We sincerely thank the reviewers for their very helpful feedback on this paper. We address all their comments and suggestions below.

Reviewer comments are in black.

Our responses are in blue.

Where practical/necessary, we provide a screenshot of the track-changes document to show the changes that we have made (in outlined boxes). In these, text that is removed is ~~struck through~~ and coloured red, while new text is underlined and coloured blue.

Ian Ashpole (on behalf of both authors).

<u>Anonymous Referee #1</u>

Received and published: 25 January 2020

This study examines how MOPITT version 7 TIR-NIR retrievals compare over land versus water over and around a single coastal city – Halifax, Canada. The authors examine the L3 product as well as the L2 product averaged separately for land and water over the same grid box. They generally find that retrievals over water retrievals are higher than over land. Most of this difference can be accounted for by differences in averaging kernels. While this paper has numerous strengths (nearly flawless grammar, and very detailed), I think a step back is needed to understand the overarching goal of this study or central question to be answered. This may involve a scope change from e.g., a very local to a global scale, examining additional MOPITT products, and/or including additional measurement results or results from a model the authors run. While I cannot recommend publication in its current form, I encourage resubmission when the comments below are addressed.

General comments (Some of these general comments may supersede specific comments below).

<u>G1:</u> The paper is quite detail oriented, but at times there is so much repetition and additional words and phrases that it makes it difficult to read. I think the length could be cut down by as much as 50%. Part of this could be through wording changes, part of it by reorganization, and part of it fewer numerical details that could be succinctly summarized in Tables. Aim for describing things in context rather than the large number of uncommon symbols and abbreviations.

Thank you for the feedback. We have done a thorough edit which entailed removing a lot of the repetition and 'over-explaining' and aiming to be much more succinct. We do not repeat values or numbers that are clear in figures or tables that we are discussing, apart from when it is absolutely essential to emphasise the point. We have also removed some of the acronyms to make certain sections easier to follow (especially Section 3.1.3). We sincerely hope that this makes the article easier to read.

<u>G2:</u> One of the major conclusions of this paper is caution in examining retrievals around coastlines throughout the world. This conclusion is not supported by this paper. Instead this paper focuses on only one coastal grid box of hundreds across the globe. All references that imply potentially large differences elsewhere need to be removed or a study needs to be undertaken to look at all coastlines.

We apologise for making these claims. There are statements along these lines on 3 separate occasions in the original submission (in the abstract, introduction, and conclusion), and we have toned the wording down in all cases.

Abstract:

> the findings also apply to MAM and SON. ~~Although we focus only on the city of Halifax, our results imply potentially large differences in the results of near-surface CO analysis using the L2 and L3 datasets for other cities that are situated within a coastal L3 gridbox, among which are some of the most populous in the world.~~ The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface CO.

Introduction:

> likely to be targets for analyses of temporal trends in air quality indicators. The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface CO~~are therefore relevant to each of these cities~~.

Conclusion:

> in order to maximize the information content of L3 data in coastal gridboxes. Although our study has only focused on the city of Halifax, the results suggest that similar studies be performed for other coastal L3 gridboxes before using MOPITT surface CO, ~~are of relevance to all coastal L3 gridboxes, which~~ since these contain 6 of the top 10 and 43 of the top 100 agglomerations by population and are therefore likely targets for analysis of temporal changes in air pollution indicators such as CO, especially near to the surface. The

G3: MOPITT version 8 has been available for a year, but version 7 was used. While there is nothing wrong with using an older version, V8 should at least be mentioned with a note that the difference may or may not remain in V8. Further, when comparing land and water soundings the TIR product is probably a better choice than TIR-NIR as NIR bands are only used over land. Retrievals from other seasons should also be considered for completeness, even if not examined in as much detail.

There are 3 components to this comment, which we address separately:

MOPITT v8: Unfortunately, this product was released after our analysis was completed. Thank you for accepting that work with the previous product is still valid. We have included references to v8 in a couple of locations where it is relevant, e.g, in Section 2.1.1:

> release of new product versions, with enhanced validation statistics against in situ CO observations. The work presented in this manuscript is based on MOPITT Version 7 (V7) products (Deeter et al., 2017). We analyse both L2 and L3 products (as outlined below). It should be noted that MOPITT Version 8 products have been released very recently, incorporating an improved radiance bias correction method to address a documented drift and geographical variability in retrieval bias compared to in-situ measurements (Deeter et al., 2019). It remains to be seen whether the impacts of land-water retrieval sensitivity contrasts documented in this study remain in this newest product version.

TIR-only product: Our justification for choosing the joint TIR-NIR product is the greater LT sensitivity it offers compared to TIR-only, as stated by several references that we cite. The fact that NIR bands are only used over land contributes to the land-water LT sensitivity contrast that then creates problems when L2 land and water retrievals are averaged together, as is the case for coastal L3 gridboxes. This is the main point we make in the paper. We have performed a supplementary analysis using the TIR-only product and find that the land-water contrast remains, albeit more weakly. This supports our claims made in Section 3.1.2 that a key source of difference in retrieval sensitivity in the lower troposphere is thermal contrast differences over land and water, as supported by previous studies (e.g. Deeter et al. 2007). We now mention the TIR-only product in our Data and Methods section (Section 2.1.1), and that we expect our analysis with the joint TIR-NIR product to represent an upper-limit on the land-water retrieval differences owing to the inclusion of the NIR data (see screenshot from track-changes document below). We include the analysis with TIR-only data in our Supplementary Material (SM1), and also refer to it in Section 3.1.2 when explaining the cause of the land-water sensitivity contrast, pointing out that while thermal contrast in the LT plays an important role, the NIR contribution being limited to retrievals over land will also have an effect (although we also note that a sensitivity contrast exists even in the TIR-only data).

> lower troposphere (LT), to which TIR radiances are typically less sensitive (Pan et al., 1995, 1998). NIR radiances can, however, only be exploited in daytime scenes over land. Our results are based on analysis of the TIR-NIR combined product, owing to its greater sensitivity to LT CO compared to the TIR- and NIR-only products which are also available (e.g. Deeter et al., 2017). Owing to the increased LT sensitivity from NIR radiances being limited to retrievals over land, we expect that the results presented here show an upper bound on the retrieval differences between surface types within our coastal L3 gridbox of focus, and the consequent effects on sample statistics and temporal trends that we outline. Differences are still found in the
>
> 4
>
> TIR-only product, however, and we outline these in Supplementary Material 1. We restrict our analysis to

Retrievals from other seasons: Thank you for raising this. We have now included a sub-section that briefly discusses the MAM and SON seasons at the end of the results and discussion section (“Section 3.3. Consideration of MAM and SON”), which ultimately demonstrates that the findings for DJF and JJA hold for these seasons too. We show mean averaging kernels to demonstrate that in MAM there are strong sensitivity differences in the lower troposphere (similar to JJA), while in SON the sensitivity contrast is much smaller (similar to DJF). We then demonstrate how the temporal changes detected in surface level VMRs in the different datasets studied are affected by these differences, with the datasets based on retrievals over water (what we call L3W and the original L3 dataset) underestimating the change in MAM (as in JJA), whereas there is very little difference between datasets in SON (like in DJF).

G4: A major conclusion seems to be that averaging kernels need to be accounted for. When they are, the differences in retrieved amounts over land compared to water decrease significantly. The need to account for averaging kernels has already been known in the remote sensing community for decades.

This is true. In our paper we demonstrate a practical consequence of this, for a widely-used data product. The point that we make in our paper is, essentially, that the information content of 1^{o} x 1^{o} Level 3 gridded MOPITT products in a coastal location is significantly affected by the fact that the averaging kernels are different for the retrievals over land and water from which it is made. We show that this has significant consequences for the results of temporal trend analysis with the data – the type of analysis which the MOPITT dataset is well-suited to owing to its long timespan. We therefore strongly believe that this is worthy of communication to the scientific community.

G5: A recommendation in this paper is that all soundings over water should be discarded. This would represent a significant loss of information. The argument is that individual land soundings have greater information content (which is not surprising as land soundings use both TIR and NIR, when water soundings can only use TIR). If this argument were extrapolated further, one might say to only use soundings with degrees of freedom of signal (DFS) of say 1.8 or larger. While this would also maximize information content of individual soundings, it would likely decrease the information content from the MOPITT record of the Earth system as a whole. An atmospheric model is needed to substantiate the advice of discarding all retrievals over water.

The point that discarding retrievals = a loss of information content about the Earth system as a whole (which we interpret as meaning e.g. a loss of spatial and/or temporal coverage) is a fair one. We have now made sure to mention this when talking about different guidelines to filter the MOPITT product, as it is a practical consequence that is important to bear in mind (see screenshot below). However, we are not alone in recommending that certain retrievals are discarded (we cite multiple references to the MOPITT team making such recommendations), and we feel that our recommendations are simply a logical extension of this. On the recommendation that additional data (i.e. from an atmospheric model) are needed to substantiate the advice of discarding all retrievals over water: we agree that a study about the effects of discarding certain retrievals on overall information content about the Earth system as a whole would be highly interesting and valuable. However, this is beyond the scope of the work that we present, which we hope is an acceptable contribution to the scientific literature in itself.

Unfortunately, such filtering does lead to an overall loss of available retrievals for analysis, reducing the effective temporal and spatial coverage of the data.

2.1.3. Study area, time period, and MOPITT data processing in this study

G6: P-values are used throughout, but their implications often need more consideration. A small p-value may indicate a statistical difference in the mean, but does not say why the difference appeared. In this study it seems the difference can mostly be accounted for by differences in averaging kernels (which are already known to be important). When 93% of days do not have the right data, it makes it difficult to draw conclusions.

We think that there are two issues raised by this comment, which we address separately:

The use of p-values: We use p-values to point out where mean differences are statistically significant; or where trends identified in regression lines are either significantly different from zero, or significantly different from one-another. Where p-values are not significant, we do not say much about the data, since we can't be sure that differences aren't due to chance/sampling. It is our understanding that use of p-values in this way is common practise? We do not use p-values to say *why* differences appeared – we explain possible causes for the significant differences once they have been identified (i.e. linking significant differences in retrieved surface level VMRs over land and water to averaging kernel differences).

The issue of missing data: the high proportion of missing data, especially in DJF, is frustrating, but also unavoidable when using these data. It certainly makes it difficult to draw conclusions about the time period covered *as a whole*, but it does not preclude the making of conclusions about the temporal subset of data being studied – especially where differences are found to be statistically significant. We now make sure to point out that due to the high proportion of missing data, our conclusions should not be taken as being representative of the time period as a whole. We do this in the Data and Methods section (2.1.3) and also re-iterate it in Section 3.2.2 when discussing the temporal trends we identify.

e.g. from Section 2.1.3:

that were made during daytime hours. This yields a timeseries with one observation per day, when retrieval data were available within this gridbox. There are no retrievals available on 91 % of all days in DJF and 83 % of all days in JJA for the period covered. This is a result of both 1) MOPITT's polar orbit limiting temporal resolution to ~3 days over most of the globe; and 2) on days when the satellite's swath does encompass Halifax, retrievals either not being made due to cloud coverage, or discarded due to data quality issues. While this does not prevent a meaningful comparison of available retrievals, it does mean that caution is needed when using them to draw conclusions about the time period covered as a whole, which is something that we do not attempt to do. ~~(there are no retrievals available on 93 % of all days in DJF and 83 % of all days in JJA~~

Specific comments

P2L41: CO is the only target gas from MOPITT (CH4 cannot and will not be retrieved from the observations).

We have removed the mention of "primary target gas":

the composition of the Earth's atmosphere from space. The ~~primary~~ target gas for MOPITT is carbon monoxide (CO), which ~~-~~ is ~~E~~emitted from a range of anthropogenic (e.g. fossil fuel use) and natural (e.g.

P2L43: A reference is needed for CO lifetime.

We have included a reference to Duncan et al. 2007 (JGR: "Global budget of CO, 1988–1997: Source estimates and validation with a global model")

monoxide (CO), which ~~-~~ is ~~E~~emitted from a range of anthropogenic (e.g. fossil fuel use) and natural (e.g. wildfires) sources, produced via the oxidation of methane and other volatile organic compounds, and has~~with~~ an atmospheric lifetime of weeks to months depending on season and location (e.g. Duncan et al., 2007).~~,~~

P2L43: Volatile organic compounds contribute (indirectly) to about half of CO in Earth's atmosphere.

We now include VOCs in the list of CO sources mentioned (see screenshot above).

P2L63: What is meant by "information content" here? DFS? Shannon information content? (They are related, so maybe this is meant to be generic?)

We mean DFS. Since information content is discussed in detail in the next paragraph, we do not feel that further clarification is needed at this stage.

P5L130: Include a reference to the a priori.

We are not entirely sure what the reviewer is referring to here. The source of the a priori is mentioned in the preceding sentence (CAM-Chem CTM) along with a reference (Lamarque et al. 2012). Perhaps the reviewer is pointing out that we forgot to mention that the retrieval also requires a priori information for surface temperature and emissivity? We have now mentioned this in the text.

P5L133: Watch your usage of "layers" versus "levels" here and throughout. Retrievals are on layers, but are reported on levels for MOPITT. Note there are only 8 layers from 900 to 100 hPa (when surface pressure is greater than 900 hPa). The uppermost layer is 100 to 50 hPa.

We have modified the text so as to only refer to 'levels' (or 'levels of the profile', where elaboration is necessary for clarity) throughout, when talking about MOPITT profile levels. Thank you for pointing out the missed details about the number of layers and the 50 hPa cap to the uppermost layer – we have clarified this in the text.

P5L141: Equation 1 is missing the error term.

We have used the form of the equation that is used in e.g. Deeter et al. 2017 (we have now referenced this accordingly when introducing Equation 1). We already outline in the preceding sentences that "The AK matrix…depends on the radiance weighting functions, instrument error covariance matrix, and a priori covariance matrix", and feel that introducing the error term here would add further complication to an already quite dense paper.

P6L163: "Is generally advised against" – please provide a reference.

The statement is made in the V6 data quality summary[1], in the "Data Filtering" paragraph on Page 2. However, since the data quality summary is not peer-reviewed, and no reference is given therein to support the statement, we have decided to remove the "is generally advised against" statement and stick to what is factual, i.e. that averaging together retrievals with different sensitivity profiles *will* dilute information coming from MOPITT radiances with information coming from the a priori. This then leads on to the next section about data filtering guidelines.

The averaging together of retrievals with significantly different sensitivity profiles – as could be the case when averaging retrievals over land and water – ~~is generally advised against, as this~~ serves to dilute the information coming from the MOPITT observed radiances with information coming from the a priori, thus

P6L165: These supposed guidelines need to be clarified. While such restrictions may increase the average information content of individual soundings, the restrictions may decrease the information content of the system as a whole.

This is a fair and often overlooked point. We have included a statement at the end of the paragraph in Section 2.1.2 that outlines these guidelines to point this out (see screenshot in response to your comment G5 above). We feel that saying any more is beyond the scope of this study, since it is a trade-off that obviously depends on the purposes for which the data are being used. This point has been discussed further in response to the G5 comment above.

P6L173: Maybe this many significant figures are what are reported in the census, but it seems like too many. What if someone moves to Halifax?

[1] https://eosweb.larc.nasa.gov/sites/default/files/project/mopitt/quality_summaries/mopitt_level3_ver6.pdf

We have re-stated this as "with a population in excess of 315,000…":

> longitude = -63.58º~~3~~, latitude = 44.65º~~1~~). Situated on the Atlantic coastline, Halifax is the major economic center in Atlantic Canada, with a population in excess of 315,000~~16,701~~ in the urban core of Halifax Harbour

P6L174: Briefly describe the pollutants.

We have included a list of the trace gases monitored in the referenced study.

> (from 2016 census statistics). The pollution environment of Halifax, which is an intermediate port city, was characterized in detail by Wiacek et al., (2018 – trace gases of focus include SO_x, CO_2, CO, NO_x, O_3, HC, and PM); briefly, it showed no exceedances of regulated gaseous contaminants, but nevertheless a substantial contribution of shipping emissions that is comparable to or greater than emissions from the city's vehicle fleet and a nearby 500 MW power plant. All available MOPITT V7 L2 and L3 TIR-NIR files ("MOP02J"

P6L177: It appears that 2 more years of MOPITT V7 data are now available.

This was not the case at the time the bulk of the research was undertaken, and we feel that the length of the data record included in the study is enough to support the conclusions made.

P7L205: A brief explanation about the 4 MOPITT pixels is needed here.

This has been done.

> to the L3 product that they create, we filter these based on pixel number (each pixel corresponds to one of MOPITT's four along-track detectors) and channel-average signal-to-noise ratio (SNR), as is done at the V7 L3 processing stage to improve L3 information content by excluding observations from specific detector elements on MOPITT's detector array that were found to exhibit greater retrieval noise than the other

P8L223: I dislike the use of "true" for model values. Please modify throughout.

We have replaced all instances of $X_{true,sim}$ with $X_{tr,sim}$. We experimented with using different words altogether (i.e. $X_{model,sim}$) but found it to be less logically consistent with Eq 1 (when the 'real' true profile is referred to and to which we compare $X_{tr,sim}$ in the discussion of Section 3.1.3), and also the use of 'model' created confusion with what are actually the simulated profiles (i.e. X_{sim})! We hope that this compromise is agreeable.

P13: Watch significant figures throughout.

We have taken care to limit values presented to a relevant/sensible number of significant figures.

Table 3: Is the purpose of p-values here to show that CO levels are changing in the MOPITT record?

The p-values cited in this table quantify the probability that the trend is zero (i.e. whether the trend can be considered statistically significant or not), as outlined in the table's caption.

> **Table 3** Results from W~~O~~LS regression analysis of seasonal mean L3W, L3L, L3O, $L3O_{(water)}$ and $L3O_{(mixed)}$ timeseries ~~$L3_O$, $L3_L$ and $L3_W$ timeseries~~ for selected profile levels in DJF and JJA. Trend corresponds to the gradient of the WLS best-fit line~~Units for TCO are mol cm^{-2}, all other levels are ppbv. m = gradient of OLS best-fit line~~; SE = standard error of trend ~~gradient~~; ~~p~~ P-value = probability that the trend~~gradient~~ is zero; % change y^{-1}= mean percentage change in retrieved CO per year, calculated from ~~O~~WLS regression model

Briefly describe why OLS was used.

In this section we now use WLS (on the recommendation from a different reviewer) since it is less susceptible to outliers. We start the section by saying "to identify and compare temporal trends…" We are unsure what further description the reviewer would like?

Figure 6: "Next page" (?)

Thanks for spotting this – it is erroneous and has been removed.

Figure 8: Are the bounding boxes shown correct for a 1 degree box?

Correct – now explained in the caption to Figure 8. Thanks for pointing this out.

Anonymous Referee #2

Received and published: 26 January 2020

Ashpole and Wiacek analyze a MOPITT CO Level 3 (L3) pixel over one location – Halifax, Canada – and compare it to the Level 2 (L2) retrievals within the same 1 degree pixel. They use this coastal location to highlight instrument sensitivity differences between water and land that impact near-surface and profile analysis with the joint TIR- NIR product. The influence of different surface-types (land or water) on CO retrievals and the resulting trends is investigated. The authors find that sensitivity differences account for retrieval differences in JJA, but a CO gradient is likely the reason for differences in DJF. While the MOPITT team already provide recommendations to maximize information content for a studied region, Ashpole and Wiacek demonstrate the practical implications of retrieval differences. The study suggests L2 profile data over land is more appropriate than L3 profile for the small region around the coastal city of Halifax, particularly for the lower troposphere. The authors present a valuable supplementary guide for users of the MOPITT product.

Overall the analysis shows attention to detail and generally good statistical practices. The manuscript is well written and language is clear with very few technical corrections. I have some recommendations for reducing verbosity. Additionally, I recommend the following major comments be addressed before the manuscript is considered for publication.

Major comments:

1. Section 3.2: In my opinion, the most logically inconsistent part of the manuscript is the comparison of all L3O pixels with L3L and L3W . The surface pixel type (land, water or mixed) is given in the MOPITT level 3 product, and users should use information to filter data over such locations. I think the strength of this analysis would be to show that the improvement in information content if you use the L2 land data over coastal locations. Also - is there a difference between using L2 water and the L3 water retrievals, i.e. the extra days you gain from the “mixed" L3 pixel when creating L3W. The authors need to clearly justify why they are comparing with the combination of L3 pixel-types, or alternatively compare against the water-only pixels from Level 3.

We thank the reviewer for raising this, and apologise for not considering it in the original submission. We have completely re-written Section 3.2, and feel that the analysis benefits from now considering the subsets of L3 data separately. Our main findings are unaffected (if anything they are strengthened): even the days classified as ‘mixed’ in the original L3 product (which we refer to in the text as $L3O_{(mixed)}$), which represents the best option in the case of surface level retrievals in JJA owing to the lack of days where the original L3 product is classified as ‘land’, are characterised by markedly lower variability in retrieved surface level VMRs than in the L2 land data (which we call L3L), and the temporal trend identified remains a significant underestimation. We also evaluate the difference between using the L2 water data (which we call L3W), and the L3 days when the surface index is classified as ‘water’ ($L3O_{(water)}$), pointing out that, yes, temporal coverage is greater in L3W due to data effectively being reclaimed from days when the L3 surface index is ‘mixed’, but that this appears to make little significant difference when it comes to temporal trend analysis.

2. Trend analysis: The MOPITT profile layers are known to have drift (Deeter et al., 2017, doi:10.5194/amt-2017-71). Surface drift is about -0.7% per year, 800 hPa is -1% per year. The UT levels have positive drift - both the 400 and 200 hPa levels drift at greater than 1% per year. Drift should be corrected before trend analysis is performed. I did not see this mentioned in the manuscript.

We now take care to mention the drift in profile layers in Section 3.2.2 (see screenshot below). We do not correct our data for drift, which we feel is justified since we are simply comparing trends identified in the same dataset, as opposed to doing a detailed study about changing CO concentrations over the city of Halifax *per se*, and this is outlined in our discussion. We also include the drift values from Deeter et al. 2017 in our Table 3, for reference – most of the trends we identify appear to exceed the drift, at least for the timeseries that has the greatest sensitivity to the true profile (i.e. either L3L or L3W).

Material 8. It is important to note that MOPITT profile measurements are known to have a drift (Deeter et al., 2017), and this should be corrected for in the data if the focus of analysis is to use them to quantify temporal changes in CO over time. Since the intention of the WLS trend analysis presented here is more illustrative, namely to demonstrate trend differences in the data, we have not corrected for this drift. The results should therefore not be taken out of this context (as well as bias correction, verification against a range of other datasets would be required, especially given the large proportion of missing data). We do however provide the reported drift values in Table 3 for context, which shows that the majority of the trends that we have identified appear to be stronger than the measurement drift (at least for the dataset that has greatest retrieval sensitivity at the respective level of the profile). As noted in Section 2.1.1, the measurement drift

21

has been significantly reduced in the latest version of the MOPITT products to be released (Version 8; Deeter et al., 2019).

3. Ordinary Least Squares analysis is highly susceptible to outliers and end points. End points may in particular be impacting the slope calculations for L3L data in JJA (Figure 10, bottom panel). Instead, I recommend that the trend analysis be performed with weighted least squares (WLS), weighted by the standard deviation, which is much less susceptible to outliers. Additionally, the seasons that only have one day per 3 months can be de-weighted by using sufficiently large standard deviation.

Thank you very much for this suggestion. We have replaced our OLS results with results from WLS analysis. Our main findings are unaffected.

4. Please mention somewhere (for example on P6, L189) why the analysis is statistically valid, even though only 7% of the potential DJF days and 17% of the potential JJA days have measured CO.

This is a fair point. The high proportion of missing data, especially in DJF, is frustrating, but also unavoidable when using these data. It certainly makes it difficult to draw conclusions about the time period covered *as a whole*, but it does not preclude the making of conclusions about the temporal subset of data being studied – especially where differences are found to be statistically significant. We now make sure to point out that due to the high proportion of missing data, our conclusions should not be taken as being representative of the time period as a whole. We do this in the Data and Methods section (2.1.3 – screenshot below) and also re-iterate it in Section 3.2.2 when discussing the temporal trends we identify (see screenshot in response to your comment #2).

> that were made during daytime hours. This yields a timeseries with one observation per day, when retrieval data were available within this gridbox. There are no retrievals available on 91 % of all days in DJF and 83 % of all days in JJA for the period covered. This is a result of both 1) MOPITT's polar orbit limiting temporal resolution to ~3 days over most of the globe; and 2) on days when the satellite's swath does encompass Halifax, retrievals either not being made due to cloud coverage, or discarded due to data quality issues. While this does not prevent a meaningful comparison of available retrievals, it does mean that caution is needed when using them to draw conclusions about the time period covered as a whole, which is something that we do not attempt to do. ~~(there are no retrievals available on 93 % of all days in DJF and 83 % of all days in JJA~~

5. I am uncomfortable with the authors extending this analysis to other regions (e.g. P2, L35; P20, L605 to L607). Before extending these results to other coastal locations, other types of locations need to be analyzed (tropical, temperate). Please add a comment that recommends similar analysis be performed for other coastal sites before using MOPITT surface CO.

We apologise for making these claims, and are appreciative of the suggested alternative wording. There are statements along these lines on 3 separate occasions in the original submission (in the abstract, introduction, and conclusion), and we have toned the wording down in all cases:

Abstract:

> the findings also apply to MAM and SON. ~~Although we focus only on the city of Halifax, our results imply potentially large differences in the results of near-surface CO analysis using the L2 and L3 datasets for other cities that are situated within a coastal L3 gridbox, among which are some of the most populous in the world.~~ The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface CO.

Introduction:

likely to be targets for analyses of temporal trends in air quality indicators. The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface CO~~are therefore relevant to each of these cities~~.

Conclusion:

in order to maximize the information content of L3 data in coastal gridboxes. Although our study has only focused on the city of Halifax, the results suggest that similar studies be performed for other coastal L3 gridboxes before using MOPITT surface CO, ~~are of relevance to all coastal L3 gridboxes, which~~ since these contain 6 of the top 10 and 43 of the top 100 agglomerations by population and are therefore likely targets for analysis of temporal changes in air pollution indicators such as CO, especially near to the surface. The

6. There was a change in MOPITT processing after the cooler failure in May 2001. Although a homogenized record is attempted, there remains a small step-change in data. I suggest to use data only from the latter part of the record starting August 2001, especially for the trend analysis where step changes could have large impact.

Thanks for pointing this out. We have re-done all analysis, only using data after August 2001 as recommended. The main results are unaffected by this change, although there are small differences in the specific details. We outline the reason for data truncation in our Data and Methods section:

fleet and a nearby 500 MW power plant. All available MOPITT V7 L2 and L3 TIR-NIR files ("MOP02J" and "MOP03J" files, respectively) were downloaded from the NASA Earthdata portal (https://search.earthdata.nasa.gov). There is a small inconsistency in the data record before and after an instrumental reconfiguration in 2001 (Drummond et al., 2010); we therefore discard all data prior to this reconfiguration. The remaining data~~is~~ covers the period 2001~~0~~-08~~3~~-25~~03~~ to 2017-03-05. At the time of

7. The authors have missed an opportunity to quantify the improvement in retrievals due to the inclusions of NIR between L3L and L3W . In the JJA comparisons, when the CO gradients and a priori suggest there are no expected differences in the CO.

In response to a comment from another reviewer, we have now conducted additional analysis using the TIR-only dataset, to demonstrate that the land-water contrast is present in that dataset and is therefore not purely a result of additional information gained from the NIR being limited to retrievals over land in the joint TIR-NIR product. This is outlined in Supplementary Material 1. However, we do not feel that we can go as far as quantifying the improvement in retrievals due to the inclusion of NIR between L3L and L3W without significant extra analysis which will distract from the main point of the paper.

We do explicitly address the role of NIR in Section 3.1.2. (Climatology of land-water retrieval sensitivity differences), however. We outline that LT sensitivity enhancements due to the inclusion of NIR is limited to retrievals over land and that this therefore likely has an impact on the LT land-water sensitivity contrast. However, our supplementary analysis with TIR-only data (as shown in Supplementary Material 1) shows that the LT land-water difference in AKs is comparable to what it is in the joint TIR-NIR product used in our study, highlighting that thermal contrast differences must be important.

> for the relatively greater overall true profile sensitivity (indicated by DFS values) in JJA for L3L than in DJF for ~~JJA than in~~ L3W ~~in DJF~~. Since our analysis is conducted using the joint TIR-NIR product, it is important to bear in mind that the benefit of enhanced LT sensitivity due to the incorporation of NIR is limited to retrievals over land, so this will also have an impact on the AK differences presented above. However, a land-water retrieval sensitivity contrast of comparable magnitude to that presented here is also evident in the TIR-only product, reinforcing the primary role of thermal contrast differences (see Supplementary Material 1).

