

# NONLINEAR INTERPOLATION OF MISSING POINTS WITHIN THE HOURLY *FOF2* TIME SERIES

N M Francis<sup>1</sup>,  
A G Brown<sup>2</sup>,  
P S Cannon<sup>1</sup>,  
D S Broomhead<sup>3</sup>.

<sup>1</sup>Radio Science And Propagation Group,  
<sup>2</sup>Signal and Information Processing Group,  
Defence Evaluation and Research Agency,  
Malvern, Worcestershire, WR14 3PS, UK.  
Email: nmfrancis@dera.gov.uk

<sup>3</sup>Department of Mathematics, UMIST, P.O. Box 88,  
Manchester, M60 1QD, UK.

## ABSTRACT

This paper proposes a novel technique for the prediction of solar-terrestrial data sets that contain a significant proportion of missing data points. Radial Basis Function (RBF) Neural Networks (NN) are adopted for this purpose and their advantages over their Multi-Layer Perceptron (MLP) counterparts are outlined.

## 1.0 INTRODUCTION

Non-contiguous data sets are a common problem with respect to geophysical time series. Simple linear interpolation is undesirable because it can disrupt the delicate non-linear dynamics within the data, thereby degrading the performance of the modelling process. A non-linear interpolative scheme is presented that minimises the effect of interpolation upon any given modelling process. An interpolated model has been implemented for the prediction of hourly foF2 values from 1971 to 1973 for Slough station, UK. The time series as a whole contains 6.6% missing data points. The first 23,000 points have been used to train the model, while the remaining 3304 points were used to test the ability of the model to generalise to unseen data. The effectiveness of the model has been assessed relative to the performance of a non-interpolated model, as well as persistence and recurrence interpolated models, as shown in figure 1.

## 2.0 ANALYSIS TECHNIQUE

An RBF NN has been employed to provide the core modelling capability for the interpolation process [Francis *et al.*, 1998]. The most important advantage of this network over the more commonly used Multi-Layer Perceptron (MLP) architecture is the ability of the RBF to find a globally optimum solution, to a time series prediction problem, in a single pass process. MLPs can only produce locally optimum solutions though an iterative process.

Figure 2 details the basic, non-interpolated modelling process, which acts as a bootstrap for the subsequent interpolation of missing data points. Figure 3 details the interpolation process. Two iterations of the interpolation process were sufficient to achieve the optimal results. Simple nonlinear techniques, such as interpolating missing points using the prediction output from the RBF NN, were not found to give stable solutions.

Three simple reference interpolation schemes were adopted for the purpose of comparison with the nonlinear gap-filling method. The first technique, the zero interpolation scheme, sets all missing points in the de-measured and normalised time series to a value of zero. The second uses persistence to fill in missing data points. The third and final technique uses the 24-hour recurrent structure of the hourly foF2 time series to replace missing points. In addition, the complete vector only RBF model is also included. However, the complete vector test set is much smaller than the interpolated vector test set, due to incomplete vector rejection. This means that the results are not directly comparable with the other techniques.

## 3.0 RESULTS

Figure 1 shows a graph of root mean square (RMS) error, on the interpolated test vector set, plotted against number of hours ahead for the prediction, for each of the prediction schemes. Using any of the proposed interpolation schemes, the vector rejection rate falls from 60% to 6.6%. Residual rejection occurs because training vectors are still rejected if there is a gap at the corresponding point in the required output time series. The output time series is identical to the input time series for a self-prediction problem.

The graph clearly shows that the nonlinear interpolation technique consistently produces the best results in terms of relative predictive accuracy. It shows a 2.3% improvement over the recurrence model, which characterises the dominant diurnal variation of foF2, for

each of the one to thirty hour ahead predictions. In turn, the recurrence model shows a similar improvement relative to the persistence model. The zero interpolation model falls significantly behind these reference models, indicating the necessity of adopting an approach to deal with the problem of missing data points in geophysical time series.

It is interesting to note that the interpolated model produced lower prediction errors, relative to the complete vector only model, when tested on the complete vector only test set instead of the interpolated vector test set. This shows the positive benefits that can be accrued, in terms of improved network training, through interpolation of the input time series. This point is further strengthened by the fact, on the same test set, the zero interpolation model performed less accurately than the complete vector only model. It would appear that rejecting incomplete data vectors might be preferable to including them in a model, if incomplete vectors provide the network with poor training material.

