Technical note: Detecting physically unrealistic outliers in ACE-FTS~~satellite-based~~ atmospheric measurements

P.E. Sheese[1], C.D. Boone[2], and K.A. Walker[1,2]

[1]Department of Physics, University of Toronto, Toronto, ON, M5S 1A7, Canada

[2]Department of Chemistry, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

Corresponding author: K.A. Walker

Abstract

The ACE-FTS (Atmospheric Chemistry Experiment – Fourier Transform Spectrometer) instrument on board the Canadian satellite SCISAT has been~~ing~~ observing the Earth's limb in solar occultation since its launch in 2003. Since February 2004, high resolution (0.02 cm^{-1}) observations in the spectral region of 750-4400 cm^{-1} have been used to derive volume mixing ratio profiles of over 30 atmospheric trace species and over 20 atmospheric isotopologues. Although the full ACE-FTS level 2 data set is available to users in the general atmospheric community, until now no quality flags have been assigned to the data~~ in order to guide the users~~. This study describes the two-stage procedure for detecting physically unrealistic outliers within the data set for each retrieved species, which is a fixed procedure across all species. Since the distributions of ACE-FTS data across regions (altitude/latitude/season/local time) tend to be asymmetric and ~~multi-~~multimodal, the screening process does not make use of the median absolute deviation. It makes use of volume mixing ratio probability density functions, assuming that the data, when sufficiently binned, are at most tri-modal and that these modes can be

represented by the superposition of three normal, or log-normal, distributions. Quality flags have been assigned to the data based on retrieval statistical fitting error, the physically unrealistic outliers described in this study, and known instrumental/processing errors. The quality flags defined and discussed in this study are now available for all level 2 versions 2.5 and 3.5 data and will be made available as a standard product for future versions.

Introduction

One of the most common techniques for screening out anomalous data from a data set is to calculate the set's mean (μ) and standard deviation (σ). Data that are outside the limits of $\mu \pm k\,\sigma$, where k is some constant, are deemed to be outliers. Another common, and similar, method is to use the median and MAD (Median Absolute Deviation) [*Leys et al*., 2013; *Toohey et al*., 2010; *and references therein*], in place of the mean and standard deviation respectively, where,

$$\mathrm{MAD} = median_i\left(\left|x_i - median_j(x_j)\right|\right). \quad (1)$$

This method is much less sensitive to extreme outliers, as the presence of outliers typically has an insignificant effect on the median value. It can be used as an efficient tool in detecting outliers for data that are normally distributed. However, ~~this technique assumes that~~using this value as a method of detecting outliers can be ineffective if the data being analyzed are ~~not multi-~~ multimodal and/or are asymmetrically distributed about the median. In the case of data that ~~is~~are multimodal or asymmetrically~~ or multi-modally~~ distributed and ~~consists~~ contain ~~of~~ multiple extreme outliers, it is likely that neither the σ nor the MAD will be an appropriate estimate of the variation, or scale, of the measurements. In such cases, they should be avoided in the detection of outliers [*Rousseeuw and Croux*, 1993].

In satellite remote sensing~~ of the atmospheric limb, measurements inherently suffer from~~

~~sampling biases. Therefore~~, isolating geographical/seasonal/local time regions where global satellite-based data are ~~normally~~ symmetrically distributed and uni-modal can be difficult and/or tedious. This is due to the fact that the atmosphere doesn't necessarily behave in a predictable, periodic manner. Measurements grouped into a given altitude and latitude and month and local time bin can be driven away from typical~~normal~~ behaviour by any number of factors~~—~~ (e.g., the polar vortex, a solar proton event, a sudden stratospheric warming, biomass burning, presence of polar stratospheric clouds, etc.), thereby altering the "typical" distribution of observed measurements, and hence the probability density function (pdf) of the trace species concentration.

The often used method for detecting outliers of employing the MAD does not explicitly make use of a pdf, but it does make the assumption the pdf is symmetric about the median value. Other often used methods, such as Peirce's criterion [*Peirce*, 1852; *Ross*, 2003] and Chauvenet's criterion [*Chauvenet*, 1871], explicitly make use of a pdf, however they assume that the pdf is a Gaussian distribution. It should also be noted that in atmospheric science, the use of pdfs is not uncommon in tracer and validation studies. *Lary and Lait* [2006], in their introduction, give excellent examples of different types of tracer studies; and studies such as *Migliorini et al.* [2004], *Lary and Lait* [2006], and *Wu et al*. [2008] have demonstrated that pdfs can be used as a validation tool, where pdfs as measured by different atmospheric sounders are inter-compared rather than inter-comparing co-located measurements.

The ACE-FTS (Atmospheric Chemistry Experiment – Fourier Transform Spectrometer [*Bernath et al*., 2005]) instrument, on board the Canadian satellite SCISAT, is a solar occultation, high spectral-resolution (0.02 cm^{-1}) Fourier transform spectrometer operating between 750 and 4400 cm^{-1}. ACE-FTS observations are used to derive volume mixing ratio

(VMR) profiles of over 30 atmospheric trace gases, as well as profiles of over 20 subsidiary isotopologues of atmospheric species [*Boone et al*., 2005]. SCISAT was launched in 2003 and ACE-FTS has been providing consistent measurements since February 2004. Atmospheric profiles range in altitude from ~5-110 km, depending on the species, with a vertical resolution of ~3-4 km and sampling of 1~~2~~-6 km.

This study outlines the repercussions of screening data based on the σ~~standard deviation~~ or the MAD given non-normally distributed data and discusses a two-step process for detecting outliers that is currently carried out on the ACE-FTS level 2 data set. All data presented in this study are ACE-FTS level 2 version 3.5 (v3.5) [*Boone et al*., 2013] spanning February 2004 to February 2013, however the same processes have been~~are~~ used for detecting outliers in version 2.5 (v2.5) data. The main differences in v3.5 from v2.5 are,

- Amended sets of microwindows for all molecules, and an increase in the number of allowed interferers in the retrievals;
- Improvement in temperature/pressure retrievals, leading to a reduction in unphysical oscillations in retrieved temperature profiles;
- Inclusion of $COCl_2$, COClF, H_2CO, CH_3OH, and HCFC-141b and removal of HOCl and ClO VMR retrievals.

Physically unrealistic outliers can occur in the ACE-FTS level 2 for a number of different reasons. Many of these are often caught prior to being added to the level 2 database, such as outliers due to exceedingly noisy spectra, ice contamination affecting an occultation, and a variety of processing errors. However, these aren't always caught by pre-screening, and other factors, not accounted for in the pre-screening, can contribute to the presence of outliers, for example, poor statistical fitting or convergence onto an unrealistic solution in the retrieval, inaccurate pressure and temperature a priori information.

The outlier detection and subsequent data flagging procedures discussed in this study have only been performed on the ACE-FTS level 2 data products that have been interpolated onto a 1-km altitude grid (between 0.5 and 149.5 km) [*Boone et al.*, 2005]. The philosophical approach for ~~flagging~~ identifying data as potential outliers was one of caution, in that it is better to keep some “bad” data (likely to be physically unrealistic) than to reject “good”, or “true”, data (likely to be physically realistic).~~.~~ ~~As well, i~~It was also desired that the approach be consistent for all subsets of data being analyzed, i.e. tolerance levels, regional limits, etc. should be the same for all species, for all seasons, at all altitudes. For the remainder of this study, these physically unrealistic data will be referred to as “unnatural” outliers, and the data that are likely to be physically realistic yet still seemingly outlying as “natural” outliers. All data that are not unnatural outliers will be referred to as inliers.

Detection method and results

All distributions of data discussed in this section represent the February 2004-February 2013 data, and all VMRs are given in parts per volume (ppv).

Global satellite-based measurements of trace gases in the atmosphere typically are not symmetrically~~normally~~ distributed and are often multimodal. Different regions are governed by different, varying processes, and therefore analysis of the data is typically carried out by breaking down the data into different altitude, latitudinal, etc. bins. Figure 1 shows all the ACE-FTS H_2O data at 17.5 and 35.5 km and the corresponding measurement distributions. For both subsets of H_2O data, inlier limits were determined for $\mu \pm 3\sigma$ and $median \pm 3$ MAD $\times$ 1.428 (1.428 is the scale factor for the MAD to equal the σ~~a consistent estimate of the variation~~

assuming a normal distribution [*Rousseeuw and Croux*, 1993]). These limits are plotted in Fig. 1a and Fig. 1c and highlight two key points: first, using the standard deviation when there are extreme outliers can allow for the acceptance of data that ~~should clearly be rejected~~are most likely physically unrealistic. Second, using the MAD on multimodal or asymmetrically ~~or multi-modally~~ distributed data can lead to the rejection of physically realistic~~“good”~~ data. For instance, as shown in Fig. 1a, the lower cut-off using the MAD of 2.76 ppm clearly excludes the low H_2O concentrations that are observed in Antarctic (austral) spring. As can be seen in Fig. 1b and Fig. 1d, the H_2O data at both altitude levels are not normally distributed.

