Nonlinear Principal Component Analysis for the Radiometric Inversion of Atmospheric Profiles by Using Neural Networks

Fabio Del Frate and Giovanni Schiavon

Abstract—A new neural network algorithm for the inversion of radiometric data to retrieve atmospheric profiles of temperature and vapor has been developed. The potentiality of the neural networks has been exploited not only for inversion purposes but also for data feature extraction and dimensionality reduction. In its complete form, the algorithm uses a neural network architecture consisting of three stages: 1) the input stage reduces the dimension of the input vector; 2) the middle stage performs the mapping from the reduced input vector to the reduced output vector; 3) the third stage brings the output of the middle stage to the desired actual dimension. The effectiveness of the algorithm has been evaluated comparing its performance to that obtainable with more traditional linear techniques.

Index Terms—Atmospheric profiling, macrowave radiometry, neural networks.

I. INTRODUCTION

THE RATIONALE for a reduction of dimensionality in remote sensing inversion problems mainly lies on the fact that, in many cases, reducing the number of input and output variables can lead to improved performances for a given data set. The mapping fixed by the data may be better specified in a lower-dimensional space, and this compensates for the loss of information [1]. For these reasons inversion algorithms for the retrieval of atmospheric profiles based on the expansion of the unknown quantities in terms of a base of principal components [2] or natural orthogonal functions (NOF) [3] have been proposed [4], [5]. A few functions describe each profile with a satisfactory level of accuracy, and therefore the retrieval process reduces to the estimation of the coefficients of the expansion.

Artificial neural networks (NN) have been recognized as being a powerful tool for remote sensing of atmosphere as well [6]. Their use in statistical estimation is often effective because they can simultaneously address nonlinear dependencies and complex statistical behavior.

A first attempt to combine the properties of both NN and NOF is reported in [7]. Atmospheric profiles were expanded on a base of natural orthogonal functions, and the coefficients of the expansion were estimated with a neural network from

The authors are with the Dipartimento di Informatica, Sistemi e Produzione, Università Tor Vergata, I-00133 Roma, Italy (e-mail: delfrate @disp.uniroma2.it, schiavon@disp.uniroma2.it).

Publisher Item Identifier S 0196-2892(99)07195-8.

the brightness temperatures measured by a seven-channel microwave radiometer. The results show that the combined algorithm, if compared with a linear regression technique, can yield a more accurate retrieval and is more flexible and robust.

In this paper, a nonlinear principal component analysis based on NN replaces the NOF algorithm. This analysis can be applied to either the input or the output vector, so that, in the inversion phase, the nonlinear principal components of the atmospheric profiles can be derived from the nonlinear principal components of the radiometric measurements. This inversion is again performed by a neural network.

II. NONLINEAR PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA), or NOF expansion, is a technique for mapping multidimensional data into lower dimensions with minimal loss of information [8]. The mapping, from R^m to R^f , has the form

$$[T] = [Y][P] \tag{1}$$

where [Y] is the $n \times m$ matrix of the original data (*n* observations, *m* variables), [T] is the $n \times f$ (f < m) matrix of the components (features), and [P] is the $m \times f$ matrix which performs the mapping. The columns of [P] are the eigenvectors corresponding to the *f* largest eigenvalues of the covariance matrix [B] ($m \times m$) of [Y]

$$[B] = \frac{1}{n} [Y]^T [Y] \tag{2}$$

solution of the eigenvalue problem

$$[B][P] = [P][\lambda] \tag{3}$$

where $[\lambda]$ is a $f \times f$ diagonal matrix whose elements are the f largest eigenvalues of [B]. These eigenvectors represent the first f natural orthogonal functions. The factorization (linear transformation) of [Y]

$$[Y] = [T][P^T] + [E]$$
(4)

is optimal in the sense that the Euclidean norm of the $n \times m$ residual matrix [E], norm([E]), is minimized for the given number of factors f. The reconstruction of the data is performed by

$$[Y'] = [Y] - [E] = [T][P^T].$$
(5)

Manuscript received November 9, 1998; revised June 7, 1999. This work was supported in part by Agenzia Spaziale Italiana (ASI).



