Dear Dr. Khosrawi,

Thank you very much for your positive response and the opportunity to address the minor revisions you suggested. We have carefully considered each of the points and have made the necessary revisions to the manuscript. Below, we provide a detailed response to each comment, outlining the changes made in the revised manuscript.

Comments to the replies to the referee comments:

1. I think the term ensemble spread should be still used since this is a commonly used term in the meteorological/assimilation community. Instead you should explain somewhere that it corresponds to the standard deviation.

   - The term "ensemble spread" is used for the first time on L139, where we now added the explanation that it corresponds to the standard deviation:
     "The ensemble spread (corresponding to the standard deviation) of the CH_4^{tot} tracers was sampled at..."

2. Section 3.4: If you refrain from adding these results in the manuscript, doesn't this mean that you have the results already? In your first sentence you state that you do not know the outcome.

   - To clarify, we have indeed carried out preliminary tests with the assimilation of hourly real observations but the results were not convincing as they showed unrealistically large increments and partly very large differences from the assimilation of afternoon data. Although in the idealized setup the assimilation of hourly data gives better results, it remains uncertain whether this is true for real measurements, especially for observations collected in stable nighttime boundary layers, which are very difficult to model. This uncertainty is due to two open questions: (i) whether the ensemble spread reliably reflects the real uncertainties, and (ii) whether the reference simulation is free of significant nighttime biases, the latter being probably the more critical issue.

     A much more detailed investigation of the nighttime model performance would be needed to address these questions, and we felt that this would be outside the scope of the present paper. Including these results without such a detailed analysis could lead to misleading conclusions. Therefore, we decided to limit the scope to the more robust results from the idealized setup.
     We have added a short explanation in the discussion section as to why the hourly data assimilation results are not included, and clarify the limitations or uncertainties associated with the assimilation of real hourly measurements:

"While the assimilation of hourly data provided improved results in the idealized experiment, the preliminary result for real measurements revealed unexpectedly large differences from the results obtained with the assimilation of daytime data only. These real-world tests with hourly data raised concerns about the inclusion of nighttime observations in particular, which are very challenging for the model to represent correctly. Accurate inversion depends on the ensemble spread reliably capturing uncertainties and the model being free of significant nighttime biases. Given the need for further investigation of these issues, we have limited our demonstration to the afternoon/night averages to ensure more robust conclusions."

General comments on the manuscript:

1. Consider restructuring Section 2. Ten pages are needed to reach the result section. Make subsections for the real and idealized experiments, where the respective data sets are described.

   - We appreciate your suggestion to restructure Section 2, and we understand the intent to make the content more accessible and structured for readers.
     Given the nature of this manuscript, which extensively tests a new methodology, a detailed description of the experimental setup is essential. The length of the methods section reflects the need for precision in describing the methodological details. As the results section spans 10 pages, we believe that the ratio of methods to results is appropriate for a paper of this type.
     In the methods section, we have already included subsections that distinguish between the idealized and real experiments, as well as subsections on pseudo and real observations. We believe that further subdivisions would not improve readability.
     Thus, after careful consideration, we believe that the current structure already provides a clear and coherent organization of both the real and idealized experiments.

2. Discuss better the limitations of the experiment with real data. How can these be optimized? Wouldn't this be the aim? Which regions have been considered? Europe? You should clearly state in the text which regions are considered.

   - The setup of the real data experiment is very similar to the idealized experiment described in subsection 2.4. This includes the domain over Europe. The ICON grid is described in Section 2 and is valid for all inversions. However, to make this clearer, we have added the following sentence to Section 2.4:

     "We used the same setup for the ICON simulations with a grid over Europe and the same state vector in the inversions.

We further extended the discussion of the real data application with the following discussion about its limitations:

"We have applied the flow-dependent mdm to real observations assuming that the results are improved compared to the static mdm in a similar way as in the idealized setup. However, while the ensemble spread in the idealized setup accurately captures all transport uncertainties, this is not guaranteed in a setup with real data. Further analysis would be needed to verify whether the ensemble spread provided by the ECMWF EDA NWP system adequately reflects the differences between the measurements and the simulations. Furthermore, in the idealized setup only random uncertainties are accounted for, whereas in reality there may also be systematic transport errors. Systematic errors could result, for example, from a misrepresentation of vertical mixing in stable nocturnal boundary layers, which are known to be particularly difficult to simulate and, at the same time, to have a large impact on near-surface concentrations."

Specific comments on the revised manuscript:

L89: Why perturbed? Do you mean unperturbed? Isn't the perturbation of the background coming from the emissions?

- We do mean "perturbed" here. The global ensemble we use to drive our simulations contains perturbed meteorology as well as perturbed CH_4 concentrations. With these perturbed CH_4 concentrations we drive our background tracer at the boundaries. In addition, the meteorology is different for each ensemble member, which also contributes to the ensemble spread in CH_4^{BG}. However, no emissions are added to the CH_4^{BG} tracer; this only affects CH_4^{emis}.
  In the global ensemble, on the other hand, the perturbed CH_4 concentrations result from perturbed emissions, perturbed meteorology and perturbed observations.
  We have removed the sentence explaining how the CH_4 concentrations are perturbed in the global ensemble. This is now explained in detail in the following subsection. We think that this makes the text clearer, since we now only refer to the ensemble in our simulations:

  "Unlike our previous study with a single forward simulation, we created a meteorological ensemble (see Sect. 2.1) with 10 ensemble members driven by perturbed meteorological boundary conditions and model physics, as well as by perturbed CH_4 boundary conditions from the same global ensemble simulation that provided the meteorological boundary conditions. ~~In this global ensemble, the CH_4 concentrations are perturbed as a result of perturbed meteorology and perturbed emissions.~~ Each ensemble member contains two CH_4 tracers: a background tracer (CH_4^{BG}) representing the perturbed CH_4 boundary conditions, and an emission tracer (CH_4^{emis}) representing the additional CH_4 emitted within our European model domain."

