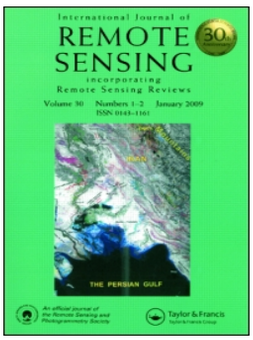


International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

# Volcanic $SO_2$ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm

Dmitry S. Efremenko , Diego G. Loyola R., Pascal Hedelt & Robert J. D. Spurr

**To cite this article:** Dmitry S. Efremenko , Diego G. Loyola R., Pascal Hedelt & Robert J. D. Spurr (2017) Volcanic $SO_2$ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm, International Journal of Remote Sensing, 38:sup1, 1-27, DOI: 10.1080/01431161.2017.1348644

**To link to this article:** http://dx.doi.org/10.1080/01431161.2017.1348644



Published online: 22 Aug 2017.

INTERNATIONAL JOURNAL OF REMOTE SENSING, 2017
VOL. 38, NO. S0, 1–27
https://doi.org/10.1080/01431161.2017.1348644

Taylor & Francis
Taylor & Francis Group



# Volcanic $SO_2$ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm

Dmitry S. Efremenko [a], Diego G. Loyola R.[a], Pascal Hedelt[a] and Robert J. D. Spurr[b]

[a]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany; [b]RT Solutions, Inc., Cambridge, MA, USA

**ABSTRACT**

Precise knowledge of the location and height of the volcanic sulphur dioxide ($SO_2$) plume is essential for accurate determination of $SO_2$ emitted by volcanic eruptions. Current $SO_2$ plume height retrieval algorithms based on ultraviolet (UV) satellite measurements are very time-consuming and therefore not suitable for near-real-time applications. In this work we present a novel method called the full-physics inverse learning machine (FP-ILM) algorithm for extremely fast and accurate retrieval of the $SO_2$ plume height. FP-ILM creates a mapping between the spectral radiance and the geophysical parameters of interest using supervised learning methods. The FP-ILM combines smart sampling methods, dimensionality reduction techniques, and various linear and non-linear regression analysis schemes based on principal component analysis and neural networks. The computationally expensive operations in FP-ILM are the radiative transfer model computations of a training dataset and the determination of the inversion operator – these operations are performed off-line. The application of the resulting inversion operator to real measurements is extremely fast since it is based on calculations of simple regression functions. Retrieval of the $SO_2$ plume height is demonstrated for the volcanic eruptions of Mt. Kasatochi (in 2008) and Eyjafjallajökull (in 2010), measured by the GOME-2 (Global Ozone Monitoring Instrument – 2) UV instrument on-board MetOp-A.

**ARTICLE HISTORY**
Received 5 April 2017
Accepted 19 June 2017



## 1. Introduction

During volcanic eruptions, sulphur dioxide ($SO_2$) and various other gases are emitted to the atmosphere. Depending on the type of eruption, ash is also emitted, and this is a major threat to local populations and aviation safety. $SO_2$ is a robust indicator for volcanic activity, and it can serve as a proxy for the emission of volcanic ash, since $SO_2$ can be collocated initially with the ash. Timely global measurement of volcanic $SO_2$ can thus provide critical information for aviation hazard mitigation, when used by the volcanic ash advisory centres (VAACs) (Carn et al. 2009). Volcanic $SO_2$ has an impact on local air quality, and it also poses a threat to aviation safety on a long timescale: $SO_2$ causes sulphidation in aircraft engines, which can lead to total engine failure if there is a

**CONTACT** Dmitry S. Efremenko dmitry.efremenko@dlr.de Remote Sensing Technology institute (IMF), German Aerospace Center (DLR), 82234, Oberpfaffenhofen, Germany

long-time exposure. Volcanic $SO_2$ and sulphate also have an effect on global climate. When released into the atmosphere they are subject to wet and dry deposition as well as oxidization to sulphate aerosols (Robock 2000). In the lower troposphere, $SO_2$ and sulphate aerosols have a lifetime of about 3 days (Lee et al. 2011; Myles, Meyers, and Robinson 2011). When injected into the stratosphere by explosive volcanic eruptions, the $SO_2$ lifetime is several weeks, whereas sulphate aerosols can reside for over a year (Forster et al. 2007; von Glasow, Bobrowski, and Kern 2009), affecting Earth's radiative forcing by reflection of solar irradiation and through induced changes in the albedos and lifetimes of clouds. For example, the 1991 Mount Pinatubo eruption released about 20 megatons of $SO_2$ into the atmosphere and caused an estimated average global cooling of about 0.3–0.5°C for several years (McCormick, Thomason, and Trepte 1995; Robock 2000).

Satellite measurements in the ultraviolet (UV) wavelength range between 305 and 330 nm provide the most sensitive measurements of $SO_2$ in the atmosphere. This allows for the detection and monitoring of volcanic and anthropogenic $SO_2$ emissions on a daily global basis (see, e.g. Fioletov et al. 2013). Backscattered UV Earthshine spectra in this wavelength range contain quantitative information about the total $SO_2$ loading and the altitude of the $SO_2$ layer (Yang et al. 2009); a change in the $SO_2$ amount causes a direct change in the optical depth, whereas a change in the altitude influences both the number of photons passing through the $SO_2$ layer and the layer absorption optical thickness.

In recent years, a number of retrieval algorithms have been developed for the interpretation of satellite-based measurements of reflected and scattered solar radiation in the infrared (IR), microwave and ultraviolet/visible (UV/vis) range. The main application of the method described in this paper is for the retrieval of $SO_2$ plume height during volcanic eruptions, based on UV measurements by nadir-viewing remote-sensing instruments such as GOME-2 (Global Ozone Monitoring Instrument – 2).

$SO_2$ total vertical columns can be retrieved easily using, for example the differential optical absorption spectroscopy (DOAS) method (see, e.g. Rix et al. 2012) by making explicit or implicit assumptions about the vertical distribution of $SO_2$; DOAS is fast enough for near-real-time (NRT) retrievals. However, retrieving plume height information is more challenging. Current plume height retrievals in the literature make use of direct fitting techniques comparing forward model computations to observations (see, e.g. Yang et al. 2009, 2010; Nowlan et al. 2011, for UV; or Clarisse et al. 2008, for IR $SO_2$ plume height retrievals). Unfortunately, these fitting methods are very time-consuming and thus not suitable for NRT applications. These limitations will be even more stringent for the upcoming generation of satellite sensors (e.g. on-board the Sentinel-5 Precursor, Sentinel 4, and Sentinel 5 platforms), having unprecedented spectral and spatial resolution and associated large increases in the amount of data. The objective of this research is to develop a fast yet accurate algorithm for the retrieval of $SO_2$ plume height from UV Earthshine measurements.

## 2. Full-physics inverse learning machines

Volcanic $SO_2$ plume height retrieval is intrinsically an ill-posed inverse problem (i.e. the retrieved values are highly sensitive to changes in the UV Earthshine measurements). In

the classical approach (Tikhonov and Arsenin 1977), the inverse problem is solved by reducing it to an exercise in optimization. The main idea behind this method is to find the state vector that minimizes the residual between simulated data and measurements. A non-linear inverse problem is solved iteratively (Rodgers 2000; Doicu, Trautmann, and Schreier 2010). Assuming an a priori state vector **x**, a non-linear forward model is linearized about **x**. Then, the linearized model can be easily inverted and a new estimation for the state vector can be found. This iterative approach is widely used for trace gases retrieval as well as for estimating aerosol and cloud properties (Efremenko et al. 2016). However, this inversion method is very time-consuming, due to repeated calls to complex radiative-transfer (RT) forward models that simulate radiances and Jacobians (i.e. matrices of the first-order partial derivatives of spectral radiances with respect to **x**), and subsequent inversion of relatively large matrices. These considerations motivate the development of alternative inversion techniques for remote sensing real-time applications.

Most machine learning algorithms do not consider the optimization problem explicitly. Rather, they *learn* from a given dataset and make predictions regarding parameters of interest. In this context, we have developed a new type of algorithm designed for solving inverse problems, called full-physics inverse learning machines (FP-ILMs). Conceptually, the FP-ILM consists of a training phase, wherein the inversion operator is obtained using synthetic data generated using a radiative transfer model (RTM, which expresses the 'full-physics' component), and an operational phase, in which the inversion operator is applied to real measurements. The main advantage of the FP-ILM over the classical optimization approach is that the time-consuming training phase involving complex RT modelling is performed off-line; the inverse operator itself is robust and computationally simple.

Figure 1 is a schematic representation of the FP-ILM used for estimating plume height values. During the training phase, a training dataset is computed using a full-physics forward model, which in our case is the RTM (see Section 2.2 for details). The RTM computes simulated radiances as a function of the following $n = 8$ input parameters: the sulphur dioxide ($SO_2$) total column density and plume height, the surface albedo, the surface height, the ozone ($O_3$) total column, the solar zenith angle (SZA), the viewing zenith angle (VZA), and the relative azimuth angle (RAA) – see Section 3 for details. The FP-ILM has the following attributes:

(1) The number of samples and the grid for input data are determined optimally by smart sampling techniques, thereby assuring the best coverage of the input and output spaces for a given number of samples (Loyola, Pedergnana, and Gimeno Garcia 2016).
(2) In order to capture the essential features of the simulated data and to avoid 'over-dimensionality' (the so-called Hughes effect, Hughes 1968), the simulated spectral data are compressed using an appropriate dimensionality-reduction technique.
(3) The mapping between the dimensionality-reduced spectral simulations and the plume height is captured via machine learning.

The design of FP-ILMs is centred around two dilemmas. The first is the trade-off between variance and bias (Raviv and Intrator 1996): efforts to decrease the bias are

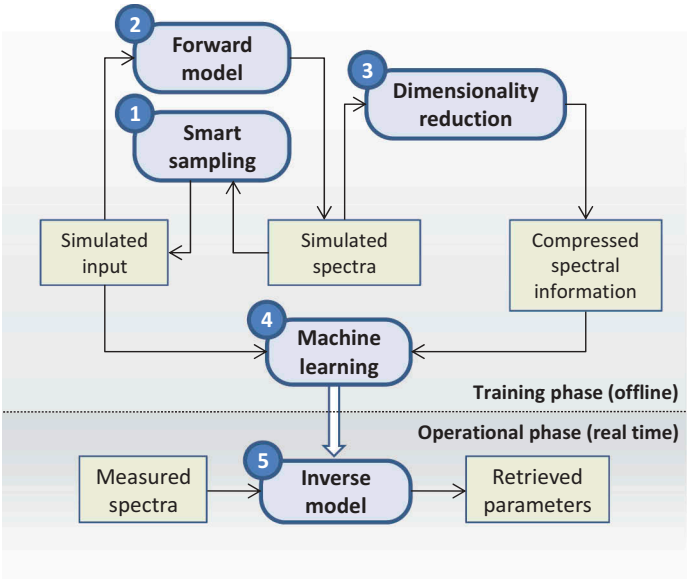



**Figure 1.** Schematic representation of the full-physics inverse learning machine algorithm. A smart sampling method provides an optimal numerical grid for the input data. Then, a forward model computes the simulated spectra for a set of simulated input data. The high-dimensional spectral data are transformed to a space of fewer dimensions using linear (e.g. principal component analysis (PCA), independent component analysis (ICA)) or non-linear (e.g. neural network, kernel PCA) mapping. Machine learning algorithms are applied to capture dependencies between compressed spectral information and the parameter to be estimated. The resulting inverse operator is then used to process the measurements.



likely to result in higher variance and vice versa. The second dilemma is the trade-off between accuracy and stability: there is a risk of over-fitting when the inverse model is affected more by random errors and noise rather than by any underlying relationships between measurements and state vector elements. In practice, it is advantageous to combine several techniques exploiting their best features. For example, in Loyola (2006) the dimensionality reduction is done using non-linear PCA, accompanied by a multi-neural network approach for solving the inverse problem. In Blackwell and Chen (2009), linear regression schemes are used to predict bias, while neural networks are used to reduce the variance. In this paper, the retrieval algorithm initially proposed by Loyola (2006) is further developed and applied to the problem of plume height retrieval.

