 2 lists the root mean square error (RMSE) in μmol/mol~~and~~ , the NMBE in %, and the correlation coefficient for all comparisons of the observations with the model results. NMBE is negative for all cases and ranges from -4.1 % to -5.3 %. NMBE as well as RSME are lower for C1 (0.08 μmol/mol) as for C2 (0.10 μmol/mol), which confirm the assumption of higher mean amplitude similarity (standard deviation) on June, 6 as on June, 7.

3.2.2 Comparison with in-situ measurements

Figure 8 shows the HALO in-situ measurements of methane J1 and J2 in black, along with the simulated CH4_FX of CM7 and CM2.8 in red and blue, respectively. In addition, atmospheric pressure along the flight track is plotted in hPa and indicates

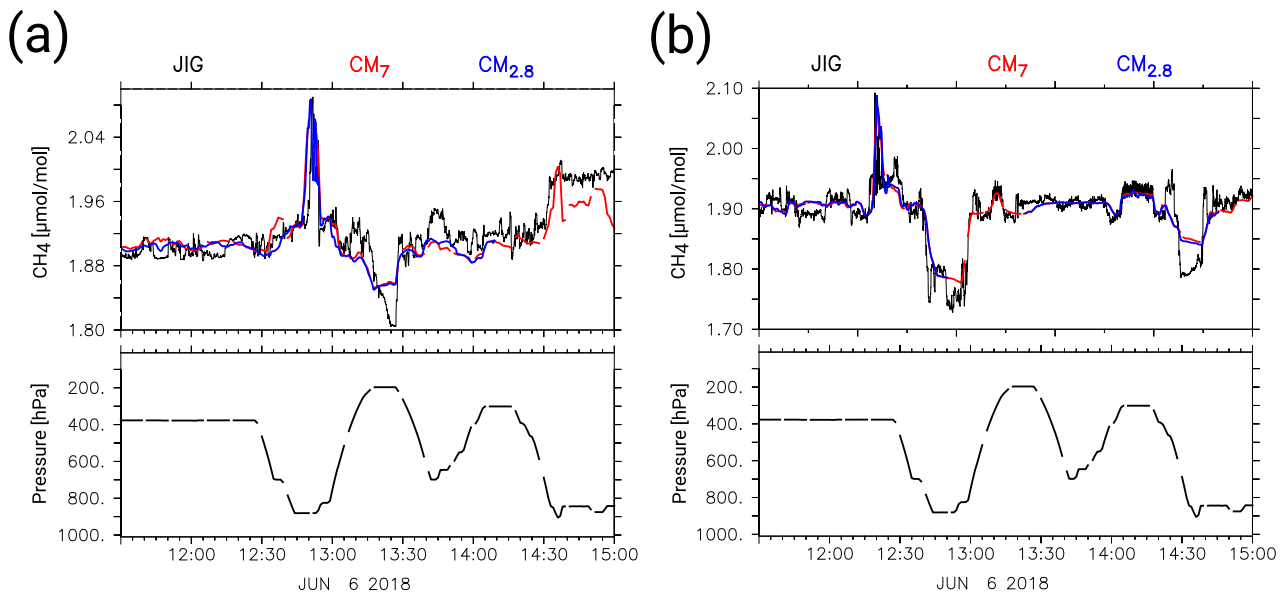

Figure 8. HALO in-situ sampled methane mixing ratios (black lines) of J1 (left) and J2 (right), as well as the S4D submodel output at flight level for CM7 (red dots) and CM2.8 (blue dots). All mixing ratios are in μmol/mol and the model results are bias corrected. Below the methane mixing ratios, the atmospheric pressure along the flight track is shown in hPa. The time axis displays the time of the specific flight in UTC.

the changes in the flight altitude. The pressure follows a steep up and down movement between 900 hPa and 200 hPa, which is because both flight paths were chosen to sample the vertical profile of methane in the atmosphere. The flight routes of J1 and J2 cover the USCB, but also parts which are lying outside the smallest model domain (see Fig. 9). Gaps in the simulated CM2.8 mixing ratios mark the temporal leaving of the smallest area.

Overall, observed and simulated methane mixing ratios correlate closely with atmospheric pressure. Consequently methane correlates negatively with flight altitude. Large scale patterns and amplitudes are very similar in both model instances as well as in the observations. Table 3 lists RMSE, NMBE and the correlation coefficient of J1 and J2 for the comparison with CM7 and CM2.8. NMBE have similar values ranging from -0.29 % to 0.08 %. Although in very good agreement, the model is not able to simulate the small scale fluctuations measured in the background methane at 400 hPa. Moreover, the model does not resolve the fine structure of the observations around 200 hPa. This can be seen in Fig. 8 (a) at 13:20 UTC and in Fig. 8 (b) at 12:00 UTC and 14:30 UTC. As the mixing ratio at these altitudes is strongly influenced by the boundary conditions of the global model, we would not expect the model to be able to reproduce these features. In contrast, methane variability at lower altitudes is well represented in CM7 and CM2.8. In general, CM7 and CM2.8 are in good agreement. RMSE for J1 is 0.02 μmol/mol for CM7 and CM2.8. For the comparison with J2, RMSE is similar with 0.02 μmol/mol for CM7 and 0.03 μmol/mol for CM2.8.

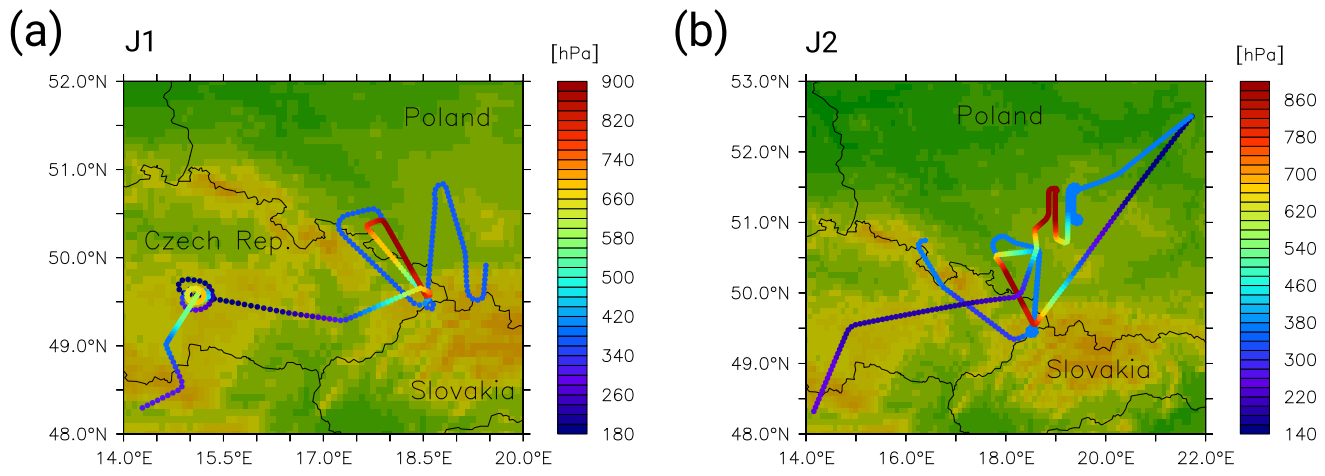

Figure 9. D-ADLR flight routes for J1 (a) and J2 (b). The color bar refers to the atmospheric pressure at flight level.

Table 3. Summary of the results of the statistical analysis of J1, J2, P4 and P5 compared to the the simulated CH4_FX mixing ratios (bias corrected). Listed are the root mean square error (RMSE) in μmol/mol, the normalized mean bias error (NMBE) in %, and the correlation coefficient (R) for the model domains CM7 and CM2.8.

Flight	RMSE (CM7)	RMSE (CM2.8)	NMBE (CM7)	NMBE (CM2.8)	R (CM7)	R (CM2.8)
J1	0.02	0.02	-0.29	-0.06	0.78	0.70
J2	0.02	0.03	0.08	-0.04	0.88	0.77
P4	0.06	0.07	1.30	1.88	0.77	0.84
P5	0.05	0.05	1.17	1.18	0.76	0.68

Here, we discuss the D-FDLR flights P4, P5 and P2. ~~The remaining~~ All other D-FDLR observations and their comparison to the model results are shown in the Supplement. Figure ~~8~~ 10 shows the comparison of the methane in-situ measurements derived with the Picarro CRDS on board of D-FDLR. The results shown are for the two flights P4 and P5. Both measurement flights aimed to sample the emissions of all methane sources within the USCB. The flight routes surround the USCB and follow a back-and-forth pattern along a horizontal track downwind of the mines, crossing the methane plume several times at different heights (see ~~Supplement for details on the flight pattern) . Figure 8~~ Fig. 11 (a) and (b)). Figure 10 (a) and (b) compare the simulated CH4_FX tracer mixing ratios along the flight tracks to the observations. Pattern similarity is good for both flights and background methane shows little variability~~, but results of CM7 and CM2.8 are equally biased towards lower mixing ratios~~. Table 3 lists the respective RMSE in μmol/mol~~and~~ , the NMBE in %, and the correlation coefficient for the comparison to both model instances ~~with~~ of P4 and P5. On June, 6 in the morning, the NMBE is ~~-4.4~~ 1.30 % for CM7 and ~~-3.6~~ 1.88 % for CM2.8. ~~Despite the negative bias, peak~~ Peak mixing ratios of CM7 and CM2.8 reach values close to or higher than those of the observations ~~,~~ and around 10:15 UTC CM2.8 mixing ratios ~~even~~ clearly exceed those of the observations. Although, generally

