Near-real time detection of unexpected atmospheric events using Principal Component Analysis on the IASI radiances

Adrien Vu Van\textsuperscript{1,2}, Anne Boynard\textsuperscript{1,2}, Pascal Prunet\textsuperscript{2}, Dominique Jolivet\textsuperscript{3}, Olivier Lezeaux\textsuperscript{2}, Patrice Henry\textsuperscript{4}, Claude Camy-Peyret\textsuperscript{5}, Lieven Clarisse\textsuperscript{6}, Bruno Franco\textsuperscript{6}, Pierre Coheur\textsuperscript{6} and Cathy Clerbaux\textsuperscript{1,6}

\textsuperscript{1}LATMOS/IPSL, Sorbonne Université, UVSQ, CNRS, Paris, 75005, France
\textsuperscript{2}SPASCIA, Ramonville-Saint-Agne, 31520, France
\textsuperscript{3}HYGEOS, Lille, 59000, France
\textsuperscript{4}CNES (Centre National d’Etudes Spatiales), Toulouse, 31400, France
\textsuperscript{5}IPSL, Institut Pierre-Simon Laplace, Paris, 75005, France
\textsuperscript{6}Université libre de Bruxelles (ULB), Spectroscopy, Quantum Chemistry and Atmospheric Remote Sensing (SQUARES), Bruxelles, 1050, Belgium

Correspondence to: Anne Boynard (anne.boynard@latmos.ipsl.fr)

Abstract. The three IASI instruments on-board the Metop family of satellites have been sounding the atmospheric composition since 2006. More than 30 atmospheric gases can be measured from the IASI radiance spectra, allowing the improvement of weather forecasting, and the monitoring of atmospheric chemistry and climate variables.

The early detection of extreme events such as fires, pollution episodes, volcanic eruptions, or industrial releases is key to take safety measures to protect inhabitants and the environment in the impacted areas. With its near real time observations and good horizontal coverage, IASI can contribute to the series of monitoring systems for the systematic and continuous detection of exceptional atmospheric events, in order to support operational decisions.

In this paper, we describe a new approach for the near real time detection and characterization of unexpected events, which relies on the principal component analysis (PCA) of IASI radiance spectra. By analyzing both the IASI raw and compressed spectra, we applied a PCA-granule based method on various past well documented extreme events such as volcanic eruptions, fires, anthropogenic pollutions and industrial accidents. We demonstrate that the method is well suited to detect spectral signatures for reactive and weakly absorbing gases, even for sporadic events. Consistent long-term records are also generated for fire and volcanic events from the available IASI/Metop-B data record.

The method is running continuously, delivering email alerts on a routine basis using the near real time IASI L1C radiance data. It is planned to be used as an online tool for the early and automatic detection of extreme events, which was not done before.

1 Introduction

Atmospheric composition is changing fast locally and globally, under natural and anthropogenic influences combined. Fire activity and local urban pollution are likely to increase in a warming climate (Hart, 2022). With their potential consequences on society and health, monitoring the events that impact atmospheric composition becomes increasingly important.
Since the end of 2006, the IASI mission has been probing the troposphere from satellite to monitor the atmospheric composition globally, onboard of 3 successive Metop satellites (Clerbaux et al., 2009). Observation records and trends are available for several infrared absorbing species, such as methane (CH₄) (García et al., 2018), carbon monoxide (CO) (George et al., 2009), ammonia (NH₃) (Van Damme et al., 2021), ozone (O₃) (Dufour et al., 2018; Wespes et al., 2019) and dust (Capelle et al., 2014; Clarisse et al., 2019). As the first goal of this mission is to feed meteorological forecast using data assimilation, radiance Level 1C (L1C) data are received in near real time, around 2-3 hours after the overpass of the satellite. This makes the detection of exceptional events possible, potentially right after they occur, such as large biomass burning fires (Turquety et al., 2009; R'Honi et al., 2013), anthropogenic pollution episodes (Boynard et al., 2014) or volcanic eruptions (Wright et al., 2022). With more than 1.2 million of radiance spectra per instrument per day, the search for local extreme events in near real time is not straightforward. A limitation is also associated with the lack of data when clouds are present in the field of view, as the usual retrieval algorithms fail to properly derive atmospheric concentrations for trace gases. Cloudy data are hence filtered.

Soon after the launch of the first IASI instrument, it has been suggested to use the principal component analysis (PCA) method to reduce data volumes by reconstructing the radiances using only the leading eigenvectors (Matricardi, 2010). This compression not only allows to heavily decrease the data volume but also to ease the data dissemination. Now available through the EUMETSAT (European organization for the exploitation of METeorological SATellites) Advanced Retransmission Service (EARS-IASI), the PCA method allows meteorological centers to directly assimilate the principal components (Collard et al., 2010; Matricardi et al., 2014; Guedj et al., 2015). It was also demonstrated that using reconstructed IASI radiance results in a substantial reduction of random instrument noise for the analysis of trace gases such as NH₃ or sulfur dioxide (SO₂) (Atkinson et al., 2010). However, it was decided to continue the distribution of the entire radiance spectra (8461 spectral channels) as one of the concerns in the use of the PCA method, for atmospheric chemistry studies, was the detection of spectral features associated with minor trace gases linked with rare events in the reconstructed spectra. Examples are volcanic eruptions, which all differ in terms of gas and type of ash emitted, and hence not enough representative cases were available in the training set. The same holds for biomass burning fires releasing different amounts of specific species depending on the type of vegetation burned. With the advent of the second and third IASI instrument together with the improvement of retrieval algorithms over time, a number of short- and long-lived trace gases were identified in the IASI spectra above or downwind from strong emission sources (Clarisse et al., 2011; De Longueville et al., 2021).