#### 4.0 CONCLUSIONS

In summary, this paper describes a novel and general scheme to deal with the problem of non-contiguity in geophysical time series. This scheme can be used to improve the viability and accuracy of the underlying non-interpolated predictive model, which is currently being used to provide real-time predictions. This baseline nonlinear scheme provides a significant improvement upon simple interpolation schemes that rely upon persistence or recurrence properties of the observed ionospheric parameter.

Currently, the interpolated model parameters are the optimal parameters for the complete vector only model, due to processing constraints. By re-optimising the model parameters at every step, it is likely that further significant gains in predictive accuracy can be made.

Finally, in instances where simple climatological models cannot be used to provide sensible interpolations of missing points, this technique will find the best interpolation solution for that particular problem. In such a case, the resultant improvement may be greater than for the prediction of foF2, where the dominant diurnal variation of this ionospheric parameter is closely modelled by a simple recurrence model.

© British Crown Copyright 1998/DERA

Published with the permission of the controller of Her Britannic Majesty's Stationary Office.

#### 5.0 REFERENCES

Brown, A. G., N. M. Francis, D. S. Broomhead, P. S. Cannon, A. Akram, 'Is there Evidence for the Existence of Nonlinear Behaviour within the Interplanetary Solar Sector Structure?', *Accepted by JGR*, 1998.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Prediction of the Ionospheric Parameter, foF2, on hourly, daily and monthly time-scales', *Submitted to JGR*, 1998.