8. Figure 6 and associated discussion. I suggest to split the logXtrue - logXap into different average profiles for land and water because Figure 7 suggests that in JJA both difference would be negative while for DJF, the land difference would be negative but the water would be positive. This would help support the overall argument that DJF sees real CO gradients, while JJA is impacted by sensitivity differences.

Thank you for the suggestion, but in this instance however, we have chosen not to modify the analysis. Our justification is as follows: the analysis of simulated profiles presented in Figure 5 and discussed in Section 3.1.3 is intended only to demonstrate how averaging kernel differences can affect retrieved VMRs. Thus, AKs are the only variable between simulated retrievals for land and water. We feel that by introducing further variables, in the form of different Xtrue and Xapr profiles, the main point about AK control would be lost. Admittedly, this could be an additional analysis component (i.e. a second set of simulated profiles for analysis – perhaps with AKs held constant over land and water?), but we do wonder whether it would distract from the main thread of the paper and whether this is really necessary, especially given comments from a different reviewer about the length of the paper. This is obviously something that we can be flexible about however, if the reviewer feels strongly that this is an angle we should include.

Minor comments:

P1, L9: Add latitude and longitude to location in abstract.

This has been done:

> We compare MOPITT Version 7 (V7) Level 2 (L2) & Level 3 (L3) carbon monoxide (CO) products for the 1° x 1° L3 gridbox containing the coastal city of Halifax, Canada (longitude = -63.58°, latitude = 44.65°),

P1, L18 L30 and elsewhere in manuscript: Change "surface profile level" to just "surface level" - otherwise it should be "surface level of the profile".

Thanks for the suggestion – we have changed most instances in the manuscript to "surface level", apart from a few which we have changed to "surface level of the profile" where the elaboration is useful.

P2, L41: Remove "primary" from "The primary target" because MOPITT only retrieves CO.

This has been done:

> the composition of the Earth's atmosphere from space. The ~~primary~~ target gas for MOPITT is carbon monoxide (CO), which ~~–~~ is ~~E~~emitted from a range of anthropogenic (e.g. fossil fuel use) and natural (e.g.

P2, L46: Add in a sentence about the secondary production of CO from VOCs.

This has been done:

> monoxide (CO), which ~~–~~ is ~~E~~emitted from a range of anthropogenic (e.g. fossil fuel use) and natural (e.g. wildfires) sources, produced via the oxidation of methane and other volatile organic compounds, and has~~with~~ an atmospheric lifetime of weeks to months depending on season and location (e.g. Duncan et al., 2007).~~,~~

P2, L48: "...since launch in December, 1999".

This change has been made:

> et al. (2013) for a comparison of CO trends from four satellite instruments), the unique strength of MOPITT lies in its nearly unbroken record of observations since launch in December 1999. This makes MOPITT data

P4, L114 to 116: It is important to note here that the improved sensitivity occurs only over land for the joint TIR-NIR product. Also, largest difference would consequently be expected between land and water with the joint product, so the study presented in this manuscript is expected to show an upper bound on the differences in retrievals between land-types.

Thanks for pointing this out. We now outline this in the text:

lower troposphere (LT), to which TIR radiances are typically less sensitive (Pan et al., 1995, 1998). NIR radiances can, however, only be exploited in daytime scenes over land. Our results are based on analysis of the TIR-NIR combined product, owing to its greater sensitivity to LT CO compared to the TIR- and NIR-only products which are also available (e.g. Deeter et al., 2017). Owing to the increased LT sensitivity from NIR radiances being limited to retrievals over land, we expect that the results presented here show an upper bound on the retrieval differences between surface types within our coastal L3 gridbox of focus, and the consequent effects on sample statistics and temporal trends that we outline. Differences are still found in the

4

TIR-only product, however, and we outline these in Supplementary Material 1. We restrict our analysis to

P4, L127: The simulation climatology resolution is 1.9◦ × 2.5◦, not 1◦ × 1◦.

Thanks for this clarification – the change has been made:

are derived from a monthly CO climatology for the years 2000–2009, simulated with the Community Atmosphere Model with Chemistry (CAM-chem) chemical transport model at a spatial resolution of 1.9º x 2.51º (Lamarque et al., 2012) and then spatially and temporally interpolated to the time and location of the

P6, L172: Add degree symbols to the latitude and longitude locations.

This has been done:

Our analysis is based on MOPITT retrievals over the city of Halifax in Nova Scotia, Canada (Figure 1: longitude = -63.58º3, latitude = 44.65º1). Situated on the Atlantic coastline, Halifax is the major economic

P9, L274: I think it should be "MT and UT".

Thank you for spotting this! The change has been made:

~~retrievals obtained at the same time within the 1º x 1º L3 gridbox containing Halifax~~. Mean ΔRET values are closer to zero and ~~generally~~ less significant in the MT and ULT (represented by the 600 hPa and 300 hPa

P10, L299: I was confused by “the widest part of each AK" - Did you mean in the X-direction or the Y-direction? Please clarify.

Apologies for the confusion – we have modified this part of the AK description to make it clearer:

> profiles in L3L and L3W analysed in the previous section. Each curve corresponds to a row of the AK matrix and represents the sensitivity of the corresponding level of the retrieved profile to each level of the true CO profile, with ~~the width of each AK giving a measure of the vertical resolution associated with a specific level of the retrieved profile;~~ the widest part of each AK in the x-direction (when a peak is evident) indicatinges the portion of the true profile that the corresponding level of the retrieved profile is most sensitive to. The sum of the elements in each AK row ~~(AK_{rowsum})~~ represents the overall sensitivity of the retrieved

P11, L328: Do you mean differences in the MT?

We apologise that this sentence is unclear. We did mean to refer to the LT here, and were implicitly referencing the fact that the 300hPa AK reaches negative values at the surface. We have reworded this to make the point explicitly and (hopefully) improve clarity:

> level, which is around 3 times lower than that over land. As in DJF, UT AKs are quite similar, except for relatively small differences at the surface, where the 300 hPa AK over land actually indicates negative sensitivity. ~~in the LT.~~ Differences in mean DFS values for retrievals in L3L and L3W are greater in JJA (-

P12, L346 to L347: Reword to be a little clearer, e.g. "The Tskin difference approaching 20◦ K between JJA L3L and DJF L3W likely..."

We have reworded this for clarity:

> inverted temperature profile. The T_{skin} increase approaching 20° K between DJF and JJA likely also accounts for the relatively greater overall true profile sensitivity (indicated by DFS values) in JJA for L3L than in DJF for ~~JJA than in~~ L3W ~~in DJF~~. Since our analysis is conducted using the joint TIR-NIR product, it is important

P12, L370: Reword to clarify, e.g. "Available CAMSRA data covers 2003-2016, so a subset of the available MOPITT data (2000-2017) are considered in simulation experiments."

We have reworded this for clarity:

retrievals are solely a result of differences in AKs. BecauseOwing to available CAMSRA data only coversing the years 2003-2016, as opposed to MOPITT data being available for 2000-2017, only a (large) subset of the retrieval pairings considered in earlier sections (which span the period 2001-2017) are could be simulated.

P13, L377: Remove the "respectively" explanation in brackets because there is only one number.

This has been done.

P13, L378: “...sensitivity to Xtrue,sim in L3L than L3W :..." P15, L463: Change all-caps TRUE.

We have renamed $X_{true,sim}$ to $X_{tr,sim}$ on recommendation of another reviewer, so this no longer seems relevant?

P16, L472 to L473: This is the first time the data is separated in this way (L3L>L3W and L3W >L3L). Mention why this is necessary. Why is it not separated this way in the other sections - for example Figure 2 shows a mean response.

We have included justification for this data separation in the text (see screenshot below). To us, this seemed like the logical way to answer the question of whether circulation differences could be generating the CO gradients that MOPITT is detecting. In the other sections it is the mean of the L3L and L3W datasets that is being compared, and shows clear differences.

We compare composite mean wind patterns across Nova Scotia using ERA-Interim data for days when retrieved surface level VMRs in L3W are greater than in L3L (L3W > L3L) and days when they are less (L3W < L3L), since a clear shift in wind direction on these days would support the case that atmospheric transport plays a role in generating differences in retrieved CO amounts over land and water. to illustrate the likely atmospheric transport direction of CO (and other atmospheric constituents) in the region. These are

P16, L491 to L496: This description needs a little clarification. Could it be explained that the density of days where L3W <L3L is higher at the beginning of the record, while the density of days where L3W >L3L is higher at the end of the record for JJA. In DJF, the density of L3W >L3L is consistent across the record.

Thank you for pointing out the lack of clarity here. On closer inspection, what we had previously written was not entirely accurate, so we have reworded accordingly, with your suggestion in mind. Our main point still holds however, that days when L3W < L3L are concentrated at the start of the timeseries in JJA.

While there is no obvious circulation difference at the surface in JJA between days when L3W > L3L and L3W < L3L, there is a difference in evidence that these days actually cover different parts of the distribution of these days throughout the analysed MOPITT timeseries, which spans 167 JJA seasons from 20010 to 2016. 16 of the 19 dDays when L3W < L3L occurtend to fall towards the start in the first half of the timeseries (i.e. before 2009)mean year in which the retrieval was performed = 2004), whereas days when L3W > L3L are spread tend to more evenly throughout (33 of the 65 days (51%) occur before 2009)be later (mean retrieval year = 2008). In DJF there is no such difference, with roughly 50% of days occurring before and after 2009 in each case. This is something we explore further in Section 3.2.2.The same is not true in DJF (mean retrieval years = 2008 and 2007 respectively). As we show in Section 3.2.2, this appears to be a function of true LT CO concentrations over Halifax decreasing across the time period studied, which is detectable in JJA by the retrievals in $L3_L$ owing to their greater near-surface sensitivity, but not by the retrievals in $L3_W$, which are tied to an a priori value that is too low at the start of the timeseries and too high at the end.

P17, Section 3.2.1: Discuss/show p-values for all comparisons between means.

We have rewritten this section entirely, and mention where the mean differences are statistically significant.

P18, L549: Consider discussing as “gradients for L3W " etc. rather than introducing more acronyms. This would also help when reading Table 3 as the new acronyms are not used there.

We have rewritten this section entirely, and remove the acronyms, instead discussing “trends in L3W” etc.

P20, L604: “... before analysis of profile values in order to...".

This change has been made:

4. Conclusions.

Users of MOPITT products are advised to filter the data before analysis of profile values in order to maximize the influence of satellite measurements and minimize the impact of a priori CO concentrations on results (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015). In particular, it is advised that retrievals

P20, L605: “In particular, when investigating coastal cities, it is advised that retrievals over land..."

In the line that the reviewer refers to here, we mention the previously-advised filtering criteria from the MOPITT Algorithm Development Team (2017) and Deeter et al. (2015) (references to these have been added). Coastal cities are not mentioned in this advice. So the suggestion is invalid in this case.

Table 3: Add units for m and SE.

This has been done.

Figure 4: Add error bars for standard deviation.

This has been done.

Figure 8: “Yearmean" is an unusual metric. I suggest “Center year".

We have removed this from the figure and no longer discuss Yearmean, instead quantifying the number of days for each sample in the first and second halves of the timeseries.

Figure 9: What are the error bars?

(Note that this is now Figure 8) These are boxplots, so the errorbars correspond to the range of values covered between the minimum value (excluding outliers) and q1; and between q3 and the maximum value (excluding outliers). We have stated in the caption that these are boxplots to avoid confusion.

Suggestions to reduce verbosity:

Thank you for these!

P5, L155: Remove end of sentence from “in the case of" onwards - redundant.

This change has been made.

each L2 retrieval is tagged according to whether it was performed over land, water, or a combination of the two ("mixed") ~~in the case of 22 x 22 km L2 footprints that straddle the coastline or contain significant bodies of water~~. The surface index of each L3 gridbox is then based on the L2 retrievals that fall within the relevant

P8, L227-L228: Remove "For simulated..." onwards - redundant.

This change has been made.

the mean of the a priori fields that correspond to the temporally coincident L3L and L3W pairings ~~(APR_{L3-L} and APR_{L3-W}, respectively)~~; and A is the retrieval averaging kernel from L3L or L3W. ~~For simulated $L3_L$ ($L3_W$), the averaging kernel from $L3_L$ (AK_{L3-L}) ($L3_W$; AK_{L3-W}) is used.~~ Thus, any differences between each pair of simulated retrievals ($X_{sim,L3-L}$ and $X_{sim,L3-W}$) are solely a result of differences in A, since $X_{true,sim}$ and $X_{apr,sim}$ are identical for both. Simulations are initially performed on log_{10}(VMR) for consistency with the

P12, L348 to L361: This is a repeat of earlier and later information. Consider whether any or all of it is really needed here.

We have removed a lot of the speculation and forward referencing from this section, but keep some of the summarising as we feel it is useful to pull the key results together at this stage before digging deeper.

That the retrieval sensitivity contrast between L3L and L3W is most pronounced in the LT is consistent with the finding that retrieved CO profiles in L3L and L3W show the greatest ~~contrast~~ differences in the LT ~~(Section 3.1.1, Figure 2). In DJF, the retrieval is more sensitive to the true profile in the LT in $L3_W$ (Figure 3, top row) and retrieved LT VMRs are greater in $L3_W$ than $L3_L$. In JJA, the retrieval is more sensitive to the true profile in the LT in $L3_L$ (Figure 3, bottom row), and retrieved LT VMRs are less in $L3_L$ than $L3_W$. This implies that in DJF, LT CO retrievals in $L3_L$ are weighted more heavily towards an a priori profile in which LT VMRs are lower than in the true CO profile; while in JJA, LT retrievals in $L3_W$ are weighted more heavily towards an a priori profile in which VMRs are greater than in the true profile. We explore this point in detail in Section 3.2.2~~. Although mean retrieval sensitivity in L3L and L3W converges with altitude, differences do exist from day-to-day but they are neither as large, nor as skewed in favour of retrievals over land or water (depending on season) as in the LT (see Supplementary Material 2), where there is a well understood thermal contrast mechanism creating the systematic land-water sensitivity contrast ~~(see Supplementary Material 1)~~. Likely causes for this could be day-to-day changes in atmospheric conditions (i.e. temperature or water vapour profiles), or random instrumental or retrieval noise.

P12, L366 to L369: Remove sentence starting "For each L3L-L3W ..." because repeats information from the methods section and is not needed here. Can just start sentence on line 369 with "Recall that any differences..."

This change has been made.

> outlined in Section 2.2. ~~For each $L3_L$- $L3_W$ simulated retrieval pairing ($X_{sim,L3\text{-}L}$ and $X_{sim,L3\text{-}W}$), the true and a priori profiles are common to both (recall the negligible mean difference between $L3_L$ and $L3_W$ a priori profiles in Figure 2), but the retrieval AKs from $L3_L$ and $L3_W$ are used (i.e. for simulated retrievals in $L3_L$ ($L3_W$), the AK from $L3_L$ ($L3_W$) is used). Thus,~~ Recall that any differences between each pair of simulated retrievals are solely a result of differences in AKs. Because~~Owing to~~ available CAMSRA data only covers~~ing~~

P13, L382 to L383: Ending of sentence starting "as would be expected from..." can be removed because it is a repeat of L378 to L380.

This change has been made.

> much closer to $X_{true,sim,sfc}$ (32.21 ppbv higher vs 57.13 ppbv higher). Both ~~Both~~ $X_{sim,L3L,sfc}$ and $X_{sim,L3W}$ ~~values~~ indicate that $X_{true,sim,sfc}$ is lower than $X_{apr,sim,sfc}$ at the surface, but $X_{sim,L3\text{-}L,sfc}$ gives the closer estimate, as would be expected ~~from the greater $AK_{diag,sfc}$ and $AK_{rowsum,sfc}$ values from $L3_L$~~. In both cases, a portion of the

P13, L384 to L391: Unsure that this adds any important points to the discussion, consider removing or shortening.

We have removed a lot of the explanation here as we agree that it does not add anything important to the discussion – we simply mention the interlevel correlation of the retrieval, which it is important to be aware of.

> would be expected ~~from the greater $AK_{diag,sfc}$ and $AK_{rowsum,sfc}$ values from $L3_L$~~. In both cases, a portion of the overall ~~$X_{sim,sfc}$~~ departure~~s~~ from $X_{apr,sim,sfc}$ at the surface level (in other words, "value added" over the a priori) originates at other ~~profile~~ levels of the profile, and not the surface level itself. This is a result of the surface level AK_{sfc} being nonzero at other ~~profile~~ levels, and is a function of the inter-level correlation of the original retrieval, which is linked to the a priori covariance matrix used in the retrieval (Deeter et al., 2010). ~~Because AK_{sfc} has a well-defined peak at the surface $L3_L$ (Figure 5), while AK_{sfc} over $L3_W$ actually indicates greatest retrieval sensitivity (albeit very weak) to a broad region from 900 – 400 hPa, the surface accounts for a greater proportion of the $X_{sim,L3\text{-}L,sfc}$ departure from $X_{apr,sim,sfc}$ (55 % vs 25 % – see Appendix B for full derivation of these values) while the $X_{sim,L3\text{-}W,sfc}$ departure from $X_{apr,sim,sfc}$ has a greater contribution from the lower-to-mid troposphere as a whole.~~

P16, L496: "As we show in Section 3.2.2,..." - Is this really needed here, or could you leave it till Section 3.2.2.

We have removed the lengthy explanation and now just state that "this is something we explore further in Section 3.2.2".

> ~~retrieval year = 2008)~~. In DJF there is no such difference, with roughly 50% of days occurring before and after 2009 in each case. This is something we explore further in Section 3.2.2.~~The same is not true in DJF (mean retrieval years = 2008 and 2007 respectively). As we show in Section 3.2.2, this appears to be a function of true LT CO concentrations over Halifax decreasing across the time period studied, which is detectable in JJA by the retrievals in $L3_L$ owing to their greater near-surface sensitivity, but not by the retrievals in $L3_W$, which are tied to an a priori value that is too low at the start of the timeseries and too high at the end.~~

P17, L510 to L511: Remove end of sentence beginning "primarily because..." - repeat of information in the beginning of the sentence.

Section 3.2 (which this comment refers to) has been entirely re-written.

P17, L511 to L514: Remove "Thus, there is..." - it is unnecessary.

Section 3.2 (which this comment refers to) has been entirely re-written.

P17, L524 to L526: Consider removing "(mean surface level..." to "...248 for L3O)".

Section 3.2 (which this comment refers to) has been entirely re-written.

P19, L581 to L587: There is no significant difference between trends at 300 hPa , so there is no need to explain why they might be different. This paragraph just needs to mention the main point that trends at 300 hPa are positive.

Section 3.2 (which this comment refers to) has been entirely re-written, and we take care to limit our discussion to significant results.

Table 2: AKsfc are in Figure 5 so can be removed. Consider adding other measures from Table 2 to Figure 5 and removing Table 2.

We have removed Figure 5 since the AK values are in Table 2. We choose to keep Table 2 as opposed to Figure 5 as it seems like the better way to present the data in this case.

The track-changes article file begins on the next page…

Impact of land-water sensitivity contrast on MOPITT retrievals and trends over a coastal city

Ian Ashpole[1] and Aldona Wiacek[1,2]

[1]Department of Environmental Science, Saint Mary's University, Halifax, Canada
[2]Department of Astronomy and Physics, Saint Mary's University, Halifax, Canada

Correspondence to: Ian Ashpole (ian.ashpole@smu.ca)

Abstract.

We compare MOPITT Version 7 (V7) Level 2 (L2) & Level 3 (L3) carbon monoxide (CO) products for the 1° x 1° L3 gridbox containing the coastal city of Halifax, Canada (longitude = -63.58°, latitude = 44.65°), with a focus on~~for~~ the seasons DJF and JJA, and highlight a limitation in the L3 products that has significant consequences for the temporal trends in near-surface CO identified using those data. Because this gridbox straddles the coastline, the MOPITT L3 products are created from the finer spatial resolution L2 products that are retrieved over both land and water, with a greater contribution from retrievals over water because more of the gridbox lies over water than land. We create alternative L3 products for this gridbox by separately averaging the bounded L2 retrievals over land (L3L) and water (L3W) and demonstrate that profile and total column CO (TCO) concentrations, retrieved at the same time, differ depending on whether the retrieval took place over land or water. These differences (ΔRET) are greatest~~, and most significant,~~ in the lower troposphere (LT), where mean retrieved VMRs are greater in L3W than L3L, with maximum mean differences of ~~11.~~11.6 % (14.3 ppbv, $p = 0.001$)~~4 % (14.9 ppbv, p = 0.116) at the 900 hPa level in DJF, and 10.8 % (12.4 ppbv, p = 0.005)~~ at the surface ~~profile~~ level in JJA. Retrieved CO concentrations are more similar, on average, in the middle and upper troposphere (MT and UT), although large differences (in excess of 4~~5~~0 %) do infrequently occur. ~~Significant (p < 0.1)~~ TCO is also greater in L3W than L3L~~differences of ~5 % are also found~~ in both seasons. By analyzing L3L and L3W retrieval averaging kernels and simulations of these retrievals, we demonstrate that, in JJA, ΔRET is strongly influenced by differences in retrieval sensitivity over land and water, especially close to the surface where L3L has significantly greater information content than L3W. In DJF, land-water differences in retrieval sensitivity are much less pronounced and appear to have less of an impact on ΔRET, which analysis of wind directions suggests is more likely to reflect differences in true profile concentrations (i.e. "real" differences). The original L3 timeseries for the 1° x 1° gridbox containing Halifax (L3O) corresponds much more closely to L3W than L3L, owing to the greater contribution from L2 retrievals over water than land. Thus, in JJA, variability in retrieved CO concentrations close to the surface in L3O is suppressed compared to L3L, and a~~they~~ declining~~e~~ trend detected using Weighted Least Squares (WLS) regression analysis is~~at a~~ significantly slower in L3O

(strongest surface level trend identifiable = -1.35 (±0.35) ppbv y^{-1}) than L3L (~~rate (surface profile level trends of -1.16 (±0.32) ppbv y^{-1} vs~~ -2~~3~~.~~2~~85 (±0.60~~8~~) ppbv y^{-1}~~, from $L3_O$ and $L3_L$ respectively~~). This is because contributing L2 retrievals over water are closely tied to a priori CO concentrations used in the retrieval, owing to their lack of near-surface sensitivity in JJA, and these are based on monthly climatological CO profiles from a chemical transport model and therefore have no yearly change (surface ~~profile~~ level trend in L3W = -0.60~~51~~ (±0.33~~28~~) ppbv y^{-1} ~~in L3W~~). Although our analysis focuses on DJF and JJA, we demonstrate that the findings also apply to MAM and SON. ~~Although we focus only on the city of Halifax, our results imply potentially large differences in the results of near-surface CO analysis using the L2 and L3 datasets for other cities that are situated within a coastal L3 gridbox, among which are some of the most populous in the world.~~ The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface CO.

1 Introduction

The Measurement of Pollution in the Troposphere (MOPITT – Drummond et al., 2010, 2016; Acronyms defined in Appendix A) instrument is one of a large fleet of satellite-borne instruments capable of observing the composition of the Earth's atmosphere from space. The ~~primary~~ target gas for MOPITT is carbon monoxide (CO), which~~.~~ is ~~E~~emitted from a range of anthropogenic (e.g. fossil fuel use) and natural (e.g. wildfires) sources, produced via the oxidation of methane and other volatile organic compounds, and has~~with~~ an atmospheric lifetime of weeks to months depending on season and location (e.g. Duncan et al., 2007).~~,~~ CO is therefore vital to monitor as a pollutant in its own right, as a tracer of local and transported pollution sources, and also because it plays an important role in atmospheric chemistry, i.e. as a precursor to ozone formation and a primary sink for the hydroxyl radical. While multiple sensors observe CO (see e.g. Worden et al. (2013) for a comparison of CO trends from four satellite instruments), the unique strength of MOPITT lies in its nearly unbroken record of observations since launch in December 1999. This makes MOPITT data very valuable for the analysis of temporal trends in CO concentrations (e.g. He et al., 2013; Worden et al., 2013; Strode et al., 2016).

MOPITT retrieves coarse vertical resolution CO profiles in the troposphere by inverting observed upwelling radiances at thermal infrared (TIR) and near infrared (NIR) wavelengths (Deeter et al., 2013). These profiles are integrated to give CO total column amounts (TCO). In addition to several other inputs, MOPITT's optimal estimation retrieval algorithm requires a priori information – among which is a description of the most probable state of the CO profile and its variability – to obtain physically realistic results (Pan et al., 1998; Rodgers, 2000; the retrieval algorithm is outlined in more detail in Section 2.1.1). The proportion of information about CO concentrations in each individual retrieval that comes directly from the satellite measurement, as opposed to the a priori, is highly variable. It depends on scene-specific factors such as surface temperature, thermal contrast in the lower troposphere, and the actual ("true") CO loading itself, as well as on instrumental noise (e.g. Deeter et al., 2015). This complicates the interpretation of retrievals, thus placing great importance on the analysis of retrieval averaging kernels (AKs), which represent the sensitivity of each retrieved profile point to the true CO profile and quantify the overall information content of the retrieval (as described in detail by, e.g., Deeter et al. (2007, 2015) and Rodgers (2000)). The lower the retrieval information content, the closer the retrieved CO loading will be to the a priori, which is based on a climatological model value. Retrievals with little information content should thus be treated with caution in any analysis.

In general, the greatest information content is associated with daytime retrievals over land, during the summer season (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015). This is where and when thermal contrast conditions are typically greatest, maximizing the instrument's ability to sense CO absorption in the lowermost layers of the troposphere against the hot surface emission background (Deeter et al., 2007; Worden et al., 2010). To ensure that analyses involving MOPITT data are not biased by retrievals that have a heavy reliance on the a priori (in other words, a low information content), it is therefore suggested that users of MOPITT data consider excluding from analysis retrievals obtained during winter months, over water, and also from certain other geographical areas where retrieval information content is known to be low, i.e. over mountainous regions, where the effects of geophysical noise reduce information content relative to flatter terrain (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015). Deeter et al. (2015) specifically emphasizes such filtering in the analysis of long-term CO trends, since inclusion of retrievals with a heavy a priori weighting will weaken any real trends in the data. This occurs because the a priori CO is based on monthly climatologies of modelled CO amounts and is therefore variable by month but not by year (Deeter et al., 2014).

MOPITT data are available as Level 2 (L2) products, where each individual retrieval at 22 x 22 km spatial resolution is available for analysis; and Level 3 (L3) products, which are a 1^{o} x 1^{o} area-averaged version of the individual L2 retrievals that fall within each gridbox (with some filtering criteria applied – see Section 2.1.2). At the heart of this study is the fact that some L3 gridboxes straddle the coastline. L3 products for such gridboxes can therefore be based on L2 retrievals over both land and water (see Figure 1), the information content of which can differ greatly (e.g., Deeter et al., 2007). In this study, we demonstrate, for a coastal L3 gridbox, how well known and well characterized differences in retrieval sensitivity over land and water can lead to significant differences in the L2 retrieved profiles that are averaged together to create the L3 products (Section 3.1). We outline the impact that this has on the statistics of the resulting L3 CO profiles, and demonstrate the consequences that it has for temporal trend analysis with the L3 dataset, when compared to the results of the same analysis applied to the underlying L2 data that can be filtered by surface type to maximize information content (Section 3.2). This is an important issue to be aware of for two reasons: firstly, owing to their smaller file size, L3 data are better suited to long timeseries analysis than L2 data (~25 MB vs ~450 MB respectively, for a single daily, global file). Working with L3 data requires fewer computing resources and, arguably, less technical expertise, making the MOPITT data more readily accessible to a greater number of users who are potentially less well-positioned to scrutinize the data. Secondly, 6 of the top 10 and 43 of the top 100 largest agglomerations by population in the world (population data taken from www.citypopulation.de, valid at time of writing) lie within a coastal L3 gridbox, and it is such cities that are likely to be targets for analyses of temporal trends in air quality indicators. The results that we report here suggest that similar analyses be performed for other coastal cities before using MOPITT surface COare therefore relevant to each of these cities.

This paper is structured as follows: In section 2 we outline the data and methods used in this study, giving an overview of the MOPITT instrument, retrieval, and surface type classification that is relevant to our work. In Section 3 our results are presented and discussed, and conclusions are drawn in Section 4.