This paper describes the potential of the PCA applied on the IASI L1C (apodized radiance) data for the automatic, near real time detection and characterization of exceptional events. The paper is organized as follows: Section 2 describes the IASI instrument and the dataset used in this study. Section 3 describes the PCA method. In Section 4, an innovative approach based on the PCA method and IASI data granules is presented, which allows spectral characterization of species in near real time. In Section 5, different case studies of exceptional past events are discussed, such as volcanic, fire, and anthropogenic pollution episodes, along with industrial accidents, detected by IASI/Metop-A and -B. Finally, conclusions are given in Section 6.

2 The IASI radiance data
IASI is a Fourier transform infrared spectrometer, which records the thermal infrared (TIR) radiation emitted by the Earth and the atmosphere, between 645 cm$^{-1}$ and 2760 cm$^{-1}$, with 8461 channels sampled every 0.25 cm$^{-1}$ and a spectral resolution of 0.5 cm$^{-1}$. An example of IASI spectrum along with the absorption band of several species is illustrated in Fig. 1.

In this work IASI-A and IASI-B are used as a combined dataset. The IASI-A dataset is used for the study of events before the launch of IASI-B and for creating the PCA training database (described hereafter), and the IASI-B complete dataset is used for data after 2013 to present. The two datasets have been shown to be highly consistent with no significant drifts over time (García et al., 2016).

Each IASI instrument provides more than 1.2 million spectra per day. IASI L1C data are disseminated by EUMETSAT in 3-minute files (called “granule” hereafter) less than 3 hours after each overpass. Each granule contains 22 or 23 IASI scan lines with 120 pixels per line. With a wide swath width of ~2200 km, global observations are provided twice a day, at 9:30 AM and 9:30 PM local time. IASI has an instantaneous field of view (FOV) at nadir with a spatial resolution of 50 km x 50 km, composed of 2 x 2 circular pixels (IFOV), each corresponding to a 12 km diameter footprint on the ground at nadir (Clerbaux et al. 2009).

Figure 1: Top panel: Example of IASI spectrum. Middle and bottom panels: radiative transfer simulations for the main and weaker infrared absorbers, respectively.

The atmospheric concentrations of some species are routinely retrieved from the spectral signatures (George et al., 2009; Clarisse et al., 2011; Van Damme et al., 2013) and distributed through the AERIS database (iasi.aeris-data.fr). Some exceptional events have been studied in detail such as the 2010 Russian fires (R’honi et al., 2013), pollution in the North China Plain (Boynard et al., 2014), and SO$_2$ anthropogenic pollution (Bauduin et al., 2014, 2016).

3 The Principal Component Analysis Method
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 \( y \) of dimension \( m \) (where \( m \) is the number of channels and \( n \) is the number of observations). Let denote \( N^{-1} \bar{y} \) the mean and \( S_e (m \times m) \) the covariance of the normalized ensemble of spectra \( N^{-1}Y \). \( N \) is the noise normalisation matrix and is defined as the square root of \( S_y (m \times m) \) the instrument noise covariance matrix associated to the IASI spectra.

The PCA is based on the eigen decomposition of the matrix \( S_e \):

\[
S_e = E \Lambda E^T \tag{1}
\]

where \( E \) is the matrix \( m \times m \) of eigenvectors and \( \Lambda \) the diagonal matrix of their associated eigenvalues.

The representation of a measured spectrum \( y \) in the eigenspace \( E \) is obtained by:

\[
p = E^T N^{-1} (y - \bar{y}) \tag{2}
\]

\( p \) (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 \( S_e \), one can represent the spectrum in the eigenspace by a truncated vector of principal component scores, \( p^* \) of rank \( m^* \) (\( m^* < m \)). \( p^* \) is thus a compressed representation of \( y \). The reconstructed spectrum, \( \tilde{y} \) (dimension \( m \)) is given by:

\[
\tilde{y} = \bar{y} + NE^* p^* \tag{3}
\]

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

\[
r = N^{-1} (y - \bar{y}) \tag{4}
\]

By definition, if \( m^* \) is taken equal to \( m \), \( \tilde{y} = y \) and the residual is the null vector. In nominal cases if the truncation rank is carefully chosen, \( r \) 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.
3.2 Construction of the training database

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. For this study, around 120000 IASI/Metop-A L1C spectra were selected during a full year (which was chosen as a nominal year for avoiding excessive occurrence of extreme events such as fires and volcanoes) on the global scale. The database contains spectra associated with a good quality flag in order to only keep reliable data, acquired indifferently during the day and the night, over land and sea, and regardless of the cloud cover. For each month of the year 2013 spectra were selected every five days (1, 6, 11, 16, 21 and 26 of each month). 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:

- between 90 and 75° only one spectrum is selected
- between 75 and 60°, two spectra are selected
- between 60 and 45°, three spectra are selected
- between 45 and 30°, four spectra are selected
- between 30 and 15°, five spectra are selected
- between 15 and 0°, six spectra are selected

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).

3.3 Number of eigenvectors

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 plot the principal component score (\(p\)) spatial correlation as a function of eigenvector rank (Atkinson et al., 2009). In this study, the estimation of \(m^*\) is based on the analysis of the eigenvalues. The eigenvalues (sorted in descending order) 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 their magnitude and behavior. 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.

4 The IASI-PCA granule-extrema (GE) based method

4.1 Granule maxima and minima

The near real time detection of exceptional events is performed on the IASI granule. The choice of applying the method on the granule is convenient for the near real time aspect as it represents 3 minutes of IASI data which are received every few 1-2 hours by the antenna.

Each granule contains ~2700 radiance spectra, from which the corresponding IFOV-residuals are computed based on the IASI-PCA method. For each granule, the largest positive and negative residual value for each spectral channel is recorded in two arrays, called hereafter “Granule Maxima” (GMA) and “Granule Minima” (GMI). GMI and GMA are defined as pseudo-residuals of dimension 8461 (the number of radiance channels) and represent the spectral envelope of the statistics of residuals over the granule. Physically, the GMI (GMA) pseudo-residual is associated with reconstruction errors of spectral absorption (emission) lines. Since the method is based on the granule extrema (GMI and GMA), the method is therefore called: IASI-PCA-GE, with GE standing for Granule-Extrema. It is important to note that these pseudo-residuals associated with a granule are different from the individual IFOV-residual associated with each IFOV.