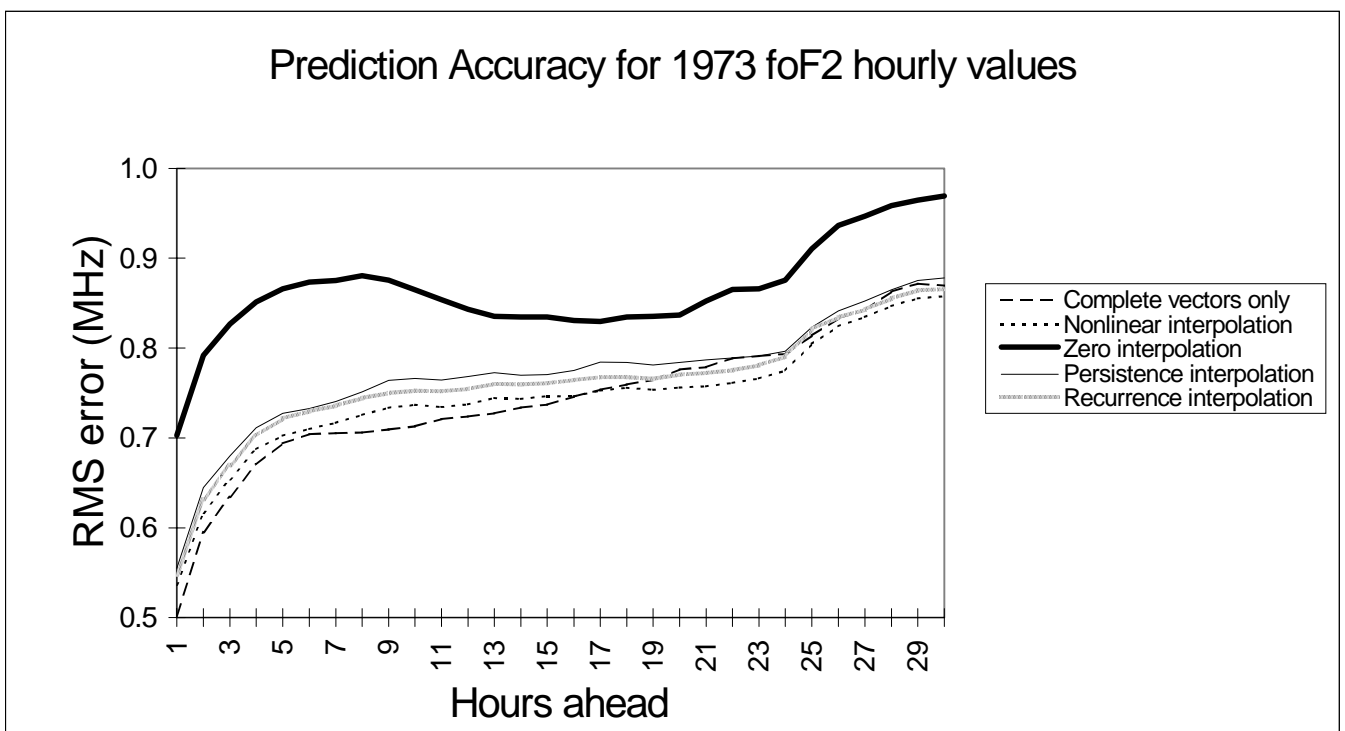
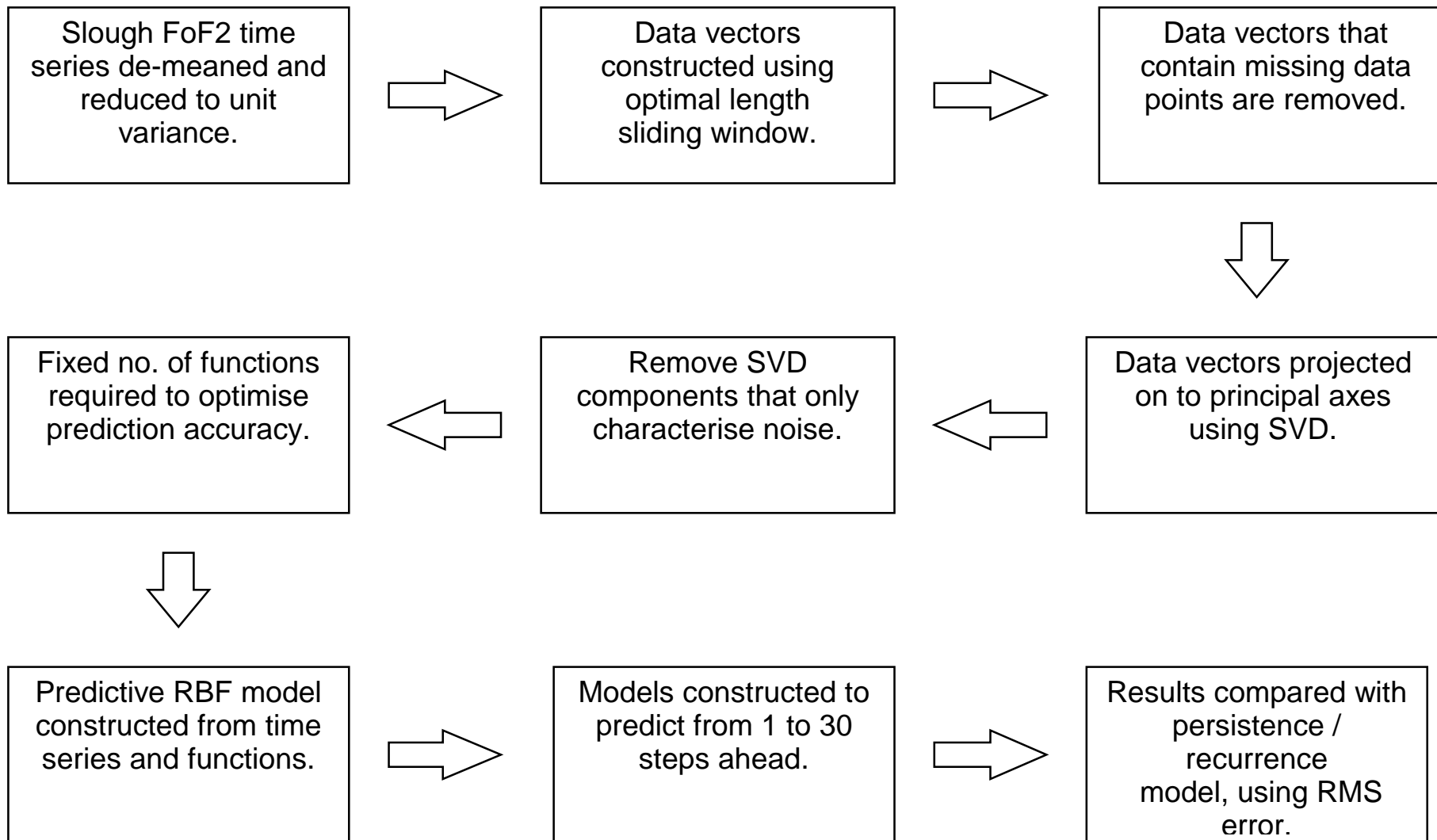


