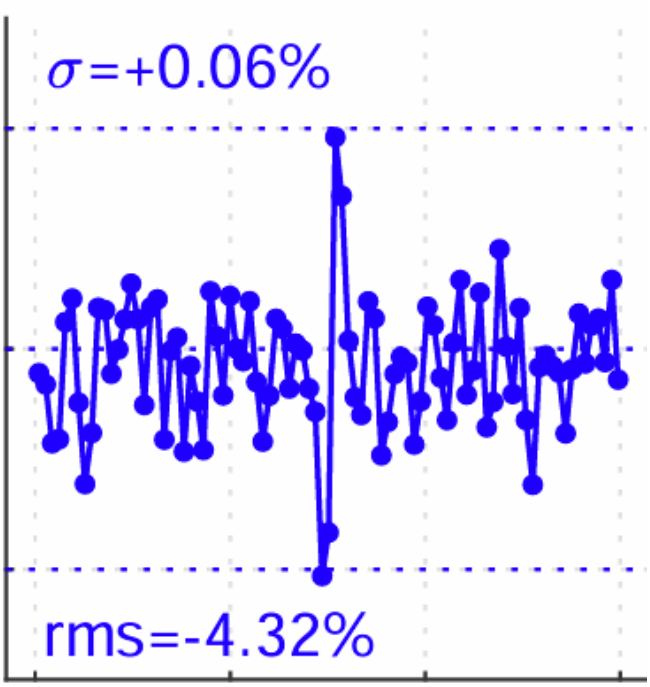
This article evaluates non-ideal performance characteristics for a high-resolution FTS instrument located in Addis Ababa. The paper is well written and is generally thorough, but I did have one major issue with it, and that is the presence of systematic residuals for HBr measurements visible in Figures 13 and 14 of the paper that might indicate a potential systematic error in the ILS determination.

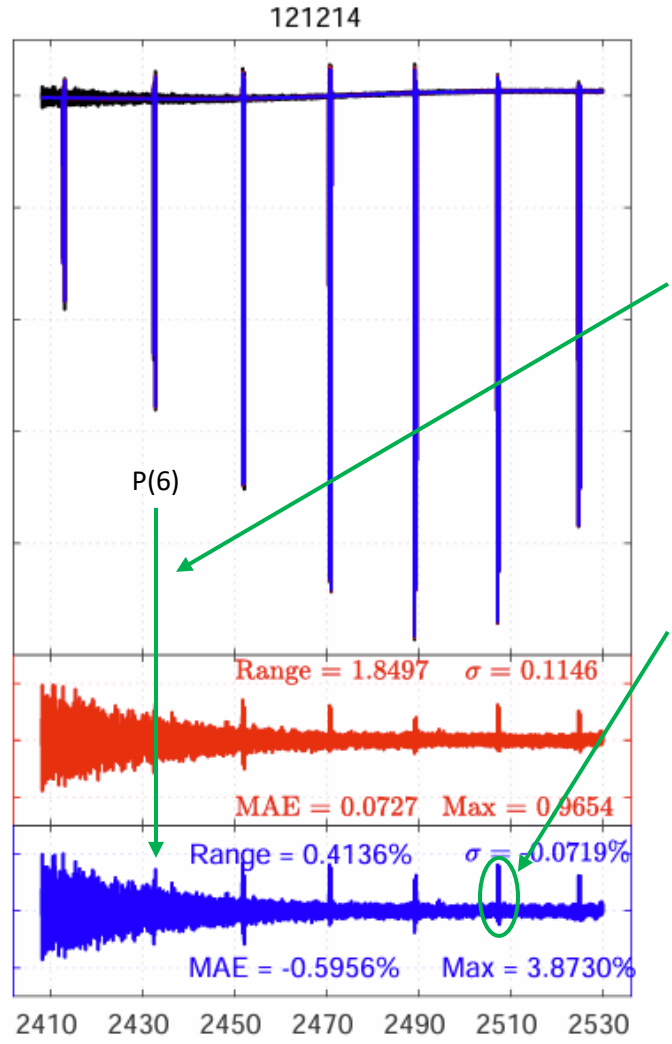


Above: an excerpt from Figure 14 of the article, showing the HBr P(6) residuals for the measurement on January 17th, 2011 using the modified ILS.

There is a systematic feature in the residuals that is commonly referred to as an "s-shaped" pattern. This pattern occurs in the residuals when the calculated and measured spectra are shifted relative to each other. To me, that suggests that the wavenumber calibration might be slightly off. The amplitude of the s-shaped residuals also changes over the years, suggesting that the misalignment drifts with time.