Figure 2 illustrates an example of GMA and GMI pseudo-residuals for an intense fire event that occurred in Australia on 1 January 2020. The GMI pseudo-residual (bottom panel) is characterized by detectable spectral features associated with a poor reconstruction around 700, 950, 1100 or 2100 cm\(^{-1}\). Using spectroscopic database allows to associate some of these strong peaks with contribution of different atmospheric components (see Section 5 for the identification of the molecules). Similar spectral features can be seen in the GMA pseudo-residual (top panel) albeit in emission and less intense.
Figure 2: Granule Maxima (GMA) (top) and Granule Minima (GMI) (bottom) pseudo-residuals obtained from a granule of IASI/Metop-B L1C data on 1 January 2020 over Australia.

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 (21500 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 (F1 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 (F2 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 F2 threshold is used in the processing of each granule selected after applying the F1 threshold. It is applied only on channels of interest associated with a strong absorption of a molecule, which are identified in
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.](image-url)
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.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Spectral range (cm$^{-1}$)</th>
<th>Peak position (cm$^{-1}$)</th>
<th>GMI day</th>
<th>GMI day</th>
<th>GMI night</th>
<th>GMA day</th>
<th>GMA night</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCN</td>
<td>711.50 - 713.50</td>
<td>712.50</td>
<td>-4.42</td>
<td>-4.42</td>
<td>-4.41</td>
<td>4.10</td>
<td>4.06</td>
</tr>
<tr>
<td>C$_2$H$_2$</td>
<td>729.25 - 730.00</td>
<td>729.50</td>
<td>-4.01</td>
<td>-4.01</td>
<td>-3.92</td>
<td>3.94</td>
<td>3.88</td>
</tr>
<tr>
<td>C$_4$H$_8$O</td>
<td>744.25 – 744.75</td>
<td>744.50</td>
<td>-4.13</td>
<td>-4.13</td>
<td>-4.10</td>
<td>3.77</td>
<td>3.76</td>
</tr>
<tr>
<td>HONO</td>
<td>790.25 – 790.75</td>
<td>790.50</td>
<td>-4.09</td>
<td>-4.09</td>
<td>-4.08</td>
<td>4.18</td>
<td>4.06</td>
</tr>
<tr>
<td>NH$_3$</td>
<td>966.00 - 968.00</td>
<td>967.00</td>
<td>-8.01</td>
<td>-8.01</td>
<td>-4.60</td>
<td>4.46</td>
<td>4.70</td>
</tr>
<tr>
<td>C$_2$H$_4$</td>
<td>949.00 - 950.50</td>
<td>949.25</td>
<td>-4.41</td>
<td>-4.41</td>
<td>-4.39</td>
<td>4.29</td>
<td>4.25</td>
</tr>
<tr>
<td>CH$_3$OH</td>
<td>1033.00 - 1033.75</td>
<td>1033.50</td>
<td>-4.35</td>
<td>-4.35</td>
<td>-4.27</td>
<td>4.40</td>
<td>4.30</td>
</tr>
<tr>
<td>HCOOH</td>
<td>1104.50 - 1105.75</td>
<td>1105.00</td>
<td>-6.06</td>
<td>-6.06</td>
<td>-4.69</td>
<td>4.47</td>
<td>4.26</td>
</tr>
<tr>
<td>HNO$_3$</td>
<td>1325.75 - 1326.25</td>
<td>1326.00</td>
<td>-6.93</td>
<td>-6.93</td>
<td>-6.43</td>
<td>6.01</td>
<td>6.38</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>1344.50 - 1346.50</td>
<td>1345.00</td>
<td>-7.52</td>
<td>-7.52</td>
<td>-4.92</td>
<td>4.38</td>
<td>4.46</td>
</tr>
<tr>
<td>CO</td>
<td>2111.00 – 2112.25</td>
<td>2111.50</td>
<td>-6.89</td>
<td>-6.89</td>
<td>-4.72</td>
<td>4.58</td>
<td>4.28</td>
</tr>
</tbody>
</table>

4.3 Towards a detection of extreme events in near real time

Right after the reception of each IASI 3-minutes granule, the two GMA/GMI pseudo-residuals are calculated as well as other statistics of the residual over the granule. Then the two different thresholds defined in Section 4.2 are applied to the GMA/GMI pseudo-residuals in order to localize the pixels potentially associated with an event and the associated channels. In case of anomalies (i.e., threshold overrun) in the GMA/GMI pseudo-residuals, an alert is set-up along with the targeted channels identified. The corresponding absorbing species with their spectral range are identified in the following together with the associated peak position of the associated channel, and the spatial distribution map of the detected pixels in the 3-minute granule is produced. This allows to visualize and further study exceptional events. The IASI-PCA-GE method was validated for past and documented events, four of which are described hereafter. It is now running continuously, delivering email alerts on a routine basis using the near real time IASI L1C radiance data. Most of these alerts are associated with fires and volcanic eruptions.
5 Case studies

This section presents a demonstration of the IASI-PCA-GE method for several past extreme events. The method is applied to IASI/Metop-A and the IASI/Metop-B L1C radiance data. Table 2 gives a brief description of the case studies presented hereafter.

Table 2: Brief description of the four case studies analyzed in this section.

