Particulate Matter Prediction of Both Steady-State and Transient Operation in Diesel Engines

Jiamei Deng

School of Mechanical and Automotive Eng.

Kingston University, London SW15 3DW

Email: j.deng@kingston.ac.uk

Abstract:

Diesel engines produce a variety of particles generically classified as diesel particulate matter (PM) due to incomplete combustion. The increasingly stringent emissions regulations require that engine manufacturers must continue to reduce the PM. The ability to predict the PM emissions is one of the key technologies that could be used in a PM reduction strategy. This paper describes a predictive technique that can be used as a virtual sensor for monitoring PM emissions in both steady and transient states for a medium or heavy-duty diesel engine. The predictive structure is stable over a broad range of engine operation points. The input parameters are chosen based on the PM formation mechanism, physical knowledge of the process and an insight into the underlying physics. Principal component analysis (PCA) is used to reduce the dimensionality of the inputs of a non-linear autoregressive model with exogenous inputs (NLARX) from 9 inputs to 5 inputs. PCA not only reduces the input number, but also improves the performance of the prediction model. The results show that the

NLARX model could predict the particulate matter successfully with R^2 above 0.99 with only 5 inputs.

1. INTRODUCTION

Diesel engines produce a variety of particles generically classified as diesel particulate matter (PM) due to incomplete combustion. The increasingly stringent emissions regulations require that engine manufacturers must continue to reduce the PM. The technologies available for PM reduction may be classified into two groups: The first method is based on after-treatment systems, such as a diesel particular filter (DPF), to trap particles that are then burnt in a controlled way from time to time. The second method is based on combustion modification technologies consisting of both injection technologies and real-time advanced control systems that modify combustion systems to reduce PM without substantial increases in NO_x concentrations. For the first method, accurately measuring PM is crucial for the DPF life. If regeneration is too frequent, the DPF life will be shortened. If it is regenerated less often and there is too much mass accumulated on the DPF, the regenerating temperature will be very high thereby shortening the DPF life. The second method must use measurements or estimates of the PM signals for control purposes. Both methods heavily rely on the measurement of PM and therefore a PM model with good prediction performance is of fundamental importance. The ability to predict the PM emissions is one of the key technologies that could be employed to reduce PM emissions.

The question addressed in this paper is how best to represent the production of PM using commonly available engine parameters. The target application is a single engine type where the combustion process is defined, but where the details of air flow, EGR, injection structure and variable boost remain to be defined in detail. While in-cylinder conditions are vital to describe the details of particulate production, the process itself is fundamentally defined by

the engine controls themselves, hence the paper's focus on use of the set of control variables. Given an engine design, the precise pattern of particulate formation is a function of air flow, fuel delivery and EGR, all of which are determined by the engine's controlled variables. One practical use of such a modelling capability is to understand the effect on particulate emissions of a particular choice of control strategy. This capability is of great significance in an engine development.

The principal goal has been to demonstrate that the particulate emissions of the engine under test are capable of being represented and accurately reproduced. Our assertion is that if the modelling process is successful, the generic nature of this approach means it is very likely to be applicable readily to other engine types.

Diesel engine PM Prediction has always been a major challenge to the industry [1, 2, 3]. Computational fluid dynamics (CFD) based PM models are computationally intensive and are not suitable for control purpose and real time measurement. Recently, neural networks have been used in a wide variety of automotive applications. Neural networks have been successfully used for emissions prediction [4]. He et al. [5] built a model that considers several engine parameters such as boost pressure and exhaust gas recirculation (EGR) and it generates several outputs including PM emissions. Maass et al presented a smoke prediction neural network model using a three-layer autoregressive model with exogenous inputs (NLARX) model to predict PM [6]. Bose and Kumar [7] use fuzzy logic to predict the engine emissions, but one of the inputs is cylinder peak pressure, which is difficult to obtain in normal engines. For all those application, there is still no systematic method to design an experiment and decide inputs of the neural networks.

How to choose effective inputs based on easily obtained parameters is of great importance for building a successful PM model. Schilling et al designed a real time model to predict NO_x emission whose inputs are start of injection, injection duration, injection duration, air mass flow and boost pressure. They chose these inputs based on a sensitivity analysis [8]. Thompson et al used ten inputs to predict emissions without doing any analysis [9].

The advantage of neural networks is their ability to be used as an arbitrary function approximation mechanism without the knowledge of the complex underlying process, which is an economic way to obtain the measurement. In this paper, various kinds of experiment, such as random walk (RW), constant speed load acceptance (CSLA), non-road transient cycle (NRTC), idle (zero torque) to full throttle (maximum torque) change with minimum time (ITF) are designed to catch all kind of characteristics of PM production. The model is developed to estimate the PM based on the easily available sensors in engine operation. The model could be used for indicating the DPF regeneration time, optimising engine operation and control actions. Such a model must be simple to implement, easily trained or re-trained and produce good prediction performance over unseen data. Moreover, to be used for different diesel operation, the model should also convey rich system information. In general, developing PM models must address some unique problems:

(1) The basic PM equations could not be used directly as many variables are either unmeasureable or unreliable, the choice of structure of such a model will be very difficult.

(2) The model that captures both steady state and transient state of PM is challenging because of the required versatility and the need to address the different physical and chemical processes that develop during transient operation.

(3) There is no systematic method to be used to choose available parameters and minimise input variables.

Therefore, this paper will use NLARX to overcome the problem (1) and (2). Problem (1) may be overcome by using 'black-box' models such as artificial neural networks (ANN). Problem (2) will be overcome by designing various kinds of experiment to validate the model robustness that together catch the full dynamics of PM production in diesel engines. In this paper, the minimum number of input variables that needs to use available parameters according to Problem (3) will be overcome using PCA and selective testing of measurements, which will

(1) use available data to fit a model,

(2) use additional data to validate the model,

(3) include different inputs if the performance is unacceptable, and then

(4) fit a new model again.

The reduction in the dimensionality of the inputs, in reality, means that the reduction of the sensor numbers. The modelling process comprises three major stages:

(1) The first stage includes the study of PM formation by analysing the chemical reactions of the air and fuel in the diesel cylinder for the purposes of identifying the significant manipulated variables that contribute to the PM formation and production.

(2) The second stage is to construct the different models based on the selected input parameters. Different approaches are used to assess the suitability of the selected input parameters.

(3) By both analysis and trial-and-error, more model inputs are selected to ensure acceptable performance if the model is not good enough when using the inputs selected during (2).

(4) The number of the input parameters is minimised by using PCA.

In this paper we demonstrate this four step modelling process: firstly, different experiment methods are reported that are designed to catch both dynamic and steady state aspects of diesel engine dynamics. The input parameters are decided according to the PM formation mechanism, physical knowledge of the process, an insight of the underlying physics and trial-and-error. PCA is used to reduce the dimensionality of the inputs of NLARX from 9 inputs to 5 inputs.

NLARX is trained and validated to predict the PM. The engine is operated in different transient and steady states and the data recorded. The resulting data is normalised and processed into a training and validation set. This paper reduces input parameters based on PCA, which is very useful to reduce the number of sensors employed and therefore the ultimate cost in production. These methods could be borrowed by any other kind of the modelling techniques. The results demonstrate that a model is able to cover both steady and transient states with an R-squared value in both training and validation of 0.99, implying both accuracy and generality. For the future legislation which is linked to number density and size distribution of soot particles, this model approach will be suitable for identifying the model of the number density of nano particles based on experimental observations. Maybe the equation (2) could be linked with the number density of nano particles.

In Section 1 a brief background of the research is introduced. Section 2 describes the mechanism of PM formation. Section 3 introduces the experimental facility for data

collection. Section 4 explains the detailed procedure for data collection and preparation. Section 5 explains the implemented model architecture and PCA. Section 6 shows the results of modelling and discusses the choice of input parameters based on PCA. Section 7 presents the conclusions and new research ideas emerging from the work.

2. PHYSICAL PM MODEL

In diesel engines, the fuel is directly injected into the cylinder prior to the ignition and combustion processes or during the combustion processes. The air flow to the cylinder is controlled through boost pressure and EGR and is designed to enable the complete combustion. A full phenomenological modelling exercise either for ignition or combustion would be very complicated [10]. This paper follows the same theoretical analysis of the paper [10] beginning with an assumption of the combustion process as:

Air+Fuel-------->PM+ other products.

The rate *r* of appearance of a reaction product is

$$r = A[m_f]^c [m_a]^d \exp(-\frac{E}{RT}), \qquad (1)$$

where m_f is the unburned fuel mass fraction, m_a is the unburned oxidizer mass fraction, *R* is the universal gas constant, *c*, *d*, *E*, *A* are all constants. Activation energy E and pre-exponential factor *A* are found by experiments. Such a model, while physically consistent is impractical as the basis for a control oriented model.

To aid simplification, the assumption is made that the chemical reaction constant depends on temperature. The order of the reaction depends on the reaction pressure. However, the combustion pressure is not constant. It is reasonable to link the reaction order to the cylinder

pressure which in turn is dependent on the volume variation due to the mixture of air and fuel in the cylinder. The volume variation is linked to the engine speed. Therefore the particulate emissions are ultimately linked to the engine speed and load as follows:

$$r = A(N)[m_f]^{c(N)}[m_a]^{d(N)}, \qquad (2)$$

where *N* is the engine speed. It is this theoretical analysis that is the basis for the initial input parameters for model identification. This model is generic to all engines using compression ignition [10]. PM is highly correlated with in-cylinder conditions. Both the fuel and air components of Equation (1) are related to load. Air supply is related to the boost pressure which is linked to exhaust conditions and in turn to engine load. Fuel quantities are directly related to load. Load is implicitly represented in the variables which are pertinent to explanations of the formation of particulate matter.

3. EXPERIMENT FACILITY

The engine employed in this study is a Cat® C6.6 ACERT^{TM} medium-duty off-highway engine. This is a 6 cylinder engine with a Cat common rail fuel system. The engine calibration used in this work produces up to 159kW at rated speed (2200rpm) with peak torque of 920Nm at 1400rpm. The engine has been modified with a high pressure loop EGR system and a variable geometry turbine for experimental purposes. The engine used in this paper was modified and therefore not representative of a production engine.

The engine is fully instrumented to measure air, fuel and cooling system pressures, temperatures and flow rates. Emissions data is gathered principally using an AVL 415 smoke and 439 opacity meters (offering steady state and transient measurement respectively) and a

Horiba 9100 exhaust gas analyser measuring nitrous oxide, carbon dioxide, carbon monoxide, unburnt hydrocarbons and oxygen. The temperature and pressure at the exhaust port are measured for each of the six cylinders. This is supported by exhaust manifold, post turbocharger, pre/post EGR cooler/control valve temperature and pressure measurements. Figure 1 shows the engine facility.

Figure 1: Schematic graph of the engine and operation station

4. DATA COLLECTION

All the engine parameters are recorded in a 1Hz and 10Hz sampling frequency under these conditions. For an initial model set-up four test cycles are operated on the engine. The engine control module is set to standard calibration mode. These test cycles are: different transient and steady data set (RW, NRTC, CLSA and ITF). The different dynamics of these cycles could be seen from Figure 2, 3, 4, and 5. The data is rearranged for training and validation purpose according to the respective transient and steady characteristics of each cycle.

4.1 Random Walk (RW)

The test result for RW is shown in Figure 2. RW is a mathematical formalisation of a trajectory that consists of taking successive random steps. Often, RW is assumed to be Markov chains. A Markov chain is a sequence of random values. The probabilities of the Markov chain at a time interval depend upon the value of the number at the previous time. The transition probability from one state to another state is a conditional probability. Suppose we know the current state, at a certain moment it moves to a new state, where the probability of moving from state X_i to state X_j is P_{ij}. P_{ij} is the transition probability.

Given a starting point, the distance from one point in the path to the next is constant and the direction from one point in the path to the next is chosen at random, and no direction is more probable than another. Often, the walk is in discrete time, and indexed by the natural numbers, as in X_0, X_1, $X_{2...}$. In this paper, two different distances are chosen: one is called slow pace and another one is fast pace. Both torque and speed are subject to the random walk variation.

(1) **Slow pace random walk test:** the test duration is 6218 sec, which is shown in the upper part of Figure 2.

(2) **Fast pace random walk test:** the test duration is 1001 sec, which is shown in the lower part of Figure 2.

