

Spectral High Resolution Feature Selection for Retrieval of Combustion Temperature Profiles

Esteban García-Cuesta¹, Inés M. Galván², and Antonio J. de Castro¹

¹ Physics Department

² Computer Science Department

Carlos III University -Avenida de la Universidad, 30 - 28911 Leganés (Madrid), Spain
egc@fis.uc3m.es, igalvan@inf.uc3m.es, decastro@fis.uc3m.es

Abstract. The use of high spectral resolution measurements to obtain a retrieval of certain physical properties related with the radiative transfer of energy leads a priori to a better accuracy. But this improvement in accuracy is not easy to achieve due to the great amount of data which makes difficult any treatment over it and its redundancies. To solve this problem, a pick selection based on principal component analysis has been adopted in order to make the mandatory feature selection over the different channels. In this paper, the capability to retrieve the temperature profile in a combustion environment using neural networks jointly with this spectral high resolution feature selection method is studied.

1 Introduction

Progress in optoelectronic technologies during last decade has led to the fabrication of new sensors to measure the radiated energy focused on new measurement concept based on high spectral resolution measurements. High resolution measurements leads a priori to better accuracy in retrieval of physical properties in radiative transfer of energy (RTE) problems. However, as the number of data increases, it makes more difficult the use of conventional data regression techniques to retrieve the physical information involved in the problem. One problem is related with the amount of samples needed to cover this high dimensionality space, which is so-called the curse of dimensionality. Also, the high dimensionality increase the complexity of regression models used to retrieve the information and consequently the number of operations to solve it. This make impractical the use of these kind of techniques when the dimensionality is large and encourage the use of new techniques for this kind of inverse problems.

The industrial fuel fired furnace is a context where optoelectronic technologies have some advantages over conventional temperature monitoring devices such thermocouples. These advantages are mainly three: it is not intrusive, it does not disturb the measurement, and it can undergo the harsh furnace environment. In this context it is very important to have devices that monitor and control the combustion process in order to minimise pollutant emissions as well as to optimise energy losses. Flame temperature appears, among others, as a very important parameter to be monitored[1][2][3][4].

The retrieval of temperature profile from high resolution measurements of radiative transfer of energy is related to ill-posed problems or under-constrained since we are trying to retrieve a continuous function from a finite measurements[5].

The use of artificial neural networks as inverse model for the RTE seem to be an alternative to retrieve the temperature profile . One important advantage of neural networks is their speed. Once the neural network has been trained, the results of the inversion method are almost instantaneous compared to regression models. Some other advantages over classical physical-statistical techniques are the not need of good initial conditions for the inversion model and the independence of a rapid direct model for iterative inversion algorithms[6].

The purpose of this study is to show an approximation to retrieve the temperature profile of combustions processes composed by CO₂ and water vapor from optical spectroradiometric measurements. The use of artificial neural networks and the improvement obtained with an specific selection of spectra channels in this kind of problems will be tested based on the accuracy of the results obtained for the retrieval.

Two approaches are presented in this paper to validate the use of neural networks and the improvement obtained using statistical techniques to introduce a priori knowledge in the reducing dimensionality process. In a initial approach a multilayer perceptron neural network has been trained using all the data belonging to a spectrum which implies a huge number of input nodes including redundancies and possible noisy information. A priori, as more data is included better results would be expected, but in the other hand this high dimensionality could prevent an appropriate performance of MLP.

To prevent this disadvantages using high spectral resolution measurements a refine approach to reduce the dimensionality is proposed. There are different ways to make a reduction of the dimensionality: feature extraction(linear or not linear), transformation of data (principal components analysis projections) and feature selection[7]. Because of the importance of semantic interpretation, a feature selection is adopted instead of methods which find a mapping between the original feature space to lower dimensional feature space. Principal Components Analysis (PCA)[8][9], and typical methods as B2 and B4 to select a subset of original variables[10] has been tested without successful results. Here, a feature selection with a pick peak selector is proposed which tries to spread the selection according to a priori physical knowledge. Then a MLP is trained with the selected channels reducing the complexity of the network and improving it's performance.

The the paper is organised as follows: In Sect. 2 a context description of retrieval of temperature profiles in flames is made, describing the simulations carried out and the obtained results. This section includes also the use of PCA and peak selection to reduce the dimensionality of the input space for the MLP. Discussion and conclusions are presented in Sect. 3.

