Reviewer comments are italicized

Author responses are indented and unitalicized

Changes to the manuscript text are in blue

The authors would like to thank the anonymous reviewers for their time and attention to this manuscript. Your comments and suggestions have been greatly appreciated and have significantly improved the paper.

Response to anonymous referee 1

I found the results to be of some though limited interest because they show how a method previously developed for correcting TROPOMI data on a global scale can be customized for local application to improve the correction.

The authors would like to clarify that this is a novel method that uses similar elements to previous corrections, and it does not primarily address global vs. local differences, but rather monthly vs. annual correction values. As monthly corrections are more spatially dependent, we executed this correction on a more local scale. Theoretically this method could be scaled up, but would likely have to be executed separately on local domains and later pieced together for the most valid results.

But I'm puzzled by the motivating premise of this paper that the TROPOMI retrieval does not account for seasonally variable surface albedo. I believe that it does because surface albedo is co-retrieved with methane for every spectrum.

The reviewer is correct that some seasonality is incorporated into the existing albedo correction algorithms. The albedo retrieval has high time resolution, but the basis of generating a correction factor in the operation product does not. We have gone back through the manuscript to clarify this and to explain more clearly that the seasonality that we describe correcting comes not from the incorporation of daily surface albedo measurements, but instead a changing relationship between the surface albedo SWIR (as well as other variables) and the correction value. For example: the relationship between sunlight angle, aerosol optical thickness, and the correction value is likely quite different in the summer than in the winter in a way that the static correction value curve constructed by Lorente et al. does not take into account.

In many regions, especially those with agriculture, surface reflectance depends on season, but corrections for this dependence are calculated from long-term averages.

When applied to methane retrievals on a seasonal basis, we show that some residual albedo effects are still apparent and may thus bias seasonal data.

In this work we demonstrate that seasonal or monthly averaged methane retrievals continue to be biased by albedo effects after the implementation of these correction algorithms.

When the data are shown by season, this is no longer true - Pearson correlations with an absolute value greater than 0.1 indicate that there exists some correlation between the SWIR surface albedo and the albedo-corrected methane retrieval. This correction algorithm does account for some seasonality because the TROPOMI retrieved variables include the surface albedo SWIR which is used in the Lorente et al. algorithm to calculate a correction value. The seasonal correlation reappears after the Lorente et al. correction because this correction assumes that the relationship between surface albedo SWIR and correction value is static over time.

Regardless of the reason, the shifting albedo and seasonally variable albedo effect biases methane retrieval data from TROPOMI at finer time scales.

The Lorente et. al. correction, which handles seasonality with a static correction based on SWIR surface albedo, significantly improves upon the uncorrected data, but preserves the seasonal trend in the data, demonstrating larger, positive correlations in the winter months and cycling through the seasons.

I'm also concerned about the use XCH4 as a predictor variable (which turns out to be the most important predictor) because it makes the correction to the variable depend on the variable itself – the authors expressed concern that using wind speed as predictor variable would propagate as an aliasing factor in inversions, but using XCH4 as a predictor variable seems worse in that regard.

Reviewer 2 brought up similar concerns, so this is obviously an important point. The only reason to not include the XCH4 and XCH4_corrected variables is if there were a danger of erroneous retrievals, which would be flagged and removed anyway. The fact that these variables are always of high importance is reassuring that the built-in algorithm is mostly correct all of the time. If there were variably much larger differences in the TROPOMI retrieved values and the predicted values, then the importance of these variables would likely decrease substantially.

As for the issue of circular inputs: wind speeds are avoided because wind speed is an input in the flux divergence method of quantifying methane emissions. We did not want to use wind speed to predict the methane concentration, then use wind speed again to calculate methane emissions. However, the XCH4 and XCH4_corrected variables are only used in the calculation of methane concentration. The model-predicted value of methane concentration is used in the calculation of methane emissions. This is difficult to describe and understand in text, so please reference the following cartoon describing this, which has also been added to the SI:

