NONLINEAR GAS TURBINE MODELLING: A COMPARISON OF NARMAX AND NEURAL NETWORK APPROACHES

N. Chiras, C. Evans and D. Rees
School of Electronics, University of Glamorgan, Pontypridd, CF37 1DL, Wales, UK
Tel: 0044-1443-482817  Fax: 0044-1443-482541  e-mail: nchiras@glam.ac.uk

In this paper two nonlinear modelling approaches are employed to derive single nonlinear models for a Rolls Royce aircraft gas turbine. The first approach is based on the estimation of a NARMAX model using conventional structure selection and parameter estimation techniques, and the second approach is based on the use of feedforward Multilayer-Perceptron (MLP) neural networks. The performances of the models derived by the two approaches are demonstrated using a range of engine tests and by analysing their static and dynamic behaviours.

INTRODUCTION

The modelling of gas turbine dynamics has been the subject of considerable study since the early days of jet propulsion. Gas turbines are now extensively used in aero, marine and industrial applications. With such widespread and increasing applications, the accurate modelling of such engines is an important issue.

Modelling of gas turbines is required both in the development and operational stages of an engine’s life. Models can be used to predict the engine’s performance, design and test the performance of the engine control systems and predict normal engine behaviour for on-line fault detection purposes. Models are usually validated against real data and in some cases a physical interpretation of the parameters can be made. This allows initial assumptions about the engine characteristics to be checked.

This paper deals with the nonlinear relationship between the fuel flow and shaft speed dynamics of a Rolls Royce Spey Mk202 aircraft gas turbine. Although no longer in service, the Spey possess the same characteristics, for control purposes as a modern engine such that of the EJ200 fitted to the Eurofighter¹.

THE GAS TURBINE

A gas turbine is made up of three basic components: a compressor, a combustion chamber and a turbine. Air is drawn into the engine by the compressor, which compresses it and delivers it to the combustion chamber. Within the combustion chamber the air is mixed with fuel and the mixture is ignited, producing a rise in temperature and hence an expansion of gases. These gases are exhausted through the engine nozzle but first pass through the turbine, which is designed to extract sufficient energy from them to keep the compressor rotating.

A schematic diagram of the Rolls-Royce Spey engine tested in this work is shown in Figure 1. It can be seen that both the compressor and the turbine are split into low pressure (LP) and high pressure (HP) stages. The HP turbine drives the HP compressor and the LP turbine drives the LP compressor. These are connected by concentric shafts, which rotate at different speeds, denoted \( N_{\text{H}} \) and \( N_{\text{L}} \). These shaft speeds are the primary outputs of a gas turbine, from which the internal engine pressures and thrust can be calculated. The estimation
of a nonlinear model capable of representing the relationship between the engine fuel feed and the shaft speeds throughout the engine’s operating range will be investigated in this paper.

Work conducted by Jackson\textsuperscript{2} for Rolls Royce plc, showed that the higher-order nonlinear thermodynamic models derived from the engine physics could be reduced to linear models of the same order as the number of engine shafts. The models were first linearised at a series of operating points, using small perturbations, and then a model reduction procedure was employed. Work done by Hill\textsuperscript{3} examined the application of time-domain methods to estimate discrete z-domain engine models. Even though linear discrete models with good input-output properties were estimated, problems with the application of time-domain techniques were reported\textsuperscript{4}.

More recently, teams at the universities of Glamorgan, Birmingham and Sheffield utilised the same data to investigate various identification techniques on a Rolls-Royce Spey engine. Data were gathered under sea-level static conditions at the Defence Evaluation & Research Agency (DERA) at Pyestock. Multisine and inverse repeat maximum length binary sequences (IRMLBS) were used at amplitudes of up to ±10% of the steady state fuel flow ($W_f$). In addition higher-amplitude signals, such as triangular waves and three-level periodic signals with input amplitudes of up to ±40% $W_f$, were used. These caused the HP shaft speed to vary between 65% and 85% of its maximum value (%N_H)

Evans et al.\textsuperscript{5,6,7} from the University of Glamorgan, used frequency techniques on the multisine and IRMLBS data to estimate linear models at different operating points. The errors due to noise and nonlinearities were assessed and found to be small for these small-signal models. The fact that the dc gains and the dynamics of these models change with operating point showed that the gas turbine is nonlinear, so the need is apparent for a more complete gas turbine description using nonlinear models.