2 Temperature Retrieval Using Neural Networks

Neural networks techniques can be used as an approach to solving problems of fitting experimental data. In the context of approximation of nonlinear maps, the architecture most widely used is the MLP. It is relatively quite straightforward to use and it has been proven by different authors that they are universal approximators ([11], [12]), in the sense that any continuous function may be represented by a MLP with one hidden layer. Here the MLP is used as inverse model for the RTE.

In previous works, neural networks have been used to invert the RTE in atmospheric problems[13]. However, important differences can be pointed out in relation to the combustion problem involved in this work. Values and variation ranges for temperature and gas concentrations are different. Moreover, the optic path length is unknown from the point of view of transmitter-receptor.

In this context of industrial fuel fired furnace, given a high spectral spectrum energy measurement from a sensor (spectroradiometer) which are related to the Eq. (1), the goal is to recover the temperature profile inside a hot gas cloud. The dependence of the energy measured by the sensor and temperature and optical-path length of the hot gas is expressed by the RTE

$$R_\nu = (I_0)_\nu \tau_\nu(z_o) + \int_{z_0}^{\infty} B_\nu\{T(z)\} \frac{d\tau_\nu(z)}{dz} dz \quad (1)$$

where ν_i are the different channels of energy, $B_\nu\{T(z)\}$ is the Planck function which indicates the radiance emitted by a blackbody at temperature T and τ_ν is the hot gas transmission at channel ν between the sensor and the depth z . The value of τ depends in a non-linear way of T and z ; $\tau = -exp^{[KTz]}$, where K is a constant provided by HITRAN/HITEMP database[14].

The problem to obtain the temperature profile from such a spectrum is not straightforward. Energy emission at each channel depends in a non-linear way on parameters like the spatial distribution of temperature and the gas cloud width. Moreover each channel emission depends in a different way on these parameters.

In the next, the procedure to obtain the data sets to train MLPs is explained. After, the results of different experimental simulations are shown. In an initial approach, the whole spectrum variables are used as input to MLPs. The refined approach try to reduce the input dimensionality of the network using the pick peak selector to make a feature selection.

The performance of different approaches is measured, in one hand, in terms of mean square error over the training and test data sets and, in other hand, in terms of average temperature error per profile and for the hottest cell. The hottest cell is used as criterion because its retrieval is the most difficult due to the fact that energy emitted by this cell is absorbed by the others which behave as a mask.

2.1 Experimental Data Sets

The data set is composed of large number of synthetic emission spectra generated with a computer code developed at University Carlos III (CASIMIR)[15] based on the well known HITRAN/HITEMP[14] spectral database. The total number of cases simulated are 1040 covering many possible sceneries of a typical flame combustion. Data set generation has been performed under the following assumptions:

- Synthetic spectra will correspond to energy emission of hot gas cloud of width L . Temperature and gas concentrations present gradients inside the cloud.
- The spectral range selected for this data set is 2110cm^{-1} - 2410cm^{-1} . In this spectral range, the CO_2 emission band is by far the most important emission feature, being the water emission nearly negligible. Due to this fact, only the emission associated to the CO_2 will be considered.

- For retrieval scheme, we have used an spatial discretization with five cells of equal width ($L/5$). Each cell has an average value of temperature and gas concentration.
- The concentration profiles for carbon dioxide and water vapour will keep unchanged for the whole data set. Numerical values for these concentrations have been selected from typical combustion experiments.
- Four basic temperature profiles have been chosen to simulate different temperature gradient. The step between the temperature of two consecutive synthetic flames is $\Delta T = 50$ K, with a variation in the hottest cell between 540 K and 1140 K. And for each of these variations of temperature, a variation of cell's length is done. These variations have an step of $\Delta w = 0.02$ meters for each cell which means a total step variation $\Delta W = 0.1$ meters, covering a range between 0.1 and 2 meters. These profiles have been adjusted to a spatial discretization of five cells. All the value ranges for temperature and length have been chosen to be representative for hot gases clouds associated to fossil fuel combustion.
- Experimental noise of spectra has not been simulated in order to extract pure features associated with physical characteristics.