<table>
<thead>
<tr>
<th>Type</th>
<th>Location</th>
<th>Date</th>
<th>AM/PM</th>
<th>Instrument</th>
<th>Observed molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volcanic Eruption</td>
<td>Ubinas/Peru</td>
<td>20/07/2019</td>
<td>AM</td>
<td>IASI-B</td>
<td>SO₂, HNO₃</td>
</tr>
<tr>
<td>Fires</td>
<td>Australia</td>
<td>01/01/2020</td>
<td>AM</td>
<td>IASI-B</td>
<td>HCN, C₂H₂, C₂H₄, HCOOH, CO, NH₃, C₄H₄O, CH₃OH</td>
</tr>
<tr>
<td>Anthropogenic pollution</td>
<td>China</td>
<td>13/01/2013</td>
<td>PM</td>
<td>IASI-A</td>
<td>NH₃, SO₂, CO</td>
</tr>
<tr>
<td>Industrial accident</td>
<td>Iraq</td>
<td>24/10/2016</td>
<td>PM</td>
<td>IASI-B</td>
<td>SO₂, HNO₃</td>
</tr>
</tbody>
</table>

For each event, we identify the molecules in the outliers through analysis of the residual statistic, in order to assign the spectroscopic feature characteristic of the corresponding species, over a granule and applying the IASI-PCA-GE method. We also provide distribution maps to illustrate the spatial distribution of the target event. When available, the maps are compared to the existing retrieved IASI products (CO: Hurtmans et al., 2012; NH₃: Van Damme et al., 2021; CH₃OH and HCOOH: Franco et al., 2018; C₂H₂: Franco et al., 2022; C₂H₄: as yet unpublished; HCN: Rosanka et al., 2021; SO₂: Clarisse et al., 2012).

5.1 Volcanic eruption events

Volcanic eruptions have a major impact on atmospheric composition. SO₂, which has several strong absorption bands in the TIR spectral range, is the most common molecule observed in the volcanic plume (Clarisse et al., 2012). Several other species were previously observed by satellites in volcanic eruptions such as hydrochloric acid (HCl) (Clarisse et al., 2020), hydrogen sulfide (H₂S) (Clarisse et al., 2011) and sulfuric acid (H₂SO₄) (Ackerman et al., 1994; Karagulian et al., 2010), which can be injected in the stratosphere in case of high-altitude eruption (Rose et al., 2006; Millard et al., 2006).

5.1.1 The Ubinas (Peru) case study

The IASI-PCA-GE method was applied to several volcanic eruptions. Here, we illustrate the findings for the eruption in Ubinas, Peru on 20 July 2019 (Venzke et al., 2019). Instituto Geofisico del Perú (IGP) mentioned that seismic activity suddenly increased during June 2019 and remained high during July 2019 with important ash emissions causing the evacuation of the population in some areas affected by ashfall. Figure 4 illustrates the normalized GMI pseudo-residual obtained during this volcanic eruption corresponding to a granule taken in the area of the plume during daytime. A large difference between the reconstructed spectra and raw spectra is seen in the SO₂ ν₃ band around ~1371
cm\(^{-1}\) and \(\sim\)1377 cm\(^{-1}\) which is in agreement with results of Clarisse et al. (2008, 2012) showing the sensitivity of the \(v_3\) band. Indeed, the peak found at 1371.50 cm\(^{-1}\) is associated with the presence of SO\(_2\) plume in the upper troposphere/lower stratosphere (\(\sim\)14 km, 150 hPa) between 0.5 DU and 200 DU (saturation) (Clarisse et al., 2011). Such detection is expected in this case due to the high quantity of SO\(_2\) emitted. It is worth noting that other peaks in the GMI pseudo-residual also show strong absorptions, which were initially associated with HNO\(_3\). Even if this constituent has previously been reported in volcanic plumes in some active degassing volcanoes (Mather et al., 2004), peaking in the GMI at \(\sim\)763 cm\(^{-1}\), \(\sim\)879 cm\(^{-1}\) and \(\sim\)897 cm\(^{-1}\), and \(\sim\)1326 cm\(^{-1}\) associated with \(v_8\), \(v_5\), \(2v_9\), \(v_3\) and \(v_2\) nitric acid absorption bands, respectively, it has never been observed by remote sensing before. As the analysis of the IASI HNO\(_3\) L2 products shows no HNO\(_3\) enhancement, further investigations were performed to identify where the signature comes from.

![Image](image.png)

**Figure 4:** Top: Example of GMI pseudo-residual calculated from IASI/Metop-B L1C data during a volcanic eruption in Ubinas, Peru on 20 July 2019 in the morning (AM orbit). Bottom: HITRAN spectroscopic parameters associated with the absorption of HNO\(_3\) and SO\(_2\) are shown in blue and in orange, respectively.

The HNO\(_3\) detection by the IASI-PCA-GE method was further investigated by applying the whitening method proposed by De Longueville et al. (2021). The use of a covariance matrix, calculated from a set of IASI spectra shows similar results as those found with the IASI- IASI-PCA-GE method. However, using a covariance matrix excluding the SO\(_2\) absorption band, no HNO\(_3\) spectral feature was found. This suggests that no nitric acid is present in the plume. The features found in the HNO\(_3\) absorption band by the IASI-PCA-GE method is likely related to SO\(_2\) features given that the SO\(_2\) \(v_3\) absorption band superimposes with the HNO\(_3\) \(v_3\) band.
Furthermore, other spectral signatures remain difficult to characterize in the 1200 – 1300 cm$^{-1}$ spectral domain. This spectral range corresponds to the absorption of different volcanic compounds such as ash, aerosols and other possible volcanic molecules such as H$_2$S or H$_2$SO$_4$ (Karagulian et al., 2010) but is also sensible to strong H$_2$O absorptions.

After applying the threshold filters defined in Section 4.2 to the GMI pseudo-residual, the spatial distribution of the pixels associated with outliers can be mapped. Figure 5 shows a plume of SO$_2$ (left) in Southeast America, with large signal intensity values reaching around -150 in the center of the plume. The spatial distribution of the retrieved IASI SO$_2$ L2 operational products (right) also shows the plume located in Southeast America and is in excellent agreement with the SO$_2$ plume detected from the IASI-PCA-GE method.

