Response to reviewer 2

Thank you very much for the very good comments. In the revised manuscript, we tried to follow the reviewer's comments as possible as we can. Responses for the comments are addressed below.

Major comment

1) The manuscript does not follow the usual titles, which is fine by me except for the fact that I would like to see a section shortly describing the different instruments/data: IASI, GANAL, and I would also transfer to that section the RI information and maybe also the radiative transfer code if it is re-used from a previous publication and with only a short description (see also another major comment on the RT itself)

⇒ The word "Optimal" was removed from the title. A short explanation for IASI was added at the end of section 1 (L70-73). For GANAL, Web address of the numerical weather prediction system was added as well as a short description (L322-325), GANAL is the results of the global assimilation analysis by using the atmospheric state of Global Spectral Model (GSM) as *a priori*.
We would like to omit the detail of our RT code (MBCRM) because the volume of the entire manuscript getting large. Instead, results for the difference from line-by-line calculation for clear sky atmosphere (Fig. 1), and a comparison between MBCRM and RTTOV for cloudy atmosphere by using a volcanic ash model in RTTOV (Fig.2), were added.

2) Please be quantitative when assessing the quality of a result, do not use "good agreement", "good fit", "agree well" or similar but provide numbers (RMS, or other more relevant depending on the case)

⇒ It was one of the most difficult points in writing this manuscript. Estimation of RI model is easier for ash plumes which show large spectral variation (V-shape) in the measured brightness temperature (BT) because the difference between the RI models appears clearly in RT calculations. At the same time, however, RMS between measurements and calculations tend to enlarge because the applied RI model and ash physical parameters in the RT calculations are imperfect and small difference in the RI causes large difference in the calculated BT spectra (BTS). Therefore, absolute value of RMS is not always the appropriate indicator to evaluate the RI models, though it can show the difference in fitting between the RI models. Although a new relevant parameter that precisely shows the degree of matching is possible, we could not derive it so far.
In the revised manuscript, we added the plots of mean and standard deviation of RMS over pixels of each eruption event (Fig. 9) to follow the reviewer's suggestion.

3) The strength of "V-shaped" spectral feature of volcanic ash aerosols (or the slope on each side of that V-shape) depends on the refractive index, the effective radius, the optical thickness and the plume altitude. In the cited ref Clarisse and Prata 2016, the Fig 10 shows clearly the dependence of the split-window BTD

with optical depth and effective radius for a specific refractive index, and in Fig. 3 of the same reference the authors also show the impact of the plume altitude, for a specific refractive index (and others have reached the same conclusions e.g. Maes et al, 2015, doi:10.3390/rs8020103). In this manuscript under review, it seems that the problem is sometimes taken too "lightly", not considering this complex relationship between the 4 parameters and the fact that their retrieval might well not be independent, and multiple solutions may be plausible. For example this sentence, line 62: "The particle size of ash clouds, which can be determined from the negative BTD between two infrared split window channels" -> yes, if you know the optical depth (or thickness) and refractive index.

⇒ Some sentences were modified (L64-66), and a detailed explanation for the VAC parameters (P_{top}, r_{eff}, τ_c) and O3/SO2 retrieval was added (L374-379, L383-387). To avoid the *a priori* dependence, forward calculations are performed for all combination of the three VAC parameters by changing the value of each parameter within the assumed range. The VAC parameters which give the minimum RMS are derived for each RI model, and the minimum RMSs are compared between different RI models. As the reviewer commented, multiple solutions may be possible depending on the applied RI model. In this work, we discussed the residual of BT difference between measurements and calculations (minimum RMS) which cannot be removed by the change of the VAC parameters. Therefore, we don't take some VAC parameters lightly in our retrieval calculations. However, we did not discuss much about the results of the VAC parameters because we analyzed only for the measurement pixels of large split-window BTDs and the results may disagree with the typical/average properties of the measured VAC. The estimated VAC parameters are provided as the supplementary materials for verification of our radiative transfer calculations. Another important issue is that we are assuming a homogeneous ash layer within the area of IASI footprint. This assumption may cause an underestimation of the VAC top height. In our test calculations for some measurements in this work, the estimated VAC top heights increased 1-2 km when we applied ash fraction (same as cloud fraction) 0.7. A short comment was added in the revised manuscript (L410-414).
Regarding Maes et al. (2016), comments were written at response No. 36.

4) The radiative transfer code: the authors state that they "developed an original radiative transfer code", using a demonstrated approach. This is fine, but it is very unclear to me if the authors re-used the same code, or wrote a new code. In the first case, I would not think appropriate to write the sentence I cited here. In the second case, I would like to see some "validation" of the radiative code. What is stated in lines 88-89 is unclear (I don't understand against what the calculated spectra were compared) and not sufficient to demonstrate that the radiative transfer works as expected. I would like to see comparisons with at least one other well-established RT code.

