Using neural networks to describe tracer correlations

D. J. Lary1,2,3, M. D. Müller1,4, and H. Y. Musa3

1 Global Modelling and Assimilation Office, NASA Goddard Space Flight Center, USA
2 GEST at the University of Maryland Baltimore County, MD, USA
3 Unilever Cambridge Centre, Department of Chemistry, University of Cambridge, UK
4 National Research Council, Washington DC, USA

Received: 8 October 2003 - Accepted: 10 November 2003 - Published: 12 November 2003
Correspondence to: D. J. Lary (david.lary@umbc.edu)

Abstract

Neural networks are ideally suited to describe the spatial and temporal dependence of tracer-tracer correlations. The neural network performs well even in regions where the correlations are less compact and normally a family of correlation curves would be required. For example, the CH4-N2O correlation can be well described using a neural network trained with the latitude, pressure, time of year, and CH4 volume mixing ratio (v.m.r.). In this study a neural network using Quickprop learning and one hidden layer with eight nodes was able to reproduce the CH4-N2O correlation with a correlation coefficient of 0.9995. Such an accurate representation of tracer-tracer correlations allows more use to be made of long-term datasets to constrain chemical models. Such as the dataset from the Halogen Occultation Experiment (HALOE) which has continuously observed CH4 (but not N2O) from 1991 till the present. The neural network Fortran code used is available for download.

1 Introduction

The spatial distributions of atmospheric trace constituents are in general dependent on both chemistry and transport. Compact correlations between long-lived species are well-observed features in the middle atmosphere, as for example described by Fahey et al. (1989); Plumb and Ko (1992); Loewenstein et al. (1993); Elkins et al. (1996); Keim et al. (1997); Michelson et al. (1998); Rinsland et al. (1999); Strahan (1999); Fischer et al. (2000); Muscari et al. (2003). The correlations exist for all long-lived tracers – not just those which are chemically related. The tight relationships between different constituents have led to many analyses where measurements of one tracer are used to infer the abundance of another tracer. These correlations can also be used as a diagnostic of mixing (Schoeberl et al., 1997; Morgenstern et al., 2002) and to distinguish between air-parcels of different origins (Waugh and Funatsu, 2003).

Of special interest are the so-called “long-lived tracers”: constituents such as ni-
trous oxide (N₂O), methane (CH₄), and the chlorofluorocarbons (CFCs) that have long lifetimes (many years) in the troposphere and lower stratosphere, but are destroyed rapidly in the middle and upper stratosphere.

The correlations are spatially and temporally dependent. For example, there is a "compact-relation" regime in the lower part of the stratosphere and an "altitude-dependent" regime above this. In the compact-relation region, the abundance of one tracer is uniquely determined by the value of the other tracer, without regard to other variables such as latitude or altitude. In the altitude-dependent regime, the correlation generally shows significant variation with altitude (Minschwaner et al., 1996) (Fig. 1d).

The description of such spatially and temporally dependent correlations are usually achieved by a family of correlations. However, a single neural network is a natural and effective alternative.

2 Motivation

The motivation for this study was preparation for a long term chemical assimilation of Upper Atmosphere Research Satellite (UARS) (Reber et al., 1993) data starting in 1991 and coming up to the present. For this period we have continuous version 19 data from the Halogen Occultation Experiment (HALOE) (Russell et al., 1993) but not observations of N₂O as both ISAMS and CLAES failed. In addition we would like to constrain the total amount of reactive nitrogen, chlorine, and bromine in a self-consistent way. Tracer correlations provide a means to do this by using HALOE CH₄ observations.

3 Neural Networks

Computational neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between elements (Fig. 1b).

Commonly neural networks are trained so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are used, in this supervised learning, to train a network. Batch training of a network proceeds by making weight and bias changes based on an entire set (batch) of input vectors. Incremental training changes the weights and biases of a network as needed after presentation of each individual input vector. Incremental training is sometimes referred to as "on line" or "adaptive" training.

Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems. It is well established that multilayer feedforward networks are universal approximators (Hornik et al., 1989; Castro and Delgado, 1996; Ying, 1998).

In this study we use neural networks (Peterson et al., 1994) to describe the temporal and spatial dependence of tracer correlations (Fig. 1). To find the optimum neural network configuration a range of network architectures were considered containing between one and two hidden layers with between one and sixteen nodes in each hidden layer. A range of updating procedures were also used including back-propagation, Manhattan learning, Langevin Learning, Quickprop and Rprop. Each network was trained for 10⁶ epochs. The details of the different learning methods can be found in (Peterson et al., 1994). A variety of activation functions were used. Non-linear activation functions performed best, and the most successful is shown below in Eq. (1).

