C484/046/94

Prediction of emissions from a turbocharged passenger car Diesel engine using a neural network

C J Brace, BEng, MIMechE **M Deacon,** BEng, MIMechE and **N D Vaughan,** BSc(Eng), PhD, MIMechE **S J Charlton***, BSc, PhD, MIMechE, **Prof. C R Burrows,** DSc, PhD, FIMechE
School of Mechanical Engineering, University of Bath

*currently of Cummins Engine Company Incorporated, Columbus, Indiana, USA

SYNOPSIS A method of using steady state engine data to predict transient behaviour has been developed and applied to a 1.8l DI TCi passenger car engine. The empirical model uses emissions data obtained from a steady state engine test facility to predict values under real running conditions. Current work is aimed at validating predictions against transient experimental data.

NOTATION

AFR	Air -fuel ratio
BSFC	Brake specific fuel consumption
EPIC	Electronically Programmed Injection Control
CO	Carbon monoxide
CTXE	Continuously variable transaxle under electronic control
CVT	Continuously variable transmission
DI	Direct injection
HC	Hydrocarbons
MLP	Multi-layer perceptron
NOx	Oxides of nitrogen
TCi	Turbocharged and intercooled
a	Activation of a synapse
w	Weight of a synapse
ϕ	Bias of a neuron

1 INTRODUCTION

The environmental impact of passenger cars is a major problem which is being addressed with increasing urgency in Europe. In particular the contribution to atmospheric pollution from the motor car emphasises the need to reduce vehicle emissions. The passenger car of the future must aim to have an exhaust virtually free of toxic substances but must also have significantly improved overall fuel economy in order to limit the production of carbon dioxide.

Improvements can undoubtedly still be made to engine design (**1,2**) but this is not the only approach. Considerable additional benefit can be obtained by the adoption of an engine control strategy such that the required output power is always produced in the most advantageous region of the torque - speed map. This requires the use of a continuously variable transmission to decouple the fixed relationship which otherwise exists between vehicle and engine speed with a stepped ratio transmission. Good drivability characteristics must be retained if a project of this nature is to be successful.

One of the major components in work of this type is clearly an accurate transient engine model. A comprehensive model is required for the drive cycle simulation work. The model is also required to supply data to the driveline controller, both during simulation and in the vehicle. A model has been developed to meet this requirement for an integrated driveline control project running at the University.

The project includes experimental and computer simulation studies used to develop the necessary control algorithms. Reduced order models are used to investigate the relative performance of candidate control strategies by simulation. Selected control methods are then tested experimentally in the laboratory to determine the emission and fuel economy improvements. The modified system is also installed in a vehicle for the conventional ECE15 drive cycle test and assessment of drivability.

1.1 HARDWARE

The chosen prime mover is an experimental Ford 1.8DI TCi Diesel. This engine has significant fuel economy advantages over the IDI generation of engines. It is equipped with exhaust gas recirculation for improvement of NOx emissions.

The fuel injection equipment is the Lucas EPIC system (**3**). It consists of an electronic engine management module controlling a high pressure rotary pump for small high speed DI diesel engines. The controller also schedules the EGR valve opening. The system is currently fitted to the turbocharged 2.5DI installed in Ford Transit vans.
The exhaust gasses are treated using a Johnson Matthey oxidation catalyst. This will reduce hydrocarbon (HC), carbon monoxide (CO) and particulate emissions but not the oxides of nitrogen.

Data has been collected using a steady state rig at Bath. The rig has closed loop control of all important variables. Data collection is computer controlled, including the emissions levels.

2 SYSTEM MODEL

2.1 MODELLING STRATEGY

The system is modelled using the Bath*fp* simulation software developed at the University (**4**). System components are represented by icons linked to FORTRAN or C code. A simulation is then built up using the icons as building blocks to construct the complete system. Interfaces between components are standardised for compatibility. The code behind the icons is linked automatically by the software to form a system model. This can then be used interactively and the results presented graphically. As users develop applications they may need to write their own models of new components. This is facilitated by a series of utilities which allow the generation of icons and code. Component models can range from a simple empirical or instantaneous model to a highly complex dynamic distributed parameter model. The new components can be integrated easily with existing structures.

The complete vehicle system is built up in this way. Emissions and dynamic performance can be predicted during transient manoeuvres such as the ECE drive cycle. A block diagram of the simulation is shown in Figure 1. Controller development will concentrate on optimal control techniques to obtain a balance between particulates, hyrocarbons, oxides of nitrogen (NOx) and fuel economy whilst maintaining good drivability characteristics.