The data can be separated further into bins based on latitudinal regions and local times. For example, Fig. 2 shows H_2O and O_3 sunset data at 30.5 and 35.5 km, separated into different latitude regions (0-30°S, 30-60°S, and 60-90°S), with dashed lines representing best fits to Gaussian~~normal~~ distributions. These regions are representative of bins often used to partition atmospheric data. Figure 2 exemplifies that using a given bin definition that leads to data with a symmetric and uni-modal distribution~~symmetrically distributed data~~ at one altitude level doesn’t necessarily lead to a symmetric and uni-modal distribution of~~symmetrically distributed~~ data at all altitude levels, nor across all species. For instance, in Fig. 2a the 35.5 km O_3 distributions in all three latitude regions are fairly symmetric. However the 35.5 km H_2O distribution, ~~(~~ Fig. 2c~~)~~, in the mid-latitudes is highly skewed, and in the high latitudes the distribution is tri-modal.~~,~~ ~~and i~~In Figs. 2b and d we see bimodal, asymmetric distributions for both O_3 and H_2O in the 30-60°S and 60-90°S regions at 30.5 km. For high-latitude data in many species’ data sets, distributions can be bimodal due to observing inside and outside of the vortex, and therefore it is not possible to find sub-regions (based on season, latitude, or local time) that will always exhibit symmetric distributions.

Therefore, the ACE-FTS data screening process takes an approach that does not require the distribution of any subset of data to be symmetric or containting just one mode.~~.~~

Initially, all data are pre-screened. Any occultation that contains errors due to previously known issues (e.g. unrealistic N_2O concentrations due to a convergence failure for occultations with low water levels, ice buildup on the detectors during early mission occultations, bad spectra used in the calibration, level 0-1 processing errors, etc.) are removed prior to analysis. A full list of known issues is given on the ACE validation website, https://databace.scisat.ca/validation. Then, for each species, at each altitude level, any data point with an absolute value greater than 10,000 times the median of all absolute values is rejected. Absolute values are used, as ACE-FTS VMR retrievals are allowed to be negative, and therefore, in some cases the median of the actual values could be very close to zero.

The screening processes starts by analysing the data's ~~probability density function~~pdfs. The normalized ~~probability density function~~pdf of data subset *x*, $pdf(x)$, multiplied by the number of data points, *N*, gives you the expectation density function (edf)~~expected number of data points~~ at a given value of *x*,

$$edf\cancel{E}(x) = N \times pdf(x). \quad (2)$$

The total integral of ~~function $E(x)$ is~~ the edf is equal to *N*, and the integral between any two values of *x* gives the number of expected data points within that range. For determining unnatural outliers, we want to find the values of *x* where the integral between infinity (negative and positive) and x_{lim} (lower and upper values) is less than 1.~~expectation distribution.~~ Anywhere that the integral (from infinity) of the edf~~expectation distribution~~ is less than 1 is most likely a statistical outlier, as no data points are expected to be measured beyond the~~at those~~ values of x_{lim},~~x,~~ given the pdf. Therefore, the criterion for excluding data can be any value of *x* where

$\int_{-\infty}^{x} E(x')dx'$ or $\int_{x}^{\infty} E(x')dx'$ ~~$E(x)$~~ is less than or equal to 1. This is similar to Peirce's criterion [*Peirce*, 1852; *Ross*, 2003] and Chauvenet's criterion [*Chauvenet*, 1871], which both assume a normally distributed p~~robability density function~~df. The tolerance level can be varied to suit the desired acceptance level of possible outliers. For ACE-FTS data, a tolerance level, determined empirically, of 0.025~~10^{-4}~~ is used, which corresponds to a 97.5~~99.99~~% confidence of an excluded data point being an outlier—i.e., any value x~~lim~~ where $\int_{-\infty}^{x} E(x')dx'$ ~~$E(x) < 10^{-4}$ i~~ or $\int_{x}^{\infty} E(x')dx'$ is less than 0.025 is rejected.

This method, however, requires determining an analytical solution for the data's edf~~expectation distributions~~. For each of the 50+ ACE-FTS retrieved species, at each altitude level, the data is separated into sunset and sunrise occultations, in order to separate into similar local conditions, as well as into four different latitude regions: 60-90°S, 0-60°S, 0-60°N, and 60-90°N. Due to the SCISAT orbital geometry, the majority of ACE-FTS measurements are at high latitudes, and therefore each latitudinal bin has roughly the same number of profiles. The distribution of each subset is then fit to a Gaussian mixed distribution, using three Gaussian distributions. This assumes that the data is at most tri-modally distributed. Since it is not uncommon for distributions of atmospheric measurements to be log-normal, the fit is done in log-space. The fit is performed using the Matlab statistical toolbox, which uses an Estimation Maximization algorithm [*McLachlan and Peel*, 2000]. In an effort to avoid fitting to extreme outliers an ad-hoc "Olympic"-style method is employed, whereby the data set's five lowest and five highest values are excluded in the fit. Figure 3 shows the O_3 distribution at 30.5 km in the 60-90°S and 60-90°N regions, along with the fitted e~~xpectation distribution~~dfs and the three Gaussian distributions derived in the fit for both cases. ~~It should be noted that prior to fitting a data subset to a Gaussian mixed distribution, profiles affected by previously determined issues~~

~~either with the instrument or with the data processing have been excluded, and extreme outliers are screened out by assuming that any data points with absolute values greater than 10,000 times the median of the subset's absolute values are outliers~~.

Figure 4 shows three examples of ACE-FTS sunset data distributions—NO_2 at 60-90°S and 30.5 km, CH_4 at 0-60°N and 17.5 km, and N_2O at 60-90°N and 20.5 km—and the corresponding fitted e~~xpectation distribution~~dfs. These were chosen in order to illustrate typical results for commonly used ACE-FTS data. The average root-mean-square error (RMSE) between the e~~xpectation distribution~~dfs and actual distributions, over all species and data subsets, is 6% and has a 1 σ deviation of 2%. ~~In the case of rare extreme events present in the data, which tend to be under-sampled in ACE-FTS data, the effect on the distribution can be to skew a tail end of the distribution, driving the shape of the tail away from Gaussian. An additional ad hoc method has been implemented to ensure that no "true" data is excluded in the screening process when rare extreme events occur that are not properly accounted for in the fit. For each subset, the standard deviation is calculated for the inlying data, where $E(x) > 10^{-4}$. This standard deviation we will call σ_{in}. Original upper and lower limiting values, x_l, are calculated, where $E(x_l) = 10^{-4}$, and both the upper and lower limiting values are extended by σ_{in}. Hence, $x_{lim}^{up} = x_l^{up} + \sigma_{in}$ and $x_{lim}^{low} = x_l^{low} + \sigma_{in}$.~~ Figure 5 shows the inliers and unnatural outliers as determined by the e~~xpectation distribution~~dfs for the subsets shown in Fig. 4. As can be seen, not all subsets contain many extreme outliers, e.g. NO_2 at 30.5 km (Fig. 5a), which only has one detected outlier. When there are obvious outliers, this method does exclude the most extreme outliers, although perhaps not all unnatural outliers. For instance, several (potentially) anomalously low values, near 0.75 ppm, in the CH_4 data (Fig. 5b) remain as inliers. This is in part due to the relatively lax tolerance level of ~~10^{-4}~~0.025, which is more likely to leave in outliers than if a larger value (but still less than 1) wer~~eas~~ chosen.

It should be noted that screening using the e~~xpectation distribution~~df is a hard-limiting filter.~~,~~ Therefore, using it in the manner described above~~ which~~ doesn't necessarily reject data that are non-physically anomalous for a given season. To screen the data ~~of~~ for this type of moderate outlier, the 15-day ~~running mean~~ $\cancel{(\mu_{15})}$median and a 15-day ~~running standard deviation~~ $\cancel{(\sigma_{15})}$variation scale are calculated for each subset, excluding outliers as determined from the e~~xpectation distribution~~dfs. Even on a 15-day timescale, ACE-FTS subset data can have distributions that are bimodal. In many cases, the primary mode is sampled much more frequently than the secondary mode, and therefore, without careful consideration, data within the secondary mode can be erroneously screened as unnatural outliers. To avoid this, we need a variation scale that is sensitive to outliers (unlike the MAD), but not overly sensitive to outliers (like the σ). For this we define a variation scale that is similar to the MAD, only more sensitive to outliers—the MeAD,

$$\mathrm{MeAD} = mean_i\left(\left|x_i - median_j(x_j)\right|\right). \qquad (3)$$

