https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

Differences in MOPITT surface-level CO retrievals and trends from Level 2 and Level 3 products in coastal grid boxes

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

MOPITT retrievals are more sensitive to near-surface CO when performed over land than water. Data users are therefore advised to discard retrievals performed over water from analyses to limit the a priori influence on results. Level 3 (L3) products are a 1° x 1° gridded average of finer resolution Level 2 (L2) retrievals. For coastal grid boxes, these are retrievals that are either performed over land, water, or a combination of the two, on any given day. L3 data users therefore have limited ability to filter for retrievals performed over water for these grid boxes. The consequences that this has on retrievals and their temporal trends in "as-downloaded" L3 data (L3O) are examined in this paper, for all coastal L3 MOPITT grid boxes ($n = 4299$), by comparison to separate land- and water-only grid box averaged L2 retrievals (L3L and L3W, respectively). First, it is established that mean retrieved VMRs in L3L and L3W differ by over 10 ppbv, significant ($p < 0.1$) at 60 % of the coastal grid boxes. Trends are also stronger in L3L (mean difference between 0.28 ppbv y^{-1} and 0.43 ppbv y^{-1}), with the L3L – L3W trend difference significant at 36 % of grid boxes. These L3L-L3W differences are clearly linked to retrieval sensitivity differences, with L3W being more heavily tied to the a priori CO profiles used in the retrieval, which is a model-derived monthly mean climatology. On days when L3O is created from the averaging together of L2 retrievals over both land and water ($L3O_M$), the result is VMRs that are significantly different to L3L for 75 % of grid boxes where the L3L – L3W difference is also significant, 45 % of all coastal grid boxes. Just under half of the grid boxes that featured a significant L3L – L3W trend difference also see trends differing significantly between L3L and $L3O_M$. Factors that determine significance of difference between $L3O_M$ and L3L include proportion of the surface covered by land/water, and the magnitude of sensitivity contrast. Comparing the full L3O dataset to L3L, it is shown that if L3O is filtered so that only retrievals over land ($L3O_L$) are analysed, there is a huge loss of days with data. This is because L2 retrievals over land are routinely discarded during the L3O creation process, for coastal grid boxes. The problem can be lessened by also retaining $L3O_M$ retrievals, but the resulting L3O "land or

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

mixed" (L3O_{LM}) subset still has less data days than L3L for 61 % of coastal grid boxes. Moreover, as already shown, these additional days with data feature some influence from retrievals made over water that can affect results. Coastal L3 grid boxes contain 33 of the 100 largest coastal cities in the world, by population. Focusing on the L3 grid boxes containing these cities, it is shown that mean VMRs in L3O_L and L3L differ significantly for 11 of the 27 cities that can be compared (there are no L3O_L data for 6 of the cities). The L3L – L3O_{LM} mean VMR difference exceeds 10 (22) ppbv for 11 (3) of the 33 cities, significant in 13 cases. 9 of the 18 cities where WLS analysis can be performed in L3O_L feature a trend that is significantly different to L3L. The trends in L3O_{LM} and L3L differ significantly for 5 of the 33 cities. It is concluded that a L3 product based only on L2 retrievals over land would be of benefit to MOPITT data users, given the clear and sometimes significant differences in mean CO VMRs and trends that can be obtained for coastal grid boxes using L2 products in which retrievals performed over water can be more easily discarded.

1. Introduction

Carbon monoxide (CO) is directly emitted into the atmosphere from anthropogenic (e.g. fossil fuel burning) and natural (e.g. wildfire) sources, and also produced via the oxidation of hydrocarbons in the atmosphere. With an atmospheric lifetime of weeks to months (e.g. Duncan et al., 2007), it is an important tracer of pollutant transport and indicator of emission sources. While a health concern in its own right at high enough concentrations, CO also plays an important role in atmospheric chemistry, for example as a precursor to ozone formation and a primary sink for the hydroxyl radical. Atmospheric CO concentrations have decreased since the start of the 21st century, with a slowdown in the rate of decline observed in recent years (Buchholz et al., 2021). Trends also show substantial spatial variability (Hedelius et al., 2021). Satellite instruments have been central to our understanding of global change in CO concentrations, with the Measurement of Pollution in the Troposphere (MOPITT – Drummond et al., 2010, 2016) instrument well suited to this task, providing a nearly-unbroken and consistent data record since the year 2000.

MOPITT observes upwelling radiances at thermal infrared (TIR) and near infrared (NIR) wavelengths and uses these in an optimal estimation retrieval algorithm to retrieve coarse vertical resolution CO profiles, which are integrated to give total column amounts. Among multiple additional inputs required by the retrieval algorithm, a priori CO profiles – which describe the most probable state of the CO profile at a given location – are necessary to constrain the retrieval to physically reasonable limits (Pan et al., 1998; Rodgers, 2000; the retrieval algorithm is outlined in more detail in Sect. 2.1). For the most recent iterations of MOPITT products, these a priori CO profiles are based on a monthly climatology from a chemical transport model. The degree to which a given MOPITT retrieval reflects information obtained from the observed radiances – known as

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

"information content" – is highly spatially and temporally variable, depending 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). The lower the retrieval information content, the closer the retrieved CO loading will be to the a priori; a model value.

Retrievals that take place over water are known to have a lower information content than retrievals that take place over land. This is due to weak thermal contrast near to the surface hampering the instrument's ability to sense CO absorption in the lowermost layers of the troposphere (Deeter et al., 2007; Worden et al., 2010). It is therefore recommended that MOPITT data users exclude these retrievals from any analyses they perform, to ensure that results are not biased by retrievals that have a heavy reliance on the a priori (MOPITT Algorithm Development Team, 2018; Deeter et al., 2015). Such filtering is specifically emphasised where the focus of analysis is the identification of long-term CO trends, because any real trends in the data will be weakened by the inclusion of retrievals that are tied heavily to the a priori (Deeter et al., 2015). This is because the a priori CO profiles are taken from monthly modelled CO climatologies: for a given location and day of the year, they will be the same every year and therefore feature no temporal trend (Deeter et al., 2014).

MOPITT data are available as either Level 2 ("L2") or Level 3 ("L3") products. L2 products contain each individual retrieval, at ~22 x 22 km spatial resolution. L3 products are a 1° x 1° gridded area-average of the individual L2 retrievals that fall within each grid box (see Fig. 1), with some filtering criteria applied. One criterion is the surface type over which the L2 retrievals were performed – either land, water, or "mixed". If more than 75 % of the bounded L2 retrievals were performed over the same surface type then only those retrievals are averaged to create the L3 product and the rest are discarded; otherwise, all bounded L2 retrievals are averaged, and the L3 product is given the surface type classification of "mixed" (L3 surface type classification is explained in more detail in Sect. 2.2). This creates a problem for L3 grid boxes that overlay coastlines: To a greater or lesser extent, these L3 products will have some contribution from L2 retrievals performed over water, as shown in Fig. 1. L3 product users have limited capability to discard them, at least without sacrificing temporal resolution (because each L3 grid box only has a single retrieval per day). By contrast, with L2 products it is possible, for the same coastal grid boxes, to choose to retain only the retrievals performed over land. In practical terms, this means that, for coastal L3 grid boxes, valuable retrieval information over land, available in L2 products, can be lost to users of L3 products.

With a focus on the coastal L3 grid box containing the city of Halifax, Canada, Ashpole and Wiacek (2020) demonstrate the consequences of this loss of retrieval information in L3 products. They compare the results of analyses performed using L3 data and L2 data whereby only bounded retrievals performed over

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

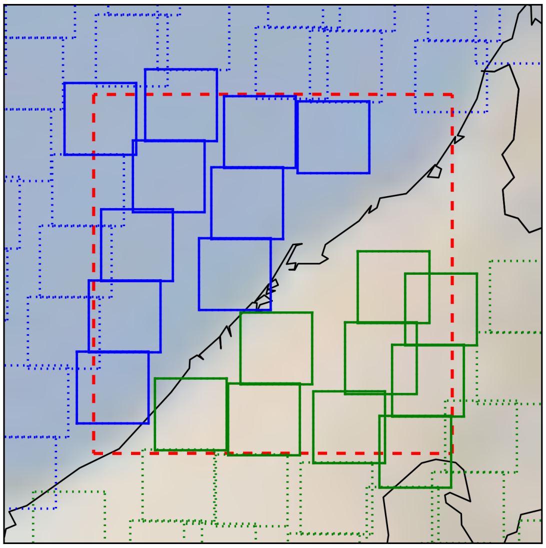


Figure 1. Example of coastal L3 grid box (red dashed box) and bounded L2 retrievals from which the L3 products for that grid box are created. Blue (green) boxes correspond to L2 retrievals with a surface index of "water" ("land"). Note that only L2 retrievals with a midpoint that falls within the boundaries of the L3 grid box will be used in L3 creation for that grid box. These are indicated by solid blue/green outlines – those not included in L3 creation for this grid box are shown with dotted blue/green outlines. More information on surface indexing and L3 product creation is given in Sect. 2.2. "Coastal" L3 grid box classification is outlined in Sect. 2.3. The coastal L3 grid box visualized here contains the city of Dubai (~centre = 55.296º E, 25.277º N), which features in the case study analysis of Sect 3.4. Background shading is from Nasa Blue Marble imagery.

land were retained, and find significant differences in both seasonal mean statistics and the magnitudes of trends identified in surface-level CO. These differences are a direct result of the L3 products being dominated by L2 retrievals over water, which feature a weaker trend than the L2 retrievals over land, demonstrably due to a greater a priori influence owing to their reduced true-profile sensitivity. In their conclusions, Ashpole and Wiacek (2020) suggest that L2 retrievals over water should not contribute to L3 products for coastal grid boxes, which would be consistent with previous data filtering recommendations (MOPITT Algorithm Development Team, 2018; Deeter et al., 2015). The primary aim of this paper is to explore the extent of the difference that this would make on a global scale. This is necessary to understand for two reasons: firstly, L3 data are better suited to long timeseries analysis than L2 data owing to their smaller file size (~25 MB vs ~450 MB respectively, for a single daily, global file). It cannot be overlooked that working with L3 data thus requires fewer computing resources and less technical proficiency. L3 products thus make the MOPITT data more easily accessible, especially to less-expert users, who may lack the expertise required to scrutinize the data for potential a priori bias. Secondly, many of the world's largest agglomerations are situated within a

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

coastal L3 grid box (5 of the top 10 and 33 of the top 100 largest agglomerations by population; derivation outlined in Sect. 2.5), making these likely targets for analyses of air quality indicators, especially their changes over time.

This paper presents a comparison of results from analyses performed using L3 data products and separate land-only and water-only area averages from L2 products for all MOPITT L3 grid boxes that overlay coastlines. Section 3.1 demonstrates the magnitude of the sensitivity difference for retrievals over land and water, zooming in to focus on coastal grid boxes, the classification of which is outlined in Sect. 2.3. Section 3.2 links the sensitivity contrast to differences in mean CO volume mixing ratios and their temporal trends for L2 retrievals performed over land and water within coastal L3 grid boxes, and evaluates the effect that the averaging together of these retrievals has on the statistics and trends in resulting L3 "mixed" values. Section 3.3 quantifies the proportion of L2 retrievals performed over land within coastal L3 grid boxes that are lost to L3 products, before finally comparing statistics and trends in L3 and L2 products for all coastal L3 grid boxes, outlining the magnitude and significance of differences for the 33 largest coastal cities in the world (Sect. 3.4).

2. Data and Methods

2.1. MOPITT Instrument and retrieval overview

Carried on board the polar-orbiting NASA Terra satellite that was launched in December 1999, MOPITT began measuring CO in March 2000 and has provided near-continuous measurements to date. With a native pixel resolution of ~22 x 22 km at nadir and a swath width of ~640 km, it offers near global coverage roughly every 3-days, crossing the equator at ~10:30 and ~22:30 local time. The instrument is a gas correlation radiometer that measures radiances in two CO-sensitive spectral bands: the TIR at 4.7 µm, which is sensitive to both absorption and emission by CO and can provide information on its vertical distribution in the troposphere; and the NIR at 2.3 µm, 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 (Drummond et al., 2010; Pan et al., 1995, 1998). For the work presented here, the TIR-NIR combined MOPITT product is used, owing to its demonstrably greater sensitivity to CO loadings near to the surface than the TIR- and NIR- only products which are also available (Deeter et al., 2013). Note, however, that retrievals over water and at night are limited to the TIR band only due to the lacking NIR signal. This analysis is based on daytime-only retrievals (more information on data selection and preparation is given in Sect. 2.4).

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Multiple other sources describe the retrieval algorithm in detail (e.g., Deeter et al., 2003; Francis et al., 2017). In short, it uses optimal estimation (Pan et al., 1998; Rogers, 2000) and a fast radiative transfer model (Edwards et al., 1999) to invert measured radiances and retrieve the CO volume mixing ratio (VMR) profile on 10 vertical layers. The vertical grid consists of 9 equally spaced pressure levels from 900 to 100 hPa (the uppermost level covers the atmospheric layer from 100 to 50 hPa), with a floating surface pressure level (if the surface pressure is below 900 hPa, less than 10 profile levels are retrieved). Retrieved values represent the mean CO VMR in the layer immediately above that level. These profile measurements are then integrated to provide total column CO amounts. Retrievals are only performed for scenes free of cloud (cloud clearing is based on coincident MODIS observations and MOPITT's own radiances).

