Atmos. Chem. Phys., 9, 5321–5330, 2009
www.atmos-chem-phys.net/9/5321/2009/

Atmospheric Chemistry and Physics

Technical Note: Functional sliced inverse regression to infer temperature, water vapour and ozone from IASI data

U. Amato[1], **A. Antoniadis**[2], **I. De Feis**[1], **G. Masiello**[3], **M. Matricardi**[4], **and C. Serio**[3]

[1]Istituto per le Applicazioni del Calcolo "Mauro Picone" CNR, Napoli, Italy
[2]Laboratoire Jean Kuntzmann, Université Joseph Fourier, Grenoble, France
[3]Dipartimento di Ingegneria e Fisica dell'Ambiente, Università della Basilicata, Potenza, Italy
[4]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

Received: 7 January 2009 – Published in Atmos. Chem. Phys. Discuss.: 23 March 2009
Revised: 10 June 2009 – Accepted: 14 July 2009 – Published: 29 July 2009

Abstract. A retrieval algorithm that uses a statistical strategy based on dimension reduction is proposed. The methodology and details of the implementation of the new algorithm are presented and discussed. The algorithm has been applied to high resolution spectra measured by the Infrared Atmospheric Sounding Interferometer instrument to retrieve atmospheric profiles of temperature, water vapour and ozone. The performance of the inversion strategy has been assessed by comparing the retrieved profiles to the ones obtained by colocating in space and time profiles from the European Centre for Medium-Range Weather Forecasts analysis.

1 Introduction

The development of satellite high-spectral resolution infrared spectrometers is expected to improve quality and density of retrieval of atmospheric parameters. Both Numerical Weather Prediction and Earth's monitoring are expected to benefit from these new modern sensors.

The pioneer of this new generation of instruments has been the Interferometric Monitor for Greenhouse (IMG) gases sensor (Kobayashi et al., 1999), that flew on the ADvanced Earth Observing Satellite (ADEOS) platform from August 1996 to June 1997. At present, high-resolution infrared sensors on operational meteorological polar orbiters include the Atmospheric Infrared Sounder (AIRS) on the second Earth Observing System (EOS) polar orbiting platform, EOS-Aqua, launched in April 2002 (Aumann and Pagano, 1994), and the Tropospheric Emission Spectrometer (TES) on the AURA satellite launched in 2004 (Beer et al., 2001). In March 2002, the European Space Agency launched Envisat, an advanced polar-orbiting Earth observation satellite which provides measurements of the atmosphere, ocean, land, and ice. It is equipped with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), that is a Fourier transform spectrometer for the measurement of high-resolution gaseous emission spectra at the Earth's limb and operates in the near to mid infrared (Fischer et al, 1996, 2000).

The new arrived in the family of high resolution infrared sensors is the Infrared Atmospheric Sounding Interferometer (IASI) (EUMETSAT, 1998) on the first European Meteorological Operational Satellite (METOP/1) launched in 19 October 2006. IASI is a Fourier Transform Spectrometer based on a Michelson Interferometer coupled to an integrated imaging system that observes and measures infrared radiation emitted from the Earth in the spectral range 3.62–15.5 µm (645–2760 cm^{-1}), covering the peak of the thermal infrared and particularly the intense CO_2 band around 666 cm^{-1}, with an apodized resolution of 0.5 cm^{-1} and a spectral sampling of 0.25 cm^{-1}. IASI characteristics have been specified to get observations, which are compatible in terms of sampling, resolution, accuracy and overall performances with the mission objectives of providing improved information on temperature, water vapour, ozone, cloud top pressure and temperature, cloud cover and cloud optical properties.

The large amount of observed data needs algorithms for the radiative transfer equation and its inversion specifically designed for IASI. In this paper we describe a statistical regression methodology for temperature, water vapour and ozone, (T, q, o), which exploits IASI observations. The

Correspondence to: I. De Feis
(i.defeis@iac.cnr.it)

methodology is based on a suitable statistical dimension reduction technique, FSIR (Functional Sliced Inverse Regressions. FSIR (Amato et al., 2006) generalizes the well known Principal Component Analysis (PCA) (Jolliffe, 2002, e.g.) or Empirical Orthogonal Function (EOF) approach (see the recent review on EOF regression methods by Serio et al., 2009) and allows one to deal with functional models. Functional data analysis is about the analysis of information on curves or functions, and the radiance spectrum is a curve evaluated at fixed points that depend on the design of the interferometer.

FSIR needs to be trained on a suitable set of pairs: radiances, profiles. We have selected the profiles from the well known ECMWF (European Centre for Medium-range Weather Forecasts) Chevallier data base (Chevallier, 2001). The Chevallier data base documentation can be downloaded from the EUropean Organisation for the Exploitation of METeorological SATellites (EUMETSAT) Satellite Application Facilities (SAF) Scientific Report web page: http://www.eumetsat.int/groups/pps/documents/document/002197.pdf; also proper documentation and the data-base as well can be obtained directly by F. Chevallier at f.chevallier@ecmwf.int. IASI synthetic radiances have been computed using the radiative transfer code σ-IASI (Amato et al., 2002).

The code σ-IASI has been extensively validated with the use of the NAST-I instrument (Cousins and Gazarick, 1999, e.g.). NAST-I is the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Airborne Sounder Testbed Fourier Transform spectrometer, flying onboard the NASA aircrafts, ER-2 and Proteus. An extensive retrieval exercise with σ-IASI has been performed for the CAMEX/3 experiment (Convection and Moisture Experiment 3) (Carissimo et al., 2005, 2006, e.g.). More recently, σ-IASI has also been used within the EAQUATE campaign (European AQUA Thermodynamic Experiment), for analysis of NAST-I spectra (Proteus Flight) recoded over the Mediterranean Sea (Taylor et al., 2008; Grieco et al., 2007). The forward module σ-IASI has been also validated with the use of AIRS (Atmospheric Infrared Radiometer Sounder) data, flying onboard the Aqua satellite (Saunders et al., 2007).

The application of FSIR to IASI data has been exemplified through a series of IASI soundings recorded over the tropical basin. These observations have been FSIR-regressed for (T, q, o). To simplify the comparison of retrieval products with *truth* data, only clear-sky, sea-surface IASI soundings have been analyzed in this work. Truth data have been derived from the ECMWF analysis for the same date and location as the IASI soundings. The difference (retrieval-ECMWF) for temperature, water vapour and ozone has been evaluated, which has allowed us to assess the retrieval performance of FSIR. The paper also provides a comparison with a conventional EOF or PCA regression scheme (see also Serio et al., 2009).

The paper is organized as follows. Section 2 will deal with the mathematical aspects of FSIR. Application of FSIR to IASI data will be described and discussed in Sect. 4. Finally conclusions will be drawn in Sect. 5.