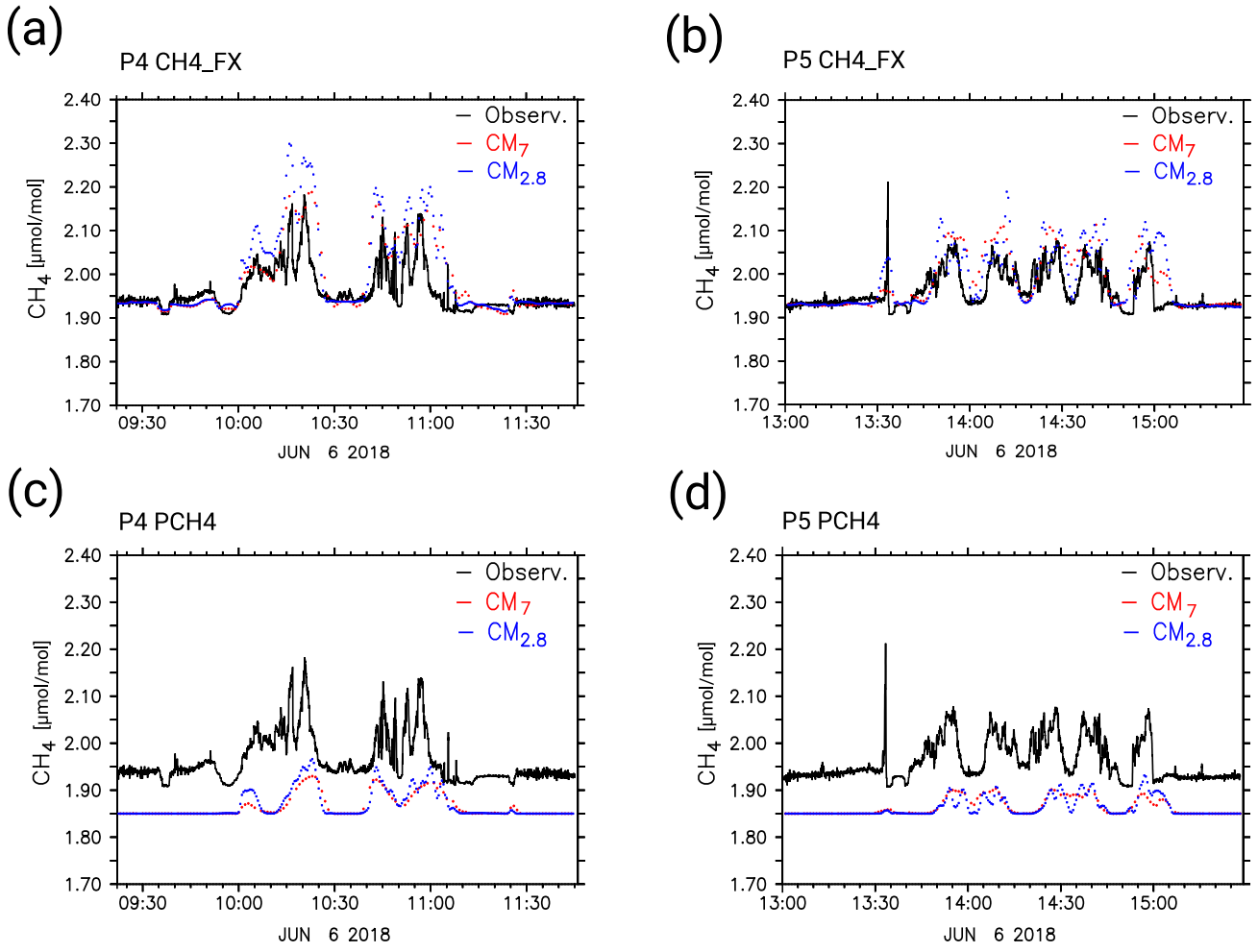

Figure 10. D-FDLR in-situ sampled CH_4 mixing ratios (black lines) of P4 (left) and P5 (right), as well as the S4D submodel output at flight altitude for CM7 (red dots) and CM2.8 (blue dots). Panels (a) and (b) show the comparison with the CH4_FX tracer, whereas (c) and (d) shows the comparison with the PCH4 tracer. All mixing ratios are in μmol/mol. CH4_FX is bias corrected with 0.108 μmol/mol and an offset of 1.85 μmol/mol is added to PCH4 for a better visualization. The time axis displays the time of the specific flight in UTC.

in good agreement, CM7 and CM2.8 differ from each other from 10:00 UTC to 10:30 UTC~~. Due to the finer resolution, ~~, where CM2.8 ~~better resolves the small fluctuations observed within the plume and~~ shows larger methane peaks than CM7. Regarding the afternoon flight (P5) model results again represent the observations well, in terms of time and location of the peaks. At 13:30 UTC the measured mixing ratios display a sharp increase, which is not distinctly presented in the model results. These high mixing ratios were taken very close to a specific coal mine, which is apparently not resolved well in CM2.8, and even less in CM7. In general, observed peaks in the afternoon flight are lower than ~~the observed peaks~~ those of the morning flight. This is also seen in the ~~simulation results. CM7 and CM2.8 differ from each other in such a way that CM2.8 better resolves the finer~~

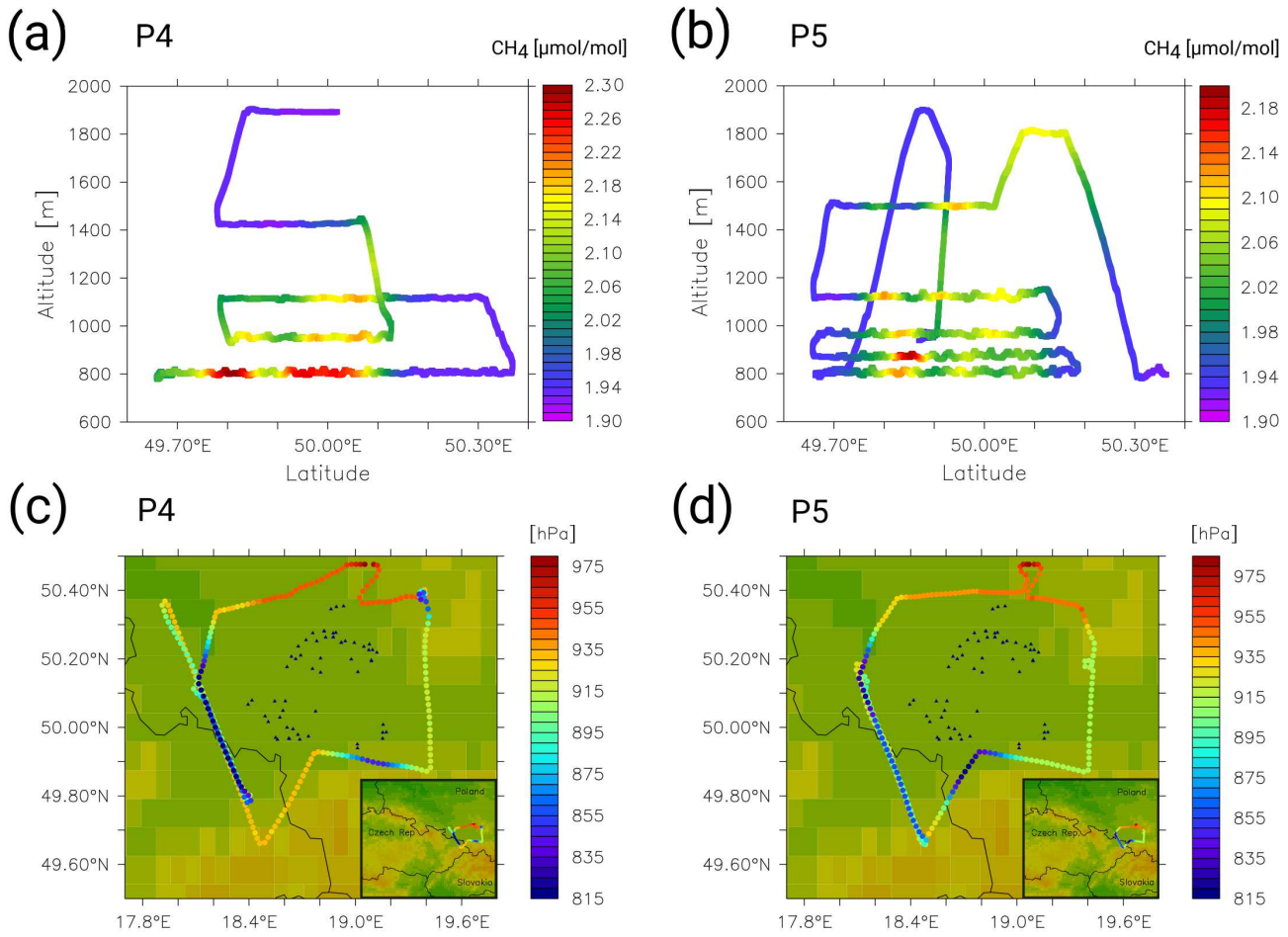

Figure 11. D-FDLR flight pattern in the USCB for P4 (left) and P5 (right). The upper panels show the vertical profile of the flight sections on June, 6th 2018 between 10:10 and 11:20 UTC (a) and between 13:25 and 15:10 UTC (b). The color bars refer to the bias corrected CH4_FX mixing ratios in μmol/mol simulated by CM2.8. The lower panels show the flight routes in the USCB for P4 (c) and P5 (d). Here, color bars refer to the atmospheric pressure at flight level in hPa. Blue triangles show the location of the coal mining ventilation shafts.

~~structure of the observed mixing ratios. Again, in CM2.8 methane peaks approach~~ model results. Again, simulated methane peaks exceed the observational peaks, but not as significantly, as seen for P4. ~~Figure 8~~ The NMBE is consequently lower for P5, with 1.17 % and 1.18 % for CM7 and CM2.8 respectively.