In addition to the measured radiances, the retrieval requires multiple inputs including meteorological data, surface temperature and emissivity, and, of direct relevance to this study, a priori CO profiles, which are necessary to constrain the retrieval to physically reasonable limits. These a priori CO profiles come from a monthly CO climatology (years 2000-2009), simulated with the Community Atmosphere Model with Chemistry (CAM-chem) chemical transport model (Lamarque et al., 2012) at a spatial resolution of 1.9^{o} x 2.5^{o}, which is then spatially and temporally interpolated to the time and location of each individual MOPITT observation. A priori profiles for a given location and day of the year are therefore the same every year and feature no temporal trend. To understand the physical significance of the MOPITT CO retrievals, it is necessary to examine the retrieval Averaging Kernels (AKs), available with all MOPITT data products, which quantify the sensitivity of the retrieved vertical profile to the "true" vertical profile. The lower the retrieval sensitivity, the greater the a priori weighting. Two different components of AKs are analysed in this paper: AK rowsums, which represent the overall sensitivity of the retrieved profile at the corresponding pressure level to the whole true profile; and AK diagonal values, which represent the sensitivity of the retrieved profile at the corresponding pressure level to the same level of the true profile (e.g. the AK diagonal value for the surface level of the retrieved profile represents its sensitivity to the surface level of the true profile).

From time-to-time, new MOPITT products become available as improvements are made to the retrieval algorithm and radiative transfer model, yielding superior validation statistics compared to earlier product versions (Worden et al., 2014). This analysis uses MOPITT Version 8 (V8) products (Deeter et al., 2019). Note that Version 9 (V9) products became available shortly after this study was completed. V9 features cloud screening improvements that yield additional retrievals over land in comparison to V8 (the exact percent change varies significantly with geography). Validation results are comparable to V8. It is expected that the main conclusions of this paper to hold for V9, since the land-water sensitivity contrast

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

remains and L3 processing method appears to be unchanged. An overview of MOPITT V9 is given by Deeter et al (2021).

2.2. MOPITT surface type classification

To aid in filtering and interpreting retrievals, all MOPITT data products are distributed with a range of diagnostic fields. As retrieval information content is known to be variable depending on the type of surface over which it is performed (Deeter et al., 2007), L2 retrievals are given a surface index according to whether they were performed over land, water, or a combination of the two ("mixed"). For a given 1° x 1° L3 grid box, how the L2 retrievals that fall within its boundaries are processed to produce the L3 product depends on how their surface indexes vary: If more than 75 % of the bounded L2 retrievals have the same surface index, only those retrievals are averaged to produce the L3 gridded value, and the L3 surface index is set to that surface type (the other L2 retrievals are discarded). Otherwise, all L2 retrievals available in the L3 gridbox are averaged together and the L3 surface index is set to "mixed", as is the case in the example shown in Fig. 1 (this information is taken from the MOPITT Version 6 L3 data quality summary[1], which at the time of writing, is the most recent data quality summary to detail exactly how L3 data are created). Note that the L2 VMR profiles that are averaged to produce the L3 retrieval are first converted to log(VMR) profiles, then averaged, and the mean log(VMR) profile is then converted back to a VMR profile.

Each L3 grid box only has one retrieval per day. This dictates that where the grid box overlies both land and water, its surface index will vary through time, depending on the population of L2 retrievals from which it is created. The make-up of this population can also vary from day-to-day due to factors such as cloud cover, and screening for data quality issues: on day n the population could be predominantly L2 retrievals over land, on day $n+1$ it could be predominantly L2 retrievals over water, and on day $n+2$ it could be an even mix of the two. Given that the averaging together of retrievals with significantly different sensitivity profiles – as could be the case when averaging retrievals over land and water – serves to dilute the information coming from the MOPITT observed radiances with information coming from the a priori and is therefore discouraged (MOPITT Algorithm Development Team, 2018; Deeter et al., 2015; Deeter et al., 2007); and that MOPITT data users are advised to exclude retrievals over water from analyses owing to the known reduced sensitivity, this introduces two potential problems for L3 data taken from coastal grid boxes: firstly, discarding all L3 retrievals with the surface index of water will result in a loss of temporal coverage; secondly, L3 retrievals

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

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

with a surface index of mixed feature some contribution from L2 retrievals over water. The consequences of both of these problems are explored in this paper.

2.3. Coastal grid box classification for this study

Since the focus of this paper is on "coastal" L3 grid boxes, it is first necessary to isolate these from the remaining "land-only" or "water-only" L3 grid boxes in the MOPITT data set. The initial step is to identify all grid boxes that have a surface index of "mixed" at least once during the study period. This indicates that the ground area within those grid boxes was both land and water. However, analysis of the global distribution of L3 grid boxes featuring a surface index of mixed revealed that, in addition to actual coastlines, a large proportion of inland grid boxes that are clearly not coastal ("false coastal") are given the surface index of mixed at least some of the time (Fig. 2a). The reason for this is unclear, but it could be for real physical reasons, such as land grid boxes sporadically flooding, or due to issues in the retrieval schemes caused by e.g. cloud screening problems or the presence of surface ice cover. One characteristic of these false coastal grid boxes is that, compared to the total number of days with L3, the relative frequency with which they are flagged as land is very high (expressed as the ratio "n_days($L3O_L$/L3O)", plotted in Fig. 2b). This relative frequency is much lower for "true" coastal grid boxes, to be expected given prior knowledge of 1) the fact that these grid boxes span both land and water surface types; and 2) how the surface index is determined for L3 data (as outlined in Sect. 2.2). Following iterative threshold testing, L3 coastal grid boxes are classified as grid boxes that:

1. Have at least one classification of "mixed" during the study period
2. Have an n_days($L3O_L$/L3O) ratio < 0.5.

The distribution of coastal grid boxes identified using these criteria is shown in Fig. 2c. Most false coastal grid boxes are removed, although there are still some erroneous classifications evident, mostly in the north of Canada and Russia. However, placing a more restrictive threshold on the n_days($L3O_L$/L3O) ratio to remove these areas has diminishing returns since it results in the rejection of more true coastal grid boxes. These criteria therefore strike a balance between minimising false and maximising true classifications.

Applying these criteria to the MOPITT L3 data yields 4299 coastal grid boxes, from a total of 64800 L3 grid boxes (6.6 %). This mask is applied to all data, and only those L3 grid boxes that remain are classified

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

as coastal. Only data for these coastal grid boxes are analysed in this study (with the exception of global L3 maps in Sect. 3.1.1).

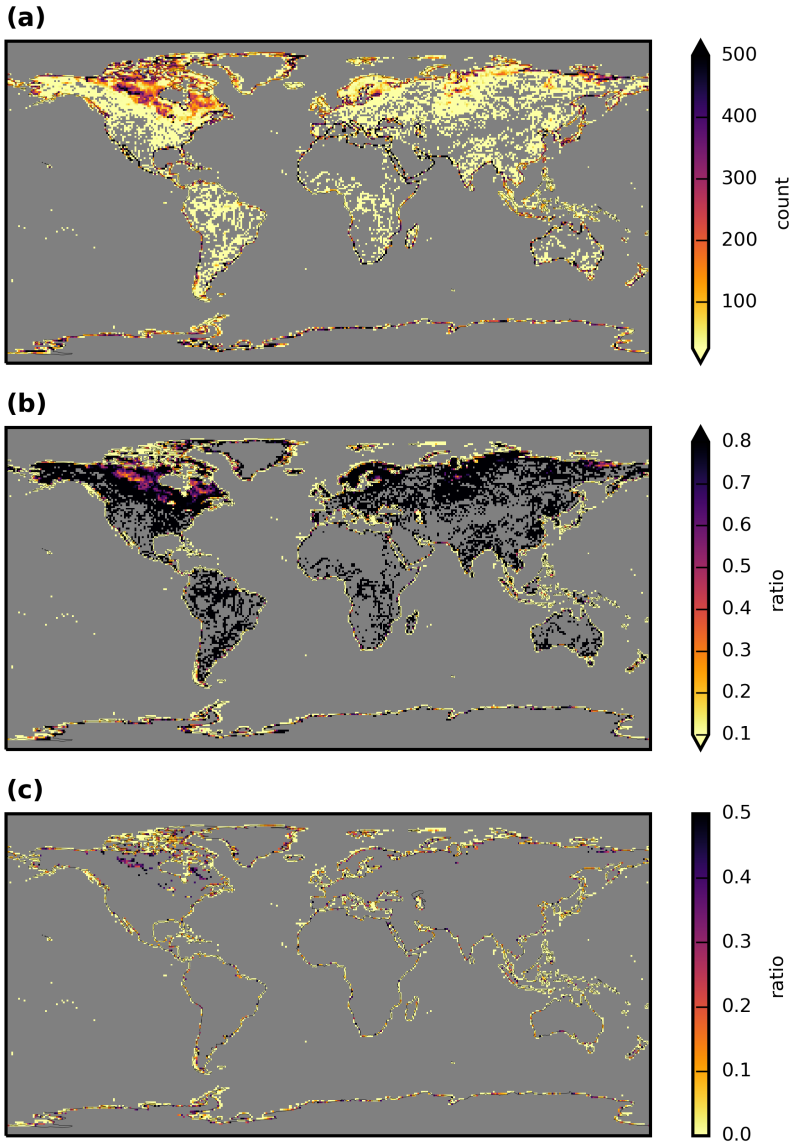


Figure 2. Maps showing the stages of derivation of the coastal L3 grid box mask applied in this paper to MOPITT data. **(a)** Frequency with which L3 grid boxes are given the surface index of "mixed", calculated from daily data between 2001-08-25 and 2019-02-28. **(b)** Frequency with which L3 grid boxes that have a surface index of "mixed" at least once in panel a have the surface index of "land", compared to the total number of days with which L3 data are available for that grid box (expressed as n_days($L3O_L$/L3O)). **(c)** As b, but with a threshold of n_days($L3O_L$/L3O) < 0.5 applied. This is the coastal L3 grid box mask used in this paper.

2.4. MOPITT datasets analysed, and data processing methods

All available MOPITT V8 Level 2 (L2) and Level 3 (L3) TIR-NIR files ("MOP02J" and "MOP03J" files, respectively) were downloaded from the NASA Earthdata portal (https://search.earthdata.nasa.gov). Although the data record begins in March 2000, analysis is restricted to the period from 2001-08-25 to 2019-02-28. Data prior to 2001-08-25 are discarded due to an instrumental reconfiguration in 2001 creating an inconsistency in the data record (Drummond et al., 2010). Data post 2019-02-28 are flagged as "beta" at the time of writing, their use in scientific analysis (especially for examining long-term records of CO) being discouraged until final processing and calibration occurs (MOPITT Algorithm Development Team, 2018). For clarity, the original, "as-downloaded" L3 timeseries is referred to as "L3O" for the remainder of this paper. Only retrievals that were performed during daytime hours are retained (daytime and nighttime retrievals are stored as separate fields in MOP03J files). For this analysis, separate subsets of L3O are created according to surface index: L3O land-only ("$L3O_L$"), L3O water-only ("$L3O_W$"), L3O mixed ("$L3O_M$"), L3O land-or-mixed ("$L3O_{LM}$"). When the L3O dataset is analysed with no filtering by surface index applied, it is referred to as "$L3O_{NF}$".

The first step of L2 data processing for this study is to filter the retrievals as is done at the L3 processing stage. This involves:

- Discarding all observations for Pixel 3 (this corresponds to one of MOPITT's four detectors);
- Discarding all observations where both (1) the channel 5A signal-to-noise-ratio ("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)

This filtering takes place because observations from specific elements on MOPITT's detector array were found to exhibit greater retrieval noise than the other elements, and their inclusion therefore lowered overall L3 information content (MOPITT Algorithm Development Team, 2018). Only daytime L2 retrievals are retained, using a solar zenith angle filter of < 80^o.

From the remaining set of filtered L2 retrievals, separate area averages are taken for those with a surface index of land and water, for every 1^o x 1^o L3 grid box. This effectively creates two new L3 "land only" and "water only" products, which are referred to herein as "L3L" and "L3W". For clarity of analysis, remaining L2 retrievals with a surface index of mixed are discarded. These make up a very small proportion of the overall L2 retrievals (e.g. < 5 % for the grid box containing Halifax, analysed in Ashpole and Wiacek, 2020). Note that, as with the creation of L3O, L2 VMR profiles for each L3 grid box are first converted to

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

log(VMR) profiles before averaging, and the mean log(VMR) profile is then converted back to a VMR profile to give the final L3L and L3W retrievals.

From these L3O, L3L, and L3W datasets, only grid boxes that are classified as "coastal" using the coastal grid box masked outlined in Sect. 2.3 are analysed.

2.5. Statistical methods used for this study, and additional data sources

For every coastal L3 grid box, two separate timeseries from each of the L3O, L3L, and L3W datasets are analysed. In Sect. 3.1 and 3.2 the timeseries analysed only contain days where L3L and L3W are both present and the L3O surface index is mixed ("$L3O_M$"). In Sect. 3.3 and 3.4 the full timeseries from each dataset is analysed. Descriptive statistics are calculated from both timeseries across the whole study time period, and also for individual years (full years only – 2002 to 2018 inclusive).