In this case, maximum load request is reduced from 100% to 70% of rated torque in order to avoid engine stall.

4.2 Constant Speed Load Acceptance (CSLA) test

Figure 3 shows the results of the CSLA test. The duration of the test is 45 minutes. The speed is changed from 1000 RPM to 2200 RPM in 200RPM increments. For each speed step, a step change towards the maximum torque value is imposed and the response is recorded over a holding time. The ramp times and holding times for each test point are defined in Table 1.

Table 1: CSLA ramp time definition

4.3 Idle to Full Throttle (ITF)

Figure 4 is the test result of ITF. The engine response is recorded from idle (no torque) to full throttle (peak torque).

4.4 Non-Road Transient Cycle (NRTC)

The NRTC is an engine dynamometer transient driving schedule of total duration of about 1200 seconds. The speed and torque during the NRTC test is shown in Figure 5. The NRTC is a cycle that was devised by the U.S. Environmental Protection Agency to represent the range of operating conditions of off-highway machinery. It is the standard test cycle for Tier 4 emissions standards. For this project the motivation for this choice of cycle is twofold. Firstly, experience has shown that this is one of the most challenging cycles in terms of emissions modelling. Secondly, meeting emission formation requirements under the NRTC cycle is also a major concern to engine manufacturers. The current trend is to design engines which pass legislative emission test by a small margin, but where that margin must be provably robust against deterioration in engine systems. For this reason the use of reliable and highly accurate emissions models is of critical importance.

Figure 5 is the result of NRTC. The test is completed with 70% maximum load and full speed range covering a wide range of engine transients in different frequencies and combinations. The load limitation is due to formulation of the speed governor which is at the time of conducting the experiments was not fully optimised for speed following in the NRTC.

Figure 2: Random walk of fast & slow- Training and validation part distribution

Figure 3: CSLA Test - Training and validation part distribution

Figure 4: Idle ramp test training and validation ramps

Figure 5: NRTC test

4.5 Test Cycle Data Processing

The test cycle data, once acquired is processed into a training and validation set. Each cycle is split into training (T) and validation (V) parts. These parts are recompiled as shown in the scheme in Figure 6.

Figure 6: Illustration of training and validation data

Figure 7: Training set including RW, NRTC, ITF and CSLA

Figure 8: Validation set including RW, NRTC, ITF and CSLA

The resulting training and validation sets are shown in Figures 7 and Figure 8. Each set contains the same amount data of each test cycle. The feature density is high covering a wide scope of engine operation behaviour in both steady-states and transient operation. The curves show the complete range of the present engine from 800-2300RPM and 0-900Nm for speed and torque, respectively. Each data set is initially processed to give a 1 Hz sampling frequency for the initial model identification.

5. NLARX MODEL AND PRINCIPAL COMPONENT ANALYSIS

5.1 NLARX model

The field of virtual sensing has become more and more popular with growing systems complexity such as in combustion engine control. Its origin lies in the field of estimators which are specified through physical and numerical relations whereas virtual sensors are characterized through black-box approaches such as neural networks.

Neural networks can be split into the following three categories:

(1) single-layer feedforward networks (SLFN),

(2) multi-layer feedforward networks (MLFN),

(3) recurrent neural network (RNN).

The chosen network structure or architecture is crucial for the output performance. Depending on the systems characteristic: linear or non-linear, static or dynamic, the network needs to be designed accordingly. Here, the prediction of PM is recognised as highly dynamic and non-linear that implies a recurrent network structure has to be chosen to offer sufficient predictive capability. The NLARX structure can accommodate the dynamics of the system by feeding previous network outputs back into the input layer. It also enables the user to define how many previous output and input time steps are required for representing the systems dynamics best. In this paper a NLARX model is applied as it is suitable for non-linearity of the problem. Although an important result of approximation theory is that a three-layer feed-forward neural network with sigmoid activation functions in the hidden layer and linear activation functions in the output layer has the ability to approximate any continuous mapping to arbitrary precision, provided that the number of units in the hidden layer is sufficiently large [11]. However, the performance of feed-forward neural networks is limited due to limitations to the number of units in the hidden states. Performance is further limited by the memory of personal computers. It is for this reason that, SLFN and MLFN have not formed part of the work reported in this paper.

Figure 9: NLARX canonical structure

A typical structure of an NLARX model is illustrated in Figure 9. The inputs are represented by u(n) and the outputs are described by y(n). The formulation of this NLARX model can be described as:

$$y(n) = F(y(n-1),\ldots,y(n-ny),u(n),\ldots,u(n-nu+1)), \quad (3)$$

where *ny* is number of past output terms used to predict the current output, *nu* is number of input terms used to predict the current output.

Each output of an NLARX model is a function of regressors which are transformations of past inputs and past outputs. Usually this function has a linear block and a nonlinear block. The model output is the sum of the outputs of the two blocks. Typical regressors are simply delayed input or output variables. More advanced regressors are in the form of arbitrary user-defined functions of delayed input and output variables.

NLARX model training can be cast as a non-linear unconstrained optimization problem:

$$\min_{\theta} F_M(\theta, Z_M) = \frac{1}{2M}\sum_{k=1}^{M}\left\|y(k) - \hat{y}(k\,|\,\theta)\right\|^2 \quad (4)$$

where $Z_M = [y(k), u(k)]_{k=1,\ldots,M}$ is a training data set, *y(k)* represents the measured output which is the measured PM in the training set, $\hat{y}(k\,|\,\theta)$ is the NLARX output which is predicted PM, $\|.\|^2$ is 2-norm operation, θ is a parameter vector, where $\theta = [\theta_1, \cdots, \theta_i, \cdots, \theta_p]$ and *p* is the number of parameters·

The training process is described as follows. Given a neural network described by Equation (3), there is an error metric, that we refer to as performance index of Equation (4), that is to be minimized. This index is a representation of the approximation of the network to some given training patterns. The task will be to modify the network parameters θ to reduce the index $F_M(\theta, Z_M)$ over the complete trajectory to achieve the minimal value. In this paper the neural networks are trained using gradient descent algorithms while the initial value of θ is perturbed several times in order to avoid the local minimal solution· The gradient descent methods will calculate the vector $\nabla_\theta F_M$ whose elements are $\frac{\delta F_i}{\delta \theta_i}$ $(i = 1, \cdots, i, \cdots, p)$. The

training algorithm will find the parameters of the network for which the performance index has reached a desirable value. Given a vectorising trajectory for the network output and training patterns, the performance index is the Euclidean norm of the error matrix of the whole training batch for the output PM.

5.2 Principal Component Analysis (PCA)

PCA is used to reduce the input numbers of the NLARX model. PCA generates a new set of variables, which are principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, meaning that no redundant information exists. The set of principal components forms an orthogonal basis for the space of the data. While there are many ways to construct an orthogonal basis for several columns of data, PCA assumes that the directions along which the variations are the largest include the most information about the model. The most common approach to construct PCA is to use the standardized linear projection, which construct the maximum variance in the projected space [12, 13]. Suppose there is a q-dimensional data set D, the n principal axes P_1, P_2, P_3,..., P_n, with $n<q$, are orthogonal axes onto which the retained variance is maximum in the projected space. Generally, P_1, P_2, P_3,..., P_n can be given by the n leading eigenvectors $\lambda_i$of the sample covariance matrix $C = 1/s \sum_{i=1}^{s}(d_i - m)^T(d_i - m))$, where $d_i \in D$, (i=1,2,…S), m is the sample mean, and s is the number of samples, so that:

$$CP_i = \lambda_i P_i.$$

The n principal components of D are de-correlated in the projected space. Most of the information in the observation vectors is contained in the subspace spanned by the first m principal axes, where $m<q$. Therefore, each original data vector can be represented by its principal component vector with dimensionality m [12].

The first principal component is a single axis in space. When each variable is projected on that axis, a new variable is formed, whose variance is the maximum among all possible choices of the first axis. The second principal component is another axis in space, which is perpendicular to the first axis. The variable on this axis is projected on this second axis and the second new variable is formed whose variance is the maximum among all possible choices of this second axis. This process continues until all variables are used. The full set of principal components is as large as the original set of variables. However, it is natural for the sum of the variances of the first few principal components to exceed 80% of the total variance of the original data. After studying plots of these few new variables carefully, a deeper understanding of the driving forces that generated the original data could be understood. Therefore, dimensionality could be reduced.

6. MODELLING EXERCISE AND PRINCIPAL COMPONENT ANALYSIS

6.1 Initial Model Identification with seven Inputs

For the initial model identification seven inputs were chosen:

(1) torque/ load,

(2) engine speed,

(3) intake manifold temperature,

(4) mass air flow,

(5) air-to-fuel ratio,

(6) boost pressure,

(7) exhaust pressure**.**

The initial parameters were chosen based on the model analysis of Equation (2), a physical knowledge of the process and an insight into the underlying physics. From Equation (2), we

know that the PM is mainly affected by engine speed, unburned fuel and air fraction, which are not independent variables. Therefore, the inputs will include all the parameters that affect these variables. The air-to-fuel ratio (AFR) provides an indicative threshold of smoke formation. Both internal and external EGR will change the combustion temperature, thus affecting the PM formation. Since, EGR flow is not measured directly, exhaust and boost pressures are used instead as inputs. It is believed that this is equivalent to having EGR flow rate as an input. Another important input is compressor mass air flow as it affects the fresh-air supply to the engine. Emission formation rate is also dependent on the engine speed and torque transient responses and thereby these are included as inputs. The effect of rail pressure and start of injection are also included because of their effect on the combustion process.

Three different approaches were tested for this initial model identification in order to find a suitable model structure. Each set was reduced to 1 Hz data and the inputs and outputs are normalised for data range reduction. The tree different modelling approaches are:

(1) single NLARX structure,

(2) three-layer parallel NLARX structure,

(3) fuzzy Model.

Each model was trained on training set (Figure 7) and validated against the validation set (Figure 8). The correlation between the desired measured test data and the model predicted output is determined through the coefficient of determination R^2 that is shown in Equation (3):

$$R^2 = 1 - \frac{\sum_{t=1}^{M} (y(t) - \hat{y}(t))^2}{\sum_{t=1}^{M} (y(t) - \bar{y})^2}, \quad (3)$$

where y describes the desired output data, $\hat{y}(.)$ the prediction and $\bar{y}$ the mean value of the output data. The coefficient of determination shows the explained variability of the systems output by the regression model. A result of $R^2=1$ means an exact model has been found whereas with an R^2 value of 0 shows there is no correlation between the system and the model output.

R^2 is probably the most popular measure of fit in statistical modelling and is commonly used in regression analysis to indicate how well the model predicts responses for new observed data or in the validation. Its use can prevent overfitting the model because it is calculated using new observed data (in the validation data) not included in model estimation (in the training data). Overfitting refers to models that appear to explain the relationship between the predictor and response variables for the data set used for model calculation but fail to provide valid predictions for new observed data. If the R^2 in the training data is much greater in value than that in the validation data, there is overfitting in the model, which must be rectified. Larger values of predicted R-squared suggest models of greater predictive ability. There is no minimum value for R^2. An acceptable minimum is determined by the particular application.

6.2 Initial modelling with single NLARX structure

The data sampling rate is 1 Hz. A single NLARX model was used for training and validation purpose. The training of R^2 is 0.88 and the validation of R^2 is 0.67. The visual correlation of training and validation data for this single NLARX model can be found in Figure 10. For this case a single NLARX model appears not to work and investigation proceeded to a three-layer NLARX.

.

6.3 Three-layer approach with NLARX structures

The data sampling rate is 1 Hz. The three-layer NLARX structure was presented in [6], which partitions the amplitude of the PM into three layers. Each layer is used to train a separate NLARX model and then the sum of three NLARX outputs forms the predicted output as described in [6]. The training of R^2 is 0.86 and the validation of R^2 is 0.69. It is obvious that the R^2 values are too low to indicate a reliable model. The visual correlation of training and validation data for this three-layer NLARX can be found in Figure 11. It is concluded that neither the single layer nor the three-layer NLARX works for this case. Therefore, fuzzy modelling is used as an alternative.

6.4 Fuzzy modelling

The data sampling rate is 1 Hz. For the fuzzy model, the training of R^2 is 0.88 and the validation of R^2 is 0.75 for the fuzzy model. The visual correlation of training and validation data for this fuzzy model can be found in Figure 12.