⇒ Same as the response No. 1. A plot of channel RMS between MBCRM and Line-by-Line calculations under clear sky conditions (Fig.1), and a comparison between MBCRM and RTTOV for cloudy atmosphere

(Fig.2), were added.

5) Particle size: in the cited reference Clarisse and Prata 2016 it is written that "IR sounders are highly sensitive to the effective radius of the distribution within the range 0.5-5 μ m" (page 198), and this reference is cited to justify the particle size retrieval done in the current manuscript. Then in the results (Table 3) 6 over 21 cases end up with an average particle size smaller than 0.5 μ m. How much can we trust these results?

⇒ We used a dataset of IASI measurements which show large negative ΔBT_{split}. These data may be extreme cases and the results of the retrieved r_{eff} may differ from that of typical or average ash property for the target eruptions. As far as the IASI measurements shown in this study are concerned, there is a problem with the applied RI models if the effective radius r_{eff} smaller than $0.5\ \mu\mathrm{m}$ is unrealistic and if our radiative transfer calculations are not very wrong. As shown in Fig.11d in the revised manuscript, the brightness temperature spectrum (BTS) changes when the ash r_{eff} changes from $0.5\ \mu\mathrm{m}$ to the smaller, and the effect mainly appears in wavenumber range $1070-1230\ \mathrm{cm}^{-1}$. This leads that the results of r_{eff} strongly depend on the absorption property of the applied RI model in this wavenumber range, and the BT difference between the two ends of the ozone band ($1000\ \mathrm{cm}^{-1}$ and $1070\ \mathrm{cm}^{-1}$) is a good indicator. If we use the RI model which has a feature of the absorption index k as $k_{1000\mathrm{cm}\hat{}1} \geq k_{1070\mathrm{cm}\hat{}1}$, a drop of BT from $1000\ \mathrm{cm}^{-1}$ to $1070\ \mathrm{cm}^{-1}$ in the IASI measurement can be explained only by a small size of r_{eff}. As shown in Figs. 3-4, most of the RI models by Reed et al. (2018) and Prata et al. (2019) have $k_{1000\mathrm{cm}\hat{}1} \geq k_{1070\mathrm{cm}\hat{}1}$ features. In our BTS simulation, the features of the BTS for Calbuco_A (Fig.13c) and Puyehue-Coedon Caulle (Figs. 16c, g) could be simulated by small particles with $r_{eff} < 0.5\ \mu\mathrm{m}$ though the results of RMS are relatively large comparing to the other volcanic eruptions in Table 2. To obtain a retrieval results of $r_{eff} \geq 0.5\ \mu\mathrm{m}$ in the cases of Puyehue-Coedon Caulle, an RI model which has a strong absorption peak as that of rhyolite model (PLRHY in Fig. 5) by Pollack et al. (1973) is necessary. Although the results of r_{eff} were $0.7-0.8\ \mu\mathrm{m}$ (Supplemental Material 2), the results of RMS by using PLRHY were relatively large because of the disagreements of the simulated BTS in other wavenumber range.

For this issue, we think that a further investigation is necessary. At least, we don't know the study about the results of BTS simulations which successfully represented the whole measured BTS with large negative ΔBT_{split} for Puyehue-Coedon Caulle and Calbuco cases as those of Fig. 10c and Fig. 13 by using $r_{eff} \geq 0.5\ \mu\mathrm{m}$.

6) At the end of page 7, it is said that the method "aims to select pixels showing sparse VACs comprised of small particles". I am unsure to understand this. As mentioned before the large slope in the V-shape does not ensure small particles, it could also be high optical thickness and/or high altitude ash aerosols. Second, why would it target only "sparse VAC"? In the selected spectra, I see some with a very strong ash signature,

which I would not refer to as sparse ash (for example fig 9c with about 20K drop in BT along the V-shape)

⇒ In the revised manuscript, the sentence was slightly modified (L339-340). As a general tendency, negative ΔBT_{split} becomes very weak when the ash r_{eff} of VAC is large (approximately more than 5 μm, as shown in Fig. 11a, and Figure 4 of Clarisse and Prata (2016) page 197). In that condition, the difference of the minimum RMS between the applied RI models becomes small, and it is difficult to discriminate the results of different RI models. On the other hand, strong V-shape can be realized even for a high optical depth of the VAC. As we discussed in this section, such optically thick ash plumes show similar BT spectrum as those of the optically thin ash plumes which locate above the optically thick ice/water clouds (MC). Since the contamination of MC may causes a large error for our RI estimation, we applied Eq. (2) to reduce the probability of the MC contamination in the measurement dataset for the analysis. The condition of Eq. (2) rejects the data of optically thick ash plumes as well as MC contaminated ash plumes. As the results, we select the data of optically thin ash plumes comprised of small ash particles by applying the conditions of $\Delta BT_{split} < -2$ K and Eq. (2).