To identify and compare temporal trends for each coastal grid box in the datasets outlined above, weighted least squares (WLS) regression analyses is performed on yearly mean values, weighted by the inverse of the standard deviation of the measurements used in the yearly mean (i.e. $1/\sigma$). For years that contain just a single retrieval, the weighting is set to 1/100000 to de-weight them in the fit. If there are more than 2 years in a timeseries for a given grid box that have no data, the regression analysis is not performed. WLS is preferred over OLS because it is less sensitive to outliers. For simplicity, no other trend detection methods – e.g. the Thiel-Sen slope estimator – are applied to corroborate the trends that are detected with WLS. Such extra steps would be necessary if the actual trend values were the focus of this study; however, the aim of this trend analysis is instead to identify whether the same method can yield different results depending on which of L3O, L3L or L3W is analysed.

To determine whether two trends identified are significantly different, their difference is evaluated using the Z test 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. $p < 0.1$). In addition, two trends are classified as being significantly different if $Trend_1$ is

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

significantly different to zero ($p < 0.1$) but $Trend_2$ is not ($p > 0.1$), and vice-versa (i.e. the conclusion would be that $Trend_1$ is not zero, but $Trend_2$ may be).

A list of the top 100 largest agglomerations by population in the world is obtained from http://www.citypopulation.de/ (valid at time of writing). 33 of these are situated in a coastal grid box, according to the classification in Sect. 2.3. Time series of L3L, L3W, and L3O are extracted from each of these grid boxes for the analysis in Sect. 3.4.

3. Results and Discussion

3.1. Land-water contrast in MOPITT sensitivity

This section demonstrates the land-water sensitivity contrast in MOPITT retrievals at levels throughout the vertical profile, and examines the magnitude of the difference within coastal L3 grid boxes.

3.1.1. Global context

Figure 3 shows long-term mean maps for the retrieval sensitivity metrics AK diagonal value, AK rowsum, and retrieved minus a priori VMR ("VMR ret-apr") at selected profile levels, created from L3O data averaged across the entire study period (September 2001 – February 2019, inclusive). All indicators show that retrieval sensitivity is greater over land than water in the lower troposphere ("LT"; represented by the surface, 900 hPa and 800 hPa profile levels), with sharp differences evident at almost all land-water boundaries. The sensitivity contrast clearly decreases in strength with height. By mid-tropospheric levels ("MT"; represented by 600 hPa profile level), AK diagonal values and rowsums reach greater values on average over water than land. Some strong land-water gradients remain present in VMR ret-apr fields, most notably over North Africa, the Arabian peninsula, and south-east China, but on average these values are much more similar across land and water than in the LT. No clear land-water contrast is evident in the upper troposphere ("UT"; represented by the 300 hPa profile level), with retrieval sensitivity instead varying more with latitude, decreasing towards both poles (a companion to Fig. 3 with an altered colour bar to better show spatial patterns in AK diagonal values and rowsums at MT and UT levels is provided in the Supp. Mat. (SM1)).

AK diagonal values and rowsums show that retrieval sensitivity increases across both land and water with height. It is lowest at the surface level, with little information content in the retrieval over water (AK

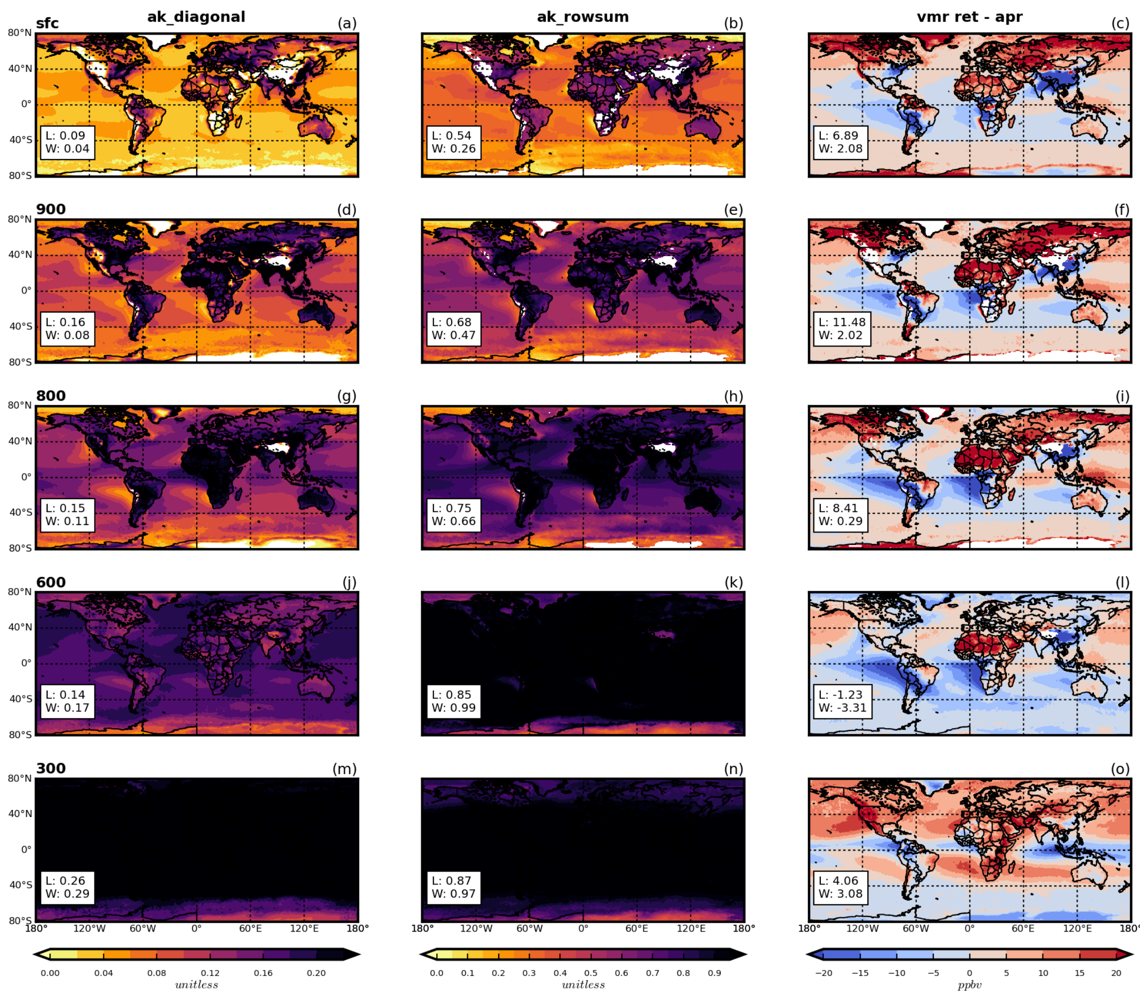


Figure 3. Mean sensitivity metrics from MOPITT L3 data, averaged across the entire study period (September 2001 – February 2019, inclusive). Shown are AK diagonal values (left column), AK rowsums (center column) and VMR retrieved minus a priori values (right column) for the following levels of the retrieved profile: surface (top row), 900 hPa (second row), 800 hPa (third row), 600 hPa (fourth row), and 300 hPa (bottom row). Values in white boxes correspond to mean values across all land ("L") and water ("W") L3 grid boxes.

diagonal values and rowsums over water are less than half what they are over land, on average). There is high spatial variability over land: AK diagonal values and rowsums reach values comparable to those at higher profile levels in some sensitivity hotspots (e.g. parts of central Europe, east Asia, eastern USA and tropical west Africa), while being more comparable to values over water in other areas. By 800 hPa, AK diagonals

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

and rowsums over water reach values comparable to or greater than those reached over land at the surface level, in most places.

Spatial patterns in retrieved minus a priori VMRs are slightly more complex to interpret, because they are influenced both by retrieval sensitivity and the accuracy of the a priori. For example, while values close to zero can indicate a retrieval that is heavily weighted by the a priori and therefore low retrieval sensitivity, they can also indicate that the true VMR is close to the a priori value. Despite this, retrieved minus a priori VMR values clearly reach more strongly positive or negative values over land than water in the LT, with the contrast becoming less pronounced with height. Furthermore, there are clear land-water changepoints in the LT, especially in the LT. This further demonstrates the impact of the land-water contrast in retrieval sensitivity.

3.1.2. Analysis of coastal L3 grid boxes

Scatterplots of sensitivity metrics at selected profile levels, for coastal L3 grid boxes only, are shown in Fig. 4. Specifically, these plots show the sensitivity of the L2 land and water retrievals that are bounded by the 1° x 1° L3 grid boxes and used to create the L3O data. The values that are plotted correspond to the long-term mean from the L3L and L3W datasets for these grid boxes. For this comparison, the L3L and L3W means are only calculated from days when L2 retrievals over both land and water are present and the L3O surface index is mixed. This minimises potential differences in the true CO profiles that could arise due to temporal variations for the L3L and L3W pairings being compared. Furthermore, it allows for the analysis of the resulting $L3O_M$ data on these days with knowledge of the parent L2 retrievals over land and water and their differences. For ease of interpretation, the absolute retrieved minus a priori VMR values are plotted, i.e. ignoring whether the result is positive or negative. However, the results hold if using signed values, and a duplicate of Fig. 4 with signed retrieved minus a priori VMR values is included in the Supp. Mat. for reference (SM2).

The AK diagonal value and rowsum plots clearly demonstrate greater sensitivity over land (L3L) than over water (L3W) at LT levels (a point below the diagonal line on these panels indicates greater values in L3L) for the majority of grid boxes, with the difference decreasing into the MT and UT. Correspondingly, retrieved VMRs also deviate more greatly from their a priori values in L3L than L3W in the LT, with smaller land-water differences in the MT and UT. Mean values are significantly different ($p < 0.005$) apart from AK diagonal values and retrieved minus a priori VMR at 300 hPa ($p = 0.13$ and 0.07 respectively). Sensitivity metrics are generally better correlated in the MT and UT than at LT levels.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

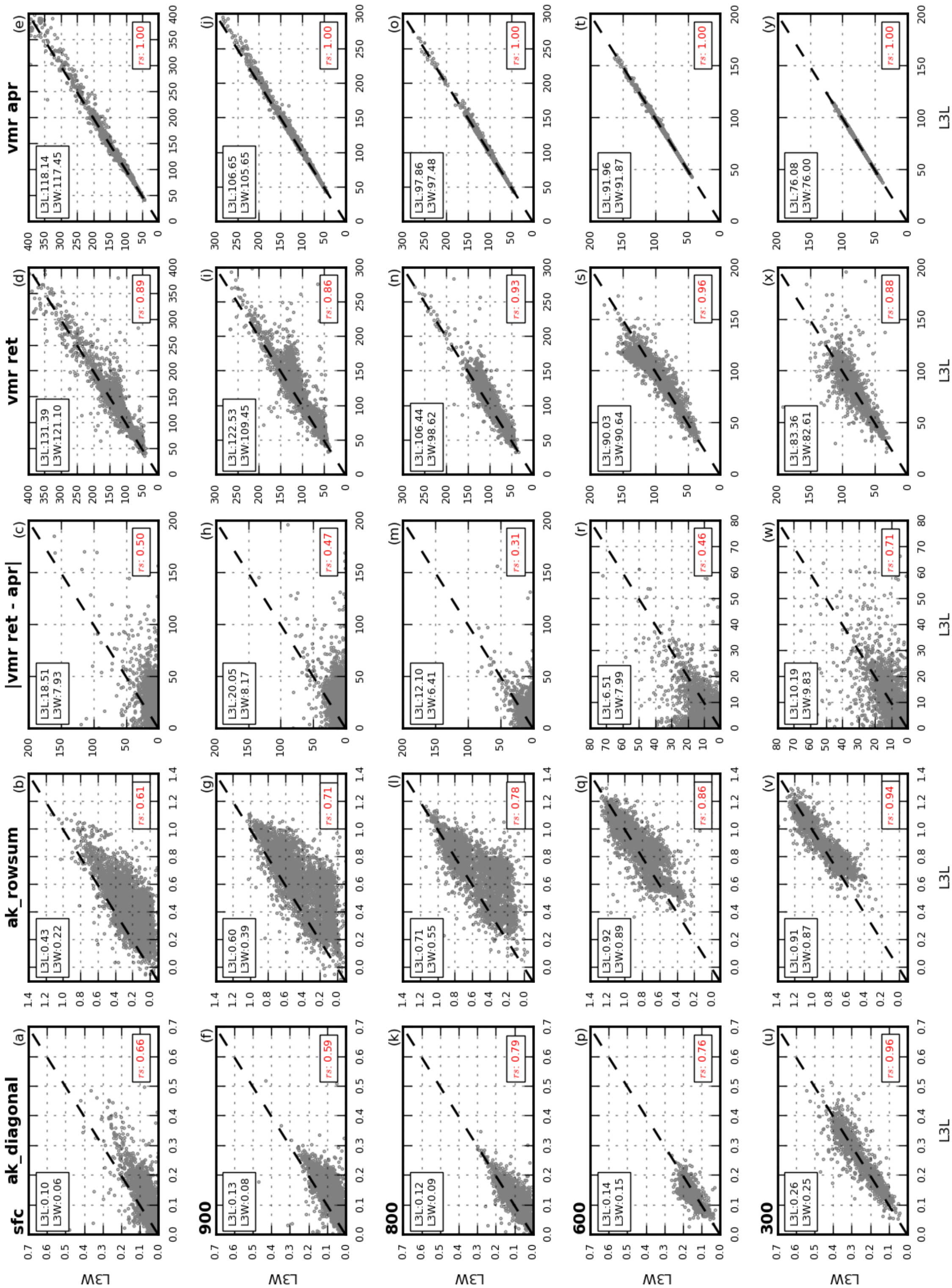


Figure 4. Mean sensitivity metrics and VMRs (retrieved and a priori) from coastal L3 grid boxes. Values compared in the scatterplots are mean values from matched L3L and L3W retrievals within these grid boxes. "Matched" means that only days when both L3L and L3W are present, and the L3O surface index is mixed, are used to create the mean values analysed. Shown are AK diagonal values (left column), AK rowsums (second column), absolute VMR retrieved minus a priori values (third column), retrieved (fourth column) and a priori (fifth column) VMRs, for the following levels of the retrieved profile: surface (top row), 900 hPa (second row), 800 hPa (third row), 600 hPa (fourth row), and 300 hPa (bottom row). Values in boxes in the top-left corner of each panel correspond to mean values across all L3L and L3W grid boxes. These means are significantly different using a 2-tailed t-test (unequal variance) with $p < 0.005$ in all cases except ak_diagonal at 300 hPa where $p = 0.13$, vmr_ret_minus_apr at 300 hPa where $p = 0.07$, vmr_ret at 600hPa where $p = 0.30$, vmr_ret at 300hPa where $p = 0.11$. No vmr_apr mean differences are significant. Values in the bottom-right corner of each panel correspond to the Spearman's rank correlation coefficient ($p < 0.005$ in all cases).