Figure 10 (c) and (d) show the comparison between the simulated PCH4 values along the flight track to P4 and P5. The black line ~~shows~~ illustrates the observed CH_4 mixing ratios in μmol/mol and the red and blue dots show the model results for CM7 and CM2.8, respectively. As PCH4 only considers the point source emissions without any background nor other methane source emissions, one can assume that the enhancements seen in the model and in the measurements originate from the ventilation shafts. Smaller variations within the background methane are consequently not present in the model results and

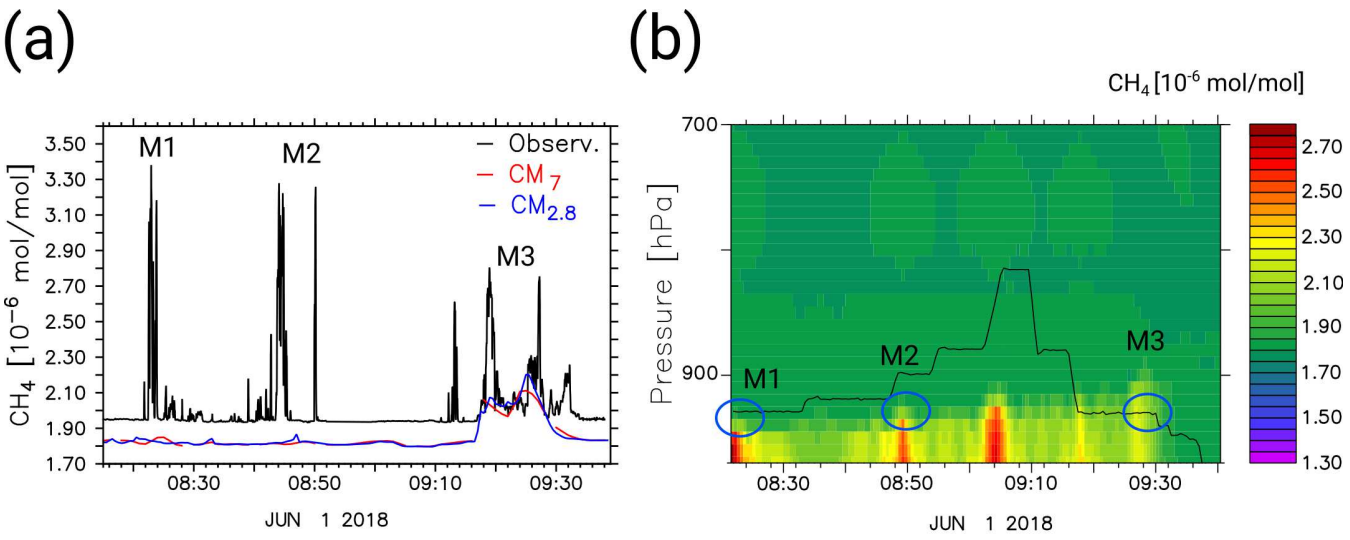

Figure 12. Panel (a) shows D-FDLR in-situ sampled CH_4 mixing ratios (black lines) of P2, as well as the S4D submodel output of CH4_FX at flight altitude for CM7 (red) and CM2.8 (blue). Panel (b) displays the corresponding flight altitude of P2 (black line) and the simulated profile (CM2.8) of the methane mixing ratio along this flight track. All mixing ratios are in μmol/mol. The time axes display the time of the flight in UTC. M1, M2 and M3 mark specific methane peaks seen in the observation (a) and in the simulated methane profile (b), but not necessarily in the sampled S4D output at flight altitude (a).

stay at a constant level of zero. To allow a better comparison with the observations we added a constant offset of 1.85 μmol/mol in both plots. The simulated PCH4 mixing ratios show a positive correlation with the major observed methane peaks. Although they do ~~not show~~ have the same amplitudes, all methane elevations are simulated by the model. On June, 6 in the morning, CM2.8 values exceed CM7 values and clearly show a more distinct structure. In the afternoon this difference is even more remarkable. CM2.8 is able to simulate the variability ~~very~~ more precisely, whereas CM7 does not resolve the smaller patterns seen in the observations (e.g. at 14:30 UTC). The result for PCH4 contrasts with the CH4_FX tracer, where peak emissions exceed the observations. We therefore compared the point source emissions of CoMet ED v1 to the anthropogenic emissions in the EDGAR v4.2FT2010 inventory. Whereas the point source emissions sum up to only 465 kt/a, the EDGAR v4.2FT2010 emissions, summed over all corresponding grid cells, are 1594 kt/a. 96.40 % of these emissions are attributed to the fugitive solid fuels of EDGAR sector 1B1.

Figure ~~9~~ 12 (a) shows the comparison of the D-FDLR in situ observations of P2 ~~at~~ on June, 1st 2018 to the CH4_FX mixing ratios simulated by CM7 and CM2.8. Again, a ~~constant offset~~ systematic bias between observations and model results exists. Until 09:00 UTC atmospheric conditions were mostly stable and D-FDLR flew a back-and-forth pattern at a distance of 20 km downwind of the southwestern cluster of USCB mines. The very high observed mixing ratios around 08:22, 08:45 and 08:50 UTC result from the only slightly diluted plumes. Those enhancements (M1 and M2) are barely detectable in the S4D output. The M3 enhancement was sampled further to the north, downwind of the northern USCB mines. Panel (b) shows the

corresponding simulated (CM2.8) methane profile along a flight track following curtain and the flight altitude in black. The model results in (b) show elevated methane mixing ratios at M1 and M2. The methane peaks are below the ~~simulated boundary layer~~ top of the simulated planetary boundary layer height (PBLH) and below the flight track of P2. Consequently they are not visible in the S4D results (a). These findings indicate, that the simulated ~~boundary layer~~ planetary boundary layer (PBL) during the morning is too low. ~~Contrary~~On the contrary, the observed methane peaks between 09:10 and 09:35 UTC (a) can be seen in the S4D results. Here, the ~~boundary layer~~ PBL already extended towards higher altitudes and the flight track is crossing the simulated methane plume (see panel (b)). Overall, CM7 and CM2.8 show smaller methane mixing ratios than observed.

~~Summary of the results of the statistical analysis of P4 and P5 compared to the the simulated CH4_FX mixing ratios. Listed are the root mean square error (RMSE) in μmol/mol and the normalized mean bias error (NMBE) in %. for the model domains CM7 and CM2.8. **Flight RMSE (CM7) RMSE (CM2.8) NMBE (CM7) NMBE (CM2.8)**P4 0.10 0.10 -4.2 -3.6P5 0.09 0.10 -4.3 -4.3~~

~~D-FDLR in-situ sampled CH_4 mixing ratios (black lines) of P4 (left) and P5 (right), as well as the S4D submodel output at flight altitude for CM7 (red dots) and CM2.8 (blue dots). Panels (a) and (b) show the comparison with the CH4_FX tracer, whereas (c) and (d) shows the comparison with the PCH4 tracer. All mixing ratios are in μmol/mol. An offset of 1.85 μmol/mol is added to PCH4 for a better visualization. The time axis displays the time of the specific flight in UTC.~~

~~Panel (a) shows D-FDLR in-situ sampled CH_4 mixing ratios (black lines) of P2, as well as the S4D submodel output of CH4_FX at flight altitude for CM7 (red) and CM2.8 (blue). Panel (b) displays the corresponding flight altitude of P2 (black line) and the simulated profile (CM2.8) of the methane mixing ratio along this flight track. All mixing ratios are in μmol/mol. The time axes display the time of the flight in UTC. M1, M2 and M3 mark specific methane peaks seen in the observation (a) and in the simulated methane profile (b), but not necessarily below in the sampled S4D output at flight altitude (a).~~

~~Figure 10 shows the HALO in-situ measurements of methane J1 and J2 in black, along with the simulated CH4_FX of CM7 and CM2.8 in red and blue, respectively. All mixing ratios are displayed in μmol/mol. In an extra panel below the methane mixing ratios, atmospheric pressure along the flight track is plotted in hPa and indicates the changes in the flight altitude. The pressure follows a steep up and down movement between 900 hPa and 200 hPa, which is because both flight paths were chosen to sample the vertical profile of methane in the atmosphere. The flight route of J1 and and J2 covered the USCB, but also parts which are lying outside the smallest model domain. Gaps in the simulated CM2.8 mixing ratios mark the temporal leaving of the smallest area. Overall, observed and simulated methane mixing ratios correlate well with atmospheric pressure. Consequently methane correlates negatively with flight altitude. Large scale patterns and amplitudes are very similar in both model instances as well as in the observations. Again, a constant bias towards smaller mixing ratios exists in the model results. It seems to be constant throughout the vertical levels. Table **??** lists RMSE and NMBE of J1 and J2 for the comparison with CM7 and CM2.8. NMBE have similar values ranging from -5.6 % to -5.9 %. Although in very good agreement, the model is not able to simulate the small scale fluctuations measured in the background methane at 400 hPa. Moreover, the model does not resolve the fine structure of the observations around 200 hPa. This can be seen in Fig. 10 (a) at 13:20 UTC and in Fig. 10 (b) at 12:00 UTC and 14:30 UTC. As the mixing ratio at this altitudes are strongly influenced by the boundary conditions of the global model, we would not expect that the model is able to reproduce these features. Contrary, methane variability at~~

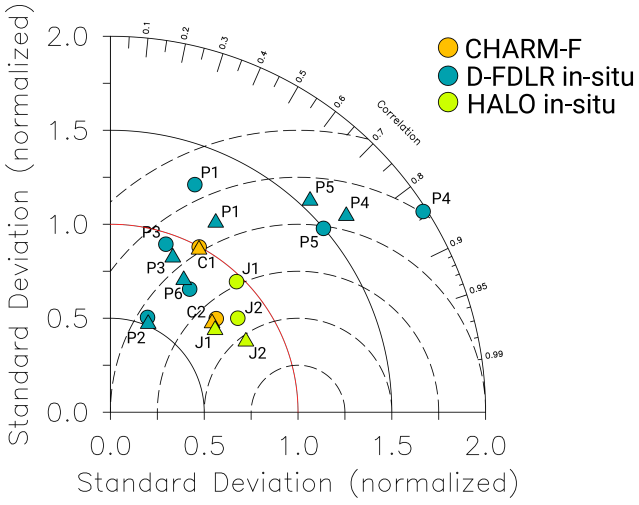

Figure 13. Taylor diagram summarizing normalized standard deviation (radius), correlation coefficient (angle) and centered NRMSE (dashed semi circles). Results show the comparison between observations and model domain CM7 (triangles) and CM2.8 (circles). Comparisons to C1 and C2 are displayed in orange, those to P1 until P6, in blue and those to J1 and J2 shown in green. P7 is outside the diagram.

