SPATIAL NOISE-AWARE TEMPERATURE RETRIEVAL FROM INFRARED SOUNDER DATA

David Malmgren-Hansen[†], Valero Laparra[‡], Allan Aasbjerg Nielsen[†] and Gustau Camps-Valls[‡]

[†]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark
[‡]Image Processing Laboratory (IPL), Universitat de València, Spain

ABSTRACT

In this paper we present a combined strategy for the retrieval of atmospheric profiles from infrared sounders. The approach considers the spatial information and a noise-dependent dimensionality reduction approach. The extracted features are fed into a canonical linear regression. We compare Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF) for dimensionality reduction, and study the compactness and information content of the extracted features. Assessment of the results is done on a big dataset covering many spatial and temporal situations. PCA is widely used for these purposes but our analysis shows that one can gain significant improvements of the error rates when using MNF instead. In our analysis we also investigate the relationship between error rate improvements when including more spectral and spatial components in the regression model, aiming to uncover the trade-off between model complexity and error rates.

Index Terms— Infrared Atmospheric Sounding Interferometer (IASI), Minimum Noise Fractions, Principal Component Analysis (PCA), Statistical retrieval.

1. INTRODUCTION

> *"Perfection is achieved not when there is nothing more to add, but when there is nothing more to take away."*
> — *Antoine de Saint-Exupry: Terre des hommes.*

Temperature and water vapour atmospheric profiles are essential meteorological parameters for weather forecasting and atmospheric chemistry studies. Observations from high spectral resolution infrared sounding instruments on board of satellites provide for retrieval of such profiles. However, it is not trivial to retrieve the full information content from radiation measurements; accordingly, improved retrieval algorithms are desirable to achieve optimal performance for existing and future infrared sounding instrumentation.

EUMETSAT, NOAAA, NASA and other agencies are continuously developing product processing facilities to obtain L2 atmospheric profile products from infrared hyperspectral radiance instruments, such as IASI. One of the retrieval techniques commonly used in L2 processing is based on linear regression, which is a valuable and very computationally efficient method. It consists of performing a canonical least squares linear regression on top of the data projected onto the first principal components or Empirical Orthogonal Functions (EOF) –known in statistics as PCA– of the measured brightness temperature spectra (or radiances) and the atmospheric state parameters. To further improve the results of this scheme for retrieval, nonlinear statistical retrieval methods, as well as nonlinear pre-processing methods [1], can be applied as an efficient alternative to more costly optimal estimation (OE) schemes. These methods have proven to be valid in retrieval of temperature, dew point temperature (humidity), and ozone atmospheric profiles when the original data are used [2, 3]. However, they are costly to train and do not consider spatial correlation between radiances neither the noise information.

Recently, in [4], a high improvement on the performance of retrieval methods was reported when applying standard compression algorithms to the images. Although this result may appear counter-intuitive since compression implies reduction on the amount of information in the images, a certain level of compression is actually useful because: 1) compression removes information but also noise, and it could be useful to remove the components with low signal-to-noise ratio.; and 2) spatial compression introduces in a simple way information about the neighboring pixels. The use of Minimum Noise Fractions (MNF) employed here is a simpler and more mathematically elegant way to take advantage of both properties simultaneously. MNF is specifically designed to sort components according to the signal-to-noise ratio (SNR) [5]. The way we apply MNF here also enforces the inclusion of spatial information as noise is estimated by the residuals of fitting a quadratic surface locally. In this work we compare the effect of using PCA or MNF when retrieving temperature profiles using IASI data. We will show that MNF is better suited for this task. Moreover since PCA and MNF are both linear and unsupervised transformations, using MNF do not introduce any modification in the data processing pipeline.

The remainder of the work is organized as follows. Section §2 describes the data set collected and the pre-processing for dimensionality reduction and spatial filtering. Section §3 reviews the two decomposition methods used in the work. Section §4 gives empirical evidence of performance of the proposed scheme for spatial, noise-aware retrieval of atmospheric parameters. We conclude in §5 with some remarks and outline for the further work.