This analysis clearly shows how L2 retrievals that are averaged together to create the L3O data over coastal grid boxes on days when the surface index is mixed have differing degrees of sensitivity, especially in the LT. This is explicitly cautioned against in the MOPITT data user's guide (MOPITT Algorithm Development Team, 2018). The remainder of this paper focuses on the surface-level of the retrieved profile, since the LT is where discrepancies are greatest, and the cause of this sensitivity contrast is well established (as outlined in the introduction).

3.2. Differences in retrieved VMRs and temporal trends, and their relation to the land-water sensitivity contrast

3.2.1. L3L vs L3W

Retrieved VMR comparison between L3L and L3W

In addition to the clear land-water LT sensitivity contrast in coastal grid boxes, there are (sometimes large) differences in the retrieved VMRs (Fig. 4; Fig. 5a (black boxplots)). The retrievals performed over land yield surface-level VMRs that are over 10 ppbv greater than over water, on average. As with sensitivity, land-water differences in retrieved VMRs decrease higher up in the profile. Although a decrease in VMRs from land to water might be expected in the LT, given that most CO sources are land-based, the land-water difference in LT retrieved VMRs is well over 10 times greater than in the a priori VMRs used for the retrievals (also shown in Fig. 4; mean difference = 10.29 vs 0.69 ppbv, respectively, for surface level). Furthermore,

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

this assumption only seems reasonable where large CO sources are proximal to the coastline, as it is unrealistic to expect such large gradients in background CO (which coastal grid boxes far from large CO sources are more likely to represent) across a relatively small distance – as is verified by the a priori VMR comparison which suggests the land-water difference in CO concentrations should be negligible for the majority of the grid boxes compared.

Greater land-water sensitivity differences tend to be associated with greater retrieved VMR differences. Figure 5b shows the distribution of retrieved surface level VMR differences (L3L – L3W) stratified by the corresponding surface level AK rowsum difference. Larger retrieved VMR differences are clearly associated with greater AK rowsum differences (some degree of spread in the results is expected, since the relationship also depends on the accuracy of the a priori, as outlined previously). Absolute retrieved VMR difference values are shown in Fig. 5b for clarity, since L3L – L3W can be either positive or negative depending on whether a priori VMRs used in the retrieval are greater or less than the "true" VMR being retrieved, which complicates the analysis. The corresponding plot with raw values (i.e. not discarding the +/- sign) is included in the Supp. Mat. however, and the same conclusions can be drawn based on this figure (SM3).

Of the 3971 coastal grid boxes that are compared, 60 % (2379) show a significant difference ($p < 0.1$, determined using a 2-tailed student's t-test) in mean VMRs in L3L and L3W (Fig. 5a). Compared to grid boxes where the mean VMR difference is not significant, there are several notable differences (detailed in Table 1). As expected from the previous analysis, the land-water sensitivity contrast is greater when mean VMRs are significantly different than when not. This is evident in AK rowsum and VMR retrieved minus a priori differences (the magnitude of difference between significance subsets is around 50 % and 100 %, respectively). Interestingly, the AK difference is due to sensitivity being lower over water when VMRs are significantly different than when they are not; sensitivity over land is similar in both subsets. This may be explained as follows: when sensitivity over water is especially low, as is the case in the significant-difference subset, the retrieved VMR will be heavily weighted by the a priori and unable to match the variation present in the more sensitive retrieval over land. As sensitivity over water increases, this a priori weighting weakens and the retrieved VMR will more closely track the retrieval over land, resulting in a less significant difference. Also of note, a priori VMRs are much lower when retrieved VMRs are significantly different than when they are not, on average. Considered alongside the greater retrieved minus a priori differences, this suggests that the a priori VMR could be a less accurate estimate of the "true" VMR for the significant-difference subset, whereas it is closer to reality when retrieved VMRs over land and water are not significantly different. Intuitively, this makes sense: for a hypothetical situation where the a priori VMR is a perfect match for the "true" VMR, and both are uniform across a coastal L3 grid box, retrievals over the land and water grid box

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

435

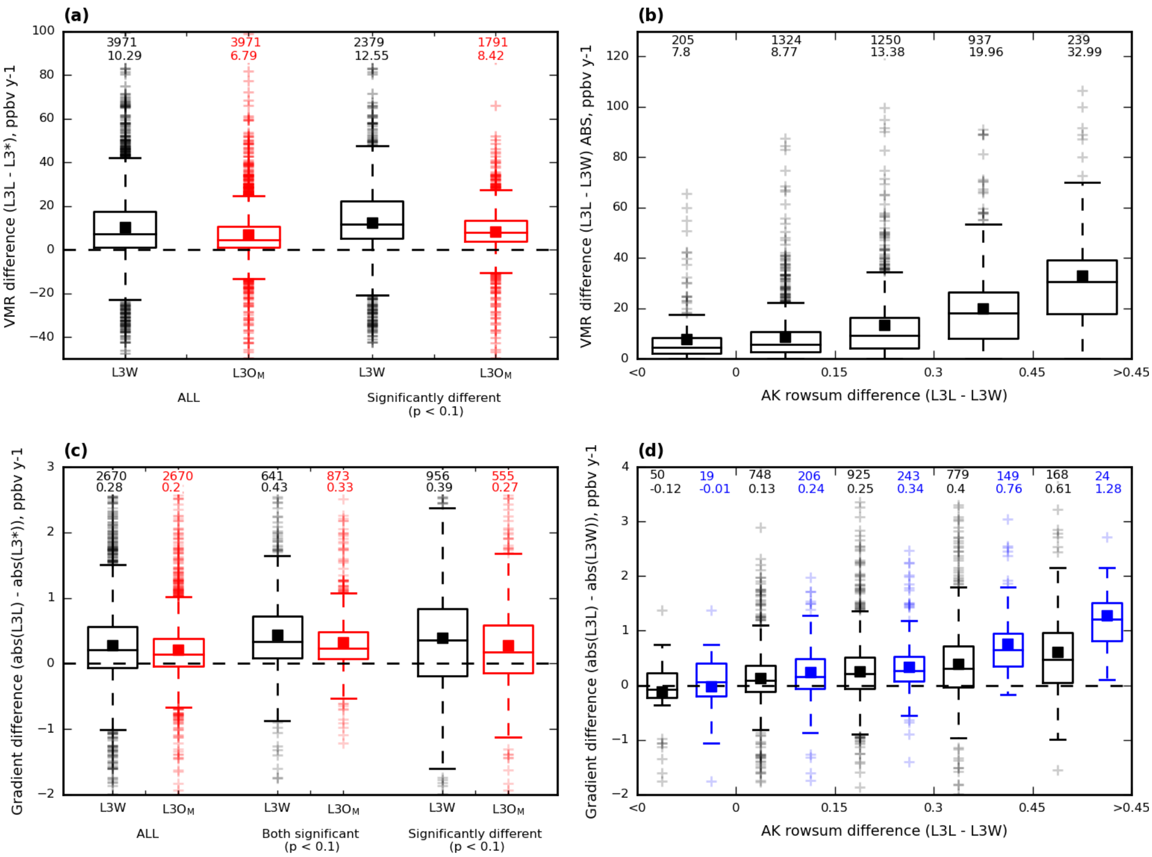


Figure 5. Boxplots showing how mean VMRs and trends from WLS analysis compare for coastal L3 grid boxes, calculated from matched retrievals within these grid boxes. "Matched" means that only days when both L3L and L3W are present and the L3O surface index are mixed are used to create the mean values analysed. Mean values are represented by filled squares, and values above the boxplots correspond to number of grid boxes with data for that boxplot, and the mean value, respectively. **(a)** Mean VMR differences for L3W (black) and $L3O_M$ (red) compared to L3L (L3L – L3* in both cases). Shown are the differences for all coastal grid boxes, and only for those grid boxes where the difference is significant ($p < 0.1$), determined using a 2-tailed t-test. **(b)** Absolute mean VMR differences between L3L and L3W, stratified according to corresponding AK rowsum difference (L3L – L3W in both cases). **(c)** Absolute differences in gradients detected using WLS regression analysis for L3W (black) and $L3O_M$ (red), compared to L3L (L3L – L3* in both cases). Shown are differences for all coastal grid boxes where WLS analysis could be performed, for grid boxes where both trends compared are significantly different to zero ($p < 0.1$), and for grid boxes where the trend difference is significant ($p < 0.1$). **(d)** Absolute differences in gradients detected using WLS regression analysis between L3L and L3W, stratified according to corresponding AK rowsum difference (L3L – L3W in both cases). Shown are the differences for all coastal grid boxes where WLS could be performed (black), and only for those grid boxes where the detected trend is significant ($p < 0.1$) in both L3L and L3W (blue).

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

Table 1. Mean values for selected variables from L3L and L3W for coastal L3 grid boxes, matched retrievals only. "Matched" means that only days when both L3L and L3W are present and the L3O surface index are mixed are used to create the mean values analysed. Mean values are calculated and presented separately according to the results of a 2-tailed student's t-test (unequal variance) performed on mean retrieved VMR values in L3L and L3W (n = 3971).

	P < 0.1 (n=2379, 60 %)			**P > 0.1 (n=1592, 40 %)**		
	land	water	d	land	water	d
Mean vmr_ret	129.97	117.41	12.55	133.52	126.60	6.90
Mean vmr apr	113.78	113.18	0.61	124.65	123.83	0.83
Mean ret-apr	16.18	4.24	11.94	8.87	2.77	6.09
Mean ak rowsum	0.43	0.18	0.24	0.44	0.27	0.16

subsets would be expected to be identical irrespective of any differences in retrieval sensitivity over those surfaces. To summarise: assuming "true" VMRs are similar over land and water within coastal L3 grid boxes, differences in retrieved VMRs depend not only the sensitivity of the retrieval, but also on the accuracy of a priori VMRs used in the retrievals.

Trend comparison between L3L and L3W

Temporal trends detected in L3L and L3W are now compared. The aim here is to demonstrate that there are trend differences between L3L and L3W datasets for coastal gridboxes that can be related to AK differences. An underlying assumption is that the temporal trend in "true" VMRs should not vary much across a 1° x 1° L3 grid box. Hedelius et al. (2021) lends credence to this assumption with the finding that CO trends are similar within regions spanning a few thousand kilometres (L3 grid boxes are ~ 100 km^2), and that trends within urban areas are generally indistinguishable from the trend of the broader region encompassing the urban area, despite an assumption that urban trends should exceed the regional background due to a concentration of CO emissions here. Gradients obtained from WLS regression on each dataset are compared, subtracting the trend in L3W from the trend in L3L. For clarity, differences between the absolute trend values (i.e. ignoring the +/- sign of the trend) are presented, since this shows the degree of difference in the trend magnitude, irrespective of trend direction. A positive trend difference in this case signifies a stronger (faster) trend in L3L than L3W.

Atmospheric Measurement Techniques Discussions

Table 2. Descriptive stats corresponding to the WLS trends detected in L3L, L3W, and L3O_M that are compared in the boxplots of Fig. 5c.