To determine which network architecture and updating procedure was most suitable each configuration was tried in turn and the correlation coefficient between the actual solution and the neural network solution were computed (the correlation coefficient being a normalized measure of the linear relationship strength between variables). The configuration with the highest correlation coefficient was chosen.
3.1 The CH$_4$-N$_2$O Correlation

Fig. 1a shows the CH$_4$-N$_2$O correlation from the Cambridge 2D model (Law and Pyle, 1993a,b) overlaid with a neural network fit to the correlation. The neural network used was a feed-forward multilayer perceptron type with Quickprop learning (Peterson et al., 1994). There were four inputs, one output, and one hidden layer with eight nodes. A non-linear activation function was used, namely

$$g(x) = \frac{1}{1 + \exp(-2x)}$$  \hspace{1cm} (1)

The training dataset contained 1292 patterns, sampling the input space completely as shown in Fig. 1. The network was constrained for $10^6$ epochs (iterations).

The correlation coefficient between the actual solution and the neural network solution was 0.9995. Figure 1 panel (b) shows how the median fractional error of the neural network decreases with epoch (iteration). Both CH$_4$ and pressure are strongly correlated with N$_2$O as can be seen in panels (c) and (d). Latitude and time are only weakly correlated with N$_2$O as can be seen in panels (e) and (f). Even though the correlation with time of year and latitude is relatively weak it still does play a role in capturing some of the details of the CH$_4$-N$_2$O correlation in Panel (a).

A polynomial or other fit will typically do a good job of describing the CH$_4$-N$_2$O correlation for high values of CH$_4$ and N$_2$O. However, for low values of CH$_4$ and N$_2$O there is quite a spread in the relationship which a single curve can not describe. This is the altitude dependent regime where the correlation shows significant variation with altitude (Minschwaner et al., 1996).

Figure 1c shows a more conventional fit using a Chebyshev polynomial of order 20. This fit was chosen as giving the best agreement to the CH$_4$-N$_2$O correlation after performing fits using 3687 different equations. Even though this is a good fit the spread of values can not be described by a single curve. However, a neural network trained with the latitude, pressure, time of year, and CH$_4$ volume mixing ratio (v.m.r.) (four inputs) is able to well reproduce the N$_2$O v.m.r. (one output), including the spread for low values of CH$_4$ and N$_2$O.

3.2 Scaling

Variable scaling often allows neural networks to achieve better results. In this case all variables were scaled to vary between zero and one. If the initial range of values was more than an order of magnitude then log scaling was also applied. In the case of time of year the sine of the fractional time of year was used to avoid a step discontinuity at the start of the year.

4 Conclusions

Neural networks are ideally suited to describe the spatial and temporal dependence of tracer-tracer correlations. Even in regions when the correlations are less compact. Useful insight can be gained into the relative roles of the input variables from visualizing the network weight assignment. The neural network Fortran code used is available for download.

Acknowledgements. It is a pleasure to acknowledge: NASA for a distinguished Goddard Fellowship in Earth Science; the Royal Society for a Royal Society University Research Fellowship; the government of Israel for an Alon Fellowship; the NERC, EU, and ESA for research support.

References


Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359-366, 1989. 4


Strahan, S. E.: Climatologies of lower stratospheric NOx and O3 and correlations with N2O based on in situ observations, J. Geophys. Res. (Atmos.), 104, 30 463-30 480, 1999. 2


Ying, H.: General takagi-sugeno fuzzy systems with simplified linear rule consequent are universal controllers, models and filters, Information Sciences, 108, 91-107, 1998. 4
Quickprop NN with 1 layer and 8 nodes

![Graph showing the relationship between CH₄ and N₂O](image)

Fig. 1. The neural network used to produce the CH₄-N₂O correlation in Panel (a) used Quickprop learning and one hidden layer with eight nodes. The correlation coefficient between the actual solution and the neural network solution was 0.9995. Panel (b) shows how the median fractional error of the neural network decreases with epoch (iteration). Both CH₄ and pressure are strongly correlated with N₂O as can be seen in panels (c) and (d). Latitude and time are only weakly correlated with N₂O as can be seen in panels (e) and (f). Even though the correlation with time of year and latitude is relatively weak it still does play a role in capturing some of the details of the CH₄-N₂O correlation in Panel (a).

Fractional Error for Quickprop NN with 1 layer and 8 nodes

![Graph showing the fractional error over epochs](image)

Fig. 1. (b) Continued.
Fig. 1. (c) Continued.

Fig. 1. (d) Continued.
Fig. 1. (a) Continued.

Fig. 1. (f) Continued.