Review 1

The replies to the comments are highlighted in green.

General comments

This study examines whether current observation network is capable for detecting future potential changes in CH4 emissions in the Arctic. Arctic is important as vast amount of carbon is stored and could be released as Arctic warming proceeds, leading to positive climate feedback and enhance global warming. The authors use FLEXPART to generate synthetic observations (using current knowledge of fluxes and meteorological data) and examine whether those data can be used to detect scenario emission changes by an analytical inverse model. Inverse modelling has been widely used to quantify current and history of greenhouse gas budgets, but this study attempts to implement it also for studying future changes. This is a novelty.

The study challenges important questions in climate change, but I have few doubts and questions regarding their choice of methods. Particularly,

- The authors study the Arctic CH4 emission changes in 35 years – this is rather short considering the processes of e.g. permafrost thaw, and in comparison to other scenario studies (which are often up to 2100). Because of the relatively short study period, the CH4 emissions are needed to be increased unrealistically fast (20 % yr-1), as authors point out as well, and therefore, the credibility of the results are weak. The choice of length and the increasing rate of emissions need to be justified, and at least add implications for more realistic changes.

The trend of 20% increase per year presented in the study is indeed very drastic. We have chosen this increase to represent a true "methane bomb" event, where large amounts of CH_4 are released in a relatively short period of time. Other, possibly more realistic trends, starting with an increase of 0.1% per year, were indeed simulated as part of this experiment. However, as similar results in terms of network detectability were obtained with smaller trends, we have chosen to present only the results from the highest trend scenario in this paper, as it shows that there is a problem with the ability of the networks to detect even very large changes.

We propose to add the following in section 3.5 to clarify:

“For each of these zones, positive trends are applied separately on wetland and anthropogenic emissions. Oceanic CH_4 emissions are only increased in the sub-region that contains the ESAS, as these are difficult to detect with the surface networks.

The trends are hereby varied between a 0.1 and 20% increase per year for anthropogenic and wetland emissions and between 1 and 100% for oceanic sources. As the results obtained are applicable to both lower and higher trend scenarios, we focus only on the highest selected increase (20% for wetlands and anthropogenic sources and 100% for oceanic fluxes), as this is also the most representative for a "methane bomb" event.”

- The above point leads to a conceptual question about “methane bomb”. In Introduction (P2 L7 – P2 L16), you use this term for both gradual and sudden methane release from the Arctic. As I understood, this study is about the gradual and continuous changes, and this needs to be clarified (Abstract, Introduction and Method).

In literature, the term “methane bomb” has indeed been used for both steady and sudden increase in CH_4 emissions. In this study we define a “methane bomb” event as a sudden and steep increase in methane emissions releasing large amounts of CH_4 over a few years. A similar definition has e.g. been used by Schuur et al. (2022).
We added clarifications of this term in the introduction and abstract.

- If I understood correct, you have generated synthetic data based on present/past prior emission information and meteorological data, which are used to constrain the future scenario fluxes. This would mean that observations would try to adjust emissions to current emission level. So the “detection limit” is when the observations cannot anymore constrain the fluxes to current emission level (+ uncertainty limit), i.e. the limit where observations cannot “see”. Is this correct, and what you aim to do? I would assume that it would be more meaningful if you generate synthetic mixing ratio data based on future emission scenarios, and constrain some prior fluxes with that data. With this, you could see if we can detect emission changes even if there are “missing information” in prior fluxes.

The generated synthetic data is indeed based on present and past prior emissions data, however, as described in the manuscript, we apply trends on these current emission scenarios to generate possible future emission scenarios which we define as the *truth*, thus generating a positive trend on future observations. The prior state of the fluxes is given by current flux estimates and remains unchanged throughout the whole period under study since this is the best guess that we have for the actual CH_4 fluxes in the Arctic region. Ideally, the synthetic observations should move these prior flux estimates to the “true” flux estimates, which is not the case: the prior fluxes are not brought close enough to the true fluxes by assimilating data representative of the truth. This is observable even in the first years of the period under study, where the true state is well within the uncertainty ranges of the prior emissions. Therefore we conclude that the problem lies within the limitations of the practical implementation of data assimilation (i.e., in this case, the sparseness of the network).