L91,92: Why does CH4tot=CH4bg+CH4emis represent transport uncertainty?

- With this sentence, we express that the ensemble spread in our simulated CH_4 concentrations (and in the end we are interested in the total concentration of CH_4^{BG} + CH_4^{emis}) reflects the transport uncertainty and does not reflect uncertainties in emissions due to disturbed emissions. Although, CH_4^{emis} is the result of emissions in the domain, its ensemble spread is only driven by changes in the transport, whereas the emissions are identical for all ensemble members.

L94 and L97: "10 ensemble members" and "10 members" written -> repetition?

- In one case, the "10 members" refers to the number of members used in the ICON-ART simulations, and in the other case, it refers to the number of members used in the global ensemble. These do not have to be the same; one could also use a selection of the available members for a simulation. Here, however, we use all 10 available members, which is why the numbers are the same.

L105: GOSAT, IASI and TROPOMI -> abbreviations of the instrument names should be introduced.

- We now write out the full names of the instruments and provide the abbreviation in parentheses:

  "~~GOSAT, IASI, and TROPOMI retrievals of CH₄ were assimilated.~~"
  "Greenhouse Gases Observing Satellite (GOSAT), Infrared Atmospheric Sounding Interferometer (IASI), and TROPOspheric Monitoring Instrument (TROPOMI) retrievals of CH_4 were assimilated."

L106: the "driving data" has not been introduced. Which data is this? CH_4^{bg}, CH_4^{emis} or something different?

- By "driving data" we mean the meteorological variables and the CH_4 concentrations for the boundary conditions of CH_4^{BG}, which we describe in Sect. 2. We add this in brackets on line 106 and start the sentence with "In the ICON ensemble simulations" to make this clearer:

  "~~Besides the perturbed driving data, we also applied perturbations to model physics tuning parameters as implemented in ICON for the ensemble data assimilation scheme of the German weather service.~~"

  "In the ICON ensemble simulations, in addition to the perturbed driving data (meteorological variables + CH_4 concentrations to drive the CH_4^{BG} tracer), we also applied perturbations to model physics tuning parameters as implemented in ICON for the ensemble data assimilation scheme of the German weather service."

L139: Here the term "spread" is still used. This contradicts to your reply to the referee comments.

- We had overlooked "ensemble spread" at this point, when revising the text. However, in line with your first comment, we will leave it at "ensemble spread" and instead add that the ensemble spread corresponds to the standard deviation:

  "~~The ensemble spread of the CH$_4$^{tot} tracers was sampled…~~"

  "The ensemble spread (corresponding to the standard deviation) of the CH$_4$^{tot} tracers was sampled…"

L161-163: Introduce abbreviations used.

We introduce the abbreviations JSBACH-HIMMELI, GCP and GFED:

"~~…peatlands and mineral soils from JSBACH-HIMMELI (Raivonen et al., 2017; Reick et al., 2013) (version 2), inland water (provided by Université Libre de Bruxelles to the GCP-CH4 data set; Saunois et al., 2020), termites (Saunois et al., 2020), ocean (Weber et al., 2019), and biofuels and biomass burning (GFED 4.1s; van der Werf et al., 2017) as well as geological emissions (Etiope165 et al., 2019) (scaled to a global total of 15 Tg).~~"

"… peatlands and mineral soils from JSBACH-HIMMELI (Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg coupled to HelsinkI Model for Methane build-up and emission for peatlands; Raivonen et al., 2017; Reick et al., 2013) (version 2), inland water provided by Universite Libre de Bruxelles to the Global Carbon Project (GCP) CH$_4$ data set; (Saunois et al., 2020), termites (Saunois et al., 2020), ocean (Weber et al., 2019), and biofuels and biomass burning from the Global Fire Emission Database 4.1s (GFED; van der Werf et al., 2017) as well as geological emissions (Etiope et al., 2019) (scaled to a global total of 15 Tg)."

L178: real -> What data is included in which experiment? Where is the nighttime data included? Didn't you in your answer state that only daytime data was used for the real experiments?
For the experiment with real observation pseudo observations are used, isn't this a contradiction?

- In the real data application, we follow the standard setup where we assimilate the afternoon means (or nighttime means for mountain station). This is expressed by the sentence in L178. To make this clear, we have added in parentheses (e.g., as in "fc01") to indicate the corresponding inversion in the synthetic setup, which also assimilates afternoon and night means. The "nighttime" part refers only to the assimilation at mountain stations, as described in Sect. 2.3.
  We do not use pseudo-observations anywhere in the application with real data and it is not clear to us where we would have written this in the manuscript.

L211ff: Here you use the term standard deviation.
L267: Here "spread" used.

- In line with your first comment, we now use the term "ensemble spread" throughout the manuscript, except for the figure captions, which use "std" in their labels. Since we explain at the beginning that by "ensemble spread" we mean the standard deviation, we see no problem in using both terms in the manuscript as synonyms.

Figure 6 caption: Figure label "std", but caption "spread".

- The label and caption are brought into agreement by using "standard deviation" also in the caption.

L369: Add which figure.

- Done, also for L373.

Technical corrections:

L88: Space between full stop and "In the global" missing.
L165: according -> according to
L220: a similar principles -> a similar principle
L291: supports -> support ? Please check the sentence.
L395: add comma to the sentence?

- We addressed all 5 technical corrections. The sentence in L88 has already changed due to an adjustment to a previous comment of yours.