2 The regression model

Functional Sliced Inverse Regression (FSIR) (Amato et al., 2006) is a statistical tool to reduce the dimensionality based on the Sliced Inverse Regression (SIR) (Li, 1991), that permits to deal with functional models. It generalizes the Principal Component Analysis (PCA) using inverse regression.

In functional regression problems, one predicts a response variable Y from a set of variables $R_1, \ldots, R_d$ that are discretizations of a same curve $\boldsymbol{R}$ at points $\nu_1, \ldots, \nu_d$, that is $R_j = \boldsymbol{R}(\nu_j),\ j = 1, \ldots, d$, where the discretization points ν_j lie in some interval. In our case Y is the geophysical variable to be retrieved (surface temperature, temperature or gas concentration in a fixed layer) and $R_1, \ldots, R_d$ are the measurements, i.e., the radiances at wavenumbers $\nu_1, \ldots, \nu_d$. Let us consider the model

$$Y = m\left(\langle \beta_1, \boldsymbol{R} \rangle, \ldots, \langle \beta_K, \boldsymbol{R} \rangle\right) + \varepsilon\,, \tag{1}$$

where m is a smooth link function of the K-dimensional Euclidean space, E^K, into the one-dimensional space or real axis E^1, that indicates what kind of relation exists between the geophysical variable to retrieve and the radiance and is assumed to be linear; ε is noise assumed independent of the curve $\boldsymbol{R}(\nu)$; $\{\beta_i(\nu), i{=}1, \ldots, K\}$ are K orthonormal functions with respect to the usual inner product defined in the space $L_2([\nu_{\min}, \nu_{\max}])$ and denoted with the symbol $< \cdot, \cdot >$.

Let $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ be the covariance operator of $\boldsymbol{R}(\nu)$ and $\boldsymbol{\Sigma}_e$ the covariance operator of the conditional expected value, $E(\boldsymbol{R}|Y)$, of the radiance given the geophysical variable to retrieve.

Provided that $K<d$, with d denoting the number of points in which the radiance curve is known, Eq. (1) says that the regression function depends on $\boldsymbol{R}(\nu)$ only through K linear functionals of the explanatory process $\boldsymbol{R}(\nu)$. Hence, to explain the dependent variable Y, the space of d explanatory variables can be reduced to a space with a smaller dimension K. The dimension reduction methods aim at finding the dimension K of the reduction space and a basis defining this space. The functions β_i, $i{=}1, \ldots, K$, are called effective dimension reduction (edr)-directions and the space they generate is the edr-space.

FSIR is able to work with this model, indeed it yields d directions $\hat{\beta}_i$ and corresponding (eigen-)values $\hat{\lambda}_i$ which allow one to rank the importance of $\hat{\beta}_i$, where $\hat{\beta}_i$ and $\hat{\lambda}_i$ denote the estimated values. Obviously the $\beta_i(\nu)$ are determined in the values ν_j, $j{=}1, \ldots, d$. The technique takes advantage of the method of inverse regression. Here the aim is not to estimate $E[Y|\boldsymbol{R}{=}r]$ but the reverse $E[\boldsymbol{R}|Y{=}y]$, a one-dimensional regression problem that avoids *the curse of dimensionality*. This curse means that high-dimensional spaces have too few data for local averaging, that is the risk or expected squared error of estimation of a nonparametric regression estimator increases rapidly with the dimension p and to maintain a given degree of accuracy of an estimator, the sample size must increase exponentially with the dimension p. This does not happen with parametric techniques, i.e. least squares regression, whose risk will decay to zero at a rate of n^{-1}, where n is the number of observations. With inverse regression instead of having one d-dimensional regression problem we have d one-dimensional regression problems which do not suffer from that curse. The key of FSIR is the connection between the edr-space and the inverse regression curve given by the covariance operator of the inverse regression curve $\mathbf{\Sigma}_e$. Indeed it is possible to prove under some mild assumptions (Amato et al., 2006) that the eigenvalue-eigenvector decomposition of the operator $\mathbf{\Sigma}_{\boldsymbol{R}}^{-1}\mathbf{\Sigma}_e$ permits to identify a basis for the edr-space. Unfortunately the inverse of $\mathbf{\Sigma}_{\boldsymbol{R}}$ is not bounded, therefore we consider $\mathbf{\Sigma}_{\boldsymbol{R}}^{-1/2}\mathbf{\Sigma}_e\mathbf{\Sigma}_{\boldsymbol{R}}^{-1/2}$; in particular, we use the fact that $\mathbf{\Sigma}_{\boldsymbol{R}}^{-1/2}\mathbf{\Sigma}_e\mathbf{\Sigma}_{\boldsymbol{R}}^{-1/2}$ has the same eigenvectors as $\mathbf{\Sigma}_{\mathbf{R}}^{1/2}\mathbf{\Sigma}_e^{+}\mathbf{\Sigma}_{\boldsymbol{R}}^{1/2}$, where $\mathbf{\Sigma}_e^{+}$ is the Moore-Penrose generalized inverse operator (Groetsch, 1974).

Let us now consider a dataset of N radiance spectra measured at wavenumbers ν_j, j=1, ..., d, and corresponding profiles Y_N, that is $(\boldsymbol{R}_n, Y_n)$, n=1, ..., N, with $\boldsymbol{R}_n$ being vectors of dimension $d\times 1$. Then after centering the data, $\mathbf{\Sigma}_{\boldsymbol{R}}$ can be estimated by

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{R}_n \boldsymbol{R}_n^t \,. \tag{2}$$

For the estimate of $\mathbf{\Sigma}_e$ we may proceed in the following way. Let $M_Y(\nu)=E(\boldsymbol{R}|Y)$ and $\hat{M}_Y(\nu)$ be the wavelet smoothing of $\boldsymbol{R}(\nu)$ with design points the Y_n's, n=1, ..., N, obtained through the BINWAV estimator (Antoniadis and Pham, 1998). Then we consider the following estimate for $\mathbf{\Sigma}_e$:

$$\hat{\mathbf{\Sigma}}_{e,N} = \frac{1}{N}\sum_{n=1}^{N} \hat{\boldsymbol{M}}_{Y_n} \hat{\boldsymbol{M}}_{Y_n}^t \,, \tag{3}$$

with $\hat{\boldsymbol{M}}_{Y_n}=(\hat{M}_{Y_n}(\nu_j))_{j=1\ldots,d}=(E(R_j|Y=Y_n))_{j=1\ldots,d}$ a vector of dimension $d\times 1$ for each n=1, ..., N.