2. Data and Methods

2.1 MOPITT

2.1.1 Instrument and retrieval overview

MOPITT has been making routine observations almost continuously since March 2000. It is carried on board the polar-orbiting NASA Terra satellite, with a nominal altitude of ~705 km and an equatorial overpass time of ~10:30 and ~22:30 local time. The instrument is a nadir-viewing gas correlation radiometer, with a ground resolution of 22 x 22 km. It observes radiances in two CO-sensitive spectral bands: the TIR at 4.7 µm, and the NIR at 2.3 µm. The TIR band is sensitive to both absorption and emission by CO and yields information on its vertical distribution in the troposphere (Pan et al., 1995, 1998). The NIR band measures reflected solar radiation, which constrains the CO total column amount and yields information on CO concentrations in the lower troposphere (LT), to which TIR radiances are typically less sensitive (Pan et al., 1995, 1998). NIR radiances can, however, only be exploited in daytime scenes over land. Our results are based on analysis of the TIR-NIR combined product, owing to its greater sensitivity to LT CO compared to the TIR- and NIR-only products which are also available (e.g. Deeter et al., 2017). Owing to the increased LT sensitivity from NIR radiances being limited to retrievals over land, we expect that the results presented here show an upper bound on the retrieval differences between surface types within our coastal L3 gridbox of focus, and the consequent effects on sample statistics and temporal trends that we outline. Differences are still found in the

TIR-only product, however, and we outline these in Supplementary Material 1. We restrict our analysis to daytime-only retrievals (more information on data selection in Section 2.1.3).

Multiple other sources describe MOPITT's CO retrieval algorithm in detail (e.g., Deeter et al., 2003; Francis et al., 2017). Briefly, it employs optimal estimation (Pan et al., 1998; Rogers, 2000) and a fast radiative transfer model (Edwards et al., 1999) to invert radiance measurements performed by the instrument to obtain CO concentrations. Additional inputs required include meteorological data (profiles of temperature and water vapour), surface temperature and emissivity, and satellite viewing geometry for the radiative transfer model; and a priori CO profiles, to constrain the inversion to physically reasonable limits. For latest MOPITT product versions, meteorological fields are extracted from the NASA Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) reanalysis product; and a priori CO profiles are derived from a monthly CO climatology for the years 2000–2009, simulated with the Community Atmosphere Model with Chemistry (CAM-chem) chemical transport model at a spatial resolution of 1.9º x 2.5~~1~~º (Lamarque et al., 2012) and then spatially and temporally interpolated to the time and location of the MOPITT observation. As it is a multi-year climatology, the a priori features no yearly trend, i.e. values for a given location and day of the year are the same every year. Surface temperature and emissivity values are retrieved from the radiance measurements (the retrieval also requires a priori information for these measurements). Retrievals are only performed for cloud-free scenes, with cloud screening based on collocated Moderate Resolution Imaging Spectroradiometer (MODIS) observations and MOPITT's own radiances. CO profiles are retrieved on 10 vertical levels, with 9 equally spaced pressure levels from 900 to 100 hPa (the uppermost level covers the atmospheric layer from 100 to 50 hPa) and a floating surface pressure level. Where the surface pressure is below 900 hPa, only 8 profile levels are retrieved. Reported values represent the mean CO volume mixing ratio (VMR) in the layer immediately above that level. Retrievals are initially performed on a $\log_{10}$(VMR) scale, owing to large CO variability in the atmosphere.

Averaging kernels are produced for each retrieval and distributed with the data. The AK matrix (A) quantifies the the sensitivity of the retrieved vertical profile to the "true" vertical profile, and depends on the radiance weighting functions, instrument error covariance matrix, and a priori covariance matrix. Its relationship to the retrieved profile (X_{rtv}), the "~~'~~true"~~'~~ profile (X_{true}), and the a priori profile (X_{apr}) is expressed as follows (e.g. Deeter et al., 2017):

$$X_{rtv} = X_{apr} + A(X_{true} - X_{apr}) \quad (1)$$

Thorough analysis of AKs is essential for understanding the physical significance of MOPITT's CO retrievals. We discuss AKs in more detail in section 3.1.2.

The MOPITT retrieval algorithm is subject to continuous development, in line with improvements in understanding of the changing instrumental characteristics and geophysical factors that affect the retrieval sensitivity, and with periodic updates to the radiative transfer model (Worden et al., 2014). This prompts the release of new product versions, with enhanced validation statistics against in situ CO observations. The work presented in this manuscript is based on MOPITT Version 7 (V7) products (Deeter et al., 2017). We analyse both L2 and L3 products (as outlined below). It should be noted that MOPITT Version 8 products have been released very recently, incorporating an improved radiance bias correction method to address a documented drift and geographical variability in retrieval bias compared to in-situ measurements (Deeter et al., 2019). It remains to be seen whether the impacts of land-water retrieval sensitivity contrasts documented in this study remain in this newest product version.

2.1.2. Surface type classification

Both L2 and L3 data files come with a range of diagnostic fields and values, in addition to the averaging kernel matrix, that can be used for filtering and interpreting retrievals. Of particular importance is the surface index flag. Because retrieval information content is variable depending on surface type (Deeter et al., 2007), each L2 retrieval is tagged according to whether it was performed over land, water, or a combination of the two ("mixed") in the case of 22 x 22 km L2 footprints that straddle the coastline or contain significant bodies of water. The surface index of each L3 gridbox is then based on the L2 retrievals that fall within the relevant 1° x 1° grid boundaries (Figure 1). Where more than 75 % of the bounded L2 retrievals have the same surface index, only those retrievals are used to produce the L3 gridded value (the other L2 retrievals are discarded) and the L3 surface index is set to that surface type. Otherwise, all L2 retrievals available in the L3 gridbox are averaged together and the L3 surface index is set to "mixed" (this information is taken from the MOPITT Version 6 L3 data quality summary[1] – at the time of writing, no V7 L3 data quality summary was available).

The averaging together of retrievals with significantly different sensitivity profiles – as could be the case when averaging retrievals over land and water – is generally advised against, as this serves to dilute the information coming from the MOPITT observed radiances with information coming from the a priori, thus increasing the dependence of the resulting CO profile values on the a priori profile. In fact, guidelines to maximize the information content of MOPITT data and minimize the influence of the a priori are to restrict analysis to daytime observations over land during the summer season, since this is when thermal contrast conditions are greatest, thus maximizing the instrument's ability to sense CO in the lowermost layers of the troposphere (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015; Deeter et al., 2007).

[1] available here: https://eosweb.larc.nasa.gov/sites/default/files/project/mopitt/quality_summaries/mopitt_level3_ver6.pdf

Unfortunately, such filtering does lead to an overall loss of available retrievals for analysis, reducing the effective temporal and spatial coverage of the data.

2.1.3. Study area, time period, and MOPITT data processing in this study

Our analysis is based on MOPITT retrievals over the city of Halifax in Nova Scotia, Canada (Figure 1: longitude = -63.58º~~3~~, latitude = 44.65º~~1~~). Situated on the Atlantic coastline, Halifax is the major economic center in Atlantic Canada, with a population in excess of 315,000~~16,701~~ in the urban core of Halifax Harbour (from 2016 census statistics). The pollution environment of Halifax, which is an intermediate port city, was characterized in detail by Wiacek et al., (2018 – trace gases of focus include SO_x, CO_2, CO, NO_x, O_3, HC, and PM); briefly, it showed no exceedances of regulated gaseous contaminants, but nevertheless a substantial contribution of shipping emissions that is comparable to or greater than emissions from the city's vehicle fleet and a nearby 500 MW power plant. All available MOPITT V7 L2 and L3 TIR-NIR files ("MOP02J" and "MOP03J" files, respectively) were downloaded from the NASA Earthdata portal (https://search.earthdata.nasa.gov). There is a small inconsistency in the data record before and after an instrumental reconfiguration in 2001 (Drummond et al., 2010); we therefore discard all data prior to this reconfiguration. The remaining data~~is~~ covers the period 2001~~0~~-08~~3~~-25~~03~~ to 2017-03-05. At the time of writing, more recent data are flagged as "beta" files, which await a future retrospective processing after the annual hot calibration becomes available, and their use in scientific analyses is discouraged (Deeter et al., 2017). For clarity and brevity, we restrict our main analyse~~i~~s and discussion to the winter (DJF) and summer (JJA) seasons, since these best encapsulate the different thermal contrast conditions over land and water, when compared to the intermediate (MAM and SON) seasons. For completeness, we demonstrate that our findings also hold for MAM and SON in section 3.3.

We extract L3 data for the 1º x 1º gridbox that contains the city of Halifax, and retain only the observations that were made during daytime hours. This yields a timeseries with one observation per day, when retrieval data were available within this gridbox. There are no retrievals available on 91 % of all days in DJF and 83 % of all days in JJA for the period covered. This is a result of both 1) MOPITT's polar orbit limiting temporal resolution to ~3 days over most of the globe; and 2) on days when the satellite's swath does encompass Halifax, retrievals either not being made due to cloud coverage, or discarded due to data quality issues. While this does not prevent a meaningful comparison of available retrievals, it does mean that caution is needed when using them to draw conclusions about the time period covered as a whole, which is something that we do not attempt to do. ~~(there are no retrievals available on 93 % of all days in DJF and 83 % of all days in JJA for the period covered – this is a result of 1) MOPITT's polar orbit limiting temporal resolution to ~3 days over most of the globe; and 2) retrievals either not being made due to cloud coverage, or discarded due to~~

~~data quality issues, on days when the satellite swath encompasses Halifax).~~ For clarity, we refer to the~~is~~ original, "as-downloaded" L3 timeseries as L3O for the remainder of this paper, owing to the way that we process the L2 data (explained below). Because this gridbox straddles the coastline, the L3O surface index varies each day. The surface classification breakdown of the L3O timeseries is given in Table 1a. "Water" is the modal classification in both seasons, followed by "mixed". L3O is only classified as "land" on one occasion each season. This is most likely due to the fact that more of the L3 gridbox is situated over water than land (Figure 1). The ratio of water to mixed observations is far greater in DJF than in JJA. This may be due to preferential cloud coverage over land in winter, and/or could be linked to the misidentification of snow/ice coverage on the surface as cloud during cloud screening (identifying the exact cause for this difference is beyond the scope of this paper).

We select all L2 retrievals that fall within the 1° x 1° L3 gridbox that contains the city of Halifax (lower-left corner: -64° E, 44° N; upper right corner: -63° E, 45° N). Because we directly compare the L2 retrievals to the L3 product that they create, we filter these based on pixel number (each pixel corresponds to one of MOPITT's four along-track detectors) and channel-average signal-to-noise ratio (SNR), as is done at the V7 L3 processing stage to improve L3 information content by excluding observations from specific detector elements on MOPITT's detector array that were found to exhibit greater retrieval noise than the other elements (MOPITT Algorithm Development Team, 2017; Deeter et al. 2017). Specifically, these filters exclude the following: all observations for Pixel 3; and all observations where both (1) the channel 5A SNR < 1000 and (2) the channel 6A SNR < 400. 5A and 6A correspond to the average radiances for MOPITT's length-modulated cell TIR and NIR channels, respectively. Finally, we only retain daytime retrievals, using a solar zenith angle filter of < 80°.

From this subset of L2 retrievals, we take separate area averages for those with a surface index of land and water, creating two timeseries that are effectively new L3 "land only" and "water only" products, for days when MOPITT retrievals over Halifax are available. We herein refer to these as L3L and L3W, respectively. For clarity of analysis, we discard remaining L2 retrievals with a surface index of mixed (these account for ~5 % of the total L2 retrieval subset). The number of individual L2 retrievals that are averaged together each day to create L3L and L3W is given in Table 1b. From this, it is clear that there are around double~~a greater~~ the number of L2 retrievals over water than land within the L3 gridbox containing Halifax, which explains the dominance of water in the L3O surface classification (Table 1a), and also ~~implies~~ means that L2 retrievals over water will have a greater weighting in L3O than L2 retrievals over land on days when the surface index is mixed.

2.2. Retrieval Simulation

To demonstrate how MOPITT retrieved CO concentrations are affected by retrieval sensitivity (Section 3.1.3), we simulate pairs of L3L and L3W retrieved profiles that are obtained concurrently (i.e. retrieved on the same day – on some days, one of L3L or L3W is missing) as follows:

$$X_{sim} = X_{apr,sim} + A(X_{tr,sim\,\sout{true,sim}} - X_{apr,sim}) \tag{2}$$

For each simulated retrieval: $X_{tr\sout{ue},sim}$ is taken from the Copernicus Atmospheric Monitoring Service (CAMS) reanalysis (CAMSRA – see section 2.3), for the model gridbox that contains Halifax, for the corresponding month and year of the observed retrieval (because the CAMSRA data are monthly mean values); $X_{apr,sim}$ is the mean of the a priori fields that correspond to the temporally coincident L3L and L3W pairings~~ (APR_{L3-L} and APR_{L3-W}, respectively)~~; and A is the retrieval averaging kernel from L3L or L3W.~~ For simulated $L3_L$ ($L3_W$), the averaging kernel from $L3_L$ (AK_{L3-L}) ($L3_W$; AK_{L3-W}) is used.~~ Thus, any differences between each pair of simulated retrievals ($X_{sim,L3\sout{-}L}$ and $X_{sim,L3\sout{-}W}$) are solely a result of differences in A, since $X_{tr\sout{ue},sim}$ and $X_{apr,sim}$ are identical for both. Simulations are initially performed on $\log_{10}$(VMR) for consistency with the MOPITT retrieval algorithm, and then converted back to VMR scale for analysis.

2.3. Additional datasets

The CAMSRA dataset to simulate retrievals is described by Inness et al., (2019). For the CAMSRA gridbox containing Halifax (horizontal resolution = 1° x 1°), we extract CO volume mixing ratios for levels 1000 hPa – 100h Pa at 100h Pa intervals, which correspond to the MOPITT levels of the profile~~ levels~~. The CAMSRA dataset has no "~~'~~surface"~~'~~ level~~profile level~~, so we take the 1000 hPa ~~profile ~~level (the lowest level available in the dataset) to correspond to MOPITT's floating surface level. At the time of writing, CAMSRA data are only available for the years 2003-2016.

Information on mean wind patterns across Nova Scotia and the surrounding area is taken from the European Centre For Medium-Range Weather Forecasts (ECMWF) ERA-Interim dataset (horizontal resolution = 0.75° x 0.75°; see Dee et al., (2011) for a dataset overview). We analyse daily mean u and v vector winds for the following levels: 10-metres (the closest level to the surface for winds in the dataset), 850 hPa, and 500 hPa (which correspond roughly to the lower- and mid- troposphere, respectively). In addition, we extract monthly mean temperature profile data (at 100 hPa intervals, plus the skin temperature and 2-metre air temperature variables) for the closest model gridboxes to Halifax that exclusively cover land and ocean, in order to illustrate the typical "land-only" and "water-only" temperature profiles that correspond to the MOPITT L2 retrievals over land and water that are analyzed.

3. Results and Discussion.

3.1. Impact of retrieval sensitivity differences on temporally coincident L3L and L3W retrievals.

In this section we compare the L3L and L3W CO retrievals and demonstrate where and when there are differences in retrieved CO concentrations that are clearly linked to differences in retrieval sensitivity over land and water. We restrict our analysis to days when the L3O surface index is "mixed" and both L3L and L3W retrievals are present, in order to minimize any potential differences in the true profile between land and water (there are a couple of days in the L3O timeseries when one or the other of L3L or L3W is missing, even when the L3O surface index is "mixed", owing to the presence of L2 retrievals with a surface index of "mixed", which we have discarded). Thus, an underlying assumption here is that land-water differences in the true profile for retrievals contributing to L3L and L3W are small, owing to the fact that they are retrieved in close spatial proximity to eachother (i.e. within the same 1° x 1° gridbox) and at the same time. We test this assumption in Sections 3.1.4 and 3.1.5.

3.1.1. Climatology of land-water retrieval and a priori differences.

Figure 2 shows the percentage difference between temporally coincident retrieved VMRs for selected levels of the profile ~~levels~~ and for CO total column (TCO) amounts (ΔRET) in L3L and L3W. Positive (negative) differences indicate that retrieved VMRs/TCO are greater (less) in L3W than L3L. Differences are expressed as percent values, rather than differences in measurements units, so that we can display profile and TCO retrievals on the same plot (profile units are ppbv, TCO units are molecules cm^{-2}).

In both seasons, mean retrieved VMRs are greater in L3W than L3L in the lower troposphere (LT – surface, 900hPa and 800hPa ~~profile~~ levels), with a maximum mean difference of 11.6 % (14.3 ppbv) at the surface level in JJA ~~in DJF of 11.4 % (14.9 ppbv, p = 0.116) at the 900hPa level, and in JJA of 10.8 % (12.4 ppbv) at the surface level~~, ~~where~~ the only profile location where the mean difference is ~~highly~~ significant (p = 0.001~~5~~). The spread of ΔRET values is comparable in both seasons at these levels ~~at the surface and 900 hPa levels (although larger outliers exist in JJA), and much larger at 800 hPa in JJA than DJF~~, with a clear skew towards positive values. ~~in all three cases. This demonstrates that,~~ Thus, although ~~on any given day~~ retrieved LT VMRs in L3L may occasionally exceed those in ~~and~~ L3W by over 20%, ~~can differ by as much as ~50 % in the LT and especially near to the surface,~~ they are usually greater over water than land ~~in L2 retrievals obtained at the same time within the 1° x 1° L3 gridbox containing Halifax~~. Mean ΔRET values are closer to zero and ~~generally~~ less significant in the MT and U~~L~~T (represented by the 600 hPa and 300 hPa profile levels respectively), indicating that differences in retrieved VMRs are not as persistent at higher altitudes. However, the spread in ΔRET remains large at these altitudes, with retrieved VMRs in L3L and L3W differing by over ± 40 % on individual days (with outliers exceeding ± ~~7~~60 %). TCO is ~~significantly~~ greater in L3W than L3L in both seasons, significantly so in JJA ~~(mean difference of 6.2 % in DJF ($p < 0.1$) and 3.9 % in JJA ($p < 0.06$))~~. This~~,~~ ~~which~~ is consistent with the most persistent VMR differences occurring in the LT, where atmospheric densities are greatest, thus contributing a relatively greater amount to the total column than MT and UT ~~profile~~ levels.

Our assumption in this section is that L2 retrieved CO concentrations obtained within the same 1° x 1° L3 gridbox should be similar. We may actually expect retrieved CO amounts in L3L to be greater than those in L3W due to CO sources existing on land, particularly within the city of Halifax. One reason for the ΔRET values instead indicating higher concentrations over water could be differences in the a priori profiles (ΔAPR) used in the corresponding retrievals. The L2 retrievals over land and water have different a priori profiles owing to spatial interpolation of the 1° x 1° model climatology to the 22 x 22 km footprint of the MOPITT L2 retrieval. However, as Figure 2 demonstrates, ΔAPR values are small in comparison to ΔRET, with mean difference values very close to zero and a maximum range of 10.4~~9.5~~ % (occurring at the surface level in JJA). Moreover, the sign of mean ΔAPR does not match that of mean ΔRET at several levels in both seasons (i.e. retrieved VMRs in L3W are greater than in L3L, but a priori VMRs are less). It therefore appears unlikely that a priori profile differences are responsible for the observed differences in retrieved CO concentrations in L3L and L3W.

3.1.2. Climatology of land-water retrieval sensitivity differences.

An alternative explanation for the observed ΔRET could be differences in retrieval sensitivity over land and water, quantified by the retrieval AK matrix. Figure 3 compares the mean AKs corresponding to the retrieved profiles in L3L and L3W analysed in the previous section. Each curve corresponds to a row of the AK matrix and represents the sensitivity of the corresponding level of the retrieved profile to each level of the true CO profile, with ~~the width of each AK giving a measure of the vertical resolution associated with a specific level of the retrieved profile;~~ the widest part of each AK in the x-direction (when a peak is evident) indicating~~es~~ the portion of the true profile that the corresponding level of the retrieved profile is most sensitive to. The sum of the elements in each AK row ~~(AK_{rowsum})~~ represents the overall sensitivity of the retrieved profile at the corresponding pressure level to the whole ~~the~~ true profile; values close to zero indicate that the retrieval is relatively insensitive to the true profile and therefore closely tied to the a priori profile, while the converse is true as the rowsum approaches one. The mathematical trace of the AK matrix (i.e. sum of the diagonals) gives the degrees of freedom for signal (DFS) of the retrieval, which is a measure of the number of independent pieces of information (in other words, "information content") in the retrieval from the measurement, with respect to the true profile. When DFS values approach two, this is interpreted as the retrieval being able to resolve CO in two independent atmospheric layers.

There are some clear differences between the mean AKs over land (from L3L) and water (from L3W) shown in Figure 3. In DJF, AKs for LT (especially surface level) retrievals reach greater values in the LT in L3W than L3L, indicating that sensitivity to the true profile at these levels is actually greater in L3W (surface level AK peak = ~0.145 for water, ~0.0910 for land, both at the surface level). This is reflected in greater rowsum values for LT AKs in L3W than L3L ~~(0.55 vs 0.44 for surface level AKs, to 0.82 vs 0.74 for 800 hPa level AKs, respectively)~~. Differences in MT (600 hPa) and UT (300 hPa) AKs are much less pronounced, with sensitivity to the true profile actually becoming slightly greater in L3L than L3W higher up in the troposphere. In JJA, the mean LT and MT AKs are qualitatively much more different between L3L and L3W than in DJF. LT rowsums are significantly greater in L3L ($p < 0.001$ at the surface, 900 hPa and 800 hPa profile levels, see Supplementary Material 21) and closer to one, signifying that these retrievals contain much more true profile information than those in L3W ~~(AK$_{rowsums}$ = 0.15 vs 0.70 at the surface level to 0.67 vs 0.94 at 800 hPa, in L3$_W$ and L3$_L$ respectively)~~. The AK shapes also indicate that the respective retrievals are sensitive to different parts of the true profile. In L3L, surface and LT AKs peak at either the surface (surface level AK) or at 900 hPa (both the 900 hPa and 800 hPa level AKs) and decline towards the UT, while MT AKs indicate relatively equal sensitivity throughout the LT, MT and the lower levels of the UT. In L3W on the other hand, LT and MT AKs (excluding the largely insensitive surface level) indicate relatively little sensitivity in the lowest profile levels and peak in the MT. The surface level AK in L3W actually indicates close to zero sensitivity throughout the profile ~~(AK$_{rowsum}$ = 0.15)~~, aside from a very weak peak at the surface level, which is around 3 times lower than that over land. As in DJF, UT AKs are quite similar, except for relatively small differences at the surface, where the 300 hPa AK over land actually indicates negative sensitivity. ~~in the LT.~~ Differences in mean DFS values for retrievals in L3L and L3W are greater in JJA (-0.121) than DJF (0.07), highlighting the greater land-water sensitivity contrast in JJA and~~but~~ also reflecting the switch in surface type exhibiting the greatest LT retrieval sensitivity between the seasons.

The differences in surface and LT AKs for MOPITT retrievals in L3L and L3W discussed above can be accounted for primarily by the differing LT thermal contrast conditions over land and water, as explored in detail by Deeter et al. (2007). The sensitivity of MOPITT retrievals to CO in the LT is predominantly controlled by the thermal contrast between the surface skin temperature (T_{skin}) and the surface air temperature (T_{sfc}), and by the tropospheric temperature profile. Seasonal mean temperature profile data from ERA-Interim for the nearest land-only and water-only model gridboxes to Halifax show clear differences in DJF and JJA (Figure 4). In DJF, T_{skin} is around 6° K warmer than the 2-meter air temperature (T_{2m}, which we use as a proxy for T_{sfc} as this is the lowest model level) over water, and a further degree warmer than the air at 1000 hPa, whereas the temperature gradient is weak/slightly inverted over land, with T_{skin} less than a degree warmer than T_{2m}, which is actually slightly cooler than the air at 1000 hPa. Correspondingly, the lowest couple of retrieval levels indicate greater sensitivity to the surface and LT over water (in L3W) than land (in L3L) in DJF (Figure 3). Temperature profiles converge towards the MT, as do AKs. In JJA on the other hand, there is a clear gradient between T_{skin} and the overlying air on land, while the ocean surface is actually cooler than the air above, up to a height of 900 hPa. As a result, surface and LT sensitivity is greater in L3L than in L3W in JJA, with the sensitivity of retrievals in L3W approaching zero close to the surface owing to this inverted temperature profile. The T_{skin} increase approaching 20° K between DJF and JJA likely also accounts for the relatively greater overall true profile sensitivity (indicated by DFS values) in JJA for L3L than in DJF for ~~JJA than in~~ L3W ~~in DJF~~. Since our analysis is conducted using the joint TIR-NIR product, it is important to bear in mind that the benefit of enhanced LT sensitivity due to the incorporation of NIR is limited to retrievals over land, so this will also have an impact on the AK differences presented above. However, a land-water retrieval sensitivity contrast of comparable magnitude to that presented here is also evident in the TIR-only product, reinforcing the primary role of thermal contrast differences (see Supplementary Material 1).

That ~~-~~the retrieval sensitivity contrast between L3L and L3W is most pronounced in the LT is consistent with the finding that retrieved CO profiles in L3L and L3W show the greatest ~~contrast~~ differences in the LT ~~(Section 3.1.1, Figure 2). In DJF, the retrieval is more sensitive to the true profile in the LT in $L3_W$ (Figure 3, top row) and retrieved LT VMRs are greater in $L3_W$ than $L3_L$. In JJA, the retrieval is more sensitive to the true profile in the LT in $L3_L$ (Figure 3, bottom row), and retrieved LT VMRs are less in $L3_L$ than $L3_W$. This implies that in DJF, LT CO retrievals in $L3_L$ are weighted more heavily towards an a priori profile in which LT VMRs are lower than in the true CO profile; while in JJA, LT retrievals in $L3_W$ are weighted more heavily towards an a priori profile in which VMRs are greater than in the true profile. We explore this point in detail in Section 3.2.2~~. Although mean retrieval sensitivity in L3L and L3W converges with altitude, differences do exist from day-to-day but they are neither as large, nor as skewed in favour of retrievals over land or water (depending on season) as in the LT (see Supplementary Material 2), where there is a well understood thermal contrast mechanism creating the systematic land-water sensitivity contrast ~~(see Supplementary Material 1)~~. Likely causes for this could be day-to-day changes in atmospheric conditions (i.e. temperature or water vapour profiles), or random instrumental or retrieval noise.

3.1.3. Control of ΔRET by land-water sensitivity differences

To demonstrate that retrieval sensitivity differences over land and water can lead to the observed differences between CO profiles in L3L and L3W that are retrieved at the same time, we simulate and compare the pairs of retrieved CO profiles in L3L and L3W that are analysed in the preceding sections, using the transformation outlined in Section 2.2. For each L3$_L$- L3$_W$ simulated retrieval pairing ($X_{sim,L3-L}$ and $X_{sim,L3-W}$), the true and a priori profiles are common to both (recall the negligible mean difference between L3$_L$ and L3$_W$ a priori profiles in Figure 2), but the retrieval AKs from L3$_L$ and L3$_W$ are used (i.e. for simulated retrievals in L3$_L$ (L3$_W$), the AK from L3$_L$ (L3$_W$) is used). Thus, Recall that any differences between each pair of simulated retrievals are solely a result of differences in AKs. BecauseOwing to available CAMSRA data only coversing the years 2003-2016, as opposed to MOPITT data being available for 2000-2017, only a (large) subset of the retrieval pairings considered in earlier sections (which span the period 2001-2017) are could be simulated.

We first demonstrate the sensitivity effect with a case study on 2013-08-18. For brevity, we focus only on the simulated surface level retrievals ($X_{sim,L3-L,sfc}$ and $X_{sim,L3-W,sfc}$). Profiles (a priori ($X_{apr,sim}$) and ""truth"" ($X_{true,sim}$)) and surface level AKs (AK_{sfc}) used in the simulation, and the resulting $X_{sim,L3-L}$ and $X_{sim,L3-W}$, are given in Table 2. For brevity, we only focus on the surface level. AK_{sfc} from L3L and L3W are also shown in Figure 5. In this example, $X_{true,sim}$ is considerably lower than $X_{apr,sim}$ at the surface level, (-60.36 ppbv; $X_{true,sim,sfc}$ and $X_{apr,sim,sfc}$ respectively) and several features of the AKs indicate greater sensitivity to $X_{true,sim}$ in L3L and L3W: the AK value is significantly greater at the surface component of AK_{sfc} ($AK_{diag,sfc}$) is significantly greater in L3L than L3W (0.23 vs 0.01), and the rowsum ($AK_{rowsum,sfc}$) is over 5 times as high (0.84 vs 0.15). Correspondingly, $X_{sim,L3-L,sfc}$ is much24.92 ppbv lower than $X_{sim,L3-W,sfc}$ at the surface and much closer to $X_{true,sim,sfc}$ (32.21 ppbv higher vs 57.13 ppbv higher). Both Both $X_{sim,L3L,sfc}$ and $X_{sim,L3W}$ values indicate that $X_{true,sim,sfc}$ is lower than $X_{apr,sim,sfc}$ at the surface,, but $X_{sim,L3-L,sfc}$ gives the closer estimate, as would be expected from the greater $AK_{diag,sfc}$ and $AK_{rowsum,sfc}$ values from L3$_L$. In both cases, a portion of the overall $X_{sim,sfc}$ departures from $X_{apr,sim,sfc}$ at the surface level (in other words, "value added" over the a priori) originates at other profile levels of the profile, and not the surface level itself. This is a result of the surface level AK_{sfc} being nonzero at other profile levels, and is a function of the inter-level correlation of the original retrieval, which is linked to the a priori covariance matrix used in the retrieval (Deeter et al., 2010). Because AK_{sfc} has a well-defined peak at the surface L3$_L$ (Figure 5), while AK_{sfc} over L3$_W$ actually indicates greatest retrieval sensitivity (albeit very weak) to a broad region from 900 – 400 hPa, the surface accounts for a greater proportion of the $X_{sim,L3-L,sfc}$ departure from $X_{apr,sim,sfc}$ (55 % vs 25 % – see Appendix B for full derivation of these values) while the $X_{sim,L3-W,sfc}$ departure from $X_{apr,sim,sfc}$ has a greater contribution from the lower-to-mid troposphere as a whole.