~~lower altitudes is well represented in CM7 and CM2.8. In general, CM7 and CM2.8 are in good agreement. RMSE for J1 is 0.12 μmol/mol for CM7 and 0.11 μmol/mol for CM2.8. For the comparison with J2, RMSE is similar with 0.11 μmol/mol for CM7 and CM2.8.~~

~~Summary of the results of the statistical analysis of J1 and J2 compared to the the simulated CH4_FX mixing ratios. Listed are the root mean square error (RMSE) in μmol/mol and the normalized mean bias error (NMBE) in %. for the model domains CM7 and CM2.8. **Flight RMSE (CM7) RMSE (CM2.8) NMBE (CM7) NMBE (CM2.8)**J1 0.12 0.11 -5.9 -5.7J2 0.11 0.11 -5.7 -5.6~~

~~HALO in-situ sampled methane mixing ratios (black lines) of J1 (left) and J2 (right), as well as the S4D submodel output at flight level for CM7 (red dots) and CM2.8 (blue dots). All mixing ratios are in μmol/mol. Below the methane mixing ratios, the atmospheric pressure along the flight track is shown in hPa. The time axis displays the time of the specific flight in UTC.~~

3.2.3 Taylor Diagram

Taylor diagrams combine three statistical metrics to better compare and interpret different model performances. They summarize standard deviation (radial distance from the origin), correlation coefficient (angle) and centered ~~RSME~~ RMSE (dashed semi circles) in a single diagram (Taylor, 2001). Thanks to the normalization of standard deviation and centered RMSE (NRMSE), metrics become non-dimensional and different model results can be compared to each other. The point on the horizontal axis displaying a normalized standard deviation of 1 outlines the point where model results fit perfectly the observations. Figure ~~11~~ 13 shows the results of the statistical analysis of CM7 (circles) and CM2.8 (triangles) compared to the observations (Table 1).

~~Although simulations visually suit well , the pattern and amplitude of C1 (see Fig. 7 (a)), correlation coefficient and centered NRMSE are rather low. Amplitudes and pattern statistically differ~~ The three data sets differ in measuring technique, as well as in their geographical extent, duration and daytime of sampling. This makes it difficult to define an overall model skill. The meteorological situation, such as wind conditions or convection, and topographical features may lead to further uncertainties. The comparison to J1 and J2 is the best in the Taylor diagram. Results are close to the reference point on the horizontal axis (normalized standard deviation = 1.0) and correlation coefficients are very high, especially for CM7. The measurements cover larger areas outside the USCB including for example the Czech Republic. Other than for C and P observations, the horizontal distribution of the methane plumes only plays a minor role. The JIG samples focus on the vertical gradient of methane in the atmosphere, which is well represented in the model (Fig. S1 in the Supplement). The comparisons to C are reasonably well, too. The C1 pattern statistically differs from the observations, which may be due to a temporal or ~~spacial~~ spatial shift of the plume in the model at the beginning of June, 6. ~~Contrary, the~~ But normalized standard deviations are ~~almost the same. The standard deviation of C2 is larger than for C1, but here amplitude and pattern better fit to the observations. Overall~~ close to 1 and CM7 and CM2.8 ~~show similar performance. Model results corresponding to the smaller scale in-situ measurements P1, P4 and P5 have higher standard deviations than the observations. CM2.8 centered NRMSE are always larger than those of CM7. This is also~~ agree equally with the observations. Since CHARM-F measures the total column average mixing ratio, mismatches between actual and simulated PBLH, are less apparent. The comparisons to the smaller scale P observations assess the models ability to represent regional scale features like coal mine emission plumes. As described in Sect. 3.2.2~~where~~ , the skill also depends on how good the model actually simulates the PBLH. We can further see the highest variability between CM2.8 ~~, contrary to CM7, clearly exceeds the observed amplitudes. P3~~ and ~~P6 are close to the reference line and P2 shows very low amplitudes compared to the observations, which is the result of very high methane mixing ratios M1 and M2 (see Sect 3.2.2) seen in the observations but not in the modelresults. Correlation coefficients do not show a specific pattern. Except for C1, all comparisons show lower correlation coefficients than J and C comparisons.~~ CM7 for the comparisons with P. Depending on the grid size of the model, the very localized methane enhancements can be either more diluted or more intensified in the model results. And, mixing ratios sampled very close to the ventilation shafts are often not resolved by the model. Furthermore, high wind speeds, for example during the sampling of P7 ~~is not presented in the diagram as its~~ (see Supplement), lead to a low correlation with MECO(n) and a normalized standard deviation ~~is~~ larger than 2. ~~The comparison to J1~~ P7 is consequently not present in the Taylor diagram. Finally, C and J data were sampled on June 6 and ~~J2 is the best in this diagram. Results are closest to the reference point on the horizontal axis (normalized standard deviation = 1.0). Contrary to the comparisons with the P data sets, the CM2.8 standard deviation is closer to the observed standard deviation as the one of CM7. Correlation coefficients are very high, especially for CM7~~7, where wind conditions were stable. Additional data would be necessary in order to compare the models' skill on days with less perfect conditions, such as for P.

4 Evaluation of Forecast Skill

A good forecast should be able to simulate both, amplitude and pattern variability, of the observed methane mixing ratios in the atmosphere. To identify the temporal evolution of the forecast skill with each forecast day, we therefore ~~calculated~~ calculate skill scores (after Taylor, 2001) that consider standard deviation and correlation coefficient. We used the two different skill scores

$$S_{\mathrm{V}} = \frac{4(1+R)}{(\sigma_{\mathrm{f}} + 1/\sigma_{\mathrm{f}})^2 (1+R_0)} \tag{3}$$

and

$$S_{\mathrm{C}} = \frac{4(1+R)^4}{(\sigma_{\mathrm{f}} + 1/\sigma_{\mathrm{f}})^2 (1+R_0)^4} \tag{4}$$

which either emphasize the similarity of the amplitudes, or the similarity of the patterns. R is the correlation coefficient between forecast and observation, R_0 is the maximum attainable correlation coefficient, and σ_{f} is the ratio of the standard deviation of the forecast to that of the observation. We assume R_0 to be 1, although in reality maximum correlation coefficients between observations and simulation can not be reached due to differences in spatial and temporal resolution. The skill ranges between 0 and 1, with small values indicating low skill and high values indicating high skill. We use the analysis simulation as a reference observation to evaluate a theoretical forecast skill. As the forecasts are branched from the analysis simulation we aim to quantify the deviation of the forecast from the analysis with increasing forecast day. The results are discussed in Sect. 4.1. In order to find the ~~expected~~ actual skill of the forecast, we further compare the different forecast days to the observations C1, C2, J1, J2, P4 and P5. Section 4.2 describes these results.

4.1 Theoretical Forecast Skill

We ~~compared~~ compare every single forecast day out of six forecast days to the analysis simulation ~~and calculated~~ (between June, 1^{st} and 22^{th}, 2018) and calculate a daily skill score at each point on the respective two dimensional model grid. The skill ~~was~~ is calculated for the simulated CH4_FX values. In order to compare CM7 and CM2.8, the analyzed area only covers the area obtained by removing the outermost 15 grid points of the CM2.8 domain (relaxation area). Figure ~~12~~ 14 assigns to each forecast day the average percentage of the area which reveals a skill score larger than 0.7. The results are shown in red for CM7 and in blue for CM2.8. Panels (a) and (b) refer to the different skill scores S_{V} and S_{C}, respectively.

On forecast day I to III, CM7 shows slightly larger values than CM2.8. This is most terse for S_{C}, which puts greater emphasis on the correlation coefficient. However, differences between the two model instances are rather small. The forecast skill is very large at forecast day I. Here, the forecasts are branched from the analysis simulation, which results in a good agreement between reference and forecast. Both skills decrease with increasing forecast day, whereby the skill in (b) shows a steeper decrease than the skill in (a). This suggests that the correlation between forecast and analysis reduces faster than the similarity of amplitudes. For S_{V}, the area which exceeds a threshold of 0.7, covers about 65 % and 40 % at forecast day II and III, whereas for S_{C} it only covers about 50 % and 25 %, respectively. From forecast day IV onwards, less than 20 % (S_{V}) or 10 % (S_{C}) of the

area reveal a skill larger than 0.7. The graphic shows the evolution of the theoretical forecast skill with increasing forecast day for CM7 (red) and CM2.8 (blue). The vertical axis displays the average area (in %), according to the smallest model domain, with a Skill Score > 0.7. The horizontal axis shows the specific forecast days 1 to 6. The Skill Scores are calculated for each day at each grid point from 1^{st} to 22^{th} June 2018 (CoMet 1.0). It compares the CH4_FX total column mixing ratios of the forecast simulations to the CH4_FX total column mixing ratios of the analysis simulation. The error bars indicate the interval which contains 95% of all skill scores per day. S_V (a) emphasizes variability and S_C (b) emphasizes the correlation.

The Taylor diagram shows the comparison of the CH4_FX total column mixing ratios of the forecast simulations with the CH4_FX total column mixing ratios of the analysis simulation sampled with the SCOUT submodel at a selected location. Different symbols display the specific forecast days: Day 1 - green dots, day 2 - purple triangles, day 3 - red diamonds, day 4 - light blue stars, day 5 - blue crosses, day 6 - yellow circles. Correlation coefficient (angle), normalized standard deviation (radius) and normalized centered root mean square error (dashed semi circles) are calculated for each day from 10^{th} to 25^{th} August 2017.

