Response to Referee #2

The authors would like to thank the referee for her/his review. Below are our responses to the comments brought up by the referee. Referee's comments and our replies are marked in blue and in black, respectively. In italic are the changes made in the manuscript.

General comments

The paper describes a method for detection of unusual IASI spectra based on the noise normalised (PC) reconstruction residual. Each channel of the noise normalised reconstruction residual is a linear function of the radiances. 11 channels corresponding to absorption lines of selected gases are chosen for the detection (in the appendix a larger set of channels (ranges) is presented but the relation and role of each of these channel sets is not clear). While, by construction, the method is well suited to identify unusual spectra, the allocation to specific molecules can cause false detections as illustrated in the paper with the (false) HNO3 detection in a volcanic plume.

The set of 11 spectral intervals corresponding to absorption lines of selected gases were defined to identify the potential presence of molecules involved in an exceptional event. However in this work, only one spectral channel per molecule associated with the strongest absorption peak was used for the detection. For clarity concern, we added in Table 1 the spectral ranges and the infrared absorption peaks associated with the 11 species and we removed Table A1.

Specific comments

Section 3.1 needs some corrections. N is defined as the instrument noise covariance matrix, which is consistent with its use in equation 1, but not in the following equations, where it should be the matrix square root of the instrument noise. Line 102: this is not a projection (look up the definition). Line 107: "conservatively"?? Line 115: "optimal number" optimal in what sense? Line 97: is it really necessary to give a formula for the covariance matrix, especially since this formula is not the best way to actually compute it.

Section 3.1 has been modified in order to correct and improve the description of the methodology. Section 3.1 of the revised manuscript is as follows:

3.1 Basic concepts

The PCA method for high spectral resolution sounders, such as IASI, is described in Atkinson et al. (2008). This method is well suited to efficiently represent the amount of information contained in the 8641 IASI channels. It relies on the use of a dataset of thousands of spectra representing the full range of atmospheric conditions from which the principal components are calculated, the so-called "training database".
One considers an ensemble Y of n IASI radiance spectra $\boldsymbol{y}$ of dimension m (where m is the number of channels and n is the number of observations). Let denote $\boldsymbol{N}^{-1}\overline{\boldsymbol{y}}$ the mean and $\boldsymbol{S}_{\epsilon}\,(m \times m)$ the covariance of the normalized ensemble of spectra $\boldsymbol{N}^{-1}Y$. $\boldsymbol{N}$ is the noise normalisation matrix and is defined as the square root of $\boldsymbol{S}_{\boldsymbol{y}}(m \times m)$ the instrument noise covariance matrix associated to the IASI spectra.
The PCA is based on the eigen decomposition of the matrix $\boldsymbol{S}_{\epsilon}$:

$$\boldsymbol{S}_{\epsilon} = \boldsymbol{E}\,\boldsymbol{\Lambda}\,\boldsymbol{E}^{T} \qquad (1)$$

where $\boldsymbol{E}$ is the matrix m x m of eigenvectors and $\boldsymbol{\Lambda}$ the diagonal matrix of their associated eigenvalues.
The representation of a measured spectrum $\boldsymbol{y}$ in the eigenspace $\boldsymbol{E}$ is obtained by:

$$\boldsymbol{p} = \boldsymbol{E}^{T}\boldsymbol{N}^{-1}(\boldsymbol{y} - \overline{\boldsymbol{y}}) \qquad (2)$$

$\boldsymbol{p}$ (dimension m) is the vector of the principal component scores.