My question is, how are potential misalignments between measured and calculated spectra handled in the analysis? Are misalignments corrected before (or while) performing the least squares fitting with LINEFIT? Is a wavenumber shift applied to the calculated or the measured spectrum to bring them into alignment? Or is there a stretching applied to the wavenumber scale of the measurement, which would be physically more rigorous than using a shift? There is no mention in the article how such things are handled, and the presence of s-shaped residuals seems like a red flag to me. There is a danger that "fudge factors" in the analysis (fitting for HBr partial pressure, total pressure in the cell, and effective temperature in the cell) could empirically reduce the residuals resulting from a misalignment without necessarily improving the quality of the analysis result (e.g., the accuracy of the ILS determined with LINEFIT).

I also note that the selected HBr line, P(6), is in a region where the signal-to-noise ratio is significantly lower than other HBr lines being measured. Part of the purpose of this choice of line for Figure 14 seems to be to avoid other lines where the noise is lower and systematic features in the residuals are not so effectively masked.



An excerpt from Figure 13. The selected line, P(6), features higher noise levels, which obscures somewhat the systematic nature of the residuals from HBr, but that is like sweeping the problem "under the rug."

How about using a line with better signal-to-noise ratio and see if the systematic feature can be reduced in the analysis?

I would suggest using a line with less noise, such as the line with the systematic feature in the residuals circled in the above figure. Systematic features in the residuals illuminate potential problems in the ILS determination, if you cannot explain their origin.

Does the analysis include determining a stretching factor for the wavenumber scale of the measurement? If not, I would suggest it should be implemented. For example, you could Fourier interpolate the spectrum onto a fine wavenumber grid (to reduce sampling issues), find the locations of the HBr peaks in the interpolated spectrum, and use the HBr line position values from HITRAN to calibrate the wavenumber spacing in your measured spectrum.

Things will get a bit more complicated if there are non-Voigt (speed dependence or line mixing) contributions to the HBr line shape, in that the systematic residuals may not go away entirely unless you have the appropriate non-Voigt parameters for the lines. Although if that were the case, it should be a known issue, since HBr is routinely used as a calibration standard for NDACC instruments.

To reiterate, the presence of systematic residuals for the HBr makes me nervous that some factor (such as incomplete accounting in the analysis for misalignments between measured and calculated spectra) is polluting the ILS determination. I would suggest you not shy away from using the line with lower noise, where the systematic feature is more prominent (relative to the noise) than what we see for P(6). Is there some adjustment you can make in the analysis to reduce the systematic residuals? If you cannot obtain something approaching flat residuals in the analysis, it makes it difficult to fully trust the results.

Sometimes, the reader is expected to infer some definitions from context, which could make full understanding a challenge. In Figure 3, "theoretically ideal" is not explicitly defined but I suppose is readily deduced. "Nominally ideal" is also not explicitly defined. I assume it means the only difference from the theoretically ideal instrument is self apodization from off-axis rays in the instrument (i.e., the finite field of view effect), but that is an assumption. There are two configurations. The "nominal configuration" is mentioned in Section 3.8 as representing "ideal conditions," but the text suggests to me that it uses an ILS derived from LINEFIT [line 393: *In the LINEFIT ILS retrieval, the nominal configuration assumed zero offsets, no wavenumber shifts, and excluded spectral channeling*], which implicitly includes the impact of non-ideal effects like instrument misalignment. So, does the nominal configuration use the ILS from LINEFIT or not? Section 3.8 discusses a "misaligned setup," and in the next section starts talking about the "modified configuration," which we need to infer is the same thing with a different label. On line 456, the text discusses metrics for "measured conditions," which are indicated to be the same as nominal conditions, yet the remainder of the text employs the phrasing of nominal rather than measured, so it wasn't clear why different labeling was used there.

Some of the introductory material seems superfluous in that it does not directly relate to quantities being measured in the study. For example, formalizing separating the contributions from different sources on modulation efficiency and phase error in Section 2.6 plays no role in the analysis. Similarly, Section 2.5 derives an equation for the tolerance to lateral shifts, but that equation is never used in the analysis. The material would be appropriate in other forums (like a book delving into non-ideal aspects of an FTS), but it seemed like some of the theory could be trimmed without losing any necessary information for understanding the article.