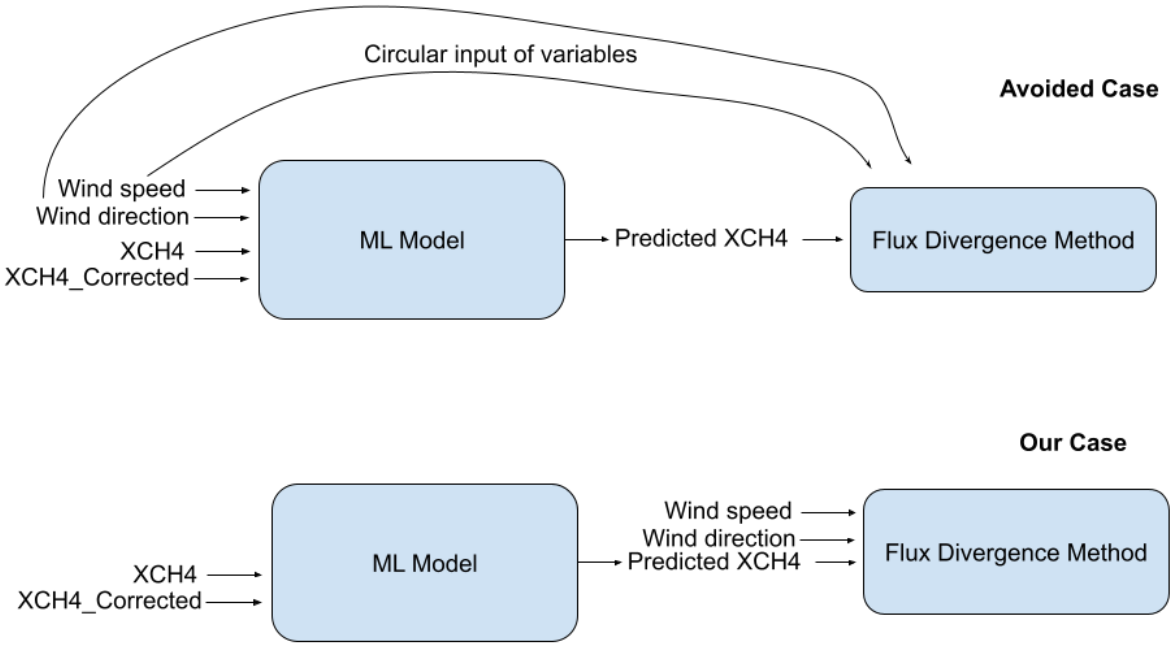

Figure S5. Demonstration of circular input of wind speed variables. The top panel shows the input of wind speed and wind direction into the ML model, and again into the flux divergence method, potentially causing double counting problems. The same issue is not seen with the XCH4 and XCH4_corrected values because those are not incorporated into the flux divergence method - only the model-predicted XCH4 is an input.

There is also a lot of chatty prose and repetition in this paper, some of which does not follow standard scientific practice of conciseness and focus. We don't need to hear about Mark Twain or Horace Greeley.

These brief quotes are meant to illustrate important points within the manuscript.

1) The Greely quote is meant to showcase how important agriculture is in this region and how the landscape is grassland/high desert that has been converted to farmland. In our opinion, this is a potent point to make, demonstrating why this agriculturally-minded albedo correction is especially important in this region.

2) We provided the flow rates of the two rivers in this region and compared it to the Mississippi river to give a numerical idea of scale. However, we believe some readers might connect more with a verbal illustration. Twain's quote disparaging the waterways was meant to illustrate just how small and pitiful these rivers are in reality. If you're unfamiliar with the North and South Platte rivers, "river" is an extremely generous term, and the quotation was meant to give that impression.

We understand the reviewer's concern, so we have cut the Greeley quotation. However, since the Twain quotation directly addresses a reviewer's later concern, we have decided to keep this quotation along with removing repetitive text throughout the manuscript.

Line 12, Abstract: not sure what 'many satellite products' means. UV/Vis retrievals indeed often assume fixed surface albedo, but that wouldn't apply to the TROPOMI methane retrieval which calculates its own surface albedo and would thus account for seasonality, unless I'm missing something.

We have removed the ambiguous "many satellite products" and we have refined the language around the surface albedo correction. As this is the abstract, we cannot fully explain the deficiencies of current albedo corrections here, but we have also modified the language elsewhere to better explain the phenomenon we are attempting to solve.

In many regions, especially those with agriculture, surface reflectance depends on season, but corrections for this dependence are calculated from long-term averages.

Later in the results and discussion section we describe

This correction algorithm does account for some seasonality because the TROPOMI retrieved variables include the surface albedo SWIR which is used in the Lorente et al. algorithm to calculate a correction value. The seasonal correlation reappears after the Lorente et al. correction because this correction assumes that the relationship between surface albedo SWIR and correction value is static over time.

We believe this better describes what the reviewer expects; that the Lorente et al. correction is performed based on retrieved albedo values which will fluctuate seasonally. We add that the long-term nature of the correction calculation leaves out important seasonality in the relationship between the SWIR and the XCH4, which the reappearance of the correlation between these values in the winter data confirms, must change over time.