Fig. 1. Two level temperature profile dataset distribution with linear correlation.

In nonlinear principal component analysis (NLPCA) [9], the mapping into feature space is generalized to allow arbitrary nonlinear functionalities. Analogous to (1), the mapping is expressed in the form

$$[T] = [G]([Y]) \tag{6}$$

where [G] is a nonlinear vector function, composed of f individual nonlinear functions, $[G] = \{G_1, G_2, \dots, G_f\}$, analogous to the columns of [P]. The inverse transformation, restoring the original dimensionality of the data, analogous to (5), is implemented by a second nonlinear vector function $[H] = \{H_1, H_2, \dots, H_m\}$

$$[Y] = [H]([T]). (7)$$

Analogous to PCA, the functions [G] and [H] are selected to minimize norm([E]).

If nonlinear correlation between variables exists, NLPCA will describe the data with greater accuracy and/or by fewer components than PCA. A simple example can help to understand this statement.

Consider the set of atmospheric temperature profiles, described by the two temperature values at ground level and at 500 m height, which is plotted in Fig. 1. Each point represents a different profile and some samples are drawn in Fig. 2 for reference. The points in Fig. 1 are mainly distributed along a straight line, showing little dispersion in the orthogonal direction. The PCA identifies the two principal axes drawn in the figure as dashed lines, and the use of only the first principal component for the representation of the data leads to a variance in the reconstruction error of 0.03 (versus a value of 13.79 using only the least principal component), while the variances of the original temperatures were 8.39 and 5.43, respectively.

A different set of profiles is shown in Figs. 3 and 4. In this case the points in Fig. 3 are mainly distributed around an ellipse, representing profiles with greater variability in the slope and with ground temperature inversions (Fig. 4). In



Fig. 2. Samples of profiles from the dataset of Fig. 1.



Fig. 3. Two level temperature profile dataset distribution with nonlinear correlation.

this case, again, the two principal axes are drawn as dashed lines, but a very poor approximation of the data is obtained using only one principal component, with a variance in the reconstruction error of 8.04 and 12.56 for the first and the least component, respectively. If, on the other hand, the mapping to the feature space is performed by the nonlinear transformation which maps rectangular coordinates to polar with respect to the center of the ellipse, a much better approximation is obtained using only the polar angle θ . In fact, a variance in the reconstruction error of 0.04 is obtained in this case, while the use of only the radius ρ leads to a value of 41.22. For this reason, the angle θ can be considered our first nonlinear principal component.

III. AUTOASSOCIATIVE NEURAL NETWORKS

To generate the functions [G] and [H] of previous section, a basis function approach can also be used. Cybenko [10] has



Fig. 4. Samples of profiles from the dataset of Fig. 3.

shown that functions of the following form are capable of fitting any nonlinear function [v] = f[u] to an arbitrary degree of precision

$$v_k = \sum_{j=1}^{N_2} w_{jk2} \sigma \left(\sum_{i=1}^{N_1} w_{ij1} u_i + \theta_{j1} \right)$$
(8)

where $\sigma(x)$ is any continuous and monotically increasing function with $\sigma(x) \to 1$ as $x \to +\infty$ and $\sigma(x) \to 0$ as $x \to -\infty$. A suitable function is the sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}.\tag{9}$$

Equations (8) and (9) are the describing equations for a feedforward NN with N_1 inputs, a hidden layer containing N_2 nodes with sigmoidal transfer functions, and a linear output node for each kth component of [v]. In (8), w_{ijl} represents the weight on the connection from node *i* in layer *l* to node *j* in layer l + 1. The θ are nodal biases, treatable as adjustable parameters like the weights. More details on NN of this type are given in [11]. Here, we emphasize that (8) and (9) are showing that multilayer neural networks can be used to perform nonlinear dimensionality reduction, thereby overcoming some of the limitations of NOF linear algorithm.

