“Assessment of the theoretical limit in instrumental detectability of Arctic methane sources using 13C atmospheric signal” *by* Thibaud Thonat et al.

Reviewers’ comments are in italic blue.
Responses are in normal black font. **Changes in the text are in black bold.**

Response to Anonymous Referee #1 –

We are very grateful to Referee #1 to have reviewed the manuscript and submitted helpful comments and suggestions to improve both the study and the text.
Here we respond to the reviewer point by point.

GENERAL COMMENTS
In my opinion, the scientific value of being able to detect methane emissions from wetlands in the Arctic is limited. We know that those emissions exist, and that they are important. More interesting is to be able to improve their quantification. For that, detection is not a sufficient requirement. The detection of regional trends would add significant understanding, but for that the requirements will be different. The question is not only about single measurement precision, but also the minimum number of measurement sites needed. This also brings in the dimension of data averaging, reducing the requirements depending in the statistics of the errors, the measurement frequency, and the temporal resolution that is needed. The conditions that are used to define ‘detectability’ in this study are not well motivated. Since the required measurement performance will depend on the details of the scientific questions that the measurement should help to answer, however, I think that to quantify the expected amplitude of variation is a more important outcome. It is possible to turn this into requirements, but then the purpose should be more clearly defined, and the inevitable limitations should be discussed as well.
To address this first part of the general comment, we acknowledge that the detectability definition used in the submitted text (based on the signal departure from the background) was not suitable for observations analysis. We have redefined the “detectability” from an inverse modeling point of view. As a result we now analyze the amplitude of the variations of the total simulated signal – corresponding to what would be measured in the atmosphere. We compare the amplitude of the simulated signal to some instrument precision (called threshold). Then we determine which source (including boundary conditions) contributes the most to the variation in the simulated/expected signal. We acknowledge that this is a first order contribution as sources may overlap in time and space.
An important distinction is found between remote, and regionally to locally influenced stations. Since the signal amplitudes differ between those sites, so will the measurement requirements. Yet the abstract and conclusion sections generalize the requirements to a single set. It should be made clearer what kinds of sites are addressed by the numbers that are listed (rather than just a statement that the requirements will vary between sites).
This second part is also addressed in the revised manuscript: we include variations in the signal due to boundary conditions and regional/local sources. Also Figure 5 has been modified and presents the potential detection at all

stations, allowing to have a quick look at which station is able to detect which source depending on the instrument uncertainty. We now include more discussion in the text and more details of the results in the abstract and conclusions.

More useful would be to distinguish between applications. For some applications the requirements may be less stringent, especially if a larger number of cheaper sensors are deployed.

This part has also been addressed when clarifying the detectability definition. We now clearly state that this study focus on signal that could help regional inverse modeling in better quantifying methane emissions. So that we define the detectability based on daily signal – used in regional inverse model.

Over land, the amplitude of the signal will depend strongly on the altitude of the air inlet, and therefore the model level that is sampled. The altitudes in Table 1 probably refer more to the local orography than the height of the measurements with respect to the ground. There is a potential for increasing the significance of this work by adding the vertical dimension. What is the implication for required accuracy of towers and aircraft measurements?

Yes the altitudes in Table 1 refer to the altitude of the station not of the air inlet. Here we use the inlet altitude corresponding to each existing site, associated with the corresponding vertical level of the model (as will be done for an atmospheric regional inversion), so this should include the existing tall towers. Using aircraft measurements is not really appropriate in our framework where we consider daily means.

SPECIFIC QUESTIONS

page 3, line 140: Although not long-term, the benefit of high frequency measurements was convincingly demonstrated by Roeckmann et al (acp, 2016).

This reference has been added to the text. "**For example, Röckmann et al. (2016) have deployed high frequency isotopic measurements of both $\delta^{13}C$-CH_4 and δD-CH_4 at Cabauw in Europe and were able to identify specific events and allocated them to specific anthropogenic sources (ruminants, natural gas or landfills)**."

page 4, line 218: It seems that the detectability of biomass burning could be influenced by the use of monthly average emissions, since in reality they may vary strongly with time.

