

Review of “Deep Transfer Learning Method for Seasonal TROPOMI XCH₄ Albedo Correction” by Bradley et al.

This manuscript describes the use of a deep transfer learning (DTL) method for applying bias correction to TROPOMI XCH₄ retrievals over a region of the central U.S. The study develops on previous bias-correction schemes for TROPOMI CH₄ through use of DTL in order to produce monthly-varying models for the bias correction. The selected target region covers parts of Colorado and some surrounding states, areas with wide-ranging agricultural activity, urban areas, and undeveloped rural areas, providing good testing data for their models.

Overall, the authors have done a good job of developing and applying their model to the selected region and show quite well the impact of the monthly-varying models on the imposed bias corrections, linking them to different crop types and other landscapes. This has the potential to be of some interest to the wider community. However, in my opinion the manuscript requires some significant work before it is ready for publication.

If the authors can adapt the manuscript based on the following comments, then the work can be published in AMT.

Major comments

- 1) Throughout the paper, the authors are muddled in their explanations and terminology about the seasonality of the albedo correction in their work and the previous works by Lorente et al. and Balasus et al.. As I understand it, in both of these previous works the corrections applied to the retrieved XCH₄ do vary over time (as albedo and other variables change), it's just that the models used for calculating these corrections use all available data and do NOT vary with time. Bradley et al. are quite careless in their terminology throughout in describing this concept and it is misleading and confusing for the reader.
- 2) The above point raises important (linked) questions that Bradley et al. do not satisfactorily discuss in the manuscript:
 - What is the physical justification for applying a monthly-varying bias correction model as opposed to a fixed model? I assume that it's ostensibly that the relationship between the satellite bias and the predictor variables changes with time, but why is this exactly?
 - The monthly DTL models appear to assign different relative import to the input variables in the different seasons, but why is this the case, and what is the implication of this?
 - Why exactly is the temporally-fixed model of Balasus et al. for example, not sufficient in this case (beyond the fact that it is trained on global rather than regional data)?

These questions must be raised and discussed in my opinion. As a non-expert in ML methods, I can't comment on the technical detail, but in other data-fitting

schemes there is the concept of ‘overfitting’ one’s model to the ‘true’ or observational data. Is there a similar possibility here, where applying the DTL models separately to each month’s data might destroy the conceptual generality of the model?

- 3) The authors here build on and compare to Balasus et al.’s global correction model and they have changed 3 major factors in comparison to that model – the regional bounds covered by the training data; the use of the DTL method for monthly models; and finally, the chosen subset of predictor variables. Therefore, it’s impossible to say how much of the improvement is due to the specific regional training and DTL method, and how much is simply due to the fact that (e.g.) XCH4 and XCH4_corrected are now input variables (and other variables have been removed). Is it possible to include results from a version of your regional DTL models that keeps the same predictor variables as Belasus et al.?
- 4) The inclusion of XCH4 and XCH4_corrected is not justified in this work and should be. The authors state that they stopped using wind speed variables to avoid circular use of inputs in potential top-down application of this data, but shouldn’t the use of XCH4 itself should be avoided for a similar reason? Especially as XCH4_apriori, XCH4_corrected and XCH4 turn out to be 3 of the 5 most important predictor variables.

Specific comments

Line 28: This Karakurt et al. reference is quite dated now – can you cite more up-to-date values?

Line 36: Quantify ‘relatively large’.

Line 38: Note the assumptions and limitations of bottom-up methods also.

Lines 48-49: These statements about the limitations of current albedo corrections need a reference.

Line 59: Clarification needed here – “cropland seasonal albedo shifts are not currently compensated for in albedo corrections”. However, as I understand it, the other correction methods should indeed account for the changing albedo (which is retrieved for each sounding from the satellite and included in their correction models) – it’s the model for the magnitude of the correction which is consistent over time.

Line 76: If you want to stop including the year for Lorente et al.’s and Balasus et al.’s work from this point, you should now clearly state that the shorthand that you’re using thereafter refers to a particular paper (with year cited).

Line 76: What do you mean by ‘well-formed’?

Line 92: ‘different numbers of averaging kernels’ – Is this correct as stated? Is it clearer to say that the (number of/locations of) vertical retrieval pressure levels are different for the two satellites, and therefore the averaging kernels are different?

Line 106: Define QA.

Line 107: SWIR is defined later in the text but should be defined here (or possibly earlier?).

Line 112: Make clear here that this is not something that you do in this study, however, due to your implicit striping correction?

Line 119: Add more detail here about the co-location criteria so the reader does not have to refer to the Balasus paper.

Line 121: This Kriesel reference needs more information for the reader to find it. Indeed the author himself explicitly asks that any citation includes the URL, and states a year. http://www.dkriesel.com/en/science/neural_networks. There are other locations where the authors should be more careful and explicit with their citations.

Line 127: Are these changes really ‘minor’? See earlier major comment.

Line 128: To be clear here – what exactly is the ‘corrected XCH4’? The Lorente et al. bias correction?

Line 132: As a non-expert, I need more clarification of what the hidden layers/levels are.

Line 134: More detail needed here. Trained and tested on 80% of the data? How much of this was training data and what was the other 20% used for? How was the 80% selected? Was it evenly distributed across months – or do months potentially get different amounts of training data? Would that be an issue if so?

Lines 192 – 197: I found this information a little hard to follow as structured here. Can you make clearer?

Line 211: ‘not designed to handle seasonality’ – rephrase.

Around line 265 – what does it mean that the relative importance of predictor variables in your work appears quite different to those of Belasus et al.? (e.g. Aerosols/Surface Albedo SWIR are very important for them).

Figure 5: Discuss – why are the urban corrections so variable? Should they be relatively constant?

Line 325: To be clear – is this testing the significance of the differences between derived corrections over the different land-types?

Line 336: I still find this reasoning confusing – I apologise if I’m misunderstanding. Other corrections do take account of the changing albedo of the crops, which is retrieved by TROPOMI, right? So the reason that they’re not capturing the seasonal variability is their model? But why is the relationship between bias and albedo changing with time?

Figure 7: I think I’m right in saying that figure captions in this journal should include a description of the figure only and not ‘discussion’-type text such as this one comparing results in the different panels.

Line 346: ‘appears to have’ – you should be able to quantify this.

Line 347: ‘due to the correction’s dependence on the UoL GOSAT ... data’ – you should confirm that the Lorente data is based on GOSAT? I was under the impression that Lorente et al.’s correction was not based on GOSAT.

Line 350: ‘appear visually smoother’ this is vague and unscientific.

Lines 351 – 357: I’m slightly unclear as to what you are saying with this statement regarding the lack of necessary GOSAT data. Can you clarify further? Is it that you are not able to produce a map of correlations?

Technical corrections

Line 35: The (number/density/range?) of atmospheric measurements has expanded...

Line 43 and elsewhere : Include hyphen: “Sentinel-5 Precursor”

Line 47: “currently-used”

Line 51: top 10 U.S. states

Line 61: change to: (CAFOs)

Line 64 and elsewhere: be consistent in spelling of collocation/co-location throughout.

Line 275: “Model-predicted”

Line 277: Surely red, not brown? Additionally, it would be good if Figure S1 used the same red-blue colourbar if possible. The green colours (ranging from around +5 to -15 ppb) are hard to distinguish from each other.