Consider then a multilayer perceptron whose topology has the form shown in Fig. 5, having a left-hand stage (LS) with dinputs and a mapping (hidden) layer, one bottleneck (hidden) layer with M units (M < d), and a right-hand stage (RS) with a demapping (hidden) layer and d output units. The targets used to train the network are simply the input vectors themselves, so that the network is attempting to map each input vector onto itself. Due to the reduced number of units in the central layer, a perfect reconstruction of all inputs is not possible, in general, but the network can be trained by minimizing a sum-of-squares error of the form

$$E = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{d} \{y_k(\mathbf{x}^n) - x_k^n\}^2$$
(10)

where N is the number of the training examples, $y_k(\mathbf{x}^n)$ represents the output of unit k as a function of the input vector \mathbf{x}^n , and the quantity x_k^n represents the desired value for output unit k when the input vector is \mathbf{x}^n . Such a network is said to form an autoassociative mapping. Error minimization in this case represents a form of unsupervised training, since no independent target data is provided [12]. The network can be viewed as two successive functional mappings \mathbf{F}_1 and \mathbf{F}_2 . The first mapping \mathbf{F}_1 (performed by the LS) projects the original d-dimensional data onto an Mdimensional subspace defined by the activations of the units in the central hidden layer. Because of the presence of nonlinear units, this mapping is essentially arbitrary, and in particular not restricted to being linear. Similarly the second half of the network (RS) defines an arbitrary functional mapping from the M-dimensional space back into the original d-dimensional space. The described network performs a NLPCA. It might be thought that these arbitrary mappings can be performed by a net with one hidden layer, obtained by the one of Fig. 5 removing the mapping and demapping layers, provided that the bottleneck layer has nonlinear (sigmoid) activation functions. However, it was shown by Bourland and Kamp [13] that, in this case, the PCA analysis is the one providing the minimum error on the representation of the data in a subspace with reduced dimensionality. There is not therefore any advantage in using the net with one hidden layer to perform dimensionality reduction. Another comment is that, using the NLPCA analysis, it may be necessary to train and compare several networks having different values of M to be able to specify the proper number of components to be used. A convenient strategy might be that of exploiting the PCA properties and select this number using the linear approach. The actual representation in the lower-dimensionality space can then be realized with the NLPCA technique.

Examples of autoassociative networks capable of providing good one-factor representation of data of the kind of those reported in Fig. 3 can be found in [9], [14].

IV. THE GLOBAL INVERSION PROBLEM

Although the dimensionality reduction, operated on either the input or the output vector, can lead to an optimized definition of the inversion problem, it does not perform by itself the inversion from the measured data to the desired parameters. Therefore, three stages are needed to complete the retrieval algorithm, as shown in Fig. 6. The first neural net N_1 , which is the LS of a net, as in Fig. 5, performs the dimensionality reduction of the input. This net, processing the input data, provides the mapping \mathbf{F}_1 described in the previous section. The middle net N2 will carry out the pure inversion phase. Its input and output units are hidden units of the neural net N given by the connection of networks N1, N2, and N3. The third net N_3 (RS of a net, as in Fig. 5) executes the mapping \mathbf{F}_2 of the previous section on the output of net N₂, yielding the predicted vector after having transformed its dimension to that of the original data. The topology and the weights characterizing each net are calculated separately so that the nets can be optimized for their specific different



Fig. 5. Topology of an autoassociative multilayer perceptron.



Fig. 6. Topology of the net performing the global retrieval algorithm.

task. Note that nets N_1 and N_3 can easily implement a NOF algorithm and in the same way net N_2 can implement a linear regression model so that the net N represents a general inversion procedure including linear models as special cases.