In Evans et al.\textsuperscript{8} an extended least-squares algorithm with optimal smoothing was used by Norton, to identify time-varying transfer function models to represent large transient and non-equilibrium effects and provide a more detail insight into the slow thermal dynamics of the engine. In order to identify a model capable of representing the engine at all operating points, Rodriguez\textsuperscript{9,10} used a multiobjective genetic programming approach on the same data and allocated weights to various objectives, to assess their significance in the structure selection of Nonlinear AutoRegressive Moving Average with eXogenous inputs (NARMAX) models of the engine. A simple NARX model was identified which was able to represent both the small and large signal dynamics of the engine.

In this paper two nonlinear modelling approaches are employed to derive single nonlinear models for a Rolls Royce aircraft gas turbine. The first approach is based on the estimation of a NARMAX model using conventional structure selection and parameter estimation techniques, and the second approach is based on the use of feedforward Multilayer-Perceptron (MLP) neural networks. The performances of the models derived by the two approaches is demonstrated using a range of engine tests and by analysing their static and dynamic behaviours

**NARMAX MODELLING**

Leontaritis and Billings\textsuperscript{11} and Chen and Billings\textsuperscript{12}, introduced the NARMAX approach as a means of describing the input-output relationship of a nonlinear system. The model represents the extension of the well-known ARMAX model to the nonlinear case, and is defined as

$$y(k) = F(y(k-1),...,y(k-n_y),u(k-1),...,u(k-n_u),e(k-1),...,e(k-n_e)) + e(k)$$  \hspace{1cm} (1)
where $F$ is a nonlinear function; $y(k)$, $u(k)$ and $e(k)$ represent the output, input and noise signals respectively, and $n_y$, $n_u$, and $n_e$ are their associate maximum lags. The NARMAX representation constitutes a powerful tool for nonlinear modelling since it includes a family of other nonlinear representations such as block-structured models and Volterra series. In addition, the model is linear-in-the-parameters so that linear least-squares parameter estimation techniques can be easily applied. A well established procedure for structure selection of a NARMAX polynomial model is based on the error reduction ratio (ERR) defined in Billings et al. as

$$\text{ERR}_i = \frac{g_i^2 \sum_{k=1}^{N} w_i^2(k)}{\sum_{k=1}^{N} y^2(k)}$$  \hspace{1cm} (2)$$

where $g_i$ are the coefficients and $w_i(k)$ are the terms of an auxiliary model constructed in such a way that the terms $w_i(k)$ are orthogonal over the data records. A forward-regression algorithm is employed to select at each step the term with the highest ERR, in other words the term which contributes most to the reduction of the residual variance. The procedure is usually stopped using an information criterion such as the AIC, defined as

$$AIC = N \log_e \left( \sigma^2_e(\theta_p) \right) + kp$$  \hspace{1cm} (3)$$

where $\sigma^2_e(\theta_p)$ is the variance of the residuals associated with a $p$-term model and $k$ is a penalising factor. This procedure was found to work very well when identifying polynomial NARMAX models and it does exclude the vast majority of insignificant candidate terms by working towards the reduction of the residual variance. In addition, the structure selection procedure can be aided considerably by the use of a priori knowledge for the nature and approximate order of nonlinearity. In Chiras et al., time and frequency nonparametric analysis was applied to the gas turbine data to qualify the nonlinearity and approximate the maximum nonlinearity order. A part of this work concentrated on the analysis of a series of static tests throughout the operating range of the engine to model the static behaviour of the engine. It was shown that the static behaviour of the engine could be obtained using a single high-amplitude triangular test shown in Figure 2. In Figure 3(a) the spectrum of this signal is shown over a wide bandwidth, in order to show the high frequency noise at around 6 Hz and 9 Hz. This effect can also be seen in the time records. In Figure 3(b) it can also be seen that most of the input signal power is concentrated at low frequency. In this case, 97% of the total power of the input signal lies at the fundamental frequency of 0.005 Hz. Considering the fact that the –3dB bandwidth of the HP shaft is around 0.4 Hz, this test is not capable of exciting the engine dynamics but can serve as a pseudo-static test, since it can give a very good approximation of the static behaviour of the engine. This is shown in Figure 4, where the quadratic static polynomial of equation (4) was fitted to these data, to give a good approximation of the static nonlinearity in the engine, and indicate that a second-order term is sufficient to model the engine nonlinearity.