We now consider differences between all $X_{sim,L3-W}$ and $X_{sim,L3-L}$ pairings (ΔSIM) throughout the profile and compare these to the observed differences between temporally coincident retrievals in L3L and L3W (ΔRET) discussed previously (Figure 5a6). For ease of comparison, ΔRET values are overlaid (faint lines) for the shorter time period matching CAMSRA data availability (the 2003-2016 ΔRET patterns are very similar to those seen in Figure 2). It should be noted that we cannot expect ΔSIM to match ΔRET exactly, owing to (possibly large) differences between the $X_{true,sim}$ profiles used in the simulations and the (unknown) true profiles at the time of the actual MOPITT retrievals (X_{true}). For instance, $X_{true,sim}$ is a monthly mean value from a reanalysis model, whereas X_{true} varies by day, which should result in less variance in ΔSIM than ΔRET. In DJF, mean ΔSIM is negligible at all profile levels shown, and the range of values is far smaller than seen for ΔRET. In JJA on the other hand, ΔSIM reaches considerably larger values than in DJF and mean values are significantly different ($p < 0.05$) over land and water at all but one profile level shown (300 hPa). LT and MT ΔSIM distributions are a remarkably good match for ΔRET, given the likely substantial differences between $X_{true,sim}$ and X_{true}. In the UT, ΔSIM values are somewhat smaller, although this isn't unexpected given the smaller land-water AK differences evident in Figure 3.

That L3L and L3W simulated and retrieved profile differences over land and water are comparable in JJA is clear evidence that ΔRET is strongly influenced by the land-water sensitivity contrast in summer months, at least in the LT. However, while ΔSIM values are of a much smaller magnitude in DJF than in JJA, ΔRET is actually of a similar magnitude in both seasons. The question therefore arises as to why ΔSIM is so different from ΔRET in DJF, and why it is of much smaller magnitude in DJF and JJA. Considering the terms of Equation 2, there are two possible explanations. Firstly, differences between L3L and L3W LT and MT AKs (A in Equation 2) are much smaller in DJF than in JJA (Figure 3). This means that the deviation of $X_{sim,L3-W}$ and $X_{sim,L3-L}$ from $X_{apr,sim}$ will be more similar in DJF (resulting in small ΔSIM), as opposed to in JJA where, as was seen in the case study discussed above, $X_{sim,L3-L}$ can deviate more from $X_{apr,sim}$ than $X_{sim,L3-W}$ owing to increased sensitivity in L3L in the lower profile levels (resulting in greater ΔSIM than in DJF). In the UT, AK differences are comparable in both seasons, and ΔSIM is correspondingly similar. Secondly, the magnitude of $X_{true,sim} – X_{apr,sim}$ from Equation 2 is, on average, around 4 (3) times greater in JJA than in DJF at the surface (900 hPa) level (Figure 56b). $X_{sim,L3-L}$ therefore deviates more from $X_{apr,sim}$ than $X_{sim,L3-W}$ throughout the LT and MT in JJA owing to strong contrasts in near-surface sensitivity at these profile levels, thus yielding large ΔSIM. Conversely, closer $X_{true,sim}$ and $X_{apr,sim}$ profiles combine with small sensitivity differences to limit ΔSIM in DJF. A final, alternative explanation for ΔSIM being a poor match for ΔRET in DJF, unlike in JJA, is that sensitivity differences in L3L and L3W could have less of an impact on ΔRET in

DJF than in JJA, and that something else is responsible – for example, "real" differences between the true CO profile over land and water. This is something we explore in detail in the following sections.

3.1.4. Regional land-water contrast in L2 data

To further evaluate whether ΔRET within the L3 gridbox containing Halifax is a function of land-water sensitivity contrasts, or actually due to real gradients in the true CO profile (for example, due to offshore transport of emissions from the city of Halifax or marine-land chemistry differences), we analyse the characteristics of retrieved profiles over the broader geographical region ~~containing~~ surrounding Halifax. If ΔRET is linked to sensitivity, then we would expect there to be a clear land-sea contrast across the whole region. This is exactly what we see. Figure 6~~7~~ shows seasonal median L2 retrieved and a priori CO concentrations for the surface ~~profile~~ level of the profile, where ΔRET and L3L-L3W AK differences are greatest and most significant, for the Canadian maritime provinces and a small portion of northeast USA (note that we show seasonal median fields here as the spatial patterns are clearer than for plots consisting only of the subset of days analysed in the rest of this section. The corresponding plot for the subset of days is shown in Supplementary Material 3~~2~~, and the main findings are unchanged). Difference fields (RET-APR) are also shown. In JJA, a land-sea contrast is remarkably clear in both the retrieved and RET-APR fields. The a priori field shows elevated CO amounts emanating from the west-southwest (indicative of CO sources in northeast USA and around the Great Lakes) and decreasing quite smoothly towards the north and east. The west-southwest maxima is replicated in the retrieved field, but the smoothly decreasing gradient is clearly broken, with the land in the image characterized by lower CO values than the adjacent ocean. This contrast is enhanced further in the RET-APR field. RET-APR values close to zero indicate either very low retrieval sensitivity or closely matching retrieved and a priori values. The analysis of averaging kernels in Section 3.1.2 demonstrated the lack of retrieval sensitivity at the surface over water (in L3W) in JJA. On average, RET-APR is ~~over~~ 11.5~~0~~ ppbv lower over land than over water (determined by binning the L2 data for the region shown according to surface classification)~~, a highly significant difference (p = 0.000, from 2-tailed T-test of binned data)~~. This reinforces our earlier interpretation that, in JJA, LT retrievals in L3W are weighted more heavily towards an a priori profile in which CO concentrations are too high than LT retrievals in L3L.

The land-sea contrast is less clear in DJF, although the RET-APR field does indicate generally positive values over water and negative values over land. The contrast being less apparent in DJF compared to JJA is consistent with the smaller L3L-L3W AK differences in DJF compared to JJA. However, it is surprising that RET-APR changes sign from land (generally negative) to water (generally positive). This may be linked to some factor other than retrieval sensitivity to the true CO profile, such as errors in retrieved/a priori surface temperatures or emissivities, which are important components of the radiative transfer model used in

MOPITT's CO retrieval algorithm. If this were the case over the sea, it could explain the maxima in RET-APR values to the northwest of Halifax in the Bay of Fundy, where the water is relatively shallow and the tidal range is the highest in the world at 16.3 m, transporting large amounts of suspended sediments (which will affect emissivity). Alternatively, the difference could reflect a physical process causing elevated CO over the ocean, although this seems unlikely, as we expect atmospheric transport to minimize such a contrast.

Corresponding maps for selected other ~~retrieval~~ levels of the profile and TCO are shown in Supplementary Material 4~~3~~. In JJA, a land-sea contrast is qualitatively evident at all ~~profile~~ levels and for TCO, with the exception of 800 hPa; and in DJF a contrast is evident at 800 hPa and for TCO.

3.1.5. Can ΔRET be explained by circulation-driven horizontal gradients in the ~~TRUE~~ true CO profile?

The preceding sections presented evidence that differences in temporally coincident L3L and L3W retrieved profiles are linked to sensitivity contrasts over land and water, especially in JJA and in the LT. However, these differences could also be a result of horizontal gradients in the true CO profile. It is plausible that retrieved LT VMRs are greater in L3W than in L3L due to, e.g., offshore transportation of CO by regional winds either from Halifax or from the large polluting areas on the northeast coast of the US and around the Great Lakes (as seen in the general decline in CO amounts from the west-southwest towards the northeast in Figure 6~~7~~). Winds generally tend from the west/northwest/southwest in this area (Figure 7~~8~~), which will lead to offshore transport of continental pollution.

We compare composite mean wind patterns across Nova Scotia using ERA-Interim data for days when retrieved surface level VMRs in L3W are greater than in L3L (L3W > L3L) and days when they are less (L3W < L3L), since a clear shift in wind direction on these days would support the case that atmospheric transport plays a role in generating differences in retrieved CO amounts over land and water. ~~to illustrate the likely atmospheric transport direction of CO (and other atmospheric constituents) in the region.~~ These are shown in Figure 7~~8~~ (10-metre winds are used). There is a clear circulation difference evident in DJF. When L3W > L3L (12~~3~~ days of the 18~~9~~ in Figure 2) the wind is in an offshore direction, whereas when L3W < L3L (6 days of the 18~~9~~ in Figure 2) the wind is alongshore from the west-southwest (and noticeably weaker). This is evidence to suggest that the LT ΔRET patterns in DJF are linked to the horizontal gradients in the true CO profile, although the small sample sizes involved here dictate caution. In contrast to DJF, however, there is no clear circulation difference in JJA; the winds are generally from the west-southwest irrespective of whether L3W > L3L or L3W < L3L. This lends further support to the conclusion that LT ΔRET in JJA is strongly linked to the demonstrated land-water sensitivity contrast. The seasonal difference in these results could explain why LT (and especially near-surface) ΔRET is of comparable magnitudes in DJF and JJA (Section 3.1.1), despite smaller L3L-L3W retrieval sensitivity contrasts and ΔSIM values in DJF than JJA

(Sections 3.1.2 and 3.1.3 respectively). In other words, in DJF, ΔRET is more indicative of differences in true CO concentrations, while in JJA it is more strongly tied to differences in retrieval sensitivity. This is not to say contrasts in retrieval sensitivity do not influence LT ΔRET in DJF; just that the effect is not as strong as in JJA, when the LT sensitivity contrast is greater. We only consider the surface profile level here; the findings are consistent higher up in the LT, but the DJF circulation difference is absent in the MT (see Supplementary Material 5~~4~~), consistent with ΔRET being much smaller in the MT and above, on average.

While there is no obvious circulation difference at the surface in JJA between days when L3W > L3L and L3W < L3L, there is a difference in ~~evidence that these days actually cover different parts of~~ the~~-~~ distribution of these days throughout the analysed MOPITT timeseries, which spans 16~~7~~ JJA seasons from 2001~~0~~ to 2016. 16 of the 19 d~~D~~ays when L3W < L3L occur~~tend to fall towards the start~~ in the first half of the timeseries (i.e. before 2009)~~mean year in which the retrieval was performed = 2004)~~, whereas days when L3W > L3L are spread ~~tend to~~ more evenly throughout (33 of the 65 days (51%) occur before 2009)~~be later (mean retrieval year = 2008)~~. In DJF there is no such difference, with roughly 50% of days occurring before and after 2009 in each case. This is something we explore further in Section 3.2.2.~~The same is not true in DJF (mean retrieval years = 2008 and 2007 respectively). As we show in Section 3.2.2, this appears to be a function of true LT CO concentrations over Halifax decreasing across the time period studied, which is detectable in JJA by the retrievals in L3$_L$ owing to their greater near-surface sensitivity, but not by the retrievals in L3$_W$, which are tied to an a priori value that is too low at the start of the timeseries and too high at the end.~~

3.2. Consequences for L3O timeseries

In this section we demonstrate how the statistics of the L3O timeseries, and the results of a typical trend analysis using those data, are affected by the loss of loss of LT retrieval information from L2 products over land when the L3 products for the coastal gridbox containing Halifax are created. We do this through comparison with the L3L and L3W timeseries. Because users of L3 data are advised to filter according to surface index in order to limit their analysis to retrievals with maximal information content, we consider L3O subsets that remain after filtering the timeseries for days with a surface index of “water” and “mixed” (“L3O$_{(water)}$” and “L3O$_{(mixed)}$”), as well as the unfiltered L3O timeseries to evaluate the full range of options available to users of the products. L3O only has a surface index of “land” once each season, so this subset is omitted. We focus on the LT profile levels since this is where retrieval sensitivity differences are greatest and can be linked to differences in retrieved CO values, as shown in the previous analyses.

3.2.1. Impact on seasonal data distribution

Seasonal surface level VMR distributions for L3L, L3W and all L3O subsets are shown in Figure 8. Most strikingly, L3L is the clear outlier in JJA. Mean VMRs are significantly lower than in all other timeseries ($p < 0.1$ in all cases) and the spread of values is around twice as large, both in terms of the interquartile range and overall range (excluding outliers), this difference mostly coming from the lower end of the distribution. This is unsurprising when comparing L3L with the timeseries that are based purely on retrievals over water (L3W and $L3O_{(water)}$), given the demonstration in previous sections that retrievals over water have significantly lower information content than over land in the summer months and are therefore more closely tied to the a priori CO concentration (i.e. retrieved VMRs will vary less). However, it clearly shows how the valuable additional information on true CO content that is available in L2 retrievals over land is diluted by their averaging with retrievals over water for $L3O_{(mixed)}$ days (mean surface level AK rowsum = 0.38 for $L3O_{(mixed)}$ vs 0.69 for L3L and 0.14 for L3W, see Figure 3), effectively creating a high bias in the resulting gridded mean VMRs. The loss of retrieval information from L2 to L3 is actually exacerbated for this 1° x 1° L3 coastal gridbox containing Halifax, given that a greater number of L2 retrievals over water contribute to the gridded averages than L2 retrievals over land (as previously outlined in Table 1), primarily because more of the surface within the L3 gridbox is water than land (see Figure 1). Consequently, the $L3O_{(mixed)}$ distribution more closely resembles L3W than L3L (although the L3W-$L3O_{(mixed)}$ mean difference is still statistically significant ($p < 0.1$)). Owing to the lack of days when the L3O timeseries is created only from retrievals over land, $L3O_{(mixed)}$ represents the best option for quantifying surface level CO in JJA that is available to users of the original L3 product in this case. The optimal retrievals for this task are only available in the L2 data products, a direct result of the way that the L3 products are created.

For the same reasons discussed above, L3L also represents the outlier distribution in DJF. Unlike in JJA however, the spread of VMR values is similar in L3L and L3W, likely reflecting the fact that there is some surface level sensitivity in retrievals over both land and water in DJF, allowing for a similar degree of departure from the a priori. The main difference in the distributions is that VMRs in L3L are offset towards lower values. However, the mean difference is not significant between L3L and any of the other timeseries. Also unlike in JJA, L3L does not necessarily represent the optimal timeseries for analyzing surface level CO in DJF, since retrieval sensitivity is actually higher over water in this season. Information loss resulting from the way L3 products are created is therefore less of an issue than in JJA, owing to the dominance of retrievals over water on the L3O timeseries (mean surface level AK rowsum = 0.52 for $L3O_{(mixed)}$ vs 0.43 for L3L and 0.53 for L3W, see Figure 3). It is worth noting, however, that L3W offers ~25% more days with data than $L3O_{(water)}$ in DJF, due to the fact that it “gains” the retrievals over water that go into $L3O_{(mixed)}$. This is potentially valuable additional temporal information for users of MOPITT products.

Although the sample sizes considered here are different, because L2 retrievals over land and water are not necessarily always present on the same days (e.g. due to variable cloud coverage), the differences in seasonal data distribution discussed in this section hold if the analysis is restricted to only days when L3L and L3W are both present (see Supplementary Material 6). We have only presented analysis for the surface level of the profile here as this is where L3L-L3W differences in retrieved VMRs and retrieval sensitivity are greatest. Plots of other levels are given in Supplementary Material 7.

3.2.2. Consequences for temporal trend analysis

To identify and compare temporal trends in the timeseries considered above, we perform weighted least squares (WLS) regression analyses on respective seasonal mean profile and TCO values, weighted by the standard deviation of the measurements used in the seasonal mean. For seasons that contain just a single measurement, we use the data record standard deviation scaled by a factor of 100 so as to de-weight these seasons in the fit. All trends identified are detailed in Table 3, and WLS best-fit lines, along with boxplots of seasonal data distributions, are presented for the surface level in Figure 9, which is where the most notable differences occur, In JJA. Here, the decreasing trend identified in L3L is over four times stronger than the trend in L3W, a highly significant difference ($p < 0.01$). This is a direct consequence of surface level retrievals over water being tied closely to the a priori owing to their negligible sensitivity, which has no yearly change. This effectively masks the full magnitude of the decrease in CO that appears to be occurring, and which is better detected by retrievals over land owing to their greater sensitivity. Consequently, trends in all L3O subsets are also significantly weaker than the trend in L3L. Since it has some contribution from retrievals over land, $L3O_{(mixed)}$ provides the closest approximation of the trend in L3L, but it is still over 50% weaker – representing a decrease of 10% (19 ppbv) over the 15-year period covered by the analysis vs 21% (40 ppbv) in L3L. Compared to JJA, trends in surface level VMRs in DJF are far more similar across all the timeseries, with no significant differences between any of the trends identified. This is attributable to the much smaller differences in retrieval sensitivity over land and water in this season. Thus, while the greater number of days corresponding to the L3W timeseries makes it of potentially greater value than $L3O_{(water)}$, at least for temporal trend analysis it has no statistical benefit (in fact, users of the original L3 product would seem to get comparable results by performing the analysis on the unfiltered version of L3O).

Although the land-water retrieval sensitivity contrast remains large at the 900hPa level in JJA, the trend in $L3O_{(mixed)}$ is a closer match to L3L than at the surface, and the difference loses statistical significance. The loss of information content available in retrievals at the 900hPa level over land during the creation of L3O (mean 900hPa AK rowsum = 0.91 for L3L vs 0.64 for $L3O_{(mixed)}$) therefore does not have a statistically significant impact on temporal trends identified using $L3O_{(mixed)}$, when compared to L3L. Although the L3L-

L3W trend difference is smaller than at the surface level, L3L is still twice as strong as L3W and the difference remains statistically significant. This could be a result of either the retrieval over water still lacking sufficient information to deviate as far from the a priori as the retrieval over land, despite the increase in information content relative to the surface level (mean AK rowsums = 0.15 and 0.44 for the surface and 900hPa levels respectively); and/or the retrieval at the 900hPa level over water having a sensitivity peak higher up in the troposphere where CO concentrations may be decreasing at a slower rate than they are closer to the surface, where sensitivity peaks for the 900hPa level over land (see Figure 3). By 800hPa, trends in all timeseries in JJA have converged, consistent with the further weakening of the land-water sensitivity contrast.

Moving away from the LT, the trends outlined in Table 3 indicate that all timeseries generally agree on the broad picture: in both seasons, CO concentrations are decreasing in the LT and MT, increasing in the UT, while TCO shows a decrease in DJF and no significant trend in JJA. Although not as pronounced or significant as at the surface in JJA, in all cases there are differences in the magnitude of the identified trend. This is not unexpected given that the seasonal means being regressed differ between timeseries. As outlined in Sections 3.1.1 and 3.1.2, temporally coincident retrieved VMRs over land and water do differ in levels of the profile above the LT despite similar retrieval sensitivity at these levels; the differences are just not as systematic as in the LT. However, in addition to the surface and 900hPa levels in JJA, there are 2 other instances where trends in L3L and L3W are significantly different: 600hPa in JJA and 300hPa in DJF (in both cases there is no statistically significant trend identified in L3L, whereas the trend in L3W is significant). The cause of these discrepancies is not readily apparent given that retrieval sensitivity over land and water is highly comparable in both cases, so further investigation would be needed before we can say with confidence whether or not they have consequences for analyses using the L3O timeseries, as we have been able to do for the LT.

The main differences in trend discussed above remain if the WLS regression analysis is restricted to only days when L3L and L3W are both present. Results from this restricted analysis are shown in Supplementary Material 8. It is important to note that MOPITT profile measurements are known to have a drift (Deeter et al., 2017), and this should be corrected for in the data if the focus of analysis is to use them to quantify temporal changes in CO over time. Since the intention of the WLS trend analysis presented here is more illustrative, namely to demonstrate trend differences in the data, we have not corrected for this drift. The results should therefore not be taken out of this context (as well as bias correction, verification against a range of other datasets would be required, especially given the large proportion of missing data). We do however provide the reported drift values in Table 3 for context, which shows that the majority of the trends that we have identified appear to be stronger than the measurement drift (at least for the dataset that has greatest retrieval sensitivity at the respective level of the profile). As noted in Section 2.1.1, the measurement drift has been significantly reduced in the latest version of the MOPITT products to be released (Version 8; Deeter et al., 2019).

3.3. Consideration of MAM and SON

In this paper we have focused only on DJF and JJA for brevity and clarity. The main findings discussed also hold for MAM and SON, however. In MAM, there is a strong sensitivity contrast in the LT between L3L and L3W, similar to that seen in JJA with retrieval sensitivity much greater over land than water (Figure 10a). As in JJA, the decreasing trend detected in WLS regression analysis of L3L surface level VMRs in MAM is strongly underestimated in L3W and, consequently, in all L3O subsets (Figure 10b). In SON on the other hand, AKs indicate a much more comparable degree of LT retrieval sensitivity over both land and water; correspondingly, the detected temporal trends are similar in L3L, L3W and L3O subsets.

3.2. Consequences for $L3_O$ timeseries

In this section, we demonstrate the impact that sensitivity driven differences in L2 retrieved profiles have on the statistics of the original L3 timeseries for Halifax ($L3_O$) that they are used to create, and consequences therein for the results of a typical temporal trend analysis. We do this through comparison with the $L3_L$ and $L3_W$ timeseries. For clarity, unlike in the previous section, the analysis in this section is not only limited to days when the $L3_O$ surface index is mixed; all days are considered.

3.2.1. Impact on seasonal data distribution

In both DJF and JJA, retrieved CO VMRs at the surface in $L3_O$ share a distribution that is qualitatively much more similar to $L3_W$ than $L3_L$, with smaller (and less statistically significant) mean differences (Figure 9). This is to be expected: the $L3_O$ timeseries consists of a greater proportion of L2 retrievals over water than over land (Table 1), primarily because more of the surface within the L3 gridbox containing Halifax is water than land (owing to the location of the coastline – see Figure 1). Thus, there is a greater likelihood of the $L3_O$ surface index being classified as water (and only L2 retrievals over water being retained for the area averaging in L3 data creation), and also of more L2 retrievals over water contributing to the area average than L2 retrievals over land when the $L3_O$ surface index is mixed (see Section 2.1.2).

The difference in the distribution of surface level retrieved CO VMRs between $L3_L$ and both of $L3_O$ and $L3_W$ is most striking in JJA. This is to be expected, given that this is when the $L3_L$-$L3_W$ sensitivity differences are greatest. Although the mean difference between $L3_O$ and $L3_L$ is relatively small (5.9 ppbv, $p = 0.04$), the spread of values is far greater for $L3_L$ (standard deviation (interquartile range) = 35.4 ppbv (44.3 ppbv)) than for $L3_O$ (standard deviation (interquartile range) = 19.1 ppbv (19.7 ppbv)). This lower variability in $L3_O$ is a direct result of the fact that L2 surface level retrievals over water, which have a significantly lower information content than L2 surface level retrievals over land in the summer months and are therefore more closely tied to the a priori CO concentration (Section 3.1.2), have a greater contribution to the $L3_O$ timeseries than L2 retrievals over land. $L3_L$ contains significantly more information on the true CO concentration variations at the surface (mean surface level AK_{rowsum} = 0.70 and 0.25 for $L3_L$ and $L3_O$, respectively, p = 0.000) – albeit at the cost of over 55 % fewer observations being available for analysis during the time period covered (n = 106 for $L3_L$ vs 248 for $L3_O$) – yet this information on CO variability is lost to users of the $L3_O$ timeseries because of the way the L3 products are created, as described in Section 2.1.2. The same is not true in DJF. Although $L3_W$ has marginally more information on the true CO concentrations at the surface than $L3_O$ (mean surface level AK_{rowsum} = 0.55 and 0.54, respectively), which has its information content degraded slightly by the influence of L2 retrievals over land (mean surface level AK_{rowsum} = 0.44), the difference in CO concentrations in the two timeseries is negligible and not statistically significant (1.8 ppbv, p = 0.49), owing to the dominance of L2 retrievals over water on the $L3_O$ timeseries. Thus, unlike in JJA, there is no significant loss of information on the true CO VMR variations at the surface for users of the $L3_O$ timeseries. It is also noticeable that the spread of values is much more similar across the three timeseries than in JJA (standard deviation (interquartile range) = 24.9, 22.2, 19.9 (25.2, 25.0, 24.2) for $L3_L$, $L3_W$ and $L3_O$, respectively), suggesting a lack of systematic difference in the values that are being retrieved, in contrast to JJA.

Although the sample sizes considered here are different, because L2 retrievals over land and water are not necessarily always present on the same days (e.g. due to variable cloud coverage), these results hold if the preceding analysis is restricted to days when $L3_L$ and $L3_W$ are both present (see Supplementary Material 5). We only present analysis for the surface profile level here as this is where $L3_L$-$L3_W$ differences in retrieved VMRs and retrieval sensitivity are greatest. Plots of other profile levels and TCO are given in Supplementary Material 6.

3.2.2. Consequences for temporal trend analysis

Results from Ordinary Least Squares (OLS) regression analyses on seasonal mean profile and TCO values from the three different timeseries are presented in Table 3. OLS best-fit lines, along with boxplots of seasonal data distributions, are shown in Figure 10 for the surface profile level, where the $L3_L$-$L3_W$ difference in retrieval sensitivity and retrieved VMRs is greatest. In our analysis we focus on the gradient parameter of the best-fit line ($m_{L3\text{-}O}$, $m_{L3\text{-}W}$, and $m_{L3\text{-}L}$), since this describes the trend in seasonal mean values over time. In general, the trend direction is consistent across all 3 timeseries in both DJF and JJA at all profile levels considered (with the exception of 600 hPa in JJA, discussed below) and for TCO. The trend magnitudes m_{L3-O} and m_{L3-W} are quite closely matched, while m_{L3-L} is most often different from both. This is not surprising, given the greater influence of L2 retrievals over water on the $L3_O$ timeseries, as explained earlier.

At the surface in JJA (Figure 10), m_{L3-L} shows a decrease ~6 times greater than m_{L3-W} and almost 3 times greater than m_{L3-O}. The difference between m_{L3-L} and both m_{L3-W} and m_{L3-O} is highly statistically significant ($p < 0.01$) in both cases. This is because $L3_w$ lacks the retrieval sensitivity in JJA to detect a change in the true CO profile close to the surface and is tied to the a priori which has no yearly change. Consequently, m_{L3-W} and, by extension, m_{L3-O} significantly underestimate the declining trend in surface level CO in JJA that m_{L3-L} is able to detect owing to the greater retrieval sensitivity over land during this season. The trend underestimation persists at 900 hPa, albeit more weakly (m_{L3-L} is now only ~2 times greater than m_{L3-W} and m_{L3-O}), but the differences remain statistically significant ($p < 0.1$). By 800 hPa, this trend underestimation in m_{L3-W} and m_{L3-O} has all but gone away, likely owing to the fact that the retrieval in $L3_W$ contains a moderate amount of true profile information at this level. In DJF, trends at LT profile levels are far more similar across the three timeseries compared to JJA, owing to the much smaller retrieval sensitivity differences between $L3_L$ and $L3_W$; the differences are largest at the surface profile level (but not statistically significant), and by 800 hPa the trends have almost converged, consistent with the DJF AK comparison (Figure 3, Section 3.1.2).

At 600 hPa in JJA, m_{L3-O} and m_{L3-W} indicate a decrease of 0.55 and 0.86 ppbv y^{-1} respectively, while m_{L3-L} is effectively zero, with no statistical evidence that any trend exists ($p = 0.942$). While at this profile level the overall information content of the retrievals is effectively the same in $L3_L$ and $L3_W$ (indicated by AK rowsums 1.03 and 1.01 respectively (Figure 3)), the AK shape is completely different: retrieval sensitivity in $L3_W$ has a strong peak at 600 hPa, as opposed to the retrieval sensitivity in $L3_L$ being uniform throughout the LT, MT and into the UT. Therefore, m_{L3-W} and, by extension, m_{L3-O}, appear to be more indicative of the actual temporal trend in the MT than m_{L3-L}, which seems to be detecting a mixed signal between decreasing CO in the LT-MT and increasing CO in the UT (see below), owing to the broad sensitivity profile of the $L3_L$ retrieval at this level. The influence of L2 retrievals over land on the $L3_O$ timeseries thus clearly weakens m_{L3-O} in JJA, which loses statistical significance and indicates a decrease in CO that is ~36 % weaker than the more reliable m_{L3-W}. As in the LT, in DJF the trends at 600 hPa are highly similar, which is to be expected given the closely-matched AKs for retrievals in $L3_L$ and $L3_W$ at this profile level.