Convergence in probability of both $\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}$ and $\hat{\mathbf{\Sigma}}_{e,N}$ to $\mathbf{\Sigma}_{\boldsymbol{R}}$ and $\mathbf{\Sigma}_e$ can be found in (Amato et al., 2006). To estimate them accurately, we improve the conditioning of $\hat{\mathbf{\Sigma}}_{e,N}$ applying a projector method before performing the spectral decomposition: let $\hat{\boldsymbol{\pi}}_{k_N}$ denote the orthogonal projector into the space spanned by the k_N eigenvectors of $\hat{\mathbf{\Sigma}}_{e,N}$ corresponding to the k_N largest eigenvalues; we let $\hat{\mathbf{\Sigma}}_{e,N}^{k_N}= \hat{\boldsymbol{\pi}}_{k_N}\hat{\mathbf{\Sigma}}_{e,N}\,\hat{\boldsymbol{\pi}}_{k_N}$. Estimation of the EDR space is derived from the spectral decomposition of

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}^{1/2}\left(\hat{\mathbf{\Sigma}}_{e,N}^{k_N}\right)^{+}\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}^{1/2} \,, \tag{4}$$

where $\left(\hat{\mathbf{\Sigma}}_{e,N}^{k_N}\right)^{+}$ is the pseudoinverse matrix defined through the Singular Value Decomposition (SVD) (Golub and Van Loan, 1983; Golub, 1970). Let $(\alpha_i)_{i=1,\ldots,K}$ be the smallest eigenvalues of (4) and $\boldsymbol{\eta}_i$ the corresponding eigenfunctions, then

$$\hat{\boldsymbol{\beta}}_i^N = \frac{1}{\alpha_i}\left(\hat{\mathbf{\Sigma}}_{e,N}^{k_N}\right)^{+}\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}^{1/2}\boldsymbol{\eta}_i \,. \tag{5}$$

Summarizing the procedure for computing an estimate $\hat{\boldsymbol{\beta}}_i^N=(\hat{\beta}_i(\nu_1), \ldots, \hat{\beta}_i(\nu_d))^t$ of the EDR directions $\boldsymbol{\beta}_i$, i=1, ..., K, goes through the following steps:

Algorithm

1. calculate $\hat{M}_Y(\nu)$, the wavelet smoothing of $\boldsymbol{R}(\nu)$ with design points $Y_1, \ldots, Y_N$, using the BINWAV estimator and evaluate it in $\nu_1, \ldots, \nu_d$;

2. estimate $\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}$ by Eq. (2) and $\hat{\mathbf{\Sigma}}_{e,N}$ by Eq. (3);

3. evaluate the spectral decomposition of $\hat{\mathbf{\Sigma}}_{e,N}$ and its projection $\hat{\mathbf{\Sigma}}_{e,N}^{k_N}$;

4. evaluate the spectral decomposition of $\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}^{1/2}\left(\hat{\mathbf{\Sigma}}_{e,N}^{k_N}\right)^{+}\hat{\mathbf{\Sigma}}_{\boldsymbol{R},N}^{1/2}$ and estimate the EDR directions by Eq. (5).

FSIR can be seen as a generalization of PCA. Indeed supposing that $\mathbf{\Sigma}_{\boldsymbol{R}}$ is the identity operator, than FSIR aims at determining the directions along which to project the data by the eigenvalues-eigenvectors decomposition of $\mathbf{\Sigma}_e$, which takes into account the information about the profiles by means of a regression on the spectra; on the contrary the PCA uses only spectral information.

3 Vertical resolution of the retrieval

If the vector, $\hat{\boldsymbol{Y}}=(\hat{Y}(1), \hat{Y}(2), \ldots, \hat{Y}(M))^t$, is made up with the layer-mean estimated values of a given parameter, in order to form the profile function of the parameter, then the retrieval covariance matrix can be obtained by

$$\mathbf{\Sigma}_{\hat{\boldsymbol{Y}}} = E\left((\hat{\boldsymbol{Y}}-\boldsymbol{Y}_{\text{true}})(\hat{\boldsymbol{Y}}-\boldsymbol{Y}_{\text{true}})^t\right), \tag{6}$$

where expectation value has to be taken with respect to training data set and $\boldsymbol{Y}_{\text{true}}$ is the *true* value of the parameter. This matrix will be denoted by $\mathbf{\Sigma}_{\boldsymbol{T}}$, $\mathbf{\Sigma}_{\boldsymbol{H_2O}}$ and $\mathbf{\Sigma}_{\boldsymbol{O_3}}$ for temperature, water vapour and ozone profiles, respectively.

The retrieval covariance matrix, for a given parameter profile, can be used to analyze the spatial vertical resolution of the parameter itself, according to a methodology which has been developed by Serio et al. (2008b) and which is summarized for the sake of clarity.

The vector $(\boldsymbol{Y(1)}, \boldsymbol{Y(2)}, \ldots, \boldsymbol{Y(M)})^t$ represents the discretized version of a spatial function (i.e., temperature profile, water vapour mixing ratio profile, ozone mixing ratio profile). A strong correlation, that is a relatively high value of the covariance, between any two of the parameters means that the two have not been independently resolved by the data set, and that only some linear combination of the parameters is resolved. However, a direct examination and interpretation of the off-diagonal elements (covariances) of a covariance operator is not easy, which makes the definition of some suitable scalar index highly desirable.

To this end, it has to be considered that the retrieval correlation is determined by the non-null off-diagonal terms in the covariance operator. For a full independent retrieval (and therefore for a retrieval, which attains the maximum possible vertical spatial resolution), the covariance matrix has to be fully diagonal.

In what follows we will assume that the covariance operator has been normalized in order to obtain the correlation matrix,

$$\mathbf{\Sigma}_{\hat{Y}}(i,j) \leftarrow \frac{\mathbf{\Sigma}_{\hat{Y}}(i,j)}{\sqrt{\mathbf{\Sigma}_{\hat{Y}}(i,i)\mathbf{\Sigma}_{\hat{Y}}(j,j)}} \,. \tag{7}$$

$\mathbf{\Sigma}_{\hat{Y}}$ may be additively decomposed in its diagonal and off-diagonal components:

$$\mathbf{\Sigma}_{\hat{Y}} = \mathbf{\Sigma}_{\hat{Y},\mathrm{diag}} + \mathbf{\Sigma}_{\hat{Y},\mathrm{off}} \,. \tag{8}$$

An index which assesses the dominance of the diagonal term over the off-diagonal one might be simply defined from the norms of the matrices in Eq. (8). However, the norm is not additive. In general, we have

$$\mathrm{norm}(\mathbf{\Sigma}_{\hat{Y}}) \neq \mathrm{norm}(\mathbf{\Sigma}_{\hat{Y},\mathrm{diag}}) + \mathrm{norm}(\mathbf{\Sigma}_{\hat{Y},\mathrm{off}}) \,; \tag{9}$$

therefore, an index such as the ratio of the $\mathbf{\Sigma}_{\hat{Y}}$, diag-norm to the $\mathbf{\Sigma}_{\hat{Y}}$-norm is not well defined. It could be less or greater than one depending on the given matrix.