7) The retrieval method (pages 8-9): actually, no retrieval method is explained. It is only mentioned that parameters are "estimated". This part should be much more detailed, as all results depend on how much the retrieval can be trusted. If I understand correctly, there are 2 steps: one estimating the ash parameters (how and from which a priori values?), and then a second step to estimate O3 and SO2 parameters estimate (same questions). In those steps, surface temperature and temperature profiles are maintained constant (not retrieved) to the GANAL values – which are not described, see another comment on this later. In most of the thermal infrared retrievals, at least the surface temperature is a retrieved parameter (because a wrongly assumed Ts has a devastating impact on the retrieval results), why is it not the case here? Then, when calculating the RMS, it is very unclear why "error in the GANAL atmospheric profiles" (what exactly do you mean? Uncertainties or biases? In which profiles?) would have an impact only between 650 and 750cm-1. If there are indeed higher uncertainties/bias in that spectral range, then why use it at all in the retrieval? Finally, I would like to see a discussion on the possibility to retrieve together the altitude, optical thickness and particle size of volcanic aerosols, linked to the retrieval method use, information content, a priori and constraints of the retrieval. At this point, I am not convinced that there is enough information in the observations to retrieve it all together, except maybe with extremely strong constraints (and then the choice of a priori value is very important). But none of this is discussed.

⇒

・Some sentences for the description of the VAC parameters estimation were added (L374-379, L383-387).
・GANAL is the output of a four-dimensional data assimilation using 6-hour intervals forecasts of the Global Spectral Model (GSM) as *a priori*. Since the assimilation uses various data including satellites and ground observations, it is expected that GANAL has some accuracy for the atmospheric profiles and sea

surface temperature (SST). In this study, SST and the atmospheric profiles were used as fixed values. Although it is possible to estimate the SST by using the data of IASI channels in the vicinity of clear sky near the ash plume, we did not do that because there is a possibility to increase the error.

・Because the brightness temperature (BT) of CO_2 channels in wavenumber range $650\,\mathrm{cm}^{-1} \leq \nu \leq 750\,\mathrm{cm}^{-1}$ strongly depends on the temperature profiles, the RMS of BT between measurements and simulations can be very large depending on the applied GANAL profiles. Although the BT in this wavenumber range is important in estimating the VAC parameters (especially top height of the ash layer), the large RMS in this wavenumber range may underestimate the RMS due to difference in RI models. In such reason, we excluded the channels in this wavenumber range from the final RMS calculations.

8) The supplementary material contains a lot of information, but lacks a short description. For example, I had to guess that S1 contained the RMS. For S2, maybe some graphs would be useful (as in Fig. 7) in addition to the numbers. Those could be also in supplementary information, but I think that just the table with all the numbers is very difficult to analyse quickly.

⇒ S1 file was revised to know that contained RMS data. The plots of the VAC parameters for the eruption events were added in the same folder of S2. The data of measured and calculated BTS for the VAC in Table 3 was provided as another supplementary data (S3)

9) I do not think that it is enough to mention in the paper that the RI leading to the lowest RMS was selected; you need to provide some statistics of that RMS in the paper. I would do that in Table 2: add columns with mean RMS (not total, that does not allow comparing between different eruptions) and standard deviation on RMS, or number of RI leading to a mean RMS within a certain range (e.g. 0.15K = IASI noise from the first estimation) from the "optimal", to show if there were large differences between the results with different RI. In addition, I find much too limited to just take the minimum RMS as criterion (with exception of the additional criterion on size) to select "the best" RI. Indeed, in many cases the mean RMS for different RI is very close, as I will underline again in the specific comments by eruption. The difference in RMS between different RI in that case (when the difference is very small) could just be linked to uncertainties in the other parameters (surface and atmospheric T being the most important), and therefore the conclusion on the selected RI could be wrong.

⇒ Plots of the mean and standard deviation for each RI model were added as Fig.9. As the reviewer pointed out, the differences of the RMS between the RI models were small in some eruptions, and it was difficult to identify a specific RI model from the RI dataset used in this work. We listed 1-3 RI models in Table 2 as the representative examples. However, the best RI models may change depending on the applied atmospheric profiles and SST. Some sentences were added in sec 5 (L467-469) and summary (L1064-1067).

10) When the selected pixels are less than 5, the statistical significance is pretty low. This is mentioned in the text but should be briefly mentioned in the Table 2 caption

⇒ We eliminated the parts for Bezymianny, Rinjani, Sarichev_peak, and Zhupanovsky from the revised manuscript.