			Mean	Std	Median	IQR
All	L3L – L3W (n = 2670)	L3L	-0.55	1.27	-0.47	1.00
		L3W	-0.49	1.08	-0.34	0.65
	L3L – L3O_M (n = 2670)	L3L	-0.55	1.27	-0.47	1.00
		L3O_M	-0.51	1.03	-0.38	0.73
Both significant ($p < 0.1$)	L3L – L3W (n = 641)	L3L	-1.39	1.66	-1.15	1.08
		L3W	-1.06	1.56	-0.78	0.92
	L3L – L3O_M (n = 873)	L3L	-1.24	1.64	-1.06	1.07
		L3O_M	-1.02	1.38	-0.83	0.88
Significantly different ($p < 0.1$)	L3L – L3W (n = 956)	L3L	-0.64	1.39	-0.65	0.92
		L3W	-0.52	1.06	-0.43	0.67
	L3L – L3O_M (n = 555)	L3L	-0.69	1.36	-0.67	0.85
		L3O_M	-0.60	1.00	-0.51	0.68

On average, across all grid boxes where WLS can be performed in both datasets following the criteria outlined in Sect. 2.5 (n = 2670), trends are stronger in L3L than L3W (Fig. 5c (black boxplots)), with the range of differences around 2.5 ppbv y^{-1} (~-1 ppbv y^{-1} to 1.5 ppbv y^{-1}). When the comparison is restricted to grid boxes where both trends are significantly different to zero ($p < 0.1$; 641 of the 2670 grid boxes, 24 %), a greater proportion of those grid boxes have a stronger trend in L3L than L3W (> 75%), but the overall range of differences doesn't shift by much. The L3L – L3W trend difference is significant in 956 of the 2670 coastal grid boxes for which the analysis can be performed (36 %), with the range in differences spanning around 4 ppbv y^{-1}. Analysis of the distribution of differences between the raw trend values (i.e. including the +/- sign) is complicated because negative differences can have multiple meanings. However, trends are negative at 75 % of coastal grid boxes in both datasets (this increases to 95% when the trend in both L3L and L3W is significant), with retrieved surface CO concentrations over land decreasing at a faster rate than over water, on average (descriptive stats associated with raw trend values are detailed in Table 2).

To determine whether differences in trend can be linked to differences in retrieval sensitivity, L3L – L3W trend differences (again, based on absolute WLS trend values) are stratified by L3L – L3W surface level AK rowsum differences (Fig. 5d). As with mean VMR differences, the size of the trend difference tends to increase as the difference in AK rowsums increases. In addition, as the magnitude of rowsum difference increases in the positive direction (i.e. increasingly greater sensitivity over land), a greater proportion of trend

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

differences are positive (i.e. a stronger trend over land). This pattern is even more pronounced when restricted to grid boxes where both trends are significant (also shown in Fig. 5d).

All together, these results show that within coastal L3 grid boxes, differences in retrieval sensitivity over land and water are related to differences in temporal trends identifiable in corresponding surface-level retrievals. The relationships found in these analyses are not perfect because trend differences are sensitive to several other factors, in addition to differences in retrieval sensitivity. For example, a greater trend difference would be evident if the rate of change in "true" CO concentrations is faster than if it is slow/negligible, for a given sensitivity difference. Similarly, there should be zero trend difference if "true" CO concentration levels are stable over time, irrespective of the magnitude of difference in retrieval sensitivity. The accuracy of the a priori is a further complicating factor. However, the results presented do imply a general tendency for trend underestimation in retrievals over water within coastal grid boxes compared to retrievals over land in the same grid boxes obtained at the same times, which appears to be linked to differences in retrieval sensitivity.

3.2.2. Consequences for L3Omixed

To recap, L3O data are given the surface index "mixed" when neither land nor water is the dominant surface type of the bounded L2 retrievals, for a given retrieval time. When this is the case, the retrievals over land and water are averaged together. Users of L3O data do not have the option of choosing to only analyse the subset of retrievals made over land (L3L) or water (L3W), as was done in the preceding analysis. To do so requires the original L2 retrievals. In this section, the $L3O_M$ retrievals are compared to the L3L retrievals that were analysed in the previous section. The aim here is to demonstrate how, for some L3 grid boxes, information on "true" VMRs and temporal trends that is available in the L2 retrievals over land (L3L) is effectively lost to users of L3O data by their averaging together with the less sensitive L2 retrievals over water (L3W).

Retrieved VMRs in $L3O_M$

For long-term mean VMRs, $L3O_M$ unsurprisingly represents a mid-point between L3L and L3W, with lower VMRs and weaker trends than L3L, but a smaller difference range overall than L3W (Fig. 5a). The L3L – $L3O_M$ differences in long-term mean VMR are significant at 45 % (1791) of coastal grid boxes. All but 3 of these also see a significant difference between long-term mean VMRs in L3L and L3W. This makes sense: retrievals in L3L would not be expected to differ significantly from those in $L3O_M$ if they do not also differ

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

significantly from L3W. In total, 75 % of grid boxes that feature a significant difference between L3L and L3W also see a corresponding significant difference between L3L and $L3O_M$. There are several notable differences between this subset of coastal grid boxes ("BOTH"), compared to those that see a significant difference between L3L – L3W but not between L3L and $L3O_M$ ("L3L_L3W_ONLY"; detailed in Table 3a):

- The grid boxes of BOTH see greater retrieved VMR differences between L3L and L3W than the grid box subset of L3L_L3W_ONLY (mean L3L – L3W difference of 13.84 vs 8.67 ppbv). This is logical: $L3O_M$ only differs significantly from L3L if the underlying L3L – L3W difference is sufficiently large to persist through the averaging.
- The grid boxes of BOTH also feature a greater land-water sensitivity contrast than those of L3L_L3W_ONLY. This is indicated both by L3L – L3W AK rowsum differences, driven predominantly by decreased sensitivity over water in BOTH; and by L3L – L3W retrieved minus a priori VMR differences.
- The grid boxes of BOTH tend to have a greater proportion of their surface covered by water than land. This is quantified by comparing the mean number of L2 retrievals over land and water that are averaged together to make L3L and L3W each day ("n_ret(L3L)" and "n_ret(L3W)"), for each coastal grid box compared. A mean n_ret(L3L/L3W) ratio of 0.87 for BOTH indicates a greater water influence on $L3O_M$ than for the grid boxes of L3L_L3W_ONLY, for which a mean n_ret(L3L/L3W) ratio of 1.00 indicates a more even land/water split. Thus, $L3O_M$ more closely resembles L3W – which is significantly different to L3L – in BOTH than in L3L_L3W_ONLY.

It is easy to understand how each of these can lead to a $L3O_M$ retrieval that differs significantly from the corresponding L3L retrieval. In addition to the differences between the grid boxes of BOTH and L3L_L3W_ONLY outlined above, it is also notable that retrieved and a priori VMRs are lower in BOTH than in L3L_L3W_ONLY, and that retrieved minus a priori VMR values are greater in BOTH than in L3L_L3W_ONLY. This could imply that the a priori VMRs are closer to reality for the grid boxes of L3L_L3W_ONLY than those of BOTH, however further information on "true" VMRs is required to properly assess this.

Trends in $L3O_M$

Temporal trends detected in $L3O_M$ are now compared to those in L3L. Overall, a greater number of grid boxes feature a significant trend in both L3L and $L3O_M$ than in L3L and L3W (873 vs 641; 33 % vs 24 %),

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

540

Table 3a. Descriptive stats corresponding to matched retrievals over land and water (L3L and L3W) where the long-term mean retrieved surface level VMR in L3L and L3W is significantly different ($p < 0.1$, $n = 2379$). Grid boxes are divided into two subsets depending on whether long-term mean VMRs in L3L and $L3O_M$ are significantly different ($p < 0.1$; "BOTH") or not ($p > 0.1$; "L3L_L3W_ONLY"). n_ret(L3L) (n_ret(L3W)) = the number of L2 retrievals over land (water) used to make a retrieval in $L3O_M$. A ratio n_ret(L3L/L3W) value > 1 (< 1) implies that more of the L3 grid box surface is covered by land (water).

	BOTH (n = 1788, 75 %)			**L3L_L3W_ONLY (n = 591, 25 %)**		
Mean n_ret(L3L/L3W)	0.87			1.00		
	Land	**Water**	**L-W**	**Land**	**Water**	**L-W**
Mean vmr_ret	127.21	113.37	13.84	138.30	129.64	8.67
Mean vmr_apr	109.11	108.62	0.49	127.94	126.96	0.98
Mean ret-apr	18.11	4.75	13.36	10.36	2.68	7.68
Mean AK rowsum	0.42	0.16	0.26	0.46	0.26	0.20

Table 3b. Descriptive stats corresponding to matched retrievals over land and water (L3L and L3W) where the temporal trend detected using WLS regression analysis on yearly-mean retrieved surface level VMR in L3L and L3W is significantly different ($p < 0.1$, $n = 956$). Grid boxes are divided into two subsets depending on whether the trend in L3L is significantly different to the corresponding trend detected in $L3O_M$ ($p < 0.1$; "BOTH") or not ($p > 0.1$; "L3L_L3W_ONLY"). n_ret(L3L) (n_ret(L3W)) = the number of L2 retrievals over land (water) used to make a retrieval in $L3O_M$. A ratio n_ret(L3L/L3W) value > 1 (< 1) implies that more of the L3 grid box surface is covered by land (water).

	BOTH (n = 447, 47 %)			**L3L_L3W_ONLY (n = 509, 53 %)**		
Mean n_ret(L3L/L3W)	0.77			0.99		
	Land	**Water**	**L-W**	**Land**	**Water**	**L-W**
Mean WLS trend	-0.72	-0.58	-0.14	-0.58	-0.47	-0.11
Mean ABS WLS trend	1.18	0.76	0.42	1.04	0.68	0.35
Mean trend standard error	0.55	0.39	0.16	0.58	0.36	0.22
Mean vmr_ret	128.25	121.36	6.90	129.22	120.20	9.02
Mean vmr_apr	117.21	117.13	0.08	116.01	115.73	0.29
Mean ret-apr	11.05	4.22	6.82	13.21	4.47	8.74
Mean AK rowsum	0.46	0.22	0.25	0.44	0.20	0.24

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

and fewer see a significant difference between trends (555 vs 956; 21 % vs 36 %). The trends in L3L and L3O$_M$ are significantly different in just under half (47 %) of the grid boxes where the trend is also significantly different between L3L and L3W ("BOTH"; Table 3b). The ratio n_ret(L3L/L3W) clearly shows that these grid boxes are more water-dominated (mean ratio = 0.77) than the remaining 53 % grid boxes that feature a significant difference between trends in L3L and L3W but not L3O$_M$ ("L3L_L3W_ONLY"; mean ratio = 0.99). Additionally, detected trends in the grid boxes of BOTH are slightly stronger, with a greater difference between L3L and L3W, than for the L3L_L3W_ONLY subset. Those L3 grid boxes featuring the strongest land-water trend difference are therefore most likely to also see a significant trend difference between L3L and L3O$_M$. Again, this is logical. Unlike with the retrieved VMR comparison above, there are no clear differences in mean retrieved or a priori VMRs, nor sensitivity metrics, between these two grid box subsets (also detailed in Table 3b). However, it is not necessarily expected that there would be clear differences in these parameters for this analysis, since trend magnitudes themselves are a variable.

Most of the grid boxes where the L3L and L3O$_M$ trends are significantly different also feature a significant difference between L3L and L3W (453 of 555; 82 %). There are no clear differences between these and the remaining 18 % of grid boxes that, counter-intuitively, feature a significant difference between trends in L3L and L3O$_M$ but not between trends in L3L and L3W. However, small discrepancies are to be expected for results based on statistical thresholds, especially where the variables being compared are subject to multiple different factors (e.g. land-water surface cover ratio in L3O$_M$; land-water sensitivity contrast; retrieved VMR differences; differences in the "true" CO concentration being retrieved and its change over time).

3.3. Implications for users of L3O data

So far, this paper has shown a clear difference in retrievals over land and water for coastal grid boxes, demonstrated how results using these retrievals (L3L and L3W) differ, and outlined consequences (in terms of long-term VMR statistics and temporal trends) of averaging these retrievals together to create L3O$_M$. The full timeseries of available data in L3O is now compared with L3L and L3W, without the constraint that a retrieval needs to be present in both L3L and L3W for it to be included in the analysis. This replicates what a user of the L3O data would do, i.e., work with all available data.

Users of MOPITT data are advised to restrict their analysis to retrievals performed over land. This poses a quandary for users of L3O: what to do about days with a surface index of mixed? Therefore, the implications of choosing to include or discard these days are also considered. In the subsequent sections, the following subsets of the full L3O timeseries for each coastal gridbox are analysed: the full L3O timeseries with no filtering by surface index ("$L3O_{NF}$"); only days with a surface index of land ("$L3O_{L}$"); and days where the surface index is land or mixed ("$L3O_{LM}$" – i.e., only days with a L3O surface index of water are discarded).

3.3.1. Loss of available data

This section analyses how L3L and the L3O subsets compare in terms of the overall number of days with retrievals available for analysis at coastal grid boxes (total coastal grid boxes = 4299). To begin with, these are compared to the total number of days available for analysis in $L3O_{NF}$ ("n_days($L3O_{NF}$)"; Fig. 6a). Most strikingly, 67 % of coastal grid boxes see an L3O surface classification of land less than 5% of the time (yielding a n_days($L3O_{L}$/$L3O_{NF}$) ratio of 0.05 or less). Just less than half of these (35 % of the 4299 total coastal grid boxes) have zero days classified as land in the L3O dataset (note that as a result of how coastal grid boxes are classified (outlined in Sect. 2.3), all n_days($L3O_{L}$/$L3O_{NF}$) ratios are below 0.5 (i.e. at best, L3O has a surface classification of land on 50% of days)). The guideline to only analyse retrievals performed over land thus results in a huge loss of data for coastal grid boxes when using the L3O dataset. Importantly, retrievals over land are made on a large proportion of these filtered days; but they are either discarded altogether, or averaged together with retrievals made over water. This point is demonstrated by comparison to the n_days(L3L/$L3O_{NF}$) ratio. In contrast to a mean (median) n_days($L3O_{L}$/$L3O_{NF}$) ratio of 0.08 (0.01), a mean (median) n_days(L3L/$L3O_{NF}$) ratio of 0.44 (0.40) demonstrates the stark loss of data, something which is further highlighted by the fact that well over half (56%) of coastal grid boxes see a n_days(L3L/$L3O_{L}$) ratio > 25 (Fig. 6b), indicating that these L3 grid boxes have at least 25 times more days with retrievals made over land than are available for analysis in the L3O dataset, if filtering guidelines are followed.