V. RESULTS

The algorithm described in the previous section has been applied to retrieve atmospheric profiles from radiometric data. Two sets (training and evaluation) of 1783 and 1662 temperature and water vapor profiles, respectively, statistically generated starting from the midlatitude summer standard atmosphere, have been used [15]. All profiles contain clouds. Realistic and physically acceptable humidity and temperature irregularities as well ground-based inversions are also included. Liebe's millimeter-wave propagation model (MPM) [16] has been used to compute the brightness temperatures as would be measured with a ground-based microwave radiometer aiming at zenith at the following seven frequencies: 22.235, 23.87, 31.65, 51.25, 52.85, 53.85, and 54.85 GHz. These values define the actual channels of a new-generation radiometer designed and developed under a European Space Agency (ESA) contract by Officine Galileo, Florence, Italy [17]. To simulate noise in the radiometric channels, random fluctuations with 0.5 K standard deviation (representative of the performance of the mentioned radiometer) have been added to the brightness temperature data. Ground measurements of surface temperature and relative humidity have also been considered.

The dimensionality reduction of the inputs (measurements) could be very useful in the case of hyperspectral and/or scanning remote sensing instruments, with hundreds or even thousands of channels which can be correlated to one another. For example, Westwater *et al.* [18] perform atmospheric temperature profiling by means of a multifrequency (7) scanning (50 angles) radiometer. In this case the degree of redundancy is considerable and the measurements dimensionality is conveniently reduced from 350 to 9 using a singular value decomposition (SVD) [19].

In our case, with only seven radiometric channels, which have been chosen to give different pieces of information, and one ground measurement, the reduction of the dimension of the input vector did not show to provide significant improvement in the inversion process. For this reason it has not been applied, while seven nonlinear principal components have been used for the output (starting from the initial number of 33 levels discretizing the profile) which guaranteed a satisfactory level of representation. This means that the net used was made up by only sections N_2 and N_3 of Fig. 6.

Using sigmoidal functions a scaling procedure is generally recommended to prevent the input values from being in the saturated regions. In this study, the input variables of the autoassociative networks have then been linearly transformed to have all values comprised in the range [-1, 1]. Similarly, being the outputs units characterized by sigmoidal functions, the output variables have been scaled to have all values varying in the range [0, 1].

Several attempts have been made to properly choose the number of hidden units to insert in the LS and RS stages of the autoassociative NN which performs the NLPCA of the atmospheric profiles. The following two topologies have been selected: 33-18-7-18-33 for the case of temperature and 33-30-7-30-33 for the vapor case. The hidden layer of net N_2 of Fig. 6 has been made up by eight units for temperature and 12 units for vapor. Therefore, the resulting topology of the net performing the global algorithm has been 8-8-7-18-33 for temperature and 8-12-7-30-33 for vapor.

The rms representation accuracies, for both temperature and vapor profiles, using seven nonlinear principal components are plotted in Figs. 7 and 8 compared to those obtained with the same number of linear principal components. The behavior of PCA and NLPCA is quite similar for temperature (Fig. 7) while NLPCA achieves a reduced rms error profile in the case of vapor (Fig. 8).

As far as the complete inversion process is concerned, some examples of retrieved temperature and vapor profiles are plotted in Figs. 9 and 10, respectively. The profiles retrieved with the technique described in this study are compared with the actual profiles and with those retrieved by means of a linear technique which estimates the coefficients of the NOF expansion (PCA) of the profiles with a linear regression,



Fig. 7. The rms representation accuracies of temperature profiles using seven NOF (dashed curve) and seven nonlinear components (solid curve).



Fig. 8. The rms representation accuracies of water vapor profiles using seven NOF (dashed curve) and seven nonlinear components (solid curve).

as already performed in [5]. In general, the accuracies of the retrievals obtained with the two different techniques are comparable, while for the extreme cases, which are those with the presence of ground inversions for temperature and of thick clouds for vapor, the nonlinear technique performs better. Observe in Fig. 9 (bottom) how the temperature profile retrieved with the nonlinear inversion approaches the value of the ground inversion more precisely than the linear technique, while the estimation is essentially equivalent with no inversion (top). In Fig. 10, two samples of vapor profiles are plotted. The bottom graph shows the improved ability of the nonlinear technique to follow sharp variations in the profile, particularly at the boundaries of a thick cloud, where the linear one tend to smooth out the retrieved profile. The estimation is comparable when low cloud liquid is present, as shown in the top graph.