The following sections describe in more detail the components of the FP-ILM retrieval algorithm, with subsection numbers corresponding to the five algorithm steps indicated in blue in Figure 1.

## 2.1. *Smart sampling*

The input dataset based on $n$ input variables ($n = 8$ in our case) is characterized by a discrete set of samples (points). Sampling the input $n$-dimensional space with $k$ values per dimension requires $k^n$ samples in total; this number can be very large for high-

dimensional data. Clearly, the value of $k$ should be as small as possible, without compromising the correct representation of the dataset.

Conceptually, the easiest (though the least efficient) sampling technique is based on an uniform grid along each axis of the hypercube $[0,1]^n$. More advanced algorithms rely on quasi-random sequences with uniformly distributed points in $[0,1]^n$ as $k \to \infty$. Following Weyl (1916), the uniformity of those sequences is measured using the so-called discrepancy of the sequence. Without going into details (see, e.g. Wang and Hickernell 2000, and references therein), we just remark here that the discrepancy is low if the number of points belonging to a sub-interval of $[0,1]^n$ is proportional to the measure of the sub-interval. Low discrepancy values indicate that the points are well scattered over the hypercube. Coverage using uniform grid sampling is very poor for a low number of sampling points and high dimensionality.

Smart sampling methods developed by Loyola, Pedergnana, and Gimeno Garcia (2016) are used to cover optimally the input and output spaces with a minimum number of samples. These methods are based on special sequences (named quasi-random) that share some properties of random variables. In Loyola, Pedergnana, and Gimeno Garcia (2016), different sampling methods were analysed in the context of high-dimensional regression problems. It was demonstrated that Halton series (Halton 1960) is superior for handling high-dimensional data, yet having reasonable computational time. The required number of sample points $s$ is estimated iteratively. The smart sampling and incremental function learning procedure includes the following steps:

(1) Given the expected accuracy $\varepsilon$ and confidence $\delta$, the number of samples $s$ is taken as the minimum number from sequences

$$s(n,i) = \begin{cases} i^n & \text{if } n \leq 9, \\ 10^i & \text{if } n \geq 10, \end{cases} \tag{1}$$

that satisfy the Chernoff bound

$$s \geq \frac{1}{2\varepsilon^2} \ln \frac{2}{\delta}. \tag{2}$$

(2) The input dataset with $s(n,i)$ samples for each dimension is generated using Halton series.
(3) Using the full-physics RTM, the set of radiance spectra is computed. This is by far the most time-consuming part of the sampling generation process.
(4) The parameters of the FP-ILM are estimated using regression models.
(5) The parameter $i$ for Equation (1) is increased by 1.

Steps 2–5 are repeated until the statistical moments (mean, standard deviation, skewness, and kurtosis) of the output as well as parameters of the FP-ILM converge. Details of the smart sampling technique applied to the current problem are considered in Section 3.2.

## 2.2. *Full-physics radiative transfer forward model*

For the spectral range of interest, 310–330 nm, $SO_2$ and ozone ($O_3$) are the two principal trace gas absorbers. This range covers the $O_3$ Hartley–Huggins absorption bands, while there are strong $SO_2$ absorption features towards the lower end of this window. Figure 2 (top panel) shows the natural logarithm of the total optical

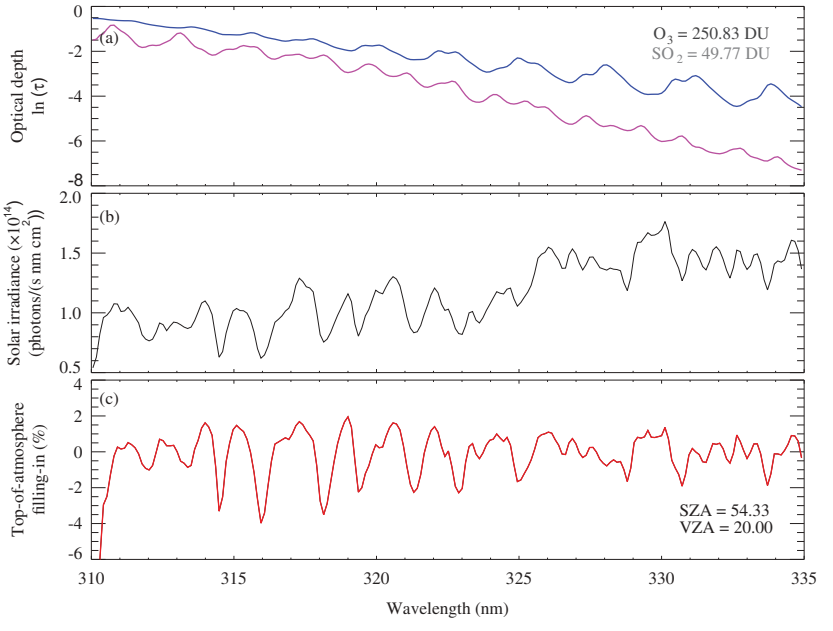



**Figure 2.** Top panel: Total optical thickness of $O_3$ and $SO_2$. Middle panel: Solar irradiance spectrum from GOME-2. Lower panel: Relative difference of a purely 'elastic' calculation (with no RRS) and a 'Raman' calculation including RRS.

thickness for $O_3$ (total column amount 250.83 DU, from a climatological profile) and $SO_2$ (total amount 49.77 DU, distributed in a Gaussian profile centred around 3.0 km, with half width 2.5 km).

Molecular scattering dominates in the UV, and the RTM must include single and multiple scattering in a multi-layer atmosphere. The solar irradiance spectrum in this wavelength range exhibits strong Fraunhofer structures (Figure 2, middle panel, is a solar spectrum from the GOME-2 instrument). Not all scattered light is 'elastic'; it is well known that Earthshine spectra show marked 'filling-in' of Fraunhofer-solar and telluric-absorber features, due to 'inelastic' (wavelength-redistributed) rotational Raman scattering (RRS) by air molecules – this filling-in phenomenon is known as the Ring effect (Grainger and Ring 1962). In our case, RRS filling is a marked interference effect, and we need an RTM that can handle it.

LIDORT-RRS (LRRS) (Spurr et al. 2008) is a discrete-ordinate RT model that has the ability to treat elastic scattering to all orders, and RRS to one order (photons are scattered just once inelastically). RRS is a small effect; about 4% of molecular scattering is inelastic (Young 1981). LRRS is based on the LIDORT models (Spurr 2008), which use discrete-ordinate methods to solve the RT equations in each layer of an optically stratified atmosphere. RRS is governed by quantum-mechanical rotational transitions for molecular oxygen ($O_2$) and nitrogen ($N_2$). For each excitation wavelength of interest, some light is lost when photons scatter inelastically out of this wavelength, and some light is gained for those photons that are scattered into this wavelength. In the UV, the redistribution range is typically ±2 nm.

LRRS is based on a perturbation approach. First, elastic-scattering RT calculations are performed at the excitation wavelength and all RRS transition wavelengths; these 'zero-order' elastic-scattering solutions are then taken as source terms for the first-order radiative transfer equation including RRS. This procedure is described in detail in Spurr et al. (2008). LRRS is slower than LIDORT, though there are procedures involving the selective grouping of rotational-Raman transitions that can speed up the calculations. The choice of Fraunhofer spectrum is very important, and it is usual to do calculations on the wavelength grid of this spectrum.

LRRS has all the capabilities of LIDORT, including treatment of solar beam attenuation in a spherical atmosphere, an exact computation of single-scattering radiation (both elastic and inelastic), the ability to model non-Lambertian surfaces, and a facility for the analytic generation of Jacobians (radiance weighting functions). Excluding RRS, the model will reproduce standard LIDORT results.

In Figure 2 (lower panel), we contrast the purely 'elastic' calculation (with no RRS) with a 'Raman' calculation including RRS, expressing this contrast as the relative difference (in per cent) between elastic and Raman radiances – this is one definition of the filling. This calculation was done for a molecular atmosphere with no aerosol scattering, with four discrete ordinates, and for solar and viewing angles in the principal plane, an albedo of 0.06, and amounts of trace species as indicated in the upper panel. It is clear that neglect of RRS can lead to very substantial radiance errors with plenty of differential structure, for example in the large Fraunhofer features over 310–315 nm.

### 2.3. *Dimensionality reduction*

Dimensionality reduction is crucial for stable and high-performance processing of spectral measurements. It excludes redundant information from the initial dataset and improves the efficiency of machine learning. This topic has been addressed frequently in the context of high-performance radiative transfer modelling (Natraj et al. 2005; Liu et al. 2006; Matricardi 2010; Efremenko et al. 2014b; Budak et al. 2015).

There is a distinction between linear and non-linear techniques for dimensionality reduction. Linear techniques embed the data into a subspace of lower dimensionality. PCA constructs a new basis-vector dataset in which the large majority of the data variance is captured. The new basis vectors are related to the eigenvectors of the covariance matrix of the initial dataset. ICA seeks linear projectors, not necessarily orthogonal to each other, though they are statistically independent.

To preserve non-linear properties of datasets and discover relationships between variables, several non-linear methods have been developed. Kernel PCA is a reformulation of conventional PCA in which the basis vectors are constructed from eigenvectors of the kernel matrix, rather than from those of the covariance matrix. Other non-linear methods are related to locally linear embedding (LLE) (Roweis and Saul 2000), in which the global non-linear structure is recovered through application of locally linear fits to optimally preserve local neighbourhood information. To find a new representation for a given ensemble of observations, artificial neural networks can be also used (Kramer 1991). More detailed review can be found in Gorban (2008), Fodor (2002), and references therein.

Linear and non-linear techniques have been inter-compared by van der Maaten, Postma, and van den Herik (2009). Results of these numerical experiments reveal that non-linear techniques perform well on selected artificial tasks. However, they hardly outperform PCA on real-world tasks. Similar conclusions were reported by Efremenko et al. (2014a), where several methods for dimensionality reduction were inter-compared in the context of accelerating radiative transfer performance. In this work, no obviously superior method emerged in the benchmarking; increasingly time-consuming and sophisticated dimensionality reduction techniques lead to more accurate results, and vice versa. Indeed, classical PCA showed good results for the 'computational time–accuracy' trade-off.

