

Review of

OCIO and BrO observations in the volcanic plume of Mt. Etna – implications on the chemistry of chlorine and bromine species in volcanic plumes

J. Gliß, N. Bobrowski, L. Vogel and U. Platt

General comments

In this manuscript, the authors present the results MAX-DOAS measurements targeting halogen oxides in the plume of Mt. Etna. What sets these measurements apart from most previous ones is that they were performed in geometries specifically targeting the spatial distribution and temporal evolution of halogen oxide concentrations within the plume. The results are interesting, as is the discussion about chemical formation mechanisms and implications for atmospheric chemistry at large (specifically methane destruction).

In my opinion, the manuscript has one major overarching deficiency. Throughout the entire paper, spatial and temporal trends are interpreted that appear to be statistically insignificant. In several places, the authors argue that the trends might still be significant, even though the values either lie below the detection limit of their measurement technique, or the observed trends are smaller than the errors associated with the individual measurements. Such argumentation does not hold up to scientific standards.

I feel that there are two avenues for improvement, one or both of which could be explored to improve the manuscript. For one, the error treatment could be improved. Section 2 is already quite extensive, and clearly the authors have put much thought into accurately assessing the errors associated with their measurements. This is very important, as much of the measurement results are close to the detection limits of the instrument. However, perhaps the error assessment is overly conservative? In any case, it stands in contrast to the interpretation of the results that occurs later on. Improving the methodology for determining realistic errors might help consolidate this apparent conflict.

The other potential path towards consolidation is a reduction of the data into larger sample sizes. As it currently stands, the authors first evaluate trends found in individual plume scans, then evaluate trends in the aggregate of all their measurements. Due to the improved statistics, the interpretation of the aggregate dataset is significantly better constrained. The question then arises whether an interpretation of individual scans is even necessary, and what value it adds to the paper. It appears the authors could arrive at most of their conclusions without the need for unfounded interpretations of individual scans. Since the manuscript is quite long already, either removing the discussion of individual scans or moving it to an appendix could streamline the study, moving the scientific findings more into the foreground.

Other than this conflict between error assessment and data interpretation, I only have two other relatively minor general comments. (1) The organization of the manuscript might be improved by more clearly separating measurement results from conclusions. As it currently stands, the conclusions section

really just repeats the previously described results. One example is the estimation of Cl atom concentrations and relatively slow CH₄ depletion – this is more of a conclusion than a measurement result, and this section could be moved. (2) The readability of the manuscript could be improved by careful English language proof-reading.

Specific comments

P25215 L26 – Here you mention volcanic risk assessment and prediction of eruptions. These are two different things. Risk assessment is the assessment of potential risks to populations, e.g. certain areas could be affected by lahars, others by ash fall. Hazard maps are a good example of risk assessments. Please explain how gas emissions and plume chemistry can be helpful for risk assessments. Do you mean direct risk to populations from toxic gases? If so, please cite an appropriate reference. With regard to eruption prediction (better: eruption forecasts), there are examples where gases helped a great deal. The references you cite here may not be the best. Especially Carroll and Holloway (1994) is a book that, as far as I am aware, does not deal with risk assessment or forecasting at all. But perhaps I am wrong? Please either cite the chapter you mean or give a better reference.

P25217 L7 – You mention that photochemistry is involved in BrO formation, but do the references that you cite really “confirm” this fact? An increase in BrO/SO₂ ratio during plume evolution doesn’t really imply a photochemical reaction at all, or does it? Perhaps the active LP-DOAS measurements performed by Kern et al (2009) at Masaya (which you mention later) are more relevant in this context?

P25218 L15 – Is a release of BrCl really only found for Cl/Br ratios of the order of 1e4? What about larger ratios? Perhaps better to say the release is only effective for Cl/Br > 1e4?

P25224 L14 – How does DOASIS calculate the Ring spectrum? Why are two Ring spectra needed? If only R4 accounts for the wavelength dependency of the Rayleigh cross-section, why is R needed?

P25225 L10 – I believe you mean an “additive polynomial in intensity space”.