The situation can be improved for L3O users by keeping days when the L3O surface index is classified as mixed, in addition to land ("$L3O_{LM}$"). Even in this best-case scenario however, $L3O_{LM}$ sees less days with data than L3L for 61% of coastal grid boxes (the ratio n_days(L3L/$L3O_{LM}$) is > 1 for 61 % of coastal grid boxes; Fig. 6b). Moreover, the large proportion of these $L3O_{LM}$ days where the surface index is mixed suffer from the averaging together of retrievals over land with retrievals over water which, as has been shown, can significantly impact the results of analyses using these data. This point is returned to in following sections.

Intuitively, it is to be expected that the ratio n_days(L3L/$L3O_{LM}$) should *never* be < 1. L2 retrievals over land obviously contribute to days when L3O is classified as land, and should, by definition, also contribute to days when L3O is classified as mixed. In these cases, L3L will therefore also be present.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

However, there are two instances where L2 retrievals over land in fact do not contribute to a L3O retrieval classified as mixed. Firstly, L2 retrievals themselves also have a classification of mixed, when the L2 retrieval does not predominantly overlie water or land. L3O can thus have a surface classification of mixed when created from bounded L2 retrievals that are either only retrieved over a mixed surface, or a combination of mixed and water: in both cases, there are no L2 retrievals over land, and therefore no L3L. Secondly, analyses performed for this paper identified numerous instances where L3O is classified as mixed, but the only contributing L2 retrievals are retrievals over water. In these instances, L3O would therefore seem to be misclassified. On days when this is the case, there will be no corresponding L3L retrieval. This is documented further in the Supp. Mat. (SM4). Attempting to quantify the extent of this misclassification influence is beyond the scope of this paper. In the vast majority of cases where a given gridbox has a n_days(L3L/$L3O_{LM}$) ratio < 1, the difference is negligible (i.e. 75 % of these grid boxes have a ratio between 0.9 and 1). Irrespective, in terms of the number of days with retrievals available for analysis, L3L is an improvement over $L3O_{LM}$ for more grid boxes than it is not.

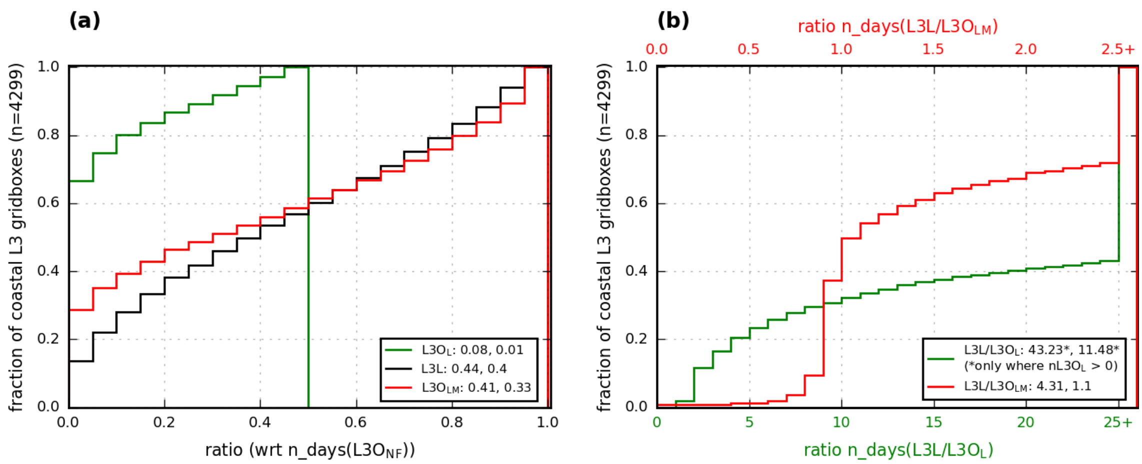


Figure 6. Cumulative frequency histograms comparing the number of days with data for different L3Osubsets and L3L at coastal L3 grid boxes. A ratio < 1 (> 1) indicates the plotted dataset has less (more) days with data than the comparison dataset, indicated on the x-axis. **(a)** $L3O_L$ (Green), L3L (black), and $L3O_{LM}$ (red) are compared to the "as-downloaded" L3O dataset, without any filtering by surface index ("$L3O_{NF}$"). Values in legend correspond to mean and median ratio for indicated dataset, respectively. **(b)** L3L is compared with $L3O_L$ (green line, bottom x-axis) and $L3O_{LM}$ (red line, top x-axis). Values in legend correspond to mean and median ratios, respectively.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

3.3.2. Scientific implications

Here, long-term mean (ltm) retrieved VMR values from the different L3 subsets are compared to L3L for all coastal grid boxes. As expected from the analyses in Sect. 3.2, all L3 subsets that have some influence from L2 retrievals over water have a ltm retrieved VMR that is below that in L3L, on average (Fig. 7a; note that the n value is different for each boxplot because not all L3 subsets are present at every coastal grid box, as discussed in Sect. 3.3.1). Unsurprisingly, the closest match to L3L is $L3O_L$ (mean difference -3.1 ppbv), with the mean difference increasing for each L3O subset as the influence of retrievals over water increases (e.g. $L3O_{LM}$ differs less on average from L3L (mean difference = 5.2 ppbv) than $L3O_{NF}$ (mean difference = 9.1 ppbv)).

Note that ltm retrieved VMRs in $L3O_L$ and L3L are not a perfect match because $L3O_L$ is only a subset of L3L for each grid box considered in the analysis: L3L may be present on a day when $L3O_L$ is not owing to the way that the L3O data are created (i.e., classified based on the ratio of L2 retrievals over land and water, with retrievals over land potentially being discarded if these are not the majority). Apart from $L3O_L$, less than 25 % of the coastal grid boxes have a retrieved ltm VMR that is greater in an L3O subset than in L3L. The range of ltm differences for each of these L3O subset comparisons to L3L exceeds 35 ppbv (excluding outliers), with over 25 % of coastal grid boxes compared having ltm differences exceeding 9 ppbv (as indicated by boxplot upper quartile values).

The percentage of coastal grid boxes that feature a significant difference between ltm retrieved VMRs in L3L and each L3O subset (indicated in red above each boxplot) is high: strikingly, it is found that, for the two subsets that L3O users could realistically choose to analyse if following data filtering guidelines, almost a quarter ($L3O_L$) or almost half ($L3O_{LM}$) of coastal grid boxes see a significant difference to L3L.

The results of WLS regression analysis on yearly mean values from each dataset are now compared. As expected from the earlier analysis, trends are strongest, on average, in L3L and $L3O_L$ – this is especially so when the comparison is restricted only to trends that are significantly different from zero ($p < 0.1$) (Table 4). These datasets also have the largest measures of spread, indicating their tendency to yield stronger trends than the other L3O subsets (and L3W), and these measures lessen for each L3O subset as the influence of retrievals over water increases (e.g. standard deviations and inter-quartile ranges are smaller for $L3O_{NF}$ and L3W than $L3O_L$). Concomitant with trends decreasing in strength as the influence of retrievals over water increases in each L3O subset, overall retrieval sensitivity also decreases, as indicated by the mean averaging kernel metrics shown in Table 4. Comparing the magnitude of trends at each coastal grid box, trends are stronger in L3L for at least 75% of grid boxes for all comparison datasets apart from $L3O_L$ (Fig. 7b). $L3O_L$ sees stronger trends than L3L on average, but the comparison of these two datasets needs to be interpreted

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

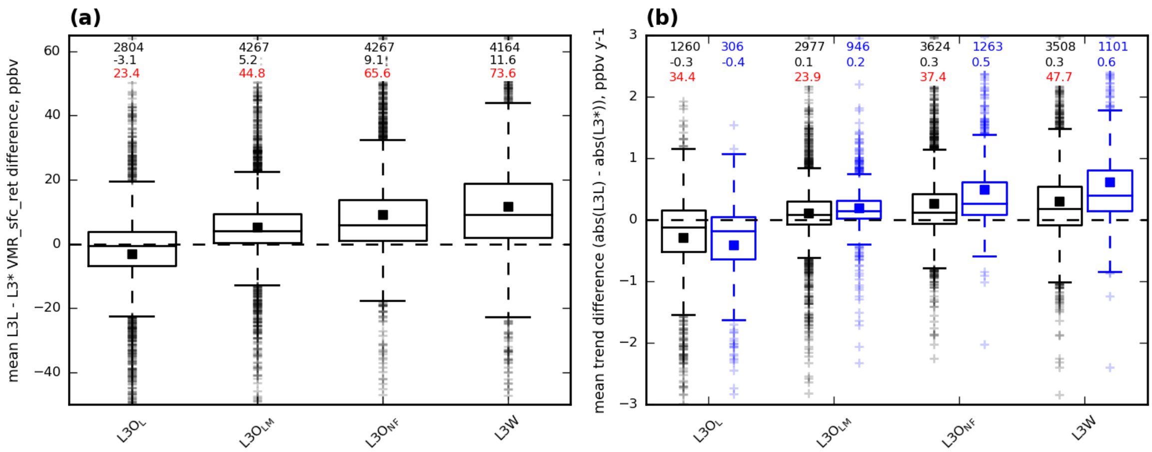


Figure 7. Boxplots showing how mean VMRs and trends compare from selected L3O subsets and L3W to L3L. Values compared are calculated from all available data across the study period. Mean values are represented by filled squares, and values above the boxplots correspond to number of grid boxes with data for that boxplot, the mean value, and the percentage of grid boxes represented in that boxplot that feature a significant difference with L3L (shown in red), respectively. The comparison is calculated as L3L – L3* in both cases; therefore a point above (below) the black y=0 line indicates that the value being compared is greater (lower) in L3L. **(a)** Mean VMR differences between L3L and the indicated L3O subset or L3W. **(b)** Differences in gradients (absolute values) detected using WLS regression analysis between L3L and the indicated L3O subset or L3W. Shown are the differences for all coastal grid boxes where WLS could be performed for both datasets compared (black), and only for the sample of those grid boxes where the detected trend is significant ($p < 0.1$) in both (blue).

Table 4. Descriptive stats corresponding to the WLS trends detected in L3L, L3W, and selected L3O subsets. Also shown are mean averaging kernel rowsums and diagonal values corresponding to the retrievals from which trends are calculated.

		L3L	$L3O_L$	$L3O_{LM}$	$L3O_{NF}$	L3W
Calculated from all gridboxes where WLS could be performed	Number of grid boxes	3624	1260	2999	4288	4169
	Mean (std) trend	-0.59 (1.22)	-0.52 (1.38)	-0.50 (0.95)	-0.54 (0.67)	-0.54 (0.66)
	Median (IQR) trend	-0.45 (0.89)	-0.46 (1.08)	-0.37 (0.67)	-0.42 (0.53)	-0.40 (0.54)
	Mean AK rowsum	0.45	0.45	0.33	0.28	0.22
	Mean AK diagonal value	0.10	0.10	0.08	0.07	0.06
Calculated only from gridboxes where WLS trend is significant ($p < 0.1$)	Number of grid boxes	1447	453	1265	2588	2499
	Mean (std) trend	-1.23 (1.55)	-1.17 (1.90)	-0.95 (1.18)	-0.79 (0.73)	-0.78 (0.72)
	Median (IQR) trend	-0.98 (0.94)	-1.09 (1.28)	-0.74 (0.75)	-0.62 (0.56)	-0.62 (0.57)
	Mean AK rowsum	0.51	0.48	0.39	0.33	0.29
	Mean AK diagonal value	0.11	0.10	0.08	0.07	0.06

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

with caution due to $L3O_L$ being a subset of L3L that features far fewer days with data, as discussed previously. Like with ltm retrieved VMRs discussed above, the percentage of coastal grid boxes that feature a significant difference between trends detected in L3L and each L3O subset is high, with over a third (almost a quarter) of the trends in $L3O_L$ ($L3O_{LM}$) being significantly different to L3L.

3.4. Illustrative examples comparing L3O and L3L: analysis of the most populous coastal cities

Ltm VMRs and temporal trends from $L3O_L$ and $L3O_{LM}$ (the L3O subsets that data users would realistically choose to analyse) are compared to those from L3L for the 33 coastal grid boxes that contain cities classified amongst the 100 most populous in the world (derivation outlined in Sect. 2.5).