Any data point with a value outside the bounds of $median\cancel{\mu}_{15} \pm 10\cancel{5.5} \times \cancel{\sigma}\mathrm{MeAD}_{15}$ are considered to be unnatural outliers. The value of 10~~5.5~~ was empirically found to maximize the number of discovered unnatural outliers without rejecting obvious natural outliers~~ly "true" data~~. ~~If outliers are detected, they are removed from the data, and a new running mean and standard deviation are calculated for the inlying data in order to determine if there are any more outliers. This process is iterated until all data points are determined to be inliers. The mean and standard deviation are used instead of the median and MAD, as it is assumed that the subsets have already been screened for extreme outliers.~~ Figure 6 shows the inliers and outliers as determined by the 15-day running values for the subsets shown in Fig. 4. Clearly this step catches moderate outliers that were not detected using the e~~xpectation distribution~~dfs, although still not all anomalous data have been screened out. The potentially anomalous values near 0.75 ppm in the CH_4 data (Fig. 6b) still remain as inliers. Stricter tolerance criteria in either the e~~xpectation distribution~~df or running ~~standard deviation~~MeAD screening process would allow for these data to be screened out; however, they were found to lead to screening out ~~"true" data~~natural outliers in other subsets of data, which would be discordant with our philosophical approach. Going back to the original case of H_2O at 17.5 km (Fig. 1a), an example of the difference between using the MeAD as opposed to the MAD in the second step is shown in Fig.7. However, now the focus is on only the Antarctic data. The unnatural outliers and remaining inlying data for Antarctic H_2O at 17.5 km are shown for the two different approaches. Fig.7a shows the results when using limiting values of $median_{15} \pm 10 \times \mathrm{MeAD}_{15}$, where the significant majority of the data points screened as unnatural outliers are likely to be physically unrealistic for their local conditions. Using the limiting values of $median_{15} \pm 10 \times \mathrm{MAD}_{15}$, Fig. 7b, leads to many more outliers being detected as unnatural outliers. Upon inspection, many of these erroneous "unnatural" outliers are most likely being erroneously rejected, especially in late 2009 where ACE-FTS is most likely routinely observing dehydrated air masses. In both cases, outliers were detected in the sunrise and sunset data sets separately.

In order to explore the response to periodic extreme events and to trends, Fig. 8~~7~~ shows the final inliers and unnatural outliers in all ACE-FTS HCN data at 9.5 km, which exhibits periodic increases that could correspond to biomass burning events [e.g. *Crutzen and Andreae*, 1990; *Pommrich et al.*, 2010]; as well as all SF_6 data at 19.5 km, which exhibits a clear positive trend throughout the time series [e.g. *Rinsland et al*., 2005; *Brown et al*., 2011]. Even in these instances of extreme events and a significant trend in the data, the outlier detection method outlined here is able~~robust enough~~ to keep the natural outliers~~data~~ as inliers. The top panels (a and d) in Fig. 8~~7~~ show~~s~~ all data points and demonstrates the extreme unnatural outliers (red dots) that can occur within the ACE-FTS data set. The middle panels (b and e) show~~s~~ the same data as the top panels, however without the more extreme unnatural outliers in order to better view the data; and the bottom panels (c and f) show~~s~~ the data with all unnatural outliers removed.

In the overwhelming majority of instances where the ACE-FTS VMR data exhibit a sudden and/or extreme change in the distribution, the unnatural outlier detection method described above does not screen out these events. Sudden stratospheric warmings cause there to be strong descent in the northern high-latitude upper atmosphere. This leads to anomalously large concentrations of NO~~ and CO~~ in the upper stratosphere-lower mesosphere, near 50 km [e.g. *Manney et al*., 2008; *Randall et al*., 2009]. Figure 9a~~8~~ shows the time series of the final inliers ~~and outliers~~ in all ACE-FTS NO~~ and CO data~~ at 55.5 km~~ and 50.5 km, respectively~~. It can be seen that ~~Again,~~ the detection method is ~~robust enough~~able to keep the ~~majority of~~ data during these extreme events as inliers. *Anderson et al*. [2012], using in situ aircraft measurements, demonstrated that in the summer there can be H_2O intrusions from the upper troposphere into the lower stratosphere at Northern mid-latitudes. As can be seen in Fig. 9b, the final inlying ACE-FTS data in the summer Northern mid-latitudes, in the lower stratosphere, do exhibit large increases in H_2O concentrations. [*Manney et al*., 2011] showed that the Microwave Limb Sounder (MLS) on the Aura satellite observes decreases in lower stratospheric HCl concentrations in the Arctic vortex each spring; and in the spring of 2011, HCl concentrations were anomalously low for a prolonged period. Figure 9c shows the final inlying ACE-FTS HCl data in the Arctic, which are consistent with the MLS findings. No instances have been found in which the unnatural outlier detection system outright rejects these types of phenomena. When sudden, extreme changes do occur, the rejection of potential natural outliers has been minimized, and the result of which is that the rejection of the detected unnatural outliers has an insignificant effect on the mean.~~ In a couple extreme cases, NO data in early 2004 at 55.5 km (Fig. 8a-c), seven data points were flagged as extreme outliers that could potentially be "true" data; and similarly in early 2011 CO data at 50.5 km (Fig. 8d-f), eight data points may have been erroneously flagged. Out of all subsets of all 50+ species/isotopologues, these were the most extreme instances of apparent rejection of true data.~~ The disadvantage~~own side~~ of not screening out rare extreme events, however, is that this method is less likely to catch sporadic systematic instrument or processing errors. Therefore, continual monitoring of both the rejected and non-rejected data statistics is necessary to determine if any such errors have occurred.

Table 1 shows what percentage of ACE-FTS level 2 v3.5 profiles contain at least one detected outlier (by either step~~method~~). For any given species, if all profiles that contained at least one outlier are rejected, less than 6% of the total number of profiles will be omitted.

Conclusions

A two-step process has been developed in order to screen all ACE-FTS level 2 data for physically unrealistic outliers. The first step fits an e~~xpectation distribution~~df, the superposition of three Gaussian distributions, to actual distributions. This fit is done in log-space. Data in the tails of the distributions where the probability of finding a data point~~that have corresponding expectation values that are~~ is less than the ~~subset~~ tolerance level are determined to be extreme unnatural outliers. The second step iteratively takes the 15-day running median and ~~standard deviation~~MeAD and screens for moderate seasonal unnatural outliers. ~~At each iteration, d~~Data that are further than 10~~5.5~~ times the MeAD~~standard deviation~~ from the median are determined to be moderate outliers.

Using these methods to screen the ACE-FTS data for unnatural outliers, a flagging system has been implemented to give ACE-FTS level 2 data users a guide for how best to use the data. Each VMR data point in each profile is flagged with an integer from 0-9. Table 2 gives the definition for each flag value. Any data with a 0 flag are~~is~~ recommended for use. In previous versions, data users were recommended that they filter out data where the percent error (the retrieval statistical fitting error divided by the retrieved value) is either greater than 100% or less than 0.01%; for legacy reasons, these data have been given a flag value of 1. It is recommended that data points with a corresponding flag greater than 2 be removed before any analysis is performed. This screening method alone may be adequate when only looking at one altitude level, however, profiles that contain an outlier at a given altitude level may also be compromised at lower altitude levels. Therefore it is recommended that any profile that contains a flag between 4 and 7 (inclusive) be removed before analysis. ~~However, screening the data using these flags should be done with caution when investigating middle to upper atmospheric NO, CO, and CO isotopologues.~~

At certain altitude levels for a given species, the data can be either noisy, with a significant number of negative values, or have a strong negative bias. In either case, since the ACE-FTS retrieval allows for negative concentrations, it is possible for valid data to have values close to zero, both positive and negative. When values are systematically near zero, the percent error becomes extremely large. Therefore, in these situations, screening the data based on the percent error may introduce a bias in the data. As such, before analysis, removing data that has a corresponding flag value of 1 is only recommended at altitude levels where the overwhelming majority of data points have a VMR value~~are~~ greater than zero.

Since the outlier detection methodology was approached with a philosophy that it is better to leave in unnatural outliers than to remove natural out~~in~~liers, there are outliers that have gone un-flagged—especially in data sets that are inherently noisy and at low altitudes (below ~10 km). Level 2 data users should use the defined quality flags as a starting point for screening the data and be aware that some unnatural outliers may still exist that could be screened out prior to analysis. I~~t is recommended that~~ f data users are~~lso avoid~~ using the MAD in an~~y~~ attempt~~s~~ to further screen the ACE-FTS level 2 data, for best results it is advised that they ensure that the distribution of the data they are screening is not multimodal nor heavily skewed.

The flag values for all ~~v2.5, v3.0, and~~ v3.5 are now available for download on the ACE-FTS website, and v2.5 flag values ~~data~~ are available upon request from the lead author and will soon be made available for download on the ACE-FTS website. It is currently expected that similar flags will be a standard product within the level 2 data of all future products.

Acknowledgements

The authors wish to thank the anonymous referees for their comments and valuable insight. This work was supported by the Canadian Space Agency (CSA). The Atmospheric Chemistry Experiment is a Canadian-led mission mainly supported by the CSA.