To quantify the relative contribution of the diagonal term (and off-diagonal term) to the norm of $\mathbf{\Sigma}_{\hat{Y}}$, let us consider the SVD of $\mathbf{\Sigma}_{\hat{Y}}$. With the usual notation, we have

$$\mathbf{\Sigma}_{\hat{Y}} = \mathbf{\Sigma}_{\hat{Y},\mathrm{diag}} + \mathbf{\Sigma}_{\hat{Y},\mathrm{off}} = \mathbf{U}\mathbf{S}_{\hat{Y}}\mathbf{V}^t \tag{10}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unit, orthogonal matrices of size M by M; $\mathbf{S}$ is a diagonal matrix whose elements are the eigenvalues, positive definite and supposed decreasingly ordered, of the operator $\mathbf{\Sigma}_{\hat{Y}}$. From the equation above, we have

$$\mathbf{U}^t\mathbf{\Sigma}_{\hat{Y},\mathrm{diag}}\mathbf{V} + \mathbf{U}^t\mathbf{\Sigma}_{\hat{Y},\mathrm{off}}\mathbf{V} = \mathbf{S}_{\hat{Y}} \tag{11}$$

which with the position

$$\begin{array}{r} \mathbf{B}_{\hat{Y},\mathrm{diag}} = \mathbf{U}^t\mathbf{\Sigma}_{\hat{Y},\mathrm{diag}}\mathbf{V} \\ \mathbf{B}_{\hat{Y},\mathrm{off}} = \ \mathbf{U}^t\mathbf{\Sigma}_{\hat{Y},\mathrm{off}}\mathbf{V} \end{array} \tag{12}$$

gives

$$\mathbf{B}_{\hat{Y},\mathrm{diag}} + \mathbf{B}_{\hat{Y},\mathrm{off}} = \mathbf{S}_{\hat{Y}} \tag{13}$$

Finally, because of the definition of norm (e.g. Golub and Van Loan, 1983), we have

$$\mathrm{norm}(\mathbf{\Sigma}_{\hat{Y}}) = \mathbf{S}_{\hat{Y}}(1,1) = \mathbf{B}_{\hat{Y},\mathrm{diag}}(1,1) + \mathbf{B}_{\hat{Y},\mathrm{off}}(1,1), \tag{14}$$

where $\mathbf{X}(1,1)$ is the element (1,1) of the given matrix $\mathbf{X}$. The formula above is fully additive and allows us to decompose the norm of the retrieval covariance matrix in its diagonal and off-diagonal contribution. Thus, a proper index which quantifies the degree of diagonalization of $\mathbf{\Sigma}_{\hat{Y}}$, that is how much the matrix is dominated by its diagonal terms, can be defined by

$$i_D = \frac{\mathbf{B}_{\hat{Y},\mathrm{diag}}(1,1)}{\mathbf{S}_{\hat{Y}}(1,1)} \,. \tag{15}$$

For a full diagonal matrix, we have i_D=1 and the retrieval is truly independent, whereas for a highly correlated matrix, we have i_D=M^{-1} and we can retrieve only the columnar amount of a geophysical parameter. As far as IASI is concerned, this is, e.g., the case for trace gases such as carbon monoxide or methane.

The index (15) can be easily re-scaled to the range 1 to M by simply redefining it as

$$i_D = M\frac{\mathbf{B}_{\hat{Y},\mathrm{diag}}(1,1)}{\mathbf{S}_{\hat{Y}}(1,1)}. \tag{16}$$

Then, i_D=1 simply means that for the retrieval at hand it is as if the full atmosphere had been divided just in one layer, that is only the columnar amount of the parameter has been resolved. On the opposite edge of the i_D scale, we have i_D=M, and the retrieval has been fully resolved on the grid mesh used to divide the atmosphere. Nearby layers can then, e.g., be used to form average quantities and, therefore, reduce the estimation error.

4 Application to IASI data

The retrieval accuracy of FSIR procedure has been assessed both on the Chevallier dataset and directly on IASI data. Analysis has focused on clear sky, sea-surface, tropical soundings. To limit the burden of the computational effort, only nadir view soundings have been considered.

4.1 Chevallier dataset

The training dataset consists of 377 ($\boldsymbol{T}, \boldsymbol{q}, \boldsymbol{o}$) tropical profiles for clear-sky and sea surface extracted from the ECMWF compilation by Chevallier (Chevallier, 2001).

These profiles are the input to the σ-IASI code (Amato et al., 2002), which has been used to calculate the related synthetic IASI spectra at nadir view angle.

The spectral ranges used to train the regression algorithm include the intense CO_2 band 645 cm^{-1}–830 cm^{-1}, the ozone band 1010 cm^{-1}–1070 cm^{-1}, the window