11) In Table 3 the mean parameters for each eruption are listed, but it makes little sense to average SO2 content, ash optical depth/thickness and plume altitude at different locations and different times (even if for the same eruption, those vary a lot with time and location). Only (maybe) the effective radius can be considered to not vary much – although it should vary as the biggest particles fall sooner.

⇒ The parameters in Table 3 are not the average but those to simulate the measured BTS for the figure of "Figure No.". We made this table to simplify the figure captions. These parameters may be useful for validations of our RT calculations though the same data can be obtained from S2 file.

12) Spectra as they are represented do not really allow seeing how big the difference is between observed and modelled. For example, in Fig 6 spectra look really similar but the simple fact that, at such a small scale, we can see the different colours means that actually the difference is probably of the order of 1K, which is significant. I would suggest to plot all spectra as in Fig. 7 (or zoom even a little bit more), with focus on the 2 slopes of the V-shape and not showing in full the CO2 O3 and H2O bands.

⇒ Ranges of X and Y axes were modified for all the plots of brightness temperature spectra.

13) In most results Figures (from Fig. 6 onwards) the legend is pretty difficult to read

⇒ The fonts of the legend in the figure were magnified.

Specific Comments

14) Line 107: is an interval of 10cm-1 enough to reproduce the fine features of the used RI?

⇒ Although the original RI data by Reed et al. (2018) and Prata et al. (2019) are given in $1\ \mathrm{cm}^{-1}$ intervals, we think a linear interpolation for the data of $10\ \mathrm{cm}^{-1}$ interval is enough to simulate the wavenumber dependence of their RI models and the resultant BT spectrums.

15) Line 112: "and artificial weak absorption features were added to them" -> this reads very weird; clearly I understand what the author means, after reading the complete manuscript, but when reaching this part it is very unclear what is meant. I think that the authors should here either detail what they mean, or clearly refer to a further paragraph.

⇒ In this study, the andesite model and the rhyolite model by Pollack et al. (1973) were prepared for comparison with Reed et al. (2018) and Prata et al. (2019) RI models. We replaced MP-A and MP-R models to the original andesite and rhyolite models (PLAND and PLRHY) by Pollack et al. (1973). All simulations for these two models were recalculated and the text was revised. Because the original PLAND gave better results than the modified model for Grimsvotn VAC, some sentences in the text were also modified.

16) Line 235: is the BTD for the selection of the IASI scenes calculated using only 2 channels? If not, the details should be given; if yes, then I strongly suggest using the average of a number of channels, to reduce the impact of noise. Indeed, at those wavenumbers, the IASI spectral noise was reported to be ~0.15K at the early years of the instrument (Clerbaux et al., doi:10.5194/acp-9-6041-2009) and a bit later ~0.25K (Hilton et al, doi:10.1175/BAMS-D-11-00027.1). Therefore, a simple BTD without any averaging may bear an error of up to 0.5K, which is 25% of your threshold.

⇒ We thank the reviewer for the information. Regarding the measurement noise, we agree with the reviewer. On the other hand, many IASI channels are affected by the narrow absorption lines in the upper atmosphere, and a simple average with neighboring channels may cause another error. In this study, two channels are used, which are considered to have high BT and little influence of the narrow absorption lines. The BTD is used as a threshold to select the IASI pixel data and has no influence to the results of the retrieval. Therefore, we didn't care much about the exact BTD value.

17) Line 250 and following -> only day-time IASI (because MODIS is used to analyse the imagery) but then Line 270 also night-time data is included in the analysis. This is unclear. At the end, is it only day-time IASI pixels or also night-time?

⇒ Night-time data were included to increase the number of data points in Fig.6 to show the distribution of the points. No night-time data was used for the radiative transfer analysis in this study. A sentence was modified (L320-321).

18) Lines 261-263: "As the temperature of the MC layer is generally lower than that of the sea surface at the same geographic location, a VAC above an MC layer tends to have a lower infrared brightness temperature than a VAC with no MC if the cloud parameters of the VAC

are the same" -> This is confusing and should be rephrased. The VAC will have the same brightness temperature in both cases (unless the temperature profile is changed). The observed spectrum is not the BT of the VAC (unless it is optically thick and the surface underneath is not seen at all), but the surface emission (either of the sea surface or of the MC) followed by the atmospheric impact of all gases and aerosols between the surface (again, sea or cloud) and the instrument. The observed BT (not the VAC BT) at the satellite is indeed usually lower if the ash plume is above a MC than above sea, on the condition that the ash cloud is not optically thick (and this happens).

⇒ Similar sentences were used in the revised manuscript (L310-314).