### 2.3.1. *Principal components*

Here, we make a short mathematical exposition to put the above considerations in the proper context. For clarity, we specify sizes of matrices using the notation $\in \mathbb{R}^{\text{rows}\times\text{columns}}$.

Let $\mathbf{y} = (y(\lambda_1), y(\lambda_2), \ldots, y(\lambda_W))$, $\mathbf{y} \in \mathbb{R}^{1\times W}$, be a row-vector of atmospheric radiances at $W$ wavelengths $\{\lambda_w\}_{w=1,\ldots,W}$. A set of $S$ spectra are assembled into a matrix $\mathbf{Y} \in \mathbb{R}^{S\times W}$ whose $i$th row is $\mathbf{y}_i$. Then, $\mathbf{y}_i$ can be represented in a new basis system as follows:

$$\mathbf{y}_i = \bar{\mathbf{y}} + \sum_{k=1}^{W} t_{ik}\mathbf{f}_k. \tag{3}$$

Here, $\bar{\mathbf{y}} = \frac{1}{S}\sum_{i=1}^{S}\mathbf{y}_i$, $\bar{\mathbf{y}} \in \mathbb{R}^{1\times W}$ is the sample mean of the spectra (the average spectrum), $t_{ik}$ is the $k$th coordinate of the vector $\mathbf{y}_i$ in the new basis system, and $\mathbf{f}_k = (f_k(\lambda_1), f_k(\lambda_2), \ldots, f_k(\lambda_W)) \in \mathbb{R}^{1\times W}$ is the $k$th basis vector. Noting that high-dimensional real data are often situated on or near a lower-dimensional manifold, the spectrum $\mathbf{y}_i$ can be projected onto the $K$-dimensional subspace ($K<W$) as follows:

$$\mathbf{y}_i \approx \bar{\mathbf{y}} + \sum_{k=1}^{K} t_{ik}\mathbf{f}_k, \tag{4}$$

or in matrix form for the initial dataset:

$$\mathbf{Y} \approx \bar{\mathbf{Y}} + \mathbf{TF}, \tag{5}$$

where $\bar{\mathbf{Y}} = \{\bar{\mathbf{y}}, \ldots, \bar{\mathbf{y}}\} \in \mathbb{R}^{S\times W}$, $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_K\}^{\mathrm{T}} \in \mathbb{R}^{K\times W}$, $\mathbf{T} \in \mathbb{R}^{S\times K}$ is the matrix whose entries are $\{t_{ik}\}_{i=1,\ldots,S}^{k=1,\ldots,K}$. Hereinafter, the superscript T stands for 'transform'.

The transformation (Equation (4)) can be done using dimensionality reduction techniques, such as PCA (Pearson 1901). In the latter, basic vectors $\mathbf{f}_k$ in Equation (4) are referred to as 'principal components' (PCs) or empirical orthogonal functions (EOFs), while the coordinates $t_{ik}$ in the new coordinate system and the corresponding matrix $\mathbf{T}$ are called 'principal component scores'. Considering a spectral decomposition for the covariance matrix $\mathrm{cov}(\mathbf{Y},\mathbf{Y}) \equiv \mathbf{C}_Y \in \mathbb{R}^{W\times W}$:

$$\mathbf{C}_Y = \mathbf{ELE}^{\mathrm{T}}, \tag{6}$$

where $\mathbf{E} \in \mathbb{R}^{W\times W}$ is the eigenvector matrix, and $\mathbf{L} \in \mathbb{R}^{W\times W}$ is the diagonal matrix of eigenvalues, the principal components are taken as $K$ eigenvectors related to the $K$ most significant eigenvalues. The principal component scores are computed as follows:

$$\mathbf{T} = (\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{F}^{\mathsf{T}}. \tag{7}$$

The number $K$ depends on the desired level of variance to be captured by the principal components. Several semi-empirical rules have been proposed for the optimal number of principal components (e.g. the broken stick model, MacArthur 1957). However, there is no universal rule for the selection of $K$; the choice is application-specific. In practice, $K$ can be significantly smaller than $W$ when $K$ principal components capture a significant part of the spectral radiance variability. Note that high-order principal components are related to noise rather than useful data and therefore it makes sense to neglect them.

By characterizing the set of measurements $\{\mathbf{y}_i\}_{i=1}^{S}$ with fewer parameters, we are led to more simple, stable, and computationally efficient inversion schemes. PCA has proved to be a very useful tool for retrieval of earth atmospheric constituent information from measurements. For instance, in Li et al. (2013), the PCA has been applied to the $SO_2$ total column retrieval, in the following manner. First, the set of principal components is defined for regions with no significant $SO_2$ loading. In the second step, the resulting principal components and $SO_2$ Jacobians are used to estimate $SO_2$ vertical column density. These authors reported that the correction of instrument-specific features is not necessary, since these have been already captured by principal components. In Timofeyev et al. (2003), the dimensionality reduction technique was applied to parameterize the aerosol extinction coefficient for incorporation into the inversion algorithm, while in Kataev and Lukyanov (2016) this method was used for carbon dioxide ($CO_2$) and methane ($CH_4$) retrieval.

## 2.4. *Machine learning*

### 2.4.1. *Linear regression*

Let us assume that we want to retrieve an atmospheric parameter $x$ (e.g. the volcanic $SO_2$ plume height in our case) from a measured spectrum $y = F(x, \mathbf{b})$, where $F$ is the forward model, and the vector $\boldsymbol{b}$ comprises atmospheric parameters (excepting $x$) and observation geometry. The corresponding inverse problem $x = F^{-1}(y, \mathbf{b})$ is a non-linear large-scale problem, which is usually solved by fitting the forward model to measurements at each spectral point.

The inversion would be easier if we retrieved the desired property not from the observations $y$, but from a smooth function of a single argument $x$. Such a function can be constructed by using the standard linear regression model $\mathbf{L} \in \mathbb{R}^{W\times 1}$ with regression coefficients $\{l_i\}_{i=1}^{W}$ between $x$ and the spectral data $y$:

$$x = c + \sum_{w=1}^{W} l_w y(\lambda_w), \tag{8}$$

where $c$ is the linear offset. For the vector $\mathbf{X} = \{x_1, x_2, \ldots, x_S\}^{\mathsf{T}} \in \mathbb{R}^{S\times 1}$, with, $x_1, x_2, \ldots, x_S$ the plume height values for $S$ cases, the corresponding relation in matrix form reads

$$\mathbf{X} = \mathbf{C} + \mathbf{YL}, \tag{9}$$

where $\mathbf{C} = \mathrm{diag}\{c, \ldots, c\} \in \mathbb{R}^{S\times 1}$. However, this technique can be unstable under data perturbation, and instrumental artefacts (instrument-associated errors) greatly affect the result.

Another approach, which is referred to as 'principal component regression' (PCR), employs the linear regression model between $\mathbf{X}$ and the principal component scores $\mathbf{T}$ of the spectral radiance:

$$x_i = c + \sum_{k=1}^{K} l_k t_{ik}; \tag{10}$$

$$\mathbf{X} = \mathbf{C} + (\mathbf{Y} - \bar{\mathbf{Y}})\mathbf{F}^{\mathrm{T}}\mathbf{L} = \mathbf{C} + \mathbf{TL}. \tag{11}$$

In this case $\mathbf{L} \in \mathbb{R}^{K\times 1}$. As the number of PC scores $K$ in most cases is significantly smaller than the number of spectral points $W$, the dimension of the linear regression model (and the corresponding inverse problem) is reduced. Moreover, since the instrument noise does not affect PC scores of low order, the whole inversion scheme is more stable.

For noisy data, the set of eigenvectors $\mathbf{F}$ must be computed for the matrix $\mathbf{C}_{\mathrm{Y}} + \mathbf{C}_{\mathrm{e}}$, rather than for $\mathbf{C}_{\mathrm{Y}}$, where $\mathbf{C}_{\mathrm{e}}$ is the noise covariance matrix. In this case, the PC scores for the noisy data are correlated and are therefore called 'projected principal components' (Blackwell and Chen 2009). If the statistics of the noise are unknown, the noise covariance matrix can be estimated by making some assumptions (e.g. Gaussian noise). For simplicity, we use the following approximation $\mathbf{C}_{\mathrm{e}} \approx a\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $a$ is the regularization parameter. Adding $\mathbf{C}_{\mathrm{e}}$ to $\mathbf{C}_{\mathrm{Y}}$ shifts the eigenvalues of the covariance matrix in the positive direction. This procedure reduces the impact of high-order principal components.

The kernel ridge regression (KRR) algorithm (Shawe-Taylor and Cristianini 2004) generalizes the PCR method; KRR has been used for predicting atmospheric profiles from the IASI (the Infrared Atmospheric Sounding Interferometer) instrument (Camps-Valls et al. 2012). One drawback of the PCR and KRR models is that the basis vectors $\mathbf{F}$ characterize the measurements $\mathbf{Y}$, while information contained in $\mathbf{X}$ is not taken into account. An alternative model that gets round this drawback is the partial least squares regression (PLSR) (Rosipal and Krämer 2006).

In Wentzell and Montoto (2003) and Maitra and Yan (2008), it was shown that PLSR leads to model-fitting with fewer PCs than required with PCR.

The PLSR method is based on the following linear decomposition both for $\mathbf{Y}$ and $\mathbf{X}$:

$$\mathbf{Y} \approx \breve{\mathbf{Y}}\mathbf{Q}^{\mathrm{T}}, \tag{12}$$

$$\mathbf{X} \approx \breve{\mathbf{X}}\mathbf{W}^{\mathrm{T}}, \tag{13}$$

where $\breve{\mathbf{Y}} \in \mathbb{R}^{S\times K}$ and $\breve{\mathbf{X}} \in \mathbb{R}^{S\times K}$ are the $\mathbf{Y}$-scores and $\mathbf{X}$-scores, respectively, while $\mathbf{Q}$ and $\mathbf{W}$ are the corresponding basic vectors. The $\mathbf{Q}$ and $\mathbf{W}$ vectors are constructed in such a way to maximize the correlation between $\breve{\mathbf{Y}}$ and $\breve{\mathbf{X}}$ (the correlation between $\breve{\mathbf{Y}}$ and $\breve{\mathbf{X}}$ is expected to be stronger than that between $\mathbf{T}$ (or $\mathbf{Y}$) and $\mathbf{X}$). To find new coordinates $\breve{\mathbf{Y}}$ and $\breve{\mathbf{X}}$, the eigenvalue decomposition algorithm can be applied. The first coordinate $\breve{\mathbf{y}}_1 \in \mathbb{R}^{S\times 1}$ of $\breve{\mathbf{Y}} = \left\{\breve{\mathbf{y}}_1, \ldots, \breve{\mathbf{y}}_K\right\}$

is computed as $\breve{\mathbf{y}}_1 = \mathbf{Yq}$, where $\mathbf{q}$ is the eigenvector corresponding to the first eigenvalue of $\mathbf{Y}^T\mathbf{X}\mathbf{X}^T\mathbf{Y}$. Similarly, the first $\mathbf{X}$-score $\breve{\mathbf{X}}_1 \in \mathbb{R}^{S\times 1}$ is $\breve{\mathbf{X}}_1 = \mathbf{Xw}$, where $\mathbf{w}$ is the eigenvector corresponding to the first eigenvalue of $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. Once $\breve{\mathbf{y}}_1$ and $\breve{\mathbf{X}}_1$ have been found, we repeat this procedure for matrices $\mathbf{Y}_1 = \mathbf{Y} - \breve{\mathbf{y}}_1\breve{\mathbf{y}}_1^T\mathbf{Y}$ and $\mathbf{X}_1 = \mathbf{X} - \breve{\mathbf{y}}_1\breve{\mathbf{y}}_1^T\mathbf{X}$ to find the second scores. The process is repeated recursively until all scores are found.