2.2 ENGINE MODEL REQUIREMENT

For both simulation and control the model needs to be accurate during transients. In addition, for control purposes the engine model must be very fast running in order to give a real time update of the ideal operating point. It is possible that the engine model used in the simulation study will be more complex and thus slower running than the on board model. Speed of running is still an issue, however, if drive cycle simulation is to be possible within a realistic time scale.

For simulation the model will be fed information from other models in the system and used to predict the performance of the vehicle and emissions data. This will enable competing strategies to be evaluated.

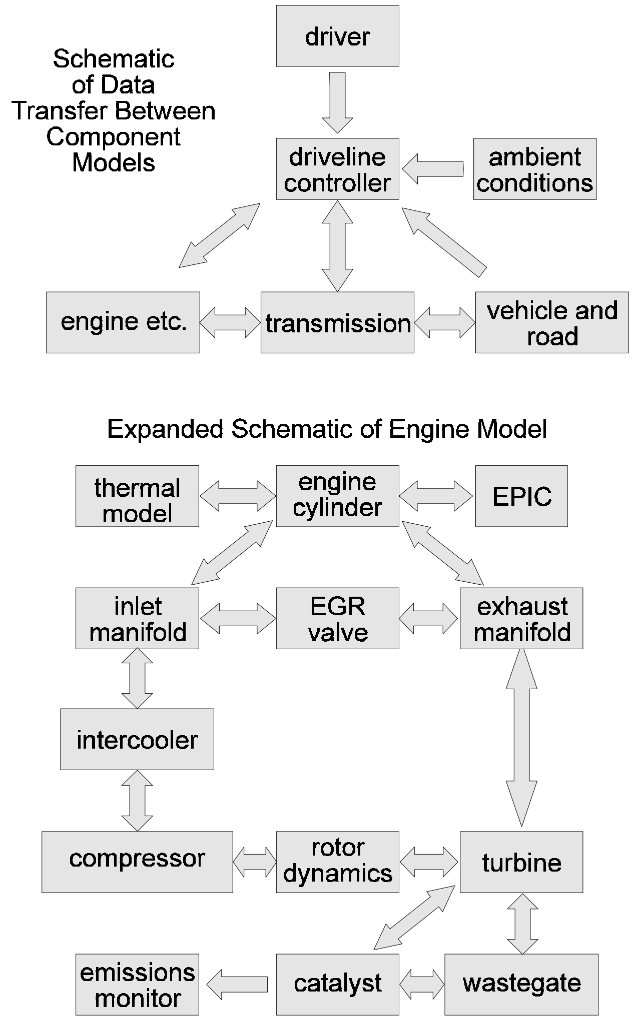


Fig. 1 Schematic of Bath*fp* models

Another instance of the model, resident in the on board controller, will enable an ideal operating point to be predicted in real time during normal vehicle operation. The accelerator pedal input is treated as a power demand by the controller. The model is then used to give the controller enough knowledge of the engine to decide on an optimum operating point. This could be accomplished by running a series of points through the model, all with the same nominal power output at the wheels. The transmission efficiency varies non linearly with speed and load. This must be taken into account when a required engine power is calculated. A weighted sum of the various emissions and BSFCs will enable the best operating point for the power demand to be determined. This point will be continuously updated according to driver demands and engine conditions. The driver demand can then be implemented in the most advantageous manner by the controller.

The weightings used to establish the relative importance of the different emissions may be varied to tune the system. In this instance the use of a catalyst to help with the HCs will enable more effort to go into avoiding NOx production. There is scope for varying the weights according to operating conditions. For example, HCs may be more important during the initial portion of the ECE15 test where the catalyst will not be operational.

3 ENGINE MODEL ASSUMPTIONS

To ensure fast run times and sufficient accuracy an empirical model has been used. A model which can use data generated using steady state rigs is clearly an advantage. This makes the task of gathering and validating the data much simpler. A primary requirement of the model, however, is that it be accurate during transients, since in reality most driving is transient. Some care is therefore necessary in the model structure.

The products of combustion can be predicted by looking at experimental data from a steady state test. A simple technique would be to determine operating speed and torque and interpolate emissions data from an engine map. There are, however, problems with this approach. In a turbocharged Diesel engine the air-fuel ratio can vary greatly from the design point during transients. Temperature and other variables have a similarly dramatic effect. A more complex approach is required.

The solution chosen is to define more fully the operating point of the engine. Speed and load are only two factors. Air fuel ratio, charge temperature, water temperature, injection timing and exhaust gas recirculation fraction are all important variables. This list is not complete, but the major factors affecting performance are defined.