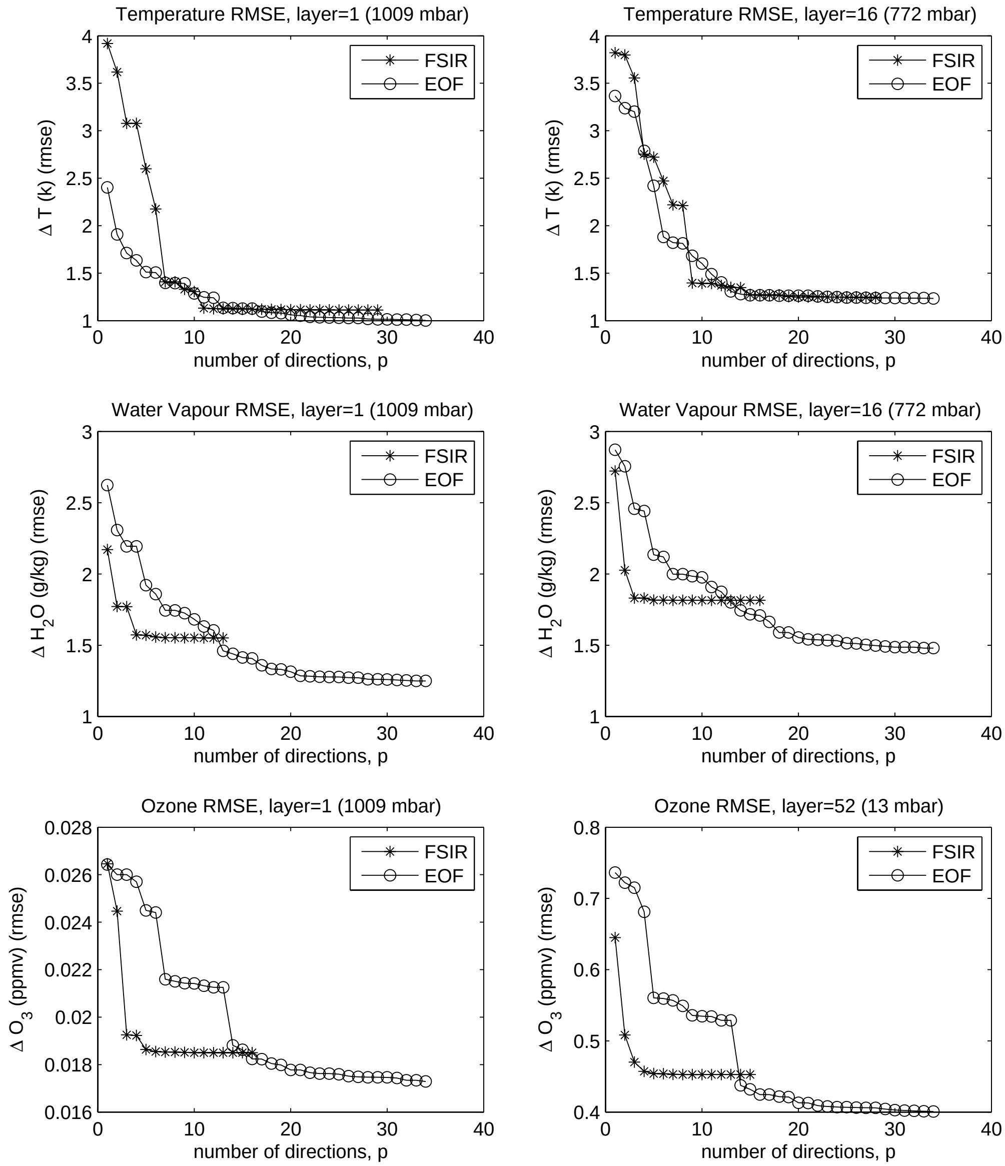

Fig. 1. Root mean square error of the retrievals as a function of the number of scores, for various atmospheric layers and parameters.

1130 cm^{-1}–1180 cm^{-1}, a portion of the ν_2 water vapour band, 1400 cm^{-1}–1700 cm^{-1}, and the C_2O/N_2O shortwave absorption band, 2000 cm^{-1}–2230 cm^{-1}, for a total of 3305 channels. Thus, the data space is made up of radiance vectors of size d=3305.

The parameter space is made up of triplets (T_j, q_j, o_j) for each given atmospheric layer. The number of layers, M_L, is the same as in σ-IASI, namely M_L=60. For temperature we have one more parameter, since we consider the surface temperature, as well.

The FSIR is implemented according to a single parameter regression algorithm, in which each single element of the triplet (T_j, q_j, o_j) is regressed vs. the EDR components.

Similarly to the EOF methodology, the FSIR approach also requires that the user specifies the number of EDR components, p, (also called scores). To this purpose we define the root mean square error or estimation error, $e(p)$, of the regression as $e(p)=E[(\hat{Y}-Y_{\mathrm{Chev}})^2]^{1/2}$, where, as before, $E[\cdot]$ means expectation value, again Y denotes a generic parameter and Y_{Chev} refers to values from the Chevallier dataset. Our algorithm is developed in the general context of a signal plus noise model (see, e.g., Eq. 1), therefore the curve $e(p)$ reaches a sort of plateau or noise floor. A suitable choice for p is the closest value to the knee of the curve, $e(p)$ (see Serio et al., 2009).

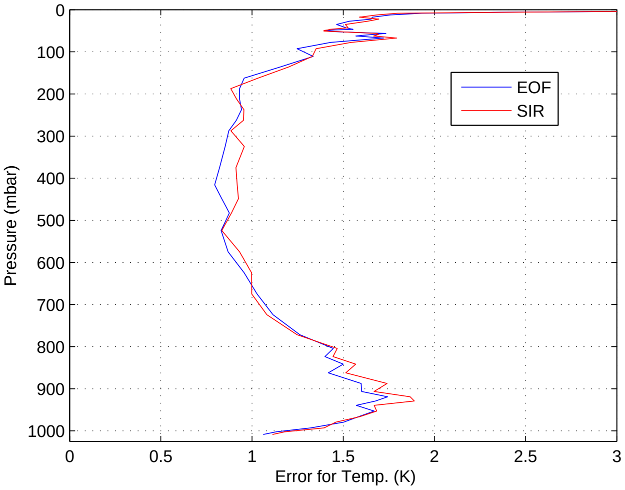

Fig. 2. Temperature root mean square error of the retrievals on the Chevallier dataset as obtained for FSIR and EOF.

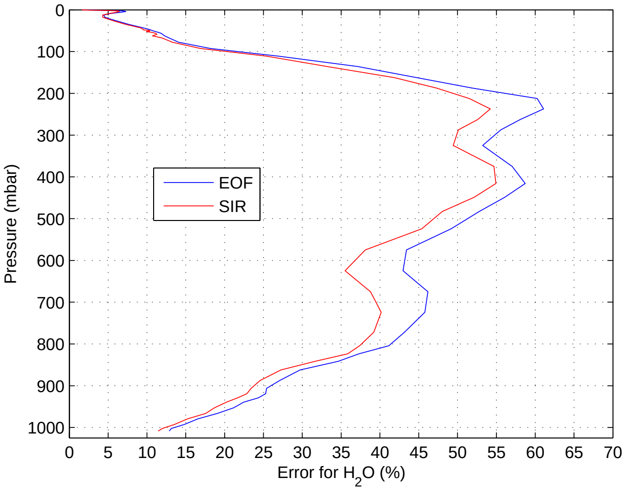

Fig. 3. Water vapour percentage root mean square error of the retrievals on the Chevallier dataset as obtained for FSIR and EOF.

Examples of these curves are shown in Fig. 1, which also provides a comparison with a PCA/EOF regression approach. Based on these curves the number of scores for which $e(p)$ reaches a plateau ranges between p=5 and p=20. Also note that FSIR tends to be much more parsimonious in terms of the number of selected scores.

Figures 2–4 exemplify the expected root mean square error for temperature, water vapour and ozone, respectively. The figures also allow us to compare the forecast skill of FSIR with that of the EOF regression. For the temperature the two regression schemes are almost equivalent in terms of expected root mean square error. For water vapour FSIR is superior to EOF, the same as for ozone.

4.2 i_D index

Using the methodology outlined in Sect. 3, we can analyze the degree of interdependency of the retrieval for temperature, water vapour and ozone, indeed the index i_D says how many independent linear combinations can be resolved from the data, and therefore how many independent pieces of information we have in the retrievals. From a statistical point of view it gives a measure of the degrees of freedom.

We give to the term degrees of freedom the usual meaning which is given to it in statistics. In its discretized form a given geophysical parameter is represented by a vector, $\boldsymbol{v}$ of size, say m. Therefore the vector $\boldsymbol{v}$ lives in a space of dimension m and we need, at least m independent equations, hence m pieces of information, to fully resolve for the geophysical parameter. However, because of data correlation and uncertainties, we have that the retrieved elements, $\hat{\boldsymbol{v}}_i$ can be correlated each to other, which means that only some linear combination of the elements of $\boldsymbol{v}$ has been resolved from the data set. The index i_D says how many independent linear combinations can be resolved from the data, and therefore how many independent pieces of information we have in the retrieved, $\hat{\boldsymbol{v}}$.