The PLSR approach can be generalized to the case when we are retrieving a set of correlated parameters (e.g. the temperature profile) rather than a single variable $x$. The corresponding method is then referred to as canonical correlations (Hotelling 1935). The use of canonical correlations in atmospheric sciences applications is summarized in Wilks (2011).

One final remark – we emphasize again that the training dataset of observations $\mathbf{Y}$ and the regression model are obtained using full-physics RTM simulations in the training phase. This step is very time-consuming, but it is done off-line.

### 2.4.2. *Non-linear regression*

As noted in the previous section, linear regression models find linear relations between the observations and parameters to be obtained. In practice, linearity is a very strong assumption for real data. Fortunately, linear regression appears to be a useful tool even in cases where the measurements are non-linearly related to the parameters to be retrieved. When a large number of spectral channels are considered, the non-linear dependencies can be excluded by appropriate linear combinations. For instance, in Weisz et al. (2007), temperature is retrieved from thermal-IR measurements using linear regression, although the governing forward-model relationship (the Planck function in this case) is non-linear.

However, when non-linear dependencies cannot be excluded, non-linear regression schemes must be applied for correct representation of the data. If non-linearities are specified by means of analytical expressions (e.g. in the form of a polynomial in wavelength), the linear regression framework can incorporate the corresponding non-linear functions. Then, the linear regression is used to estimate coefficients of these non-linear functions. If the non-linear dependence is not expressed analytically (as in our case) and the predictor form is not predetermined, then we have a non-parametric model, which can be determined from a given set of observations during the training process. The kernel regression or the neural network approach can be used to find the corresponding non-linear relation. Obviously, neural networks can be trained directly for a mapping from the observation space into the space of parameters to be estimated. Consequently, inversion using trained neural networks is extremely fast, as reported by Loyola (2006).

## 2.5. *Inverse model*

The inverse model comprises two steps: first, the measured spectra are transformed into a space of lower dimension, thereby providing a set of principal component scores. Second, the FP-ILM inverse operator obtained during the training phase is applied to the spectra in the reduced space. Note again that there are no RTM calculations, matrix

inversions, or iterative algorithms in the inverse model. The inverse value is computed using only straightforward matrix multiplications; the inverse model is simple and fast.

## 3. Retrieval of $SO_2$ plume height based on synthetic data

In this section, we discuss the organization of the training dataset, and we test the retrieval procedure based on the use of synthetic measurement data.

### 3.1. *RTM setup for spectra simulations*

For the training simulations, we calculate high-resolution reflectance spectra in the wavelength range from 310 to 335 nm with the LRRS RTM (see Section 2.2) using a high-resolution solar irradiance spectrum as input (Chance and Kurucz 2010). The calculations were done for a molecular-scattering atmosphere, with Rayleigh scattering parameterization taken from Bodhaine et al. (1999), and Raman spectroscopy from the literature (e.g. Chance and Spurr 1997). Eight discrete ordinates (in the polar half-space) were used in the multiple scattering treatment.

Simulations were performed in an eight-dimensional parameter space using the smart sampling algorithm (see Section 3.2). The parameters and associated ranges can be found in Table 1. $O_3$ profiles are classified according to the total column amount, month, and latitude zones as specified in the TOMS version 8 $O_3$ profile climatology (Bhartia 2003). The $SO_2$ profile has a Gaussian shape characterized by the total $SO_2$ loading and centred at a certain plume height, with a half width fixed to 2.5 km. Simulations were done on a pressure/temperature/height grid from the US standard atmosphere, with a finer-grid vertical height resolution of 0.25 km below 15 km to account for the Gaussian $SO_2$ plume shapes.

We note that fixed $O_3$ and temperature–pressure profiles might affect the results because they differ from the real profiles during the satellite measurements. However, the effect on the retrieved plume height is believed to be rather small; this is the subject of a future article.

Simulated reflectance spectra were then convolved with the GOME-2 instrument slit-function (available at: ftp://ftp.eumetsat.int/pub/EPS/out/GOME/Calibration-Data-Sets/Slit-Function-Key-Data/). To reduce non-linearity of the problem, the simulated spectra are normalized according to the quantity $\ln(\mathbf{y})\mu\mu_0(\mu+\mu_0)^{-1}$, where $\mu_0$ and $\mu$ are the cosines of VZA and SZA, respectively. This expression is based on the quasi-single scattering approximation for the radiative transfer equation (Afanas'ev, Efremenko, and Lubenchenko 2012, 2013).

**Table 1.** Physical parameters varied for the generation of reflectance spectra.

| Parameter | Range |
|---|---|
| SZA | 0–90 |
| VZA | 0–60 |
| RAA | 0–180 |
| Surface albedo | 0–1 |
| Surface height | 0–8 km |
| $O_3$ total column | 225–525 DU |
| $SO_2$ total column | 0–1000 DU |
| $SO_2$ plume height | 2.5–20 km |

### 3.2. *Smart sampling for creating the training dataset*

As noted in Section 2.1, in order to estimate the optimal number of samples for the training dataset, it is recommended to check convergence of the statistical moments of the output dataset (Romero et al. 2006). In addition to the mean and standard deviation, Loyola, Pedergnana, and Gimeno Garcia (2016) proposed to consider higher-order statistics such as skewness (measure of asymmetry) and kurtosis (measure of peakedness). Normalized values of these parameters are plotted as a function of the number of samples in Figure 3. All statistical moments converge with the number of samples larger than $3^8$. The final FP-ILM is trained using $4^8 = 65,536$ simulated spectra covering the wavelength domain $310 - 335$ nm. A mesh of input parameters is computed using Halton sequences. The dataset generated at the previous iteration with $3^8$ samples is used as a test dataset, from which the plume height is retrieved. Thanks to smart sampling, both the test and training datasets cover the full input space and are complementary, with no duplicated inputs.

### 3.3. *Dimensionality reduction of the training dataset*

PCA, ICA, LLE, and kernel-PCA with a polynomial kernel have been applied to the training dataset. The errors in spectral radiances reconstructed from a given number of principal component scores are similar for all methods. Figure 4 shows the mean and first three empirical orthogonal functions computed for the training dataset. To estimate the number of principal components used in the regression models, we investigated the dependence of the correlation coefficient ($r$) between the predicted plume height and the actual plume height on the number of principal components. In Figure 5, $r$ is plotted for PCR and PLSR models. As expected, the PLSR method delivers higher values of $r$ than those from PCR. For both methods, $r$ scales linearly for a small number of PCs. Using

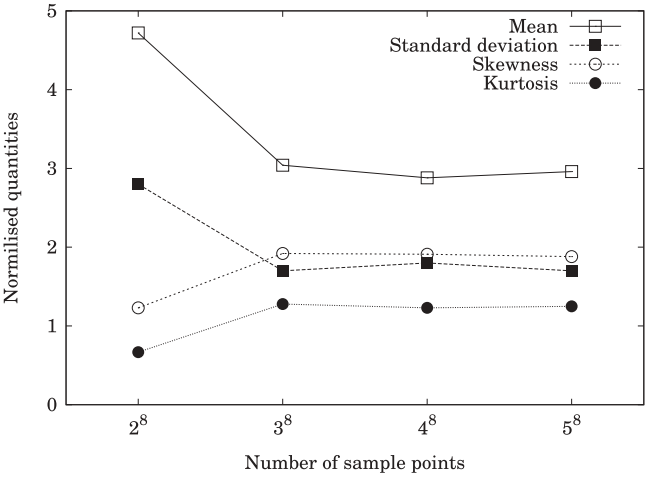



Figure 3. Statistical moments (mean, standard deviation, skewness, and kurtosis) of the radiance spectra as function of the number of sample patterns. Values have been re-scaled for convenience of display.

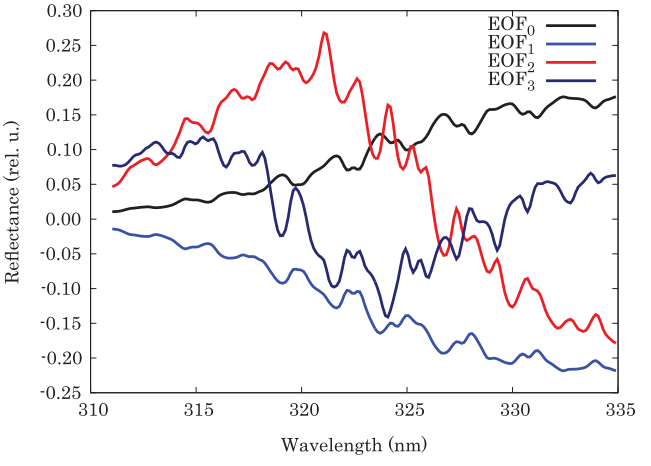



**Figure 4.** Empirical orthogonal functions computed using the training dataset. $EOF_0$ is the mean spectrum of the dataset.

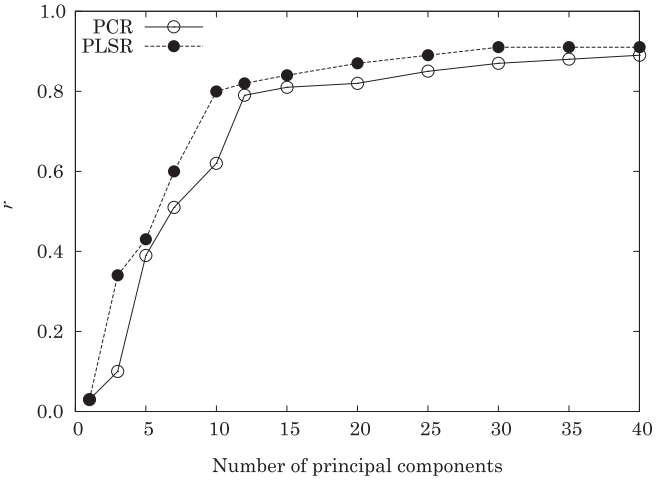



**Figure 5.** The correlation coefficient ($r$) between predicted and actual plume heights, as a function of the number of principal components used in the PCR and PLSR models.