The analysis consists in representing the multidimensional IASI spectra in a lower dimensional space, which accounts for most of the variance seen in the data. This space is spanned by a truncated set of the eigenvectors of the data covariance matrix. By noise-normalizing the spectra prior to the application of the PCA, the ability to fit the data is enhanced by avoiding giving too much weight to variance caused by noise. Giving m the number of most significant eigenvectors of* $\boldsymbol{S}_{\epsilon}$, *one can represent the spectrum in the eigenspace by a truncated vector of principal component scores,* $\boldsymbol{p}^*$ *of rank* m^* ($m^* < m$). $\boldsymbol{p}^*$ *is thus a compressed representation of* $\boldsymbol{y}$. *The reconstructed spectrum,* $\tilde{\boldsymbol{y}}$ *(dimension* m*) is given by:*

$$\tilde{\boldsymbol{y}} = \bar{\boldsymbol{y}} + \boldsymbol{N}\boldsymbol{E}^*\boldsymbol{p}^* \quad (3)$$

where $\boldsymbol{E}^*$ *is the matrix of the* m^* *first eigenvectors or principal components. We define the noise normalized residual vector* $\boldsymbol{r}$ *(dimension* m*) of the reconstruction by:*

$$\boldsymbol{r} = \boldsymbol{N}^{-1}(\boldsymbol{y} - \tilde{\boldsymbol{y}}) \quad (4)$$

By definition, if m^* *is taken equal to* m, $\tilde{\boldsymbol{y}} = \boldsymbol{y}$ *and the residual is the null vector. In nominal cases if the truncation rank is carefully chosen,* $\boldsymbol{r}$ *essentially contains noise. Several techniques exist to estimate* m^* *in order to keep the essential part of the atmospheric signal and to remove the eigenvectors containing mainly the measurement noise (e.g., Antonelli et al. (2004), Atkinson et al. (2010)).*
In the following the noise normalized residual, which is calculated for each IASI IFOV, is called IFOV-residual.

The small size (120000) of the training database is problematic because the computation of the 8461*8461 covariance matrix will be affected by instrument noise as well as unusual spectra (which can be hard to avoid).

We added in Section 3.2 some explanations on the size of the training database, as well as some additional references:

The training set includes spectra observed over different types of atmospheric/surface conditions at different scan angles and for different pixel numbers to ensure that a truncated set of eigenvectors can be adequately used to represent any observed spectrum. Additionally, if the training set is too small, the specific outcome of the random noise will not be sufficiently uncorrelated and uniform, and will therefore have an influence on the computed eigenvectors and eigenvalues. Extensive experience on IASI spectra from EUMETSAT (Hultberg, 2009, https://www.eumetsat.int/media/8306) and additional experiments with different dataset sizes show that a number of about 70000 spectra is a reasonable lower limit.

Hultberg, T.: IASI Principal Component Compression (IASI PCC) FAQ, March 2009, EUMETSAT technical note, available at https://www.eumetsat.int/media/8306, 2009.

There is no evidence presented for the usefulness of the thinning towards the poles “in order to not over-represent high latitudes”.

For a sake of clarity, the sentence has been changed to:
To avoid over-representing high latitudes, because of the large swath of IASI (~2200 km) and frequent overpasses over this area with the polar orbiting satellites, the following method was applied:

Line 131: then? And how can a random selection help to “represent all the conditions” – should be better to keep all.

This comment was taken into account and the following paragraphs were added in Section 3.2 of the revised manuscript:

To avoid over-representing high latitudes, because of the large swath of IASI (~2200 km) and frequent overpasses over this area with the polar orbiting satellites, the following method was applied:

To reach a sufficient but reasonable number of IASI spectra/IFOVs (1.3 10^6 spectra per day, 4.7 10^8 per year), 120000 IFOVs for year 2013 were randomly chosen to represent all atmospheric/surface situations (air masses, land/sea, day/night, clear/cloudy) and acquisition conditions (IASI scan mirror position and pixel number).

The reference to EUMETSAT extensive work with IASI to build the training database based on random selection has also been added to this Section 3.2 (Hultberg, 2009).

Hultberg, T.: IASI Principal Component Compression (IASI PCC) FAQ, March 2009, EUMETSAT technical note, available at https://www.eumetsat.int/media/8306, 2009.

Section 3.3: 20? Actually 1 PC is enough to “depict (sic!) most of the atmospheric variability”. While 150 PC is a good choice this is not related to any percentage of the total variance but rather the signal to noise ratio of the remaining PCs.