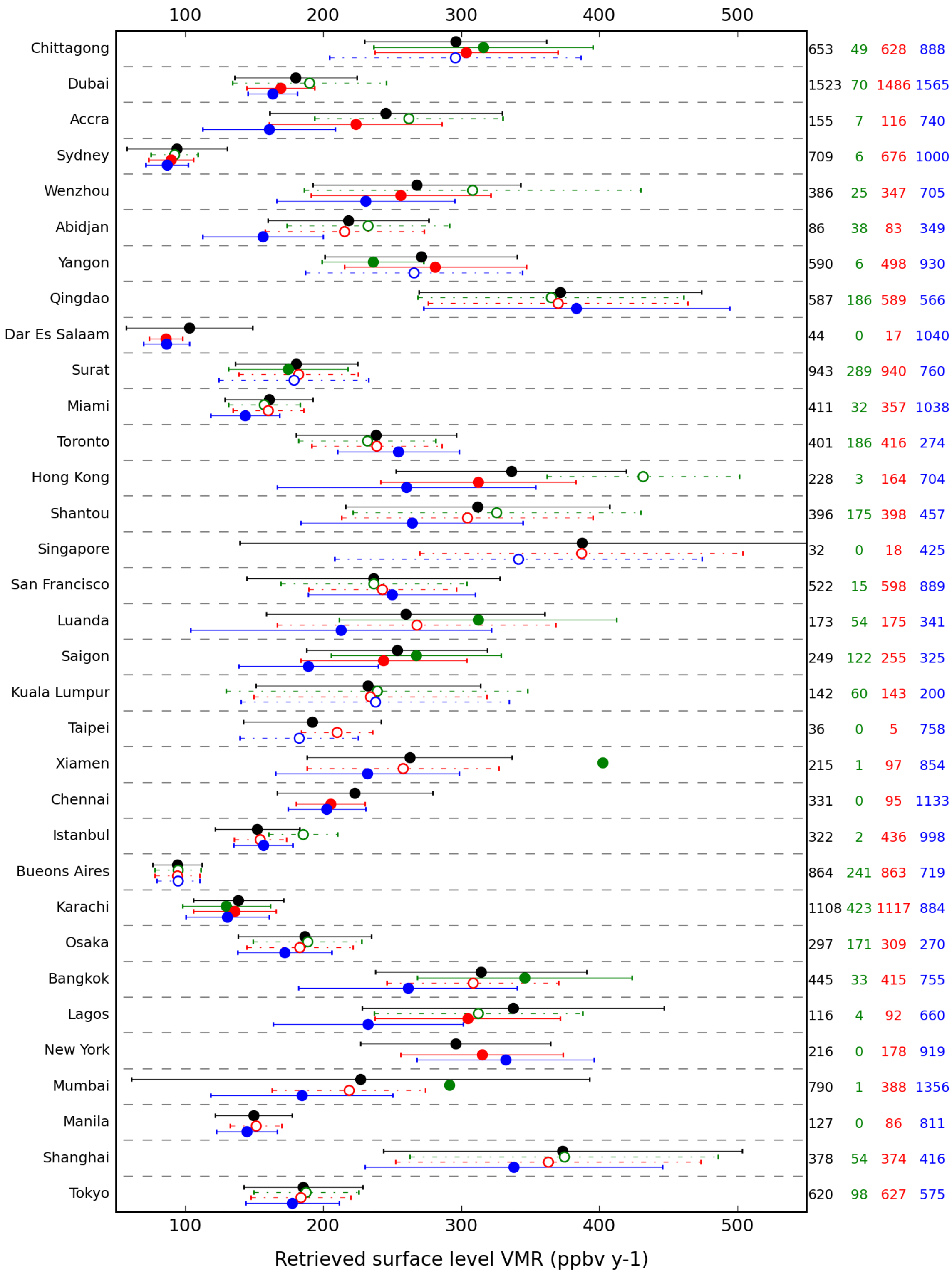
VMR comparison:

Mean VMRs calculated across the entire study period are shown in Fig. 8 for L3L, $L3O_L$, $L3O_{LM}$, and L3W (included for comparison purposes). Comparing $L3O_L$ to L3L, 6 of the 33 grid boxes analysed have no data in the $L3O_L$ subset. The mean n_days($L3O_L$/L3L) ratio for the remaining 27 cities is 0.19 (this raises slightly to 0.23 if an additional 5 cities with only a few days (< 5) of data coverage are excluded). Only a single city (Osaka) has more than 50 % of the L3L observation days in $L3O_L$. The consequence of this loss of data in $L3O_L$ is clear: mean VMR across all cities (excluding the 6 where n_days($L3O_L$) = 0) is 17.2 ppbv higher than in L3L (Table 5a). This falls to 9.8 ppbv if restricted to cities where the n_days($L3O_L$/L3L) ratio is greater than 0.05 (n=17), and 6.8 ppbv if restricted to cities where the n_days(L3OL/L3L) ratio is above 0.2 (n=11). Mean VMR is significantly different ($p < 0.1$) for 11 of the 27 cities with n_days($L3O_L$) > 0; comparing these with the remaining cities that see no significant difference, cities with a significant difference have a lower n_days($L3O_L$/L3L) ratio (i.e. relatively fewer retrieval days than L3L: ratio 0.15 vs 0.22), and much greater mean VMR differences (-36.49 vs -3.93 ppbv) (Table 5b).

$L3O_{LM}$ compares more favourably to L3L in terms of number of observations, thanks to the inclusion of days when the L3O surface index is "mixed", with a mean n_days($L3O_{LM}$/L3L) ratio of 0.85. n_days($L3O_{LM}$) > n_days(L3L) for 11 of the 33 cities, although the ratio is less than 1.05 (5 %) for all of these except San Francisco and Istanbul (ratio = 1.14 and 1.35, respectively). The L3L – $L3O_{LM}$ mean VMR difference is relatively small (3.7 ppbv, all 33 cities; Table 5a). However, this does hide some much larger discrepancies between L3L and $L3O_{LM}$ for certain cities, with the difference exceeding 10 ppbv for 11 of the 33 cities and 20 ppb for 3 of them. The difference is significant ($p < 0.1$; "SIGDIFF") for 13 of 33 cities (39

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions



https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques
Discussions

Figure 8. Comparison of long-term mean retrieved surface level VMR in L3L (black), $L3O_L$ (green), $L3O_{LM}$ (red), and L3W (blue), for the 33 largest coastal cities (ordered by population on the y-axis). The long-term mean value (in ppbv) is indicated by the filled/open circle on each row, and its standard deviation by the error bars. The L3L marker is always filled and lines are always solid. For other datasets, whether the marker is filled or not, and whether the lines are solid or dash/dot, depends on the outcome of an independent, 2-tailed t-test assuming unequal variance (aka "Welch's test") against L3L: filled markers and solid lines indicate the mean is significantly different to L3L ($p < 0.1$); open markers and dash/dot lines indicate there is no significant difference to L3L. The number of retrieval days in each time series analysed for each city is given on the right-hand y-axis, color-coded accorded to dataset.

Table 5. (a) Summary stats for mean VMRs across all cities analysed in Sect 3.4. "$L3O_X$" = L3O subset. **(b)** Selected parameters from L3 grid boxes containing cities where mean VMR in L3L and $L3O_L$ is significantly different ($p < 0.1$). **(c)** Selected parameters from L3 grid boxes containing cities where mean VMR in L3L and $L3O_{LM}$ is significantly different ($p < 0.1$).

(a)	**L3L (n = 33)**	**$L3O_L$ (n = 27)**	**$L3O_{LM}$ (n = 33)**	**L3W (n = 33)**
Mean (std) VMR_RET *(ppbv)*	235.6 (80.0)	255.2 (90.1)	231.9 (79.2)	212.2 (74.4)
Δ VMR_RET (L3L – $L3O_X$) *(ppbv)*	-	17.2	3.7	23.4

(b)	**P < 0.1 (n = 11)**	**P > 0.1 (n = 16)**
ratio n_days($L3O_L$/L3L)	0.15	0.22
Δ VMR_RET (L3L – $L3O_L$) *(ppbv)*	-36.49	-3.93

(c)	**P < 0.1 ("SIGDIFF") (n = 13)**	**P > 0.1 ("NOT_SIGDIFF") (n = 20)**
ratio n_ret(L3L/L3W)*	0.51	1.02
% days from $L3O_L$	9	20
Δ VMR_RET (L3L – L3W) *(ppbv)*	31.15	18.44
Δ AK rowsum (L3L – L3W)	0.25	0.21
Δ AK diagonal (L3L – L3W)	0.10	0.08
Δ VMR (RET - APR) (L3L – L3W) *(ppbv)*	21.66	3.22
|Δ VMR (RET - APR)| (L3L – L3W) *(ppbv)*	21.98	11.88
L3L VMR (RET - APR)	-19.82	-7.07
|L3L VMR (RET - APR)|	39.86	18.79
L3W VMR (RET - APR)	-14.75	-6.73
|L3W VMR (RET - APR)|	18.21	15.57

* n_ret(L3L) (n_ret(L3W)) = the mean number of L2 retrievals over land (water) that are averaged to make a L3L (L3W) retrieval

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

%). Compared to the subset where the L3L – $L3O_{LM}$ mean difference is not significant (n = 20, 61 %; "NOT_SIGDIFF"), the following characteristic differences are found (also detailed in Table 5c):

- The grid boxes are more water-dominated in SIGDIFF than NOT_SIGDIFF: this is evidenced by the ratio of n_ret(L3L/L3W) = 0.51 vs 1.02 respectively; and also by the fact that on average, $L3O_L$ only contributes to SIGDIFF on 9 % of days, vs 20 % of days for NOT_SIGDIFF (which means that retrievals over water contribute via $L3O_M$ more frequently to SIGDIFF than NOT_SIGDIFF).
- The L3L – L3W VMR_RET differences are larger in SIGDIFF than NOT_SIGDIFF (mean = 31.15 vs 18.44 ppbv), meaning they are less likely to be hidden by averaging to create $L3O_M$.
- Although analysis of mean averaging kernels over land and water suggest there is not a large sensitivity contrast between the SIGDIFF and NOT_SIGDIFF subsets (mean L3L – L3W rowsum (diagonal value) differences are 0.25 vs 0.21 (0.10 vs 0.08) for SIGDIFF and NOT_SIGDIFF cities, respectively), the L3L – L3W ret-apr difference, which is another indicator of sensitivity difference, is much greater for SIGDIFF than NOT_SIGDIFF: 21.66 vs 3.22 ppbv respectively (21.98 vs 11.88 ppbv if using absolute values). There is some evidence that this may be a function of the a priori values being closer to "true" VMRs in NOT_SIGDIFF: mean L3L retrieved minus a priori VMR values fall from -19.82 ppbv for SIGDIFF to -7.07 ppbv for NOT_SIGDIFF (39.86 ppbv and 18.79 ppbv respectively, if using absolute values). A similar pattern is seen in L3W, although less pronounced (-14.75 and -6.73 ppbv, respectively (18.21 and 15.57 ppbv if using absolute values)).

These findings are all consistent with what was shown in Sect. 3.2.2 when identifying factors that determine whether the averaging of L2 retrievals over land and water to create $L3O_M$ (the dominant component of $L3O_{LM}$ in all cases here, being the classification on 84 % of days, on average (max = 100 %, min = 45 %)) can yield a statistically significantly different retrieval to L3L.

Trend comparison:

The above analysis is repeated with temporal trends detected using WLS regression. The trend values, their associated standard errors, and an indication of their statistical significance ($p < 0.1$) are presented for each city in Fig. 9. Where trend information is not plotted from a dataset for a given city, this means that there were too few data points to perform the regression analysis.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

On average, the strongest trends are seen in L3O$_L$. However, this often appears as an outlier compared to the other datasets. As expected from previous sections, the weakest trends are detected in L3W, with L3O$_{LM}$ representing a mid-point between this and L3L.

Of the 18 cities where WLS analysis can be performed in L3O$_L$ according to the criteria outlined in Sect. 2.5, there are 9 cases where the trend is significantly different to that in L3L. In 3 of these cases (Dubai, Wenzhou, Bangkok), despite being significant, the trend in L3O$_L$ can be judged to be a huge over-estimate given the trend comparison to L3L (standard errors do not overlap) and the very small number of days with data compared to L3L (n_days(L3O$_L$/L3L) ratio < 0.08 in each case). There are 4 additional cities where a significant trend in L3O$_L$ appears to be an over-estimate on further analysis: Abidjan, Surat, Saigon, and Buenos Aires. The trend for these cities in L3O$_L$ is much stronger than in L3L (-1.6, -1.2, -2.2, and -0.4 ppbv/y, respectively), and the trend in L3L is also not significantly different to 0. Given the higher number of days with data in L3L however (n_days(L3O$_L$/L3L) ratio = 0.44, 0.31, 0.49, 0.28, respectively), this appears to be the more reliable result. The L3O$_L$ trend for Miami is insignificant and derived from very low n. L3O$_L$ is also the only dataset to yield an insignificant trend for Qingdao.