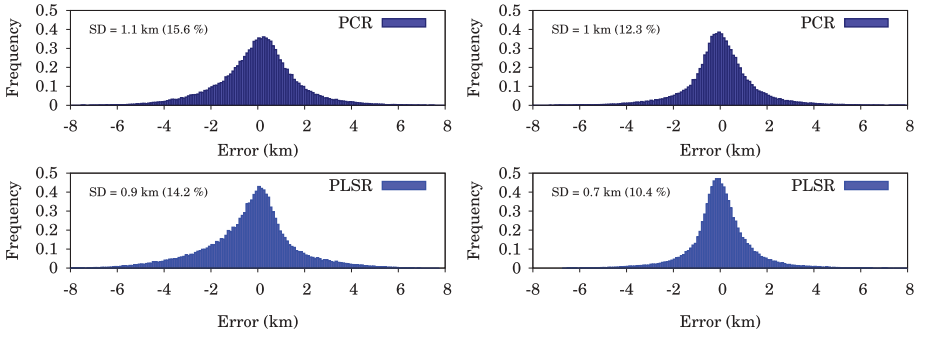


more than 15 PCs does not increase $r$ significantly. Thus, $K = 15$ principal components should be sufficient to capture the radiance information that is correlated to plume height. The difference between the maximum actual value of $r$ ($\approx 0.9$) and the theoretical maximum value (1.0) can be regarded as a measure of the error induced by using linear regression models.

## 3.4. *$SO_2$ plume height retrieval using linear inversion models*

### 3.4.1. *Retrieval from noise-free data*

In Figure 6, we show the histograms of relative errors in the prediction of the plume height using PCR and PLSR with $K = 15$ and $K = 30$ principal components, respectively.




(a) $K = 15$ (b) $K = 30$

**Figure 6.** Histogram of errors of the plume height retrieved using the PCR and the PLSR with $K = 15$ and $K = 30$ principal components. SD, standard deviation (SD for the relative error is given in brackets).

The width of the error distribution for $K = 30$ is narrower than that for $K = 15$, while the average error of the plume height retrieval for the PLSR is less than that for the PCR in both cases.

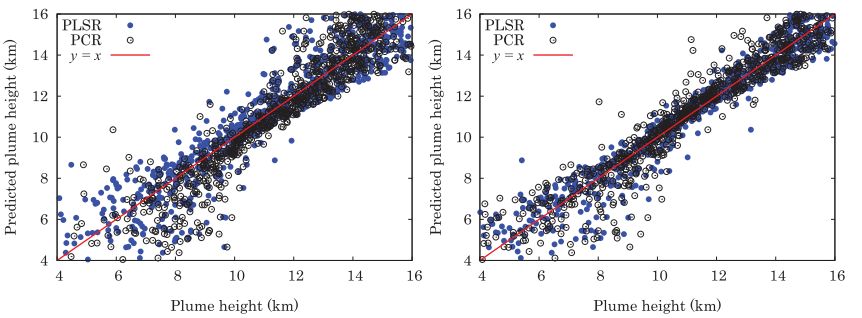
Figure 7 shows the predicted plume height values as a function of actual plume height; from the initial dataset we plot every 100th point for clarity. Points are distributed near the straight line ($y(x) = x$). The lowest errors can be found in the range of 9–13 km.

### 3.4.2. *Sensitivity to instrumental noise*

Here, we apply the PLSR and PCR models to noisy data to study the influence of noise on the retrieval results. For the signal-to-noise ratio (SNR), we use the following definition:

$$\frac{1}{\mathrm{SNR}} = \frac{\sqrt{W}\sigma}{\| \mathbf{y} \|}, \tag{14}$$

where $\sigma$ is the standard deviation and $\| \mathbf{y} \|$ is the $L^2$-norm of the spectrum.




(a) $K = 15$ (b) $K = 30$

**Figure 7.** Accuracy of the PLSR and PCR models using $K = 15$ and $K = 30$ principal components for the simulated data.

**Table 2.** Degradation of the retrieval results due to noise in the data.

| | PLSR | | PCR | |
|---|---|---|---|---|
| SNR ratio | $K = 15$ | $K = 30$ | $K = 15$ | $K = 30$ |
| | | $r$ | | |
| Noise-free | 0.84 | 0.91 | 0.81 | 0.84 |
| 1000 | 0.81 | 0.88 | 0.75 | 0.79 |
| 500 | 0.77 | 0.81 | 0.72 | 0.75 |
| 100 | 0.62 | 0.64 | 0.55 | 0.58 |
| 50 | 0.31 | 0.31 | 0.29 | 0.29 |
| | | Average relative error (%) | | |
| Noise-free | 14.1 | 10.1 | 16.6 | 14.0 |
| 1000 | 15.8 | 13.2 | 17.2 | 16.4 |
| 500 | 17.1 | 16.1 | 17.8 | 17.0 |
| 100 | 24.3 | 23.0 | 23.8 | 23.2 |
| 50 | 28.5 | 28.3 | 29.3 | 29.20 |

Note that the errors in Table 2 are averaged over all values of $SO_2$ total column in the input dataset. The plume height retrieval errors as a function of the $SO_2$ amount and the 'true' plume height are shown in Figure 8. This figure shows that the plume height can be retrieved with errors of $\pm 1$ km for high $SO_2$ total columns (>50 DU) and a plume height in the range between 6 and 18 km. This is in agreement with Yang et al. (2009) and Nowlan et al. (2011). This is a low bound for the sensitivity (the best that we can expect). In practice, plume height sensitivity will be worse. For a more comprehensive analysis, one should include errors associated with spectral calibration; this will be a topic for our future research.

Gaussian noise is added to the simulated data. The average retrieval errors and corresponding values of $r$ are shown in Table 2. Note, that $r$ decreases and the average relative error increases when the SNR decreases. For SNRs larger than 500, both PLSR and PCR models can capture the information related to the plume height, while for the SNRs less than 100, the correlation between the predicted and actual values of the plume height is low (the plume height information is hidden by noise). Unlike the situation with noise-free data, using more than 15 principal components does not lead to any improvement in retrievals with noisy data, since higher-order principal components are affected by noise.

### 3.4.3. *Sensitivity to instrument degradation*

In traditional differential optical absorption spectroscopy (DOAS) methods for retrieval of trace gas vertical columns (see, e.g. Loyola et al. 2011), the spectral data are filtered with a low-degree polynomial to improve the stability of the retrieval. This procedure is expected to reduce the uncertainty arising from imperfect instrument calibration and sensor degradation (Smedt et al. 2012) – effects that are usually strong in the UV spectral range used for the $SO_2$ column retrieval. In our simulations, we observed that filtering a linear function or a quadratic polynomial from each spectrum of the training dataset does not affect the quality of the retrieval. However, the PLSR and PCR models are still sensitive to degradation regardless of the filtering. One source of error is an imperfectly known instrumental slit function, which could suffer from temporal changes. For this paper, we used the official pre-flight GOME-2 slit-function dataset, which is certainly not equal to on-going in-flight slit-function values. Another source of error is noise increasing over time due to degradation of the instrument. In this paper, we used a fixed SNR level for training the retrieval. In a subsequent paper, we will analyse the effect of the slit function and take into account the measured noise of the spectra.

## 3.5. $SO_2$ plume height retrieval using non-linear inversion

### 3.5.1. Taking into account the $SO_2$ total column

One possible way to improve the quality of the plume height retrieval (in other words, to increase the correlation coefficient between the predicted and the actual plume height and to reduce the dispersion of the error) would be to include other input parameters in the regression scheme (such as the $SO_2$ total column). This can be done easily using neural networks. In general, neural networks can be used for establishing a non-linear mapping between a dataset of numeric inputs and a set of numeric outputs. A neural network consists of interconnected neurons, or nodes, that implement a simple, non-linear function (a sigmoid function in our case) of the inputs.

In this study, the Levenberg–Marquardt least-squares fitting algorithm is used for training the neural network. The dataset is separated into three parts:

- a training set (a set of examples used for learning);
- a validation set (a set of examples used to tune the number of hidden units or determine a stopping point for the learning process);
- a test set (a set of examples used only to assess the performance of a fully trained neural network).

Two schemes are considered. In Scheme 1, $K$=15 PC scores and the $SO_2$ total column values are used as input data, while Scheme 2 uses $K$=15 PC scores only to estimate the plume height. The average error and $r$ are given in Table 3. Clearly, Scheme 1 is superior to Scheme 2. That means that including $SO_2$ total column improves the quality of the retrieval. However, the $SO_2$ total column is unknown a priori and is estimated given the plume height values. Thus, Scheme 1 implies an iterative inversion procedure, in which the plume height and the $SO_2$ amount are to be estimated recursively. Also, noting that Scheme 2 is not substantially better than the PLSR and PCR models, we conclude that Scheme 1 is preferable to Scheme 2 for the operational algorithm.

## 3.6. $SO_2$ plume height retrieval using both linear and non-linear models

As proposed by Blackwell and Chen (2009), computational efficiency of the retrieval algorithm can be improved by separating the initial problem into linear and non-linear components. Here, the linear component is realized using linear models (e.g. PLSR), while the non-linear component is computed using neural networks. Here, we consider three neural networks:

**Table 3.** Average relative errors and correlation coefficients for the non-linear plume height inversion using $K = 15$ PCs.

| | Scheme 1 | | Scheme 2 | |
|---|---|---|---|---|
| | Average error (%) | $r$ | Average error (%) | $r$ |
| Training | 13 | 0.94 | 20 | 0.89 |
| Validation | 13 | 0.94 | 20 | 0.89 |
| Testing | 14 | 0.94 | 21 | 0.88 |

- The first neural network (NN1) predicts the plume height using 4 PC scores of the spectra;
- The second neural network (NN2) uses the plume height value predicted by the PLSR model and four PC scores of the spectra. In its turn, the PLSR model uses $K = 30$ components;
- The third neural network (NN3) predicts the plume height using 30 PC scores.

For the linear model, the regularization is implemented by adding a regularization matrix to the covariance matrix of the spectra. For neural networks, the regularization is done by training the neural network with noisy data (alternatively, error weights for input parameters can be introduced). Both methods reduce the impact of higher-order PC scores.

Table 4 lists values of $r$, the average relative errors, and training times for these three networks. The training automatically stops when the mean square errors of the validation samples start to increase.

Two conclusions can be drawn:

(1) Using PLSR prediction as an input parameter can significantly increase the accuracy of the model (72 % for NN1 against 13 % for NN2).
(2) By capturing some part of the information by the linear model and applying the neural network for the rest, we simplify the structure of the neural network. Consequently, the training time decreases significantly (12 s for NN2 against 220 s for NN3).

We conclude that the 'hybrid' model (NN2) seems to be the best in terms of computational time and accuracy.

## 4. Application to measured volcanic eruptions

In this section, the inversion operator is applied to two volcanic eruptions measured by the GOME-2 instrument.

GOME-2 is a nadir-viewing across-track scanning UV spectrometer (240–790 nm); GOME-2 (Munro et al. 2016) is currently flying on two of the three EUMETSAT polar satellites (EPS) MetOp-A (launched 2006), MetOp-B (launched 2012), and is expected to fly on MetOp-C (projected launch 2018). For these satellites, the equatorial overpass is at 9:30 h local time on the descending node of the orbit. The instrument takes regular Earthshine measurements as well as daily solar reference irradiance measurements. GOME-2 has a nominal swath of 1920 km and a corresponding ground-pixel size of 80 km × 40 km. In this configuration, global

**Table 4.** Average relative errors and correlation coefficients, various algorithms.

| Neural network type | NN1 (using 4 PCs) | | NN2 (using PLSR and 4 PCs) | | NN3 (using 30 PCs) | |
|---|---|---|---|---|---|---|
| Time of training | 10 s | | 12 s | | 220 s | |
| | Average error (%) | $r$ | Average error (%) | $r$ | Average error (%) | $r$ |
| Training | 72 | 0.63 | 13 | 0.94 | 10 | 0.98 |
| Validation | 72 | 0.63 | 13 | 0.94 | 10 | 0.98 |
| Testing | 72 | 0.62 | 14 | 0.94 | 11 | 0.96 |

coverage is achieved within 1.5 days. Since 2012, GOME-2A has operated in narrow-swath mode with a reduced ground-pixel size of 40 km × 40 km. Currently, data from both GOME-2 instruments allow for daily global coverage of atmospheric trace-gas concentrations (see, e.g. Hassinen et al. 2016). For this work, however, we investigate only volcanic eruptions measured by GOME-2 aboard MetOp-A.