Actually there was a typo in the text as we do use daily emissions from GFED and not monthly. So the detectability calculated here does take into account the strong temporal variation of biomass burning. "**monthly**" as been changed to "**daily**". However the signal of this source would also highly depend on the studied year as biomass burning has strong inter annual and spatial variability: we added a comment on this in Sect. 4: "**This study has been carried out only for the year 2012 as a test case. However, not all emissions have a high inter annual variability, such as does biomass burning. As a result, our findings should be still valid for the other sources for most of the years over a few future decades.**"

page 4, line 224: GLOGOS

Typo corrected

page 5, line 255: The d13C value of natural gas from West Siberia is known to be highly depleted (see e.g. Tarasova et al, 10.1007/s10874-010-9157-y)
We know include sensitivity tests to d13C signature for natural gas, and the isotopic signatures range between -40‰ to 50‰, with a mean value of -46‰. (**see text and Table 3).**

page 7, line 366: 'However, they are excluded from our analysis ...' But later the threshold detectability is defined from the source making the largest contribution to the signal. Shouldn't this signal include variations due to the background (it they overwhelm the regional sources this should limit the detectability).
Indeed, the signal does include variations from the background, that is our boundary conditions here (lateral and top of the model). To address this and refine our analysis, we have first deleted this sentence and then changed the way we calculate detectability. "**Here we focus on a detectability definition taken from a regional inversion point of view: regional inversion systems analyse daily signals and optimize sources depending on synoptic deviations of the observed signals compared to the simulated ones. Therefore, a measuring instrument is considered to provide useful information to the inversion only if the synoptic variability of the atmospheric signal can be detected. To that end, we compute detectability capability in Fig. 5 and Tab. 4 as follows: (1) we compute the standard deviation over a five-day running window of the simulated total isotopic signal; (2) for a set of instrument precision threshold (from 0.2 to 0.01‰ see Fig. 5 and Tab. 4), if the running standard deviation is higher than the corresponding threshold, the source with the higher running standard deviation for the same 5-day window is considered detected for that one day; (3) for each threshold, we count the number of days over the year that each source is detected. Although the total atmospheric signal integrates contributions from different sources with different isotopic signatures, we keep only the major source contributing to the signal as a first order signal.**" In this way we are able to distinguish when the variation in the signal is due to the background (boundary conditions) or to regional sources**.** For some stations (such as Churchill), close to the border of the domain, the background contributes the most to the signal variations (new Fig. 5). This is further discussed in the revised manuscript.

page 8, line 441: Wouldn't the fact that the most significant sources all lead to methane depletion limit detectability. How do you distinguish one depleted source from another? It occurs to me that the definition of detectability ought to take differences in signatures into account, rather than only single process contributions.
Indeed the atmospheric signal integrates the contributions from the different sources. Here we select the source that contributes the most to the depletion though we acknowledge that several sources may simultaneously contribute. However discussing the overlapping in time and space of the sources is challenging without any real measurements as both the emission source and magnitude and the isotopic signatures are uncertain in the model. As a result, we present here a first order signal. After the definition of our detectability, we have included the following sentence:" **Although the total atmospheric signal**

integrates contributions from different sources with different isotopic signatures, we keep only the major source contributing to the signal as a first order signal."

Table 4: Is the year dependence of the thresholds important enough to restrict it to the year 2012?
We acknowledge that multiyear simulations may strengthen the results, especially if the year 2012 were specific for any reason. However this study is a test case and more efforts will be made as soon as continuous measurements are available (which should happen soon). We expect the year dependency being important mainly for biomass burning emission detection. In the discussion, we have added the following sentence:" **This study has been carried out only for the year 2012 as a test case. However, not all emissions have a high inter annual variability, as does biomass burning. As a result, our findings should be valid for the other sources for most of the years over a few future decades**."

Figure 3: What do the triplets of numbers at each site represent?
Figures 3 has been re-arranged to facilitate its reading. The triplets have disappeared. They indicated average, low and high range of total contributions to isotopic ratios.

Figure 5: This shows that for a median wetland signature, the threshold of 0.5 per mil listed in the abstract would yield no single day of measurements. This seems to suggest that 0.5 is a too relaxed requirement.
The conclusions in the abstract have been modified accordingly to the new definition of detectability. Also we detail more the results for the different types of stations.

TECHNICAL CORRECTIONS
Page 2, line 63: carbon dioxide
page 4, line 235: ERA-Interim reanalysis
Table 2: 'Range' i.o. 'Variant'
The technical corrections have been applied