Figure 10: Single NLARX structure training and validation correlation results

Figure 11: Three-layer Parallel NLARX training and validation correlation results

Figure 12: Fuzzy model training and validation correlation results

6.5 Model Identification with nine Inputs

After exhaustive searching and training with the previous three different modelling approaches, no good results are generated using seven inputs. It is suspected that the seven inputs are not enough to cover all the dynamics of the PM formation even though these variables are supported theoretically. The ignition and combustion is a complex process for

diesel engines that is implicated in particulate formation. Based on such reasoning, two additional inputs are included in the model:

(1) common rail pressure,

(2) fuel quantity.

Those inputs may increase the information content available to predict the actual opacity output.

The new 9 inputs are:

(1) torque/ load,

(2) engine speed,

(3) intake manifold temperature,

(4) mass air flow,

(5) air-to-fuel ratio,

(6) boost pressure,

(7) exhaust pressure,

(8) common rail pressure,

(9) fuel quantity.

Inputs and outputs are normalised into a range of [0, 1]. For the new input set, both fuzzy modelling and a single NLARX structure are tested respectively. The three-layer approach is neglected here due to the similarity in results in comparison to the single NLARX approach. This leaves the assumption that the single NLARX structure is capable of predicting the full data scope. The fuzzy method resulted in slight improvement of the training correlation. However, the validation did not improve as much. This leads to of the conclusion that the

model is overfitted which can be seen in the increasing gap in the correlation values between training and validation. The training of R^2 is 0.91 and the validation of the R^2 is 0.78.

The single NLARX approach gives good results as the both training and validation R-square are above 0.96, which is for training: R^2= 0.99 and for validation: R^2= 0.96.

Here the visual correlation of the single NLARX structure is shown (Figure 13 and Figure 14).

Figure 13: Correlation results for 9 input NLARX structure

Figure 14: Illustration of regression of training and validation results for nine inputs

This result leads to the conclusion that the NLARX structure is quite capable of predicting the desired parameter. However nine parameters were considered too large a number of variables for a practical implementation. The likelihood of correlations between variables for example (1) between manifold pressure and air flow and (2) between fuel rail pressure and fuel quantity suggest that a smaller set of variables would be viable. In the following section the reduction of number of inputs is investigated to support the need for a reliable solution based on a smaller number of sensed variables.

6.6 Model Identification with PCA Pre-Processing for Inputs

With our previous good results the next step is to see whether some inputs can be neglected and how large the impact will be on the performance. For this purpose a PCA is performed on the input set of 9 inputs in order to identify the most significant inputs. This leads to a possible reduction of four inputs to a total of five for sufficient predictive accuracy of the model.

The PCA creates a listing of the variability of each input and hence its influence on the systems behaviour. The following list shows the ranking for the PCA result on the 9 inputs chosen initially. The PCA results in the following listing of principal components, starting with the strongest influencing input first:

(1) PC1 – air-to-fuel ratio,

(2) PC2 – speed,

(3) PC3 – torque/load,

(4) PC4 – exhaust manifold pressure,

(5) PC5 – common rail pressure,

(6) PC6 – intake manifold pressure,

(7) PC7 – fuel quantity,

(8) PC8 – intake manifold temperature,

(9) PC9 – mass-air-flow.

Here, PC_i represents the *i*-th principal component, where $i=1,2,\ldots,9$.

Figure 15 shows the percentage contribution of each of the principal component to the total of variance of the original data neglecting those with less than 2% contribution. From this analysis, it is concluded that only five inputs are needed for model development. Therefore, models are identified models using five and six inputs, separately.

Figure 15 Principal component analysis result

6.7 Model Identification for 6 inputs

Based on this PCA result, three inputs have been removed. However, this results in an even better result for prediction PM using NLARX model. The new inputs are:

(1) PC1 – air-to-fuel Ratio,

(2) PC2 – speed,

(3) PC3 – torque/load,

(4) PC4 – exhaust manifold pressure,

(5) PC5 – common rail pressure,

(6) PC6 – intake manifold pressure.

The correlation of the results of the single NLARX structure with six-input are shown in Figure 16 and Figure 17. It can be seen that the performance can be actually improved by reducing the numbers of inputs down to six. This is assumed to be influenced by least "waste" information added by the dropped inputs. They may contain information that does not relate to the output making it more difficult to find an optimal solution. Therefore, it is necessary to analyse the inputs using PCA. The R-squares of both training and validation are 0.9968 and 0.9965, respectively.

Figure 16: Single NLARX structure training and validation results for six inputs

Figure 17: Illustration of regression of training and validation results for six inputs

6.8 Model Identification for 5 inputs

The next step in the investigation was to identify whether there could be a further reduction on the number of variables. The result for a five input model can be seen in Figure 18 and Figure 19. The five inputs are:

(1) PC1 – torque,

(2) PC2 – speed,

(3) PC3 – air-to-fuel ratio,

(4) PC4 – exhaust manifold pressure,

(5) PC5 – common rail pressure.

The R^2 for both training and validation are 0.9961. With these training results, the inputs of the model could not be further reduced to four as it makes the predictive results unacceptable. The model using five inputs is as accurate as the model using six inputs.

It is important to place this result in context. The particulate emissions have been modelled accurately throughout the NRTC with its range of transient and steady state conditions. The five variables form an accurate basis, given the basic nature of the engine design. As the approach has been generic, the modelling process is capable of being applied to other diesel engines. The model is likely to take a different numerical form depending on the precise nature of the engine controls. What is most likely to produce a significant change in the model behaviour is a change in structure of the injection process from the pilot and main used in this engine to a more complex pattern of multiple injections. Late injections that are timed to occur during the combustion process will have the greatest effect. However, the modelling process proposed and its power of representation, has shown itself capable with the overall diesel combustion process and the new fuelling strategy is likely to manifest itself as a change in detail.

Figure 18: Single NLARX structure training and validation results for five inputs

Figure 19: Illustration of regression of training and validation results for five inputs

7. CONCLUSION

In this paper a neural network approach is presented to predict the steady and transient PM of a medium duty diesel engine. Using PCA, it turns out that only five inputs are needed to predict the engine smoke production. The results show a sufficient predictive accuracy of the NLARX structure based on five inputs. Not only is the input count reduced, the prediction

performance has been improved by using PCA. Further work will be directed into reducing the input dimensionality by analyzing the steady and transient dynamics of PM formulation and the training data set by using signal processing methods.

REFERENCES

[1] Liu, Y. Diesel engine modelling and optimization for emission reduction, PhD thesis, University of Wisconsin-Madison, 2005.

[2] Kennedy, I. M. Models of soot formation and oxidation, Progress in Energy and Combustion Science. 1997, 23, 95-132.

[3] Bockhorn, H. Soot formation in combustion: Mechanism and models, Springer-verlag, Heidelberg, 1994.

[4] Wu, B., Prucka, R. G., Filipi, Z. S., Kramer, D. M. and Ohl, G. L. Cam-phasing optimization using artificial neural networks as surrogate models—fuel Consumption and NOx Emissions, SAE 2006-01-1512.

[5] He, Y. and Rutland, C.J. Application of artificial neural networks in engine modeling, International Journal of Engine Research, 2004, 5(4), 281-296.

[6] Maass, B., Stobart , R. and Deng, J. Diesel Engine Emissions Prediction with Parallel Neural Networks, American Control Conference, 2009.

[7] Bose, N. and Kumar, N. S. Prediction of engine emissions through fuzzy logic modelling, International Conference on Computational Intelligence and Multimedia Applications, 2007.

[8] Schilling, A., Amstutz, A., Onder, C. H., and Guzzella, L. A real-Time model for the prediction of the NOx emissions in DI diesel engines, Proceedings of the IEEE International Conference on Control Applications, Munich, Germany, Oct., 2006.

[9] Thompson, G. J., Atkinson, C. M., Clark, N. N., Long, T. W., and Hanzevack E., Neural network modelling of the emissions and performance of a heavy-duty diesel engine, IMECHE Part D: Journal of Automobile Engineering, Vol 214, pp. 111-126.

[10] Ouenou-Gamo, S., Ouladsine, M., Rachid, A. Measurement and prediction of diesel engine exhaust emissions, ISA Transactions , 1998, 37, 135-140.

[11] Funahashi, K., "On the Approximate Realization of Continuous Mappings by Neural Networks," Neural Networks, 2(3), 183–192, 1989.

[12] Polat, K., Güneş, S. , A hybrid medical decision making system based on principal component analysis, *K*-NN based weighted pre-processing and adaptive neuro-fuzzy inference system, 2006, 16, 913-921.

[13] Wang, X., Paliwal, K.K., Feature extraction dimensionality reduction algorithm and their applications in vowel recognition, Journal of Pattern Recognition society, 2003, 36(10), 2429-2439.

List of captions for the illustrations

Table 1: CSLA ramp time definition

Figure 1: Instrumented engine and the operator station

Figure 2: Random walk - Training and validation part distribution

Figure 3: CSLA test - Training and validation part distribution

Figure 4: Idle ramp test- Training and validation ramps

Figure 5: NRTC test

Figure 6: Illustration of training and validation data

Figure 7: Training set including RW, NRTC, ITF and CSLA

Figure 8: Validation set including RW, NRTC, ITF and CSLA

Figure 9: NLARX canonical structure

Figure 10: Single NLARX structure training and validation correlation results

Figure 11: Three-layer parallel NLARX training and validation correlation results

Figure 12: Fuzzy model training and validation correlation results

Figure 23: Correlation results for 9 input NLARX structure with data sampled at 10 Hz

Figure 14: Illustration of regression of training and validation results for nine inputs

Figure 15 Principal component analysis result

Figure 16: Single NLARX Structure training and validation results for six inputs

Figure 17: Illustration of regression of training and validation results for five inputs

Figure 18: Single NLARX Structure training and validation results for five inputs

Figure 19: Illustration of regression of training and validation results for five inputs