Line 66: here and below, the description of machine learning seems pretentious to me. Here it's just being used as a non-parametric statistical fit.

Yes, machine learning is just another method for curve fitting. In presenting this work at conferences we consistently asked the question "who here is familiar with machine learning?" and got very little positive response, so we wanted to include a very simple definition of machine learning to make the work more accessible to a wider audience. We recognize that our intent to simplify the description made our descriptions less accurate. We have overhauled the language in this section:

Machine learning is a branch of artificial intelligence where computers are trained to recognize patterns and make decisions based on data, somewhat like how humans learn from experience. Some machine learning models are considered a 'black box' because it can be difficult to understand how they make decisions. To address this, tools like SHapely Additive exPlanations (SHAP) help provide insights into how machine learning models arrive at their predictions. (Lundberg and Lee, 2017; Rudin, 2019). Neural networks, a type of machine learning model, are inspired by the way the human brain processes information. They consist of layers of 'neurons' that work together to identify patterns in data(Martín Abadi et al., 2015).

Line 174: here and elsewhere, a driving motivation for the paper is to apply seasonality to the albedo correction from Lorente et al. 2021. I couldn't find a description of the Lorente correction but my understanding is that it is an improved polynomial spectral representation of the albedo co-retrieved with methane. If so it would have seasonality.

We are not attempting to apply seasonality to the Lorente et al correction, we are developing a novel method that provides a finer time resolution seasonal correction. You are correct as to the representation of the Lorente et al. correction and its existing seasonality. However, as discussed above, this correction was constructed using long-term averaged data and our evidence shows that the relationship between surface albedo SWIR and the correction value should change over time. Figures 2, 3, and 7 in the manuscript all provide Lorente et al. data, and all show that it is affected by seasonal albedo effects.

The authors recognize that this was not made clear in the original submission and our poor choice of language contributed to the reviewers confusion here. We have significantly updated the language of the manuscript.

In many regions, especially those with agriculture, surface reflectance depends on season, but corrections for this dependence are calculated from long-term averages.

When applied to methane retrievals on a seasonal basis, we show that some residual albedo effects are still apparent and may thus bias seasonal data.

In this work we demonstrate that seasonal or monthly averaged methane retrievals continue to be biased by albedo effects after the implementation of these correction algorithms.

When the data are shown by season, this is no longer true - Pearson correlations with an absolute value greater than 0.1 indicate that there exists some correlation between the SWIR surface albedo and the albedo-corrected methane retrieval. This correction algorithm does account for some seasonality because the TROPOMI retrieved variables include the surface albedo SWIR which is used in the Lorente et al. algorithm to calculate a correction value. The seasonal correlation reappears after the Lorente et al. correction because this correction assumes that the relationship between surface albedo SWIR and correction value is static over time.

Regardless of the reason, the shifting albedo and seasonally variable albedo effect biases methane retrieval data from TROPOMI at finer time scales.

The Lorente et. al. correction, which handles seasonality with a static correction based on SWIR surface albedo, significantly improves upon the uncorrected data, but preserves the seasonal trend in the data, demonstrating larger, positive correlations in the winter months and cycling through the seasons.

Line 294, Figure 5: it would be good to show the actual TROPOMI data before the correction. Data over rivers are bad and really should be removed from the dataset, but maybe they are not flagged as bad in the retrieval?

The effect observed in Figure 5 is due to the agriculture surrounding the rivers as described in the manuscript. Spot checking a number of files for the area, no pixels were flagged as "water" and only rarely as "some_water" the criteria definition for which is "Pixel contains water (however small the fraction), i.e. at least one of the 15 × 15 arcsecond subpixels in the SDP dataset is classified as water" in the TROPOMI documentation. This means that the area of water inside each pixel is sufficiently small as to not bias the retrieval data. "River" is a very generous term when describing the North and South Platte, where the widest parts of the waterways are on the order of

150 and 50 meters respectively. These waterways are paltry and not navigable and do not trigger any quality flags in TROPOMI retrievals. We have strengthened the language here to make this point more clear.

As figure 5 is focused on correction magnitude instead of methane concentration, we feel it is more appropriate to include uncorrected data in figure 7 instead, reproduced below.

The North and South Platte rivers are extremely small (average discharges 1,355 and 175 cu ft/s respectively, Mississippi river is 593,000 cu ft/s) and are far less in extent than one satellite pixel, making the flagging or removing of this data due to water content, unnecessary.