The overall retrieval accuracy is reported in Figs. 11 and 12, where the profiles of rms error of retrieved temperature



Fig. 9. Examples of temperature profiles from the considered data set (solid line), linear retrieval (dotted line), and nonlinear retrieval (dashed line).

and vapor, respectively, for both the linear and the nonlinear technique are plotted. The standard deviation curves, also plotted in the figures, give the *a priori* profiling accuracy without measurements. The graphs show that the rms error values of the retrieved atmospheric quantities are smaller for NLPCA than for PCA for all levels in the profiles. This means that the nonlinear technique provides a better retrieval accuracy than that of the linear one.

Fig. 13 reports the retrieval accuracy of vapor for profiles with the highest liquid content (L > 2 mm). In this case the gap between the rms error profiles corresponding to the linear and the nonlinear retrievals has grown wider with respect to Fig. 12, confirming at an ensemble level what was already observed at single-profile level when commenting Fig. 10.

It could be observed that the nonlinear method is more complex than the linear one. In fact, more adaptive coefficients



Fig. 10. Examples of vapor profiles from the considered data set (solid line), linear retrieval (dotted line), and nonlinear retrieval (dashed line).



Fig. 11. Profiles of rms error of retrieved temperature. Solid line: nonlinear retrieval, dashed line: linear retrieval, and dotted line: standard deviation of profiles from their means.



Fig. 12. Profiles of rms error of retrieved vapor. Solid line: nonlinear retrieval, dashed line: linear retrieval, and dotted line: standard deviation of profiles from their means.



Fig. 13. Profiles of rms error of retrieved vapor in case of L > 2 mm. Solid line: nonlinear retrieval and dashed line: linear retrieval.

have to be calculated. Nevertheless, the number of these coefficients can be significantly reduced by means of appropriate pruning procedures [20], which, as a secondary effect, may also improve the algorithm estimation accuracy.

Finally, we directly estimated the 33-level profiles from the seven brightness temperatures and the ground measurements by means of a multilayer perceptron with one hidden layer. As far as the estimation accuracy is concerned, the results were very close to those obtained using the NLPCA, the latter being slightly better in terms of the rms error averaged over the profile. Conversely, the inversion carried out with the direct algorithm required a training time ten times longer than that needed for the inversion in the algorithm performing dimensionality reduction. In fact, a reduction of dimensionality of the output vector, leading to a net with fewer weights, significantly alleviates the training task.

VI. CONCLUSIONS

A NN-based algorithm for the inversion of radiometric data to retrieve atmospheric profiles has been presented. In the algorithm the pure inversion phase is combined with pre/postprocessing of the data and both tasks are performed using NN. This means that the NN potentialities are exploited either to face the nonlinearities characterizing the inversion problem or to implement a nonlinear principal component analysis of the input and of the output vectors. This latter finds and eliminates nonlinear correlations in the data and has been carried out by means of autoassociative neural networks, containing a bottleneck hidden layer and trained to learn an identity mapping. The final proposed retrieval scheme illustrates an inversion procedure operating according to the described approach but also including the linear techniques as by-product.

Results obtained with the considered new methodology have been compared with those of linear algorithms. The advantages of the neural processing are concerned either with the representation of the data in a lower-dimensionality space or with the retrieval capabilities in the global inversion problem. Particularly, the considered algorithms are more precise in the estimation of ground inversions and sharp variations in the atmospheric profiles, providing results comparable with those obtained using linear techniques when profiles are closer to those characterizing the average atmospheric conditions.