List of notations

A pre-exponential factor

C covariance matrix

c, d constant for reaction rate

$c(N)$, $d(N)$ variables which change with engine speed N

E Activation energy

F_M model performance index

m the sample mean

m_f unburned fuel mass fraction

m_a unburned oxider mass fraction

N engine speed

P_{ij} the transition probability from state i to state j

P_i the principal component along axis i

R the universal gas constant

R^2 coefficient of determination

$u(.)$ input

X_i the state of Markov chain

$y(.)$ output

$\hat{y}(.)$ predicted output

$\bar{y}$ the mean value of the output

Z_M training data set

λ_i eigenvector along axis i

θ parameters vector of the model

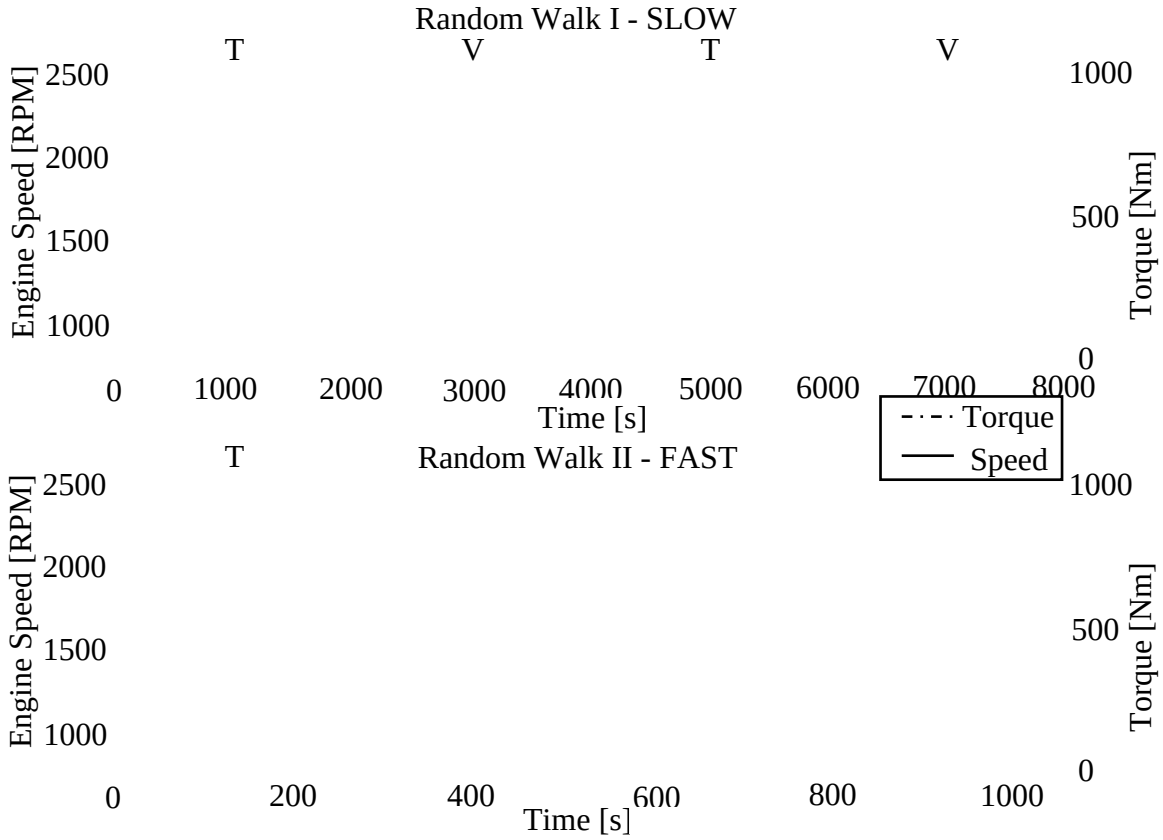


Figure 2: Random walk - Training and validation part distribution

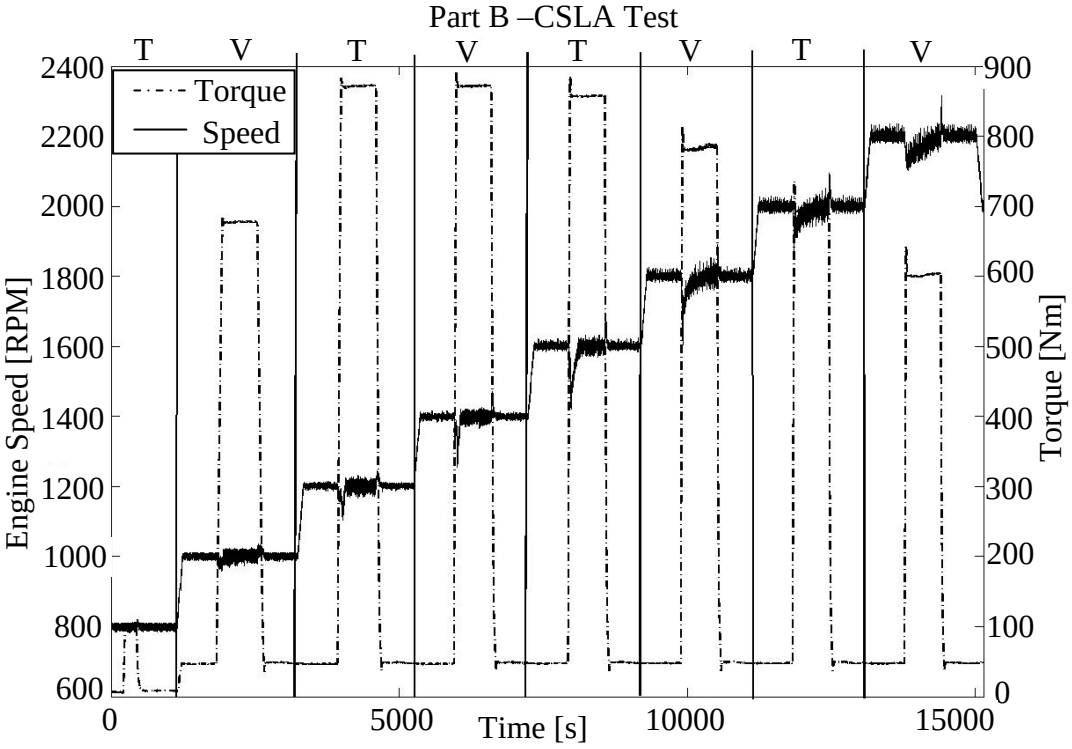


Figure 3: CSLA test - Training and validation part distribution

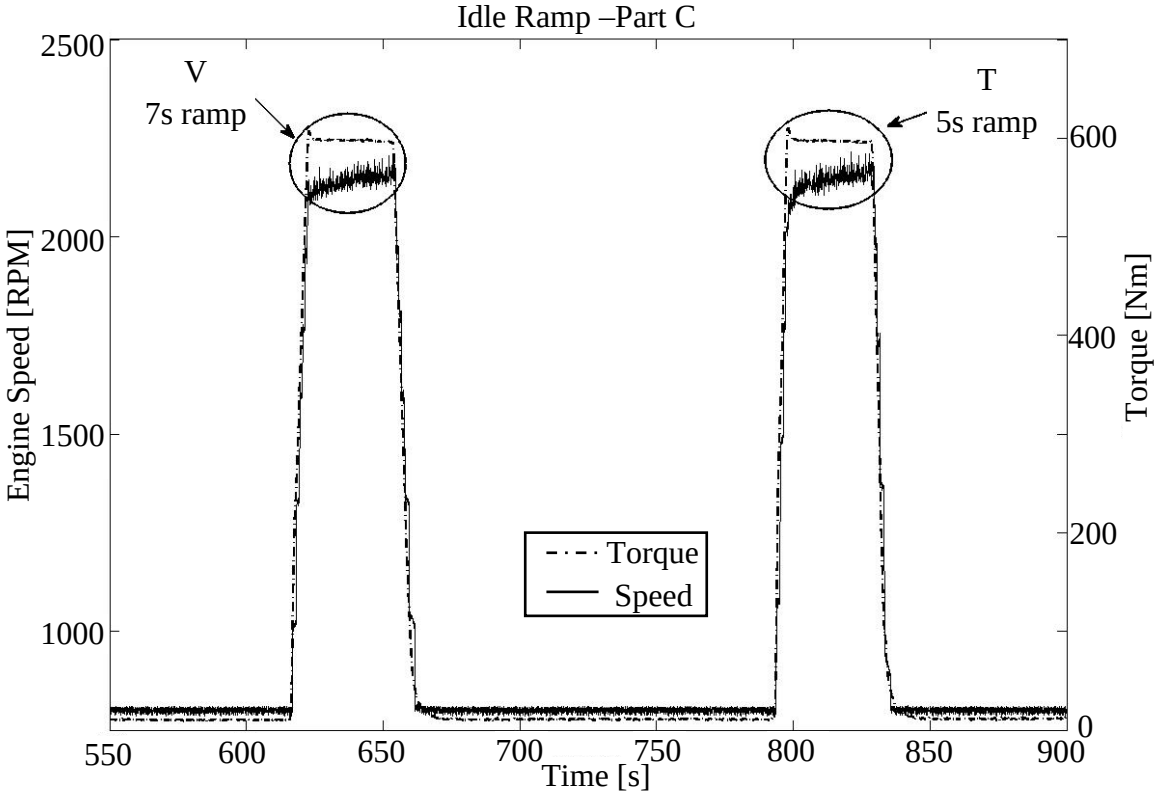


Figure 4: Idle ramp test- Training and validation ramps

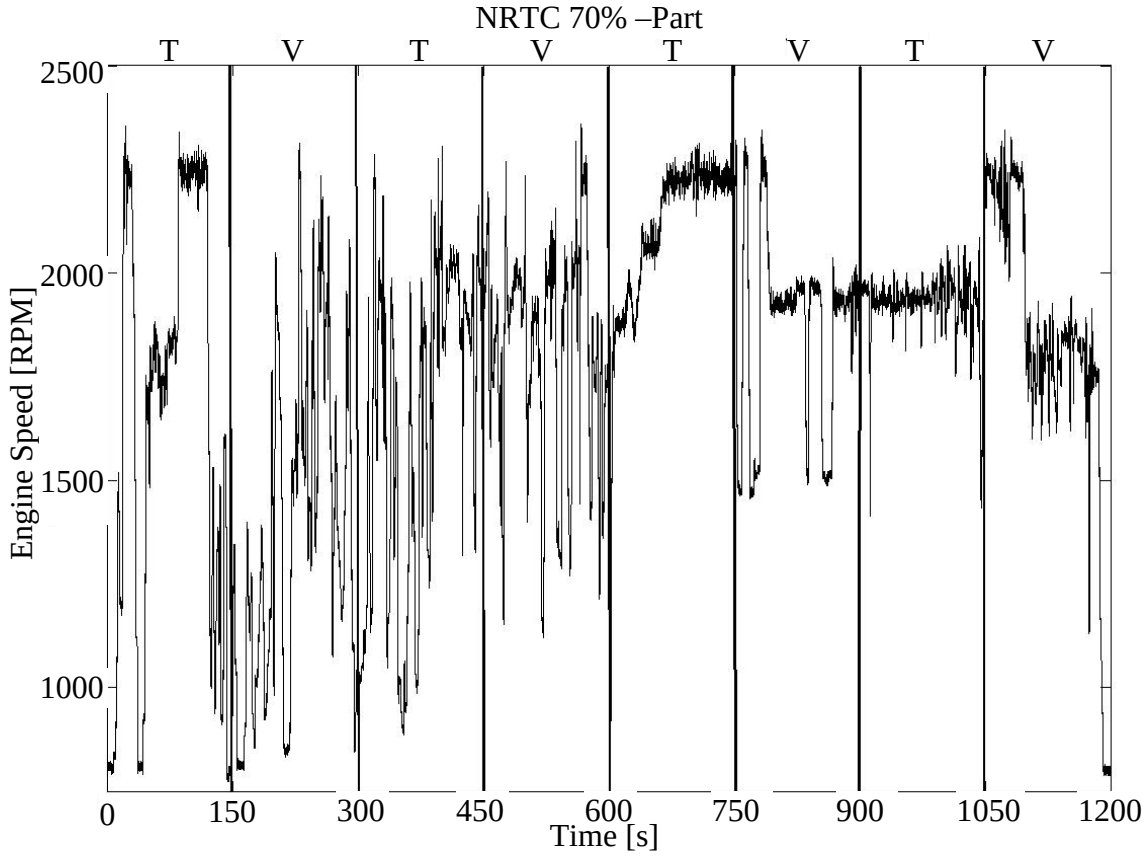


Figure 5: NRTC test

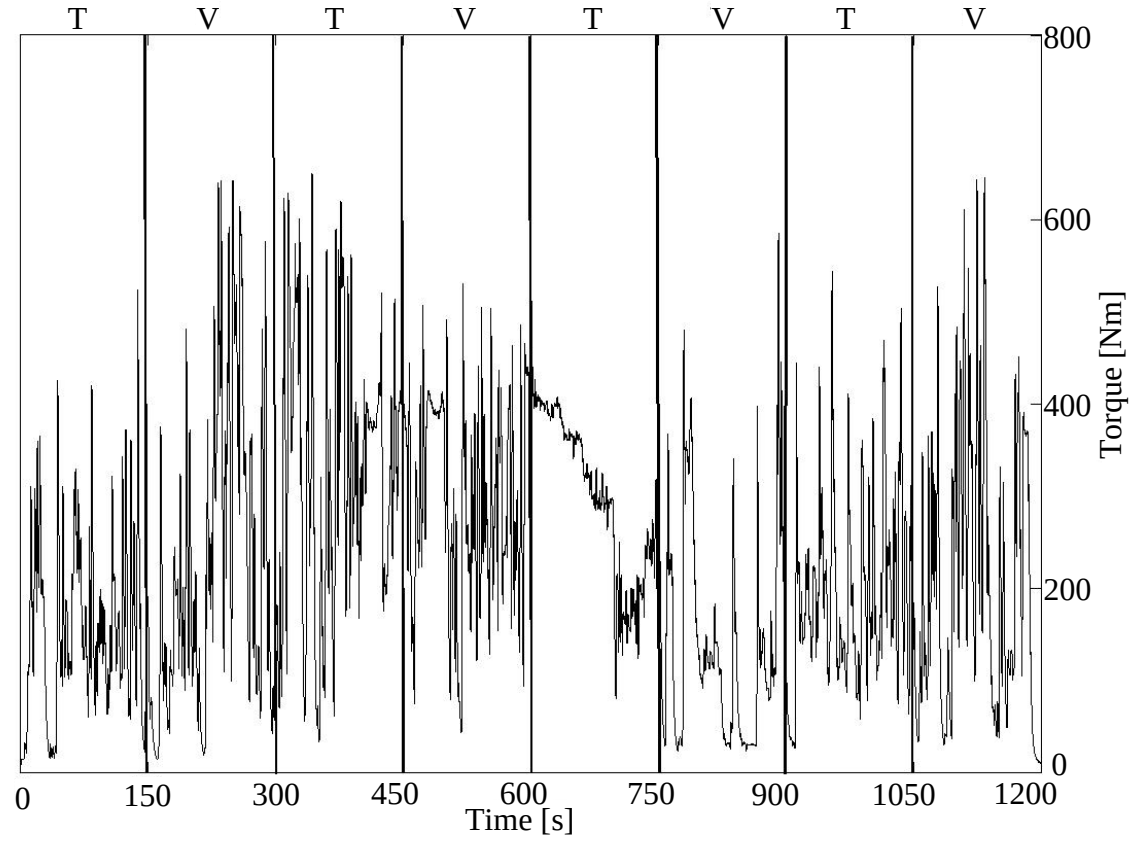
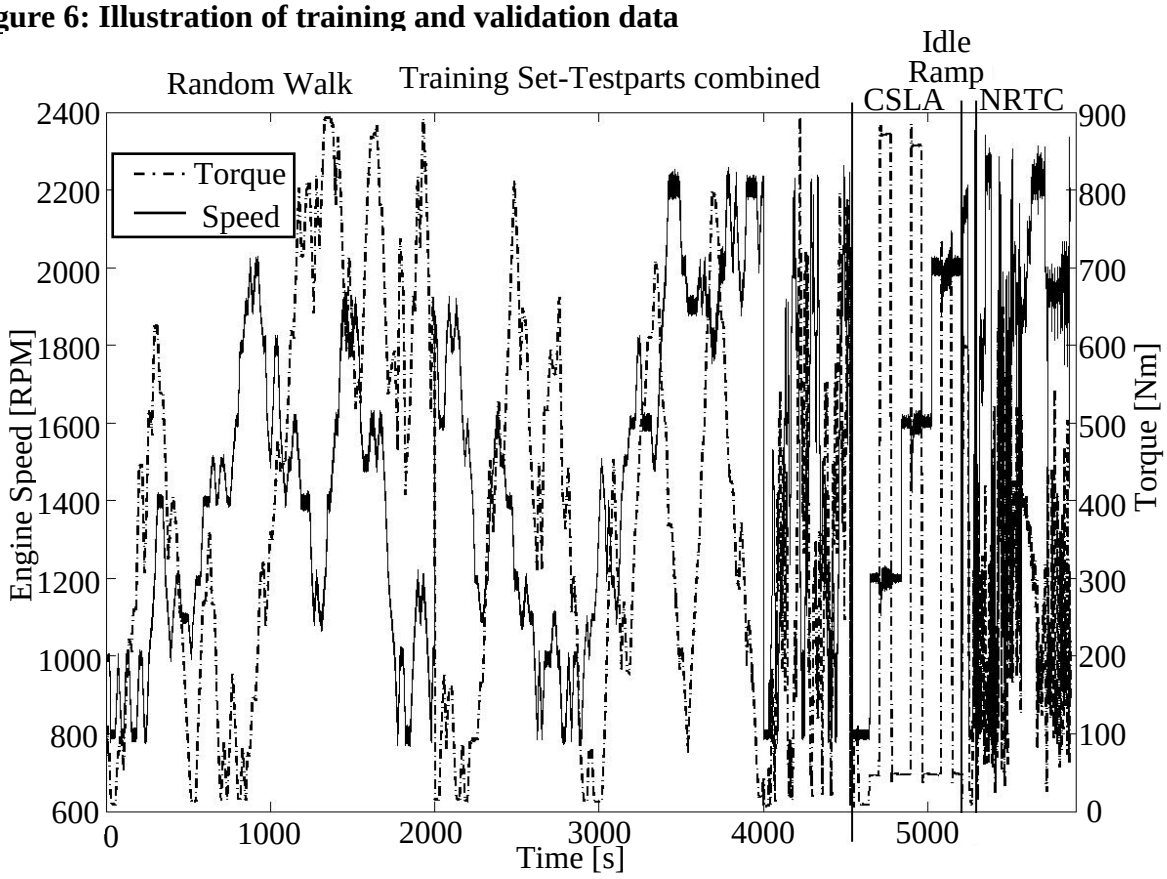


Figure 6: Illustration of training and validation data



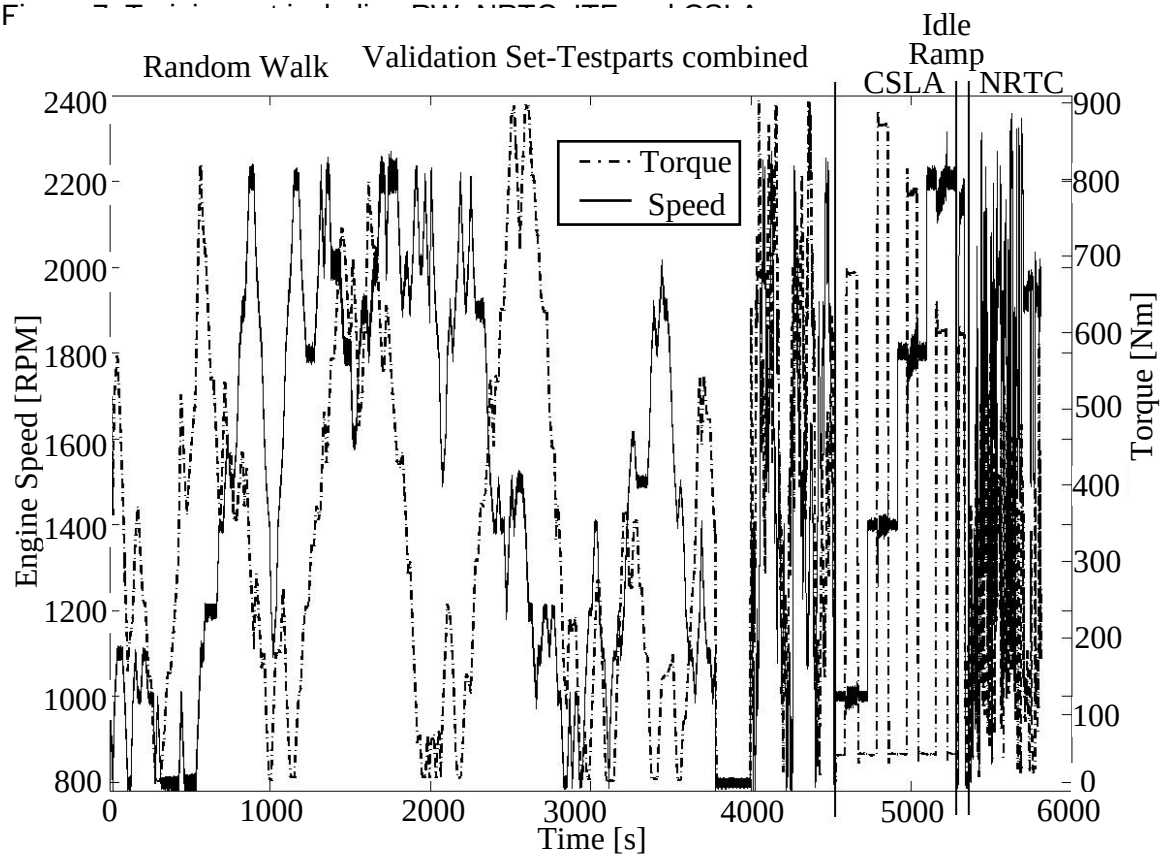


Figure 8: Validation set including RW, NRTC, ITF and CSLA

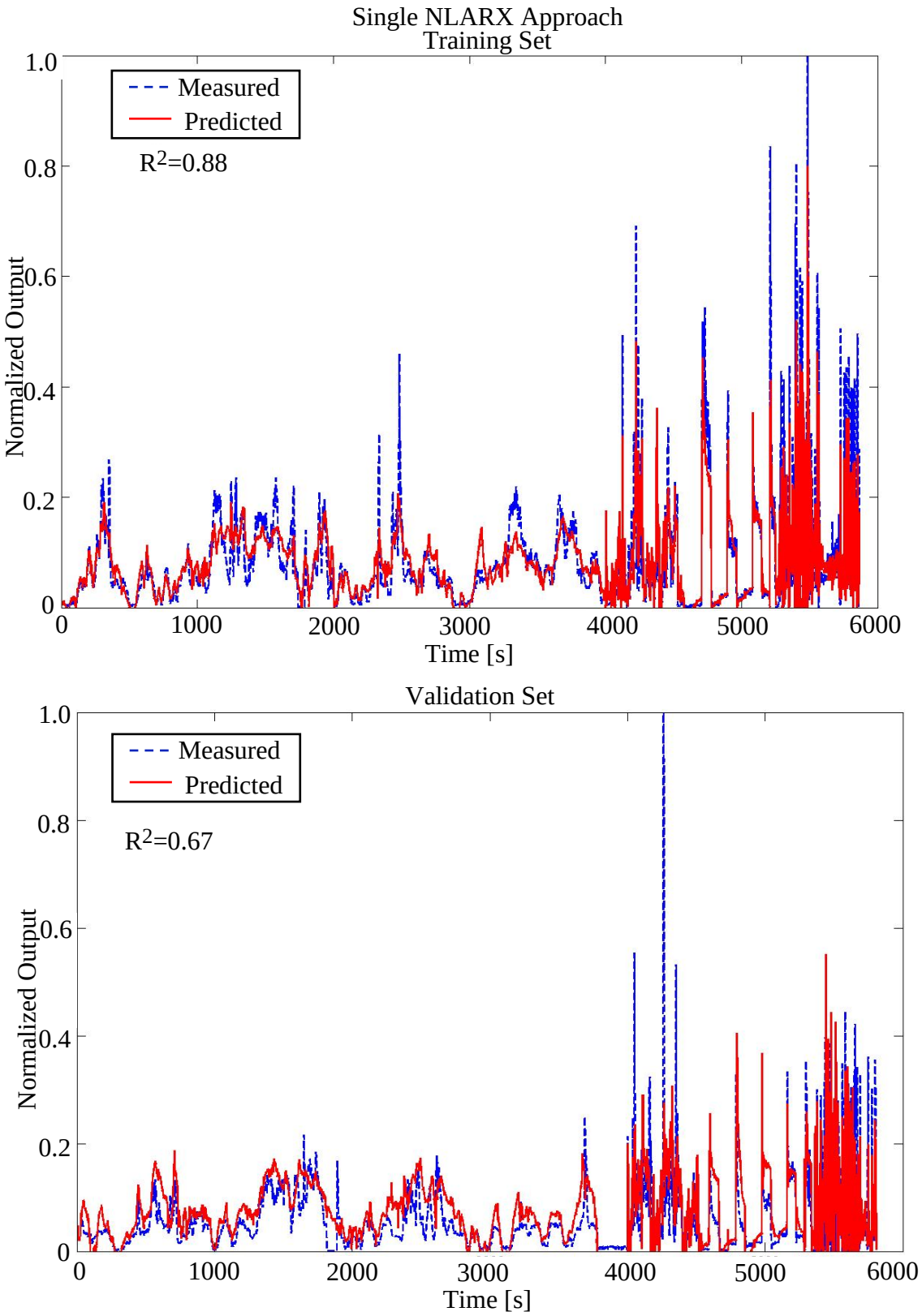
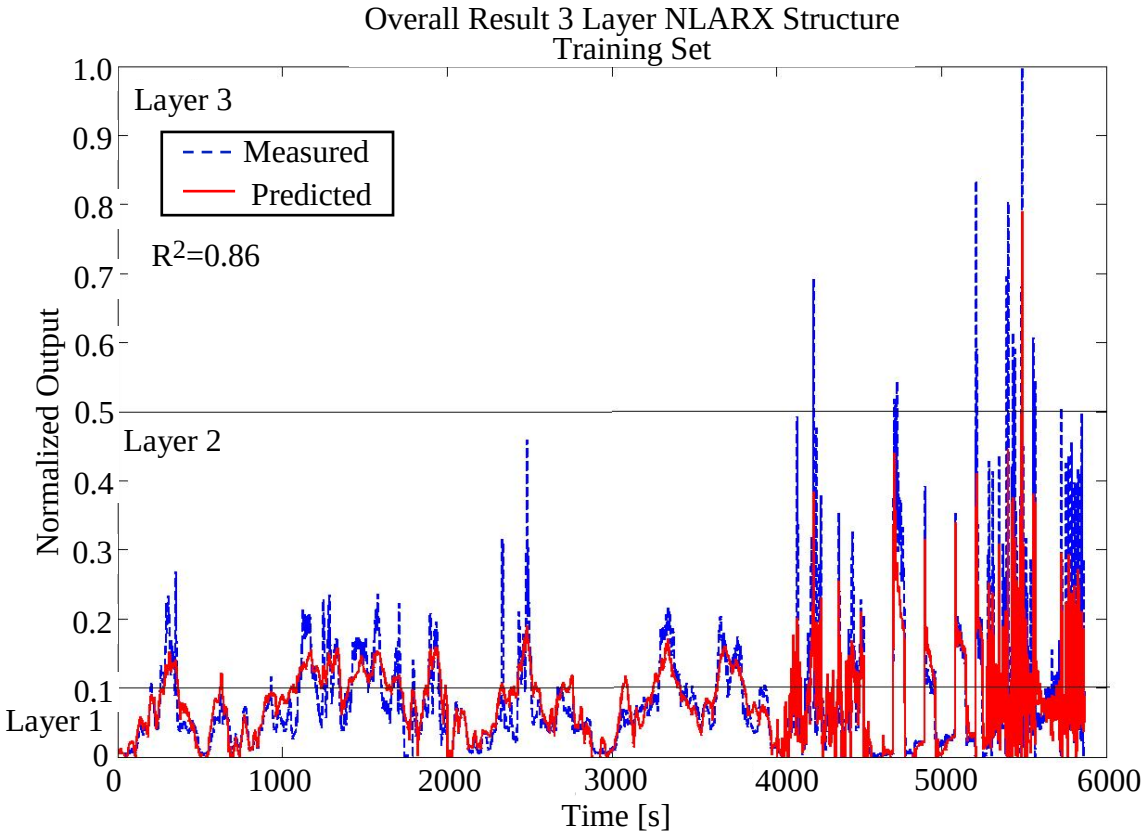