Trends at 300 hPa also display differences between the $L3_O$, $L3_L$ and $L3_W$ timeseries. In DJF, m_{L3-L} indicates a slower increase in VMRs than m_{L3-W} and m_{L3-O}, and confidence that this trend is statistically different from zero is low, unlike for the trends in the other two timeseries. In JJA, m_{L3-L} indicates a faster increase in VMRs than m_{L3-W} and m_{L3-O}, but the differences in trend magnitude are not statistically significant ($p > 0.1$). Causes for these trend differences are difficult to establish given that 300 hPa AKs are similar in

~~L3$_L$ and L3$_W$ in both seasons. It is plausible, however, that subtle AK differences, especially at LT profile levels in JJA, will influence the retrieved values enough to impact upon the trends detected in this analysis, especially if true CO concentrations are changing by > ±1 % yr^{-1} throughout the profile as this analysis suggests.~~

~~In comparison to trends at most profile levels, differences in TCO trends between the three timeseries are relatively small in both seasons. This could be because TCO retrievals are better constrained than retrievals for individual profile levels, whose retrieved VMRs depend upon the true profile sensitivity of the retrieval. The weakness and lack of statistical significance of TCO trends in JJA is possibly a result of the strong UT increase in CO offsetting the LT and MT decrease. On the other hand, the stronger and statistically significant ($p < 0.1$) decrease in TCO in DJF may be due to UT increases being weak in comparison to LT decreases. It should be borne in mind that, as neither season has DFS > 2 (Figure 3), we cannot be entirely confident about the vertical location of the trends identified in this analysis since the retrievals lack the vertical resolution to detect multiple atmospheric layers independently (although confidence is greater in JJA than in DJF, owing to greater DFS values over both surfaces in summer).~~

~~To test whether the reported trend differences occur as a result of L3$_O$, L3$_L$ and L3$_W$ sometimes sampling different days, we restricted the OLS regression analyses to days when all three datasets are present. The main results hold, although the regression parameters are more poorly constrained (i.e. larger standard error and confidence intervals, less statistical significance) owing to the smaller sample sizes. Results from this restricted analysis are discussed further in Supplementary Material 7.~~

4. Conclusions~~.~~

Users of MOPITT products are advised to filter the data before analysis of profile values in order to maximize the influence of satellite measurements and minimize the impact of a priori CO concentrations on results (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015). In particular, it is advised that retrievals over water, which are known to have lower information content than retrievals over land, are discarded. This is especially so for the analysis of temporal trends in CO concentrations, owing to the year-to-year stationarity of the a priori. However, for L3 gridboxes that straddle the coastline, the ability to apply such filtering is limited since the products will generally have some contribution from L2 retrievals that take place over both land and water. This is a direct consequence of the way that L3 products are created.

As we have explicitly demonstrated for the 1° x 1° L3 gridbox containing the coastal city of Halifax, Canada, the L2 retrieved CO concentrations, from which the L3 products are created, differ depending on whether the retrieval took place over land or water. In JJA, and especially near to the surface, this is directly linked to differences in the sensitivity of the retrievals to the true CO profile. The merging of these retrievals to create the L3 product can significantly affect the statistics of the dataset and the results of temporal trend analysis with the data, with the largest and most statistically significant effects at the surface, where land-water sensitivity contrasts are greatest. As we show, results that are more representative of changes of true CO concentrations close to the surface within the L3 gridbox containing Halifax can only currently be obtained by use of the L2 products, which can be filtered by surface type to maximize information content.

Our results suggest that L2 retrievals over land and water should not both contribute to L3 products in coastal gridboxes. This is consistent with previous data filtering recommendations (MOPITT Algorithm Development Team, 2017; Deeter et al., 2015). The horizontally averaged L3L and L3W timeseries that we have analysed in this paper are effectively L3 "land-only" and L3 "water-only" datasets, and these offer an alternative in this respect that preserves the benefits of available L3 products – namely, less computing resources and expertise required for their analysis compared to L2 products, which broadens access to the data – but offers users the flexibility to select over which surface the contributing retrievals were performed in order to maximize the information content of L3 data in coastal gridboxes. Although our study has only focused on the city of Halifax, the results suggest that similar studies be performed for other coastal L3 gridboxes before using MOPITT surface CO, ~~are of relevance to all coastal L3 gridboxes, which~~ since these contain 6 of the top 10 and 43 of the top 100 agglomerations by population and are therefore likely targets for analysis of temporal changes in air pollution indicators such as CO, especially near to the surface. The degree of information content ~~limitation~~ loss in the L3 data will depend on the relative contributions of L2 retrievals over land and water to each specific L3 gridbox, as well as on the strength of the land-water retrieval sensitivity difference, which in turn depends on scene-specific geophysical variables such as surface temperature and emissivity. Work is currently ongoing to compare the results of analyses conducted with MOPITT L2 and L3 CO data over these cities.

Data Availability

MOPITT data were downloaded from the NASA Earthdata portal (https://search.earthdata.nasa.gov/). CAMSRA and ERA-Interim data were downloaded from the ECMWF public datasets portal (https://apps.ecmwf.int/datasets/).

Appendix A: List of acronyms

ΔRET	Percentage differences in retrieved CO concentrations in $L3_L$ and $L3_W$
ΔSIM	Percentage differences in simulated retrieved CO concentrations in $L3_L$ and $L3_W$
AK	Averaging kernel

~~$AK_{L3\ L}$~~	~~Averaging kernel from $L3_L$~~
~~$AK_{L3\ W}$~~	~~Averaging kernel from $L3_W$~~
~~$APR_{L3\ L}$~~	~~A priori from $L3_L$~~
~~$APR_{L3\ W}$~~	~~A priori from $L3_W$~~
CAM-chem	Community Atmosphere Model with Chemistry
CAMS	Copernicus Atmospheric Monitoring Service
CAMSRA	CAMS Reanalysis
CO	Carbon Monoxide
DFS	Degrees of freedom for signal
DJF	December, January, February
JJA	June, July, August
L2	Level 2 data products
L3	Level 3 data products
L3L	Area-averaged L2 data for the L3 gridbox containing Halifax, for L2 retrievals over land only
L3O	Original, “~~‘~~as-downloaded”~~’~~ L3 timeseries for the 1° x 1° L3 gridbox containing Halifax
L3W	Area-averaged L2 data for the L3 gridbox containing Halifax, for L2 retrievals over water only
LT	Lower troposphere (surface – 800hPa ~~profile~~ levels of the profile in MOPITT data)
MERRA-2	NASA Modern-Era Retrospective Analysis for Research and Applications V2
~~$m_{L3\ L}$~~	~~Gradient parameter of best-fit line calculated from OLS regression analysis on seasonal mean CO concentrations from $L3_L$ timeseries~~
~~$m_{L3\ O}$~~	~~Gradient parameter of best-fit line calculated from OLS regression analysis on seasonal mean CO concentrations from $L3_O$ timeseries~~
~~$m_{L3\ W}$~~	~~Gradient parameter of best-fit line calculated from OLS regression analysis on seasonal mean CO concentrations from $L3_W$ timeseries~~
MOPITT	Measurement of Pollution in the Troposphere (instrument)
MT	Mid-troposphere (700hPa – 500hPa ~~profile~~ levels of the profile in MOPITT data)
NIR	Near infrared
SNR	Signal-to-noise ratio
TCO	Total Column CO
TIR	Thermal infrared

T_{sfc}	Surface air temperature
T_{skin}	Surface skin temperature
UT	Lower troposphere (400hPa – 100hPa profile levels in MOPITT data)
V7	MOPITT Version 7 data products
VMR	Volume Mixing Ratio
X_{apr}	A priori CO VMR profile
$X_{apr,sim}$	A priori CO VMR profile used in retrieval simulation
~~$X_{apr,sim,sfc}$~~	~~A priori surface profile level CO VMR used in retrieval simulation~~
X_{rtv}	Retrieved CO VMR profile
X_{sim}	Simulated retrieved CO VMR profile
$X_{sim,L3-L}$	Simulated retrieved CO VMR profile from L3L
~~$X_{sim,L3-L,sfc}$~~	~~Simulated retrieved surface profile level CO VMR from $L3_L$~~
$X_{sim,L3-W}$	Simulated retrieved CO VMR profile from L3W
~~$X_{sim,L3-W,sfc}$~~	~~Simulated retrieved surface profile level CO VMR from $L3_W$~~
X_{true}	“True” CO VMR profile
$X_{true,sim}$	“True” CO VMR profile used in retrieval simulation
~~$X_{true,sim,sfc}$~~	~~‘True’ surface profile level CO VMR profile used in retrieval simulation~~

~~**Appendix B**~~

~~In Section 3.1.3 we state that the surface profile level accounts for 55 % of $X_{sim,L3-L,sfc}$ departure from $X_{sim,apr,sfc}$, vs only 25 % for $X_{sim,L3-W,sfc}$. These values are calculated using the surface level row of the AK_{sfc} * Δ column from Table B1 (highlighted yellow) and the sum of that row (highlighted blue).~~

~~For $X_{sim,L3-L,sfc}$: $\frac{-0.0389}{-0.0701} = 0.55.$~~

~~For $X_{sim,L3-W,sfc}$: $\frac{-0.0018}{-0.0075} = 0.25.$~~

~~**Table B1** Case study information and full calculations for simulated $L3_L$ and $L3_W$ profile retrievals ($X_{sim,L3-L}$ and $X_{sim,L3-W}$ respectively) for 2013-08-18. Table includes $\log_{10}$(VMRs) used in the simulation and values for AK_{sfc} * Δ (Δ = $X_{apr,sim}$ – $X_{true,sim}$) in $\log_{10}$ space, to illustrate the effect of interlevel correlation on the simulated VMR retrieval. Values that are underlined indicate the maxima (by magnitude) for that column.~~

~~Simulated profile values for levels 900-100 hPa are shown in brackets because AKs and calculations for these levels are not discussed in the text and AK_{900} through AK_{100} are not shown.~~

Profile level	Input profiles						Ak_{sfc} (== **A**[1,i])		AK_{sfc} * **Δ** (log_{10})		Simulated profiles (X_{sim})			
	$X_{apr,sim}$		$X_{true,sim}$		**Δ**						$X_{sim,L3-W}$		$X_{sim,L3-L}$	
	VMR	Log_{10}	VMR	Log_{10}	VMR	Log_{10}	$L3_W$	$L3_L$	$L3_W$	$L3_L$	Log_{10}	VMR	Log_{10}	VMR
Surface	188.81	2.28	128.45	2.11	-60.36	-0.17	0.01	0.23	-0.0018	-0.0389	2.27	185.58	2.21	160.67
900 hPa	148.88	2.17	119.53	2.08	-29.34	-0.10	0.03	0.20	-0.0025	-0.0191	(2.15)	(141.49)	(2.09)	(124.02)
800 hPa	118.07	2.07	109.80	2.04	-8.27	-0.03	0.02	0.15	-0.0008	-0.0048	(2.04)	(110.76)	(2.00)	(100.25)
700 hPa	102.72	2.01	97.64	1.99	-5.08	-0.02	0.03	0.14	-0.0006	-0.0030	(1.98)	(95.37)	(1.95)	(88.72)
600 hPa	95.49	1.98	90.34	1.96	-5.15	-0.02	0.03	0.12	-0.0007	-0.0028	(1.94)	(87.92)	(1.92)	(83.64)
500 hPa	92.00	1.96	87.10	1.94	-4.91	-0.02	0.03	0.08	-0.0006	-0.0020	(1.93)	(84.57)	(1.92)	(82.57)
400 hPa	89.64	1.95	83.85	1.92	-5.78	-0.03	0.02	0.02	-0.0005	-0.0006	(1.92)	(83.46)	(1.92)	(84.12)
300 hPa	84.76	1.93	81.42	1.91	-3.34	-0.02	-0.01	-0.07	0.0002	0.0013	(1.91)	(81.73)	(1.92)	(83.91)
200 hPa	64.45	1.81	65.20	1.81	0.75	0.01	-0.01	-0.04	0.0000	-0.0001	(1.80)	(63.77)	(1.81)	(64.75)
100 hPa	26.51	1.42	28.71	1.46	2.20	0.03	0.00	0.00	0.0001	-0.0000	(1.42)	(26.41)	(1.42)	(26.52)
						Rowsum →	0.15	0.84	-0.0075	-0.0701	← Sum			

Author contributions

IA and AW jointly conceived of and designed the study. IA performed data analysis; both authors examined and interpreted the results, and prepared the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors received funding from the Canadian Space Agency through the Earth System Science Data Analyses program (grant no. 16SUASMPTN), the Canadian National Science and Engineering Research Council through the Discovery Grants Program, and Saint Mary's University. We thank the MOPITT team and ECMWF for providing the data used in this study, and would also like to thank Dylan Jones (University of Toronto) and James Drummond (Dalhousie University) for their helpful discussions, as well as two anonymous reviewers whose helpful comments greatly improved an earlier version of this paper.

References

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, a. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597. https://doi.org/10.1002/qj.828

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attié, J.-L., Packman, D., Chen, J., Mao, D. Drummond, J. R. (2003). Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument. Journal of Geophysical Research, 108(D14), 4399. https://doi.org/10.1029/2002JD003186

Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R. (2007). Sensitivity of MOPITT observations to carbon monoxide in the lower troposphere. Journal of Geophysical Research Atmospheres, 112(24), 1–9. https://doi.org/10.1029/2007JD008929

~~Deeter, M. N., Edwards, D. P., Gille, J. C., and Worden, H. M. (2015). Information content of MOPITT CO profile retrievals: Temporal and geographical variability. Journal of Geophysical Research: Atmospheres, 120(24), 12723–12738. https://doi.org/10.1002/2015JD024024~~

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., Daube, B. C. and Wofsy, S. C. (2013). Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000-2011. Journal of Geophysical Research Atmospheres, 118(12), 6710–6725. https://doi.org/10.1002/jgrd.50272

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H.M., Sweeney, C., Pittman, J. V., Daube, B. C., and Wofsy, S. C. (2014). The MOPITT Version 6 product: Algorithm enhancements and validation. Atmospheric Measurement Techniques, 7(11), 3623–3632. https://doi.org/10.5194/amt-7-3623-2014

Deeter, M. N., Edwards, D. P., Gille, J. C., and Worden, H. M. (2015). Information content of MOPITT CO profile retrievals: Temporal and geographical variability. Journal of Geophysical Research: Atmospheres, 120(24), 12723–12738. https://doi.org/10.1002/2015JD024024

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H. M., and Sweeney, C. (2017). A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product. Atmospheric Measurement Techniques, 10, 2533–2555. https://doi.org/10.5194/amt-10-2533-2017

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martínez-Alonso, S., Worden, H.M, Ziskin, D., and Andreae, M. O. (2019). Radiance-based retrieval bias mitigation for the MOPITT instrument: The version 8 product. Atmospheric Measurement Techniques, 12(8), 4561–4580. https://doi.org/10.5194/amt-12-4561-2019

Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J. (2010). A review of 9-year performance and operation of the MOPITT instrument. Advances in Space Research, 45(6), 760–774. https://doi.org/10.1016/j.asr.2009.11.019

Drummond, J. R., Hackett, J., and Caldwell, D.: Measurements of pollution in the troposphere (MOPITT), in: Optical Payloads for Space Missions, edited by: Shen-En Qian, Wiley and Sons, West Sussex, UK, 639–652, 2016.

Duncan, B. N., Logan, J. A., Bey, I., Megretskaia, I. A., Yantosca, R. M., Novelli, P. C., Jones, N.B., and Rinsland, C. P. (2007). Global budget of CO, 1988 - 1997: Source estimates and validation with a global model. Journal of Geophysical Research Atmospheres, 112(22), D22301, https://doi.org/10.1029/2007JD008459

Edwards, D. P., Halvorson, C. M., and Gille, J. C. (1999). Radiative transfer modeling for the EOS Terra satellite Measurement of Pollution in the Troposphere (MOPITT) instrument. Journal of Geophysical Research Atmospheres. https://doi.org/10.1029/1999JD900167

Francis, G. L., Deeter, M. N., Martínez-Alonso, S., Gille, J. C., Edwards, D. P., Mao, D., Worden, H. M., and Ziskin, D. (2017). Measurement of Pollution in the Troposphere Algorithm Theoretical Basis Document: Retrieval of Carbon Monoxide Profiles and Column Amounts from MOPITT Observed Radiances (Level 1 to Level 2). Atmospheric Chemistry Observations and Modelling Laboratory, National Center for Atmospheric Research, Boulder, Colorado. Downloaded from: https://www2.acom.ucar.edu/sites/default/files/mopitt/ATBD_5_June_2017.pdf

He, H., Stehr, J. W., Hains, J. C., Krask, D. J., Doddridge, B. G., Vinnikov, K. Y., Canty, T. P., Hosley, K. M., Salawitch, R. J., Worden, H. M., and Dickerson, R. R. (2013). Trends in emissions and concentrations of air pollutants in the lower troposphere in the Baltimore/Washington airshed from 1997 to 2011. Atmospheric Chemistry and Physics, 13(15), 7859–7874. https://doi.org/10.5194/acp-13-7859-2013

Inness, A., Ades, M., Agustí-Panareda, A., Barr, J., Benedictow, A., Blechschmidt, A. M., Jose Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V. H., Razinger, M., Remy, S., Schulz, M., and Suttie, M. (2019). The CAMS reanalysis of atmospheric composition. Atmospheric Chemistry and Physics, 19(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Lamarque, J. F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K. (2012). CAM-chem: Description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. Geoscientific Model Development, 5(2), 369–411. https://doi.org/10.5194/gmd-5-369-2012

MOPITT Algorithm Development Team (2017). MOPITT (Measurements of Pollution in the Troposphere) Version 7 Product User's Guide, Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder. Downloaded from: https://www2.acom.ucar.edu/sites/default/files/mopitt/v7_users_guide_201707.pdf

Pan, L., Edwards, D. P., Gille, J. C., Smith, M. W., and Drummond, J. R. (1995). Satellite remote sensing of tropospheric CO and CH_4: forward model studies of the MOPITT instrument. Applied Optics, 34(30), 6976. https://doi.org/10.1364/ao.34.006976

Pan, L., Gille, J. C., Edwards, D. P., Bailey, P. L., and Rodgers, C. D. (1998). Retrieval of tropospheric carbon monoxide for the MOPITT experiment. Journal of Geophysical Research, 103(D24), 32277. https://doi.org/10.1029/98JD01828

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, Theory and Practice, World Scientific, Singapore, 2000.

Strode, S. A., Worden, H. M., Damon, M., Douglass, A. R., Duncan, B. N., Emmons, L. K., Lamarque, J.-F., Manyin, M., Oman, L. D., Rodriguez, J. M., Strahan, S. E., and Tilmes, S. (2016). Interpreting space-based trends in carbon monoxide with multiple models. Atmos. Chem. Phys, 16, 7285–7294. https://doi.org/10.5194/acp-16-7285-2016

Wiacek, A., Li, L., Tobin, K., & Mitchell, M. (2018). Characterization of trace gas emissions at an intermediate port. Atmospheric Chemistry and Physics, 18(19), 13787–13812. https://doi.org/10.5194/acp-18-13787-2018

~~Worden, H. M., Deeter, M. N., Frankenberg, C., George, M., Nichitiu, F., Worden, J., Aben, I., Bowman, K. W., Clerbaux, C., Coheur, P. F., De Laat, A. T. J., Detweiler, R., Drummond, J. R., Edwards, D. P., Gille, J. C., Hurtmans, D., Luo, M., Martínez-Alonso, S., Massie, S., Pfister, G., and Warner, J. X. (2013). Decadal record of satellite carbon monoxide observations. Atmos. Chem. Phys. Atmospheric Chemistry and Physics, 13, 837–850. https://doi.org/10.5194/acp-13-837-2013~~

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P. (2010). Observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals. Journal of Geophysical Research Atmospheres, 115(18), 1–12. https://doi.org/10.1029/2010JD014242

Worden, H. M., Deeter, M. N., Frankenberg, C., George, M., Nichitiu, F., Worden, J., Aben, I., Bowman, K. W., Clerbaux, C., Coheur, P. F., De Laat, A. T. J., Detweiler, R., Drummond, J. R., Edwards, D. P., Gille, J. C., Hurtmans, D., Luo, M., Martínez-Alonso, S., Massie, S., Pfister, G., and Warner, J. X. (2013). Decadal record of satellite carbon monoxide observations. Atmos. Chem. Phys. Atmospheric Chemistry and Physics, 13, 837–850. https://doi.org/10.5194/acp-13-837-2013

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J., Drummond, J., Emmons, L. K., Francis, G., and Martínez-Alonso, S. (2014). 13 years of MOPITT operations: Lessons from MOPITT retrieval algorithm development. Annals of Geophysics, 56(FAST TRACK 1), 1–5. https://doi.org/10.4401/ag-6330

Table 1 (a) Surface classification of L3 data for the 1° x 1° gridbox containing Halifax (L3O) for the period 2001~~0~~-08~~3~~-25~~03~~ to 2017-03-05. Note that no retrievals are available for this L3 gridbox on 91~~3~~ % of days in DJF and 83 % of days in JJA. **(b)** Mean number of individual L2 retrievals with "land" and "water" surface index that contribute to the L3L and L3W timeseries, respectively (the standard deviation is in brackets).

	(a)				**(b)**	
Season	**Water**	**Land**	**Mixed**	**Total**	**$L3_L$**	**$L3_W$**
DJF	105	1	24	130	1.9 (±1.4)	3.6 (±1.9)
JJA	143	1	104	248	2.2 (±1.1)	3.8 (±1.1)

	(a)				**(b)**	
Season	**Water**	**Land**	**Mixed**	**Total**	**L3L**	**L3W**
DJF	101	1	23	125	1.7 (±1.1)	4.0 (±1.7)
JJA	136	1	98	235	2.1 (±1.2)	4.1 (±1.8)

Table 2 Case study information for simulated L3L and L3W profile retrievals ($X_{sim,L3L}$ and $X_{sim,L3W}$ respectively) for 2013-08-18. **Δ** = $X_{apr,sim} - X_{tr,sim}$. Values that are underlined indicate the maxima (by magnitude) for that column. Simulated profile values for levels 900-100 hPa are shown in brackets because AKs and calculations for these levels are not discussed in the text and AKs for the 900 – 100 hPa levels~~$_{900}$ through AK$_{100}$~~ are not shown.

Profile level	Input profiles			Ak_{sfc} (== A[1,i])		Simulated profiles (X_{sim})	
	$X_{apr,sim}$	$X_{true,sim}$	Δ	$L3_W$	$L3_L$	$X_{sim,L3-W}$	$X_{sim,L3-L}$
Surface	188.81	128.45	-60.36	0.01	0.23	185.58	160.67
900 hPa	148.88	119.53	-29.34	0.03	0.20	(141.49)	(124.02)
800 hPa	118.07	109.80	-8.27	0.02	0.15	(110.76)	(100.25)
700 hPa	102.72	97.64	-5.08	0.03	0.14	(95.37)	(88.72)
600 hPa	95.49	90.34	-5.15	0.03	0.12	(87.92)	(83.64)
500 hPa	92.00	87.10	-4.91	0.03	0.08	(84.57)	(82.57)
400 hPa	89.64	83.85	-5.78	0.02	0.02	(83.46)	(84.12)
300 hPa	84.76	81.42	-3.34	-0.01	-0.07	(81.73)	(83.91)
200 hPa	64.45	65.20	0.75	-0.01	-0.04	(63.77)	(64.75)
100 hPa	26.51	28.71	2.20	0.00	0.00	(26.41)	(26.52)
			Rowsum →	0.15	0.84		

Profile level	Input profiles			Surface level AK (== A[1,i])		Simulated profiles (X_{sim})	
	$X_{apr,sim}$	$X_{tr,sim}$	Δ	L3W	L3L	$X_{sim,L3W}$	$X_{sim,L3L}$
Surface	188.81	128.45	-60.36	0.01	0.23	185.58	160.67
900 hPa	148.88	119.53	-29.34	0.03	0.20	(141.49)	(124.02)
800 hPa	118.07	109.80	-8.27	0.02	0.15	(110.76)	(100.25)
700 hPa	102.72	97.64	-5.08	0.03	0.14	(95.37)	(88.72)
600 hPa	95.49	90.34	-5.15	0.03	0.12	(87.92)	(83.64)
500 hPa	92.00	87.10	-4.91	0.03	0.08	(84.57)	(82.57)
400 hPa	89.64	83.85	-5.78	0.02	0.02	(83.46)	(84.12)
300 hPa	84.76	81.42	-3.34	-0.01	-0.07	(81.73)	(83.91)
200 hPa	64.45	65.20	0.75	-0.01	-0.04	(63.77)	(64.75)
100 hPa	26.51	28.71	2.20	0.00	0.00	(26.41)	(26.52)
			Rowsum →	0.15	0.84		

Table 3 Results from W~~O~~LS regression analysis of seasonal mean L3W, L3L, L3O, $L3O_{(water)}$ and $L3O_{(mixed)}$ timeseries ~~$L3_O$, $L3_L$ and $L3_W$ timeseries~~ for selected profile levels in DJF and JJA. Trend corresponds to the gradient of the WLS best-fit line~~Units for TCO are mol cm^{-2}, all other levels are ppbv. m = gradient of OLS best-fit line~~; SE = standard error of trend ~~gradient~~; ~~p~~ P-value = probability that the trend~~gradient~~ is zero; % change y^{-1}= mean percentage change in retrieved CO per year, calculated from ~~O~~WLS regression model predicted values ~~(pred)~~ as follows:

$$\% \text{ change y}^{-1} = \left\{\left[\left(\frac{\text{Predicted}_{\text{last}}}{\text{Predicted}_{\text{first}}}\right) * 100\right] - 100\right\}/\text{ny}$$

where ny = number of years. ~~Underlined and italicized rows indicate that the p-value associated with the gradient is > 0.1.~~ The penultimate two columns correspond to the result of a significance test performed on the difference between that row's trend and the trend in L3L and L3W, respectively, as follows:

$$Z = \frac{Trend_1 - Trend_2}{\sqrt{SE_1^2 + SE_2^2}}$$

where SE_1 and SE_2 correspond to the standard errors of $Trend_1$ and $Trend_2$ respectively, and Z is the test statistic. Where Z is greater (less) than 1.645 (-1.645) the trend difference is statistically significant to at least 90 % (i.e. $p < 0.1$). Drift = the measurement drift values given in Deeter et al., (2017). No values are given for the 900 and 300 hPa levels of the profile: we therefore cite values for the 400 and 200 hPa levels to give context to the 300 hPa trends we show; and we expect that the 900 hPa level drift is somewhere between that of the surface and 800 hPa levels, which are both shown.