The engine is still regarded as a quasi steady state device. Each combustion event is treated as though it were one of a uniform series, with the engine running in the steady state. Now, however, the operating point is defined by a series of arguments rather than just a pair.

The major requirement for any successful empirical model is a large data set fully describing the engine operating envelope. Increasing the number of inputs to the model to better identify the operating point demands a significantly increased volume of data. Each of the new inputs must be varied in isolation experimentally to study the resulting change in engine performance.

4 DATA COLLECTION AND MANIPULATION

The model for this project uses performance and emissions data obtained from the steady state engine test facility at Bath. All the input parameters detailed above are recorded, with the corresponding outputs - torque, air flow, exhaust temperature, smoke, unburnt hydrocarbons, oxides of nitrogen, particulate matter and exhaust pressure.

Initially a data set was generated by the conventional mapping procedure for the design operating conditions. This was supplemented with data from a number of 'off design' operating points. Each of the engine parameters mentioned above, such as air/fuel ratio, were artificially held to values not normally observed during steady state running. All other inputs were held as close as possible to their nominal values. The points were designed to span the range of conditions expected during transients.

The data set generated by this means is extensive, and time consuming to collect. In this project an engine and a suitable rig were available. If this were not the case, or if the off design points were unattainable in the steady state an alternative source of data would be required. One such source would be an analytical model of the engine such as SPICE II (**5**). This would involve substantial computing effort, and would need to be carefully validated. If this were accomplished the theoretical model could be used to generate data instead of a rig. The run times for this work would no doubt be great, but as it would be an off line operation this does not present a major obstacle. The subsequent steps would be identical.

4.1 DATA MANIPULATION

Once data has been gathered the requirement is for an interpolation technique which can readily cope with the large multi input, multi output data set in a rapid and accurate manner. Several techniques are possible. Conventional linear interpolation could be used. Surfaces could be fitted to the data. Another tool widely used for such tasks is the artificial Neural network (**6,7**).

4.2 NEURAL NETWORKS

A neural network is a structure consisting of a large number of very simple units which combine to represent any given relationship of inputs and outputs. The name originates from the biological structures which inspired them.

The structure of a typical network is shown in Figure 2. Various forms are possible. The design used here is termed a multi layer perceptron (MLP). This is a common device for representing non linear continuous functions. It consists of three groups of nodes or neurons, the input layer, one or more hidden layers and an output layer. Each layer is fully interconnected to the next via a series of connections, called synapses. The number of hidden layers and the number of neurons in each is optional. Generally, accuracy will increase with the complexity of the network until an optimum is reached. Thereafter accuracy will diminish. The aim is to obtain the required accuracy with the most simple (and therefore quickest) network possible. There is one input neuron for each input parameter, which is presented to the neuron as a floating point number scaled from 0 to 1. These values pass via the synapses to the first hidden layer where the data is processed. Figure 3 represents one of these hidden layer neurons.

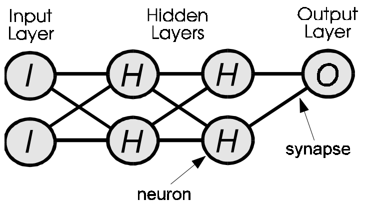
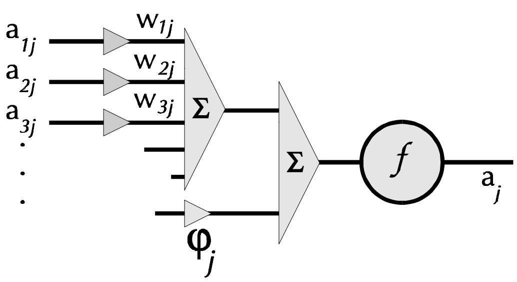


Fig 2 Schematic of a typical neural network

Each input signal is multiplied by its own weighting factor (w). The sum of these weighted inputs is then added to a bias factor (ϕ). This has the effect of increasing or reducing the importance of the neuron to the network as a whole. The resultant sum is used in a non linear activation function (commonly a sigmoidal curve) to arrive at a single output value. This process is repeated at each subsequent hidden layer. The output neurons collect signals from the last hidden layer and present a value between 0 and 1.



$$a_j = \frac{1}{1 + e^{(\Sigma a_{ij} w_{ij} + \varphi_j)}}$$

Fig 3 Hidden layer neuron

To make a network represent a real system the weights and biases are optimised by an iterative training process. The training data is presented many times and the network parameters incremented until the output values converge to be the same as the desired output with an acceptably small error. There are a number of algorithms available to do this. The technique used here is called Stochastic Back Propagation which is commonly used for modelling applications.