- The authors have examined the current and “extended” observation network, but due to the effect of the Russian war, substantial number of surface stations lack of data at current. How likely that we can still detect future changes in CH4 emissions in Eurasia? How long of data lack is critical? I think these are very important questions. You may not need to rerun all simulations without those stations, but adding a few could bring really valuable information about future Arctic CH4 study.

The lack of availability of measurement data as a result of the current war is undoubtedly an important obstacle. When we first implemented this study, this event had only just begun and

the assumption at the time was that the war and the associated restrictions would hopefully not last long. Nevertheless, we believe it is important to carry out our study outside of the current political situation, as this may change at any time.
Additionally, the data of the Russian measurement stations may not be available for many European institutions, however other countries did not interrupt their collaboration with the Russian scientific community. Insights in the methane emissions may therefore still be provided in studies carried out by scientific institutions not affected by current sanctions. Apart from that, the current measurement network in Russia is already relatively sparse, and CH_4 fluxes can only be constrained to a limited extent and in certain regions. Excluding these stations would undoubtedly remove the limited insight that we have into the Russian Arctic, even assuming that many stations could be deployed all around Russia, which can be concluded even without implementing new simulations.

We propose to clarify in section 3.2 as well as in the conclusion:

“Both the current and extended networks were selected based on their theoretical provision of CH_4 measurements, including measurements in the Russian Arctic that may not currently be accessible to the scientific communities of certain countries, as we believe it is important to conduct this work outside of ongoing political conflicts.”

“Current political differences as well as the associated sanctions are an additional obstacle regarding the accessibility of crucial CH_4 observations in the Russian Arctic and Sub-Arctic. As the network in this region is already limited, this missing information may further hamper obtaining a complete picture of ongoing processes in the Arctic, including the detection of a possible methane bomb.”

- Following the previous point, you have completely missed about the role of satellite data. I understand that it is challenging to do satellite inversion with Lagrangian models, but I would at least like to see some discussion about it. What if we have had “surface” data at satellite retrieval points?

Satellite observations have indeed a high potential to compensate for the lack of stationary observation sites in the Arctic in the future and studying an “ideal” scenario of available satellite data in high northern latitudes would undoubtedly be valuable. However, in this study, we focus on the stationary observation network since currently operating remote sensing instruments still face technical obstacles in high northern latitudes. Future satellite missions may offer a better coverage of data in the Arctic, however, these observations will not be available for several years and the final quality of these data cannot yet be predicted.

We propose to add to Section 3.2:

“In this study, we focus exclusively on stationary CH_4 measurements, as our period of study spans several decades. Other types of greenhouse gas measurements, such as satellite observations, are currently limited to providing data for only a few years and are therefore not suitable for our purposes.”

And modify the following paragraph in the conclusion:

“Therefore, efforts to integrate mobile campaigns and new-generation satellite observations into inverse modelling systems should be supported and developed further. Satellite observations in particular offer a high potential to compensate for the lack of in situ observations in the Arctic. The feasibility of using available satellite data products for inverse modelling of methane emissions in high northern latitudes was, for instance, discussed by Berchet et al., 2015 and several approaches integrating these observations in Arctic regions (e.g., TROPOMI CH_4 products, Tsuruta et al., 2023) have been implemented. However, the quality of the data provided by currently operating remote sensing instruments is hampered in high northern latitudes by factors such as high solar zenith angles, low albedo of the Arctic Ocean and limited daylight during polar nights.

However, new satellite missions (e.g., the Franco-German MERLIN project) will possibly provide large, accurate and high-resolution data sets, suitable for characterising CH_4 plumes from regional sources and better constraining methane fluxes in the Arctic.”

Specific comments

P1 L13-15 Please add references to support your argument. I agree that CH4 emissions from wetlands and other freshwater systems are probably a dominant source, but how large are the other natural sources?

References have been added to support the argument and estimates added.

P2 L3-4: Could you add information about how large are the anthropogenic CH4 emissions in the Arctic in comparison to wetland emissions? At end of Introduction: Please make it clear how many years of future scenarios you study.

Estimates as well as years were added in the introduction.