2. DATA DESCRIPTION

The Infrared Atmospheric Sounding Interferometer (IASI) data are point measurements of approximately $25\ km$ diameter with 8461 spectral components, ranging in the infrared emission spectra from 645 to $2760\ cm^{-1}$ with $0.25\ cm^{-1}$ resolution. The dataset collected for this paper consists of 4 consecutive orbits from august 2013 of which the first three are used for training the regression model and the last is used for testing.

In our problem we follow the same scheme proposed in [4]. First we remove certain bands from the spectrum that do not contains useful information for retrieval reducing the data to 4699 spectral components. Although the longitudinal distance between acquisition points increases towards equator we can reshape each orbit into a rectangular grid of 1530×60 elements. By doing so, data can be treated as an image, taking advantage of spatial relations. The dimensionality reduction transformations are calculated on the training set and applied to both the training and testing datasets.

3. DECOMPOSITION METHODS

In our analysis we consider two orthogonal transformations, PCA [6] and MNF [5]. Notationally, given an observation data matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ with n pixels of d dimensions, we aim to find a transformation to a lower dimensional representation, $d' < d$, such that the projected data preserves most of the 'information' of the input. Solutions offered by both PCA and MNF are found by solving an eigenvalue problem but where the PCA finds a solution with eigenvectors in the columns of $\mathbf{W} \in \mathbb{R}^{d \times d'}$ in direction of maximum variance, the MNF looks for the eigenvectors that minimize the noise fraction, or equivalently maximizes the signal-to-noise ratio [7, 8]:

$$\text{PCA}: \mathbf{W}_* = \underset{\mathbf{W}}{\arg\max} \left\{ \text{Tr} \left(\frac{\mathbf{W}^\top \mathbf{X}^\top \mathbf{X} \mathbf{W}}{\mathbf{W}^\top \mathbf{W}} \right) \right\}$$

$$\text{MNF}: \mathbf{W}_* = \underset{\mathbf{W}}{\arg\max} \left\{ \text{Tr} \left(\frac{\mathbf{W}^\top \mathbf{X}^\top \mathbf{X} \mathbf{W}}{\mathbf{W}^\top \mathbf{X}_N^\top \mathbf{X}_N \mathbf{W}} \right) \right\},$$

(1)

where $\mathbf{X}$ is our data matrix with each row representing a sample of a infrared spectrum and with columns corresponding to the number of spectral components. $\mathbf{X}_N$ is the corresponding noise estimation of each sample in $\mathbf{X}$. The resulting set of vectors from the PCA decomposition are orthogonal as opposed to the MNF solution which obtains orthogonality with respect to the noise covariance.

If the noise covariance matrix is known, it can be used in the MNF estimation. Often it is not the case and it has to be estimated from data. Common ways to do noise estimation in image analysis include local mean subtraction, or taking the residuals from a plane or paraboloid fit on every pixel position in the image. We follow the latter approach for our analysis with a 3×3 paraboloid residual kernel implemented as a filtering operation [9].
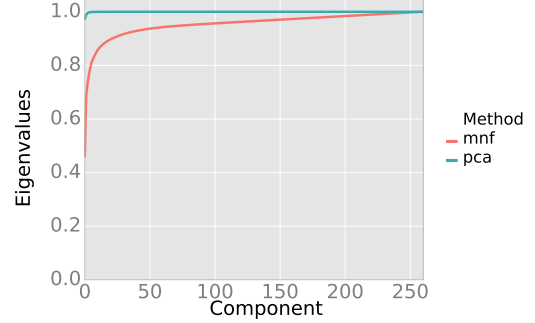

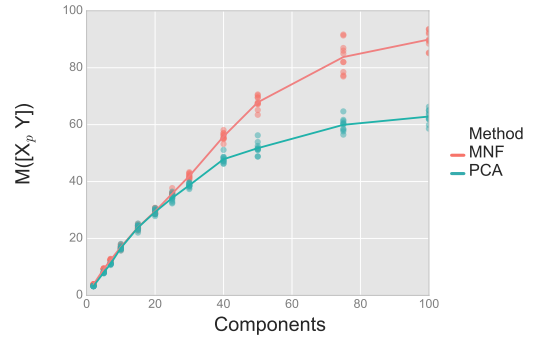**Fig. 1**. Cummulated normalized eigenvalues.