The resulting structure can predict the outcome to scenarios it has not previously encountered. Like any interpolation technique it is more accurate when interpolating, extrapolation should be avoided. Accuracy is also enhanced by increasing the amount of data presented. If highly non linear behaviour is being modelled there must be enough data to define the relationship adequately. Once trained the network is treated as a black box. Input values are presented and the outputs are collected. A network can represent a huge volume of data in only a few lines of code and runs very quickly once trained.

Many commercial software packages are available to facilitate the design and training process. For this project a package which runs in a user friendly multi tasking environment (***Windows 3.1***) was chosen (**8**). Once the network is trained the values which define it are frozen and can be written to a text file. This enables the network to be implemented as a simple subroutine in a C or FORTRAN program running under any operating system or hardware. For this project the finished network is run within the Bath*fp* UNIX based system for the simulation and as a C program in a DOS environment for the vehicle controller.

5 ENGINE SIMULATION NETWORK

The topography of the network used for the engine model is shown in Figure 4. There is one input neuron for each of the important variables describing the operating point. One hidden layer is used. There are eight outputs.

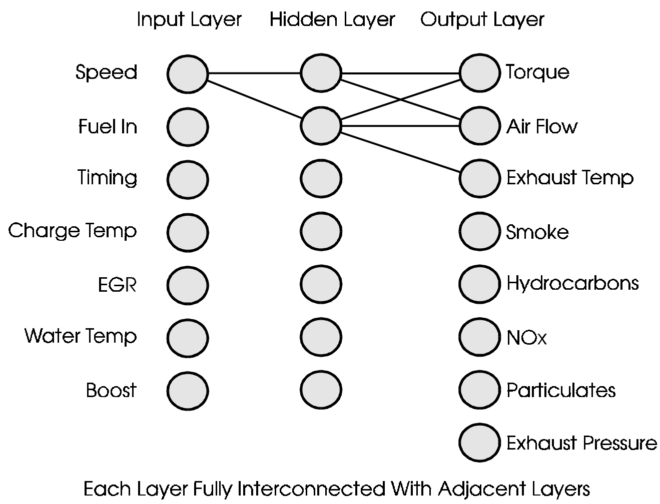


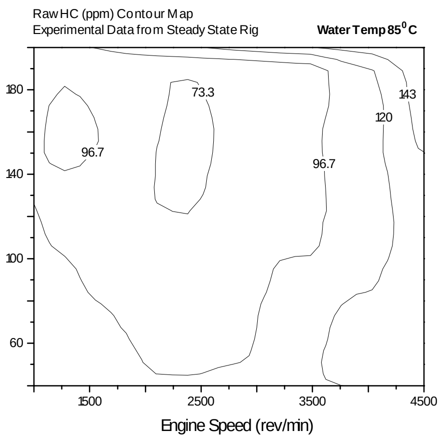
Fig 4 Neural network representation of engine

When using the trained network for simulation inputs such as boost pressure and coolant temperature are predicted by other analytically based models. These are linked within the Bath*fp* environment. A supervisory controller supplies a torque demand to the fuel pump model. This model has been developed jointly with Lucas and is based on the EPIC software used in the vehicle. The fuel pump model calculates the injection timing and fuel delivery to achieve the required torque. Fuelling will not change instantaneously, there are filters in the model to smooth the effects of a rapid change in demand. This enhances drivability by avoiding the driveline oscillations which are possible due to the very fast response of a diesel engine to a step change in fuelling. Fuelling is also limited by the boost pressure to prevent excess smoke during the initial period of acceleration before the turbocharger is up to speed. The transient torque prediction from the network will thus be significantly different from the demanded torque during rapid transients. The network prediction of torque is used to calculate the vehicle acceleration. Smoke, hydrocarbons, oxides of nitrogen and particulates are predicted as raw concentrations and combined with flow rates in a separate routine to arrive at mass flows. Exhaust temperature and pressure are used to predict turbocharger performance and catalyst temperature.

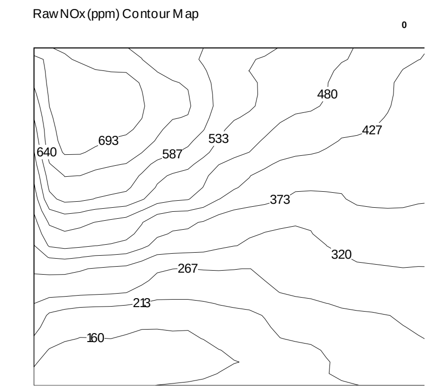
When used as an on board operating point predictor the network will take some inputs from transducers on the engine. Others such as power requirement are taken from the driveline control routine. Instantaneous fuelling data will be supplied by the EPIC system.