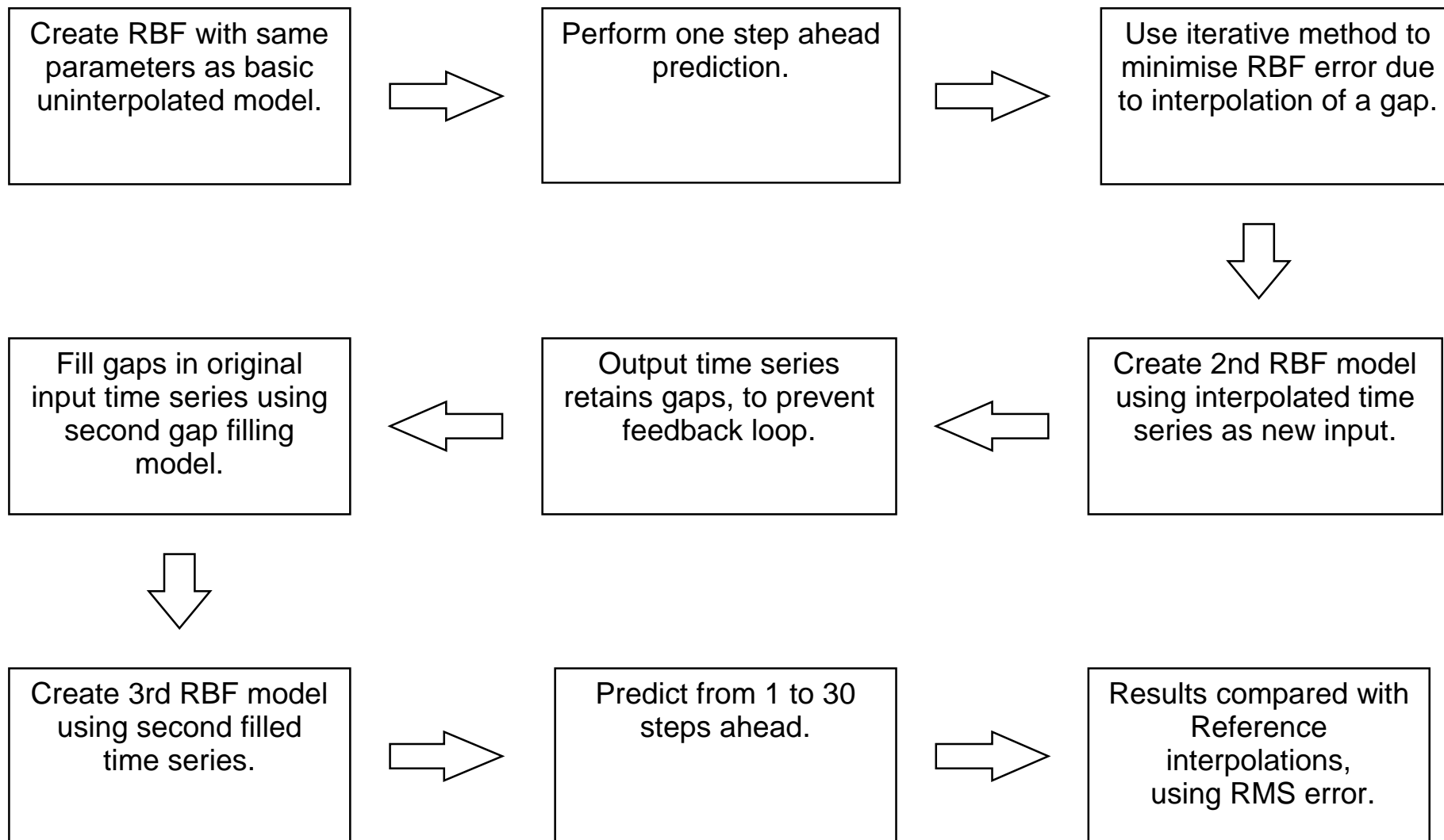
Figure 1. Relative Performance of Interpolation Techniques







**Figure 2.** Flowchart Representing Non-interpolated Modelling Process



**Figure 3.** Flowchart Representing Interpolation Modelling