Fig. 2. Multi-information between the input and output components for PCA and MNF. Ten realizations have been made for each method while including different amount of spectral components. Lines denote the mean trend of the results.

The cumulative and normalized eigenvalues for both methods are shown in Fig. 1. For PCA they represent the percentage of explained total variance and it is seen that 99% explained variance is obtained within the first 5 components. For the MNF, eigenvalues represent the signal fraction for each component [5] and less than 80% signal fraction is obtained from the first 5 components. Although this could be seen as an disadvantage one has to take into account that PCA might keep the noise information too. Therefore how the eigenvalues relates to the information necessary to predict the temperature profiles $\mathbf{Y} \in \mathbb{R}^o$ is less straight forward to estimate. We analyze in Fig. 2 this relation by using the concept of *multiinformation* [10] (also known as *total correlation*). We show the amount of multiinformation, i.e. shared information, between the projected inputs $\mathbf{X}_p = \mathbf{X}\mathbf{W}$ and the outputs $\mathbf{Y}$ using different amount of input components for each decomposition method. These values have been computed using RBIG method ([11]). We have followed similar procedure as in [12] where the amount of information contained by spatial and spectral components was analyzed for several sensor configurations. In this case, we are analyzing only the spectral information, yet including the variable to predict $\mathbf{Y}$ (temperature profiles). Fig. 2 shows the multiinformation results for PCA and MNF. Although it also includes the redundant information of the inputs, this measure can be

seen as an approximation of the information of the output that we can be obtained from the input. Note that even that MNF is not specifically designed to maximize this information, the multiinformation is bigger for MNF when using the same number of input components than for PCA. We will see in the experiments section how this behavior gives raise to improved retrieval performance.

As suggested in [4], to improve the retrieval performance it is important to remove noise from the data and to include spatial information. Fig. 3 illustrates the ability of MNF to do so. We show half orbit of data from the test set projected onto each of the 50 first components from the PCA (top row) and MNF (bottom row) decomposition. It is clear from this figure that MNF obtains smoother are less noisy projections than PCA. For instance component 38 from the PCA projection seem to contain less structure than the three following projections. This indicates that some noise components in the data have higher variance than other signal components. This behaviour repeats above the first 50 PCA components, whereas the MNF projections represents spatially smooth information in early components and gradually increase to finer details for higher components.

4. EXPERIMENTAL RESULTS

The goal of our experiments is two-fold, first to compare the effect of using PCA or MNF in the retrievals, and furthermore to uncover the trade-off between prediction performance and the number of spectral components included for each method. Dimensionality reduction is important to limit the computational load but choosing the appropriate number of components to keep is less straightforward. A lower computational load can be traded for larger amounts of training data so overfitting is prevented. Alternatively the lower number of data dimensions can enable the use of computationally heavier nonlinear models such as Kernel Ridge Regression, which has been shown to improve performance for retrieval in infrared sounder data [4].

As well as the influence of spectral sampling in temperature profile modelling we include experiments for different sizes of pixel neighborhood sampling as studies suggest this can be beneficial [12]. This means that we model the temperature profile of one sample in the IASI data from the sample plus a neighborhood of samples around it. For quadratic neighborhoods the increase of size will also lead to quadratic increase in computational load and it is therefore relevant to limit it.