> Line 460: The figures shown in the Fig. 8- 11 illustrate the progressive degradation of the FTIR instrument's performance over time

Some care should be taken here, in that the maximum optical path difference (MOPD) was increased in 2011, which essentially made it a completely different instrument. This can be seen in the narrowing of the FHWM of the ILS between 2011 and 2012 (Table 4), an expected consequence of

increasing MOPD. It also changes the sampling, as can be seen in the sampling of the P(6) line shown in Figure 14. In 2011, the peak of the line is roughly halfway between two sampled points, while in 2012, the instrument samples near the peak of the P(6) line. The interferogram ZPD peak intensity experiences a large boost in 2012, presumably because it received fresh (or thoroughly cleaned) optics and had the best possible alignment. There is a definite drop off in ZPD peak intensity between 2012 and 2013. However, I don't think it is fair to compare 2013 to 2011 and conclude that the performance has "degraded" between those two years, because at that point you are comparing two distinct instrument configurations. In 2013, it has settled into its dirty-optic, not-perfectly-aligned state that lost the signal boost it featured in 2012, but I see no reason to expect its characteristics should be the same as 2011 if it has different optics and possibly a different input aperture. It is common practise to match in the input aperture to the resolution, such that self-apodization losses in modulation efficiency (associated with the field of view radius) at MOPD is not excessive. For a larger MOPD, I would not be surprised if they used a smaller input aperture, which I presume should generate a smaller intensity at ZPD, if less light enters the instrument.

It is perhaps interesting to note that the ZPD peak intensity increases between 2013 and 2016, which is inconsistent with the suggestion that the performance is degrading over time, if one ignores the large drop between 2012 (with fresh optics and recent alignment) and 2013.

I am curious as to whether sampling issues were considered when finding the amplitudes. Figure 14 suggests a sampling drift, in that you do not have identical sampling of the HBr P(6) line in each spectrum, so there could differences in sampling for the interferogram as well. Do you Fourier interpolate the interferogram onto a finer OPD grid before determining the peak intensity?

In summary, I would like to see if the systematic features in the residuals for an HBr line with better signal-to-noise ratio than P(6) can be explained or reduced, possibly by better accounting for (non-ideal) stretching of the wavenumber scale of the measurement, if that is not currently done. I would like more concise and clear definitions of what conditions are included in the two configurations, because I cannot tell if the LINEFIT ILS is being used in the nominal configuration. It would also be nice to make clear how the approach described in this paper builds on the approach outlined in the Hase 2012 paper. What enhancements or differences are there compared to that paper?

Finally, the paper mentions many times how accurate modelling of the non-ideal characteristics of an FTS will improve atmospheric measurements. It would therefore be appropriate to include a comparison of atmospheric measurements from the instrument using nominal versus modified configuration. I would say there is no need to show that the results are better, just that they are different, along with how the differences compare to the uncertainties. However, I am hesitant to insist on adding the atmospheric results, since it will increase the length of an already lengthy paper, but perhaps it could be balanced by slimming some of the unnecessary discussion in the introduction.

Minor issues and typos:

> Line 237: Perispectives

typo

> Figure 6 caption: Data acquired on 121214

At this point, prior to Table 2, the date format has yet to be defined. The simplest solution would be to write out the date (December 14th, 2012).

>Line 321: drived

typo

>Line 331: Moreove

typo

>Line 353: the HITRAN database

Not everyone is guaranteed to be familiar with HITRAN. Perhaps a reference (and maybe a definition for the acronym)?

>Line 440:

As discussed for the interferogram ZPD peak heights previously, sampling could also play a role here for determining the asymmetry parameter. Are you interpolating onto a fine wavenumber grid before determining the sidelobe heights?

>Line 469: with the 3rd maximum showing the greatest reduction

Again the question of whether sampling is taken into account.

>Line 478: the residual MAE improves from 0.0849×10−2 to −5.9211%

I cannot interpret this statement as provided. The mean absolute error should by definition be positive, so it makes no sense to say that it "improves to -5.9%." I assume you are trying to say that MAE decreases by 5.9% for the modified configuration compared to the nominal configuration. That is not how the text reads. All the comparisons in this paragraph have a number compared to a percent value, rather than saying the stated number (obtained when using the nominal ILS) changes by the stated

percentage when using the modified ILS. Note that using 5 significant digits in the percentages implies that the precision in the residuals is good to 5 significant digits, meaning your signal-to-noise ratio should be better than 10000:1 (since the noise level limits the measurement precision). Looking at Figure 13, that is not the case. In my opinion, you should round the reported percentage change to a precision more in keeping with the signal-to-noise levels for the measurement.

>Figure 18: doesn't indicate which curve represents the nominal configuration results and which curve represents the modified configuration.

>Figure 19, bottom panel y-axis label: deference

typo