2.2 Initial Approach: Using the Whole Frequency Spectrum

In a first phase, the MLP has been designed with a number of inputs equal to the spectrum of energy dimensionality. Such that, each input neuron is associated with an energy channel value. In this case the experiments have been done with 4000 dimensions (high spectral resolution) which implies 4000 input neurons. As it has been explained in Sect. 2.1 the temperature profile to retrieve is discretized in five cells, so the output layer will have six neurons, one for each temperature cell and another one for the total length¹. We must include the length because of both parameters, temperature and optical depth or total length, have influence in the composition of the spectrum in agreement to the Beer's law.

Different architectures of MLP varying the number of hidden neurons have been trained until to reach the minimum value in validation error, not allowing overfitting. In Table 1, the mean square error over the trained and test data for different architectures are shown. The table also included the mean error per profile, the mean error on the hottest cell and standard deviation.

In any of the architectures tested the results obtained have not been good. The MLP converges in any case without good results and there is not almost difference although the best results are with 30 hidden neurons. Also could be seen how the train error and the test error are quite different possibly due to the input dimension and the bad generalization due to the ratio number of samples and high dimensionality.

2.3 Refined Approach: Feature Selection for Dimension Reduction Based on PCA

In order to introduce a priori knowledge of the problem in the MLP learning process and improve the results, a reduction using feature selection approach has been assumed.

¹ We are assuming in the discretization that all the cells have the same length so we do not include one per each cell.

Table 1. Errors for the hot gas temperature retrieval using a MLP with 4000 inputs and different architectures

Hidden neurons	MSE Train	MSE Test	Mean error per profile (K)	Mean error hottest cell (K)	Standard Deviation
10	0.00418	0.02817	21.65	19.20	16.40
20	0.00178	0.03224	21.08	19.13	14.78
30	0.00099	0.02871	18.80	17.08	13.22
60	0.00989	0.03208	25.88	21.89	17.20

This reduction try to conserve all the information of any possible scenery in a few original variables stressing the importance of that wavelengths whose influence in the temperature and length profile are important. To make this selection PCA analysis has been applied. The central idea of PCA is to reduce the dimensionality of data set where there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables[10]. The new base is composed of a set of axes which will be orthogonal between them and are calculated as a lineal combination of the old base.

Also a dimension reduction could be done using the projections of the original data over this new base, but during the experiments realized the results obtained have been always worst than with a feature selection method and consequently has been rejected. Let $C = \{e_1^M, \dots, e_n^M\}$ be a data set of n spectrum of dimension M variables. Let Σ be the covariance matrix of the data set C with dimension $M \times M$. Let V the $M \times M$ matrix with columns equal to eigenvectors of Σ and let L be the diagonal $M \times M$ matrix with the M associated eigenvalues (by definition $\Sigma \cdot V = V \cdot L$).

Table 2. Cumulated percentage of variance for spectrum data set generated

Number of PCA components	Cumulated variance
1	95.15
2	98.20
3	99.04
4	99.42
5	99.58
...	...
14	99.90

The selection of m specific channels from M variables where $m \ll M$, allows to work with lower dimensionality. This m subset of variables contains virtually all the information available in M variables. The problem then is to find the value of m , and to decide the subset or subsets of m variables are the best. Here we want to find those variables which best represent the internal variation of C to find out which channels are significant (feature selection). In other cases the linear correlation between PCs

and channels are used to interpret the physical meaning[16], or to get a first retrieval approximation[17]. To resolve the question about how many m variables we have to consider, we must check the number of PCs that account for most of the variation in a spectrum e_x of the data set C . This can also be interpreted as finding the effective dimensionality of e_x . If e_x can be successfully described by only m PCs, then it will often be true that M can be replaced by a subset m (or perhaps slightly more) variables, with a relative small loss of information[10].

The results of this analysis as cumulative percentage of variance are shown in Table 2. Between thirteen and sixteen principal component, around the 99.9% of the total variation it is covered and the spectrum could be reconstruct almost without error. Furthermore we have visualized the projections of the data set for this first principal components trying to find clusters and we have found that with five PCs we can do a first approximation clustering (k-means) by temperature and total length. It means that the projections in these first five PCs have information about temperature and length scales.

The results obtained for the temperature retrieval with typical selection methods as B2(backward selection) and B4(forward selection) were not successful. Thus, to select a subset of variables from this first five PCs, a pick peak selector has been proposed trying to search for the most important groups coefficients of each eigenspectrum² in absolute value. This pick peak selector firstly look for all the possible peaks in each eigenspectrum. Then a threshold is established, and every peak greater than this threshold will be chosen. This will allow to catch not only the most relevant information of each eigenspectrum but the mainly second, third, . . . and carry on most important variables either. To limit the number of variables to select the threshold could be adapted to adjust it.