P25225 L13 – Section 2.4.1 is very important for the rest of the manuscript, as this is where the error treatment is derived. Citing Stutz and Platt (1996), you state that the measurement error is not given by the photon noise of the optical measurement, but is instead related to absorption structures of the trace gases. Then, a seemingly arbitrary choice of correction factors U is introduced whereby U is related to the peak-to-peak value of the residual. It does not become clear from the text why these values for U were chosen. Stutz and Platt (1996) derived a relationship between the frequency of the residual, the frequency of the respective absorption cross-sections, and the correction factor U. They did not derive a relationship between the peak-to-peak value of the residual and U. Any such relationship therefore needs to be justified in the text.

For the sake of the later interpretations, it would probably be useful to discuss both accuracy and precision here. If the authors feel that the precision is greater than the accuracy of the measurement, a

separate method for deriving the measurement precision should be investigated. This could possibly be based on the magnitude of high frequency shot noise in the measurement residual.

In any case, the discussions in this section need to result in an adequate and well-motivated method for deriving the overall accuracy of an individual SCD (how close to the true SCD?), the precision of individual SCDs to one another, and the detection limit. By definition the detection limit is understood to be the limit above which a positive detection is obtained (perhaps, as you say, with 95% confidence). By definition, all measurements below the detection limit are non-detections, meaning that the SCD is equal to some unknown value below the detection limit. All measurements below detection limit are equal in this sense, and investigation of trends below the detection limit doesn't make much sense.

P25227 L12 – I recommend removing Section 2.4.3. and figure A1. If this alternate wavelength region gave less reliable results, there is really no reason to discuss it further. If the authors feel strongly about mentioning the fact that this range was found to be less suitable, then a single sentence in section 2.4.2. would suffice.

P25230 L15 – Radiative transfer effects can only be neglected for these two species if their optical depths are small, i.e. they are both “weak absorbers”. This is the case for your data, but should also be mentioned in this context.

P25231 L25 (footnote) – Why must vertical plume propagation be negligible for your approach to work? I don't understand how vertical plume rise (be it convective or buoyant) would affect the age of a plume measured at a well-defined horizontal distance from the vent. Is this because you are assuming a varying vertical wind profile?

P25233 L21 – It is unclear how SO₂ camera measurements can be used to accurately derive wind speed and wind direction, given that the images always represent a projection of the 3D scene into two dimensions. Please explain how this was done.

P25235 L2 (Eq 7) – I'm not sure why gamma is introduced here. This is simply the difference in stratospheric AMF between measurement i and measurement j. Consider using ΔAMF_{ij} or similar.

P25235 L6 to P25236 L17 – I was not able to follow the argumentation in this section. First of all, I don't understand how the range of stratospheric BrO VCDs was arrived at. Deriving a VCD according to Equation 7 assumes that this VCD is constant over the time between measurement i and measurement j. In principle, any permutation of i and j is possible, even spanning completely different days. Is this how the VCDs were determined? Or were only consecutive spectra evaluated against each other? In other words, how was the subsample of spectra used for this calculation determined, and how were the individual spectra paired? And why was this combination chosen?

In the second part of the text, a sensitivity study of stratospheric BrO VCDs on the measurements is attempted, as shown in Figure 5. However, it is a bit unclear why plotting dS_{str}/S_{meas} is useful here. Wouldn't it be better to discuss the potential errors in the BrO SCDs caused by an inaccurate stratospheric correction? For example, one might say that, based on the range of retrieved stratospheric

BrO VCDs, the 4×10^{13} molec/cm² VCD assumed for the correction has an approximate error of $\pm 2 \times 10^{13}$ molec/cm². Then, one can calculate how such an error would influence the stratospheric correction of a given BrO SCD, and adjust the previously derived SCD error appropriately in all plots. This does not appear to be the route the authors have taken here, and I do not understand their approach.

P25237 L25 – If BrO was only detected within the plume, does this mean that any potential error in the stratospheric correction smaller than the BrO detection limit?

P25238 L18 – Wouldn't errors in the wind speed cause a stretch/squeeze of the dataset, not a shift?