![Figure 5: Left: Spatial distribution of the residual values associated with SO$_2$ IASI-PCA-GE detections, using IASI/Metop-B radiance data recorded on 20 July August 2019 in the morning (AM orbit). Right: SO$_2$ total column retrievals in Dobson Units.](image)

**5.1.2 Volcanic eruption archive for IASI/Metop-B**

The time series of the SO$_2$ detections derived from the IASI-PCA-GE method is applied to IASI/Metop-B global dataset over the 2013-2022 period. Figure 6 shows the comparison of the SO$_2$ IASI-PCA-GE signal intensity with the SO$_2$ hyperspectral range indexes (HRI) product at 5 km (Bauduin et al., 2016). HRIs at 5 km are chosen because of a good sensitivity around this altitude (Clarisse et al., 2014) compared to L2 SO$_2$ concentration data that are showing concentration above 5 km (likely the high intensity volcanism). Only daily SO$_2$ extrema of both the IASI-PCA-GE method and HRI product are compared. They are spatially co-located and associated with documented volcanic events from the Global Volcanism Program, Smithsonian Institution (https://volcano.si.edu/). It is observed that both methods are able to detect not only intense eruptions but also moderate or degassing volcanic events. The largest volcanic eruptions detected during this period for both methods are Calbuco on 22 April 2015, Raikoke on 22 June 2019 and Ubinas on 19 July 2019 (Sennert, 2015, 2019, 2019b). Furthermore, for all major events (corresponding to 2810 days over 3373 days in total), an excellent correlation between HRI and IASI-PCA-GE signal intensity ($R^2 = 0.96$) is found between the two datasets.
Figure 6: Time series of SO$_2$ detections from IASI-PCA-GE method (grey) and the SO$_2$ HRI at 5 km (orange) based on the IASI/Metop-B L1C data for the 2013-2022 period. Only the daily extrema are shown in the time series.

In order to analyze and understand the differences between the two records, the correlation between the latitudes of both datasets shown in Fig. 6 are plotted (see Figure 7). An excellent location correlation between both HRI and IASI-PCA-GE methods is observed for high intensity detections. However, some discrepancies are found in case of low intensity events, corresponding to commonly active degassing volcanoes.

Some specific latitudes associated with degassing volcanoes such as Sabancaya (Moussallam et al., 2017), the Vanuatu island arc with Ambae (Ani et al., 2012), Colima and Popocatepetl in Mexico (Varley et al., 2003) and the long eruptive Kilauea volcano (Garcia et al., 2021) respectively at 15.8° S, 15.4° S, 19.5° N, 19.0° N and 19.4° N are illustrated by the black horizontal and vertical dashed lines in Fig. 7. Furthermore, some daily maxima are located around 38° S, 37.5° N et 25.2° N and are respectively related to emissions from Copahue (Reath et al., 2019), Etna (Tamburello et al., 2013 ; Ganci et al., 2012) and several Chilean volcanoes.

The daily maxima located around 56° N have been investigated and found to be associated with Kamchatka degassing volcanoes. Disperse latitudes of IASI-PCA-GE daily maxima are not consistent with the co-registered HRI maxima. These differences between the IASI-PCA-GE and HRI methods can also be explained by the relation between plume altitude/temperature not represented in the principal components that will also affect the spectral reconstruction. As a result the location of daily maxima can be different in case of low intensity detections because of the PCA overestimation (or underestimation) of atmospheric anomalies. This also results from the non-linear relationship between retrieved concentrations and PCA intensities.
It is interesting to note that both IASI-PCA-GE and HRI detections observed at around 30° N and 65° N are associated with anthropogenic emissions in the region of Sarcheshmeh Copper, one of largest industrial-mining complexes for copper that is emitting about 789.9 tons of SO$_2$ per day (Amirtaimoori et al., 2014) and over the Norilsk city, also well known for its mining and smelting industries (Bauduin et al., 2016). That finding illustrates the capacity of both methods to detect industrial emissions.

Figure 7: Comparison of latitudes corresponding to the daily maxima detected for both IASI-PCA-GE SO$_2$ signal intensity and HRI product between 2013 and 2022 with IASI-B L1C data during the day. The dashed lines show location discrepancies.

It is found that the relation between concentration and signal intensity is not linear and the PCA-based results cannot be used for an accurate quantification of SO$_2$ concentrations. Indeed, IASI-PCA-GE signals will be dependent on the molecule concentration but also on thermal contrast, and other surface parameters and atmospheric conditions. This is why discrepancies are found at high latitudes between the location of IASI-PCA-GE and HRI maxima, which are associated with eruptions in the Kamchatka region.

5.2 Fire events

Fires can be a significant source of trace gases and aerosols in the atmosphere and several species were specifically looked for in fire events: CO, NH$_3$, formic acid (HCOOH), acetylene (C$_2$H$_2$), ethylene (C$_2$H$_4$), nitrous acid (HONO), ethane (C$_2$H$_6$), acetonitrile (CH$_3$CN), methanol (CH$_3$OH), peroxyacetyl nitrate (CH$_3$CO(OONO$_2$)), hydrogen cyanide (HCN), formaldehyde (HCHO), glyoxal (CHOCHO), and CH$_4$ (Li et al., 2000; Goode et al., 2000; Sharpe et al., 2004; Coheur et al., 2009; Duflot et al., 2013; R'Honi et al., 2013; Zarzana et al., 2018, De Longueville et al., 2021). The
IASI-PCA-GE method was applied to several case studies, but only one is presented here, selected during the fire season occurring in Australia in 2019-2020.

5.2.1 The Australia case study

In Australia, fire events known as bushfires are occurring every year. Coupled with global warming and the lack of rainfall in 2019-2020, the fires were particularly intense with burned areas covering more than 186000 km$^2$. It was shown that pyro-convection allowed the plume to reach the lower stratosphere around 15-16 km (Khaykin et al., 2020). Many species were observed by ACE-FTS during that episode (e.g., Boone et al., 2020): CO, C$_2$H$_6$, C$_2$H$_2$, HCN, HCOOH, CH$_3$OH, PAN, acetone (CH$_3$COCH$_3$) and CH$_3$CN.