19) Line 266 and further: TB_obs(nu_a) is only at one wavenumber? Then as for the spit window do not forget the spectral noise. In addition for the calculation of TB_clr don't forget the model uncertainties (those linked to RT should be minimal, hopefully, but there could be significant uncertainties linked to the surface T and the T profile)

⇒ The wavenumber ν_a was estimated simply from the channel of the maximum BT in the range of $750\ \mathrm{cm}^{-1} \leq \nu \leq 900\ \mathrm{cm}^{-1}$. In this study, ν_a is used for the calculations of ΔBT_{clr} in Eq. (2). As the same reason denoted in the response No. 16, we did not care the exact value of ΔBT_{clr}. However, we agree with the reviewer's comment about the importance of the atmospheric profiles and SST.

20) Lines 270-272 Please provide a reference and more details on the GANAL data.

⇒ A simple explanation and the address of a reference for the Global Analysis (GANAL) of JMA were added (L322-325).

21) Line 279: "artificially"? = empirically?

⇒ The sentence was revised (L333).

22) Line 382: the sentence starting with "on the other hand" reads weird, it gives the impression that the MP_A should have been selected, but I guess that you mean it was better than the original but still not the best RI?

⇒ We replaced to the original RI models by Pollack et al. (1973) in the revised version.

23) Line 383: "high accuracy BTS simulations …" -> the accuracy is currently never discussed in this manuscript: you would need to provide some RMS values, and discuss those with respect to IASI

instrumental noise (and even better with respect to uncertainties in the non-retrieved parameters but maybe that is out of scope here)

⇒ The sentence was revised (L477-478).

24) Line 477: the data from Ventress et al given in the () is actually the a priori of their retrieval, not their result. The result of their OE was a particle size of about 1 μm and a plume top altitude of about 3.5km, which are I guess the values that the authors read in Fig. 5. Please rephrase correctly the whole sentence lines 476 to 479.

⇒ The values in the () were read from the left column of Fig. 5 in Ventress et al. (2016). The sentence was revised because the values were inaccurate. (L618-619)

25) Line 514: “no systematic bias was apparent” -> bias with respect to what?

⇒ The sentence was revised. (L662-663)

26) Lines 522 to 526: it should be stated (as in Fig 7 caption) that the other ash parameters were kept constant.

⇒ The figures are changed (Fig.11d,e). The sentence was modified to follow the reviewer’s comment (L671-673, L714-717).

27) Line 527-528 “Therefore, selection of the proper RI model is essential, especially for the estimation of Reff. This result also suggests that RI model selection may have a strong influence on the estimation of the ash column density, …” -> I am not sure to understand; the word “especially” in the first sentence is to emphasize that the impact is targeted on that variable, then a second sentence says that another variable is also highly impacted. I would remove “especially”. Second, in your results, the retrieved optical thickness does not seem to depend on the selected RI (Fig 7c), in contradiction with this second sentence.

⇒ The sentence was revised (L676). We intended to mention that the selection of RI model strongly influence the estimation of r_{eff} in the retrieval of the VAC parameters.

28) Figure 7: in the caption there is something weird, the d and e are described twice, and I think the first is wrong. In addition, in those Figs 7d and 7e, I would like to see the spectrum with the VAC parameters

leading to the minimum RMS (the one "selected").

⇒ The figures (Figs. 11d and 11e) were replaced for better understanding the relation between the particle sizes and RI models.

29) Eyja results: the mean RMS on all pixels for all RI range from 1.039 to 1.388K with 12/21 RI leading to a mean RMS below 1.1K, so within 0.06K from the "optimal RI". Considering the IASI spectral noise of 0.15 to 0.25K (see before) and the possible bias linked to surface temperature and atmospheric parameters, can we really conclude that only one (or two) RI is "correct" in this case? [This is connected to my Major Comment 8.2]

⇒ Some sentences were added and denoted "... RI models used here may change depending on the measurement dataset and the atmospheric conditions in the BTS simulations, such as temperature profile and surface temperature" (L471-472). We put similar sentences in the summary (L1064-1067).

30) Line 579: "In addition to Eyja…" -> This reads weird as a start for the new section, what do you mean?

⇒ The sentence was revised (L721).

31) Grimsvotn results: the top altitude seems really low with respect to what is reported in the literature; indeed Moxnes et al report an ash top height below 4km, but actually with ash up to about 4km (and matching all measurements, not only IASI). Therefore the plume altitude reported here (top at 1.6km) does not really match the cited literature, on the contrary to what is stated in line 599. A wrong retrieved ash altitude could be linked to a biased temperature profile; this is worth a check when results do not match anything previously published. For that eruption, the mean RMS on all pixels range from 0.806 to 1.198K for all RIs, with 11/21 RI leading to

⇒ According to the Figure 13 of Taylor et al. (2019) (https://doi.org/10.5194/amt-12-3853-2019), the top heights of the Grímsvötn VAC measured by CALIOP at $\mathrm{lon.}\,60-62^{\circ}N$ and $\mathrm{lat.}\sim20^{\circ}W$ on the same day (13:25UTC) were around 3 km, and the backscatter peak heights were 2 km or less. Furthermore, the result of optimal estimation height by their calculations at lat. ~60 deg. was between 1 and 2km. We think that the retrieved plume top altitude (1.7~1.8km) in our calculations are not too low. Taylor et al. (2019) was also referred in the text (L740-743).