Section 3.3. has been improved and additional information has been provided to explain and justify the choice for the truncation as follows:

Several techniques exist to estimate m in order to keep the essential part of the atmospheric signal and to remove the eigenvectors containing mainly the measurement noise. Antonelli et al (2004) define a criterium based on the spectral RMS reconstruction residuals, finding the optimal truncation rank when this value approach the spectral RMS of the instrument noise. Other methods test directly the behavior of the reconstruction score* $\sqrt{\frac{1}{m}\sum_{i=1}^{m} r_i^2}$ *as a function of the truncation rank, by looking at the second derivative of the reconstruction score as a function of the truncation rank (e.g., Hultberg, 2009) or by plotting the principal reconstruction score spatial correlation as a function of eigenvector rank (Atkinson et al., 2009). In this study, the choice of is based on the analysis of the eigenvalues. They quantify the variability explained by the corresponding eigenvectors, and the optimal number of eigenvectors needed to reproduce the signal in the raw radiances can be determined by analyzing the magnitude and behavior of the eigenvalues (sorted in descending order). In the present implementation of the PCA method we process the full IASI spectrum and use a simple method for selecting the truncation rank. The plot of the eigenvalues was examined and PCs were selected up to the point where the slope of the curve stabilized. This leads to choose the first 150 eigenvectors as done in Atkinson et al., 2010. Sensitivity tests has been performed to test the impact of using different values (from 120 to 250) on the reconstructed scores obtained on several atmospheric events (fires and volcanoes cases discussed in the next sections) and confirm this value.*

Section 4:

The details, especially regarding the detection thresholds, are hard to understand and should be rewritten. Also I miss justifications for the choices made.

Section 4.2 has been extensively modified and rewritten to improve the detection threshold definition as follows:

4.2 Detection thresholds

Two detection thresholds are defined in order to select 1) the granules associated with outliers only (which allows to gain computation time) and 2) the IFOV-residuals associated with reconstruction errors. For the definition of the detection thresholds, a dataset of 43000 IASI/Metop-B granules (21 500 granules for day-time and 21500 for night-time), containing outlier and regular spectra and chosen randomly on the first of each month between April 2013 and April 2021 is used. Note that this dataset differs from that generated for the principal component calculation as the detection method is applied on a granule basis. From this dataset, 21500 GMI and 21500 GMA pseudo-residuals are calculated for both day- and night-time conditions.

Figure 3 shows the statistical distribution of the largest minimum and maximum values for each of the 43000 GMI/GMA pseudo-residuals for all channels. The lower and upper limit of the blue box represents the 25th percentile and the 75th percentile in the data, respectively. The red line represents the median. The black lines represent upper adjacent value (UAV) and lower adjacent value (LAV), and the red crosses have been considered as "outliers" in a first analysis of the dataset. Using UAV and LAV as thresholds was observed to be too restrictive. After several tests, it has been decided to use the 25th percentile of the data to keep granules associated with potential outliers (F_1 threshold). All granules associated with GMA or GMI minimum and maximum values (in absolute values) larger than the 25th percentile of the datasets are then selected, avoiding to process granules without interesting anomalies.

A second threshold (F_2 threshold) was defined for each spectral channel based on the 99th percentile value of the GMI and GMA pseudo-residuals calculated from the 43000 granules (21500 for day-time conditions and 21500 for night-time conditions). This F_2 threshold is used in the processing of each granule selected after applying the F_1 threshold. It is applied only on channels of interest associated with a strong absorption of a molecule, which are identified in Table 1. For those channels, all IFOV-residuals associated with values larger than the F_2 threshold values are selected. The choice of the 99th percentile as the threshold value is the result of extensive tests performed both on the ensemble of statistically representative scenes (the 43000 granules) and on specific atmospheric situations of fires and volcanoes. It corresponds to the empirical compromise allowing 1) a reasonable rate of detection of extreme events (below 4%) for the processed scenes, 2) the minimization of false positive detections in the statistically representative scenes (false positive detections are empirically identified as spatially noisy i.e. isolated IFOVs) and 3) the unambiguous detection of well-identified fire and volcanic events. Values of the F_2 thresholds used for the channels of interest are provided in Table 1. In the detection processing, for each selected IFOV-residual the spectral channel associated with the detection (and thus the corresponding spectral interval and associated molecule as defined in Table 1) is recorded, along with the corresponding IFOV-residual value, the latitude and the longitude. This step allows to localize (IFOV latitude and longitude) and characterize (spectral position and corresponding IFOV-residual value) the outliers.