6 VALIDATION

A number of techniques can be used to validate the network in the steady state. Figure 5 shows an example of the training data used displayed as a map of (a) raw hydrocarbons and (b) NOx against speed and fuel demand. Fuel demand is expressed in terms of the decimal value assigned to the fuel pump displacement within the EPIC system. The value of 200 represents maximum torque, idling is achieved with a demand of around 40.



(a)



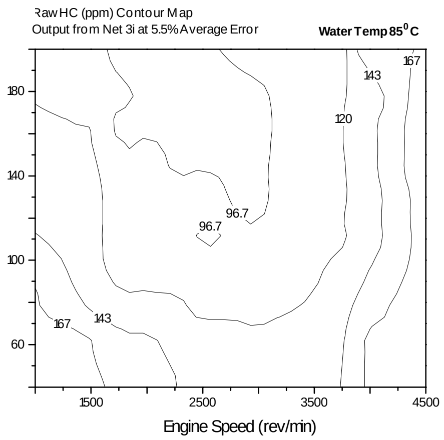
(b)

Fig 5 Experimental training data for neural networks

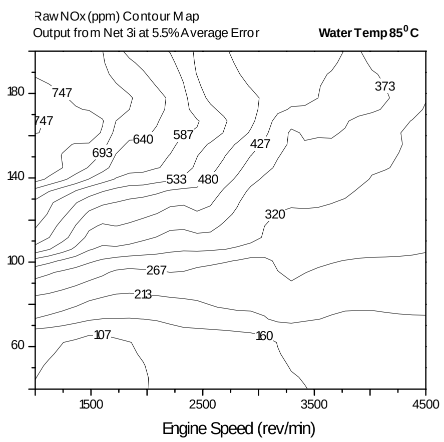
Figure 6 is the result when an early network was used to map the HC (Fig 6a) and NOx (Fig 6b) output of the engine at the design condition. This should be similar to the steady state maps used to train it (Fig 5). The obvious differences are due to poor convergence of the network. The network was not sufficiently complex to fully describe the system.

A new network was developed with more hidden layer neurons. The level of complexity for the best convergence was determined experimentally. Figure 7 shows the results of the same test using the new network. The output is a lot more realistic than from the simpler network when compared to the training data. The prediction of NOx (Fig 7(b)) is better than that for hydrocarbons (Fig 7(a)). This is likely to be due to the experimental data for hydrocarbons being more noisy. The network tends to give a smoother output than linear interpolation, which may be of benefit in some circumstances.

This network was then used to predict the HC and NOx emissions at a water temperature of 25^{o}C (Figure 8). The results are considerably different, as one would expect.



(a)

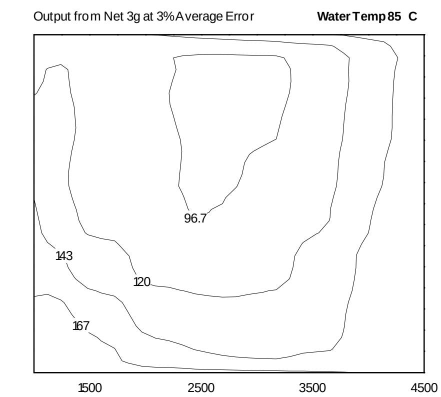


(b)

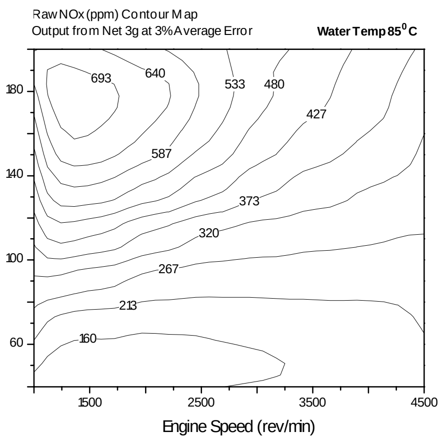
Fig 6 Output from an early network showing poor correlation with experimental data

Tests have been run at a constant speed and fuelling whilst one input parameter at a time is varied to investigate trends. Figure 9 shows the resulting variation in output if boost pressure is varied whilst all other parameters are constant. This test is perhaps a little artificial, but it allows the trends within the network to be investigated. The output values are very smooth across the range of conditions. This is one of the characteristics of the network.

The real test of the model is to verify its accuracy in the transient mode. For this purpose good quality transient experimental data is required. Data will be gathered on a transient rig being developed to perform drive cycle simulation. This rig is fully instrumented including torque and speed measurements on both input and output sides of the transmission and variables associated with the transmission, engine and catalyst. Horiba emissions equipment has been installed.