The IASI-PCA-GE method was applied to the IASI/Metop-B L1C data on 1 January 2020. Figure 8 illustrates an example of a normalized GMI pseudo-residual obtained during the Australia fire event. As expected, peaks relative to the CO absorption lines are found in the 2050-2200 cm$^{-1}$ spectral domain. Other peaks associated with the absorption of molecules are also visible: HCN with a peak at 712.50 cm$^{-1}$, furan (C$_4$H$_4$O) at 744.50 cm$^{-1}$, C$_2$H$_2$ at 729.50 cm$^{-1}$, C$_3$H$_4$ at 949.25 cm$^{-1}$, HCOOH at 1105.00 cm$^{-1}$ and 1777.00 cm$^{-1}$, CH$_3$OH at 1033.50 cm$^{-1}$, as well as peaks associated with NH$_3$ at 931.00 cm$^{-1}$ and 967.00 cm$^{-1}$.

![Figure 8](image-url): Top: Example of GMI pseudo-residual calculated from IASI/Metop-B L1C data during the intense fire event in Australia on 1 January 2020 in the morning (AM orbit). Bottom: HITRAN spectroscopic parameter associated with the absorption of different species are shown in colors.

Figure 9 (left column) shows the spatial distribution of the residual values associated with the detected species in the GMI pseudo-residual. Despite their different lifetimes, the plumes for the different species are located in the same region (around 180° E in the Pacific Ocean).

Carbon monoxide is retrieved in near real time (George et al., 2009) from IASI L1C and is used for monitoring fires (Turquety et al., 2009). In Fig. 9, CO is observed both with the IASI-PCA-GE and the L2 retrieval methods. However,
some discrepancies are found in terms of location and intensity. A few pixels are detected by the IASI-PCA-GE method in the Southeast of Australia, which is in agreement with the CO operational L2 product. However, the retrieval method is able to detect a larger plume over Australia compared to the IASI-PCA-GE method. Furthermore, a large plume is also detected over the Pacific Ocean but is missed by the IASI-PCA-GE method. Note that, the high intensity CO peaks are clearly detected in the residuals (c.f. Fig. 10). However, most of the missing pixels, in the PCA detection results, are located above sea. That could be due to the combination between the database chosen in the PCA method and the high variability in this spectral domain. Indeed, a higher thermal contrast variability is observed above land (Clerbaux et al., 2009), but the database contains spectra representing the natural variability without differencing sea and land pixels. As a result, the spectral reconstruction above sea with the PCA method will be less sensitive to spectral variations, causing a reduced sensitivity above sea. Furthermore, the spectral region between 2050 and 2200 cm$^{-1}$ has shown a large statistical distribution of extrema signals within the 21500 granules used for threshold calculation in Section 4.2 allowing to set a restrictive threshold for the outlier detection for CO. That restriction will also impact the number of detected pixels. The sensitivity of PCA reconstruction outliers to strong CO concentrations in fires should be more deeply investigated in further studies.

NH$_3$ is also retrieved in near real time (Van Damme et al., 2017) and observed in low concentration and occurrence above Australia on the 1st of January 2020 in the L2 retrievals and in low signal and occurrence in the IASI-PCA-GE method. Some pixels are detected by the IASI-PCA-GE method but are not spatially correlated with the NH$_3$ total column L2 data. A less frequent detection of NH$_3$ is expected since only low intensity peaks of NH$_3$ are found in the GMI pseudo-residual but two plumes are observed above both land and sea while L2 retrievals only show many isolated pixels.

However, for other indicators the size of the plume differs: large plumes are found for C$_2$H$_2$, C$_2$H$_4$ and HCOOH while smaller plumes are found for HCN, C$_4$H$_4$O and CH$_3$OH. Those differences can be explained by the difference between both methods. Indeed, the column maps includes the effects of radiative transfer (thermal contrast in particular), and the presence of clouds can also induce differences between both products as the retrievals are highly sensitive to clouds. For the IASI-PCA-GE method, the sensitivity for molecules detection highly depends on the selection of spectra to construct the database and the thresholds chosen for the detection.
Figure 9: Left: Spatial distribution of the residual values associated with CO, NH3, HCN, C2H2, C2H4, CH3OH, HCOOH and C4H4O detections from IASI/Metop-B L1C data during the intense fire event in Australia on 1 January 2020 in the morning (AM orbit); right: same as left for the total column L2 data. There is no map of C4H4O total column L2 data because there is no retrieval available.

5.2.2 Fire archive for IASI/Metop-B

Figure 10 illustrates the time series of the ethylene detections from IASI-PCA-GE method based on the IASI/Metop-B L1C data for the 2013-2022 period. C2H4 is a weak absorber often detected at 949.25 cm\(^{-1}\) in case of high intensity fires and is able to show many high intensity peaks attributed to fire events. In the figure, the most intense fires are
characterized by their location (name indicated in black in Fig. 10). The presence of fires was validated by comparing C$_2$H$_4$ detection to the IASI L2 CO that is shown to be a good fire tracker (Logan et al., 1981). The seasonality of fires clearly appears during summer in the northern hemisphere mainly related to fires in Canada, Russia and Siberia and during summer in the southern hemisphere with annual Australian and Indonesian fires. One of the largest detections of the 2013-2022 period is associated with the 2019-2020 Australian bushfires discussed in section 5.2.1. Note that the highest C$_2$H$_4$ intensity, observed on 29 July 2021 with a signal of 56, could not be associated with biomass burning as no other indicators are present in the PCA-residuals. The source of this C$_2$H$_4$ enhancement is likely linked to anthropogenic activities, as well as some other maxima, all located in Iran near the Iraq border. This will be further discussed in chapter 5.3.3.

Figure 10: Time series of C$_2$H$_4$ detections from IASI-PCA-GE method based on the IASI/Metop-B L1C data for the 2013-2022 period. Only the daily extrema are shown in the time series. For clarity, the time series are separated into 2 periods: 2013-2017 (top panel) and 2018-2022 (bottom panel). Some events (blue dots) are associated with sporadic industrial releases.