From a physical point of view each of these eigenspectrums does not contain only information about the temperature of one area of the gas, but also have information about the temperature and the spatial distribution over the different channels. Thus, the pick peak selector tries to get not only variables from one channel area but spread this selection over all the eigenspectrum guiding the selection.

The first 6 eigenspectrums can be seen in Fig. 1. Each one gives specific information about different channels. Because we are not trying to get the best reduction, a permissive threshold has been assumed to avoid taking out any relevant channel. Finally the number of channels selected have been 86 which are the inputs for the MLP. As in the previous approach, different number of hidden neurons has been tested with these input channels, and the results are shown in Table 3.

This refined approach improve quantitatively the results obtained using all the spectra channels. The mean temperature error retrieval is 3.36 K and 2.81 K in the hottest area for the best case. In the initial approach (see Table 1) the train error tends to decrease to 0 however the test error tends to level off which implies a bad performance in the MLP learning process. In this refined approach both errors tends to decrease together which means that the learning and generalization processes are working well and also the results obtained are better.

² Eigenspectrum is the eigenvector matrix which corresponds to the Σ matrix of the spectrum data set.

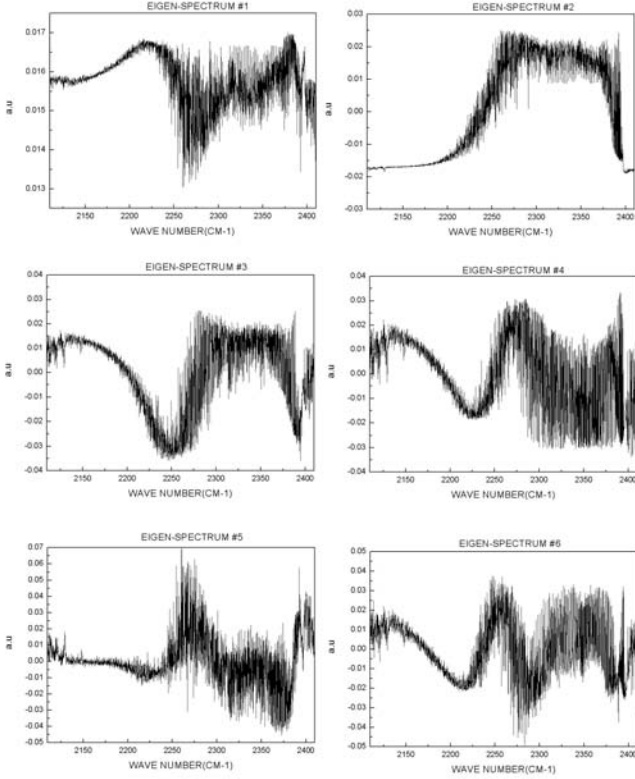


Fig. 1. First 6 eigenspectra, infrared region (2110cm^{-1} - 2410cm^{-1})

Table 3. Errors for the hot gas temperature retrieval using a MLP with 86 inputs and different architectures

Hidden neurons	MSE Train	MSE Test	Mean error per prof le (K)	Mean error hottest cell (K)	Standard Deviation
10	0.00058	0.00067	6.82	5.10	4.60
20	0.00021	0.00033	3.94	2.95	3.56
30	0.00024	0.00029	3.36	2.81	2.90
40	0.00022	0.00040	3.72	3.14	3.18

3 Discussion and Conclusions

We have presented here an efficient approach to prevent the problems related with the use of high spectral resolution measurements jointly with neural networks for inversion of radiative transfer equation in combustion processes.

All the samples data set of typical combustions emission gases used in this paper has been generated with CASIMIR[15] at high spectral resolution (0.05cm^{-1}). For this temperature sounding the range selected has been $2110 - 2410\text{cm}^{-1}$ and the number

of inputs with the resolution mentioned is 4000. This high number of inputs is a serious drawback to the use of neural networks in this framework.

The spectral feature selection has been supported by the statistical technique principal component analysis which has allowed a reduction of dimensionality of factor 46, choosing those channels specially important for temperature sounding profiles.