The inversion operator is applied to selected GOME-2 pixels with volcanic $SO_2$ plume signatures. For pixel selection, we make use of the volcanic activity flag provided by the GDP4.x operational products (Loyola et al. 2011; Hassinen et al. 2016). The volcanic activity detection algorithm is based on the SACS warning algorithm (Brenot et al. 2014) to discriminate between real plumes and background noise. For each flagged pixel, reflectance spectra are calculated from the GOME-2 Level-1 solar irradiance and Earthshine radiance spectra in the wavelength range 315–330 nm. Wavelength calibration for the solar irradiance and Earthshine spectra is performed by applying a traditional DOAS retrieval using a high-resolution solar reference spectrum (Chance and Kurucz 2010) to correct for Doppler shifts. During the retrieval, small shifts and stretches of the Earthshine spectrum as a function of wavelength across the detector array are fitted. These shift and stretch values are used to generate a wavelength-calibrated Earthshine spectrum. In this way, both the solar and Earth radiance spectra are calibrated and an appropriate reflectance spectrum determined – this is then the starting point for retrieving the $SO_2$ plume height.

The plume height values are retrieved using a hybrid algorithm incorporating the PLSR model with 10 PC scores and a neural network using 4 PC scores. Based on the retrieved plume height, total vertical $SO_2$ columns (VCDs) are calculated in a two-step approach: first, for each measurement, the $SO_2$ slant-column density (SCD) is retrieved using DOAS optical density fitting to appropriate $O_3$ and $SO_2$ cross-sections in the wavelength range 315–325 nm, and including a Ring-effect reference spectrum for RRS filling – details can be found in Rix et al. (2012). We note that for high $SO_2$ amounts, which are common to most strong volcanic eruptions, the DOAS assumption of optically thin absorption no longer holds in the range 315–325 nm. In order to compensate for this, a multi-window scheme as described in Hörmann et al. (2013) was applied to retrieve the $SO_2$ slant columns. In addition to the main window, additional fitting regions of 325–335 nm and 360–390 nm are used for moderate and very high $SO_2$ amounts, respectively. A simple threshold criterion based on retrieved SCDs will select the appropriate fit window (see Theys et al. 2017).

Second, in order to convert retrieved SCDs to final VCD values, single-wavelength air mass factors (AMFs) for each viewing and solar geometries, the surface reflection conditions, the total $O_3$ column retrieved from GOME-2 (Loyola et al. 2011), as well as the retrieved $SO_2$ plume height from the PLSR model.

In the following sections, we present results of the plume height retrieval and associated $SO_2$ VCDs for two volcanic eruptions measured by GOME-2 on MetOp-A.

### 4.1. *Eyjafjallajökull, Iceland, May 2010*

With the May 2010 eruption of the Eyjafjallajökull volcano in Iceland, large amounts of ash and gases such as water vapour and $SO_2$ were emitted into the atmosphere. Carried by winds aloft, the volcanic ash cloud resulted in the closure of airports throughout the UK and Scandinavia, and later in the rest of northern and western Europe. The volcano

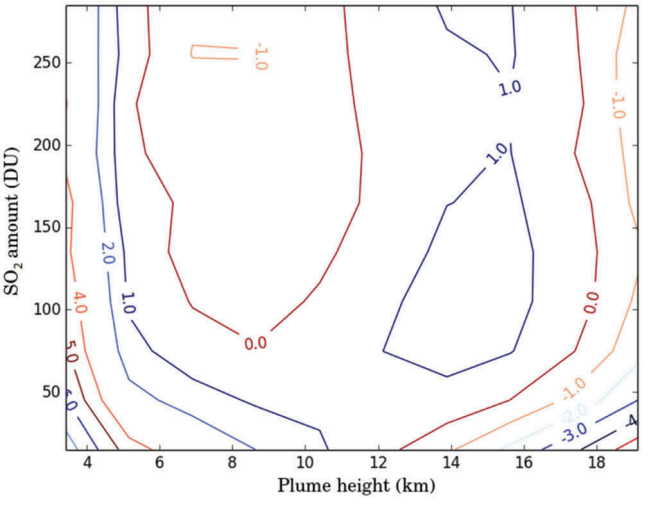



**Figure 8.** Absolute error of the retrieved plume height using the PCR model with $K = 15$ principal components as a function of $SO_2$ total column and plume height.

erupted on 20 March 2010 for the first time since 1821, and again on 14 April. The $SO_2$ plume was clearly detectable by GOME-2 for about 4 weeks after 22 April, with a maximum vertical column of not more than 15 DU. For this work, we have analysed the 5th May 2010 GOME-2 measurements, which show a strong $SO_2$ plume over Great Britain. The retrieved plume heights using the FP-ILM are in the range 6–9 km (see Figure 9), which is in good agreement with results found by Rix et al. (2012) and Flemming and Inness (2013) (both 6–10 km) based on GOME-2 UV data, and Carboni et al. (2015) (4–8 km) based on IASI IR data. Nevertheless the associated errors on the plume height are expected to be very high according to Figure 8, since the retrieved $SO_2$ total column was generally lower than 15 DU.

### 4.2. *Kasatochi, Aleutian Islands, August 2008*

Kasatochi is a stratovolcano located in the Alaskan Aleutian Island chain; it erupted on 7 August 2008 after a period of increased seismicity, with three distinctive explosive eruptions. About 2 Mt of $SO_2$ were released into the stratosphere, together with large amounts of ash, which however fell out of the atmosphere very quickly (see Yang et al. 2010). The $SO_2$ plume was transported over the entire Northern Hemisphere and was detectable by GOME-2 for several weeks.

The plume height retrieval using the FP-ILM detects the $SO_2$ plume at an altitude in the range 9–10 km, with some parts of the plume even reaching altitudes of 14 km in those parts of the plume with very high $SO_2$ amounts. Associated errors on the plume height are expected to be lower, according to Figure 8. Figure 10 shows the retrieved plume height (left panels) and the recalculated VCDs (right panels) for the first three days after the eruption, that is from 8 August (upper row) through 10 August (lower row). Although the ash loading was very high for the first few days after the eruption, we note that this had very little effect on the retrieved plume heights, which are consistent for all days.

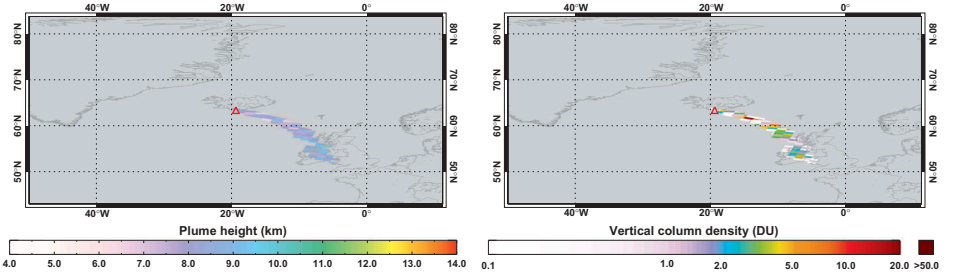



**Figure 9.** $SO_2$ plume of Eyjafjallajökull detected by GOME-2a on 5 May 2010. The left panel shows the retrieved plume heights, the right column the recalculated VCDs based on these retrieved plume heights.

The results are in very good agreement with those obtained by Yang et al. (2010) using a direct fitting approach (see Figure 6 therein); these authors also found that the bulk of the plume was situated at heights of 9–11 km. In their paper, they also compared their retrievals with CALIOP (cloud-aerosol lidar with orthogonal polarization) measurements of scattering particles and found a slightly higher plume height for scattering particles (e.g. aeresols) than their retrieved plume height. Also, the results of Nowlan et al. (2011) showed that most of the plume was situated in the altitude range of 9–10 km, with parts of the plume around 15 km (see Figure 9 therein).

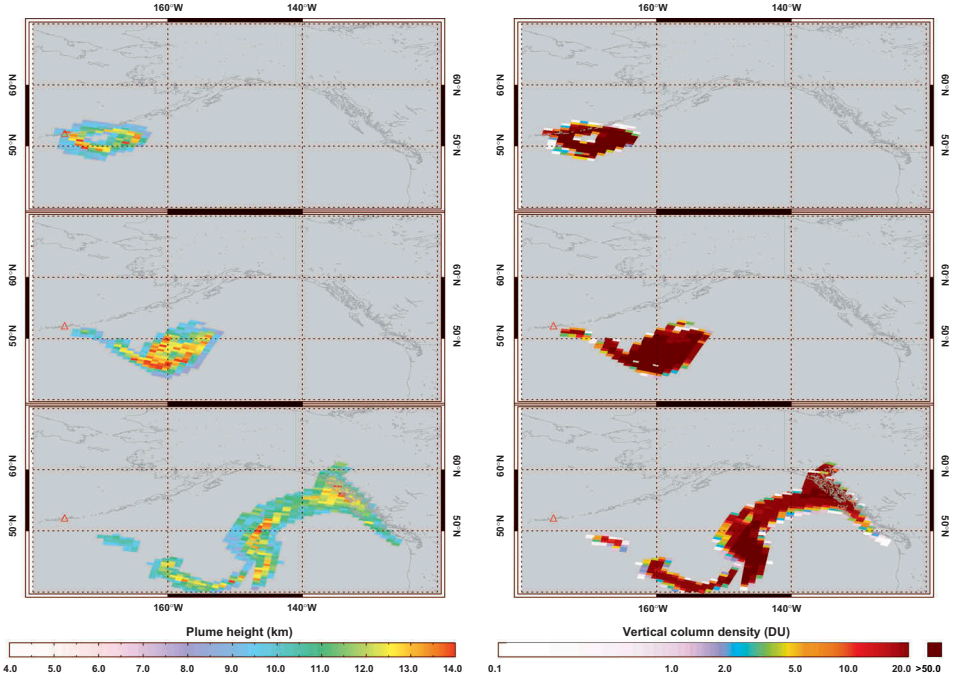



**Figure 10.** $SO_2$ plumes from Kasatochi measured by GOME-2A on 8 August (upper row), 9 August (middle row), and 10 August 2008 (lower row). The left panels show the retrieved plume heights, and the right panels the recalculated VCDs based on these retrieved plume heights.