5.3 Anthropogenic pollution events

5.3.1 High pollution in China

Boynard et al. (2014) investigated a severe pollution episode occurring in the North China Plain in January 2013. The episode was caused by the presence of anthropogenic emissions combined with low wind speed and low altitude boundary layer, leading to weak mixing and dispersion of pollutants. The ability of IASI to detect high concentrations of trace gases such as CO, SO$_2$, NH$_3$ as well as ammonium sulfate aerosol ((NH$_4$)$_2$SO$_4$) during night-time was demonstrated in case of large negative thermal contrast related to the winter season and the coal burning in China for
domestic heating. The IASI-PCA-GE method was applied on 13 January 2013 during night-time. The normalized GMA pseudo-residual obtained during the China anthropogenic pollution is illustrated in Fig. 11. In order to optimize the sensitivity of the method for a low intensity event, the $F_2$ thresholds were defined as $F_2 = 5$ for both day and night-time condition for the three species of interest (CO, NH$_3$ and SO$_2$). We clearly see a signal associated with CO, NH$_3$, and SO$_2$ spectral emission, with the largest signal for SO$_2$ (value reaching $\sim$18). The detection of SO$_2$ around $\sim$1345 cm$^{-1}$ is less frequent compared to similar detection of SO$_2$ during volcanic eruptions. This result suggests that the SO$_2$ absorption features around $\sim$1345 cm$^{-1}$ also allows the detection of SO$_2$ during anthropogenic pollution episodes, which is in agreement with the finding of Bauduin et al. (2014, 2016). Finally, the spectral features around 1180-1200 cm$^{-1}$ showing a low signal intensity are likely due to the IASI detector band 1 – band 2 inter-band domain that is well captured in the IASI-PCA-GE method and should not be associated to an anomalous atmospheric constituent.

Figure 11: Top: Example of GMA pseudo-residual calculated from IASI/Metop-A L1C data during an anthropogenic pollution event occurring in China on 13 January 2013 in the evening (PM orbit). Bottom: HITRAN spectroscopic parameter associated with the absorption of different species are shown in colors.

The spatial distribution of the residual values associated with the detected species in the GMA pseudo-residual (see Fig. 11) is presented in Fig. 12 (left). The IASI-PCA-GE method allows the spectral detection of NH$_3$, SO$_2$, and CO. However only a few pixels are detected for NH$_3$, which is due to the very low (<5) signal intensity found for that species. We see the same behavior for CO. However, a clear SO$_2$ plume characterized by a signal reaching $\sim$18 (at 1345.00 cm$^{-1}$ - see Fig. 11) is found by the IASI-PCA-GE method.

Figure 12 (right) illustrates the spatial distribution of NH$_3$ and CO total column and SO$_2$ plume altitude L2 data retrieved from the IASI/Metop-A L1C data (Clarisse et al., 2012). The retrieval and IASI-PCA-GE methods shows
different patterns. We clearly see two plumes for SO$_2$ plume altitude and CO concentrations, but only few pixels of detection are found for NH$_3$.

Figure 12: Analysis of intense fire event in China on 13 January 2013 in the evening (PM orbit) based on IASI/Metop-A L1C data. Left plots: spatial distribution of residual values associated with SO$_2$, CO and NH$_3$. Right plot: SO$_2$ plume altitude retrievals (km), and CO and NH$_3$ total column retrievals (molec.cm$^{-2}$).

5.3.2 SO$_2$ released by a sulfur plant

During the period extending from 20 October to 27 October 2016, a sulfur mine burnt in d’Al-Mishraq near Mosul, Iraq. This fire on the sulfur plant, which was set by Islamic state, caused a large emission of SO$_2$ and other sulfured species in the atmosphere, which was observed from several satellite instruments (Björnham et al., 2017). Similar plant fires occurred in June 2003 during four weeks with approximately 600 kt of SO$_2$ emitted (Carn et al., 2004). This was a major health hazard (Baird et al., 2012). Nearly thousand people were intoxicated due to toxic fire plumes, and two Iraqis died.

Figure 13 illustrates the normalized GMI pseudo-residual obtained during the Iraq industrial disaster on 24 October 2016 PM. The GMI pseudo-residual is characterized by an absorption peak at $\sim$1326.00 cm$^{-1}$ that could be assigned to HNO$_3$ and two absorption peaks associated with SO$_2$ at 1345.00 cm$^{-1}$ and 1371.00 cm$^{-1}$. The signal intensity is about 14 for SO$_2$ which suggests that the event is of low to medium intensity. However, the SO$_2$ peaks found around $\sim$1371 cm$^{-1}$ and $\sim$1377 cm$^{-1}$ are mostly seen in case of intense volcanic eruptions, suggesting that the SO$_2$ concentrations are
larger than concentrations found above most of degassing volcanoes. This suggestion for an industrial origin is well supported by Fig. 14 showing SO$_2$ total columns up to 5 DU.

The detection at $\sim$1326 cm$^{-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).

Figure 13: Top: Example of GMI pseudo-residual calculated from IASI/Metop-B L1C data during a sulfur plant fire event occurring in Iraq on 24 October 2016 in the evening (PM orbit). Bottom: HITRAN spectroscopic parameter associated with the absorption of different species are shown in colors.

The spatial distribution of the residual values associated with SO$_2$ detections is illustrated in Fig. 15. The IASI-PCA-GE method allows the spectral detection of this molecule in the region of interest four days after the fire started showing the transport of the plume on the east part of the country. Less pixels are detected by the IASI-PCA-GE method than by the L2 retrieval method. This can be explained by the fact that SO$_2$ thresholds associated with the IASI-PCA-GE method were empirically chosen to minimize false positive detections, and thus the detections of low intensity residuals can be missed.
Figure 14: Analysis of sulfur plant fire event in Iraq on 24 October 2016 in the evening (PM orbit) based on IASI/Metop-A L1C data. Left plot: spatial distribution of residual values associated with SO$_2$. Right plot: SO$_2$ total column in Dobson Unit.