† Units for TCO are mol cm^{-2}

Season	Level	Timeseries	m	SE	p value	% change y^{-1}
DJF	Surface	$L3_O$	-1.97	0.31	0.000	-0.99
		$L3_W$	-2.05	0.35	0.000	-1.02
n $L3_O$: 130		$L3_L$	-1.59	0.81	0.075	-0.85
n $L3_W$: 127	900hPa	$L3_O$	-2.77	0.31	0.000	-1.41
n $L3_L$: 32		$L3_W$	-2.83	0.35	0.000	-1.43
		$L3_L$	-2.42	0.79	0.011	-1.35
	800hPa	$L3_O$	-2.67	0.29	0.000	-1.55
		$L3_W$	-2.64	0.33	0.000	-1.53
		$L3_L$	-2.44	0.65	0.003	-1.54
	600hPa	$L3_O$	-0.95	0.44	0.047	-0.74
		$L3_W$	-0.80	0.43	0.083	-0.63
		$L3_L$	-1.24	0.62	0.071	-0.98
	300hPa	$L3_O$	1.01	0.38	0.019	1.34
		$L3_W$	1.09	0.34	0.006	1.44
		$L3_L$	*0.86*	*0.95*	*0.387*	*1.07*
	TCO	$L3_O$	-1.67E+16	4.40E+15	0.002	-0.64
		$L3_W$	-1.58E+16	4.50E+15	0.003	-0.61
		$L3_L$	-1.84E+16	8.20E+15	0.045	-0.74
JJA	Surface	$L3_O$	-1.16	0.32	0.003	-0.58
		$L3_W$	-0.51	0.28	0.092	-0.26
n $L3_O$: 248		$L3_L$	-3.28	0.68	0.000	-1.54
n $L3_W$: 243	900hPa	$L3_O$	-1.85	0.46	0.001	-1.08
n $L3_L$: 106		$L3_W$	-1.64	0.42	0.001	-0.96
		$L3_L$	-3.34	0.78	0.001	-1.83
	800hPa	$L3_O$	-1.80	0.51	0.003	-1.25
		$L3_W$	-1.97	0.58	0.004	-1.34
		$L3_L$	-2.09	0.60	0.003	-1.45
	600hPa	*$L3_O$*	*-0.55*	*0.43*	*0.225*	*-0.51*
		$L3_W$	-0.86	0.47	0.088	-0.78
		$L3_L$	*0.03*	*0.45*	*0.942*	*0.03*
	300hPa	$L3_O$	2.19	0.44	0.000	2.80
		$L3_W$	2.12	0.44	0.000	2.65
		$L3_L$	3.38	1.00	0.006	4.31
	TCO	*$L3_O$*	*-3.84E+15*	*6.30E+15*	*0.551*	*-0.16*
		$L3_W$	*-4.54E+15*	*5.90E+15*	*0.454*	*-0.19*
		$L3_L$	*-2.12E+15*	*8.60E+15*	*0.810*	*-0.09*

DJF	Level	Timeseries	Trend (ppbv)	Standard Error (ppbv)	P-value	% change y^{-1}	Sig. diff. to: L3L	Sig. diff. to: L3W	Drift (% y^{-1})
nL3W = 122	Surface	L3W	-1.63	0.33	0.000	-0.84	No	n/a	
nL3L = 31		L3L	-2.11	0.6	0.006	-1.13	n/a	No	
nL3O = 125		L3O	-1.74	0.31	0.000	-0.90	No	No	-0.69 ± 0.10
nL3O$_{(water)}$ = 101		L3O$_{(water)}$	-1.92	0.33	0.000	-0.32	No	No	
nL3O$_{(mixed)}$ = 23		L3O$_{(mixed)}$	-1.29	0.28	0.001	-0.68	No	No	
	900hPa	L3W	-2.39	0.29	0.000	-1.26	No	n/a	
		L3L	-3.14	0.48	0.000	-1.72	n/a	No	
		L3O	-2.51	0.29	0.000	-1.32	No	No	Not given
		L3O$_{(water)}$	-2.4	0.25	0.000	-1.26	No	No	
		L3O$_{(mixed)}$	-3.05	0.35	0.000	-1.61	No	No	
	800hPa	L3W	-2.36	0.3	0.000	-1.41	No	n/a	
		L3L	-2.9	0.35	0.000	-1.79	n/a	No	
		L3O	-2.47	0.29	0.000	-1.48	No	No	-1.04 ± 0.11
		L3O$_{(water)}$	-2.49	0.27	0.000	-1.47	No	No	
		L3O$_{(mixed)}$	-3.56	0.42	0.000	-2.08	No	Yes	
	600hPa	L3W	-0.962	0.39	0.027	-0.75	No	n/a	
		L3L	-2.1	0.63	0.007	-1.58	n/a	No	
		L3O	-1.21	0.41	0.011	-0.93	No	No	-0.33 ± 0.09
		L3O$_{(water)}$	-1.04	0.41	0.025	-0.80	No	No	
		L3O$_{(mixed)}$	-1.48	0.43	0.007	-1.18	No	No	
	300hPa	L3W	0.704	0.3	0.034	0.87	Yes	n/a	
		L3L	-0.507	0.65	0.456	-0.56	n/a	Yes	400 hPa: 1.15 ± 0.12
		L3O	0.58	0.35	0.122	0.71	No	No	
		L3O$_{(water)}$	0.49	0.20	0.030	0.59	No	No	200 hPa: 1.49 ± 0.13
		L3O$_{(mixed)}$	0.483	0.65	0.475	0.59	No	No	
	TCO	L3W	-2.16E+16	4.70E+15	0.000	-0.81	No	n/a	
		L3L	-2.27E+16	6.00E+15	0.000	-0.89	n/a	No	
		L3O	-2.56E+16	4.70E+15	0.000	-0.95	No	No	0.001 ± 0.001
		L3O$_{(water)}$	-2.08E+16	5.10E+15	0.001	-0.79	No	No	
		L3O$_{(mixed)}$	-3.24E+16	3.50E+15	0.000	-1.19	No	Yes	
JJA									
nL3W = 231	Surface	L3W	-0.69	0.33	0.058	-0.35	Yes	n/a	
nL3L = 101		L3L	-2.85	0.60	0.000	-1.42	n/a	Yes	
nL3O = 235		L3O	-0.99	0.28	0.003	-0.50	Yes	No	-0.69 ± 0.10
nL3O$_{(water)}$ = 136		L3O$_{(water)}$	-0.62	0.31	0.069	-0.32	Yes	No	
nL3O$_{(mixed)}$ = 98		L3O$_{(mixed)}$	-1.35	0.35	0.002	-0.68	Yes	No	
	900hPa	L3W	-1.34	0.31	0.001	-0.81	Yes	n/a	
		L3L	-2.58	0.60	0.001	-1.55	n/a	Yes	
		L3O	-1.43	0.35	0.001	-0.87	Yes	No	Not given
		L3O$_{(water)}$	-0.91	0.38	0.032	-0.56	Yes	No	
		L3O$_{(mixed)}$	-2.05	0.45	0.001	-1.23	No	No	
	800hPa	L3W	-1.55	0.41	0.002	-1.10	No	n/a	
		L3L	-1.62	0.54	0.010	-1.22	n/a	No	
		L3O	-1.47	0.39	0.002	-1.07	No	No	-1.04 ± 0.11
		L3O$_{(water)}$	-1.01	0.46	0.049	-0.75	No	No	
		L3O$_{(mixed)}$	-1.89	0.48	0.002	-1.38	No	No	
	600hPa	L3W	-0.91	0.46	0.071	-0.83	No	n/a	
		L3L	-0.38	0.47	0.440	-0.36	n/a	No	
		L3O	-0.75	0.43	0.105	-0.70	No	No	-0.33 ± 0.09
		L3O$_{(water)}$	-0.41	0.49	0.424	-0.39	No	No	
		L3O$_{(mixed)}$	-1.12	0.45	0.026	-1.02	No	No	
	300hPa	L3W	1.30	0.43	0.010	1.40	No	n/a	
		L3L	2.51	1.30	0.072	2.90	n/a	No	400 hPa: 1.15 ± 0.12
		L3O	1.38	0.47	0.011	1.52	No	No	
		L3O$_{(water)}$	0.86	0.42	0.060	0.92	No	No	200 hPa: 1.49 ± 0.13
		L3O$_{(mixed)}$	2.24	0.53	0.001	2.52	No	No	
	TCO	L3W	-9.84E+15	5.70E+15	0.109	-0.40	No	n/a	
		L3L	-1.01E+16	1.00E+16	0.331	-0.42	n/a	No	
		L3O	-8.75E+15	6.40E+15	0.192	-0.36	No	No	0.001 ± 0.001
		L3O$_{(water)}$	-7.62E+15	6.80E+15	0.282	-0.32	No	No	
		L3O$_{(mixed)}$	-1.18E+16	6.90E+15	0.109	-0.48	No	No	

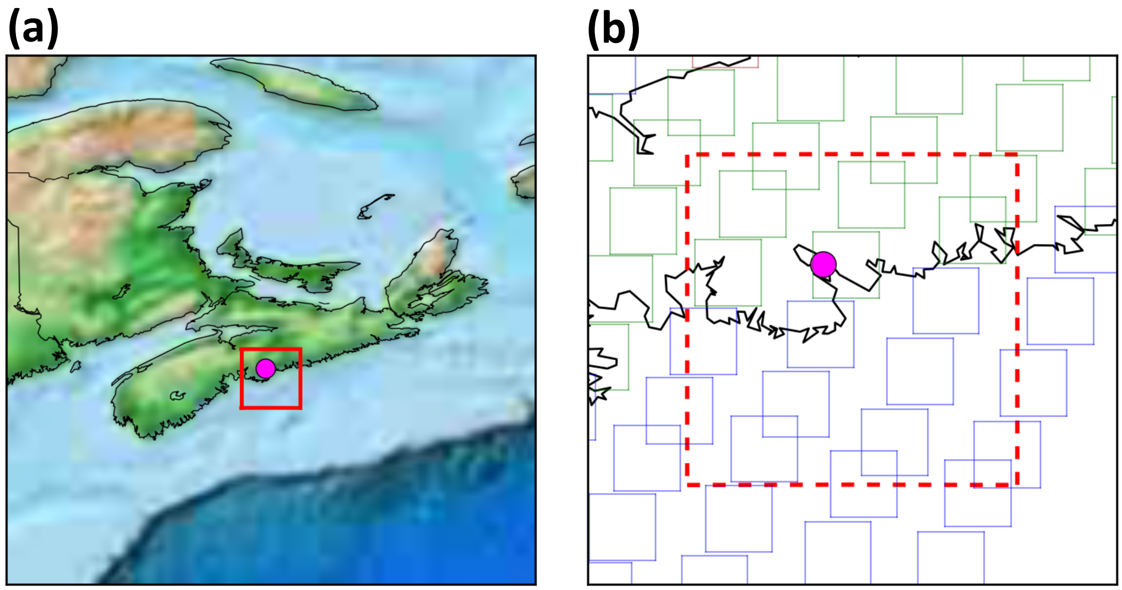
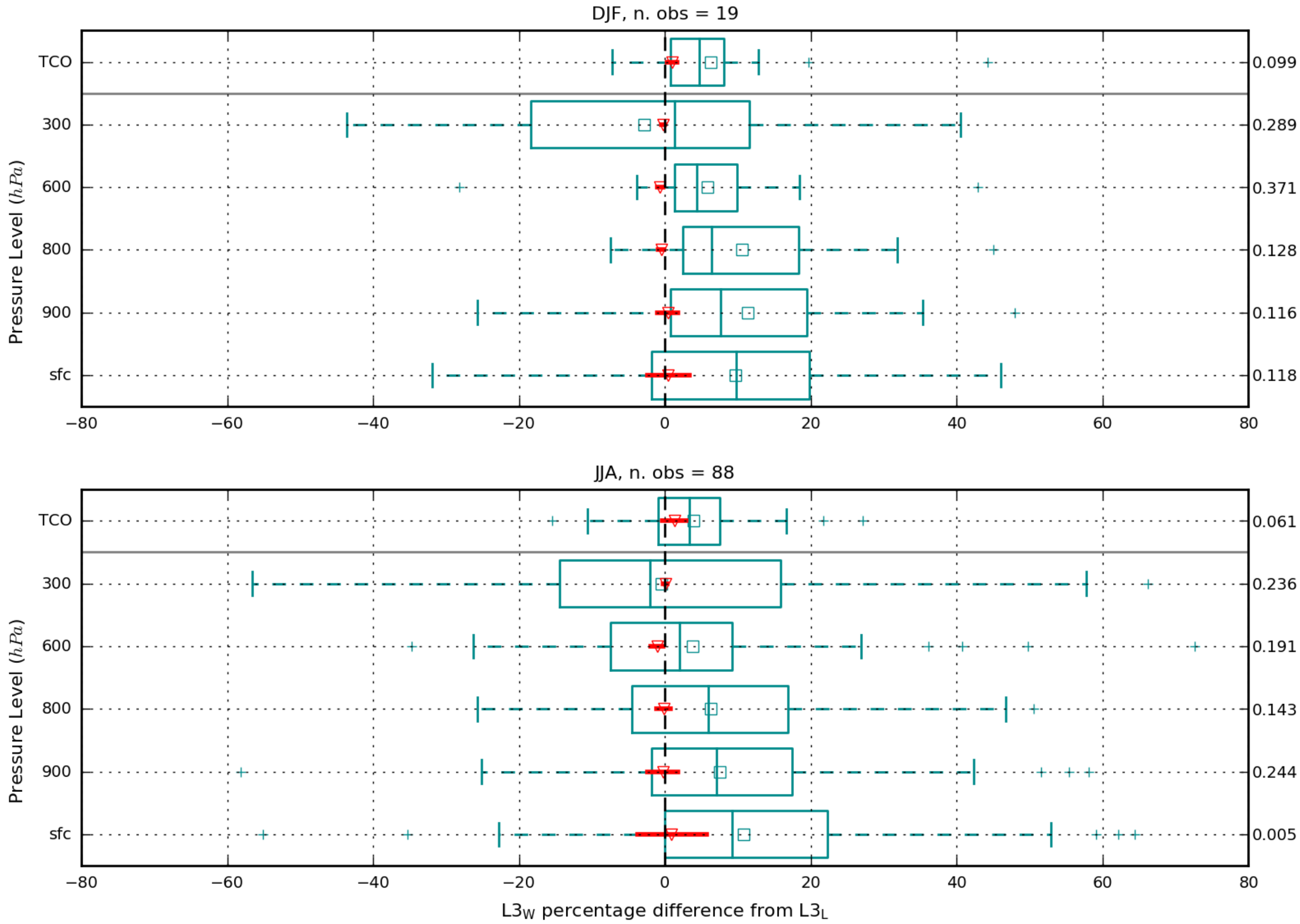


Figure 1 (a) Map of Nova scotia and the surrounding region in Atlantic Canada. Lighter colours on land indicate higher elevation. The pink icon shows the location of Halifax, the coastal city that we focus on in this study (see Section 2.1.3), and the red box shows the MOPITT L3 1^{o} x 1^{o} gridbox containing Halifax. **(b)** Map zoomed to the MOPITT L3 gridbox containing Halifax (red dashed box), with the approximate location of individual L2 retrieval footprints shown (blue boxes = L2 surface index of water, green boxes = L2 surface index of land). L2 retrieval footprints with a midpoint that falls within the boundaries of the L3 gridbox will be averaged together to create the L3 data, according to certain rules – see Section 2.1.2 for full explanation.


DJF, n. obs = 19
TCO
300
600
800
900
sfc
Pressure Level (hPa)
0.099
0.289
0.371
0.128
0.116
0.118
−80
−60
−40
−20
0
20
40
60
80
JJA, n. obs = 88
TCO
300
600
800
900
sfc
Pressure Level (hPa)
0.061
0.236
0.191
0.143
0.244
0.005
−80
−60
−40
−20
0
20
40
60
80
$L3_W$ percentage difference from $L3_L$

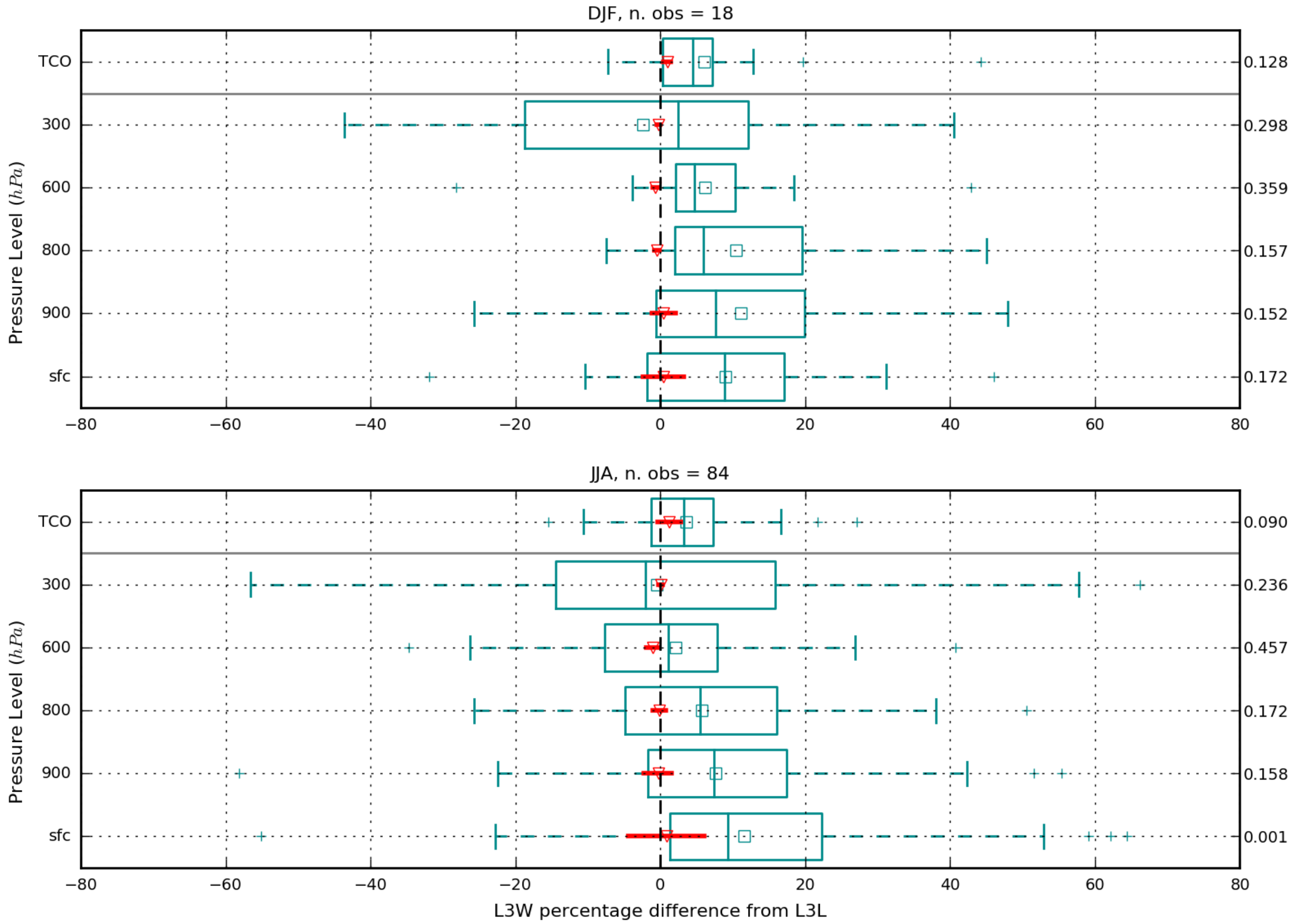
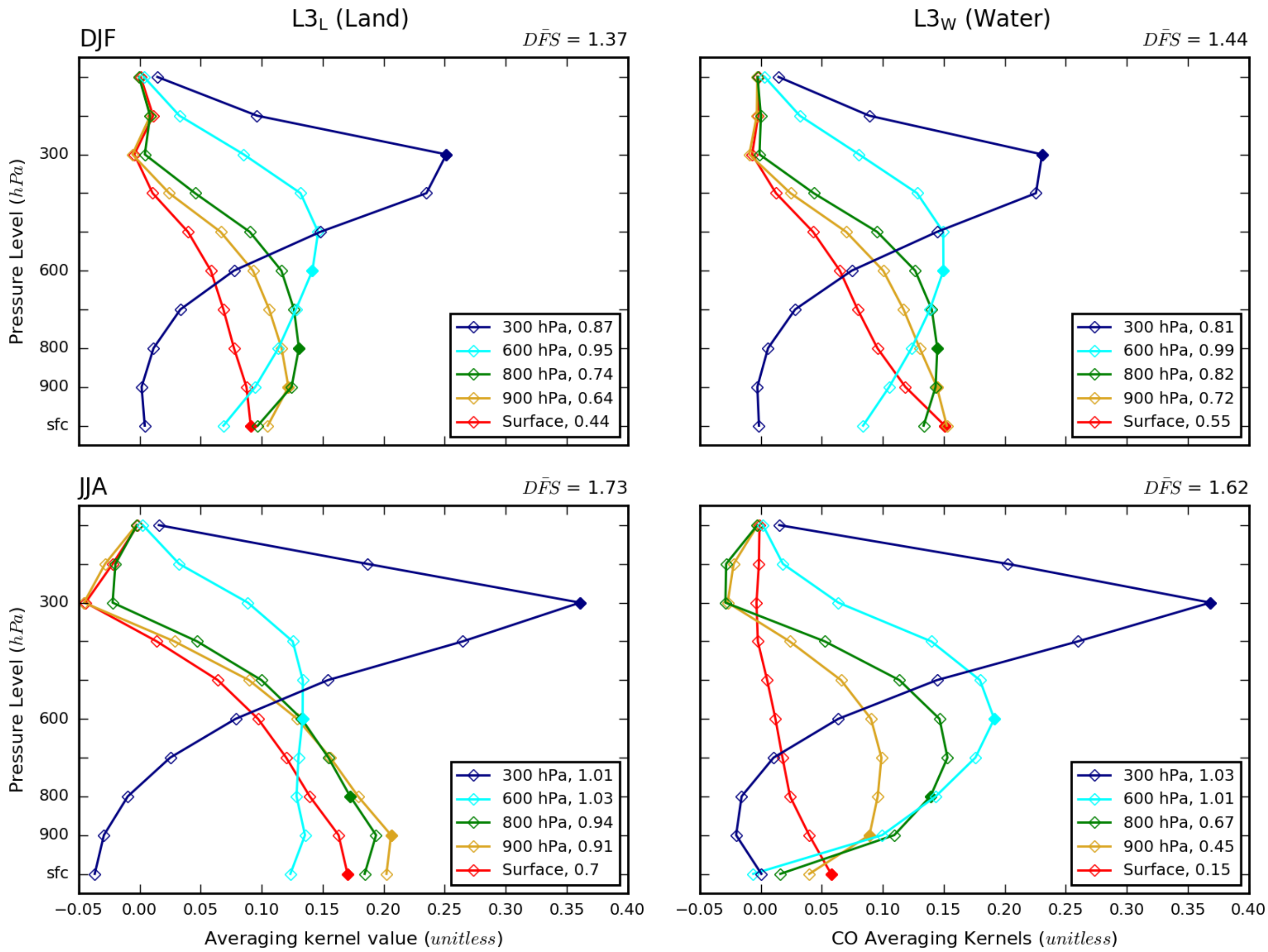


Figure 2 Distribution of percentage difference† between temporally coincident retrieved VMRs (selected profile levels) and CO total column (TCO) from L3W and L3L. Squares = mean differences, with the p value associated with each mean difference (from a 2-tailed Student's t test) given on the right-hand side y axis. Plus symbols = outliers‡. Red triangles = mean percentage difference between a priori values. Red lines = range of a priori difference values (where barely visible this means range is very small).

†Method for calculating percentage differences: $\left\{\left(\frac{L3_W}{L3_L}\right) * 100\right\} - 100$

‡ Outliers defined as: above (below) percentile 75 (25) + (-) 1.5 * interquartile range



L3L (Land)
L3W (Water)
DJF
JJA
DFS = 1.37
DFS = 1.44
DFS = 1.73
DFS = 1.62
Pressure Level (hPa)
300
600
800
900
sfc
300 hPa, 0.87
600 hPa, 0.95
800 hPa, 0.74
900 hPa, 0.64
Surface, 0.44
300 hPa, 0.81
600 hPa, 0.99
800 hPa, 0.82
900 hPa, 0.72
Surface, 0.55
300 hPa, 1.01
600 hPa, 1.03
800 hPa, 0.94
900 hPa, 0.91
Surface, 0.7
300 hPa, 1.03
600 hPa, 1.01
800 hPa, 0.67
900 hPa, 0.45
Surface, 0.15
−0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Averaging kernel value (unitless)
CO Averaging Kernels (unitless)

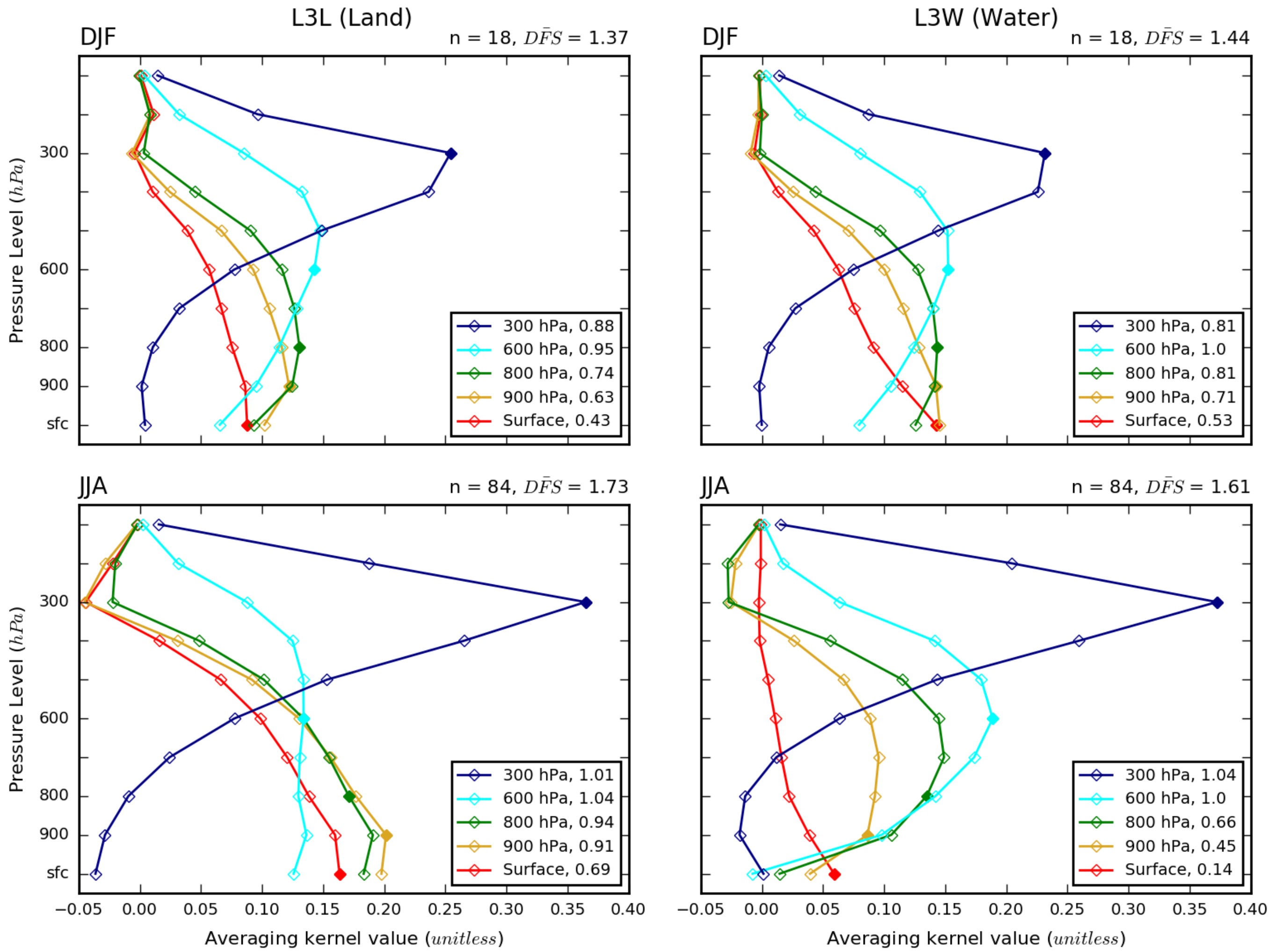
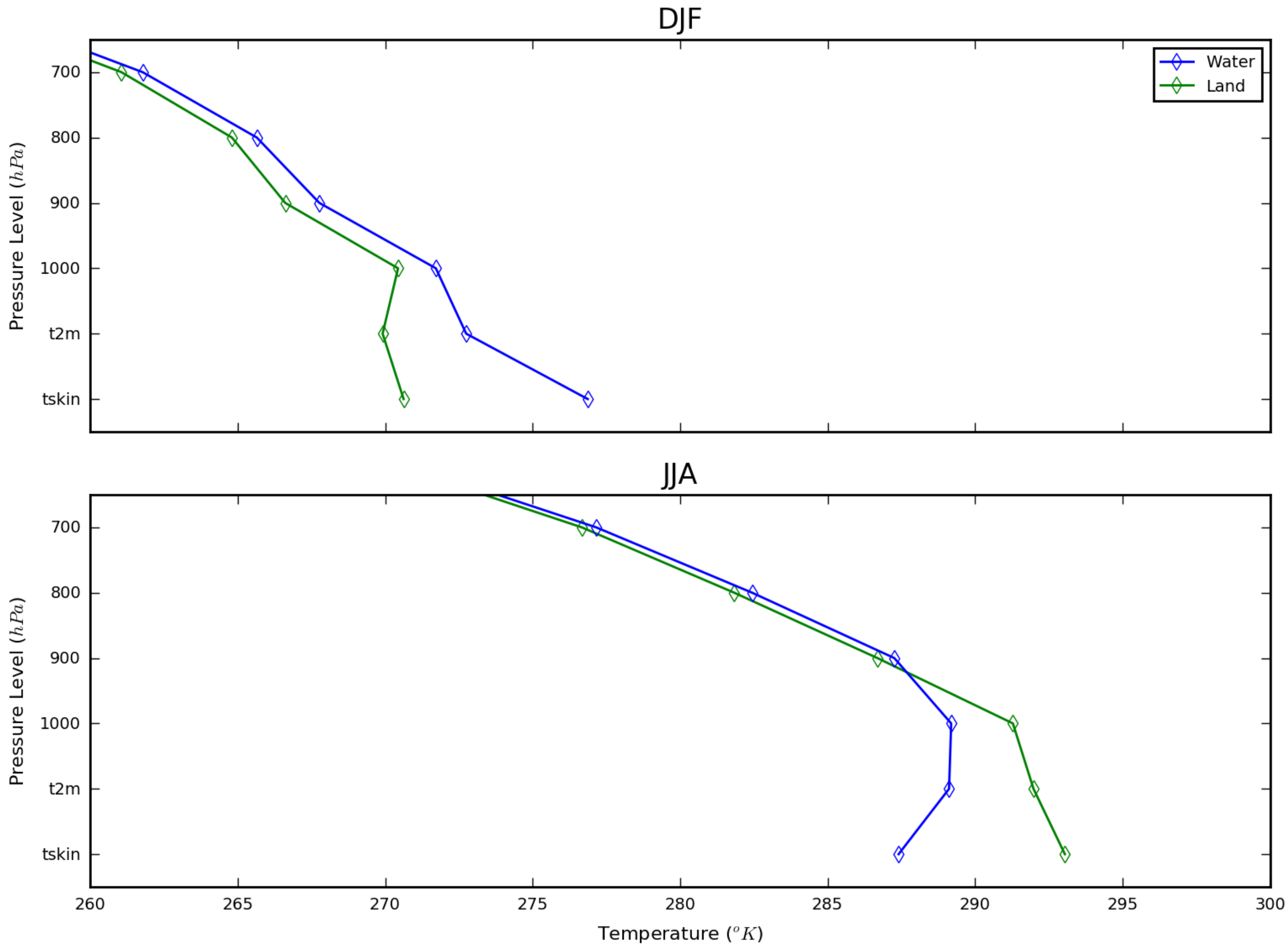


Figure 3 Mean retrieval averaging kernel (AK) rows and DFS values for L3L ("Land", left column) and L3W ("Water", right column) in DJF (top row~~; n = 19~~) and JJA (bottom row~~; n = 88~~), selected profile levels only. Filled diamonds indicate diagonal value location for that AK. Numbers in legend indicate corresponding retrieval level of AK and show the mean rowsum for that AK level.