P25238 L19 to P25239 L2 – The given discussion of Figure 7 is a good example of over interpretation of uncertain data, as mentioned in the general comments above. The authors write "Figure 7 shows an increase of the OCIO/SO₂ ratio up to a plume age of 120 s and a slight decreasing trend afterwards". I would argue that, as depicted in Figure 7, OCIO was not detected for plume ages shorter than 120 seconds, and was then constant within the uncertainty of the measurements up to an age of 200 seconds. This is a significant difference. In fact, in the presented plot, it appears that the non-detection of OCIO for ages shorter than 120 s might be simply a result of a poorer detection limit caused by lower SO₂ columns measured in this part of the plume. Therefore, as it stands, the data shown in figure 7 is in principle consistent with a constant OCIO/SO₂ ratio in the young plume. I'm not implying that I believe this to be the case, but based on this particular example dataset, one cannot rule it out. The authors go on to state that the depicted errors also include systematic errors. This may be true, but in the absence of a way of quantifying how much is systematic and how much is random, it is not possible to interpret figure 7 in the way they have done. This again points to the need to either improve the error analysis or increase the statistics by additional averaging. Or both (see general comments).

P25239 L3 to P25241 L13 – In this section, the individual scans perpendicular to the plume are evaluated. Again, I feel the data is over-interpreted. The data shown in figure 8f is essentially consistent with a constant BrO/SO₂ ratio throughout the plume, perhaps with the exception of one point measured at about 54 degrees elevation. The data shown in table A1 and described in the text is similarly inconclusive, if viewed one scan at a time. Out of the 10 scans listed in the table, only 2 have statistically relevant differences in their BrO/SO₂ ratios. In all other cases, the ranges given by Mean1, Std1 and Mean2, Std2 overlap (by the way, line 9 is missing some data). Therefore, it is questionable whether the statement "The BrO/SO₂ ratio showed an increase towards the edges in 76% of all analyzed scans" is true or not. Couldn't one just as easily say that the BrO/SO₂ ratio only showed a statistically significant increase towards the edges in 20% of the scans?

P25242 L22 – The discussion of Figure 8d is again difficult because all data is below the detection limit. As per definition, this means that OCIO could not be detected. Therefore, you cannot argue that there is an increasing trend. Either the definition of detection limit must be revised or this discussion removed.

P25242 L8 – "This could be the result of a superimposed vertical profile of BrO" What does this sentence mean? Please clarify.

P25243 L1 – This statistical analysis is exactly what I mean by reducing the data to improve statistics and derive statistically relevant trends. In my opinion, this result is by far more robust than the previously shown results and the authors should consider showing just this plot (figure 9). Note that in this plot, there is not difference in the formation time of BrO and OCIO.

P25243 L20 – It is, however, important to mention that the “slight decreasing trend” in OCIO/SO₂ ratios with distance is not statistically significant.

P25244 L10 – The statistically more robust dataset does not show a difference in formation time between BrO and OCIO. So does your data really support such a difference?

P25245 L14 – “an increase of the OCIO/SO₂ ratio might be observable”. Again, if all these values are below the detection limit, then deriving a trend is probably not valid.

I also don’t understand why the detection limit is so constant over time (particularly for BrO). Since the measurements were taken so early in the morning that the incident UV radiation was quite limited, wouldn’t one expect at least some improvement of the detection limit as the more UV light becomes available?

P25246 L14 to P25246 L25 – Here the authors write an entire paragraph about a single, extremely uncertain point of data (why is the uncertainty range not given for 0.25?). This type of excessive discussion of extremely uncertain data is contra-productive to the overall message of the paper. This point can be mentioned, but I a maximum of 2 sentences would suffice to put it in perspective.

P25247 L1 – I very much enjoyed this section. Perhaps consider moving it to the conclusions though?

P25247 L21 (Eq. 10) – I believe you need to specify that [ClO_v] on the right side of equation 10 needs to be evaluated at $t = \tau_0$, i.e. [ClO_v](τ_0).

P25250 L10 – Please motivate why these correction factors were chosen. Simply stating that they are large doesn’t help very much. In particular, why were different values chosen for different species, even though their cross-sections appear to have similar frequencies?

P25250 L5 – Please give at least the average X/SO₂ detection limits for $\tau > 3$ minutes in the text.

P25250 L13 – As it stands, the conclusions section provides no new information, instead simply summarizing what was discussed before. I recommend restructuring the manuscript and separating the observations from the conclusions (see general comments above).