Figure 13 shows the decline of the theoretical forecast skill within a Taylor plot by comparing the simulated XCH_4_FX values at a fixed location (Hotel Pustelnik in Poland, lat: 49.933024, lon: 18.799681, SCOUT output) to the analysis. 15 days of the CoMet 0.5 campaign are analyzed. Forecast day I (green circles) agrees best with the analysis. Most results gather around a point where the normalized standard deviation is almost 1, and the correlation between analysis and forecast is at its maximum. Few exceptions show lower correlation coefficients down to 0.8 but stay close to the red reference line. The correlation coefficient on day II (purple triangles) ranges from about 0.33 to about 0.97 and the normalized standard deviation lies in the range between 0.5 to 1.5. On forecast days III (red diamonds) and IV (light blue stars), most normalized standard deviations cover the same range but the correlation coefficients decrease and do not show values larger than 0.89. Similar to what we observe in Fig. 12 the correlation coefficient seems to decrease faster or at least shows more variability, whereas the standard deviation lies within a constant range. The The lower correlation could be attributed to a displacement of the simulated plume in time or space, which would also explain the fact that the normalized standard deviation remains within the given range.Results for forecast days V (blue crosses) and VI (yellow circles) show correlation coefficients below 0.7 and a large variability in normalized standard deviation. Some points are outside the diagram and consequently not shown here.

4.2 Expected Actual Forecast Skill

Figure 14 15 shows the skill score S_V calculated for the different forecast days I to VI when compared to the observations C1, C2, J1 and J2 (see panel (a)) and to the observations P1 - P7 (see panel (b)). Results for CM7 and CM2.8 are shown on the left and right side, respectively. Contrary In contrast to the theoretical skill, where S_V and S_C clearly decrease with increasing forecast day, a reduction of the skill is not obvious here. Whereas the theoretical skill score is defined to measure the skill, averaged over the entire model domain, the actual skill score compares the model results to observational data. The latter measure the downwind methane plumes, which are easier to forecast than the variability of the methane background in the overall model domain. Considering the difference between the single observations, S_V is highest for J1 and J2 with values

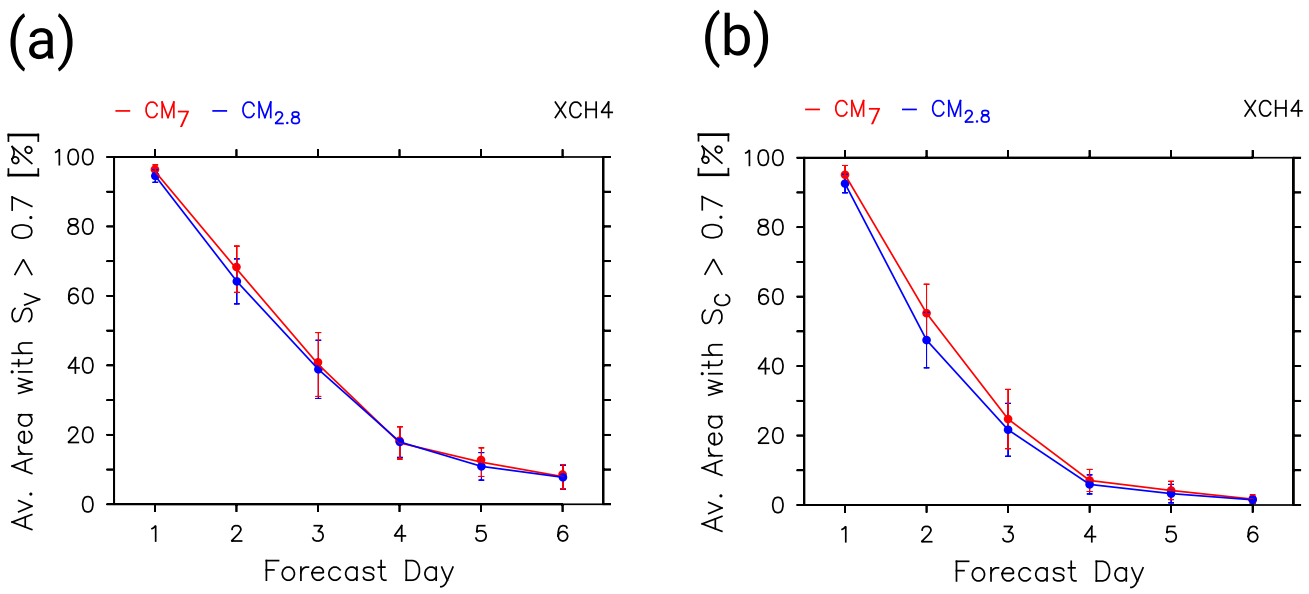

Figure 14. The graphic shows the evolution of the theoretical forecast skill with increasing forecast day for CM7 (red) and CM2.8 (blue). The vertical axis displays the average area (in %), according to the smallest model domain, with a Skill Score > 0.7. The horizontal axis shows the specific forecast days 1 to 6. The Skill Scores are calculated for each day at each grid point from 1^{st} to 22^{th} June 2018 (CoMet 1.0). It compares the CH4_FX total column mixing ratios of the forecast simulations to the CH4_FX total column mixing ratios of the analysis simulation. The error bars indicate the interval which contains 95% of all skill scores per day. S_V (a) emphasizes variability and S_C (b) emphasizes the correlation.

above 0.8. They are followed by C1 and C2 with values above 0.65 (except for C1 on forecast day I) and P1, P4, P5 and P6 mainly showing a skill between 0.6 and 0.8. The skill is lowest for the comparison with P2 and P7. S_V emphasizes the similarity of amplitude ~~height~~ between forecast and observation. This similarity seems to be highest with HALO in-situ and CHARM-F observations. ~~However,~~ As already discussed in Sect. 3.2.3, the models' skill differs between the three data sets, which is mainly due to different flight patterns and measurement techniques. J and C observations measure larger scale vertical and integrated horizontal distributions of methane, respectively. Both are well represented in MECO(n). Modelling the smaller scale features observed by P is however challenging, which is also reflected in the comparison with the forecasts. S_V does not vary significantly among the different forecast days nor does it show any specific trend. Results for C1, P4, P5 and P7 drop at forecast day V, but increase again at forecast day VI. ~~Contrary~~In contrast, P2 suddenly increases at forecast day V. Differences between CM7 and CM2.8 are rather small.

~~Figure 15~~ Figure 16 summarizes the results of the skill score S_C. S_C is generally lower than S_V, which is due to higher weighting of the correlation coefficient. Overall, skill is best for J1, J2, C2, P4 and P5, meaning that model and observations correlate well here. As mentioned in Sect. 3.2.3, conditions for measuring the downwind methane plumes on these days were favorable and model results agree better with the observations. On the contrary, P2 and P7 show again very low values (c. also

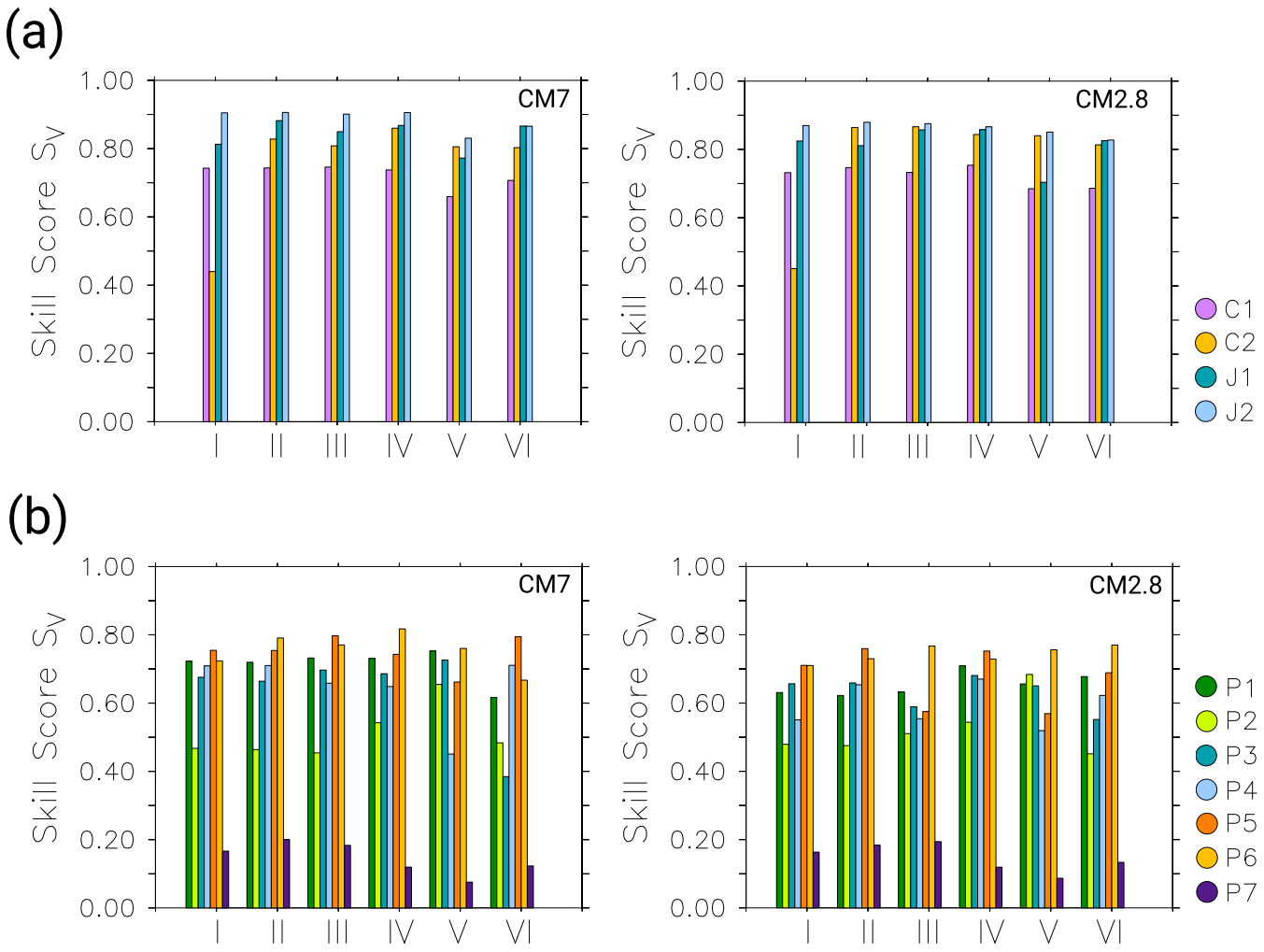

Figure 15. The bar plots show the calculated skill scores S_V for the comparison of forecast day I to VI (horizontal axis) to the observations. The colors refer to the different observations. Panel (a) displays the results for C1, C2, J1, and J2. Panel (b) shows the results for P1-P7. CM7 and CM2.8 results are shown on the left and right side, respectively.