The results presented in Sect. 2.2 with 4000 input variables without any previous treatment over data are not acceptable. The same results are reproduced for different architectures of MLP without any sensible improvement. The best results obtained in this approach have been 18.80 K mean error over the whole profile and 17.08 K on the hottest cell with a standard deviation of 13.22 K. Reference values in literature show that a temperature measurement with an error of 0.5–1% is considered an accurate measurement in the range of 1200–2000 K. The DT025 sensor is a commercial "intrusive" thermocouple with a maximum error of 0.5%. This sensor is considered to be "extremely accurate".

The digital image system[3]also has relative errors no greater than 1% in the range 1280–1690 °C. If the results for this initial approach are compared with the intrusive methods we can conclude that they are not acceptable since the relative medium error of 2.2% is sensitive bigger than 1.0%. In this case the use of neural networks as inverse model with 4000 inputs is not useful.

With the refined approach the input dimensionality space is much smaller and the results obtained have improved significantly the previous ones. The best results have a relative error on the hottest cell of 0.34% which are in the same magnitude that the ones obtained with the intrusive methods mentioned above but with the advantages explained in the introduction, as it is not intrusive and can be applied in harsh environments.

From this physical point of view these results are acceptable, and we can conclude that neural networks techniques can be applied successfully as inverse model for this problems with a priori treatment of the data. This previous treatment reduce the dimensionality of the inputs conserving the important data and limits the grades of flexibility. As consequence the MLP convergence is faster and better as has been showed in this paper.

The use of high spectral feature selection jointly with neural networks can contribute in an efficient way to retrieve the temperature profile with some advantages over classical physical-statistical techniques as speed and generalization.

Acknowledgments

This article has been financed by the Spanish founded research MEC project OPLINK::UC3M, Ref: TIN2005-08818-C04-02.

References

1. Romero, C., Li, X., K.S., Rossow, R.: Spectrometer-based combustion monitoring for flame stoichiometry and temperature control. *Appl. Therm. Eng.* **25** (2005) 659–676
2. Thakur, M., Vyas, A., Shakher, C.: Measurement of temperature and temperature profile of an axisymmetric gaseous flames using laser phase interferometer with linear gratings. *Opt. Laser Eng.* **36** (2001) 373–380

3. Lu, G., Yan, Y., Colechin, M.: A digital imaging based multifunctional flame monitoring system. *IEEE T. Instrum. Meas.* **53** (2004) 1152–1158
4. Liu, L.H., Jiang, J.: Inverse radiation problem for reconstruction of temperature profile in axisymmetric free flames. *J. Quant. Spectrosc. Radiat. Transfer* **70** (2001) 207–215
5. McCormick, N.J.: Inverse radiative transfer problems: a review. *Nuclear Science and Engineering* **112** (1992) 185–198
6. Eyre, J.R.: Inversion methods for satellite sounding data. Lecture Notes NWP Course. European Centre for Medium-Range Weather Forecasts (ECMWF) (2004)
7. Bishop, C.M.: *Neural Networks for Pattern Recognition*. Oxford University Press (1999)
8. Pearson, K.: On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2** (1901) 559–572
9. Hotelling, H.: Analysis of a complex of statistical variables into principal components. *Educ. Psychol.* **24** (1933) 417–441. 498–520
10. Jolliffe, I.T.: *Principal Component Analysis* (2nd Ed.). Springer Series in Statistics Springer-Verlag, New York (2002)
11. Cybenko, G.: Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals, and Systems* **2** (1989) 303–314
12. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. *Neural Networks* **2** (1989) 359–366
13. Aires, F., Chédin, A., Scott, N.A., Rossow, W.B.: A regularized neural net approach for retrieval of atmospheric and surface temperatures with the iasi instrument. *Journal of Applied Meteorology* **41** (2001) 144–159
14. Rothman, L.S.: The hitran molecular spectroscopic database: edition of 2000 including updates through 2001. *J. Quant. Spectrosc. Radiat. Transfer* (2003)
15. García-Cuesta, E.: *CASIMIR: Cálculos Atmosféricos y Simulación de la Transmítancia en el Infrarrojo*. University Carlos III L/PFC 01781, Madrid (in Spanish) (2003)
16. Huang, H.L., Antonelli, P.: Application of principal component analysis to high-resolution infrared measurement compression and retrieval. *J. Clim. Appl. Meteorol.* **40** (2001) 365–388
17. Aires, F.: Remote sensing from the infrared atmospheric sounding interferometer instrument. *J. Geophys. Res.* **107** (2002) ACH6–1–15