## 5. Conclusions

In this paper, we have described the FP-ILM model for fast and accurate retrieval of $SO_2$ plume height. The FP-ILM has two phases – training and operational. The training phase involves:

- Smart sampling for generating the training dataset of simulated spectra; a set that covers optimally the function to invert and at the same time minimise the number of calls to the RTM;
- dimensionality reduction for mitigating the effects of instrumental noise and improving the speed and stability of the inversion algorithms of the spectral radiance dataset;
- linear and non-linear regression models for obtaining the inversion operator.

Smart sampling reduces the total number of samples required to reproduce adequately any statistical dependencies in the training datasets. To reduce the number of predictors in the regression model and improve the stability of the algorithm, dimensionality reduction methods were applied to the spectral radiance dataset. We have considered the principal component regression (PCR) and the PLSR. Unlike PCR, PLSR computes a new set of basis vectors to maximize the correlation between input and output datasets. In our numerical tests, the PLSR model appeared to be more efficient in terms of accuracy than the PCR model, although the difference between them is not dramatic.

The regression analysis is performed in two steps. First, the linear regression model (PCR or PLSR) is applied to capture general dependencies between spectral radiances and plume height information. Second, we account for non-linear dependencies by using neural networks. One important aspect is that, as a result of extracting 'linear' dependencies by using PLSR, the neural network structure can be simplified, and the training time significantly reduced thereby. For our application, we showed that such a hybrid approach is superior to linear or non-linear regression alone. Regarding the computational time, we note that the training phase is performed off-line, while the resulting inverse operator is computationally simple.

The FP-ILM has been applied to the GOME-2 measurements for two volcanic eruptions. Results are in good agreement with the literature. However, for the FP-ILM, additional work still needs to be done to characterize properly the errors in plume height retrieval and sensitivity. In particular, errors associated with aerosols have not so far been considered. Note that the influence of aerosols on the spectrum is smooth and similar to the effect of ground albedo, in contrast to the $SO_2$ absorption-based signatures. It follows that the variability of spectral radiances due to aerosols can be (at least partially) compensated for through changes in the surface albedo. With that in mind, the FP-ILM designed for plume height retrieval should be more sensitive to differential absorption signatures rather than to the continuum signatures associated with aerosols.

The second paper is anticipated which will follow up on some of the issues raised in this paper. In particular, we plan to apply the FP-ILM to other sensors and more volcanic eruptions. The effect of temperature–pressure profiles and errors associated with spectral calibration and slit function on the retrieved plume height will be assessed.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie [Grant number 0703/89373/15/2013] and DLR programmatic [S5P KTR 2 472 046].

## ORCID

Dmitry S. Efremenko http://orcid.org/0000-0002-7449-5072

## References

Afanas'ev, V. P., D. S. Efremenko, and A. V. Lubenchenko. 2013. "On the Application of the Invariant Embedding Method and the Radiative Transfer Equation Codes for Surface State Analysis." *Light Scattering Reviews* 8: 363–423. Springer Science + Business Media. doi:10.1007/978-3-642-32106-1_8.

Afanas'ev, V. P., D. S. Efremenko, and A. V. Lubenchenko. 2012. "Determining the Applicability Boundaries of Small-Angle Approximation Solutions to the Radiative Transfer Equation for Elastic Peak Electron Spectroscopy." *Bulletin of the Russian Academy of Sciences: Physics* 76 (5): 565–569. doi:10.3103/S1062873812050036.

Bhartia, P. 2003. "Algorithm Theoretical Baseline Document, TOMS V8 Total Ozone Algorithm." Technical Report. Greenbelt, MD: NASA. http://toms.gsfc.nasa.gov/version8/version8_update.html

Blackwell, W. J., and F. W. Chen. 2009. *Neural Networks in Atmospheric Remote Sensing*. Lexington: Massachusetts Institute of Technology, Lincoln Laboratory.

Bodhaine, B. A., N. B. Wood, E. G. Dutton, and J. R. Slusser. 1999. "On Rayleigh Optical Depth Calculations." *Journal of Atmospheric and Oceanic Technology* 16 (11): 1854–1861. doi:10.1175/1520-0426(1999)0162.0.CO;2.

Brenot, H., N. Theys, L. Clarisse, J. Van Geffen, J. Van Gent, M. Van Roozendael, A. R. van der, et al. 2014. "Support to Aviation Control Service (SACS): An Online Service for Near-Real-Time Satellite Monitoring of Volcanic Plumes." *Natural Hazards and Earth System Sciences* 14: 1099–1123. doi:10.5194/nhess-14-1099-2014.

Budak, V. P., G. A. Kaloshin, O. V. Shagalov, and V. S. Zheltov. 2015. "Numerical Modeling of the Radiative Transfer in a Turbid Medium Using the Synthetic Iteration." *Optics Express* 23 (27): A829. doi:10.1364/OE.23.00A829.

Camps-Valls, G., J. Munoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet. 2012. "Nonlinear Statistical Retrieval of Atmospheric Profiles from MetOp-IASI and MTG-IRS Infrared Sounding Data." *IEEE Transactions on Geoscience and Remote Sensing* 50 (5): 1759–1769. doi:10.1109/TGRS.2011.2168963.

Carboni, E., R. G. Grainger, T. A. Mather, D. M. Pyle, G. Thomas, R. Siddans, A. Smith, A. Dudhia, M. L. Koukouli, and D. Balis. 2015. "The Vertical Distribution of Volcanic $SO_2$ Plumes Measured by IASI." *Atmospheric Chemistry & Physics Discussions* 15: 24643–24693. doi:10.5194/acpd-15-24643-2015.

Carn, S. A., A. J. Krueger, N. A. Krotkov, K. Yang, and K. Evans. 2009. "Tracking Volcanic Sulfur Dioxide Clouds for Aviation Hazard Mitigation." *Natural Hazards* 51 (2): 325–343. doi:10.1007/s11069-008-9228-4.

Chance, K., and R. L. Kurucz. 2010. "An Improved High-Resolution Solar Reference Spectrum for Earth's Atmosphere Measurements in the Ultraviolet, Visible, and near Infrared." *Journal of Quantitative Spectroscopy & Radiative Transfer* 111: 1289–1295. doi:10.1016/j.jqsrt.2010.01.036.

Chance, K. V., and R. J. D. Spurr. 1997. "Ring Effect Studies: Rayleigh Scattering, Including Molecular Parameters for Rotational Raman Scattering, and the Fraunhofer Spectrum." *Applied Optics* 36 (21): 5224–5230. doi:10.1364/AO.36.005224.

Clarisse, L., P. F. Coheur, A. J. Prata, D. Hurtmans, A. Razavi, T. Phulpin, J. Hadji-Lazaro, and C. Clerbaux. 2008. "Tracking and Quantifying Volcanic $SO_2$ with IASI, the September 2007 Eruption at Jebel at Tair." *Atmospheric Chemistry & Physics* 8: 7723–7734. doi:10.5194/acp-8-7723-2008.

Doicu, A., T. Trautmann, and F. Schreier. 2010. *Numerical Regularization for Atmospheric Inverse Problems*. Springer. doi:10.1007/978-3-642-05439-6.

Efremenko, D. S., A. Doicu, D. Loyola, and T. Trautmann. 2014a. "Optical Property Dimensionality Reduction Techniques for Accelerated Radiative Transfer Performance: Application to Remote Sensing Total Ozone Retrievals." *Journal Quant Spectroscopic Radiation Transfer* 133: 128–135. doi:10.1016/j.jqsrt.2013.07.023.

Efremenko, D. S., D. G. Loyola, A. Doicu, and R. J. D. Spurr. 2014b. "Multi-core-CPU and GPU-accelerated Radiative Transfer Models Based on the Discrete Ordinate Method."." *Computer Physics Communications* 185 (12): 3079–3089. doi:10.1016/j.cpc.2014.07.018.

Efremenko, D. S., O. Schüssler, A. Doicu, and D. Loyola. 2016. "A Stochastic Cloud Model for Cloud and Ozone Retrievals from UV Measurements." *Journal Quant Spectroscopic Radiation Transfer* 184: 167–179. doi:10.1016/j.jqsrt.2016.07.008.

Fioletov, V. E., C. A. McLinden, N. Krotkov, K. Yang, D. G. Loyola, P. Valks, N. Theys, et al. 2013. "Application of OMI, SCIAMACHY, and GOME-2 Satellite SO2 Retrievals for Detection of Large Emission Sources." *Journal of Geophysical Research: Atmospheres* 118 (19): 11,399–11,418. 2013JD019967. doi:10.1002/jgrd.50826.

Flemming, J., and A. Inness. 2013. "Volcanic Sulfur Dioxide Plume Forecasts Based on UV Satellite Retrievals for the 2011 Grimsvötn and the 2010 Eyjafjallajökull Eruption." *Journal of Geophysical Research: Atmospheres* 118 (17): 10,172–10,189. doi:10.1002/jgrd.50753.

Fodor, I. K. 2002. "A Survey of Dimension Reduction Techniques." Technical Report. doi:10.2172/15002155.

Forster, P., V. Ramaswamy, P. Artaxo, J. Berntsen, R. Betts, D. W. Fahey, J. Haywood, et al. 2007. *Changes in Atmospheric Constituents and in Radiative Forcing*, Chap. 2, 129–234. Cambridge; New York: Cambridge University Press. http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter2.pdf.

Gorban, A. N., Balázs Kégl, D. C. W., and A. Y. Zinovyev, eds. 2008. *Principal Manifolds for Data Visualization and Dimension Reduction*. Berlin Heidelberg: Springer. doi:10.1007/978-3-540-73750-6.

Grainger, J. F., and J. Ring. 1962. "Anomalous Fraunhofer Line Profiles." *Nature* 193 (4817): 762–762. doi:10.1038/193762a0.

Halton, J. H. 1960. "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals." *Numerical Mathematical* 2 (1): 84–90. doi:10.1007/BF01386213.

Hassinen, S., D. Balis, H. Bauer, M. Begoin, A. Delcloo, K. Eleftheratos, S. Gimeno Garcia, et al. 2016. "Overview of the O3M SAF GOME-2 Operational Atmospheric Composition and UV Radiation Data Products and Data Availability." *Atmospheric Measurement Techniques* 9 (2): 383–407. doi:10.5194/amt-9-383-2016.

Hörmann, C., H. Sihler, N. Bobrowski, S. Beirle, M. Penning De Vries, U. Platt, and T. Wagner. 2013. "Systematic Investigation of Bromine Monoxide in Volcanic Plumes from Space by Using the GOME-2 Instrument." *Atmospheric Chemistry & Physics* 13: 4749–4781. doi:10.5194/acp-13-4749-2013.

Hotelling, H. 1935. "The Most Predictable Criterion." *Journal of Educational Psychology* 26: 139–142. doi:10.1037/h0058165.

Hughes, G. 1968. "On the Mean Accuracy of Statistical Pattern Recognizers." *IEEE Transactions Informatics Theory* 14 (1): 55–63. doi:10.1109/TIT.1968.1054102.

Kataev, M. Y., and A. K. Lukyanov. 2016. "Empirical Orthogonal Functions and Its Modification in the Task of Retrieving of the Total Amount CO2 and CH4 with Help of Satellite Fourier Transform Spectrometer GOSAT (TANSO-FTS)." *22nd International Symposium on Atmospheric*

*and Ocean Optics: Atmospheric Physics*, edited by G. M. Gennadii and O. A. Romanovskii. nov. SPIE-Intl Soc Optical Eng.