(a)



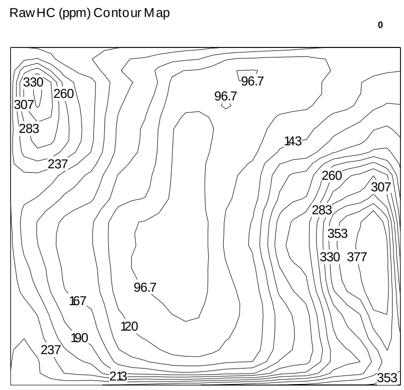
(b)

Fig 7 Output from a later network showing better training.

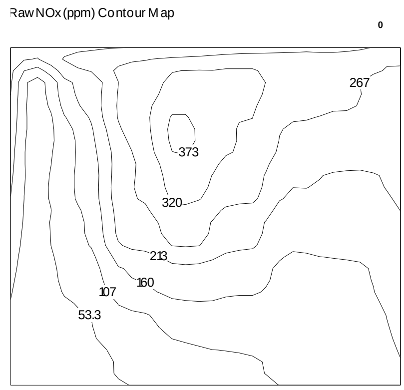
For an initial evaluation SPICE II has been used to simulate a step change in fuelling. The results are used as inputs to the network. Figure 10 shows some network input data generated by SPICE II. This data was used to evaluate two network designs. The resulting outputs are shown in Figure 11. The results shown as a solid line are from a network trained using both the design point data and the off design data discussed earlier. The dotted traces are the outputs from a network trained only on the design point data. It can be seen that the response of the first network to changes in AFR is much greater for the predictions of particulates , smoke and hydrocarbons. Since the second network has no experience of operation with varying AFR the results are likely to be very poor during periods of low boost pressure during transients. This demonstrates the necessity of an extensive data set and a multi input model.

Any discrepancies in the results using the fully trained network will probably be due to -

a. Insufficient training data in the area of operation. Basically a lack of definition in the network training set. This may include areas of extrapolation if the operating envelope was not properly defined, as in Figure 11.



(a)



(b)

Fig 8 Network output at 25^{O}C water temperature

b. Inadequate complexity. A variable not considered may be having a large effect. Examples of these factors may be fuel density or the history of the engine over the last few minutes. Currently history is partly represented by water and charge temperature inputs to the network. Effects such as short lived temperature gradients across the cylinder wall, for instance, are less easy to represent. Extra inputs can be added to the network to counter such errors. Large amounts of data gathered during steady state testing are not being used currently due to the desire to make the model as simple as possible.

c. Averaging effects. The structure of a network tends to damp out the effects of noise. This may be an advantage, for instance, when data is erratic or noisy. It could also hide some real effects and care needs to be exercised.