In Fig. 4 the results from our experiments are shown. It is seen that the RMSE improvements converges after approximately 125 spectral components. The results also show that there is a significant improvement including neighborhood pixels in the modelling of temperature profiles, but that the improvement decreases going towards larger neighborhood sizes. Figure 5 shows the resulting RMSE over the temperature profile for using 175 spectral components in the Linear regression model. Our analysis suggests to use between $125 - 175$ spectral components from a MNF decomposition and a pixel neighborhood sampling size of 3×3 or 5×5 when
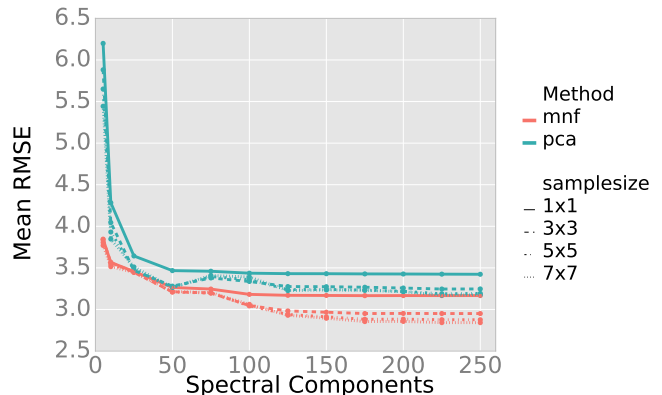

Fig. 4. Mean of RMSE on temperature profiles for different decomposition methods and spatial sample sizes, when including an increasing amount of spectral components. The Y-axis is the mean of the root-mean-square-error on prediction with $o = 90$ i.e. predition of 90 altitudes in the atmosphere given by their pressure level.

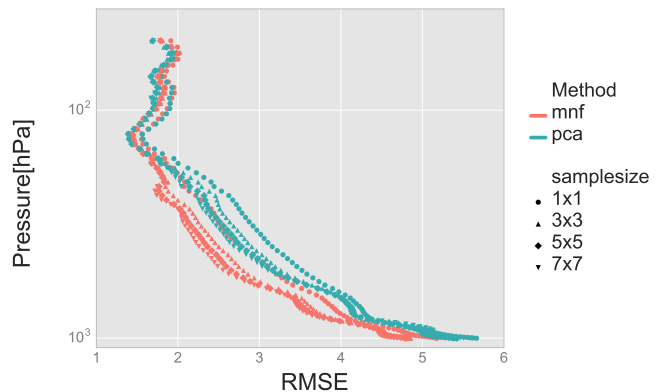performing Linear Regression on this type of data.

Fig. 5. RMSE on temperature profiles for 90 different altitudes shown by the air pressure (y-axis). In this experiment 175 spectral components was used to fit the Linear Regression model for 3 different neighborhood sampling size on PCA and MNF transformation. Measurements on clouds yields typically higher error rates for lower altitudes. Note that in this experiment nothing has been done to filter away measurements dominated by clouds.