Fig. ~~14~~15). In panel (a), CM7 and CM2.8 show a similar pattern. The skill among the different forecast days almost stays at the same level or even increases until forecast day IV. Forecast day V and VI show lower skill, with lowest values for C1 at forecast day V. The skill for J1 and J2 shows generally lower values in CM2.8 than in CM7. In panel (b) the skill is highly variable among all forecast days until day IV. On forecast day V and VI, skill decreases for all comparisons, with very low values for P7.

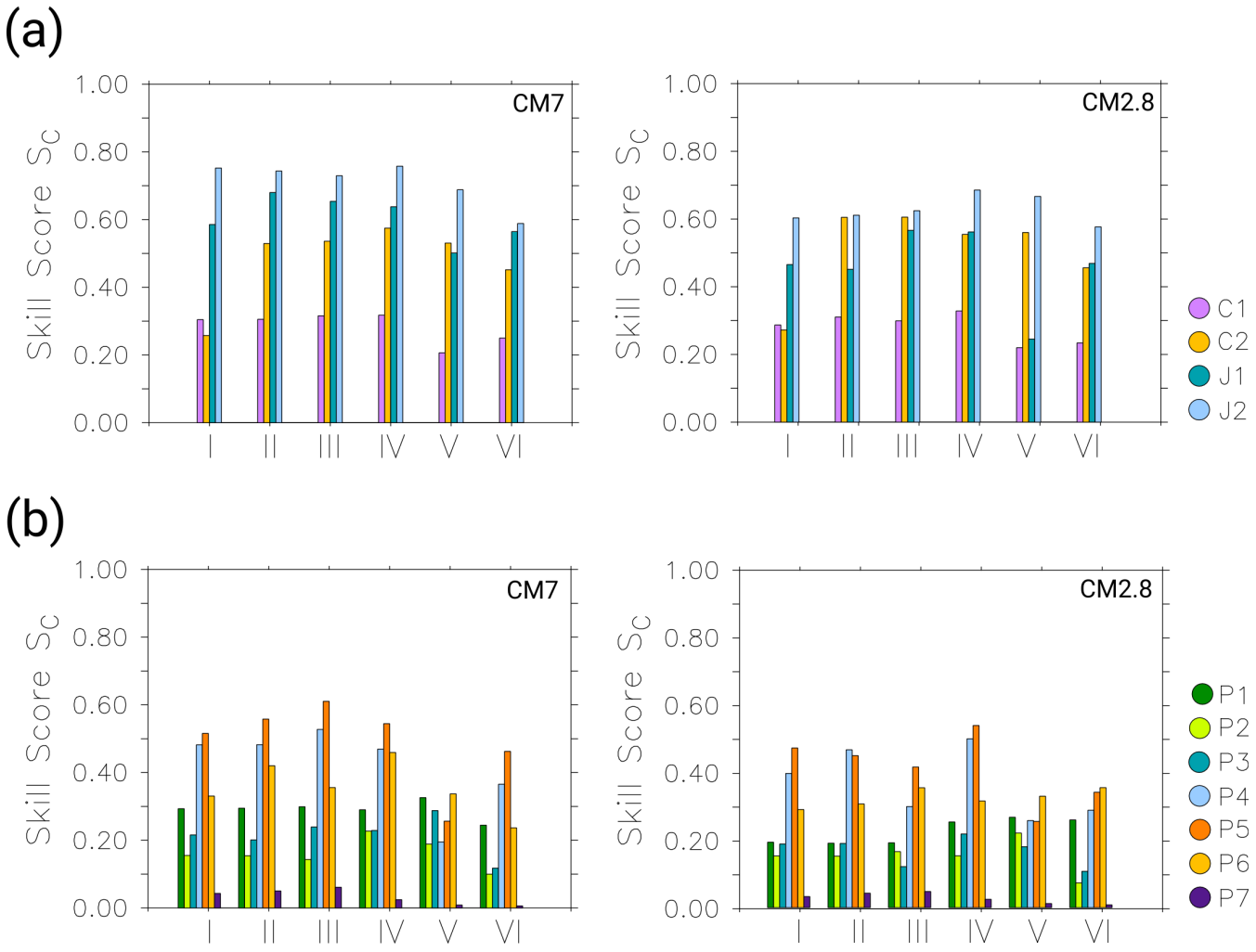

Figure 16. The bar plots show the calculated skill scores S_C for the comparison of forecast day I to VI (horizontal axis) to the observations. The colors refer to the different observations. Panel (a) displays the results for C1, C2, J1, and J2. Panel (b) shows the results for P1-P7. CM7 and CM2.8 results are shown on the left and right side, respectively.

~~The Taylor plots summarize the comparison of the forecast simulations with the observations: CHARM-F (triangles), HALO in-situ (stars) and D-FDLR in-situ (circles). Different colors display the specific forecast days: Day 2 - blue, day 3 - orange, day 4 - yellow, day 5 - green, day 6 - light blue. Dots (C1, J1 and P4) or no dots inside shapes help to distinguish between the different flights.Correlation coefficient (angle), normalized standard deviation (radius) and normalized centered root mean square error (dashed semi circles) are calculated for CM7 (a) and CM2.8 (b).~~

Figure 16 shows the comparison of the individual forecast days to the observations: C1, C2 (triangles), J1, J2 (stars) or P4 and P5 (circles) in a Taylor diagram. Here we only show the results of the six best matches between MECO(n) forecasts and observations. The colors refer to the different forecast days II to VI and the inner dots mark the results that belong to the same sampling date (e.g. J1: star with dot, J2: star without dot). Forecast day I is not considered here, as it was not available for the campaign (as described in Sect. 2.3). Panel (a) shows the results of the model domain CM7 and panel (b) shows the results of the smaller model domain CM2.8. In both cases, the comparisons with CHARM-F observations have low variability in standard deviation. All forecast days show a normalized standard deviation close to 1 (red reference line), meaning that all forecast days show similar amplitudes and match well the observations. While forecast day V (green triangles) shows the largest deviation from the reference, all other forecast days are closely grouped together and do not show a specific trend with increasing forecast day. The comparisons with the HALO in-situ observations (stars) reveal similar results. The normalized standard deviations are close to 1 and do not show a clear pattern. The normalized standard deviations of the D-FDLR in-situ comparisons (circles) are again gathered, regardless of a particular forecast day, but show more internal variability (between the different forecast days) and higher deviations from the reference line. CM7 (a) results even show that the last forecast days stay closer to the reference line, which means that their amplitude height resembles the amplitudes of the observations better. Yet, this is however just a result for two measurements sampled at the same day. The correlation coefficient varies between the different observations (data point with or without dots). And, in contrast to the normalized standard deviation, the correlation coefficient shows a decreasing trend with increasing forecast day. Forecast day V and VI mostly show lower correlation coefficients than previous forecast days II, III and IV. Apart from a few exceptions the correlation of the forecast days III and IV are slightly higher than forecast day II. The decrease of correlation with increasing forecast day is most obvious when comparing with D-FDLR in-situ observation (i.e. for the observation targeting on small scale features). Deviations in correlation coefficient seem to be slightly higher in CM7, compared to CM2.8

5 Discussion

Overall, the comparison of the analysis simulation with airborne derived measurements show that MECO(n) is able to simulate the observed methane plumes reasonably well. This is the intended result given the fact, considering that EMAC is nudged towards the ECMWF data at a coarser resolution (T42 spectral truncation), and CM50, CM7 and CM2.8 are nested into each other and only driven by relaxation at their boundaries by the next coarser model instance. Nevertheless, a continuous and constant offset of the simulated CH4_FX to all observations results from all model instances. As the bias is constant at all altitudes, it is most likely not caused by shortcomings in the vertical transport in the model. Instead, global increase of methane emissions (Nisbet et al., 2019) could explain the discrepancy between the observations and the model results that are based on EDGAR v4.2FT2010 for anthropogenic emissions and on the EMPA inventory (Frank, 2018) for all other emissions fluxes. Apart from that, the timing of the simulated peaks is in good agreement for all observations. Compared to CHARM-F observations, the simulated $X_{fl}CH_4$ peaks show similar or slightly lower amplitudes. The vertical methane gradient measured by the JIG instrument is well represented by MECO(n). Besides the smaller variations in the background methane, the model