However, our retrieval results possibly underestimate the cloud top altitude. In our radiative transfer calculations, a homogeneous ash layer (ash cloud fraction =1) is assumed within the area of the IASI pixel footprint (~ 12km at nadir). If the ash cloud covers only a portion of the pixel area, the estimated top height becomes significantly lower than the truth. Although it is difficult to estimate the ash cloud fraction from infrared sounder measurements, we would like to improve our retrieval algorithm in the future. Regarding the dependence of the estimated ash top altitude on the fraction of ash layer within the area of pixel footprint, we briefly denoted in the revised manuscript (L410-414).

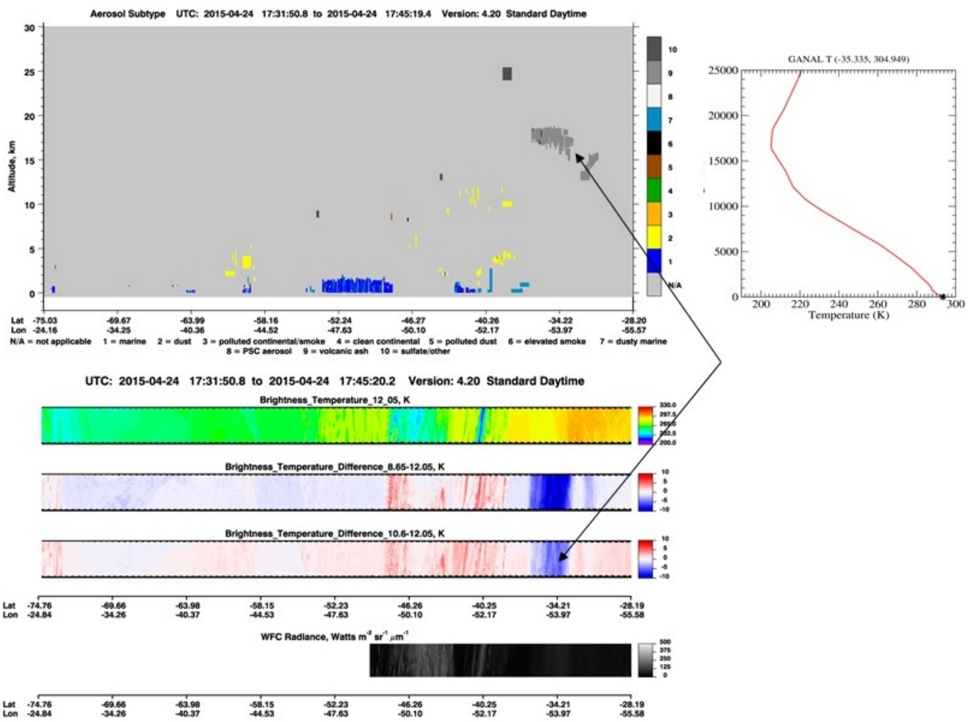
32) Lines 624-626: I am not convinced of that conclusion, because of all the comments done on the results.

⇒ The words "optimal RI model" were replaced to "some appropriate RI models" in this sentence (L768-769).

33) Calbuco results: there is no comparison with literature here; the Calbuco A mean RMS are all much higher than for other eruptions (except for the rejected RI which had a mean RMS of 1.572K, the other RI give a mean RMS between 1.823 and 3.131K) while for Calbuco B only 1 RI model leads to a mean RMS above 1K. In addition, when looking at the spectra for A and B, it is very clear that the situation is completely different. To me, the A spectrum (Fig. 9c) could hardly be for an optical thickness of only 0.14 even if at high altitude (a drop of more than 20K in radiance…) – in comparison the Kelud shows an optical thickness of 0.6 at 11km altitude for a similar radiance drop. And the different mismatches in the A spectra give me the feeling that something is not correct in the atmospheric composition, either some remaining ice clouds, or another problem in the atmosphere setup or surface temperature, or even in the radiative transfer. In any case, I think this is worth a discussion (the complete difference between A and B situations for plumes coming out of the same volcano at the same eruption) and some additional analysis to be able to trust the conclusions based on both groups of pixels.