Figure 10: Single NLARX structure training and validation correlation results


Overall Result 3 Layer NLARX Structure
Training Set
Layer 3
Measured
Predicted
R2=0.86
Layer 2
Layer 1
Normalized Output
Time [s]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1000
2000
3000
4000
5000
6000

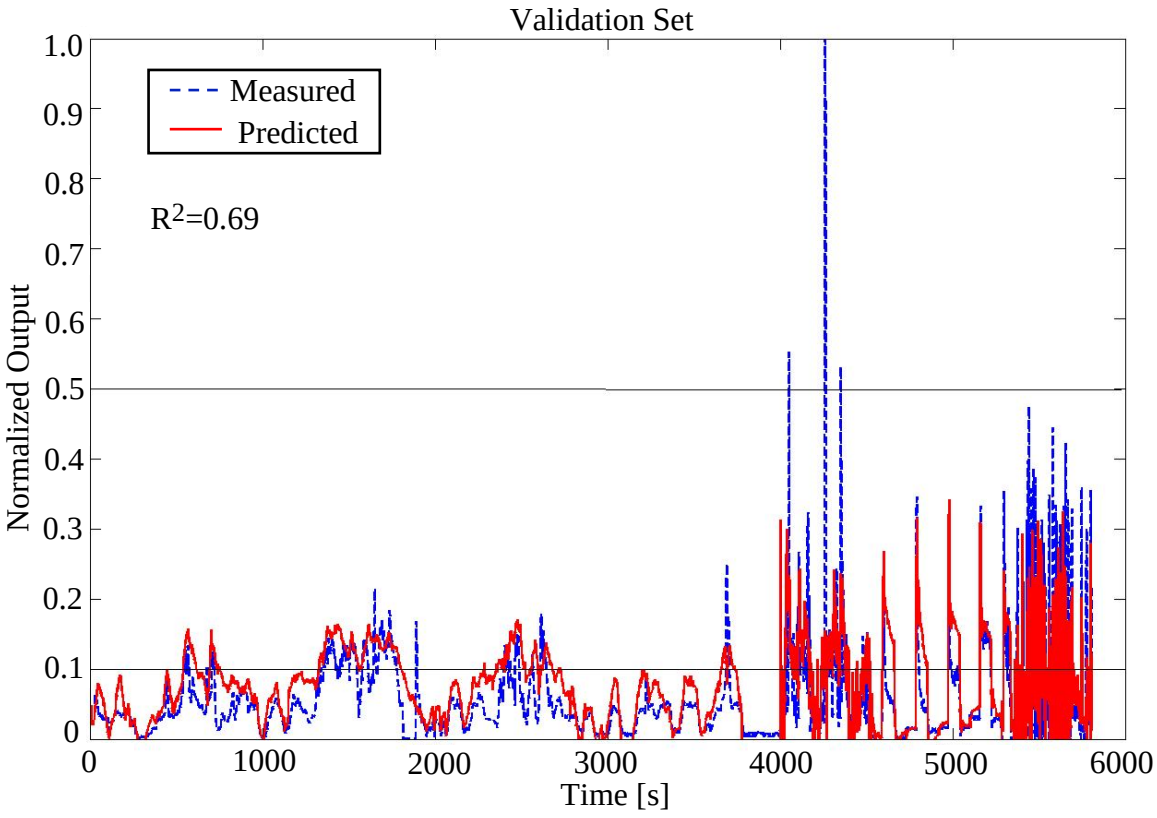

Validation Set
Measured
Predicted
R2=0.69
Normalized Output
Time [s]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1000
2000
3000
4000
5000
6000

Figure 11: Three-layer parallel NLARX training and validation correlation results