$$u(t) = 445 - 13y(t) + 0.14y^2(t)$$  \hspace{1cm} (4)$$
To this end, a concatenated data set of small signal tests is used for structure selection and parameter estimation. This is achieved by combining small-signal IRMLBS tests at different operating points to cover a range from 55% to 85% \( N_{th} \) as shown in Figure 5. The forward-regression orthogonal estimation algorithm is applied to these data and the selected terms with their associated ERRs and the coefficients of the selected model are shown in Table 1.

Table 1
Quadratic NARX Model.

<table>
<thead>
<tr>
<th>Model 1 terms</th>
<th>( \text{ERR} )</th>
<th>( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y(k-1) )</td>
<td>9.9 e-1</td>
<td>0.7453</td>
</tr>
<tr>
<td>( u(k-1) )</td>
<td>1.9 e-5</td>
<td>4.083 e-3</td>
</tr>
<tr>
<td>( y(k-2) )</td>
<td>1.3 e-5</td>
<td>0.2943</td>
</tr>
<tr>
<td>( y(k-1) \times y(k-2) )</td>
<td>3.2 e-6</td>
<td>-4.471 e-4</td>
</tr>
<tr>
<td>constant</td>
<td>8.8 e-6</td>
<td>-1.1479</td>
</tr>
</tbody>
</table>

The model quality can be assessed using higher order correlation functions\(^{16}\), but no definite conclusion can be drawn unless cross-validation is employed. Nonparametric validation is employed which consists of the simulation of the estimated model with different measured inputs and comparison of the results with the measured outputs.

Figure 6 shows a time-domain comparison between the model output and the measured gas turbine output for a small signal multisine test, where it can be seen that a very good match is achieved. Similar results are obtained when the nonlinear model is validated against small signal tests at different operating points, suggesting that the quadratic NARX model is suitable to be used in the place of the family of linear models previously estimated.

The model is also validated against high-amplitude data. The next test, shown in Figure 7, consists of a three-level sequence of period 100s with an input amplitude of \( \pm 22\% \ W_f \). This test is the most demanding available engine test since it essentially consists of a series of positive and negative large-amplitude step responses. It can be again seen that the model shows a good response suggesting that the model is capable of representing the engine dynamics at high amplitudes.

**NEURAL NETWORK MODELLING**

In this section the neural network based system identification toolbox (NNSYSID)\(^{17,18}\) for MATLAB (provided by MathWorks, Inc.) is used, to estimate a neural network model for the gas turbine. The toolbox contains a large number of functions for training and evaluating multilayer perceptron type neural networks. Several model representations closely related to linear model structures are available in the toolbox and special functions for model structure selection and validation exist so that the toolbox provides a powerful tool for nonlinear system identification. NNSYSID is independent from the System Identification and Neural Network Toolboxes (also provided by MathWorks, Inc.).

A simple definition for most types of neural networks can be formulated as in \(^{19}\): “A system of simple processing elements, neurons, that are connected into a network by a set of (synaptic) weights”. The function of the network is determined by the structure of the network, the magnitude of the weights and the mode of operation of the processing elements. The basic network element is a neuron shown in Figure 8. This is a processing element that
takes a number of inputs, applies some weights and sums them up, and feeds the result to an activation function. The inputs to the unit can be external inputs or outputs of other proceeding units. Bias inputs can also be used which represent a displacement or a constant input such as \( b \) in Figure 8. The activation function can be any kind of singular valued function, linear or nonlinear. The most popular activation functions used in system identification are the sigmoid function and the hyperbolic tangent function (tanh) shown in Figure 9. The sigmoid function is given by

\[
f_{\text{sigmoid}}(x) = \frac{1}{1 + e^{-x}}
\]  

and the closely related hyperbolic tangent function is given by

\[
f_{\text{tanh}}(x) = 2f_{\text{sigmoid}}(x) - 1
\]  

so that it makes no difference which of these two functions is used.