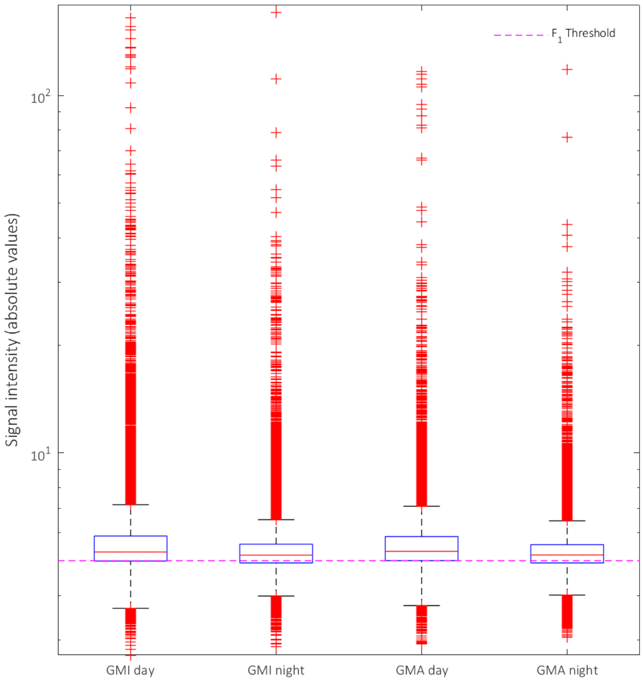

Figure 3: Distribution of normalized GMI and GMA extrema in absolute values calculated from 43000 granules (21500 for day time conditions and 21500 for night time conditions). The lower and upper limit of the blue box represents the 25th percentile and the 75th percentile in the data. The red line represents the median. The black lines represent the upper adjacent values (UAV) and lower adjacent value (LAV), and the red crosses are considered as "outliers" in the dataset. The magenta dashed line represents the F_1 threshold.

Table 1: Signal intensity thresholds (F_2) for several species for day- and night-time conditions obtained from the 99th percentile of the GMA or GMI pseudo-residuals. The thresholds are defined based on the more intense peaks associated with each molecule. Since IASI-PCA sensitivity is generally lower during night-time than during day-time, which is mainly due to thermal contrast, different thresholds for day and night conditions were defined.

Molecule	***Spectral range (cm^{-1})***	***Peak position (cm^{-1})***	***GMI day***	***GMI day***	***GMI night***	***GMA day***	***GMA night***
HCN	*711.50 - 713.50*	*712.50*	*-4.42*	*-4.42*	*-4.41*	*4.10*	*4.06*
C_2H_2	*729.25 - 730.00*	*729.50*	*-4.01*	*-4.01*	*-3.92*	*3.94*	*3.88*
C_4H_4O	*744.25 – 744.75*	*744.50*	*-4.13*	*-4.13*	*-4.10*	*3.77*	*3.76*
HONO	*790.25 – 790.75*	*790.50*	*-4.09*	*-4.09*	*-4.08*	*4.18*	*4.06*
NH_3	*966.00 - 968.00*	*967.00*	*-8.01*	*-8.01*	*-4.60*	*4.46*	*4.70*
C_2H_4	*949.00 - 950.50*	*949.25*	*-4.41*	*-4.41*	*-4.39*	*4.29*	*4.25*
CH_3OH	*1033.00 - 1033.75*	*1033.50*	*-4.35*	*-4.35*	*-4.27*	*4.40*	*4.30*
HCOOH	*1104.50 - 1105.75*	*1105.00*	*- 6.06*	*- 6.06*	*-4.69*	*4.47*	*4.26*
HNO_3	*1325.75 - 1326.25*	*1326.00*	*-6.93*	*-6.93*	*-6.43*	*6.01*	*6.38*
SO_2	*1344.50 - 1346.50*	*1345.00*	*-7.52*	*-7.52*	*-4.92*	*4.38*	*4.46*
CO	*2111.00 – 2112.25*	*2111.50*	*-6.89*	*-6.89*	*-4.72*	*4.58*	*4.28*

Why not apply thresholds for each individual spectrum, instead of the granule min and max? Faster (line 152)? No, I don't believe so.