References

Anderson, J. G., D. M. Wilmouth, J. B. Smith, and D. S. Sayres (2012), UV Dosage Levels in Summer: Increased Risk of Ozone Loss from Convectively Injected Water Vapor, *Science*, **337**, 835-839, doi:10.1126/science.1222978.

Bernath, P. F., C. T. McElroy, M. C. Abrams, C. D. Boone, M. Butler, C. Camy-Peyret, M. Carleer, C. Clerbaux, P.-F. Coheur, R. Colin, P. DeCola, M. DeMazière, J. R. Drummond, D. Dufour, W. F. J. Evans, H. Fast, D. Fussen, K. Gilbert, D. E. Jennings, E. J. Llewellyn, R. P. Lowe, E. Mahieu, J. C. McConnell, M. McHugh, S. D. McLeod, R. Michaud, C. Midwinter, R. Nassar, F. Nichitiu, C. Nowlan, C. P. Rinsland, Y. J. Rochon, N. Rowlands, K. Semeniuk, P. Simon, R. Skelton, J. J. Sloan, M.-A. Soucy, K. Strong, P. Tremblay, D. Turnbull, K. A. Walker, I. Walkty, D. A. Wardle, V. Wehrle, R. Zander, and J. Zou (2005), Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett.*, **32**, L15S01, doi:10.1029/2005GL022386.

Boone, C. D., R. Nassar, K. A. Walker, Y. Rochon, S. D. McLeod, C. P. Rinsland, and P. F. Bernath (2005), Retrievals for the Atmospheric Chemistry Experiment Fourier-Transform Spectrometer, *Appl. Optics*, **44**, 7218-7231, doi:10.1364/AO.44.007218.

Boone, C. D., K. A. Walker, and P. F. Bernath (2013), Version 3 Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, *A. Deepak Publishing*, Hampton, Virginia, USA, pp. 103-127.

Brown, A. T., M. P. Chipperfield, C. D. Boone, C. Wilson, K. A. Walker, and P. F. Bernath (2011), Trends in atmospheric halogen containing gases since 2004, *J. Quant. Spectrosc. Rad. Trans*. **112**, 2552-2566, doi:10.1016/j.jqsrt.2011.07.005.

Chauvenet, W. (1871), Manual of spherical and practical astronomy, *J. B. Lippincott Company*, Philadelphia, USA, pp. 558-563.

Crutzen, P. J., and M. O. Andreae (1990), Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles, *Science*, **250**, 1669–1678, doi:10.1126/science.250.4988.1669.

Lary, D. J. and L. Lait (2006), Using probability distribution functions for satellite validation, *IEEE Trans. Geosci. Remote Sens*., **44**, 1359–1366, doi:10.1109/TGRS.2005.860662.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata (2013), Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, *J. Exp. Soc. Psychol*., **49**, 764–766, doi:10.1016/j.jesp.2013.03.013.

Manney, G. L., W. H. Daffer, K. B. Strawbridge, K. A. Walker, C. D. Boone, P. F. Bernath, T. Kerzenmacher, M. J. Schwartz, K. Strong, R. J. Sica, K. Krüger, H. C. Pumphrey, A. Lambert, M. L. Santee, N. J Livesey, E. E. Remsberg, M. G. Mlynczak, and J. R. Russell III (2008), The high Arctic in extreme winters: vortex, temperature, and MLS and ACE-FTS trace gas evolution, *Atmos. Chem. Phys*., **8**, 505-522, doi:10.5194/acp-8-505-2008.

Manney, G. L., M. L. Santee, M. Rex, N. J. Livesey, M. C. Pitts, P. Veefkind, E. R. Nash, I. Wohltmann, R. Lehmann, L. Froidevaux, L. R. Poole, M. R. Schoeberl, D. P. Haffner, J. Davies, V. Dorokhov, H. Gernandt, B. Johnson, R. Kivi, E. Kyrö, N. Larsen, P. F. Levelt, A. Makshtas, C. T. McElroy, H. Nakajima, M. C. Parrondo, D. W. Tarasick, P. von der Gathen, K. A. Walker, and N. S. Zinoviev (2011), Unprecedented Arctic ozone loss in 2011, *Nature*, **478**, 469-U65, doi:10.1038/nature10556.

McLachlan, G., and D. Peel (2000), Finite Mixture Models, *John Wiley & Sons, Inc*., Hoboken, New Jersey, USA.

Migliorini, S., C. Piccolo, and C. D. Rodgers, C. D (2004), Intercomparison of direct and indirect measurements: Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) versus sonde ozone profiles, J. Geophys. Res., 109, D19316, doi:10.1029/2004JD004988.

Peirce, B. (1852), Criterion for the rejection of doubtful observations, *Astronomical J*., **2**, 161-163.

Pommrich, R., R. Müller, J.-U. Grooß, G. Günther, P. Konopka, M. Riese, A. Heil, M. Schultz, H.-C. Pumphrey, and K.A. Walker (2010), What causes the irregular cycle of the atmospheric tape recorder signal in HCN?, *Geophys. Res. Lett*. **37**, L16805, doi:10.1029/2010GL044056.

Randall, C. E., V. L. Harvey, D. E. Siskind, J. France, P. F. Bernath, C. D. Boone, and K. A. Walker (2009), NO_x descent in the Arctic middle atmosphere in early 2009, *Geophys. Res. Lett*., **36**, L18811, doi:10.1029/2009GL039706.

Rinsland, C. P., C. Boone, R. Nassar, K. Walker, P. Bernath, E. Mahieu, R. Zander, J. C. McConnell, and L. Chiou (2005), Trends of HF, HCl, CCl_2F_2, CCl_3F, $CHClF_2$ (HCFC-22), and SF_6 in the lower stratosphere from Atmospheric Chemistry Experiment (ACE) and Atmospheric Trace Molecule Spectroscopy (ATMOS) measurements near 30°N latitude, *Geophys. Res. Lett*., **32**, L16S03, doi:10.1029/2005GL022415.

Ross, S. M. (2003), Peirce's criterion for the elimination of suspect experimental data, *J. Eng. Tech*., **20**, 38-41.

Rousseeuw, P. J., and C. Croux (1993), Alternatives to the Median Absolute Deviation, *J. Amer. Stat. Assoc*., **88**, 1273-1283.

Toohey, M., K. Strong, P. F. Bernath, C. D. Boone, K. A. Walker, A. I. Jonsson, and T. G. Shepherd (2010), Validating the reported random errors of ACE-FTS measurements. *J.*

Geophys. Res., **115**, D20304, doi:10.1029/2010JD014185.

Wu, D. L., J. H. Jiang, W. G. Read, R. T. Austin, C. P. Davis, A. Lambert, G. L. Stephens, D. G. Vane, and J. W. Waters (2008), Validation of the Aura MLS cloud ice water content measurements, *J. Geophys. Res*., **113**, D15S10, doi:10.1029/2007JD008931.

Tables

Table 1 – Percent rejection of ACE-FTS level 2 v3.5 profiles that contain one or more detected unnatural outlier (either by running MeAD~~mean~~ or e~~xpectation distribution~~df).

Species	% reject	Species	% reject	Species	% reject
C_2H_2	1.54~~71~~	HCFC141b	1.82~~2.03~~	$C^{17}O$	1.~~95~~67
C_2H_6	1.89~~1~~	HCFC142b	1.66~~81~~	$C^{18}O$	2.~~54~~21
CCl_2F_2	2.10~~09~~	HCl	2.18~~9~~	$O^{13}CO$	~~4.46~~3.98
CCl_3F	1.53~~66~~	HCN	2.26~~50~~	$O^{13}C^{18}O$	1.~~30~~09
CCl_4	1.67~~81~~	HCOOH	1.95~~2.09~~	$OC^{17}O$	1.~~33~~24
CF_4	1.83~~2.35~~	HF	1.46~~53~~	$OC^{18}O$	4.~~90~~59
CFC113	1.33~~50~~	HNO_3	2.25~~65~~	$H^{17}OH$	2.~~95~~74
CH_3Cl	2.11~~39~~	HNO_4	2.3~~4~~9	$H^{18}OH$	~~3.08~~2.91
CH_3OH	2.40~~83~~	N_2	2.08~~62~~	HDO	2.~~76~~70
CH_4	2.81~~95~~	N_2O	3.96~~4.29~~	^{15}NNO	2.~~39~~28
CHF_2Cl	2.29~~35~~	N_2O_5	2.15~~6~~	$N^{15}NO$	2.~~37~~34
$ClONO_2$	1.7~~6~~8	NO	4.50~~91~~	$NN^{17}O$	1.~~88~~77
CO	4.03~~10~~	NO_2	2.22~~34~~	$NN^{18}O$	2.~~91~~94
CO_2	5.39~~70~~	O_2	1.81~~2.23~~	$O^{17}OO$	~~3.03~~2.98
$COCl_2$	2.08~~37~~	O_3	2.24~~40~~	$O^{18}OO$	1.~~97~~95
COClF	1.29~~35~~	OCS	1.45~~2~~	$OO^{18}O$	1.~~85~~73
COF_2	1.31~~26~~	SF_6	2.25~~2.31~~	$O^{13}CS$	2.~~17~~20
H_2CO	3.0~~4~~3	$^{13}CH_4$	~~3.00~~2.98	$OC^{34}S$	~~1.92~~2.04
H_2O	3.81~~97~~	CH_3D	~~2.01~~2.17		
H_2O_2	2.74~~3.16~~	^{13}CO	~~3.08~~2.94		

Table 2 – Definition of flag values associated with ACE-FTS level 2 data

Flag value	Definition
0	No known issues with data
1	Percent error is not within 0.01-100%, and no other category of flag applies
2	Not enough data points in the region to do statistical analysis, and percent error is within 0.01-100%
3	Not enough data points in the region to do statistical analysis, and percent error is not within 0.01-100%
4	Moderate unnatural outlier detected from running MeAD~~mean~~, percent error within limits
5	Extreme unnatural outlier detected from e~~xpectation distribution~~df, percent error within limits
6	Unnatural ~~O~~outlier detected and percent error is outside of limits
7	Instrument or processing error
8	Error fill value of -888 (data is scaled a priori)
9	Data fill value of -999 (no data)

Figures

Fig. 1 – ACE-FTS level 2 v3.5 H_2O data (left) and corresponding distributions (right). Top panel shows all data at 17.5 km, and the bottom panel shows all data at 35.5 km.