⇒ Although there may be an error due to the atmospheric profiles and surface temperature in our radiative transfer calculations, the measurement (red line in Fig. 13c) is true. From CALIPSO measurements, the ash plume around the vicinity of Calbuco_A reaches tropopause height (~15km, see below) on this day (17:31-45 UTC). To explain the ~20 K drop of BT and the maximum BT ~287 K (the surface temperature from GANAL was 293.5 K), optically thin ash plume comprised of small and felsic particles was necessary. In our retrieval calculations, similar smallest values of RMS were obtained at different VAC top heights, $h_{top} = 5 - 7$ km and $h_{top} > 10$ km. We showed the results of BTS (Fig. 13c) and the VAC parameters (Table 3) when the estimated top height $h_{top} \sim 13$ km to be closer to the true values. The same information was added in the text (L794-798).

We agree with the reviewer's last comments. In this work, we introduced Calbuco_A and Calbuco_B separately because the features of the measured BT spectra were too different to explain by a single RI model, and it was a problem to clarify the reason. We tried to explain Calbuco_A measurements taking into account the ice cloud and/ or sulfate contamination. However, the results were unsuccessful. In the revised manuscript, some sentences were added on this issue (L812-815).



34) Kirishimayama and Nishinoshima results: again, no comparison with literature to assess the obtained plume properties; those 2 eruptions lead to very good results in terms of RMS, with almost all RI leading to a RMS below or around 1K; therefore I am again not so convinced that the "optimal" is a "True optimum" and that we can really draw conclusions on the optimal RI based on these results; basically almost all RIs would be acceptable.

⇒ In the revised manuscript, we tried not to use the word "optimal" a lot, and the figures of BT spectra were also revised to see easier. Figure 9 was added to shows that some RI models can derive similar fitting results.
For Nishinoshima, some sentences were replaced because we found a report of chemical analysis for the ash particles of the same eruption.

35) Kelud results: no comparison with literature; the RMS are overall higher with a mean between 1.131 and 3.182K and here maybe the "optimal" RMS makes sense because differences are larger, but on the other hand statistics on only 4 pixels are doubtful.

⇒ A sentence was added and mentioned that statistical discussions are difficult for Kelud and

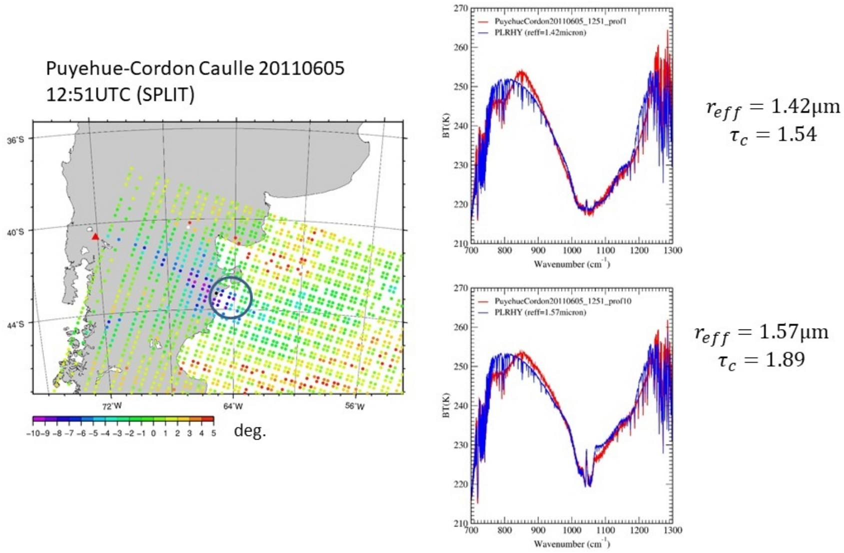
Kirishimayama because of the small number of pixels available for the analysis (L469-470).

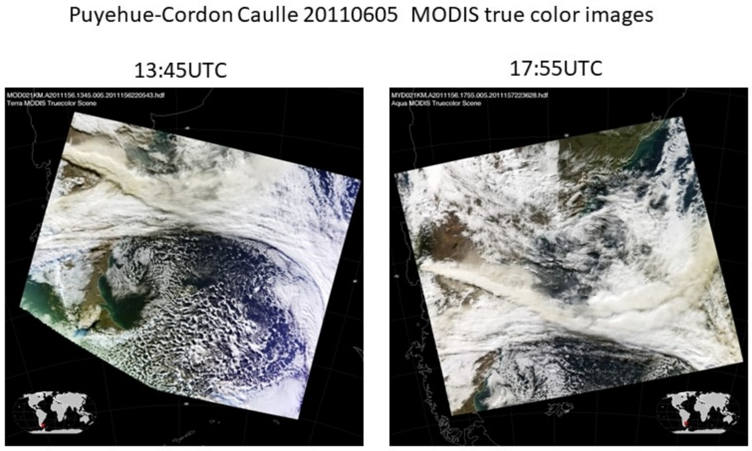
36) Puyehue results: the mean RMS is very high (1.989 to 7.08K) and the displayed spectra show that clearly something is wrong/missing in the model (atmosphere or RT), as for Calbuco A; again I am skeptical about the 0.33 optical thickness for 30K to 50K radiance drop. Such a small optical thickness in the plume centre does not match literature (e.g. the cited Klüser 2013, or Maes et al, doi: 10.3390/rs8020103, or Bignami et al, 2014 doi: 10.1109/JSTARS.2014.2320638). In addition, the retrieved particle size here is very small and does not match the literature at all (e.g. Bignami 2014 mention a particle size of 4 to 5 μ m). This needs discussion.