Kramer, M. A. 1991. "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks." *AIChE Journal* 37 (2): 233–243. doi:10.1002/aic.690370209.

Lee, C., R. V. Martin, A. Van Donkelaar, R. R. Hanlim Lee, J. C. H. Dickerson, N. Krotkov, A. Richter, K. Vinnikov, and J. J. Schwab. 2011. "SO2 Emissions and Lifetimes: Estimates from Inverse Modeling Using in Situ and Global, Space-Based (SCIAMACHY and OMI) Observations." *Journal of Geophysical Research: Atmospheres* 116: (D6): n/a–n/a. D06304. doi:10.1029/2010JD014758.

Li, C., J. Joiner, N. A. Krotkov, and P. K. Bhartia. 2013. "A Fast and Sensitive New Satellite $SO_2$ Retrieval Algorithm Based on Principal Component Analysis: Application to the Ozone Monitoring Instrument." *Geophysical Research Letters* 40 (23): 6314–6318. doi:10.1002/2013GL058134.

Liu, X., W. L. Smith, D. K. Zhou, and A. Larar. 2006. "Principal Component-Based Radiative Transfer Model for Hyperspectral Sensors: Theoretical Concept." *Applied Optics* 45 (1): 201–208. doi:10.1364/AO.45.000201.

Loyola, D. G., M. E. Koukouli, P. Valks, D. S. Balis, N. Hao, M. Van Roozendael, R. J. D. Spurr, et al. 2011. "The GOME-2 Total Column Ozone Product: Retrieval Algorithm and Ground-Based Validation." *Journal of Geophysical Research* 116: D7. doi:10.1029/2010JD014675.

Loyola, D. G., M. Pedergnana, and S. Gimeno Garcia. 2016. "Smart Sampling and Incremental Function Learning for Very Large High Dimensional Data." *Neural Networks* 78: 75–87. doi:10.1016/j.neunet.2015.09.001.

Loyola, D. G. R. 2006. "Applications of Neural Network Methods to the Processing of Earth Observation Satellite Data." *Neural Networks* 19 (2): 168–177. doi:10.1016/j.neunet.2006.01.010.

MacArthur, R. H. 1957. "On the Relative Abundance of Bird Species." *Proceedings National Academic Sciences USA* 43 (3): 293–295. doi:10.1073/pnas.43.3.293.

Maitra, S., and J. Yan. 2008. "Principal Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression." In *Discussion Papers: 2008 Discussion Paper Program - Applying Multivariate Statistical Models*, 79–90. Casualty actuarial society. https://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf.

Matricardi, M. 2010. "A Principal Component Based Version of the RTTOV Fast Radiative Transfer Model." *Quarterly Journal of the Royal Meteorological Society* 136: 1823–1835. doi:10.1002/qj.680.

McCormick, M. P., L. W. Thomason, and C. R. Trepte. 1995. "Atmospheric Effects of the Mt Pinatubo Eruption." *Nature* 373: 399–404. doi:10.1038/373399a0.

Munro, R., R. Lang, D. Klaes, G. Poli, C. Retscher, R. Lindstrot, R. Huckle et al. 2016. "The GOME-2 Instrument on the Metop Series of Satellites: Instrument Design, Calibration, and Level 1 Data Processing – An Overview." *Atmospheric Measurement Techniques* 9 (3): 1279–1301. doi:10.5194/amt-9-1279-2016.

Myles, L., T. P. Meyers, and L. Robinson. 2011. "Relaxed Eddy Accumulation Measurements of Ammonia, Nitric Acid, Sulfur Dioxide and Particulate Sulfate Dry Deposition near Tampa, FL, USA." *Environmental Research Letters* 2. doi:10.1088/1748-9326/2/3/034004.

Natraj, V., X. Jiang, R.-L. Shia, X. Huang, J. S. Margolis, and Y. L. Yung. 2005. "Application of Principal Component Analysis to High Spectral Resolution Radiative Transfer: A Case Study of the $O_2$A Band." *Journal of Quantitative Spectroscopy & Radiative Transfer* 95: 539–556. doi:10.1016/j.jqsrt.2004.12.024.

Nowlan, C. R., X. Liu, K. Chance, Z. Cai, T. P. Kurosu, C. Lee, and R. V. Martin. 2011. "Retrievals of Sulfur Dioxide from the Global Ozone Monitoring Experiment 2 (GOME-2) Using an Optimal Estimation Approach: Algorithm and Initial Validation." *Journal of Geophysical Research: Atmospheres* 116 (D18): n/a–n/a. D18301. doi:10.1029/2011JD015808.

Pearson, K. 1901. "LIII.On Lines and Planes of Closest Fit to Systems of Points in Space." *Phil Mag* 2 (11): 559–572. doi:10.1080/14786440109462720.

Raviv, Y., and N. Intrator. 1996. "Bootstrapping with Noise: An Effective Regularization Technique." *Connection Science* 8: 355–372. doi:10.1080/095400996116811.

Rix, M., P. Valks, N. Hao, D. Loyola, H. Schlager, H. Huntrieser, A. Flemming, U. Koehler, U. Schumann, and A. Inness. 2012. "Volcanic $SO_2$, BrO and Plume Height Estimations Using

GOME-2 Satellite Measurements during the Eruption of Eyjafjallajökull in May 2010." *Journal of Geophysical Research (Atmospheres)* 117: D00U19. doi:10.1029/2011JD016718.
Robock, A. 2000. "Volcanic Eruptions and Climate." *Reviews of Geophysics* 38: 191–219. doi:10.1029/1998RG000054.
Rodgers, C. D. 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Wolrd Scientific Publishing. doi:10.1142/9789812813718.
Romero, V. J., J. V. Burkardt, M. D. Gunzburger, and J. S. Peterson. 2006. "Comparison of Pure and "Latinized" Centroidal Voronoi Tessellation against Various Other Statistical Sampling Methods." *Reliability Engineering & System Safety* 91 (10–11): 1266–1280. doi:10.1016/j.ress.2005.11.023.
Rosipal, R., and N. Krämer. 2006. "Overview and Recent Advances in Partial Least Squares." In *Subspace, Latent Structure and Feature Selection*, 34–51. Springer Science + Business Media. doi:10.1007/11752790_2.
Roweis, S. T., and L. K. Saul. 2000. "Nonlinear Dimensionality Reduction by Locally Linear Embedding." *Science* 290 (22): 2323–2326. doi:10.1126/science.290.5500.2323.
Shawe-Taylor, J., and N. Cristianini. 2004. *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press.
Smedt, I. D., M. Van Roozendael, T. Stavrakou, J.-F. Müller, C. Lerot, N. Theys, P. Valks, N. Hao, and R. R. 2012. "Improved Retrieval of Global Tropospheric Formaldehyde Columns from GOME-2/MetOp-A Addressing Noise Reduction and Instrumental Degradation Issues." *Atmos Measurement Technical* 5 (11): 2933–2949. doi:10.5194/amt-5-2933-2012.
Spurr, R. 2008. "LIDORT and VLIDORT: Linearized Pseudo-Spherical Scalar and Vector Discrete Ordinate Radiative Transfer Models for Use in Remote Sensing Retrieval Problems." In *Light Scattering Reviews*, edited by A. A. Kokhanovsky, Vol. 3, 229–275. Springer Nature. doi:10.1007/978-3-540-48546-9_7.
Spurr, R., J. De Haan, R. Van Oss, and A. Vasilkov. 2008. "Discrete-Ordinate Radiative Transfer in a Stratified Medium with First-Order Rotational Raman Scattering." *Journal of Quantitative Spectroscopy & Radiative Transfer* 109: 404–425. doi:10.1016/j.jqsrt.2007.08.011.
Theys, N., I. De Smedt, H. Yu, T. Danckaert, J. Van Gent, C. Hörmann, T. Wagner, et al. 2017. "Sulfur Dioxide Retrievals from TROPOMI Onboard Sentinel-5 Precursor: Algorithm Theoretical Basis." *Atmospheric Measurement Techniques* 10 (1): 119–153. doi:10.5194/amt-10-119-2017.
Tikhonov, A. N., and V. Y. Arsenin. 1977. *Solution of Ill-Posed Problems*. Winston: John Wiley & Sons Inc.
Timofeyev, Y., A. Polyakov, H. Steele, and M. Newchurch. 2003. "Optimal Eigenanalysis for the Treatment of Aerosols in the Retrieval of Atmospheric Composition from Transmission Measurements." *Applications Optical* 42 (15): 2635. doi:10.1364/ao.42.002635.
Van Der Maaten, L. J. P., E. O. Postma, and H. J. van den Herik. 2009. "Dimensionality Reduction: A Comparative Review." Technical Report. http://lvdmaaten.github.io/publications/papers/TR_Dimensionality_Reduction_Review_2009.pdf
Von Glasow, R., N. Bobrowski, and C. Kern. 2009. "The Effects of Volcanic Eruptions on Atmospheric Chemistry." *Chemical Geology* 263 (1–4): 131–142. Halogens in Volcanic Systems and Their Environmental Impacts. doi:10.1016/j.chemgeo.2008.08.020.
Wang, X., and F. J. Hickernell. 2000. "Randomized Halton Sequences." *Mathematical and Computer Modelling* 32 (7–8): 887–899. doi:10.1016/S0895-7177(00)00178-3.
Weisz, E., H.-L. Huang, L. Jun, E. Borbas, K. Baggett, P. Thapliyal, and L. Guan. 2007. "International MODIS and AIRS Processing Package: AIRS Products and Applications." *Journal of Applied Remote Sensing* 1 (1): 013519–013519–23. doi:10.1117/1.2766867.
Wentzell, P. D., and L. V. Montoto. 2003. "Comparison of Principal Components Regression and Partial Least Squares Regression through Generic Simulations of Complex Mixtures." *Chemometrics and Intelligent Laboratory Systems* 65 (2): 257–279. doi:10.1016/S0169-7439(02)00138-7.
Weyl, H. 1916. "Ueber Die Gleichverteilung Von Zahlen Mod. Eins." *Mathematische Annalen* 77 (3): 313–352. doi:10.1007/BF01475864.
Wilks, D. 2011. *Statistical Methods in the Atmospheric Sciences*. 3rd ed. New York: Elsevier.

Yang, K., X. Liu, N. A. Krotkov, A. J. Krueger, and S. A. Carn. 2009. "Estimating the Altitude of Volcanic Sulfur Dioxide Plumes from Space Borne Hyper-Spectral UV Measurements." *Geophysical Research Letters* 36: 10. doi:10.1029/2009GL038025.

Yang, K., P. K. Xiong Liu, N. A. Bhartia, S. A. Krotkov, E. J. Carn, A. J. Hughes, R. J. Krueger, D. Spurr, and S. G. Trahan. 2010. "Direct Retrieval of Sulfur Dioxide Amount and Altitude from Spaceborne Hyperspectral UV Measurements: Theory and Application." *Journal of Geophysical Research: Atmospheres* 115: D2. doi:10.1029/2010JD013982.

Young, A. T. 1981. "Rayleigh Scattering." *Applications Optical* 20 (4): 533–535. doi:10.1364/AO.20.000533.