5.3.3 C$_2$H$_4$ sporadic emission at the border of Iran/Iraq

In Section 5.2.2 we reported that the IASI-PCA-GE method is well suited to detect biomass burning by using the C$_2$H$_4$ indicator, found in conjunction with other signatures of molecules usually associated with fire activity. Among the events that we detected, on a few occasions, we found intense signatures in the Iran/Iraq region with no other absorption than C$_2$H$_4$, which suggests that sources other than biomass burning – likely due to anthropogenic activities – are at play. The main event occurred in July 2021 and some other weaker ones are also identified in Fig. 10. By averaging IASI data over time and using a super-sampling technique, Franco et al. (2022) uncovered and identified over 300 worldwide emitters of C$_2$H$_4$, emanating from petrochemical clusters, steel plants, coal-related industries, and megacities. However, no C$_2$H$_4$ point source was formally identified in this Iran/Iraq region. But the method described in this paper is well suited to also detect sporadic events, which contrasts with the continuous emissions identified by Franco et al. (2022). Indeed, oversampling methods are well suited for the detection of regular, even weak, anthropogenic sources, but typically miss transient sources lasting for less than 24 hours. A new analysis was therefore performed on the events spotted by the IASI-PCA-GE method, which led to the identification of plumes lasting for only a few hours (see Fig. 15), for specific days as identified on Fig. 10. Although visible satellite imagery and independent online information indicate the presence of oil and gas activities in that area, no firm identification was possible, and further investigation is needed to identify the potential sources of these sporadic emissions.

Figure 15: Analysis of acetylene sporadic emission event in Iraq on 29 July 2021 based on IASI/Metop-A L1C data. Left plot: spatial distribution of residual values associated with C$_2$H$_4$ during the morning orbit. Right plot: spatial distribution of residual values associated with C$_2$H$_4$ during the evening orbit.

6 Conclusions and perspectives
This paper presents an innovative approach, based on a PCA method applied on the IASI radiance spectra, allowing the detection and characterization of exceptional events in near real time. This new method, the IASI-PCA granule extrema (GE) method, consists in focusing on extrema calculated within a given geographical region. A statistical selection is made focusing on anomalous variability in IASI channels (detection of outliers) in order to identify the contribution of specific molecules from different types of events. The method is applied to the standard three-minute granules of IASI observations allowing the near real time detection of a series of short-lived trace gases.

Using a dataset representing the full range of atmospheric conditions, we show that the PCA method is well suited to efficiently detect outliers. The analysis of the outliers allows the identification of spectral features exceeding the natural variability of several absorbing species especially for weak absorbers, emitted during fires, volcanic, anthropogenic pollution, or industrial disaster. The method is more robust than previous retrieval methods when the spectra are cloud-contaminated.

The analysis of several case studies shows a good sensitivity of the IASI-PCA-GE method, which is able to detect weak absorbers such as SO$_2$, HCN, C$_2$H$_2$, C$_2$H$_4$, CH$_3$OH, C$_4$H$_4$O and NH$_3$. We also showed that the method is well suited to detect transient events that last only a few hours/days.

Our work shows that within a granule the negative part of residuals (GMI) contains more information than the positive part represented by the GMA. However, the latter contains relevant information in case of negative thermal contrasts, allowing the detection of specific events such as the recurrent anthropogenic pollution events occurring in China in winter.

The IASI-PCA-GE method is better suited to detect spuriously emitted species. In this study, only species associated with narrow (as Q branches of C$_2$H$_2$ and C$_2$H$_4$) spectral features have been considered. Species such as PAN, CH$_3$COOH and CH$_3$COCH$_3$ characterized by broadband absorption features are more difficult to detect with the IASI-PCA-GE method. Also, unconclusive results were obtained for CO because its variability is already well captured by a truncated reconstruction due to the high variability of this species, from background conditions (50 ppb) to highly polluted areas (4000 ppb). 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.

Overall, this paper shows the capacity of PCA detection to identify different species from an event to another, especially in case of fire events, which suggest the possibility to categorize fire events based on judicious combinations of species. The method also proves useful to derive consistent long-term records for fire and volcanic events, and data will continue to accumulate over time as the method is now routinely implemented. Further work is still needed to avoid false detections, such as those associated with HNO$_3$ which are due to the correlation between different
absorption bands for the same molecule, one of them likely interfering with SO$_2$ present in the volcanic or industrial plumes.

A first version of this method is currently running continuously, delivering email alerts on a routine basis using the near real time IASI L1C radiance data. Although the method is still being tested, it is planned to be used as an online tool for the early and systematic detection of extreme events.

**Data availability statement**

IASI L2 SO$_2$, NH$_3$ and CO data can be downloaded from the AERIS portal [https://iasi.aeris-data.fr/SO2/](https://iasi.aeris-data.fr/SO2/) (https://doi.org/10.25326/42); [https://iasi.aeris-data.fr/nh3/](https://doi.org/10.25326/10); [https://iasi.aeris-data.fr/CO/](https://doi.org/10.25326/64). The VOC retrievals are processed by Franco Bruno (bruno.franco@ulb.be) and Lieven Clarisse (lieven.clarisse@ulb.be) at ULB, and available upon request.

**Author contributions**

AB and CC defined and proposed the study as part of AVV's PhD research. AVV performed the data analysis with guidance from AB, CC, PP and CCP and generated the figures. AVV, AB and CC wrote the manuscript draft. PP, CCP, BF, PFC and LC reviewed and edited the manuscript. PP, CCP, OL and DJ designed and developed the IASI PCA code. BF, PFC and LC performed the VOC retrievals. All co-authors discussed the results and contributed to the final version of the paper.

**Competing interests**

The authors declare that they have no conflict of interest.

**Acknowledgments**

A. Vu Van acknowledges funding from SPACIA SA through an ANRT CIFRE PhD grant. IASI is a joint mission of EUMETSAT and the Centre National d’Etudes Spatiales (CNES, France). The IASI Level 1C data are distributed in near real time by EUMETSAT through the EUMETCast system distribution. The authors acknowledge the AERIS data infrastructure (https://www.aeris-data.fr) for providing access to the IASI Level 1 radiance and Level 2 concentration data used in this study, and CNES for financial support.
References