results correlate well with the measured methane mixing ratios at different altitudes. By comparing the small scale D-FDLR in-situ measurements, the simulated amplitude of the peaks is mostly overestimated. This applies particularly for the CM2.8 results. Anthropogenic emissions in the EDGAR v4.2FT2010 inventory differ from the latest release EDGAR v4.3.2 (Retrieved from http://edgar.jrc.ec.europa.eu, Feb 04, 2019). Total anthropogenic methane emissions for the USCB (here: lon. 18.30°E - 19.40°E, lat. 49.90°N - 50.40°N) are 1636 kt/a in EDGAR v4.2FT2010 and only 605.6 kt/a in EDGAR v4.3.2. The overestimation of local methane plumes in the model can therefore be explained by the overestimation of methane emissions in EDGAR v.4.2FT2010. However, we also see high peaks in the methane mixing ratios of the D-FDLR in-situ observations, which are very low in the model results or not present at all. This is mainly the case when D-FDLR sampled very close to the ventilation shafts (see results for P3 in the Supplement). Here, the model can not resolve the very localized enhancements. Larger grid sizes lead to instantaneous dilution of the simulated mixing ratios close to the ventilation shafts. Or, as seen for June 2nd 2018, the simulated ~~boundary layer is too low~~ PBL is too shallow and lies below the flight altitude. Consequently, the enhanced methane mixing ratios are not present in the S4D results along the flight track at flight altitude. Previous studies (i.e. Collaud Coen et al., 2014; Mertens et al., 2016) analyzing the simulated PBLH of the COSMO model, show results opposing our findings. However, this is just a snapshot of a short term simulation and our study does not provide a detailed analysis of the PBL. Additionally, PCH4 which only considers the emissions of the ventilation shafts, is in good agreement with the CH4_FX tracer and the observed methane elevations. This indicates that the observed methane plumes actually originate from coal mining. Although the assessment of the single point source emissions is not part of the current study, it should be noted that the different PCH4 point sources can be switched on and off in the model individually. This provides a good tool to distinguish between different sources and to assign them to the different measurements. When compared to the small scale D-FDLR in-situ measurements, PCH4 correlates well with the observed methane ~~emissions. Contrary~~ mixing ratios. In contrast to the CH4_FX results, the point source tracer does not overestimate the emissions. However, it neither shows the same emission strength as the observations. PCH4 peaks have considerably lower amplitudes than observed. The reason for this discrepancy is the different emission inventories for CH4_FX and PCH4. The sum of all methane emissions in CoMet ED v1 used for PCH4 is 465 kt/a, which is less than a third of the EDGAR v4.2FT2010 emissions (summed over the corresponding grid boxes). Updated estimates of emissions from CoMet ED v2 (based on E-PRTR 2016 (Retrieved from http://prtr.eea.europa.eu, Nov 07, 2018)) indicated larger emissions of 502 kt/a and some changes in the distribution of emissions, following structural and operational changes in the mining sector over the period between reporting years (2014 and 2016 for CoMet DB v1 and v2, respectively). This implies that the simulated PCH4 using the latest emission inventory CoMet ED v2 are expected to match the observed amplitudes better. Another reason for the underestimation of the simulated PCH4 peaks might be the fact that we assume a temporally constant methane release from the ventilation shafts. But in reality the emitted amount of methane varies from day to day. This might ~~probably~~ have a small influence on the results, but would not explain the large differences between PCH4 and the observations. Overall, CM7 is able to simulated the large scale observations (HALO in-situ) and the vertically integrated methane (CHARM-F) as precisely as CM2.8. When compared to small scale measurements (D-FDLR in-situ) the model overestimates the observed peaks. This is especially true for the finer resolved CM2.8, where methane mixing ratios are larger than the mixing ratios simulated by CM7. Smaller grid cells may

catch locally enhanced methane mixing ratios in the plume, whereas coarser grid cells cover a larger portion of the methane plume and mixing ratios may be more diluted. Additionally, CM2.8 is able to better simulate the fine structure of the small scale observations. However, the differences are rather small and the observed methane peaks are well represented in both model instances.

The theoretical forecast skill illustrates the deviation of the forecast from the analysis simulation. Results show a decreasing trend with increasing forecast day. Nevertheless, correlation and amplitude similarity of a single forecast ~~days~~ day show a broad variation. The evaluation of the ~~expected~~ actual forecast skill reveals even less clear results. The amplitudes seem to be constant or at least do not show any specific trend with increasing forecast day, but the correlation between observation and forecast slightly decreases. Forecast day V seems to yield the lowest skill for P4 and P5 and also for C1 and J1, which is less obvious as for P4 and P5. Due to the fact that these observations were sampled during only one day, namely the 6th of June 2018 in the morning and in the afternoon, all comparisons for the fifth forecast day are related to the forecast simulation start date 2nd of June. Disagreement may be due specific meteorological situation of this day. In order to make a general statement about the forecast skill, it would be necessary to compare additional observations within a broader time span.

6 Conclusions

We successfully conducted 6-day-forecast simulations of methane with the on-line coupled three times nested global and regional chemistry climate model system MECO(3). The forecasts branch from a continuous analysis simulation, where EMAC is nudged towards the operational ECMWF analysis data. This is essential for appropriate initial forecast conditions. We continuously delivered the forecasts during ~~CoMet 0.5 and~~ 1.0 and analyzed the model and forecast performance with respect to the observations. The advantage of using the global/regional model system is, that we are able to simulate both, the point source emissions and the background methane. For the latter, it is essential to provide lateral boundary conditions to the nested model instances, which are consistent with the meteorology, i.e. the dynamical boundary conditions. This makes it possible to distinguish between local source emissions and fluctuations in the background methane, which is important for the quantification of different methane sources. Even though the data for Newtonian relaxation are first coarsened to a horizontal grid resolution corresponding to the T42 spectral truncation, and then nested three times down to a spatial resolution of 2.8 km, MECO(3) is able to simulate the observed methane plumes correctly. Overall, the vertically integrated values, e.g. total column average mixing ratios, and the large scale patterns, such as the vertical gradient of methane, are well represented. However, limitations exist for the simulation of small scale patterns. A bias reduction as well as a better agreement of small scale simulated methane amplitudes with the observations may be achieved by updating the applied emission inventories to the EDGAR ~~v.4~~v4.3.2 inventory for anthropogenic emissions and the latest information on point source emissions (CoMet ED v2). Another anthropogenic emission inventory, which could reduce the bias, is the CAMS-GLOB-ANT inventory (Granier et al., 2019). CAMS-GLOB-ANT extrapolates the emissions to the current year by using EDGAR v4.3.2 as a basis for 2010, and by projecting emission trends for 2011 to 2014 from the CEDS inventory (Hoesly et al., 2018) until 2018. Furthermore, we obtained decent results up to forecast day IV. The skill score calculated for all forecast ~~day~~ days is reasonable.

However, due to the limited number of comparable observations, the skill score might ~~be not~~ not be representative and its interpretation must be treated with caution. All observed methane peaks are well represented in both model domains CM2.8 and CM7. For the purpose of the field campaign, it is therefore sufficient to perform the forecasts with CM7 only.

Code availability. The Modular Earth Submodel System (MESSy) is continuously further developed and applied by a consortium of institutions. The usage of MESSy and access to the source code is licenced to all affiliates of institutions which are members of the MESSy Consortium. Institutions can become a member of the MESSy Consortium by signing the MESSy Memorandum of Understanding. More information can be found on the MESSy Consortium Website (http://www.messy-interface.org). We used MESSy v2.53

.

Data availability. The simulation results are available on request from the first author. The operational analysis and forecast data, used for nudging are licenced by the ECMWF. For further details, see https://www.ecmwf.int.

Author contributions. Astrid Kerkweg, Patrick Jöckel and Mariano Mertens developed the model system MECO(n) and Patrick Jöckel the forecast chain. Mariano Mertens set up the different model domains. Anna-Leah Nickl, Patrick Jöckel and Mariano Mertens planned and carried out the forecast simulations. Theresa Klausner (with support of Anke Roiger and Andreas Fix) provided the emission inventory CoMet ED v1. Alina Fiehn and Michal Galkowski provided the emission inventory CoMet ED v2. Axel Amediek and Andreas Fix performed the CHARM-F measurements and Axel Amediek provided the processed data. Alina Fiehn, Maximilian Eckl and Anke Roiger carried out the D-FDLR in-situ measurements and Alina Fiehn provided the processed data. Michal Galkowski and Christoph Gerbig performed the HALO in-situ measurements and provided the data. Anna-Leah Nickl performed the simulations and analyzed the model results with input from Patrick Jöckel and Mariano Mertens. Anna-Leah Nickl drafted the manuscript to which all authors contributed.

Competing interests. The author declare that they have no conflict of interest.

Acknowledgements. We gratefully acknowledge the funding of CoMet by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - ~~Project Number 316646266 -~~ SPP 1294, within the Priority Program "Atmospheric and Earth System Research with the "High Altitude and Long Range Research Aircraft (HALO)". We acknowledge the funding of the scientific data analysis by the Federal Ministry for Research and Education (BMBF) within its AIRSPACE project through grants no. FKZ 01LK1701A and FKZ 01LK1701C. We acknowledge DLR VO-R for funding the young investigator research group "Greenhouse Gases". We thank Heidi Huntrieser (DLR-Institute of Atmospheric Physics, Oberpfaffenhofen, Germany) for the helpful comments on a previous version of the manuscript. We further gratefully acknowledge Jarek Necki and Justyna Swolkien from AGH University of Science and Technology, Krakow, Poland for providing the shaft

locations for the internal databases of CoMet ED v1 and CoMet ED v2. We thank both anonymous reviewers for all the thoughtful comments and remarks, which helped to improve the publication. We thank Winfried Beer (DLR-Institute of Atmospheric Physics, Oberpfaffenhofen, Germany) for providing the java script of the forecast web page and the administration of the forecast web service. We acknowledge the computational resources provided by German Climate Computing Centre (DKRZ) in Hamburg.

References

Amediek, A., Ehret, G., Fix, A., Wirth, M., Büdenbender, C., Quatrevalet, M., Kiemle, C., and Gerbig, C.: CHARM-F – a new airborne integrated-path differential-absorption lidar for carbon dioxide and methane observations: measurement performance and quantification of strong point source emissions, Appl. Opt., 56, 5182–5197, https://doi.org/10.1364/AO.56.005182, http://ao.osa.org/abstract.cfm?URI=ao-56-18-5182, 2017.

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmospheric Chemistry and Physics, 14, 13 205–13 221, https://doi.org/10.5194/acp-14-13205-2014, https://www.atmos-chem-phys.net/14/13205/2014/, 2014.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers., Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 369 1943, 2058–72, 2011.

E-PRTR 2014, .: European Pollutant Release and Transfer Register, http://prtr.eea.europa.eu, Retrieved from http://prtr.eea.europa.eu, Feb 08, 2017.