DJF
Water
Land
Pressure Level (hPa)
700
800
900
1000
t2m
tskin
JJA
Pressure Level (hPa)
700
800
900
1000
t2m
tskin
260
265
270
275
280
285
290
295
300
Temperature (^oK)

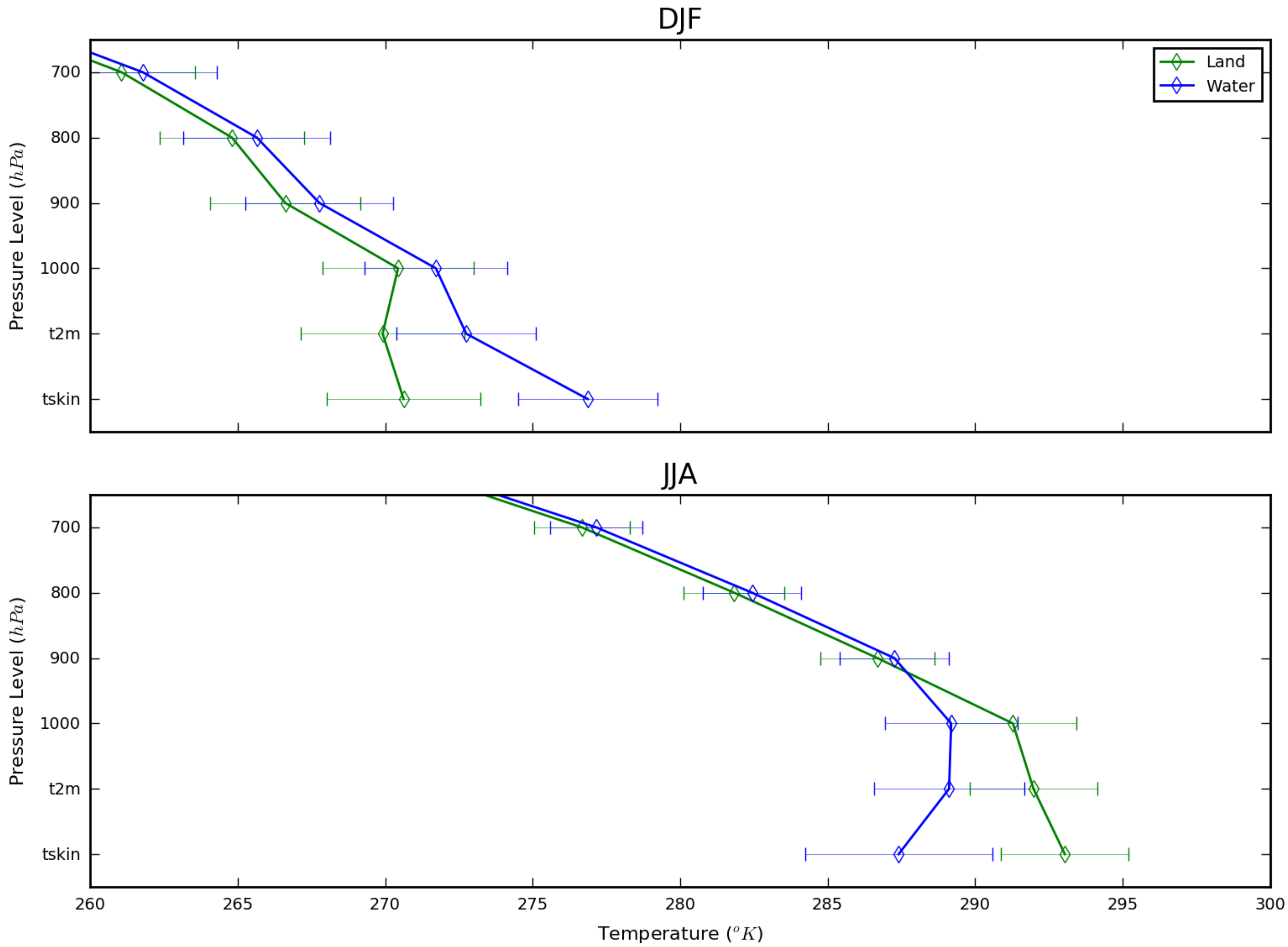


Figure 4 Mean temperature profile from ERA-Interim for the closest all-land and all-water model gridboxes to Halifax. Errorbars = standard deviation. Data are the mean of the 12:00 UTC and 18:00 UTC timesteps (this corresponds to 08:00 and 14:00 local time, the closest ERA Interim timesteps to local MOPITT overpass time of ~10:30), for 2001~~0~~-2017.

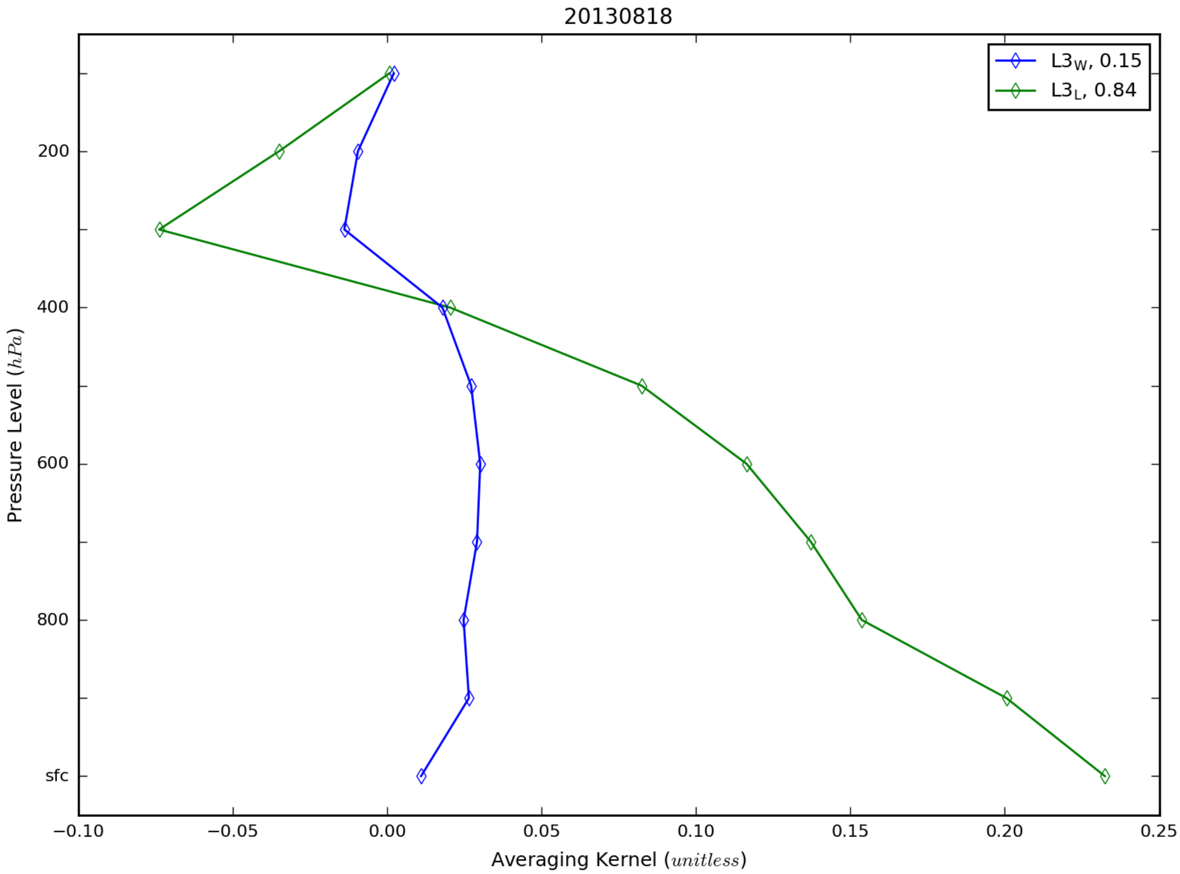
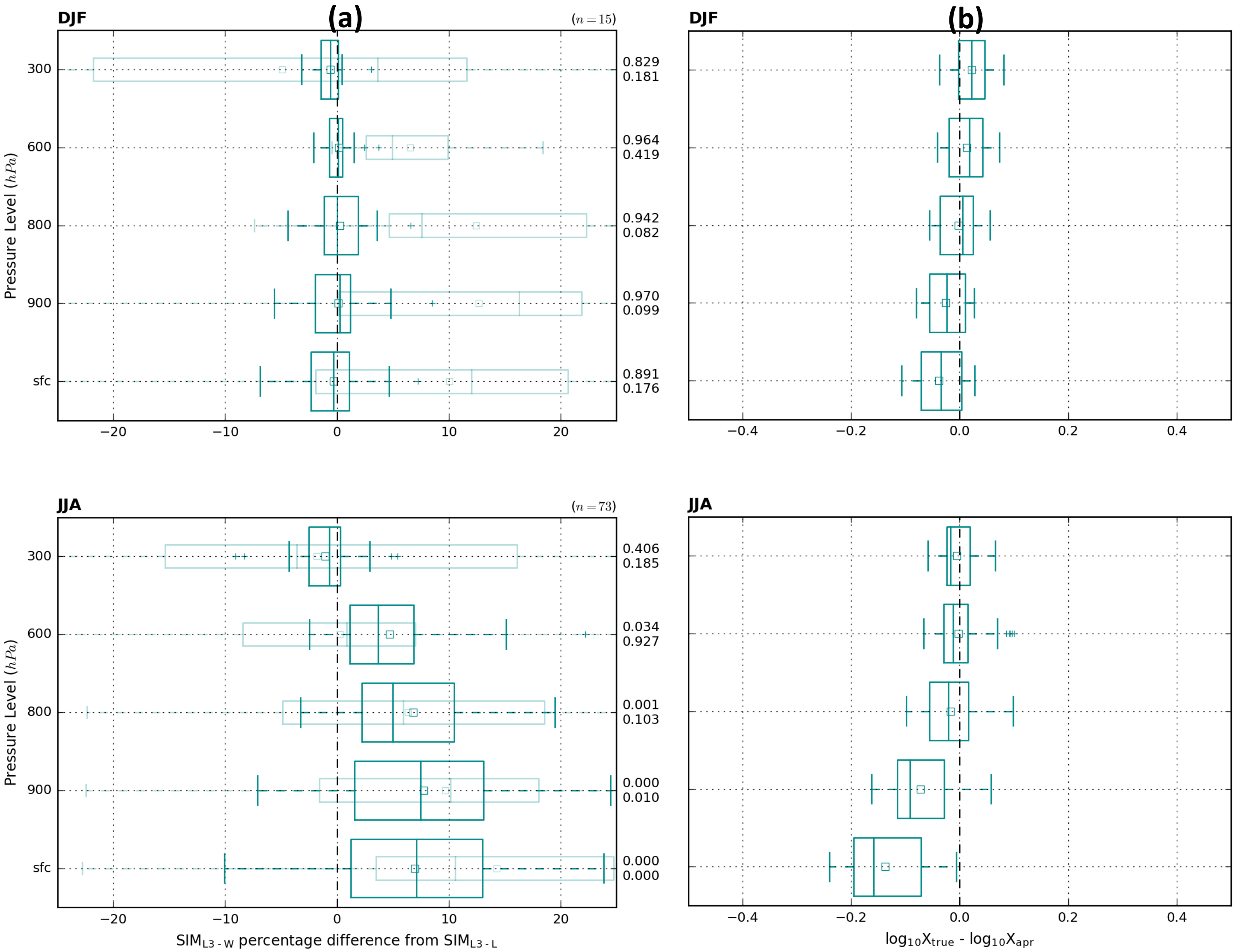


Figure 5 Surface-level averaging kernels corresponding to $L3_L$ (green lines) and $L3_W$ (blue lines) for 2013-08-18 case study. Numbers in legend show the rowsum for that averaging kernel (AK_{rowsum}).



(a)
(b)
DJF
(n = 15)
JJA
(n = 73)
Pressure Level (hPa)
300
600
800
900
sfc
0.829
0.181
0.964
0.419
0.942
0.082
0.970
0.099
0.891
0.176
0.406
0.185
0.034
0.927
0.001
0.103
0.000
0.010
0.000
0.000
−20 −10 0 10 20
−0.4 −0.2 0.0 0.2 0.4
SIM L3-W percentage difference from SIM L3-L
log10Xtrue - log10Xapr

.105

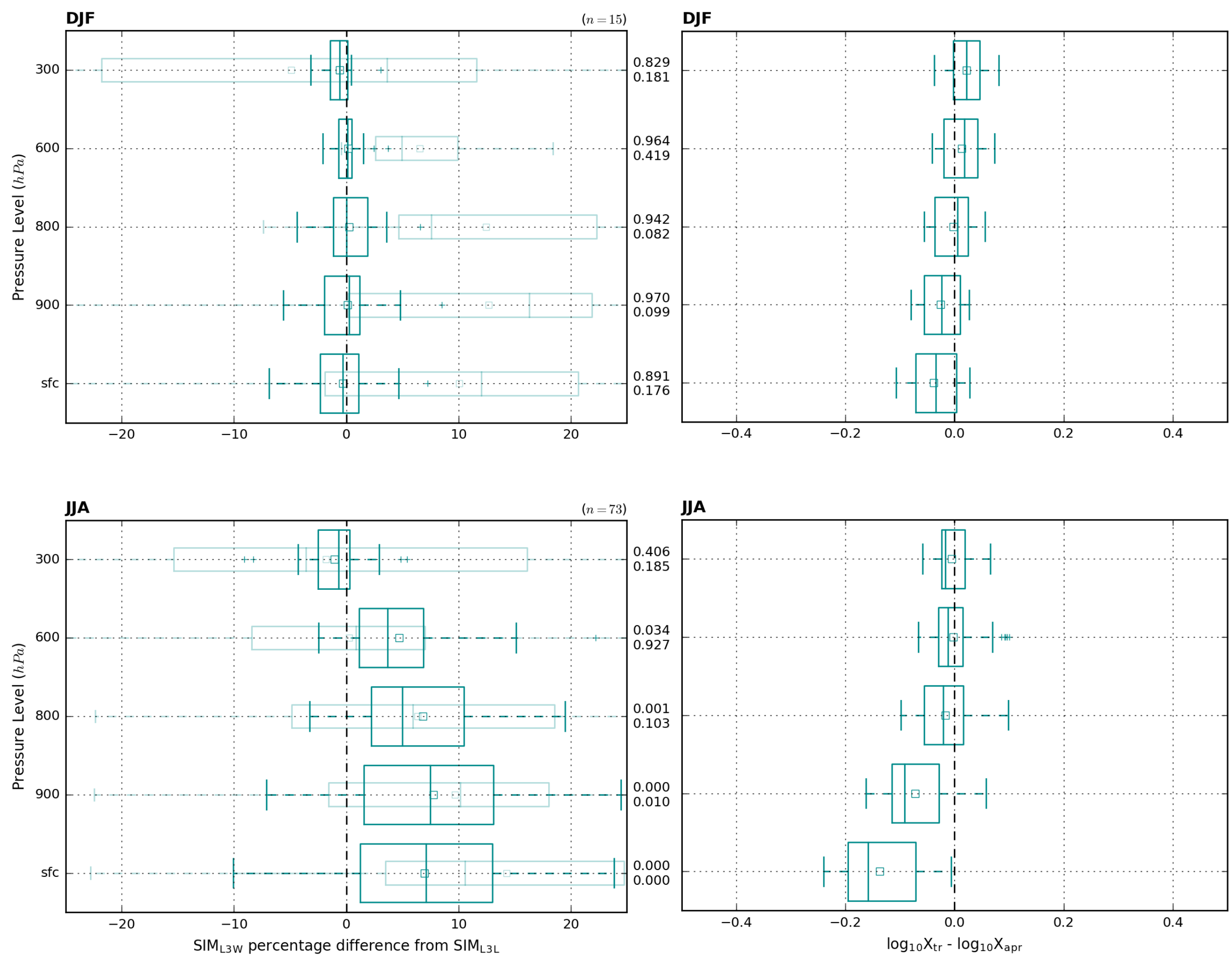
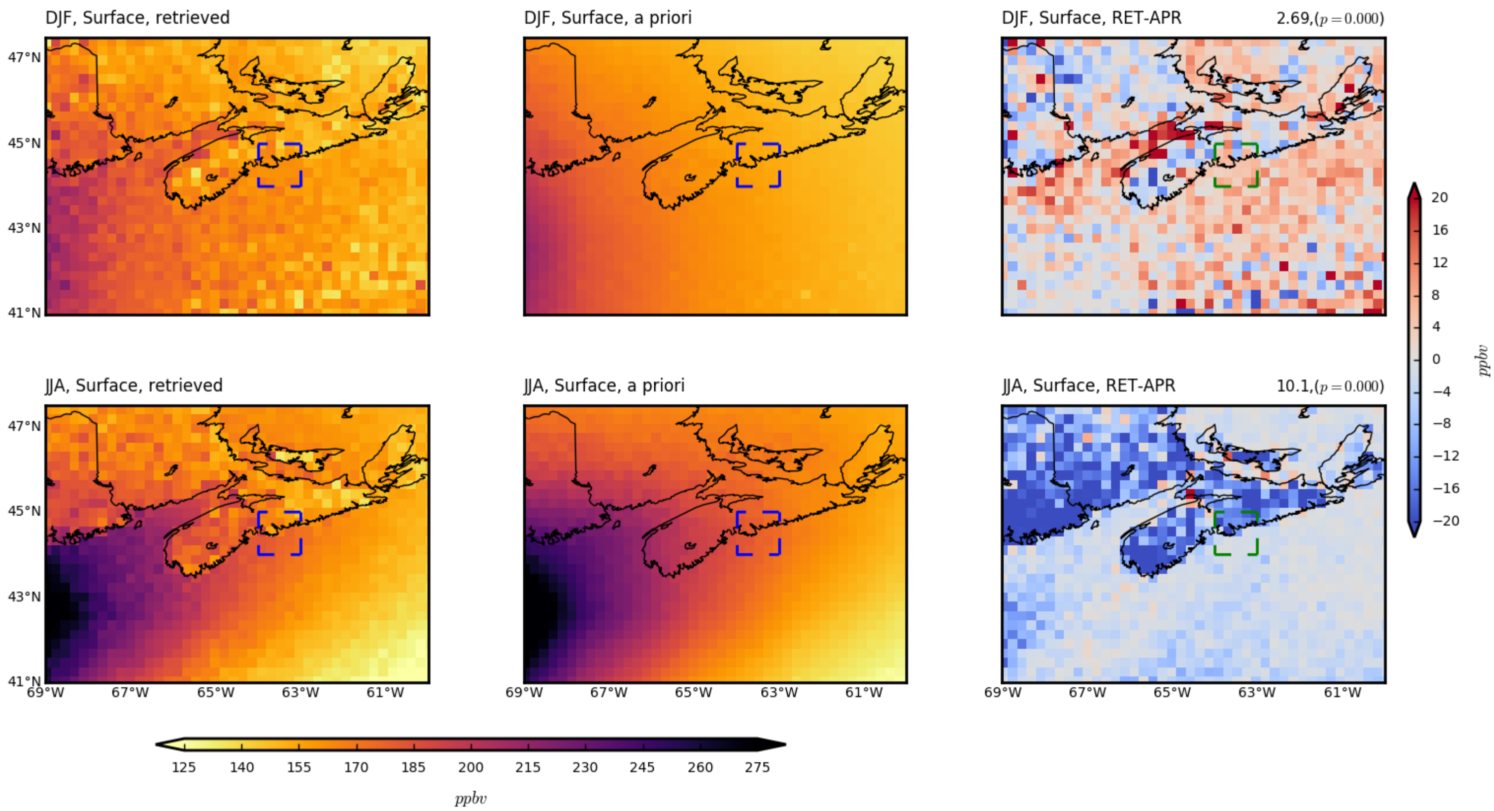


Figure 5~~6 (Next page)~~ **(a)** Distribution of percentage difference† between simulated temporally coincident VMR retrievals in L3W (SIM_{L3-W}) and L3L (SIM_{L3-L}). Squares = mean differences. The p value associated with each mean difference (from a 2-tailed Student's t test) is given on the right-hand side y axis (top row = ΔSIM; bottom row = ΔRET). Plus symbols = outliers‡. Faint shading = corresponding ΔRET boxplots for comparison. Note that the sample size is different to Figure 2 owing to the CAMSRA data used as the $X_{true,sim}$ profile only covering a subset of the MOPITT years (2003-2016 vs 2001~~0~~-2017 in Figure 2). The ΔRET boxplots overlaid cover this shortened period. **(b)** Distribution of the $\log_{10}X_{true,sim} - \log_{10}X_{apr,sim}$ values calculated during the simulation of retrieved profiles (see Equation 2) in DJF (top row) and JJA (bottom row).

†Method for calculating percentage differences: $\left\{\left(\frac{L3_W}{L3_L}\right) * 100\right\} - 100$

‡ Outliers defined as: above (below) percentile 75 (25) + (-) 1.5 * interquartile range


DJF, Surface, retrieved
DJF, Surface, a priori
DJF, Surface, RET-APR
2.69,($p = 0.000$)
JJA, Surface, retrieved
JJA, Surface, a priori
JJA, Surface, RET-APR
10.1,($p = 0.000$)
47°N
45°N
43°N
41°N
69°W
67°W
65°W
63°W
61°W
20
16
12
8
4
0
−4
−8
−12
−16
−20
ppbv
125
140
155
170
185
200
215
230
245
260
275
ppbv

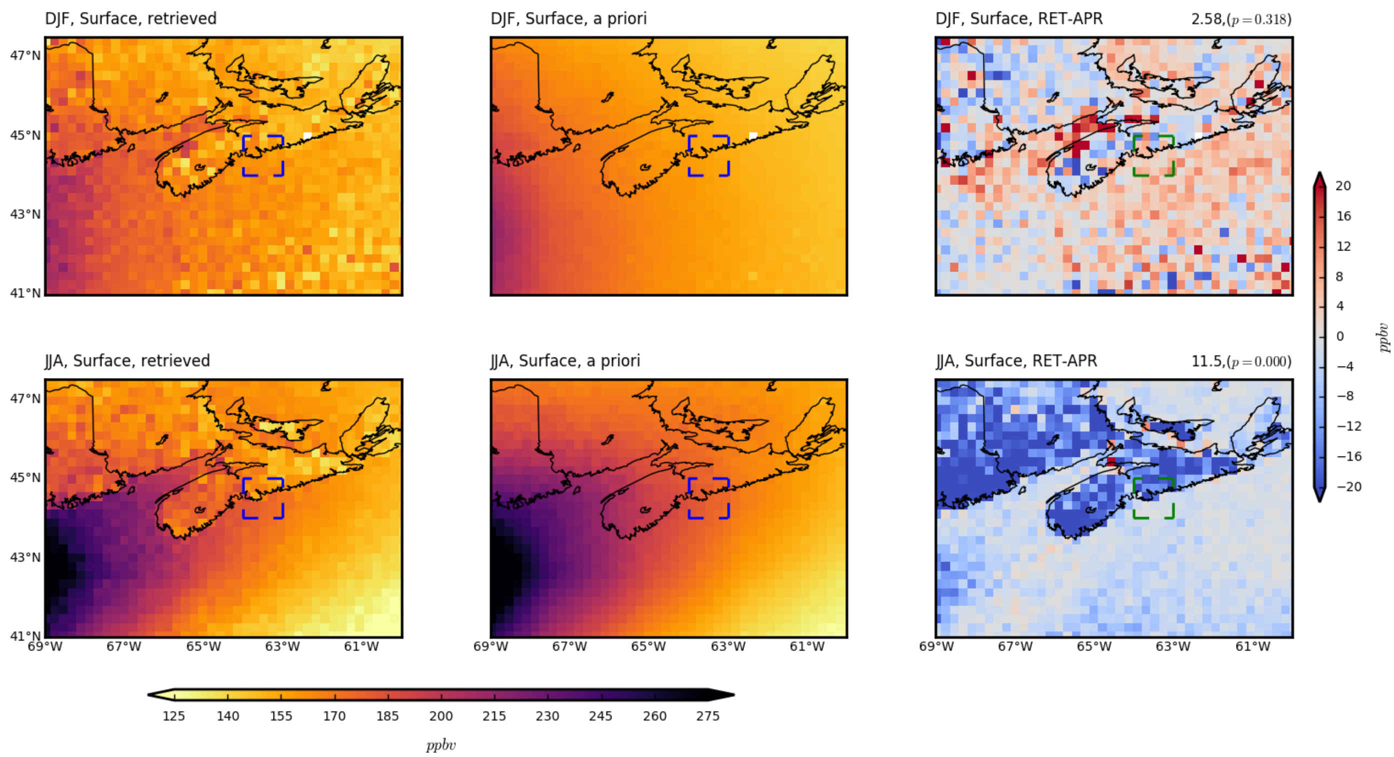
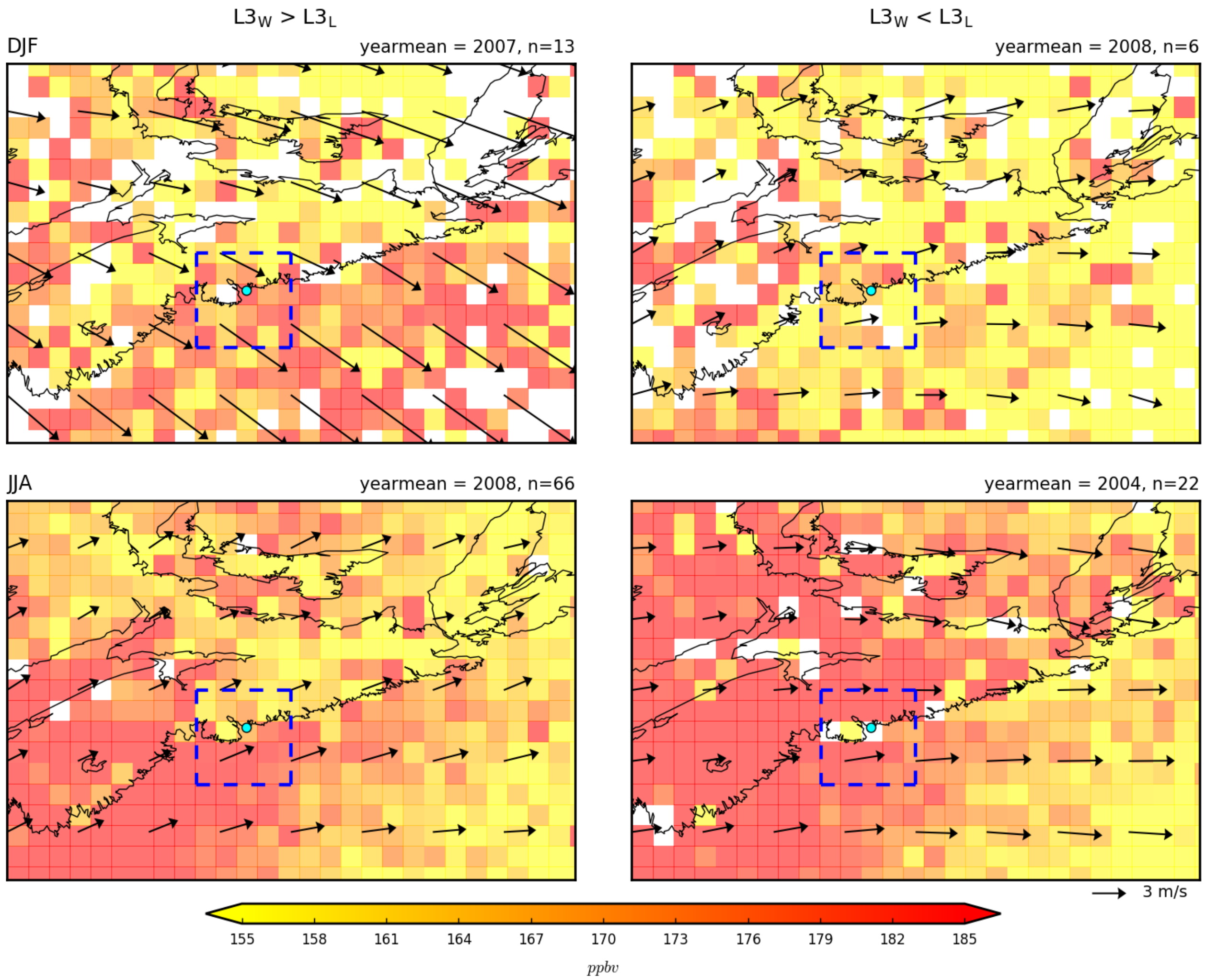

DJF, Surface, retrieved
DJF, Surface, a priori
DJF, Surface, RET-APR
2.58,($p = 0.318$)
JJA, Surface, retrieved
JJA, Surface, a priori
JJA, Surface, RET-APR
11.5,($p = 0.000$)
47°N
45°N
43°N
41°N
69°W
67°W
65°W
63°W
61°W
20
16
12
8
4
0
−4
−8
−12
−16
−20
ppbv
125
140
155
170
185
200
215
230
245
260
275
ppbv

Figure 6~~7~~ Seasonal median L2† retrieved VMR (left column), a priori VMR (centre column), and RET-APR (right column) at the surface profile level in DJF (top row) and JJA (bottom row). Values to the right above RET-APR plots = $\overline{(\text{L2 retrievals over water})} - \overline{(\text{L2 retrievals over land})}$ for plotted area (data were first binned according to L2 surface index); numbers in brackets correspond to significance of mean difference using a 2-tailed Student’s t test. Blue or green dashed square = outline of L3 grid box that contains Halifax.