⇒ As shown in the red lines in Figures 16d and 16g, the IASI measurements of 30K to 50K radiance drops are true. As well as Calbuco_A cases, simulations of these BT spectra were relatively difficult even using 21 RI models, and felsic ash particles with small r_{eff} show the smallest RMS. Although the relatively large RMS were due to the large spectral variation of V-shape in the measured BTS as mentioned at the response No. 2, some spectral discrepancies in wavenumber $700-800\ \mathrm{cm}^{-1}$ and $1070-1200\ \mathrm{cm}^{-1}$ were confirmed.

We examined the BTS simulations with reference to Bignami et al. (2014) for the ash plumes of large negative ΔBT_{split} pixels at lat.=S43-S44 and lon.=W65-W66 on 5 June 2011 by using obsidian model [Pollack et al. (1973) rhyolite: PLRHY]. It was difficult to simulate the measured BTSs by using $r_{eff}=4-5\ \mu\mathrm{m}$ particles. In the text of Bignami et al. (2014), they denotes “ash parameters are computed considering only the first 300km from the vent …” (P2792). The difference of the estimated particle sizes may be due to the difference of the distances from the vent as well as the difference of the measurements discussed by Maes et al. (2016). On the other hand, the BTS were relatively well simulated by using $r_{eff}{\sim}1.5\ \mu\mathrm{m}$ ash particles (figures below), and it was consistent with the assumption of “medium” ash particles by Maes et al. (2016). However, large discrepancies appeared in wavenumber $750-900\ \mathrm{cm}^{-1}$ and $1250-1400\ \mathrm{cm}^{-1}$. These discrepancies come from the absorption properties of the PLRHY model at $700-800\ \mathrm{cm}^{-1}$ and $1200-1400\ \mathrm{cm}^{-1}$ (k value in Fig. 5).

In our analysis, we did not use the data on 5 June 2011 because we confirmed ice/water clouds below the ash plumes on this day by MODIS images (see below). In addition, we point out that Maes et al. (2016) conclude the mixing of ice clouds to explain the feature of the measured BTS at $800-900\ \mathrm{cm}^{-1}$ (Figure 7 in their paper) by using an index of “ice slope”. We think it is difficult to simulate the measured feature by ice clouds mixing.





37) Results for the 4 remaining eruptions: I am not sure that there are enough pixels in the analysis to be relevant.

⇒ We eliminated the results for the 4 remaining eruptions in the revised version.

38) Overall the conclusions should be rewritten in light of other comments and changes to the manuscript.

⇒ The manuscript was revised as a whole and some sentences were added to follow the reviewer (L1064-1067, L1088, L1114-1115).

Minor Comments

39) Everywhere: please use BT as abbreviation for Brightness Temperature, not TB (mostly in equations)

⇒ TB was replaced to BT in the revised manuscript.

40) I find it slightly confusing to call the ash plume a “cloud”, although the difference is clearly made with Meteo Clouds. I would suggest the use of “plume”, and maybe also of “aerosols” instead of “material” when referring to atmospheric ash. At least, I would avoid “cloud parameters of the VAC” and just use “VAC parameters”. However, this is just a suggestion and if the authors feel that they really prefer the terms they first selected, I do not object.

⇒ We thank the reviewer. The words are revised whole in the text.

41) Line 22: “volcanic silicates” … here I would remove “volcanic”, because all silicates have that feature (also mineral dust, for example)

⇒ It was revised (L24).

42) Line 28: “other infrared channels”: this is unclear to me, what do you mean?

⇒ The sentence was modified to be clear (L29).

43) Lines 28 and 29: the 2 lists of references are certainly not exhaustive, maybe use “e.g.” ?

⇒ We reduced the number of references (L30-31).

44) Line 97: providing CPU time without the system on which it was run has very low significance

⇒ We deleted the sentences regarding CPU times because it was not important in the context of this paper.

45) Kelut -> Kelud

⇒ The word was replaced.

46) In the conclusions, I would define again the acronyms (and the RI data sets, shortly) so that people can read abstract and conclusions only and still understand. But we may also leave this call to the Editor / Typesetting crew.

⇒ The sentences were modified to follow the reviewer (L1053-1062).