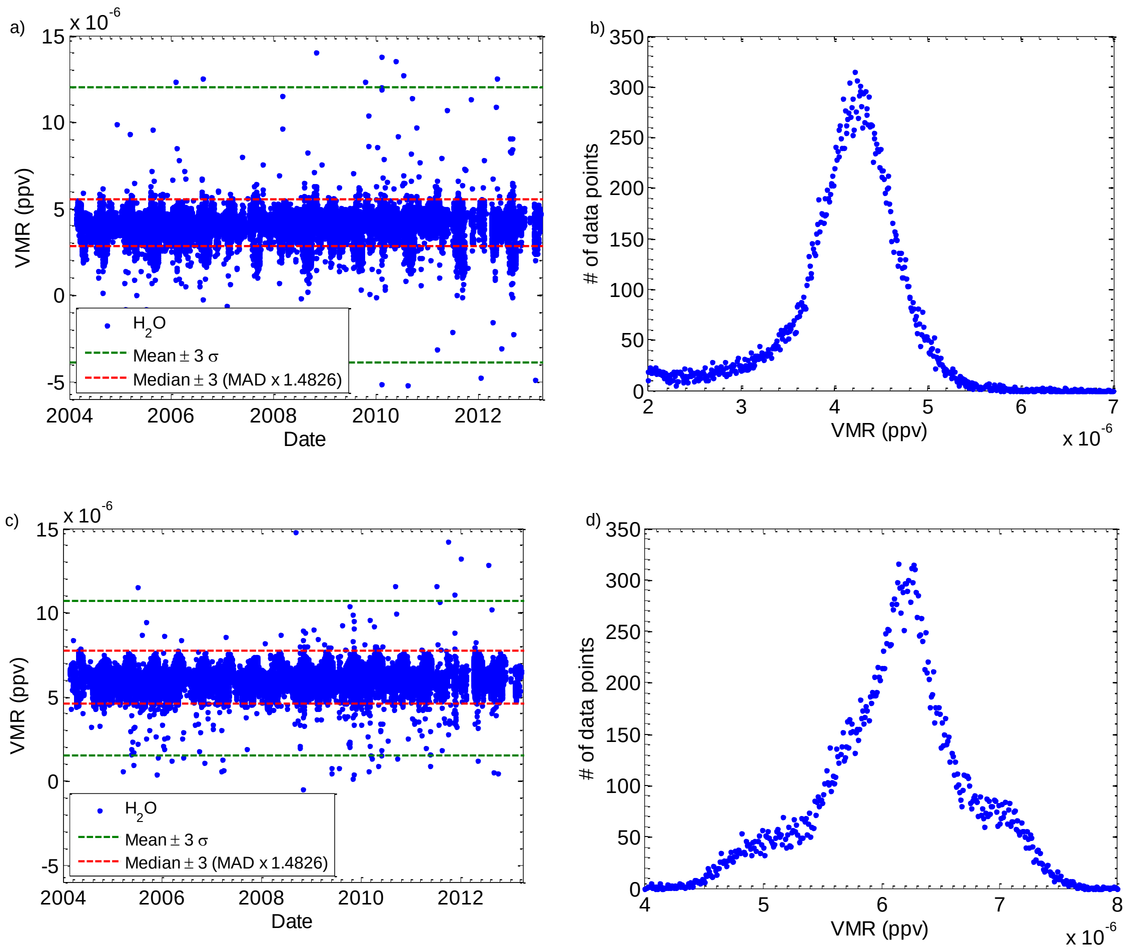


Fig. 2 – 2004-2013 ACE-FTS VMR distributions for sunset occultations (symbols~~dots~~) in the Southern hemisphere and corresponding best fits to normal distribution (dashed lines). (a) O_3 at 35.5 km, (b) O_3 at 30.5 km, (c) H_2O at 35.5 km, (d) H_2O at 30.5 km.

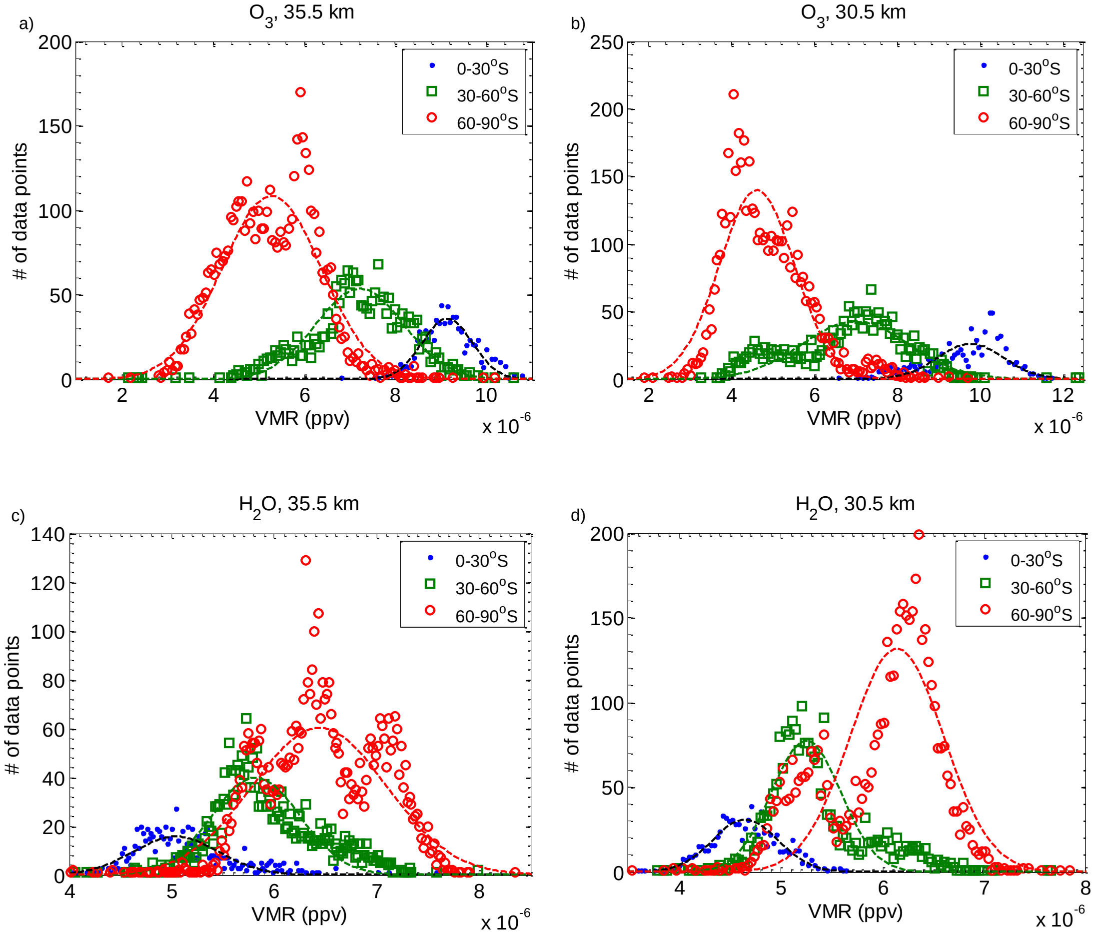


Fig. 3 – Sunrise ACE-FTS O_3 VMR distribution~~s~~ at 30.5 km (blue circles) and fitted e~~xpectation distribution~~df~~s~~ (dashed black lines) for (a) 60-90°N, and (b) 60-90°S. Dotted green lines are the fitted Gaussian distributions in calculating each of the e~~xpectation distribution~~dfs, and the fitted distributions have been normalized to the measured VMR distributions.