A number of points of contention are mentioned in the conclusions, but they are repeats of points mentioned earlier and I will not repeat my comments here. Just note the following: (1) I’m not sure that the “OCIO/SO₂ ratio showed a strong increase in the first three minutes after release”, (2) I’m not sure the “OCIO/SO₂ increase could be observed in 6 individual scans on two different measurement days”, (3) If the apparent difference in formation time is “probably due to uncertainties in the plume age determination”, but the evaluation of individual scans results in a 70% longer formation time for OCIO,

then either the uncertainty in plume age is about 70%, or the comparison of individual scans is not statistically robust, (4) only 20% of the scans given in Table A1 have BrO/SO₂ ratios that are not consistent (within 1 standard deviation) with a constant ratio throughout the plume. (5) If only 8% of the measurements are affected by uncertainties in stratospheric BrO, couldn't those just be omitted for all calculations except the photochemistry sunrise experiment? Then you wouldn't have to worry about that problem elsewhere. (5) Mentioning the single measurement of OCIO/ClO at longer plume age is questionable, because if I'm doing the math right, then the range for that point is 0.06 - 1.6.

Minor corrections

P25214 L5 – In the first sentence you define abbreviations in parenthesis (e.g. SO₂). In the next sentence you use parenthesis to present values obtained for different species, e.g. OCIO (BrO). This is confusing to read.

P25214 L8 – Recommend ending sentence after “OBrO”, starting new sentence with “None”

P25214 L19 – “... showed A BrO/SO₂ ratio...”

P25215 L12 – “CONSIDERABLY SMALLER THAN THE ATMOSPHERIC LIFETIME”

P25216 L14 – “oxides of nitrogen” = “NITROGEN OXIDES”

P25218 L14 – Recommend removing the dashes before and after “Cl releasing” as they are unnecessary.

P25219 L19 – Please start the sentence with “A key parameter...”

P25221 L4 – Replace “due to” with “USING”

P25222 L20 – “GAS-FREE”

P25224 L4 – “... background spectrum FROM LITERATURE as an FRS...”

P25231 L2 – “slightly slower THAN stratospheric OCIO...”

P25232 L14 – what does “rather linear” mean? I would say uncertainties in the wind velocity have a linear effect on the plume age uncertainty.

P25232 L18 and figure 8 – if you refer to this contribution as $\Delta\tau_{\text{wind}}$ in the text, please use the same notation in the figure.

P25233 L26 – “These data WERE used...”

P25234 L18 – “... that the STRATOSPHERIC air-mass-factor...”

P25236 L4 – “... ratio of these species TO the retrieved...”

P25237 L2 – “improved sensitivity”? Improved over what?

P25237 L3-4 – Please be more specific with regards to exactly what days the measurements were performed.

P25240 L2 – What does “suited in terms of their SO₂-plume coverage” mean?

P25242 L21 – “... performed on different days and at different times.”

P25244 L8 – “...start leveling off” is not precise enough. Please be more specific in your definition.

P25246 L1 – Throughout this section, please replace “concentrations” with “mixing ratios” as you are reporting relative mixing ratios, not absolute concentrations.

P25246 L2 – “determined AVERAGE BrO and OCIO ...”

P25246 L8 - Replace “whereas” with “with”

P25247 L8 – Positive or negative temperature dependence?

P25247 L12 – I assume you mean “All other Cl-sink reactions...”?

P25249 L8 – The typical O₃ background IS 60-80 ppb, not “should be”.

P25249 L12 – Maybe also cite some measurements of O₃ depletion, not just modeling work? For example, Kelly et al 2013, Rapid chemical evolution of tropospheric volcanic emissions from Redoubt Volcano, Alaska, based on observations of ozone and halogen-containing gases, JVGR. This would also be good in the context of discussion what “really low” O₃ concentrations would be, see line 24 on this page.

P25249 L28 – What are conditions “more favorable to Cl-activation”?

P25250 L1 – “... IO (in the UV spectral range), OIO and OBrO (in the VIS spectral range)...”

P25250 L2 – I recommend not mentioning the second spectrograph here. The fact that your instrument has 2 spectrometers was already discussed earlier.