5. CONCLUSIONS

This paper showed that using MNF is a simple and mathematically elegant way of removing the noise in the signal and at the same time taking into account spatial information. These two properties have been suggested previously as an important point when dealing with this particular data [4]. Both effects can be observed in Fig. 3, the selected features by the MNF are less noisy and spatially softer than the ones found
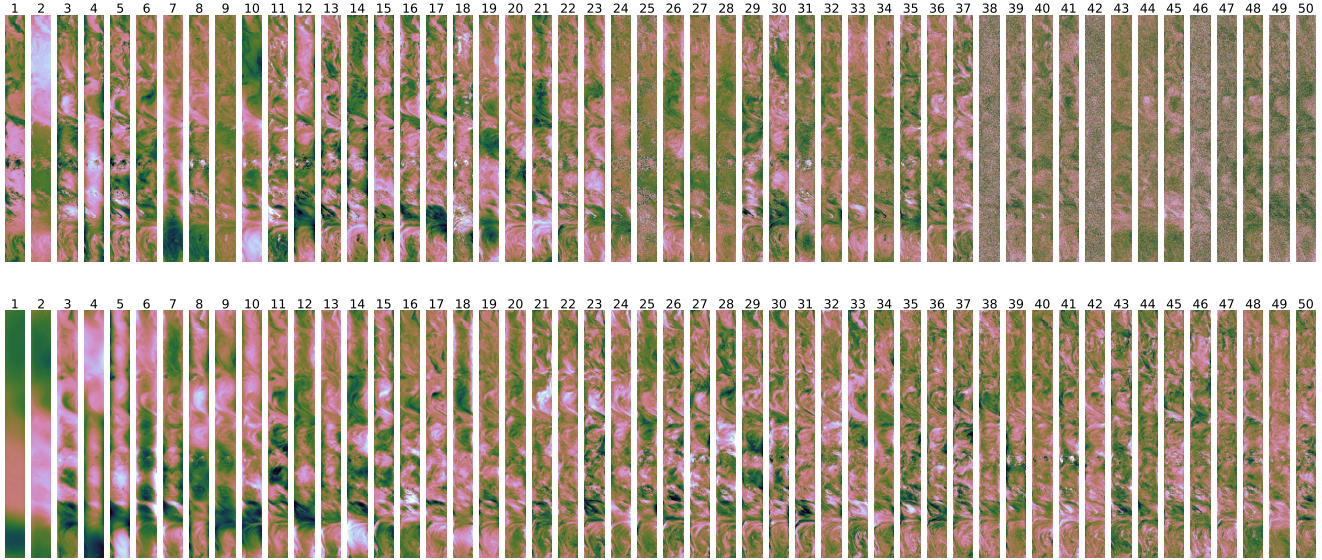
Fig. 3. Each strip is half an orbit projected onto the different components (PCA top, MNF bottom).

by PCA. We want to stress the fact that substituting PCA by MNF would not change the processing pipeline. PCA and MNF are both linear transformations so only the values of the projecting vectors should be changed. Moreover, unlike other solutions as PLS [13], PCA and MNF are unsupervised methods, i.e. are not fitted for predicting an specific variable. Therefore, although we here show the results for a particular variable (i.e. temperature), it is expected that the improvement would be consistent for the retrieval of other variables.

6. REFERENCES

[1] J. Arenas-Garca, K.B. Petersen, G. Camps-Valls, and L.K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 16–29, 2013.

[2] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1759–1769, 2012.

[3] G. Camps-Valls, V. Laparra, J. Muñoz-Marí, L. Gómez-Chova, and X. Calbet, "Kernel-based retrieval of atmospheric profiles from IASI data," in *IEEE Proc. IGARSS 11*, Jul 2011, pp. 2813–2816.

[4] J. García Sobrino, Serra-Sagristà, V. Laparra, X. Calbet, and G. Camps-Valls, "Statistical atmospheric parameter retrieval largely benefits from spatial-spectral image compression," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[5] Andrew A Green, Mark Berman, Paul Switzer, and Maurice D Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on geoscience and remote sensing*, vol. 26, no. 1, pp. 65–74, 1988.

[6] Harold Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of educational psychology*, vol. 24, no. 6, pp. 417, 1933.

[7] Allan Aasbjerg Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 612–624, 2011.

[8] Luis Gómez-Chova, Allan A Nielsen, and Gustavo Camps-Valls, "Explicit signal to noise ratio in reproducing kernel hilbert spaces," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 3570–3573.

[9] Allan Aasbjerg Nielsen, "An extension to a filter implementation of a local quadratic surface for image noise estimation," in *Image Analysis and Processing, 1999. Proceedings. International Conference on*. IEEE, 1999, pp. 119–124.

[10] M. Studený and J. Vejnarová, *The Multiinformation Function as a Tool for Measuring Stochastic Dependence*, pp. 261–297, Springer Netherlands, Dordrecht, 1998.

[11] V. Laparra, G. Camps-Valls, and J. Malo, "Iterative gaussianization: From ica to random rotations," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 537–549, April 2011.

[12] Valero Laparra and Raúl Santos-Rodríguez, "Spatial/spectral information trade-off in hyperspectral images," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1124–1127.

[13] H. Wold, "Non-linear estimation by iterative least squares procedures," *Research papers in Statistics*, pp. 411–444, 1966.