Section 4.2 was rewritten and only two thresholds are applied: a first threshold based on the granule in order to select the granules associated with outliers only (which allows to gain computation time). A 2nd threshold is applied on each individual residual in order to select the residual associated with reconstruction errors. See above for the revised Section 4.2.

The sentence line 152 was removed.

The use of 3 different thresholds seems unnecessarily complicated. And the second threshold might be counterproductive in case most of the granule is affected by a similar anomaly.

It is true the second threshold might be counterproductive in case most of the granule is affected by a similar anomaly. However, this case is very rare.

The F_2 threshold was defined to detect outliers that we were not able to associate with species for research aspects. This part of this work is not described in the manuscript and will be the subject of future work. Since the F_2 threshold is not used in this manuscript, it was removed. See above for the revised Section 4.2.

The term "signal intensity" is used without being introduced. It is simply the (absolute value of) the (noise normalised) residual and the new term does only confuse.

We changed "signal intensity" by "noise normalized residual" everywhere.

It is not clear why a second training set was used for the computation of thresholds. Why different threshold for day and night?

The training data defined for the calculation of the principal components was generated from spectra chosen randomly and associated with no extreme events. Contrarily, the dataset defined for the calculation of the thresholds was generated from granules (because the IASI-PC-GE method is based on granules) chosen randomly and associated with all atmospheric situations, including extreme event conditions. Because of this, a second training set had to be generated for the definition of thresholds. The following sentence was added in the revised manuscript:

Note that this dataset differs from that generated for the principal component calculation as the detection method is applied on a granule basis.

The IASI signal significantly differs for day- and night-time conditions, which is mainly due to thermal contrast. This directly affects the reconstruction residual amplitude, and thus IASI-PCA sensitivity is generally lower during night-time than during day-time. Using the same threshold for day and night-time would lead to fewer detection during night-time. This is the main reason for using different thresholds for day and night conditions.

We added the following sentence in the revised manuscript:

Since IASI-PCA sensitivity is generally lower during night-time than during day-time, which is mainly due to thermal contrast, different thresholds for day and night conditions were defined.

Line 442-443 and 506-507: This is what was already discussed on page 10. I feel it is wrong to talk about "artefact" and "reconstruction error" in this context. The detection method of the paper is nothing but the identification of reconstruction error. That an unusual perturbation in a limited number of channels can affect the reconstruction residual in other channels is natural and unavoidable (as you can convince yourself by looking at a two-dimensional space with 1 retained PC). Actually maybe the biggest contribution of the paper is that is shows that this kind of "cross talk" seems to be relatively rare in practice.

We agree and would like to thank the referee for his comment. The text has been modified accordingly in Section 5.3.2 and Section 6 as follows:

The detection at ~1326 cm^{-1} *is not associated to* HNO_3 *and is due to the contribution of* SO_2 *and aerosols, as already discussed in the case of Ubinas eruption (see section 5.1.1).*

Finally, as explained above concerning SO_2 *and* HNO_3*, the spectral coincidence of some of the intense spectral features of these two species can affect the reconstruction of one when the other one is highly present. In the frame of this study, this is the only identified example of confounding situations (i.e., unusual perturbation in a limited number of channels impacts the reconstruction residual in other channels) leading to false detection. Considering the high numbers and diversity of detections and extreme situations analyzed in this work, such confounding situations are rare and PCA-based detection of atmospheric events can be effectively and efficiently exploited.*

Technical corrections

Line 19: horizontal => spatial

Done

Line 120: insert "as"

Done

Line 145: "maximum of information"?

This sentence was removed.

Line 154: noise normalised "residuals"

This sentence was removed.

Line 182: I don't understand this sentence

The sentence was removed.