†These maps were created from L2 data that were interpolated to a regular 0.25° x 0.25° grid for ease of plotting.


L3W > L3L
L3W < L3L
DJF
yearmean = 2007, n=13
yearmean = 2008, n=6
JJA
yearmean = 2008, n=66
yearmean = 2004, n=22
3 m/s
155
158
161
164
167
170
173
176
179
182
185
ppbv

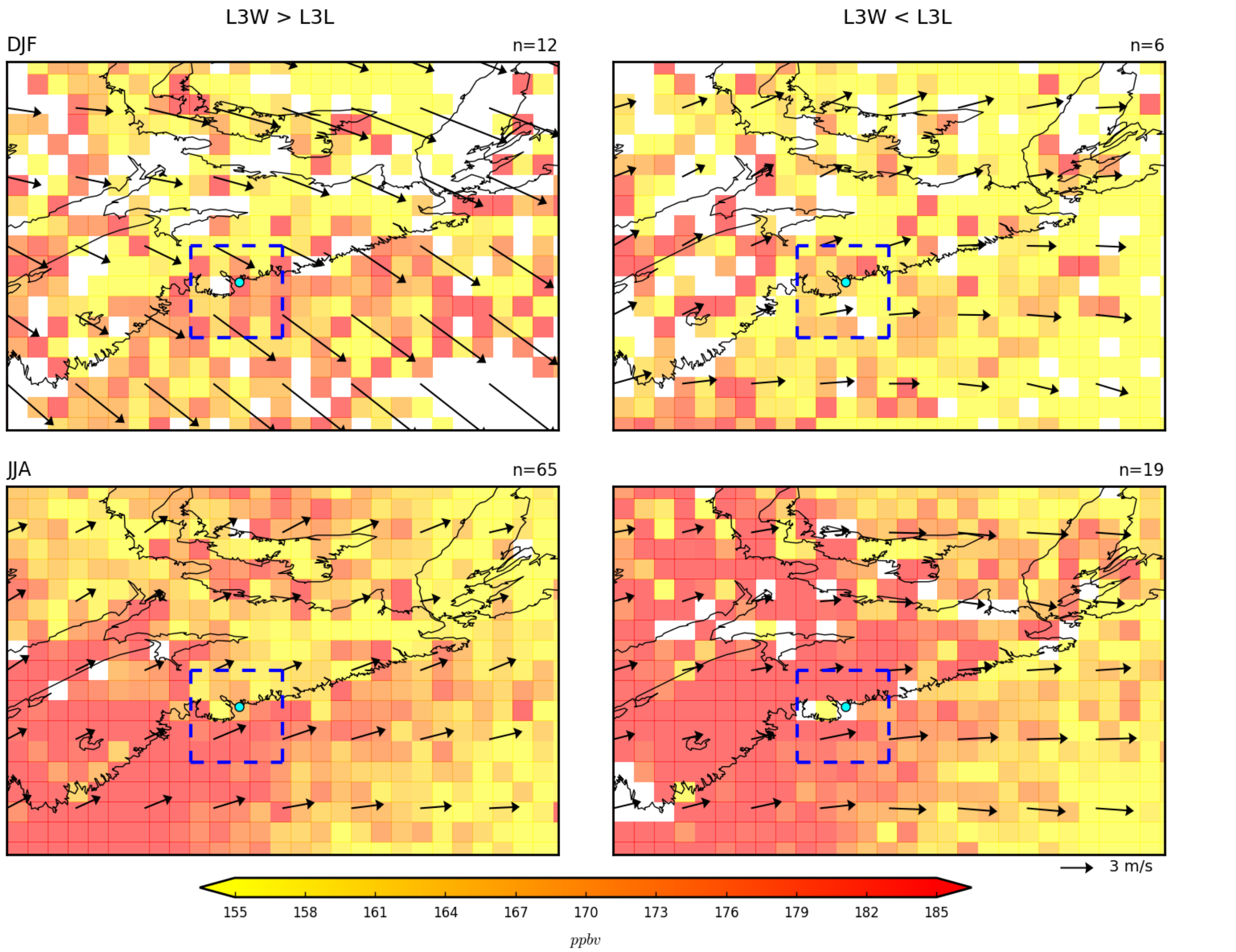
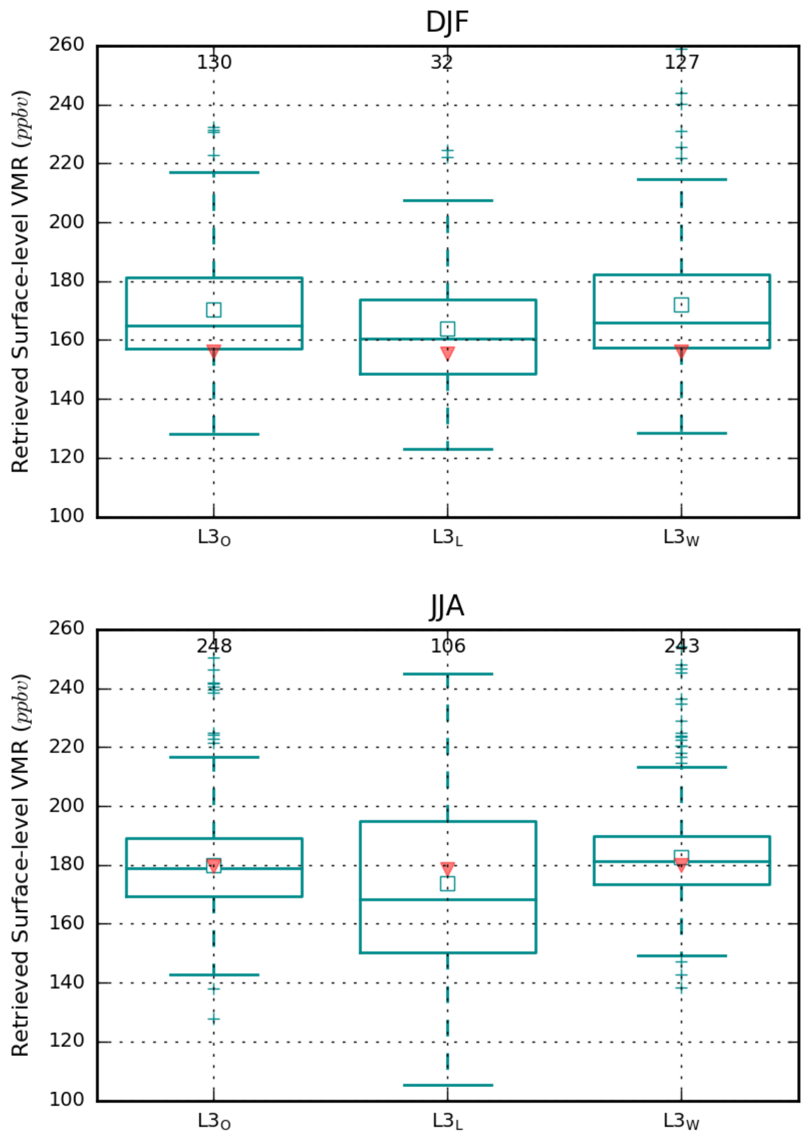


Figure 7~~8~~ Mean ERA-Interim 10-metre winds (vectors) and MOPITT L2[†] VMR at the surface profile level (shading) for days when retrieved surface level VMRs in L3W are greater than in L3L (L3W > L3L) and days when they are less (L3W < L3L). Top row = DJF; bottom row = JJA. ~~"Yearmean" corresponds to the mean year in which retrievals from the respective samples took place (study period spans 2000-2017).~~Blue dashed square = outline of L3 grid box that contains Halifax.

[†]These maps were created from L2 data that were interpolated to a regular 0.25° x 0.25° grid for ease of plotting.


DJF
130
32
127
Retrieved Surface-level VMR (ppbv)
260
240
220
200
180
160
140
120
100
$L3_O$
$L3_L$
$L3_W$
JJA
248
106
243
Retrieved Surface-level VMR (ppbv)
260
240
220
200
180
160
140
120
100
$L3_O$
$L3_L$
$L3_W$

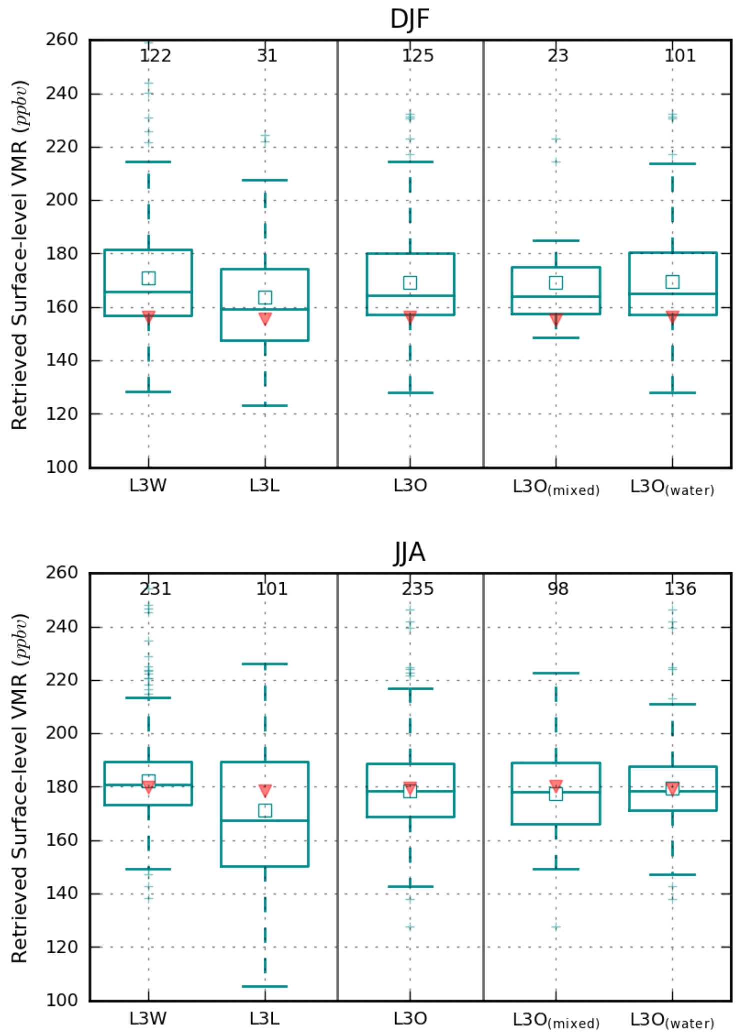
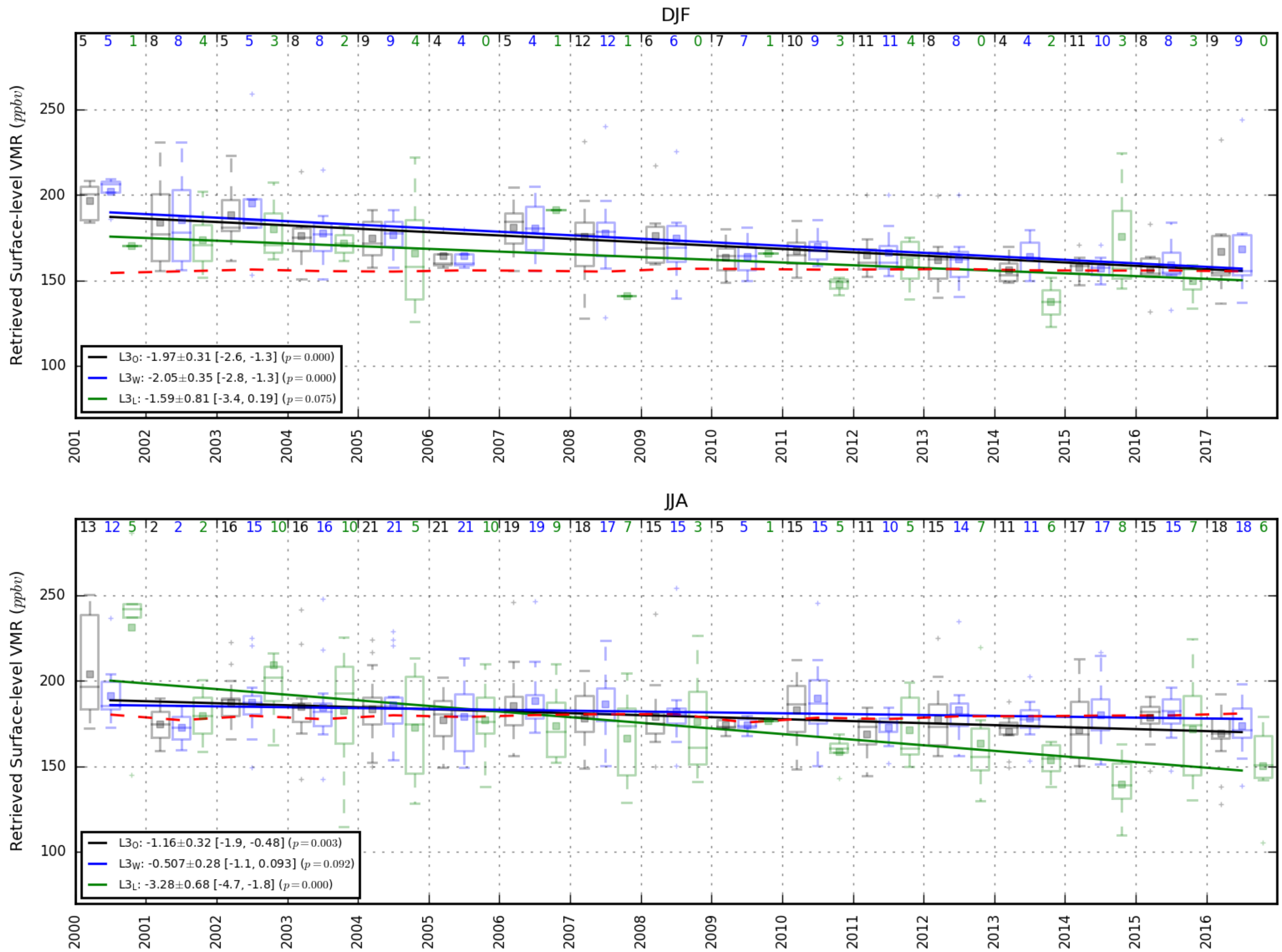


Figure 89 Boxplots showing the Sseasonal distribution of retrieved surface level VMR values from L3W, L3L, L3O, L3O$_{(mixed)}$ and L3O$_{(water)}$ L3$_L$ and L3$_W$. Squares = mean values; Red triangles = corresponding mean a priori values. Sample sizes are given below the top x axis.


DJF
Retrieved Surface-level VMR (ppbv)
L3O: -1.97±0.31 [-2.6, -1.3] (p=0.000)
L3W: -2.05±0.35 [-2.8, -1.3] (p=0.000)
L3L: -1.59±0.81 [-3.4, 0.19] (p=0.075)
JJA
Retrieved Surface-level VMR (ppbv)
L3O: -1.16±0.32 [-1.9, -0.48] (p=0.003)
L3W: -0.507±0.28 [-1.1, 0.093] (p=0.092)
L3L: -3.28±0.68 [-4.7, -1.8] (p=0.000)

150

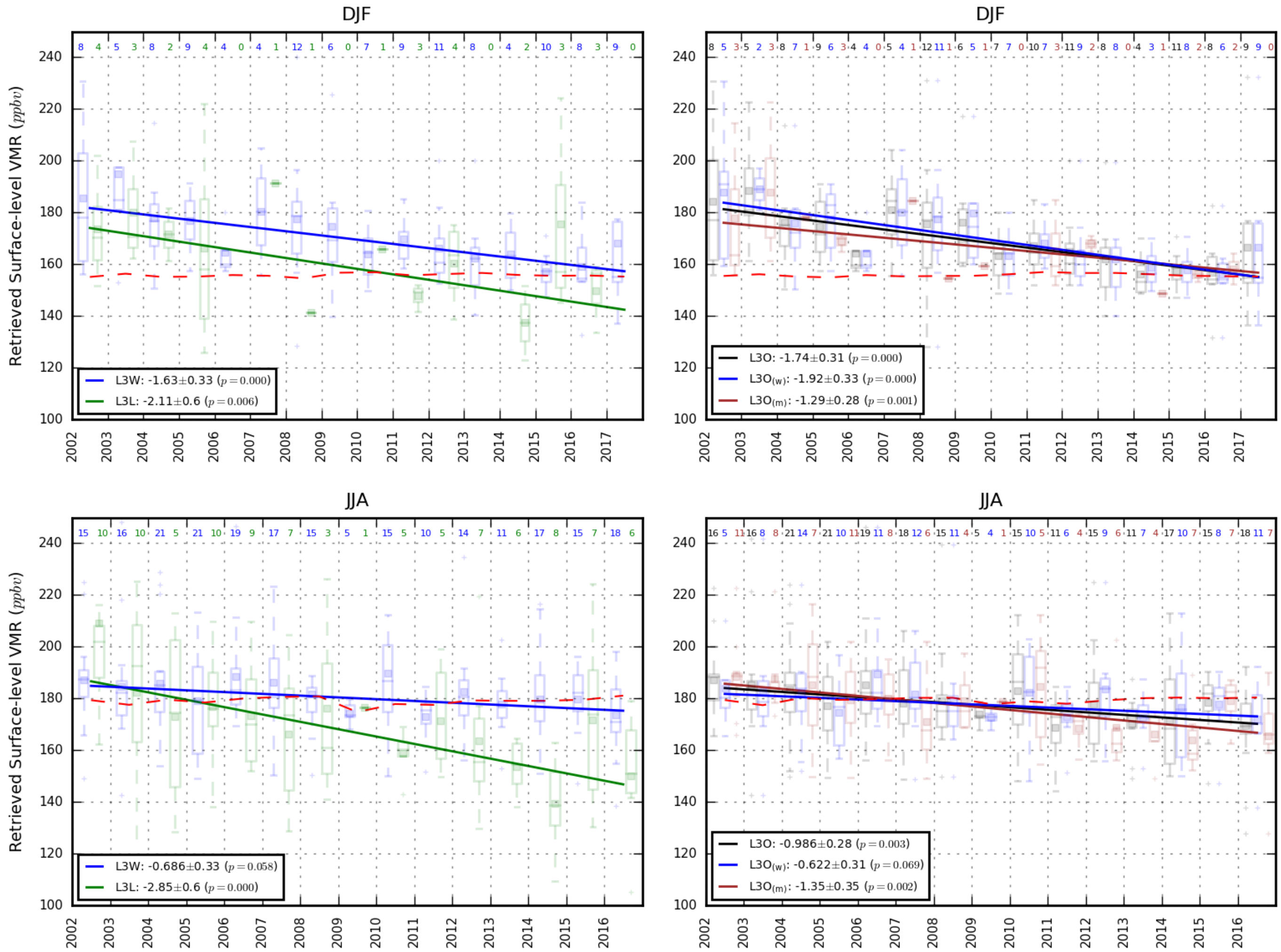


Figure 9 WLS regression best-fit lines calculated from seasonal mean retrieved surface VMR timeseries in DJF (top row) and JJA (bottom row). Left column = L3L (green) and L3W (blue); Right column = L3O (black), L3O$_{(mixed)}$ (brown) and L3O$_{(water)}$ (blue). The daily observations corresponding to each seasonal mean value are represented by colour-coded boxplots each year, and the seasonal mean value is represented by filled squares. The dashed red line is the mean of the corresponding seasonal mean a priori data from each of the timeseries in the respective panel. Colour-coded values below the top x axis correspond to the number of observations each season. Values in the legend are the value, standard error, and probability of zero value of the trend, respectively.

Figure 10 (next page) OLS regression best-fit lines calculated from seasonal mean timeseries for DJF (top) and JJA (bottom) of $L3_O$ (black), $L3_L$ (green) and $L3_W$ (blue) retrieved surface-level VMR. The daily observations corresponding to each seasonal mean value are represented by colour-coded boxplots each year (horizontal lines in middle of boxes = seasonal median value; filled squares = seasonal mean value). The dashed red line is the mean of the corresponding seasonal mean $L3_O$, $L3_L$ and $L3_W$ a priori data. Colour-coded values below the top x axis correspond to the number of observations each season. Values in the legend are the value, standard error, 95% confidence limits and probability of zero value of the gradient parameter, respectively.

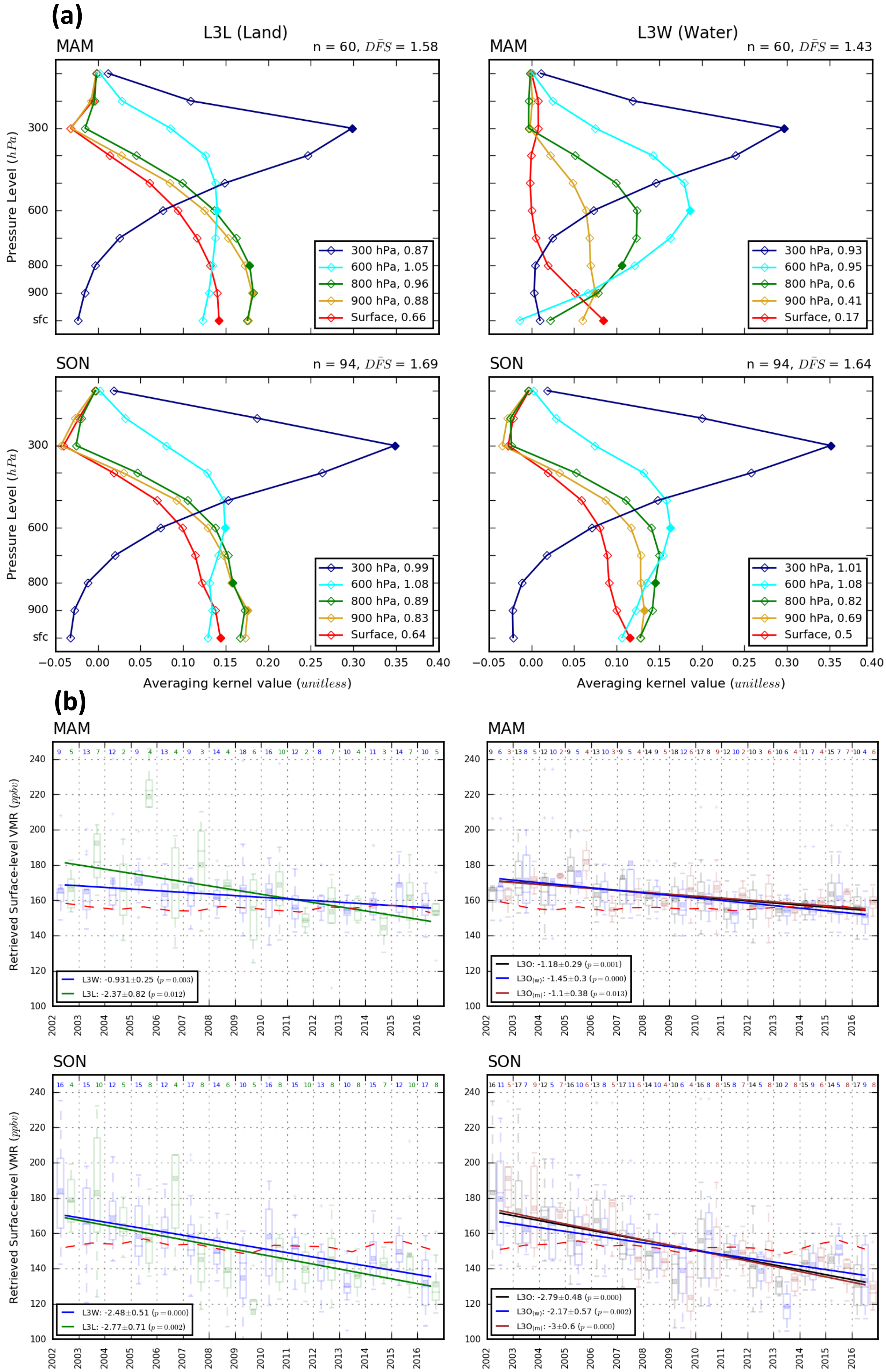

(a)
L3L (Land)
L3W (Water)
MAM
n = 60, $\bar{DFS}$ = 1.58
n = 60, $\bar{DFS}$ = 1.43
Pressure Level (hPa)
300
600
800
900
sfc
300 hPa, 0.87
600 hPa, 1.05
800 hPa, 0.96
900 hPa, 0.88
Surface, 0.66
300 hPa, 0.93
600 hPa, 0.95
800 hPa, 0.6
900 hPa, 0.41
Surface, 0.17
SON
n = 94, $\bar{DFS}$ = 1.69
n = 94, $\bar{DFS}$ = 1.64
300 hPa, 0.99
600 hPa, 1.08
800 hPa, 0.89
900 hPa, 0.83
Surface, 0.64
300 hPa, 1.01
600 hPa, 1.08
800 hPa, 0.82
900 hPa, 0.69
Surface, 0.5
−0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Averaging kernel value ($unitless$)
(b)
MAM
SON
Retrieved Surface-level VMR ($ppbv$)
100 120 140 160 180 200 220 240
L3W: -0.931±0.25 ($p=0.003$)
L3L: -2.37±0.82 ($p=0.012$)
L3O: -1.18±0.29 ($p=0.001$)
L3O(w): -1.45±0.3 ($p=0.000$)
L3O(m): -1.1±0.38 ($p=0.013$)
L3W: -2.48±0.51 ($p=0.000$)
L3L: -2.77±0.71 ($p=0.002$)
L3O: -2.79±0.48 ($p=0.000$)
L3O(w): -2.17±0.57 ($p=0.002$)
L3O(m): -3±0.6 ($p=0.000$)
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

Figure 10 (a) As Figure 3 but for MAM and SON; **(b)** As Figure 9 but for MAM and SON