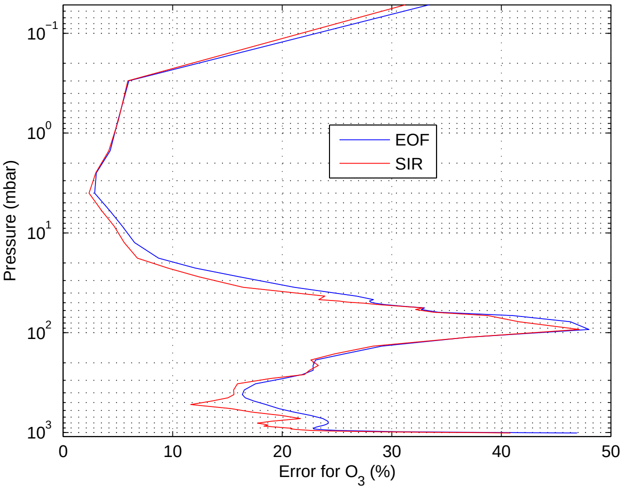

Fig. 4. Ozone percentage root mean square error of the retrievals on the Chevallier dataset as obtained for FSIR and EOF.

For temperature the value of i_D is 6.8 for FSIR and 5.6 for EOF. This means that FSIR gains more than one degree of freedom over the EOF regression. For water vapour and ozone we have 4.7 and 3.7, respectively, in the case of FSIR and 4.40 and 2.53 for EOF, respectively. The values for the index i_D were evaluated on the whole training dataset consisting of 377 tropical profiles.

It is interesting to note that the FSIR retrieval is less correlated than that produced by the EOF scheme, which means that FSIR has a better capability to reveal features and structures along the vertical.

However, in general the above i_D values are quite small. For ozone, they indicate that only two or three pieces of information are available for the retrieval. For temperature and water vapour, i_D says that only the very coarse features of the profile can be resolved. This is not a serious shortcoming for temperature, which is typically a smooth function of the altitude, but could become a serious limitations for water vapour, whose profile may be characterized by small-scale vertical structures.

In conclusion, FSIR can resolve important small structures in water vapour better than EOF. Of course, this does not mean that FSIR is, in absolute, the best retrieval method, we can only says that it is superior over EOF.

4.3 IASI retrieval and comparison with ECMWF analysis

The FSIR approach has been tested using IASI data obtained during the IASI commissioning phase.

The cloud detection scheme described in Grieco et al. (2007) was applied to tropical spectra measured over the sea surface during the 22 July 2007. This yielded a total of 603 clear sky IASI spectra. To simplify the illustration of the results only nadir view soundings have been considered.

To develop a consistent set of truth data against which IASI retrieval could be compared, ECMWF atmospheric analysis fields for temperature, water vapour and ozone were considered. These fields where co-located in space and time to the 603 IASI soundings. We used atmospheric analysis fields at 00:00, 06:00, 12:00 and 18:00 UTC on 22 July 2007. At that time, the ECMWF model was characterized by a vertical discretization of the atmosphere into 60 pressure levels and a horizontal truncation of T511. This truncation corresponds to a grid spacing of about 40 km or, equivalently, to a horizontal grid box of 0.351°×0.351°. The model has a hybrid vertical coordinate, with terrain-following coordinates in the lower troposphere and pressure coordinates in the stratosphere above about 70 hPa. Of the 60 levels in the vertical, 25 are above 100 hPa and the model top is at 0.1 hPa, corresponding to about 65 km. The vertical resolution of the analysis fields gradually decreases from 20 m at the surface to about 250 m at 1 km altitude, and about 1 km to 3 km in the stratosphere. The analysis fields were extracted from the ECMWF archive at the full T511 resolution, interpolated to a grid of points with a separation of 0.3°×0.3° and then co-located to the IASI soundings. The statistics of the difference between global radiosonde observations and ECMWF analysis in the troposphere show values of the standard deviation typically between 0.5 and 1 K for temperature and between 0.5 and 1.5 g/kg for water vapour. In addition to fields of temperature, water vapour and ozone, ECMWF fields of sea-surface temperature (SST) were also used in the study. It should be noted that these fields are based on analyses received daily from the National Centers for Environmental Prediction (NCEP), Washington DC, on a 0.5°×0.5° grid. These analyses are based on ship, buoy and satellite observations. In shallow waters, where rapid changes due to the upwelling radiation can occur close to land, the observed SST can sometimes differ as much as 5 K from the NCEP analysis.

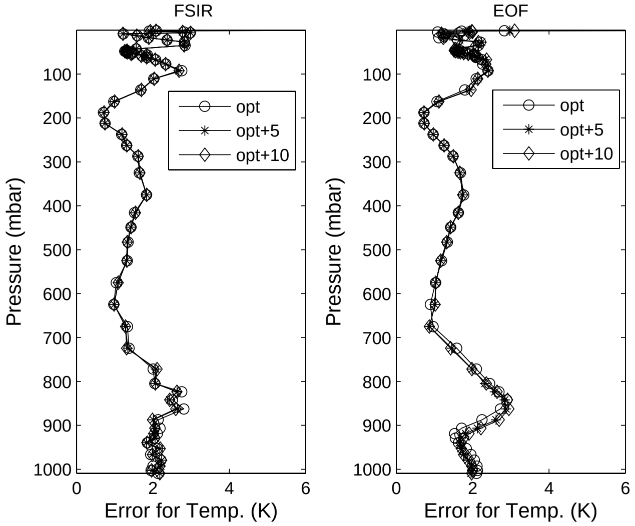

Fig. 5. Temperature root mean square difference (IASI retrieval – ECMWF) for 3 choices of the number of EDR (PC) scores for FSIR (left plot) and EOF (right plot).

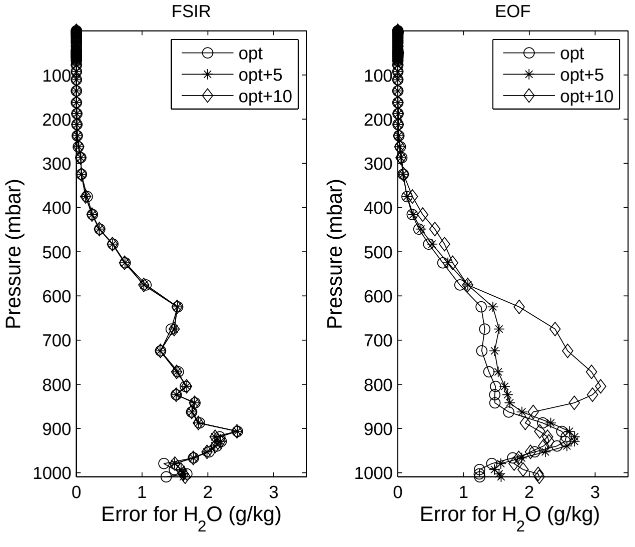

Fig. 6. Water Vapour root mean square difference (IASI retrieval – ECMWF) for 3 choices of the number of EDR (PC) scores for FSIR (left plot) and EOF (right plot).