P3 L1: Did you optimize the fluxes grid-wise or region-wise (121 sub-regions)?

The fluxes are optimised region-wise. Clarification has been added.

P6 L7: Could you clarify by “only recently”? What is the year limit you have chosen?

Since 2022. Clarification has been added.

P6 L8: “measurement of CH4 columns” is originally not measuring mixing ratios, but to be used in inversion, you will probably only use the mixing ratio data. Also, satellite data also provide CH4 column information, but those locations are probably not of satellites. This phrase should be clarified better.

We propose to rephrase:

“the stations use ground-based remote sensing instruments to obtain total column measurements CH_4”

Section 3.4:

• Could you possibly change the title to "Generating synthetic CH4 mixing ratio data"?

The title has been adjusted.

• What is the temporal resolution of your generated data? 3-hourly?

In this study, we only generated monthly measurements in order to limit the size of the observation vector.

• Initial concentrations means concentrations in each year (2008–2019)?

Yes, the initial concentrations from the years 2008 to 2019 were used to obtain the background mixing ratios.

P8 L15-18: Please specify a bit more in detail how you have come to 506 different set-ups. It is unclear from the figures/tables as well as from the text. What are the different set-ups, did you change only emission scenarios (as the sentence is is in that section), or did you also use different synthetic data? Did you use different trends, or is all inversions have same trends as presented in Table 2? It is also unclear why there are two similar figures (Figure 2 and 4). Could you possibly combine them?

Wetland and anthropogenic sources were increased in each of the 121 sub-regions as well as the 5 supra-regions, which gives 126 regions in total. We use 2 different observation networks. That gives 252 (126 x 2) scenarios for wetlands and 252 scenarios for anthropogenic sources. Oceanic sources were only increased in 1 region, using 2 different networks, which gives 2 scenarios. So in total there are 506 (252 + 252 + 2) scenarios.
A more extensive description has been added to the section.

P9 L8-9: Why "only this region should be updated by the inversion"? Is East Eurasia strictly uncorrelated with other regions? Did you strictly set it so that observations are only constraining this region? If not, it is not surprising that other regions are also affected.

Only this region should be updated because only this region is truly higher than in the prior, which should, ideally, be seen by the inversion. However, this is not the case, since the data assimilation is not perfect in our study,

P9 L10: Is it so that the posterior emissions are much lower than the truth because the observations are generated using present-day emissions?

The observations are in fact *not* generated using present flux estimates but using future emission estimates. This is explained in Section 2 of the manuscript. (See also answer to third comment of general comments.)

P10 L5-7: Is it really so that the "increase in the simulated scenarios is underestimated"? I wonder how strong are the regional correlations. Also, do you trust the "truth" or posterior estimates? You need to re-think how to put your arguments.

In our inversion set-up, we do trust the "truth" over the posterior state - since we define it as the true, future state of the fluxes. Which does not mean that these emissions are likely to occur in reality, but they are the true state of the inversion set-up and thus, the posterior state should be optimised accordingly.
We propose to clarify:

"This means that the increase retrieved in the posterior state is underestimated compared to the generated truth in the "correct" area, which is considered to be the true state of the emissions in this inversion set-up. This is partially compensated for in the total posterior by overestimations in the same emission sector in different regions."

P11 L2: By "combine", do you mean that you only show the results of the region where you modified the trends, i.e. the effect of other regions are not presented? Please make it clearer.

Yes, in the figures in section 4.2 we present 121 set-ups with elevated trends in each corresponding region. The effect on other regions from each scenario has been evaluated with the corresponding equation in section 4.2.3. We considered this the best way to present the variety of results of the different configurations. The introduction to Section 4.2 has been extended, more detailed descriptions of the maps and the corresponding calculations are in the sub-sub-sections.

P11 L 20-23: I am not sure what you wish to say. The applied trend is unrealistic, and you hoped that the inversion would detect the changes much earlier? Or you think that you should have applied a bit more realistic trend? What you mean by "more illustrative" – more, compared to what?