The most common neural network architecture and the one used in the toolbox is the feedforward multilayer perceptron network. This is constructed by ordering the units into layers, with each layer taking as input other external inputs or outputs of units in previous layers. An example of a two-layer feedforward network is shown in Figure 10 where it can be seen that the second layer produces the output, thus referred to as the output layer. The first layer is called a hidden layer since it is hidden between the external inputs and the output layer. Cybenko\textsuperscript{20} proved that a neural network with one hidden layer of sigmoidal or hyperbolic tangent units and an output layer of linear units is capable of approximating any continuous function. Based on this result the basic network architecture used in the NNSYSID toolbox consists of a two-layer feedforward neural network. The network is described by the magnitude of its weights and biases which should be determined by training the network on the estimation data. The estimation of the weights is usually a conventional estimation problem and several algorithms are implemented in the NNSYSID toolbox for this purpose.

The NNSYSID toolbox also incorporates six different model structures, five of them being motivated by their equivalence to the linear model structure. An example of the structure used in this work to model the gas turbine is given in Figure 11 where it can be seen the structure depends on the selection of the regression vector. In this case the regression vector is chosen to be of the ARMAX structure thus the name of the neural network structure is NNARMAX. It must be stressed here that this is a recurrent network since the past prediction errors depend on the model output and consequently they establish a feedback.

To this end, a NNARMAX model was trained using the data in Figure 5. In order to determine the optimum number of hidden units required to model the engine, the network was first trained using one hidden unit and the model structure was increased gradually with one hidden unit at a time. The performance of the model after each training session was evaluated using the long-term predictions on different validation data sets and calculating the sum of the squares of the prediction errors defined as

\[
V(\theta) = \frac{1}{2N} \sum_{t=1}^{N} [y(t) - \hat{y}(t | \theta)]^2
\]  

where \( \theta \) is a vector of estimated coefficients (in this case the weights), \( \hat{y}(t) \) indicates the long term predicted output and \( y(t) \) is the measured output. It is thus possible to select the most
appropriate model structure based on the performance of the model on validation data, different to the one used for estimation. Figure 12 shows the variation of this cost function with the number of hidden units. It is clear that a neural network with 8 hidden units provides the best performance on the existing validation data. Table 2 shows a comparison of the performance of this neural network against the polynomial NARX model estimated previously. It can be seen that the quadratic NARX model performs better on the high-amplitude tests and the NNARMAX model performs better on the small-amplitude tests.

Table 2
Sum of squares of the prediction errors for the two models.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Quadratic NARX model</th>
<th>NNARMAX model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multisine at 55 %N_H</td>
<td>0.0229</td>
<td>0.0110</td>
</tr>
<tr>
<td>IRMLBS at 65 %N_H</td>
<td>0.0727</td>
<td>0.0124</td>
</tr>
<tr>
<td>Multisine at 75 %N_H</td>
<td>0.0187</td>
<td>0.0163</td>
</tr>
<tr>
<td>IRMLBS at 85 %N_H</td>
<td>0.0203</td>
<td>0.0027</td>
</tr>
<tr>
<td>Three-level 58-80 %N_H</td>
<td>0.0117</td>
<td>0.0554</td>
</tr>
<tr>
<td>Triangular wave 55-85 %N_H</td>
<td>0.0968</td>
<td>0.1317</td>
</tr>
</tbody>
</table>

A better idea about the quality of the two models can be obtained if their static and dynamic behaviours are investigated. The static behaviour of the two models was derived through simulation and compared with the static behaviour of the engine derived from measured data. As it can be seen from Figure 13 the two models have similar static behaviours and they are both capable of modelling the static behaviour of the engine throughout its operating range. In order to study the dynamic behaviour of the two models at various operating points the two models were simulated with multisine signals at the four operating points for which linear models were previously estimated. Table 3 shows a comparison of the dominant modes identified from engine data using small-signal tests and frequency domain techniques, and the dominant modes identified when simulation data were generated using the two nonlinear models. It is clear from the table that the local linearisation of both models is very close to that of the engine at the different operating points.