One important aspect which has been possible to analyze with the help of the ECMWF analysis is the sensitivity of the retrieval accuracy to the choice of the number, p, of FSIR and/or PCA scores. Because of many factors, which include

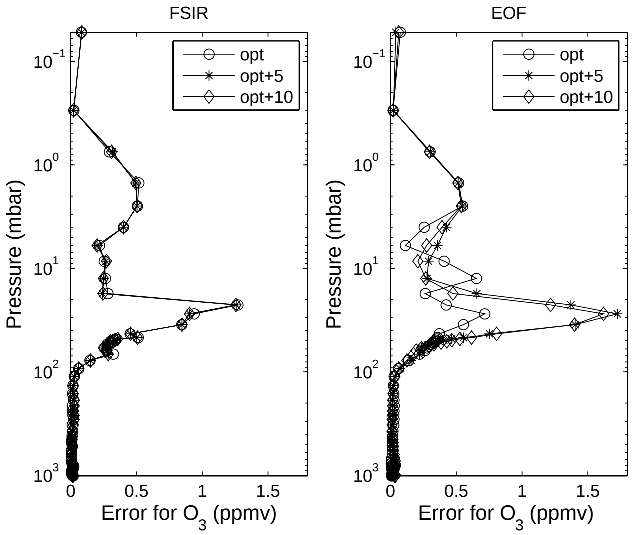

Fig. 7. Ozone root mean square difference (IASI retrieval – ECMWF) for 3 choices of the number of EDR (PC) scores for FSIR (left plot) and EOF (right plot).

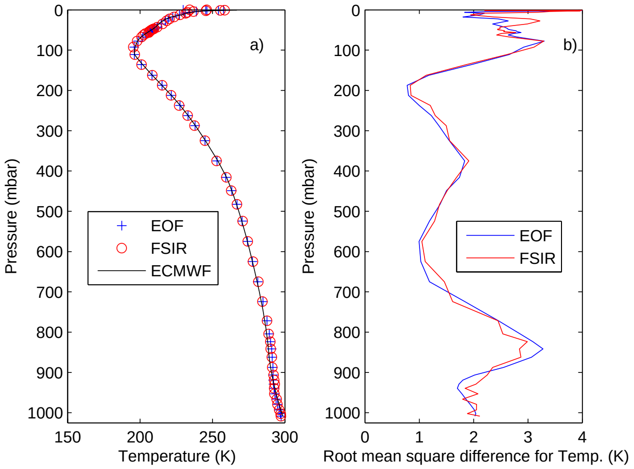

Fig. 8. **(a)** – Mean retrieved temperature profile obtained by averaging over the 603 IASI soundings and comparison with the ECMWF corresponding mean profile. **(b)** Root mean square difference (IASI retrieval – ECMWF) as obtained for EOF and FSIR methodologies.

radiative transfer accuracy, noise specifications, cloud contamination, in practice it may happen that the *optimal* choice, p_{opt}, performed based on the training data set may result to be sub-optimal. Thus, when this value is applied to real data and real conditions, a misfit may occur, which is much larger than that expected on the basis of the training data set.

To address this point we have redefined the root mean square difference, $e(p)$, as $e(p)=E[(\hat{Y}-Y_{\mathrm{ecmwf}})^2]^{1/2}$. For temperature, $e(p)$ is shown in Fig. 5 for three values of p, namely p_{opt}, $p_{\mathrm{opt}}+5$ and $p_{\mathrm{opt}}+10$. The calculations have been performed for FSIR and PCA. It can be seen that, in comparison to EOF, FSIR is more robust in terms of accuracy to variations in the number of scores, p. This is much more evident for the case of water vapour (Fig. 6) and ozone (Fig. 7).

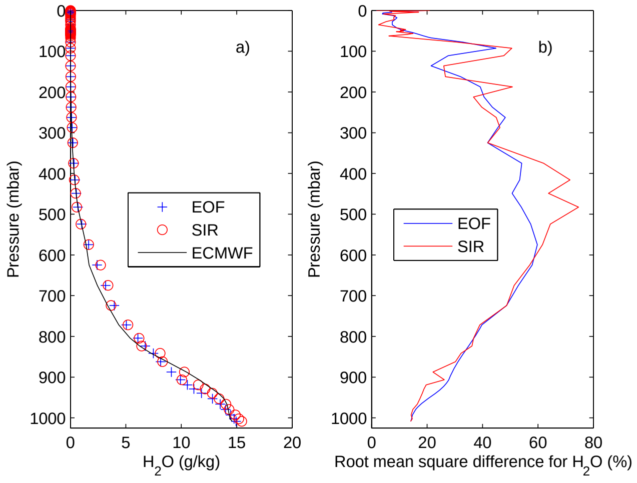

Fig. 9. **(a)** – Mean retrieved water vapour profile obtained by averaging over the 603 IASI soundings and comparison with the ECMWF corresponding mean profile. **(b)** Percentage root mean square difference (IASI retrieval – ECMWF) as obtained for EOF and FSIR methodologies.

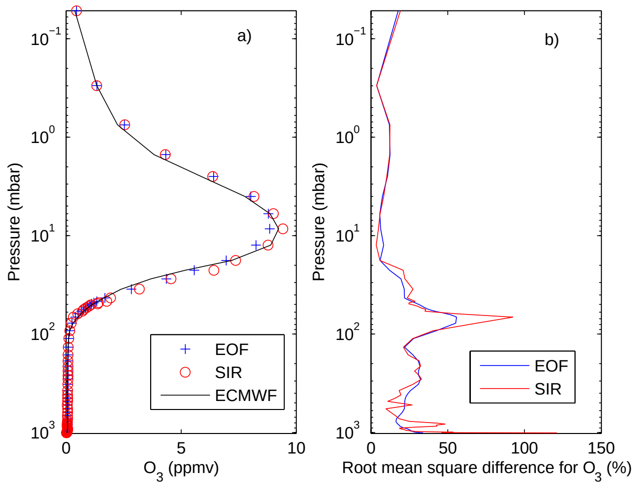

Fig. 10. **(a)** – Mean retrieved ozone profile obtained by averaging over the 603 IASI soundings and comparison with the ECMWF corresponding mean profile. **(b)** Percentage root mean square difference (IASI retrieval – ECMWF) as obtained for EOF and FSIR methodologies.

Apart from some isolated error spikes, evident in the case of ozone, FSIR also provides a more accurate retrieval, once compared to EOF. This can be seen in the three next

Figs. 8–10 which compare FSIR performance to EOF for the case of temperature, water vapour and ozone.

Figures 8–10 also suggest that FSIR accuracy for temperature is within 1–2 K, which is a bit larger than the 1 K expected for IASI. However, it is important to consider that the ECMWF analysis also has an uncertainty, which, as discussed above, is of the order of 0.5 to 1 K. Thus in $e(p)$ we also include the uncertainty of the ECMWF analysis.