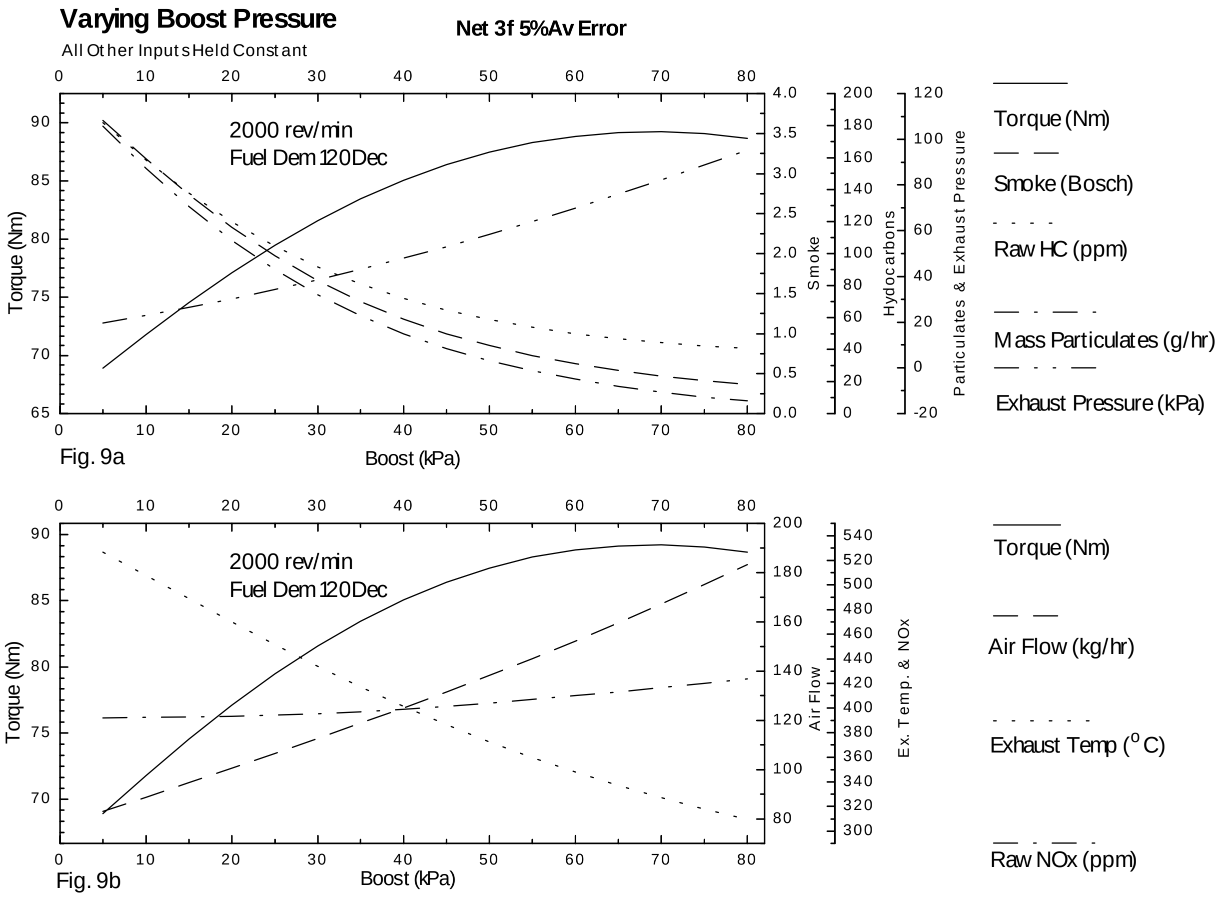


Fig 9 Effect of varying boost pressure whilst holding other inputs constant

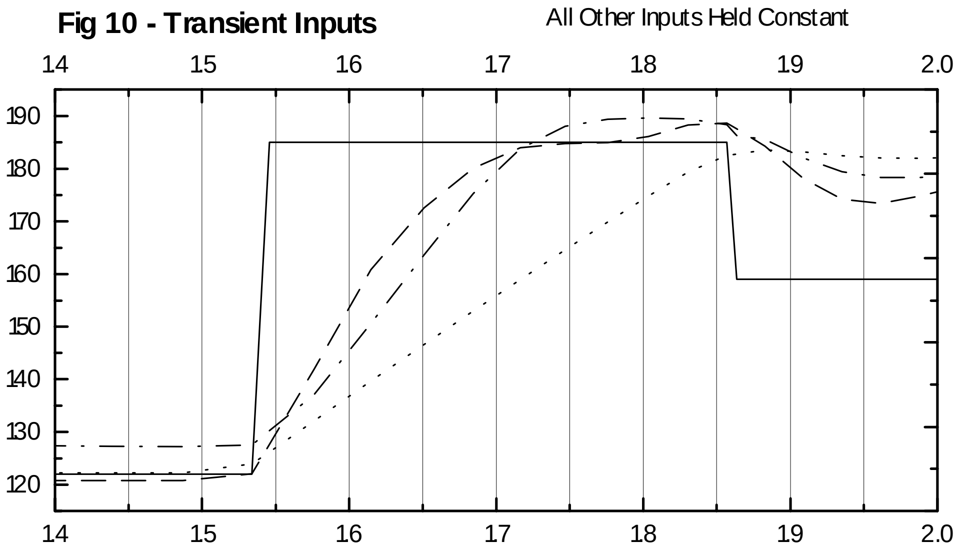


Fig 10 Transient inputs to the neural network

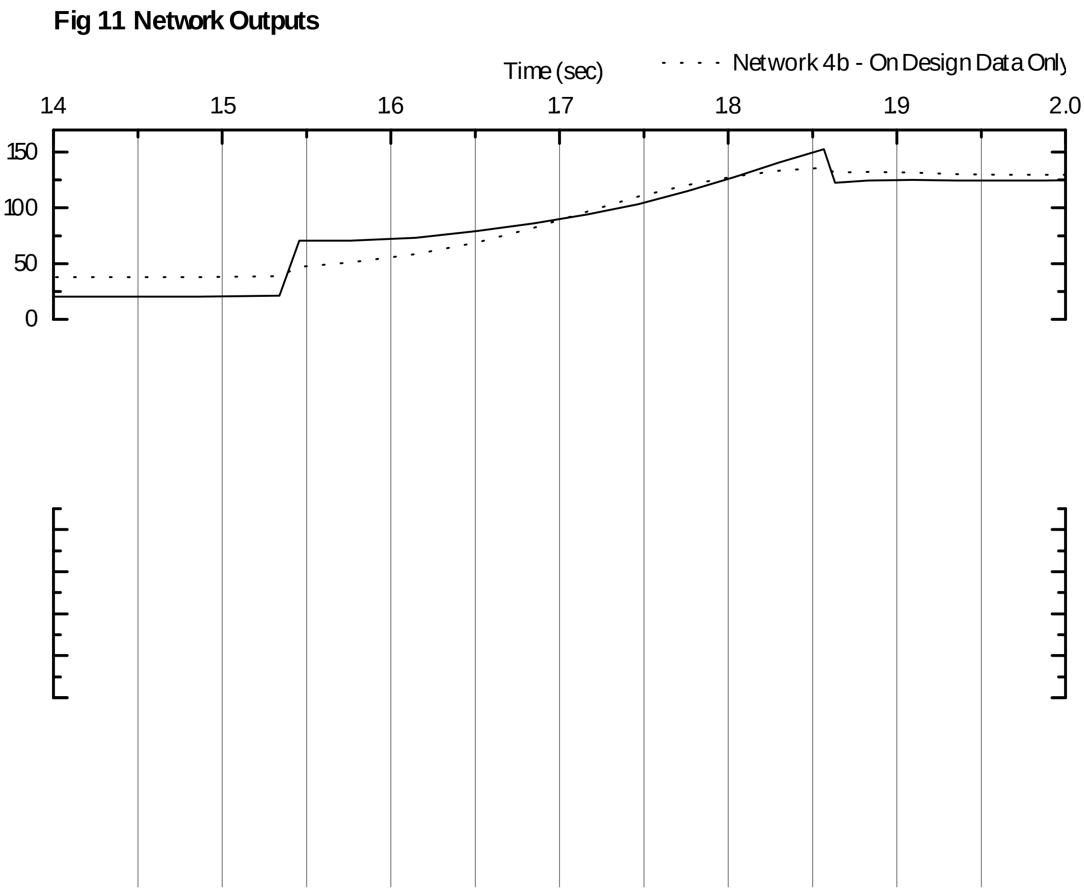


Fig 11 Transient response to a step input in pedal demand - network trained on design point data only

7 CONCLUSIONS AND FURTHER WORK

At present the network is giving a good representation of the data it was trained on. This allows it to be used with some confidence as a fast, efficient interpolation routine. Initial work using transient inputs appears encouraging. Further work needs to be done to establish the accuracy of the model during transients. When experimental data becomes available from the chassis dynamometer the engine model will be tested thoroughly and refined as necessary. This is likely to involve supplementing further the data used to train the neural network. Areas of engine operation important during transient operation will be highlighted. These areas may be influenced by the control strategy chosen for handling transients. It is also possible that more input variables are necessary to fully define the operating point. This step will be taken only if necessary as the network will run marginally slower as a result.

ACKNOWLEDGEMENTS