Table 3.
Dominant modes of estimated linear models compared with those obtained by linearisation of the two nonlinear models.

<table>
<thead>
<tr>
<th>Operating point (%N_H)</th>
<th>Estimated models from engine data</th>
<th>Locally linear models derived from the quadratic NARX model</th>
<th>Locally linear models derived from the NNARMAX model</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>-0.271</td>
<td>-0.216</td>
<td>-0.246</td>
</tr>
<tr>
<td>65</td>
<td>-0.394</td>
<td>-0.373</td>
<td>-0.329</td>
</tr>
<tr>
<td>75</td>
<td>-0.525</td>
<td>-0.515</td>
<td>-0.525</td>
</tr>
<tr>
<td>85</td>
<td>-0.800</td>
<td>-0.651</td>
<td>-0.762</td>
</tr>
</tbody>
</table>

It is also clear that the neural network model resembles the dynamic behaviour of the engine at different operating points slightly better than the quadratic NARX model. This confirms the results obtained in Table 2 where it could be seen that the neural network model performed better on the small-amplitude data. This in fact suggests that the neural network model is more accurate on the estimation data, which consisted of a combination of small-amplitude
data at different operating points as shown in Figure 5. This fine accuracy is achieved at the expense of estimating a model with 8 hidden units, which correspond to 65 weights (parameters). The neural network is thus much more complex than the quadratic NARX model, which requires only 5 parameters.

It can be concluded that the two nonlinear models are capable of modelling both the high and low amplitude engine dynamics. The neural network model seems to perform better on the small-amplitude tests at the expense of increased model complexity. The quadratic NARX model provides a very simple representation of the engine dynamics and it performs better on the high-amplitude data. The results can also be extended to the LP shaft case, where nonlinear models of similar structure (but different parameter values) were estimated.

CONCLUSIONS

The nonlinear modelling of a gas turbine was discussed in this paper. In order to identify a model capable of representing the engine dynamics throughout its operating range, two model structures and identification techniques were examined. The first was based on the polynomial NARMAX representation. Structure selection was facilitated using the error reduction ratio and 

a priori

knowledge of the engine dynamics, and a quadratic NARX model was estimated. The second approach was based on a two-layer feedforward neural network. The NNSYSID toolbox was used to train a NNARMAX network on the same estimation data.

The performance of both the polynomial NARMAX and NNARMAX models was demonstrated using a range of small- and high-amplitude signals. The static behaviour of the two models was compared with the static behaviour of the engine derived from engine data, and the dynamic behaviour of the models was compared with local modes estimated from engine data. It was shown that both models are able to represent the engine dynamics throughout its operating range. Such models could provide the basis for a global nonlinear controller for the engine.

ACKNOWLEDGMENTS

This work was conducted on data gathered at the Defence Evaluation & Research Agency at Pyestock with the support of Rolls Royce plc. The authors would like to thank all the staff involved.

REFERENCES


Figure 1. Simplified schematic of a Rolls Royce Spey engine.
Figure 2. Triangular test (a) measured input fuel flow (b) HP shaft response.

Figure 3. Input spectrum of the triangular test. (a) over a wide bandwidth (b) over the engine frequency range.

Figure 4. Input-output plot using a high-amplitude triangular wave (black), with polynomial fit (white).

Figure 5. Concatenated data set used for estimation (a) measured fuel flow (b) HP shaft response.

Figure 6. Multisine HP shaft test. (a) complete signal (b) portion of the signal. Model 1 output (dashed) measured output (solid).

Figure 7. Three-level HP shaft test. Model 1 output (dashed) measured output (solid).
Figure 8. Neuron: $a = f(Wp + b)$.

Figure 9. Commonly used activation functions (a) sigmoid, (b) tanh.

Figure 10. Two-layer feedforward neural network.

Figure 11. The NNARMAX model structure.

Figure 12. Ten different NNARMAX models showing the sum of squares of the long-term prediction errors on validation data (+).

Figure 13. Static relationship between shaft speed and fuel flow (solid bars), quadratic NARX model (solid), NNARMAX model (dashed).