The same as above can be said for water vapour: including the ECMWF uncertainty, the performance of FSIR reaches the expected accuracy of 10% only in the very deepest part of the atmosphere.

For ozone we get a very smooth retrieval, as it was expected based on the i_D index for this parameter. However, the result compares fairly well with ECMWF mean profile.

5 Conclusions

A new statistical strategy based on dimension reduction for the retrieval of atmospheric parameters from IASI radiances has been presented and discussed. Applications to IASI data have been considered for the case of tropical soundings. The new strategy, FSIR, has been also compared to a usual EOF regression scheme. The comparison shows that, mostly for gases (we have analyzed water vapour and ozone), FSIR gets a higher performance with respect to EOF. In general, FSIR seems to provide a retrieval which is better resolved along the vertical, which is particularly interesting for water vapour.

For temperature the FSIR scheme provides results quite close to the expected performance of 1 K in the lower part of the atmosphere. For water vapour the goal of 10% accuracy is reached only for the very lower part of the atmosphere.

For ozone it seems that FSIR is capable to provide at least 3 pieces of information.

We think that FSIR can provide a valid initialization scheme for physical-based inversion strategy, which is now under investigation.

Acknowledgements. IASI has been developed and built under the responsibility of the Centre National d'Etudes Spatiales (CNES, France). It is flown onboard the Metop satellites as part of the EUMETSAT Polar System. The IASI L1 data are received through the EUMETCast near real time data distribution service.

Edited by: T. Wagner

References

Amato, U., Antoniadis, A., and De Feis, I.: Dimension reduction in functional regression with applications, Comput. Stat. Data An., 50, 2422–2446, 2006.

Amato, U., Masiello, G., Serio, C., and Viggiano, M.: The σ-IASI code for the calculation of infrared atmospheric radiance and its derivatives, Environ. Modell. Softw., 17/7, 651–667, 2002.

Antoniadis, A. and Pham, D. T.: Wavelet regression for random or irregular design, Computat. Stat. Data An., 28, 353–369, 1998.

Aumann, H. H. and Pagano, R. J.: Atmospheric Infrared Sounder on the Earth Observing System, Opt. Eng., 33, 776–784, 1994.

Beer, R., Glavich, T. A., and Rider, D. M.: Tropospheric emission spectrometer for the earth observing system's aura satellite, Appl. Opt., 40, 2356–2367, 2001.

Carissimo, A., De Feis, I., and Serio, C.: The physical retrieval methodology for IASI: the δ-IASI code, Environ. Modell. Softw., 20, 1111–1126, 2005.

Carissimo, A., Grieco, G., Serio, C., Cuomo, V., Masiello, G., and Smith, W. L.: Application of sigma-IASI to NAST-I, IRS 2004: Current problems in atmospheric radiation. Proceedings of the International Radiation Symposium, Busan, Korea, 23–28 August 2004, A. Deepak Publishing, Hampton, VA, USA, 247–250, 2006.

Chevallier, F.: Sampled database of 60 levels atmospheric profiles from the ECMWF analysis, Tech. Rep., ECMWF EUMETSAT SAF programme Research Report 4, 2001.

Cousins, D. and Gazarick, M. J.: NAST Interferometer Design and Characterization, Final Report, MIT Lincoln Laboratory Project Report NOAA-26, July 13, 1999.

EUMETSAT: IASI science plan, EUMETSAT, Darmstadt, 1998.

Fischer H. and Oelhaf H.: Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb-emission spectrometers, Appl. Opt., 35, 2787-2796, 1996.

Fischer H., Blom C. E., Oelhaf H., Carli. B, Carlotti M., Delbouille L., Ehhalt D., Flaud J. M., Isaksen I., Lopez-Puertas M., McElroy C. T., and Zander R.: ENVISAT, MIPAS An instrument for ATmospheric Chemistry and Climate Research, editors C. Readings and R. A. Harris, ESA Publications Division, ESTEC, P.O. Box 299, 2200, AG Noordwijk, The Netherlands, SP-1229, 2000.

Golub, G. H. and Van Loan, C. F.: Matrix Computations, Baltimore, The Johns Hopkins University Press, 1983.

Golub, G. H. and Reinsch, C.: Singular value decomposition and least squares solutions, Numer. Mat., 14, 403–420, 1970.

Grieco, G., Masiello, G., Matricardi, M., Serio, C., Summa, D., and Cuomo, V.: Demonstration and validation of the φ-IASI inversion scheme with NAST-I data, Q. J. Roy. Meteor. Soc., 133(S3), 217–232, 2007.

Groetsch, C. W.: Generalized Inverses of Linear Operators: Representation and Approximation, Dekker, New York, USA, 1977.

Kobayashi, H., Shimota, A., Yoshigahara, C., Yoshida, I., Uehara, Y., and Kondo, K.: Satellite-Borne High-Resolution FTIR for Lower Atmosphere Sounding and Its Evaluation, IEEE Trans. Geosci. Remote Sens., 37, 1496–1507, 1999.

Jolliffe, I. T.: Principal Component Analysis, New York, USA, Springer-Verlag, 2002.

Li, K. C.: Sliced inverse regression for dimension reduction, with discussions, J. Amer. Statist. Assoc., 86, 316–342, 1991.

Saunders, R., Rayer, P., Brunel, P., von Engeln, A., Bormann, N., Strow, L., Hannon, S., Heilliette, S., Liu, X., Miskolczi, F., Han, Y., Masiello, G., Moncet, J. L., Uymin, G., Sherlock, V., and Turner, S. D.: A comparison of radiative transfer models for simulating Atmospheric Infrared Sounder (AIRS) radiances, J. Geophys. Res., 112, D01S90, doi:10.1029/2006JD007088, 2007.

Serio, C., Masiello, G., and Grieco, C.: EOF regression analytical model with applications to the retrieval of atmospheric temperature and gas constituents concentration from high spectral resolution infrared observations, in Environmental Modelling: New Research, edited by: Findley, P. N., Nova Science Publishers, Inc., 51–88, 2009.

Serio, C., Esposito, F., Masiello, G., Pavese, G., Calvello, M. R., Grieco, G., Cuomo, V., Buijs, H. L., and Roy, C. B.: Interferometer for ground-based observations of emitted spectral radiance from the troposphere: evaluation and retrieval performance, Appl. Opt., 47(21), 3909–3919, 2008b.

Taylor, J. P., Smith, W. L., Cuomo, V., Larar, A. M., Zhou, D. K., Serio, C., Maestri, T., Rizzi, R., Newman, S., Antonelli, P., Mango, S., Di Girolamo, P., Esposito, F., Grieco, G., Summa, D., Restieri, R., Masiello, G., Romano, F., Pappalardo, G., Pavese, G., Mona, L., Amodeo, A., and Pisani, G.: EAQUATE – An International Experiment For Hyper-spectral Atmospheric Sounding Validation, B. Am. Meteor. Soc., February issue, 203–218, doi:10.1175/BAMS-89-2-203, 2008.