As with mean VMRs, trends in L3O$_{LM}$ compare better than L3O$_L$ to L3L. However, there are 5 cases where L3O$_{LM}$ and L3L yield significantly different results. For 3 of these (Dubai, Hong Kong, and Istanbul), interpretation of the difference is simple: L3O$_{LM}$ is a significant under-estimate of the CO change over time, likely due to the influence of retrievals over water on the dataset (L3W yields a significantly weaker trend than L3L in all 3 cases). In the remaining 2 cases – New York and Saigon – interpretation is more complicated. For both cities, the trend detected in L3L is not significantly different from zero, whereas the trend in L3O$_{LM}$ is. Does this mean that the trend in L3O$_{LM}$ is an over-estimate? Possibly. However, in both cases, the trends are within one standard error of eachother and therefore within the range of sampling uncertainty. There are an additional 2 cities where WLS could be performed in L3L but not L3O$_{LM}$ (Dar Es Salaam and Taipei), but n_days(L3L) is so low (44 and 36, respectively) that these results are not deemed to be trustworthy.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

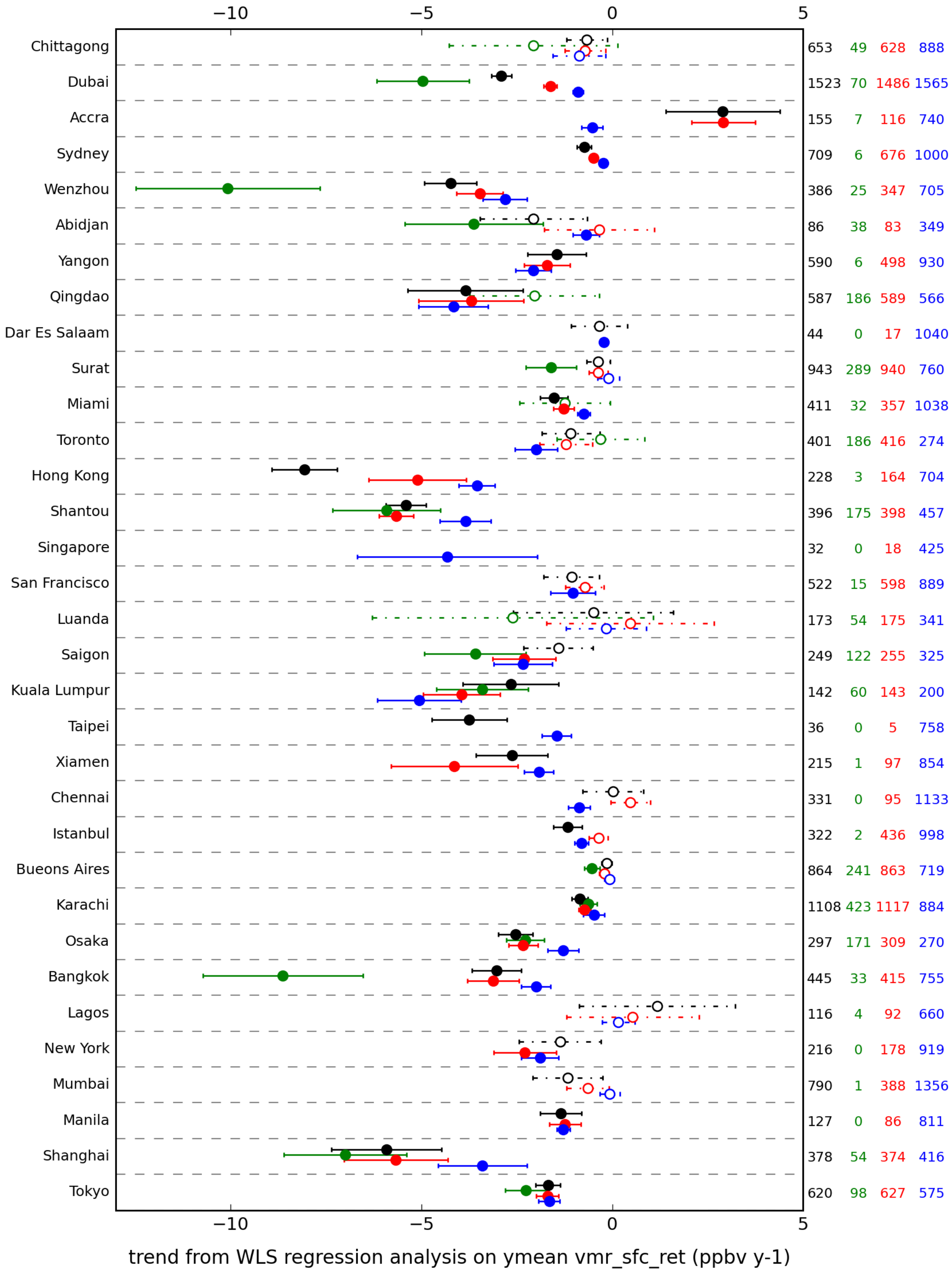


Figure 9. Comparison of temporal trend (detected using WLS, as outlined in Section 2.5) in retrieved surface level VMR in L3L (black), $L3O_L$ (green), $L3O_{LM}$ (red), and L3W (blue), for the 33 largest coastal cities (ordered by population on the y-axis). The trend value (in ppbv y^{-1}) is indicated by the filled/open circle on each row, and its standard error by the error bars. For all datasets, whether the marker is filled or not, and whether the lines are solid or dash/dot, depends on the significance of the trend: filled markers and solid lines indicate the trend is significant ($p < 0.1$); open markers and dash/dot lines indicate that the trend is not significantly different to zero. The number of retrieval days in each time series analysed for each city is given on the right-hand y-axis, color-coded accorded to dataset.

4. Summary and Conclusions

Motivated by the work of Ashpole and Wiacek (2020) which demonstrated, for the MOPITT L3 grid box containing coastal city of Halifax, Canada, that mean VMR statistics and temporal trends differ depending on whether L2 or L3 data are analysed, this paper has examined what proportion of all coastal L3 grid boxes also see differences between results from analyses performed with L2 and L3 data. While it is recommended

to MOPITT data users that analyses are restricted to retrievals performed over land owing to known sensitivity issues over water (MOPITT Algorithm Development Team, 2018; Deeter et al., 2015), such recommendations cannot practically be followed by users of L3 data for coastal grid boxes owing to the way the data are created from their bounded L2 retrievals. In short, this study haa sought to answer the question: "does it matter"? The main results are summarised below.

First, a direct comparison of the L2 retrievals performed over land (L3L) and water (L3W) that are averaged together to create L3 products on days when the L3 surface index is "mixed" ($L3O_M$) identified that:

- Retrieval information content is clearly greater in L3L than L3W. The mean L3L – L3W VMR difference is over 10 ppbv, significant ($p < 0.1$) at 60 % of the coastal grid boxes compared. Temporal trends are also stronger, on average, in L3L (mean diff = 0.28 ppbv y^{-1}, 0.43 ppbv y^{-1} if only considering trends significantly different to zero), with the L3L – L3W trend difference significant ($p < 0.1$) at 36 % of grid boxes where a trend comparison was possible. The largest L3L – L3W differences in mean VMRs and trends are clearly associated with greater differences in retrieval sensitivity.
- The resulting VMRs in $L3O_M$ are significantly different to L3L for 75 % of grid boxes where the L3L – L3W difference is also significant; this corresponds to 45 % of all coastal grid boxes

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Measurement Techniques Discussions

compared. Whether or not $L3O_M$ and L3L differ significantly depends on multiple factors including the ratio of land/water surface cover in the grid box, the strength of the land-water sensitivity contrast and VMR difference, and, potentially, the accuracy of the a priori. Just under half of the grid boxes that featured a significant L3L – L3W trend difference also see trends differing significantly between L3L and $L3O_M$. As with the mean VMR comparison, these grid boxes are more water-dominated than the subset whereby the L3L – L3W trend difference is significant but the L3L – L3OM trend difference is not. They also feature stronger L3L – L3W trend differences overall, but no other variables (such as ltm VMRs and sensitivity metrics) show clear differences.

Having established the degree of difference in $L3O_M$ and L3L retrievals that is caused directly by averaging L3L with the less-sensitive L3W, the full L3O dataset with differing surface filtering options was compared to L3L:

- If L3O is filtered so that only retrievals over land ($L3O_L$) are analysed, as has been recommended (MOPITT Algorithm Development Team, 2018; Deeter et al., 2015), there is a huge loss of data, in terms of days with data to analyse. This is a direct result of L2 retrievals over land routinely being discarded during the L3O creation process (at least for coastal grid boxes). The problem can be alleviated by also retaining $L3O_M$ retrievals, but these additional days with data feature some influence from retrievals made over water that can affect results, as outlined. The resulting $L3O_{LM}$ subset still has less days with data than in L3L for 61 % of coastal grid boxes.
- Almost a quarter (half) of coastal grid boxes see a significant difference in ltm VMR between L3L and $L3O_L$ ($L3O_{LM}$). Over a third (almost a quarter) of the trends in $L3O_L$ ($L3O_{LM}$) are significantly different to L3L.
- Focusing on the 33 largest coastal cities in the world, mean VMRs in $L3O_L$ and L3L differ significantly for 11 of the 27 cities that can be compared (40 %; there are no $L3O_L$ data for the remaining 6 cities). The L3L – $L3O_{LM}$ mean VMR difference across all 33 cities is relatively small (3.7 ppbv), but this does hide some much larger discrepancies, with the difference exceeding 10 ppbv for 11 of the 33 cities and 20 ppbv for 3 of them. The difference is significant for 13 of 33 cities (39 %). Of the 18 cities where WLS analysis can be performed in $L3O_L$, there are 9 cases where the trend is significantly different to that in L3L. The trends in $L3O_{LM}$ and L3L differ significantly for 5 of the 33 cities.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

From these results, it can be concluded that, yes, for at least a quarter of all MOPITT coastal L3 grid boxes, it does matter that there is limited capacity to filter out the influence of retrievals over water in L3 data – at least without a huge loss of temporal coverage. Demonstrably, there are significant differences in mean VMRs and temporal trends between L3O and L3L, sometimes very large. These differences could have tangible consequences, depending on the purpose for which the MOPITT data are being used. While acknowledging that this analysis has also shown that there is a sizeable proportion of coastal grid boxes where statistically, mean VMRs and trends do not differ significantly between L3L and L3O, there is enough evidence to support the suggestion from Ashpole and Wiacek (2020) that an additional L3 "land-only" product, created only from averaging bounded L2 retrievals performed over land – the L3L dataset that has been analysed in this paper – would be beneficial to the research community. This dataset would enable L3 users to maximize retrieval information content for coastal L3 grid boxes, as is currently only possible with L2 data, while also preserving the benefits of L3 products. Although this analysis has focused only on analysis of MOPITT data, it is reasonable to question whether the findings are applicable to data products from other satellite instruments that make CO retrievals based on observed thermal-infrared radiances, such as AIRS (Atmospheric InfraRed Sounder), TES (Tropospheric Emission Spectrometer), and IASI (Infrared Atmospheric Sounding Interferometer).

Data availability

MOPITT data were downloaded from the NASA Earthdata portal (https://search.earthdata.nasa.gov/). The L3L and L3W products analysed in this study are available on request from the corresponding author.

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.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

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 for providing the data used in this study.

References

Ashpole, I., & Wiacek, A.: Impact of land-water sensitivity contrast on MOPITT retrievals and trends over a coastal city, Atmospheric Measurement Techniques, 13(7), 3521–3542, https://doi.org/10.5194/amt-13-3521-2020, 2020.

Buchholz, R. R., Worden, H. M., Park, M., Francis, G., Deeter, M. N., Edwards, D. P., Emmons, L. K., Gaubert, B., Gille, J., Martínez-Alonso, S., Tang, M., Kumar, R., Drummond, J. R., Clerbaux, C., George, M., Coheur, P-F., Hurtmans, D., Bowman, K. W., Luo, M., Payne, V. H., Worden, J. R., Chin, M., Levy, R. C., Warner, J., Wei, Z., Kulawik, S. S.: Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions, Remote Sensing of Environment, 256, 112275, https://doi.org/10.1016/j.rse.2020.112275, 2021.

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.: 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, 2003.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R.: 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, 2007.

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.: 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, 2013.

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

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.: 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, 2014.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Worden, H. M.: 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, 2015.

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.: 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, 2019.

Deeter, M., Francis, G., Gille, J., Mao, D., Martínez-Alonso, S., Worden, H., Ziskin, D., Drummond, J., Commane, R., Diskin, G., and McKain, K.: The MOPITT Version 9 CO Product: Sampling Enhancements and Validation, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2021-370, in review, 2021.

Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J.: 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, 2010.

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.: 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, 2007.

Edwards, D. P., Halvorson, C. M., and Gille, J. C.: 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, 1999.

Francis, G. L., Deeter, M. N., Martínez-Alonso, S., Gille, J. C., Edwards, D. P., Mao, D., Worden, H. M., and Ziskin, D.: 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

https://doi.org/10.5194/amt-2022-90
Preprint. Discussion started: 6 April 2022

Atmospheric Research, Boulder, Colorado, downloaded from: https://www2.acom.ucar.edu/sites/default/files/mopitt/ATBD_5_June_2017.pdf, 2017.

Hedelius, J. K., Toon, G. C., Buchholz, R. R., Iraci, L. T., Podolske, J. R., Roehl, C. M., Wennberg, P. O., Worden, H. M., Wunch, D.: Regional and Urban Column CO Trends and Anomalies as Observed by MOPITT Over 16 Years, Journal of Geophysical Research: Atmospheres, 126(5), 1–18, https://doi.org/10.1029/2020JD033967, 2021.

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.: 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, 2012.

MOPITT Algorithm Development Team: MOPITT (Measurements of Pollution in the Troposphere) Version 8 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/v8_users_guide_201812.pdf, 2018.

Pan, L., Edwards, D. P., Gille, J. C., Smith, M. W., and Drummond, J. R.: 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, 1995.

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

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

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.: 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, 2010.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J., Drummond, J., Emmons, L. K., Francis, G., and Martínez-Alonso, S.: 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, 2014.