Here, we want to express that it's not so much the year of detection which is interesting, since it's based on a probably unrealistically fast methane bomb, but the smallest amount of emissions which can be detected, whenever they may happen in time.
We propose to clarify:

"Hence, it is more illustrative to analyse the smallest amount of emissions which can be detected, as shown in Figure 6b, than simply using the year of detection as an indicator. "

P12 L10-11: Is it really true that there is no influence about observations? What if there is a station over there? I would also guess that the observations in surrounding regions could affect the results.

In the oceanic regions that we describe, CH_4 emissions from wetlands are 0 in the emission data set (which is logical since wetland emissions do not occur in the Ocean). Hence, our generated *truth* is also 0 as well as the computed posterior emissions. That observation sites in these oceanic regions would further improve the detection is unlikely - since in our set-up there are no emissions to be detected in the first place.
As for the neighbouring regions, in data assimilation, any additional data is generally beneficial. So in a sense having observation sites in the Ocean could potentially bring information on wetland emissions, if the oceanic sites are sampling air masses coming from

wetland regions. The interest of these data compared to the large uncertainties in the transport and the difficulty of maintaining such oceanic sites is however questionable.

P12 L13-15: This is interesting, but could be also due to the fact that many of the extended stations are often close to the currently available stations. Also, the emission magnitude near the station is important to consider – if we add stations where emissions are small, the effect could be minor.

This is certainly true. However, in remote regions it is difficult to locate hypothetical observation sites, as they may be in regions where both construction and maintenance are not feasible. Therefore, for this study, we have chosen actual, potential observation sites where we know it is technically possible to measure CH_4 mixing ratios.
We propose to add:

“One possible reason for this could be related to the locations of the additional observation sites, as several of them are located close to operating measurement stations and/or in areas with low estimated CH_4 fluxes.”

P13 L1-2: Is there anything you could do to attribute those discrepancies to fluxes by changing some setups/uncertainties? Despite the minor effect on your results, do you still think those sites are important and could bring information about changes in trends in northern Europe or surrounding regions?

These observation sites may not be beneficial in detecting a potential methane bomb in the Arctic, which was the focus in our study, but they are certainly useful for other cases, e.g. European scale monitoring of CH_4 emissions.

Technical corrections

All technical corrections have been implemented, unless there were justified objections from our side not to make the changes (see responses to comments).

Please use same terms for generated mixing ratio data (modelled, generated, synthetic, etc..)

Please check the spaces between units, and follow the journal role.

P1 L10 Remove “temperature”

P5 L10 Section Inversion framework
Please add section number

P11 L10: annual posterior emissions in year j and region r emis aj,r

P11 L12: Please move the j and r ranges on the right hand side of the equation, i.e. emissa – emissb < e, j∈[2021, 2055], r∈[1, 121] You could put “is not fulfilled” in L10. Please also do so in Eq. 5 and 6.

We prefer to have the range of *j* and *r* below the equations, as this is a legitimate way to define them and there are no contradicting guidelines from the journal.

P11 L16: “the threshold year is generally higher” Do you perhaps mean “the year is generally later”?

P11 L25: “terms of detection limits, an increase of a few, up to 10 Tgy -1 , is necessary for statistically reliable detection.” Could you add e.g. in brackets how much they are in percentage?

P15 L18: “TROPOMI CH4” → CH4 with subscript.

Figure captions: Use (a), (b) instead of “left” “right”.

Figure 3 caption: I feel it would be more appropriate to say e.g. “Location of the sites where synthetic mixing ratio data are generated from”, as you do not use actual observations at all.

Figure 5 y-axis: Are those units really correct? For example, in the bottom panel, 100 Tg/month of CH4 from Arctic in 2020 does not sound at all realistic (even if it was annual emission). Y-axis label and caption does not have same units.

The prior annual emissions are actually in this magnitude for the “*entire region*” (which encompasses a larger area than the Arctic) shown in the supplements. This is predominantly due to the wetland emission data set we used as prior information (Poulter et al., 2017) which makes up around 50% of those emissions.
Y-label has been adjusted.

Figure 8:
• Please use more informative label in the color bars.
• The unit in color bar is [%], and color scales ranges between -103 to 103, i.e. 1000% change
in emissions. Is this correct?
• Caption for (b): “Difference between the…” → “Absolute differences between..”?