REFERENCES

- C. M. Bishop, Neural Networks for Pattern Recognition. London, U.K.: Oxford Univ. Press, 1995, ch 8.
- [2] H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psychol., vol. 24, pp. 498–520, 1933.
- [3] A. M. Obukhov, "The statistically orthogonal expansion of empirical functions," *Izv. Acad. Sci. USSR Geophys. Sci.*, pp. 288–291, 1960, English translation.
- [4] J. Askne, G. Skoog, and E. Wimberg, "Test of a ground-based microwave radiometer for atmospheric temperature profiling with meteorological applications," *Int. J. Remote Sensing*, vol. 6, pp. 1241–1256, 1985.
- [5] F. Del Frate, G. Schiavon, D. Solimini, and E. R. Westwater, "Use of natural orthogonal functions in radiometric remote sensing of atmospheric temperature, vapor and liquid profiles," in *Proc. Progress in Electromagnetics Research Symp.*, [CD-ROM], Dordrecht, The Netherlands, Norwell, MA: Kluwer, 1994, pp. 2A3/3-1–2A3/3-4.
- [6] J. H. Churnside, T. A. Stermitz, and J. A. Schroeder, "Temperature profiling with neural network inversion of microwave radiometer data," *J. Atmos. Ocean. Technol.*, vol. 11, pp. 105–109, 1994.
 [7] F. Del Frate and G. Schiavon, "A combined natural orthogonal func-
- [7] F. Del Frate and G. Schiavon, "A combined natural orthogonal functions/neural network technique for the radiometric estimation of atmospheric profiles," *Radio Sci.*, vol. 33, pp. 405–410, 1998.
- [8] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1993, ch. 6/10.
- [9] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," J. Amer. Inst. Chem. Eng. (AIChE), vol. 37, pp. 233–243, 1991.
- [10] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Contr., Signals Sys.*, vol. 2, pp. 303–314, 1989.
- [11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, pp. 318–362.
- [12] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," *Neural Networks*, vol. 2, pp. 53–58, 1989.
- [13] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biol. Cybern.*, vol. 59, pp. 291–294, 1988.

- [14] K. I. Diamantras and S. Y. Kung, Principal Component Neural Networks. New York: Wiley, 1996, ch. 6.
- [15] G. Schiavon, D. Solimini, and E. R. Westwater, "Performance analysis of a multifrequency radiometer for predicting atmospheric propagation parameters," *Radio Sci.*, vol. 28, pp. 63–76, 1993.
- [16] H. J. Liebe, G. A. Hufford, and M. G. Cotton, "Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz," in AGARD 52nd Specialists' Meeting Electromagnetic Wave Propagation Panel, Palma de Mallorca, Spain, 1993, pp. 3-1–3-10.
- [17] E. Battistelli, C. Capitani, A. Culebras, F. Del Frate, G. Schiavon, B. Arbesser-Rastburg, and J. P. Baptista, "A multifrequency mm-wave radiometer (MFR): Instrument description and retrieval algorithm," *Microwave Radiometry and Remote Sensing of the Environment*, D. Solimini, Ed. Utrecht, The Netherlands: VSP, 1995, pp. 529–536.
- [18] E. R. Westwater, Yong Han, and F. Solheim, "Resolution and accuracy of a multi-frequency scanning radiometer for temperature profiling," *Microwave Radiometry and Remote Sensing of the Environment*, P. Pampaloni, Ed. Utrecht, The Netherlands: VSP, in press.
- [19] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, ch. 2, pp. 70–73.
- [20] F. Del Frate and G. Schiavon, "Neural networks for the retrieval of water vapor and liquid water from radiometric data," *Radio Sci.*, vol. 33, pp. 1373–1386, 1998.



Fabio Del Frate received the Laurea degree in electronic engineering in 1992 and the Ph.D. degree in computer science in 1997, both from the University of Rome Tor Vergata, Rome, Italy.

From September 1995 to June 1996, he was an appointed Visiting Scientist at the Massachusetts Institute of Technology, Cambridge, and since 1998, he has been an European Space Agency (ESA) Research Fellow at the ESRIN Center, Frascati, Italy. Since 1992, his research activity has been mainly dedicated to inversion problems in active

and passive remote sensing.

Giovanni Schiavon, for a photograph and biography, see p. 968 of the March 1999 issue of this TRANSACTIONS.