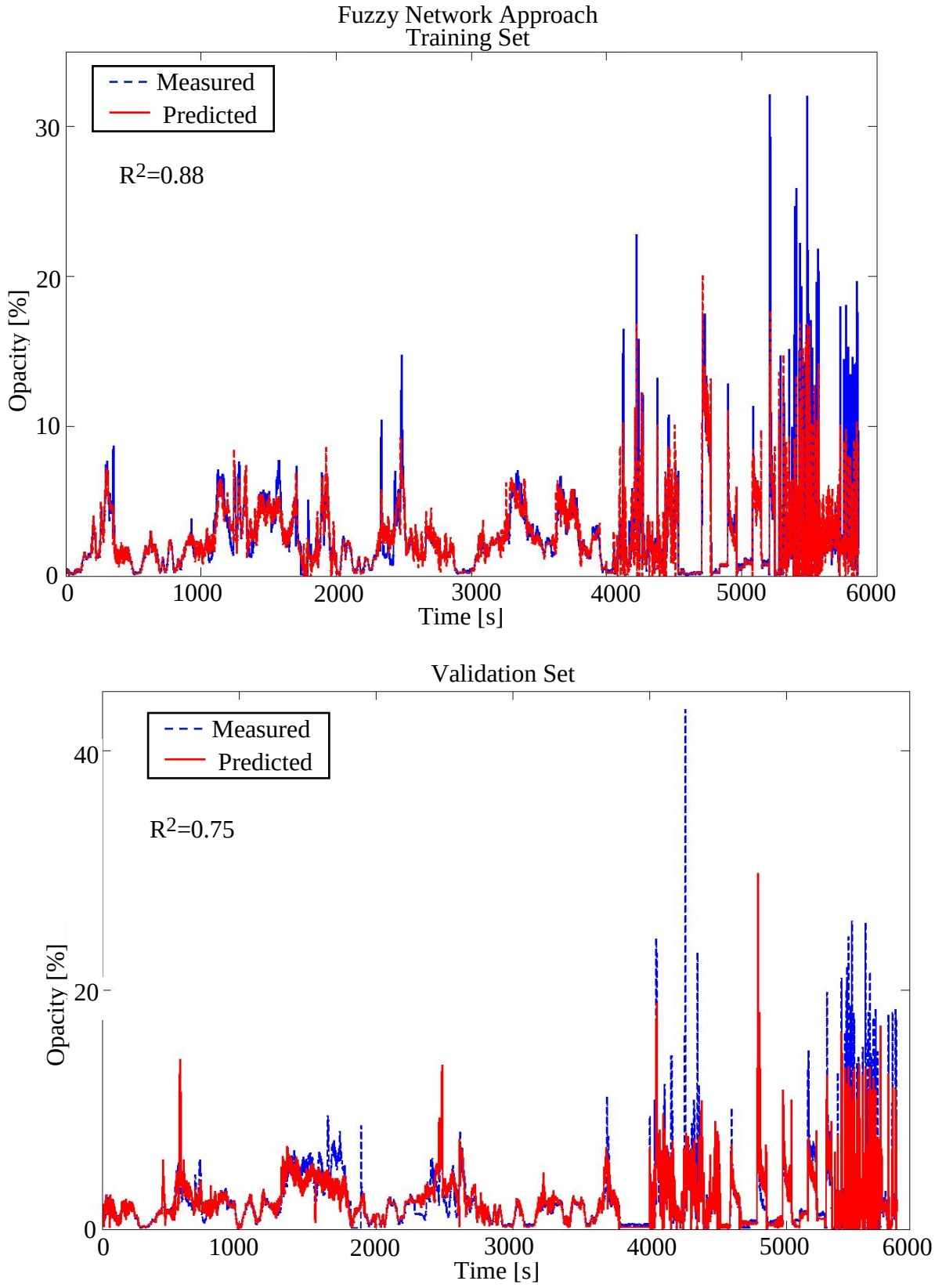


Figure 12: Fuzzy model training and validation correlation results

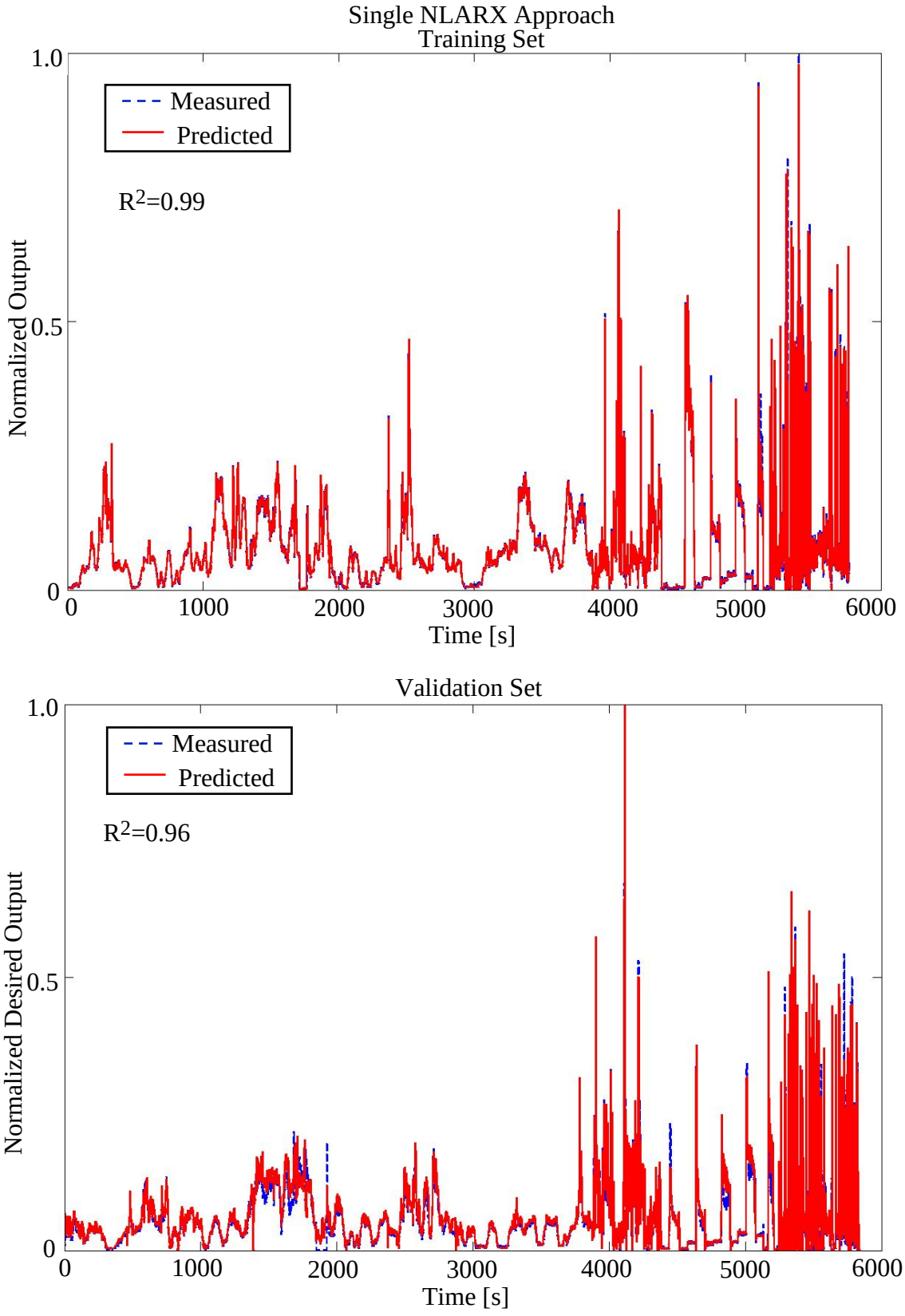


Figure 13: Correlation results for 9 input NLARX structure

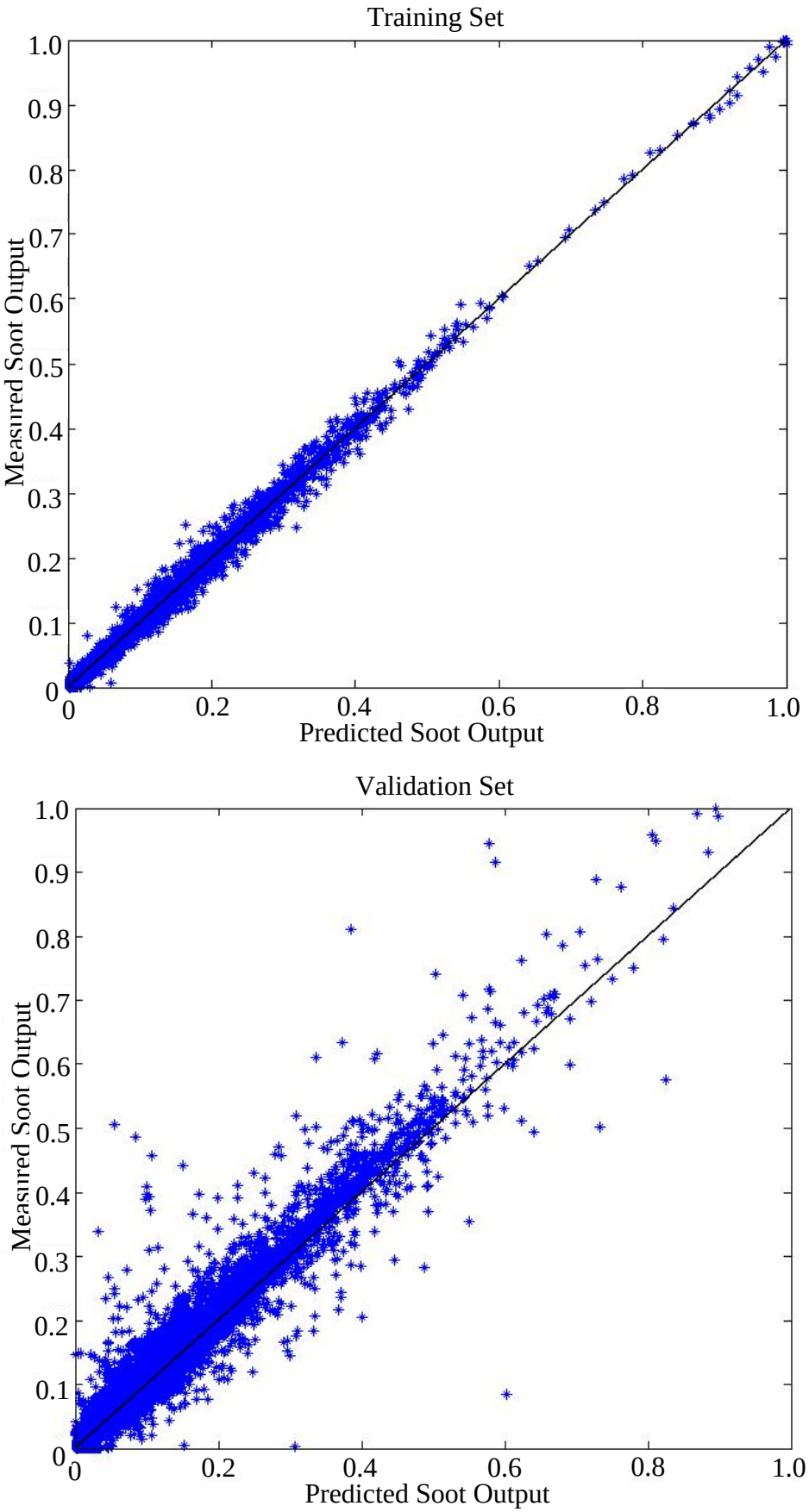


Figure 14: Illustration of regression of training and validation results for nine inputs

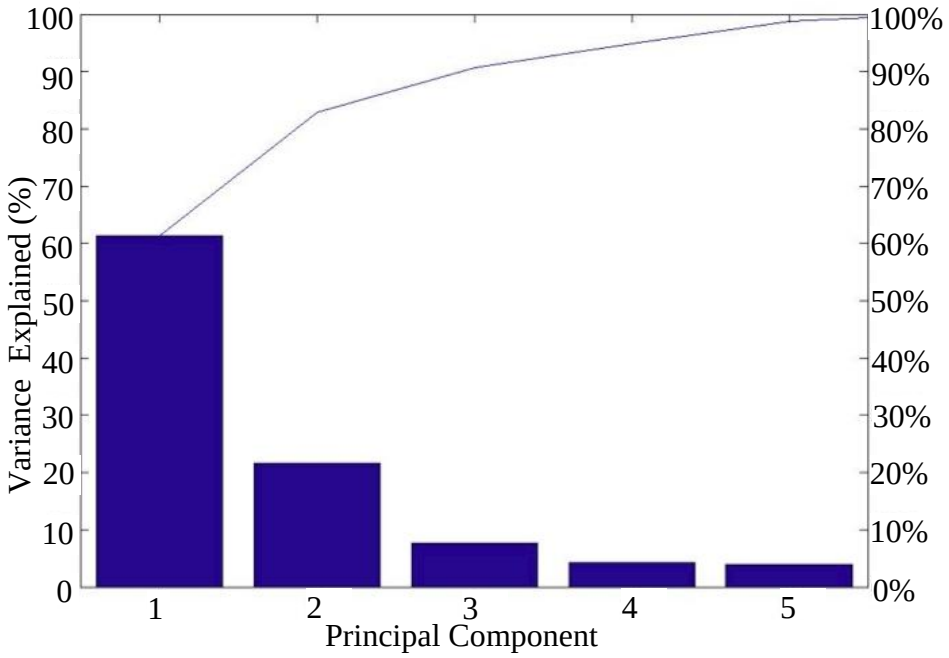


Figure 15 Principal component analysis result

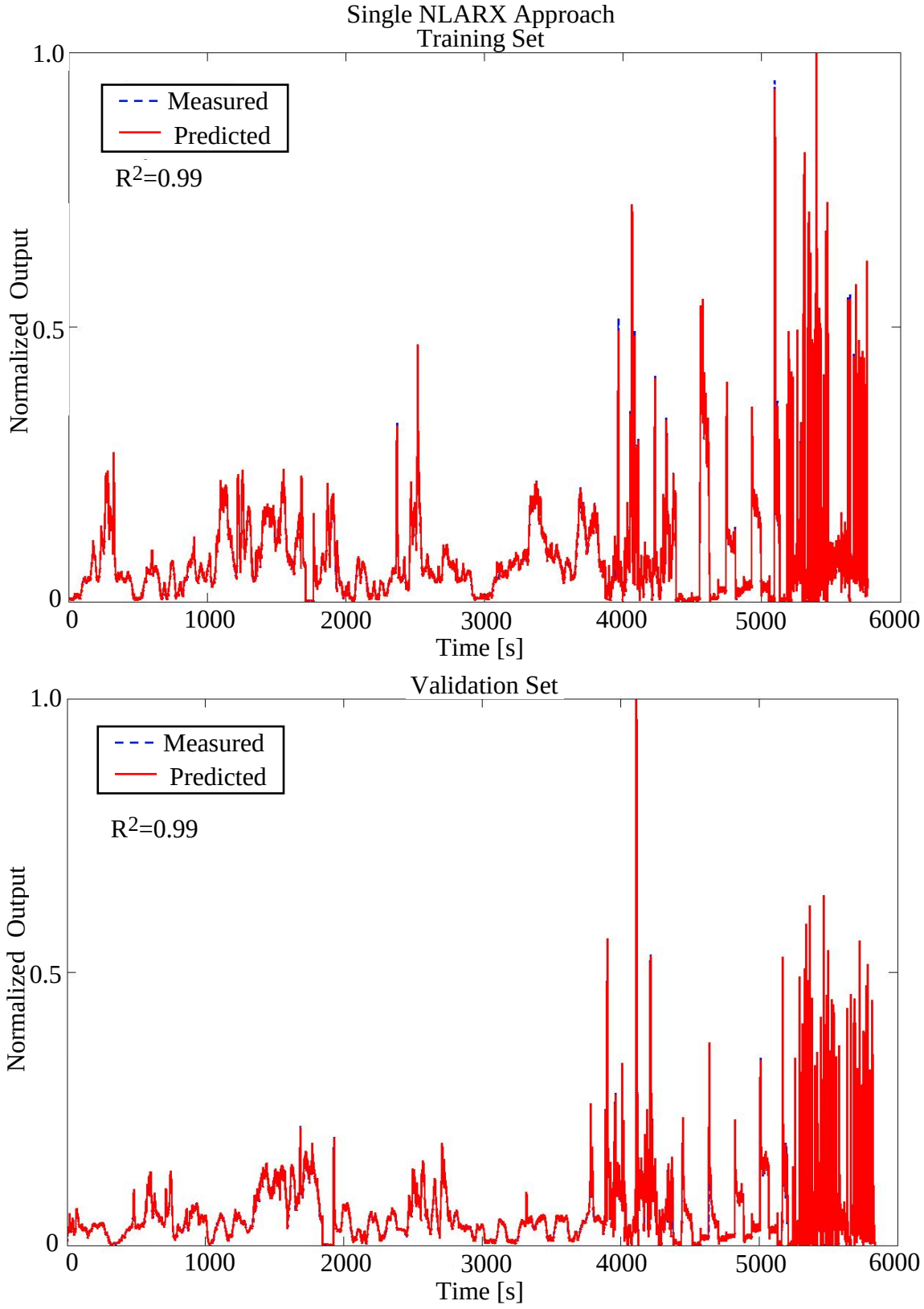
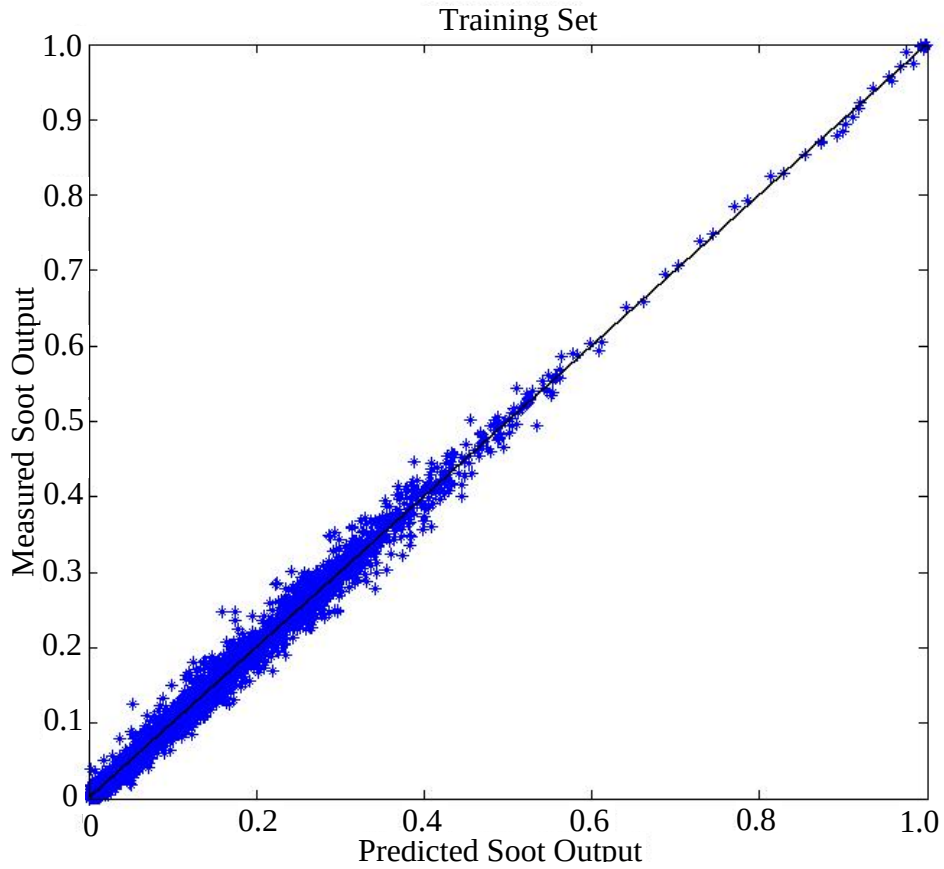