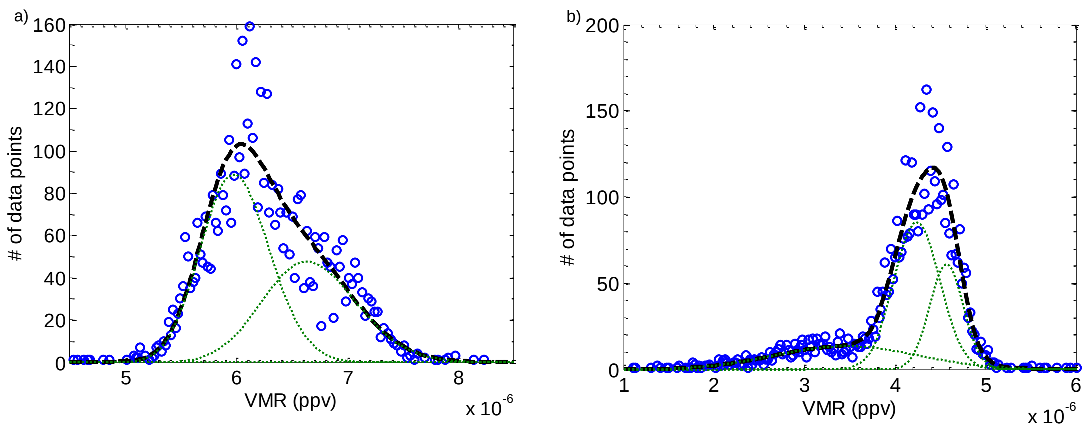


Fig. 4 – Sunrise ACE-FTS VMR distribution~~s~~ (blue circles) and fitted e~~xpectation distribution~~dfs (black dashed lines) for (a) NO_2 at 30.5 km in the latitude region 60-90°S; (b) CH_4 at 20.5 km, 0-60°N; and (c) N_2O at 20.5 km, 60-90°N. Dotted green lines are the fitted Gaussian distributions in calculating each of the edfs, and the fitted distributions have been normalized to the measured VMR distributions.

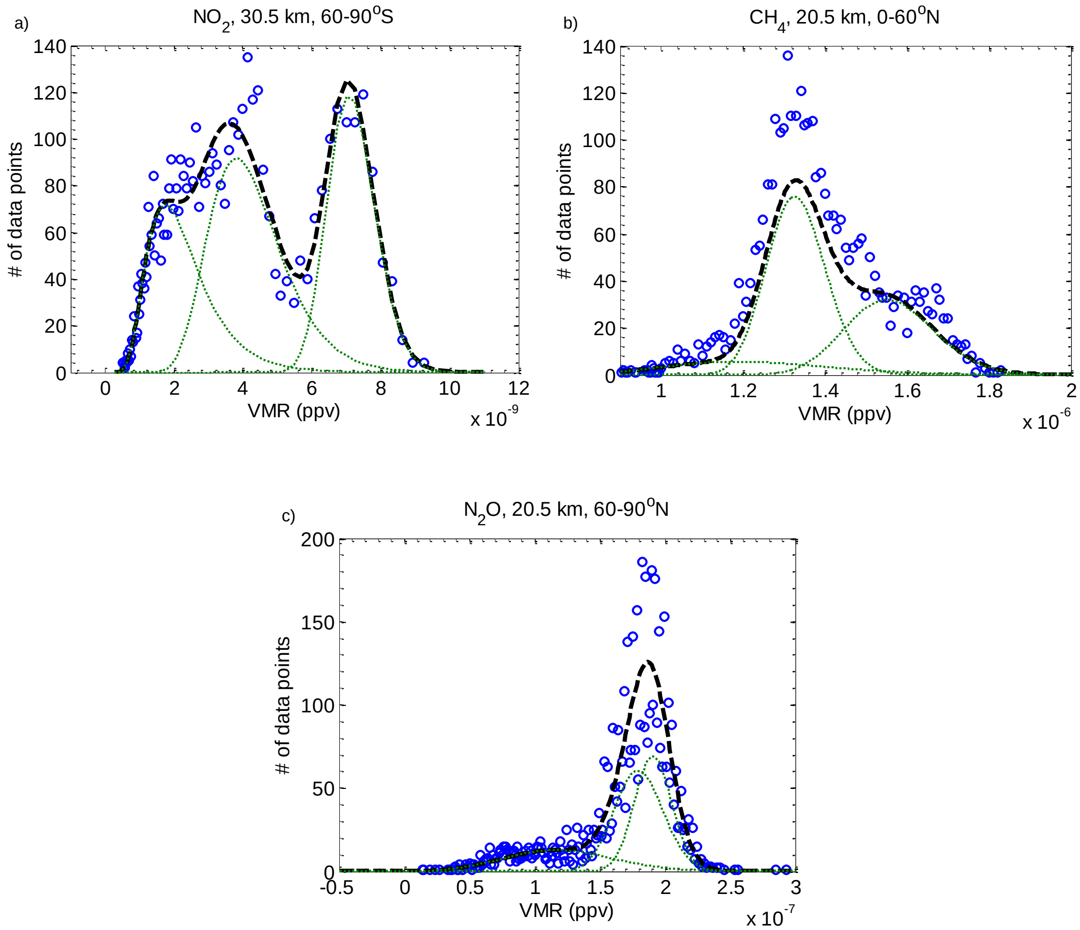


Fig. 5 – Sunrise ACE-FTS data for the same data subsets as Fig. 4. Red circles are data that have been determined to be unnatural outliers~~outlying data~~ as per the e~~xpectation distribution~~dfs, and blue dots are the inlying data.

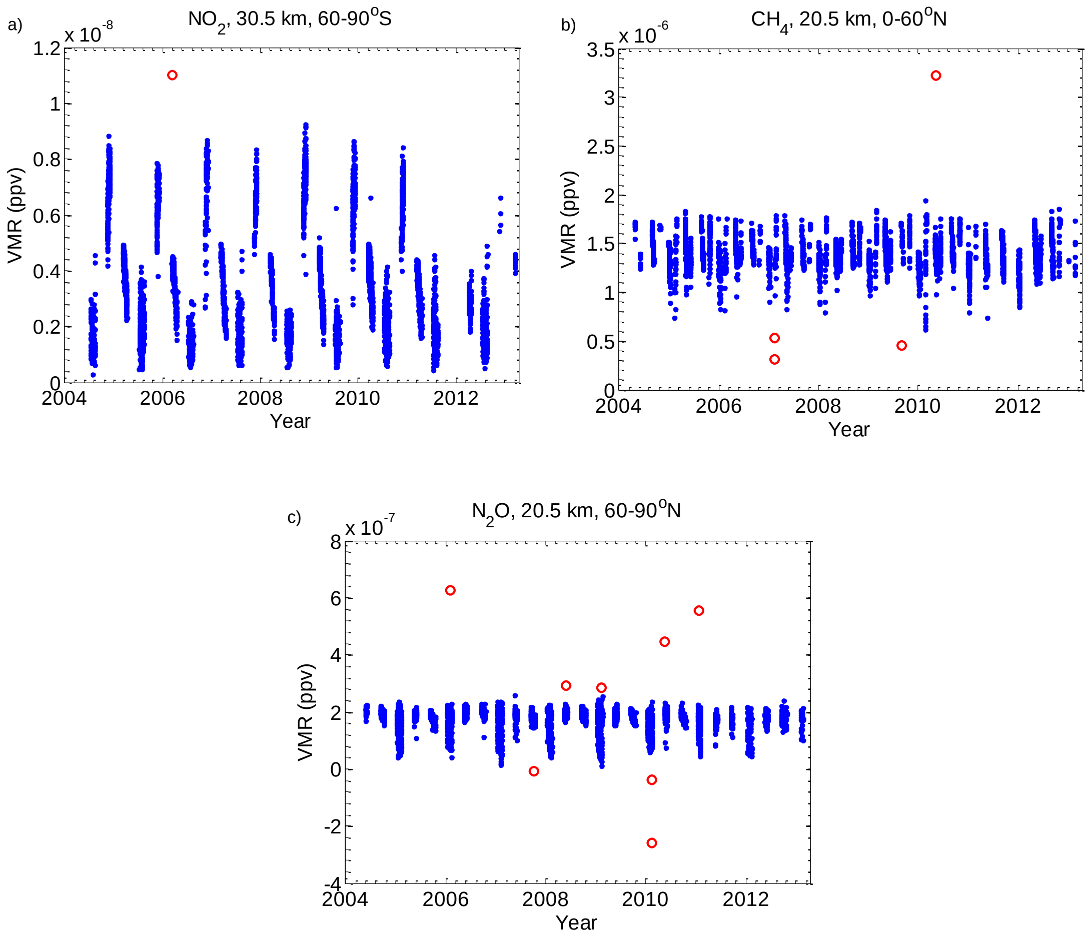


Fig. 6 – Sunrise ACE-FTS data for the same data subsets as Fig. 4. Red circles are data that have been determined to be unnatural outliers ~~outlying data~~ as per the 15-day running median and MeAD~~standard deviation~~, and blue dots are data that have been determined to be inliers.