E-PRTR 2016, .: European Pollutant Release and Transfer Register, http://prtr.eea.europa.eu, Retrieved from http://prtr.eea.europa.eu, Nov 07, 2018.

EDGAR v4.2FT2010, European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR), r.: 2013, Retrieved from http://edgar.jrc.ec.europa.eu, May 30, 2017.

EDGAR v4.3.2, European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR), r.: 2017, Retrieved from http://edgar.jrc.ec.europa.eu, Feb 04, 2019.

Filges, A., Gerbig, C., Chen, H., Franke, H., Klaus, C., and Jordan, A.: The IAGOS-core greenhouse gas package: a measurement system for continuous airborne observations of CO2, CH4, H2O and CO, Tellus Series B Chemical and Physical Meteorology B, 67, 27 989, https://doi.org/10.3402/tellusb.v67.27989, 2015.

Fletcher, S. E. M. and Schaefer, H.: Rising methane: A new climate challenge, Science, 364, 932–933, https://doi.org/10.1126/science.aax1828, https://science.sciencemag.org/content/364/6444/932, 2019.

Frank, F. I.: Atmospheric methane and its isotopic composition in a changing climate, http://nbn-resolving.de/urn:nbn:de:bvb:19-225789, 2018.

Granier, C., Darras, S., Denier van der Gon, H., Doubalova, J., Elguindi, N., Galle, B., Gauss, M., Guevara, M., Jalkanen, J.-P., Kuenen, J., Liousse, C., Quack, B., Simpson, D., and Sindelarova, K.: The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version) Report April 2019 version, null, https://doi.org/10.24380/d0bn-kx16, 2019.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, https://www.geosci-model-dev.net/11/369/2018/, 2018.

Howarth, R. W.: Ideas and perspectives: is shale gas a major driver of recent increase in global atmospheric methane?, Biogeosciences, 16, 3033–3046, https://doi.org/10.5194/bg-16-3033-2019, https://www.biogeosciences.net/16/3033/2019/, 2019.

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., and Lelieveld, J.: Technical Note: The Modular Earth Submodel System (MESSy) - a new approach towards Earth System Modeling, Atmospheric Chemistry and Physics, 5, 433–444, https://doi.org/10.5194/acp-5-433-2005, https://www.atmos-chem-phys.net/5/433/2005/, 2005.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geoscientific Model Development, 3, 717–752, https://doi.org/10.5194/gmd-3-717-2010, https://www.geosci-model-dev.net/3/717/2010/, 2010.

Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, Geoscientific Model Development, 9, 1153–1200, https://doi.org/10.5194/gmd-9-1153-2016, https://www.geosci-model-dev.net/9/1153/2016/, 2016.

Jöckel, P., Nickl, A. L., and Mertens, M.: Example of MECO(n) forecast web product, https://doi.org/10.5281/zenodo.3518926, https://doi.org/10.5281/zenodo.3518926, 2019.

Kerkweg, A. and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) – Part 1: Description of the limited-area atmospheric chemistry model COSMO/MESSy, Geoscientific Model Development, 5, 87–110, https://doi.org/10.5194/gmd-5-87-2012, https://www.geosci-model-dev.net/5/87/2012/, 2012a.

Kerkweg, A. and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) – Part 2: On-line coupling with the Multi-Model-Driver (MMD), Geoscientific Model Development, 5, 111–128, https://doi.org/10.5194/gmd-5-111-2012, https://www.geosci-model-dev.net/5/111/2012/, 2012b.

Kerkweg, A. and Jöckel, P.: The infrastructure MESSy submodels GRID (v1.0) and IMPORT (v1.0), Geoscientific Model Development Discussions, 8, 8607–8633, https://doi.org/10.5194/gmdd-8-8607-2015, https://www.geosci-model-dev-discuss.net/8/8607/2015/, 2015.

Luther, A., Kleinschek, R., Scheidweiler, L., Defratyka, S., Stanisavljevic, M., Forstmaier, A., Dandocsi, A., Wolff, S., Dubravica, D., Wildmann, N., Kostinek, J., Jöckel, P., Nickl, A.-L., Klausner, T., Hase, F., Frey, M., Chen, J., Dietrich, F., NeÌ§cki, J., Swolkień, J., Fix, A., Roiger, A., and Butz, A.: Quantifying CH_4 emissions from hard coal mines using mobile sun-viewing Fourier transform spectrometry, Atmospheric Measurement Techniques, 12, 5217–5230, https://doi.org/10.5194/amt-12-5217-2019, https://www.atmos-meas-tech.net/12/5217/2019/, 2019.

Mertens, M., Kerkweg, A., Jöckel, P., Tost, H., and Hofmann, C.: The 1-way on-line coupled model system MECO(n) – Part 4: Chemical evaluation (based on MESSy v2.52), Geoscientific Model Development, 9, 3545–3567, https://doi.org/10.5194/gmd-9-3545-2016, https://www.geosci-model-dev.net/9/3545/2016/, 2016.

Myhre, G., Shindell, D., BrÃ©on, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and natural radiative forcing, pp. 659–740, Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9781107415324.018, 2013.

Nisbet, E. G., Dlugokencky, E. J., and Bousquet, P.: Methane on the Rise—Again, Science, 343, 493–495, https://doi.org/10.1126/science.1247828, http://science.sciencemag.org/content/343/6170/493, 2014.

Nisbet, E. G., Dlugokencky, E. J., Manning, M. R., Lowry, D., Fisher, R. E., France, J. L., Michel, S. E., Miller, J. B., White, J. W. C., Vaughn, B., Bousquet, P., Pyle, J. A., Warwick, N. J., Cain, M., Brownlow, R., Zazzeri, G., Lanoisellé, M., Manning, A. C., Gloor, E., Worthy, D. E. J., Brunke, E.-G., Labuschagne, C., Wolff, E. W., and Ganesan, A. L.: Rising atmospheric methane: 2007–2014 growth

and isotopic shift, Global Biogeochemical Cycles, 30, 1356–1370, https://doi.org/10.1002/2016GB005406, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GB005406, 2016.

Nisbet, E. G., Manning, M. R., Dlugokencky, E. J., Fisher, R. E., Lowry, D., Michel, S. E., Myhre, C. L., Platt, S. M., Allen, G., Bousquet, P., Brownlow, R., Cain, M., France, J. L., Hermansen, O., Hossaini, R., Jones, A. E., Levin, I., Manning, A. C., Myhre, G., Pyle, J. A., Vaughn, B. H., Warwick, N. J., and White, J. W. C.: Very Strong Atmospheric Methane Growth in the 4 Years 2014-2017: Implications for the Paris Agreement, Global Biogeochemical Cycles, 33, 318–342, https://doi.org/10.1029/2018GB006009, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GB006009, 2019.

Rigby, M., Montzka, S. A., Prinn, R. G., White, J. W. C., Young, D., O'Doherty, S., Lunt, M. F., Ganesan, A. L., Manning, A. J., Simmonds, P. G., Salameh, P. K., Harth, C. M., Mühle, J., Weiss, R. F., Fraser, P. J., Steele, L. P., Krummel, P. B., McCulloch, A., and Park, S.: Role of atmospheric oxidation in recent methane growth, Proceedings of the National Academy of Sciences, 114, 5373–5377, https://doi.org/10.1073/pnas.1616426114, https://www.pnas.org/content/114/21/5373, 2017.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteorologische Zeitschrift, 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, http://dx.doi.org/10.1127/0941-2948/2008/0309, 2008.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model, Journal of Climate, 19, 3771–3791, https://doi.org/10.1175/JCLI3824.1, https://doi.org/10.1175/JCLI3824.1, 2006.

Sander, R., Jöckel, P., Kirner, O., Kunert, A. T., Landgraf, J., and Pozzer, A.: The photolysis module JVAL-14, compatible with the MESSy standard, and the JVal PreProcessor (JVPP), Geoscientific Model Development, 7, 2653–2662, https://doi.org/10.5194/gmd-7-2653-2014, https://www.geosci-model-dev.net/7/2653/2014/, 2014.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, Earth System Science Data, 8, 697–751, https://doi.org/10.5194/essd-8-697-2016, https://www.earth-syst-sci-data.net/8/697/2016/, 2016.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., Weiss, R., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: Variability and quasi-decadal changes in the methane budget over the period 2000–2012, Atmospheric Chemistry and Physics, 17, 11 135–11 161, https://doi.org/10.5194/acp-17-11135-2017, https://www.atmos-chem-phys.net/17/11135/2017/, 2017.

Schaefer, H., Fletcher, S. E. M., Veidt, C., Lassey, K. R., Brailsford, G. W., Bromley, T. M., Dlugokencky, E. J., Michel, S. E., Miller, J. B., Levin, I., Lowe, D. C., Martin, R. J., Vaughn, B. H., and White, J. W. C.: A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by $^{13}CH_4$, Science, 352, 80–84, https://doi.org/10.1126/science.aad2705, https://science.sciencemag.org/content/352/6281/80, 2016.

Schwietzke, S., Sherwood, O., Bruhwiler, L., Miller, J., Etiope, G., Dlugokencky, E., Englund Michel, S., A. Arling, V., Vaughn, B., White, J., and P. Tans, P.: Upward revision of global fossil fuel methane emissions based on isotope database, Nature, 538, 88–91, https://doi.org/10.1038/nature19797, 2016.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, http://dx.doi.org/10.1029/2000JD900719, 2001.

Thompson, R. L., Nisbet, E. G., Pisso, I., Stohl, A., Blake, D., Dlugokencky, E. J., Helmig, D., and White, J. W. C.: Variability in Atmospheric Methane From Fossil Fuel and Microbial Sources Over the Last Three Decades, Geophysical Research Letters, 45, 11,499–11,508, https://doi.org/10.1029/2018GL078127, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078127, 2018.

Wyzszy Urzad Gorniczy 2014, .: Retrieved from http://www.wug.gov.pl/download/5710.pdf, Feb 08, 2017.