The work described in this paper has been carried out as part of an Integrated Driveline Control project to investigate the potential for emissions reduction. The project has been undertaken by a consortium comprising the University, Ford Motor Company Limited, Lucas Automotive Limited, Johnson Matthey Catalytic Systems Division and the Department of Trade and Industry.

The authors would like to thank the other consortium members for their assistance and permission to publish this paper. Thanks are also due to I. James for his initial work on the neural network engine model and S. Carter for his work on the SPICE II modelling of the engine.

REFERENCES

(1) Horrocks , RW, Light duty diesels - the emissions challenge. Proc. Inst. Mech. Engrs. Vol. 206 D01091 1992.

(2) Horrocks , RW, Light duty diesels - an update on the emissions challenge. Worldwide engine emission standards and how to meet them. IMechE seminar May 1993

(3) Lewis, GR, The development of the Lucas EPIC system for HSDI diesel engines. IMechE seminar at National Motorcycle Museum 1992.

(4) Richards, CW, Tilley, DG, Tomlinson, SP, Burrows CR, Bath*fp* - A second generation simulation package for fluid power systems. 9th Int. Conf. on Fluid Power, Cambridge 25-27 April 1990.

(5) Charlton, SJ, 'Simulation Program for Internal Combustion Engines' User Manual, University of Bath 1986

(6) Scaife, MW, A neural network for fault recognition. SAE Int. Congress, March 1993, Session 1p23 Controls for Engines

(7) Bacon, A, Potential for engine control using neural networks. IMechE 1992 C448/057

(8) Neural Desk Users Guide, Neural Computer Sciences 1992.