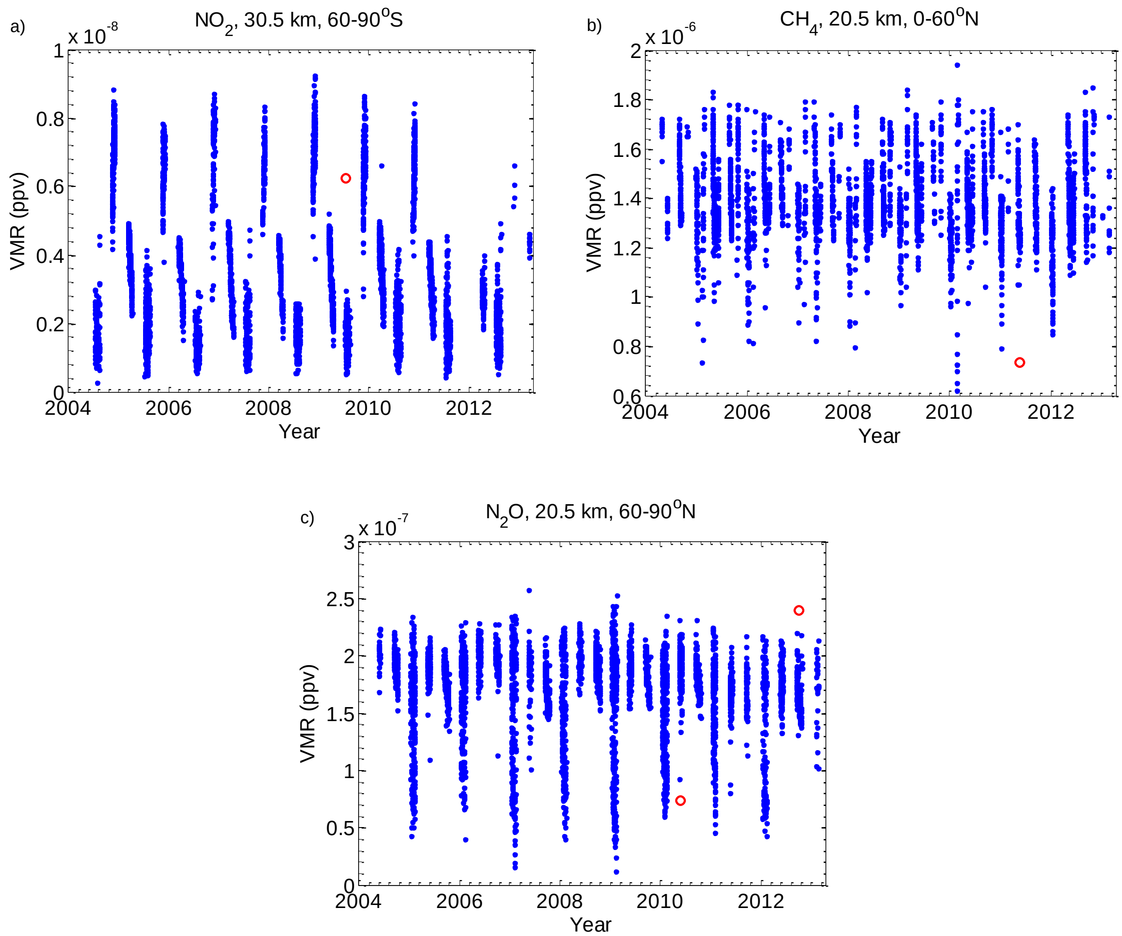


Fig. 7 – All Antarctic ACE-FTS data for H_2O at 17.5 km. Red circles are data that have been determined to be unnatural outliers following two different methods (sunrise and sunset data were analyzed separately), and blue dots are data that have been determined to be inliers. (a) Unnatural outliers determined using the 15-day running MeAD, (b) Unnatural outliers determined using the 15-day running MAD. Unnatural outliers as determined by the edfs are not shown and were not used in the analysis.

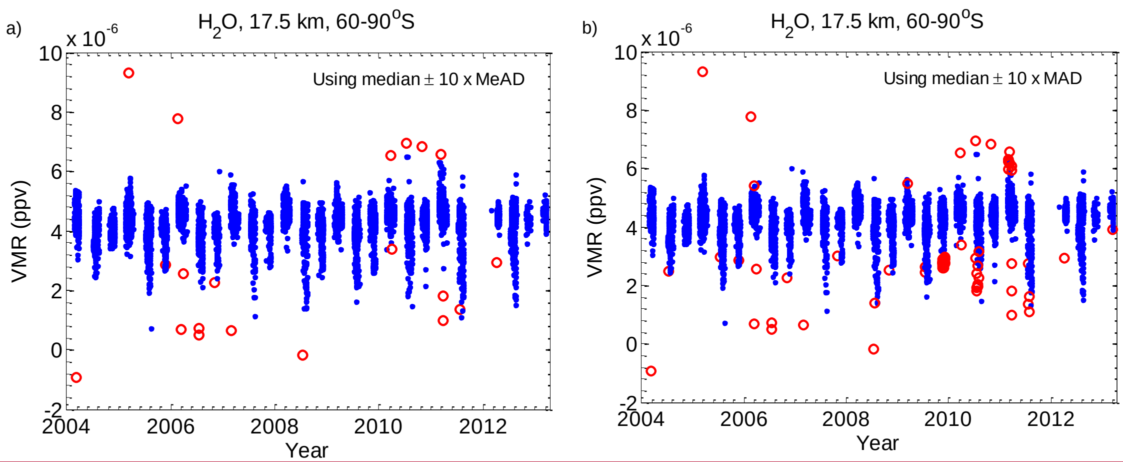


Fig. 8~~7~~ – The final inlying data (blue dots) and unnatural outliers~~outlying~~ (red dots)~~ data~~ for all ACE-FTS HCN data at 9.5 km (left) and SF_6 data at 19.5 km (right). The top panel shows all data, the middle panel is the same as the top panel only zoomed in for clarity, and the bottom panel is all data excluding the unnatural outliers.

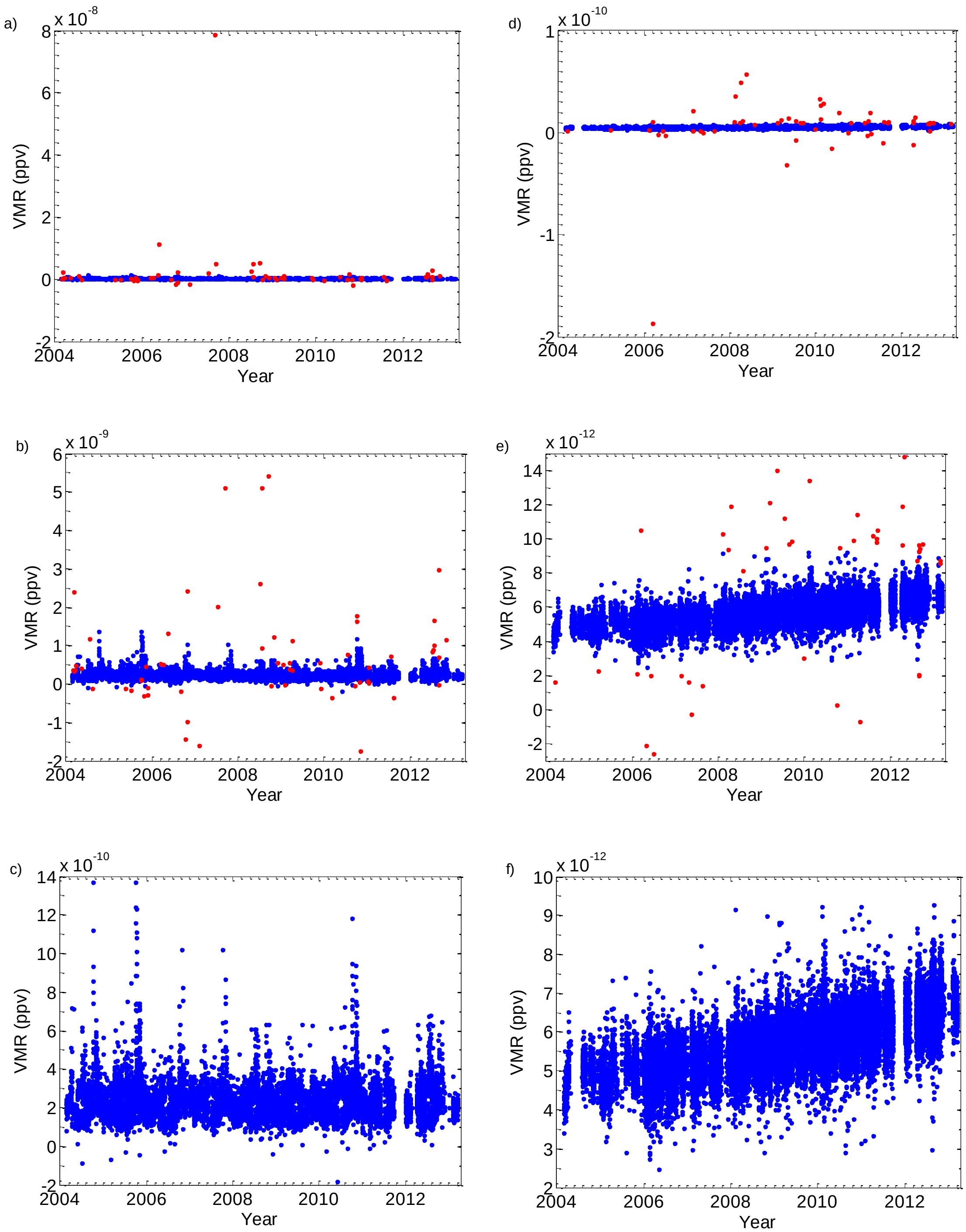
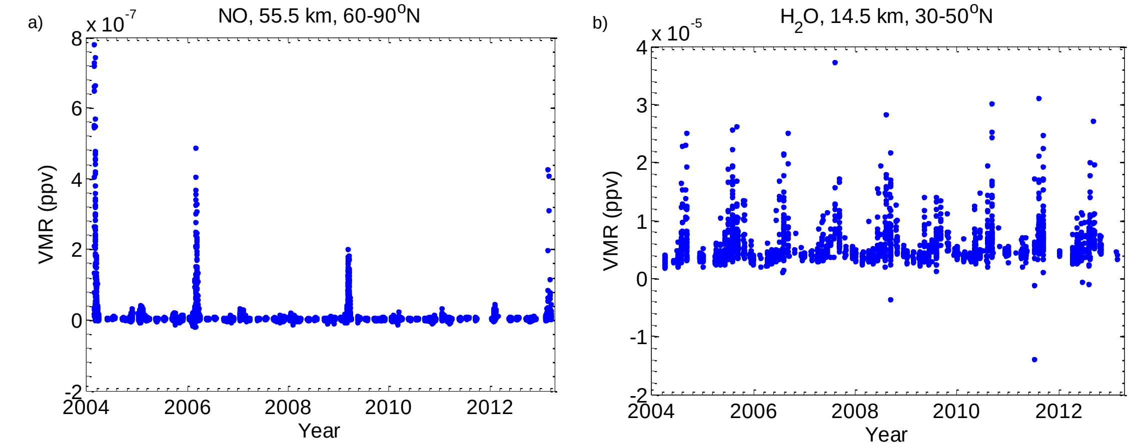
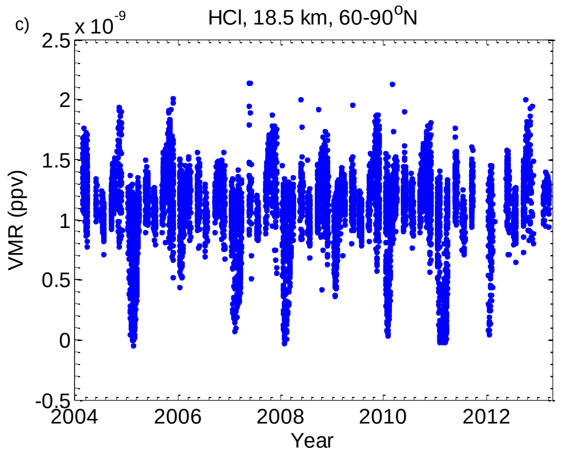
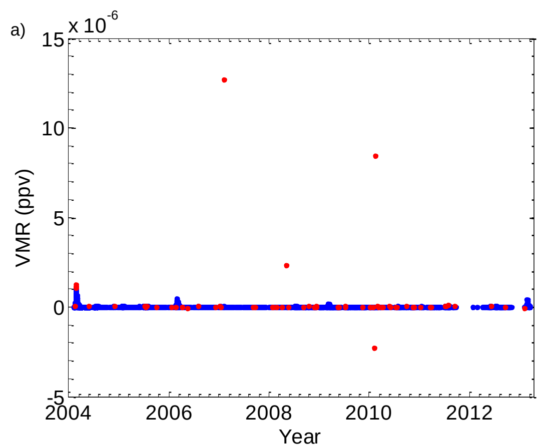
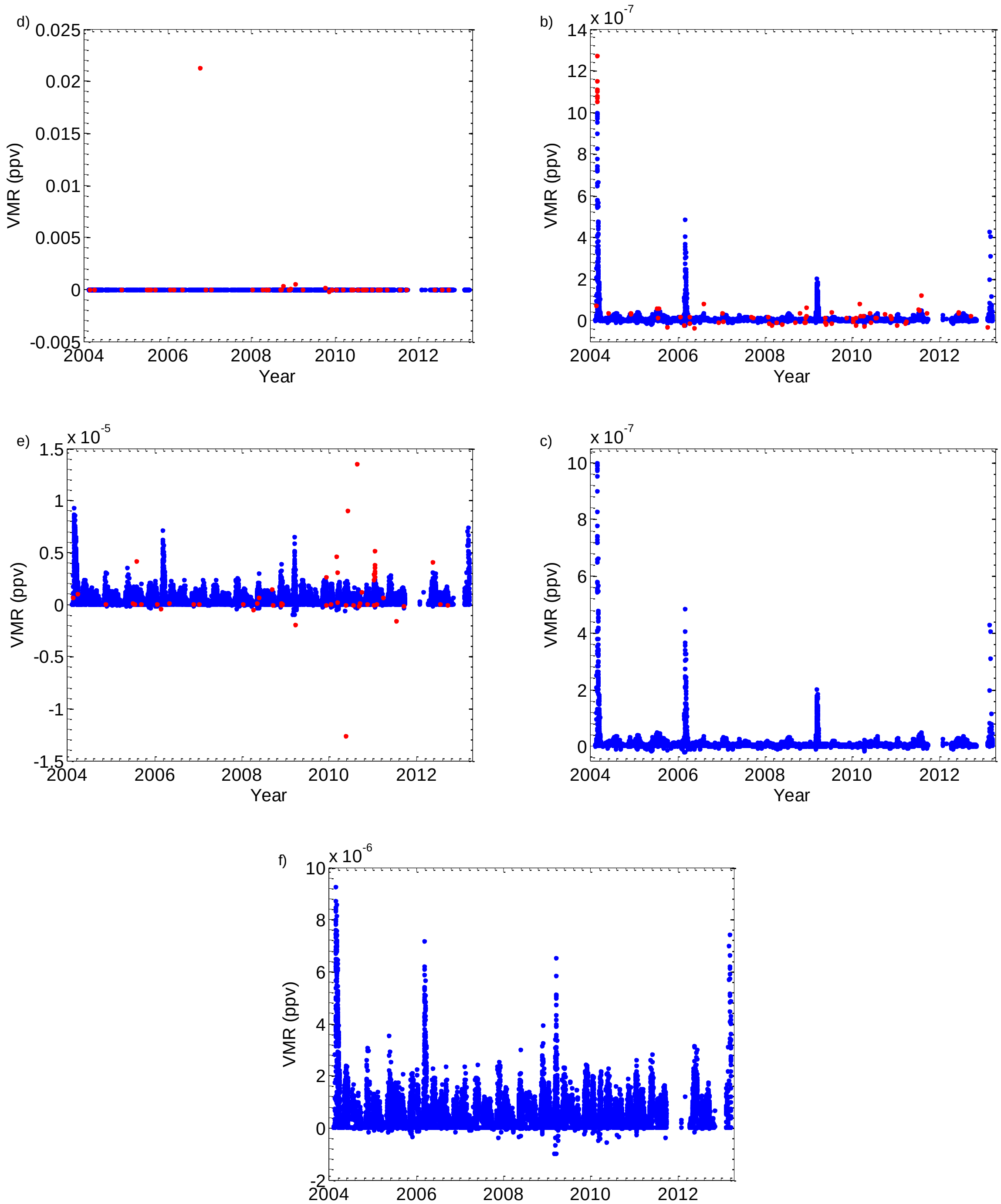


Fig. 9~~8~~ – The final inlying data for ~~(blue dots) and outlying (red dots) data for all~~ ACE-FTS (a) Arctic NO~~ data~~ at 55.5 km, (b~~left~~) mid-latitude ~~and CO~~H_2O~~ data~~ at 14~~50~~.5 km, and (c~~right~~) Arctic HCl at 18.5 km.~~ The top panel shows all data, the middle panel is the same as the top panel only zoomed in for clarity, and the bottom panel is all data excluding the outliers.~~








d)
VMR (ppv)
Year
b)
x 10^-7
VMR (ppv)
Year
e)
x 10^-5
VMR (ppv)
Year
c)
x 10^-7
VMR (ppv)
Year
f)
x 10^-6
VMR (ppv)
Year