Single NLARX Approach
Training Set
Measured
Predicted
R2=0.99
Normalized Output
Time [s]
Validation Set
Measured
Predicted
R2=0.99
Normalized Output
Time [s]

Figure 16: Single NLARX Structure training and validation results for six inputs



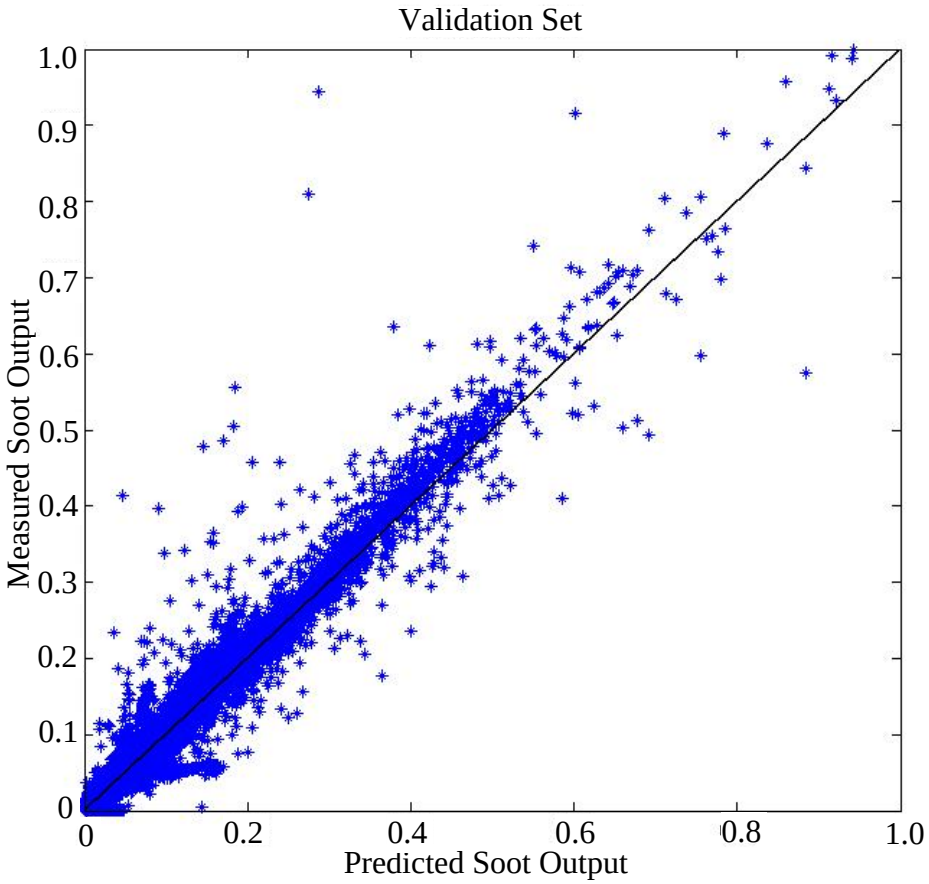


Figure 17: Illustration of regression of training and validation results for six inputs

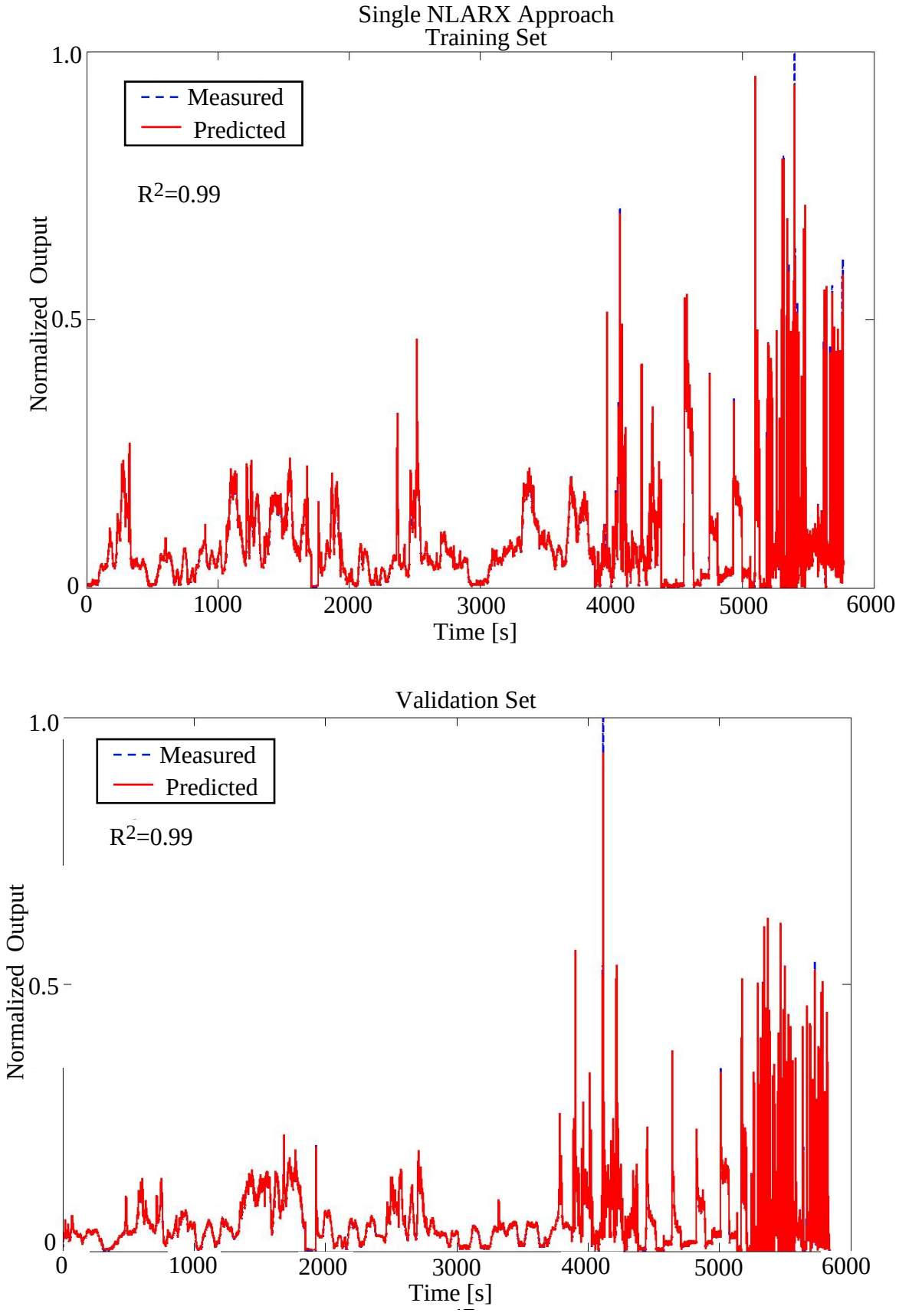


Figure 18: Single NLARX Structure training and validation results for five inputs

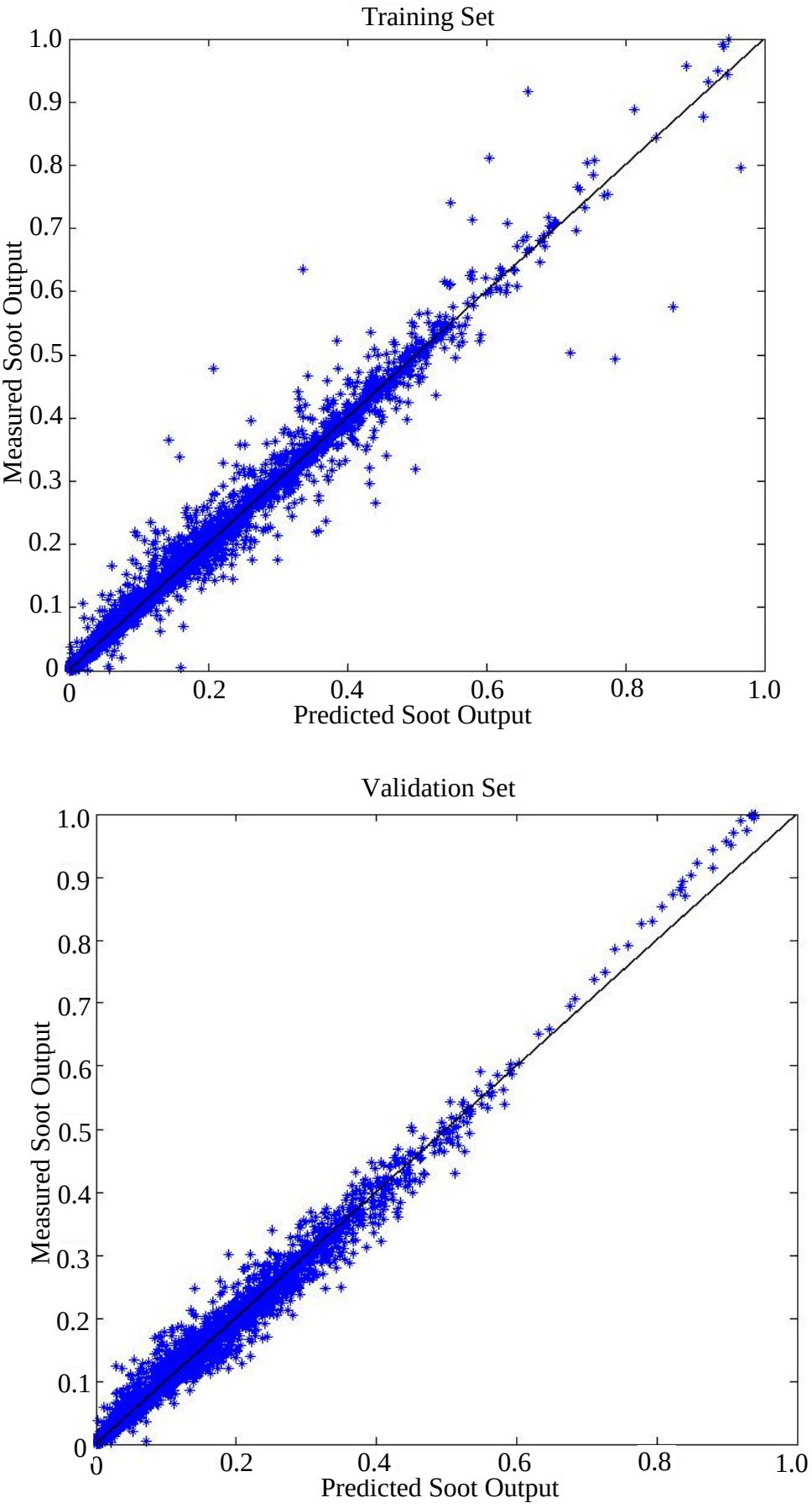


Figure 19: Illustration of regression of training and validation results for five inputs

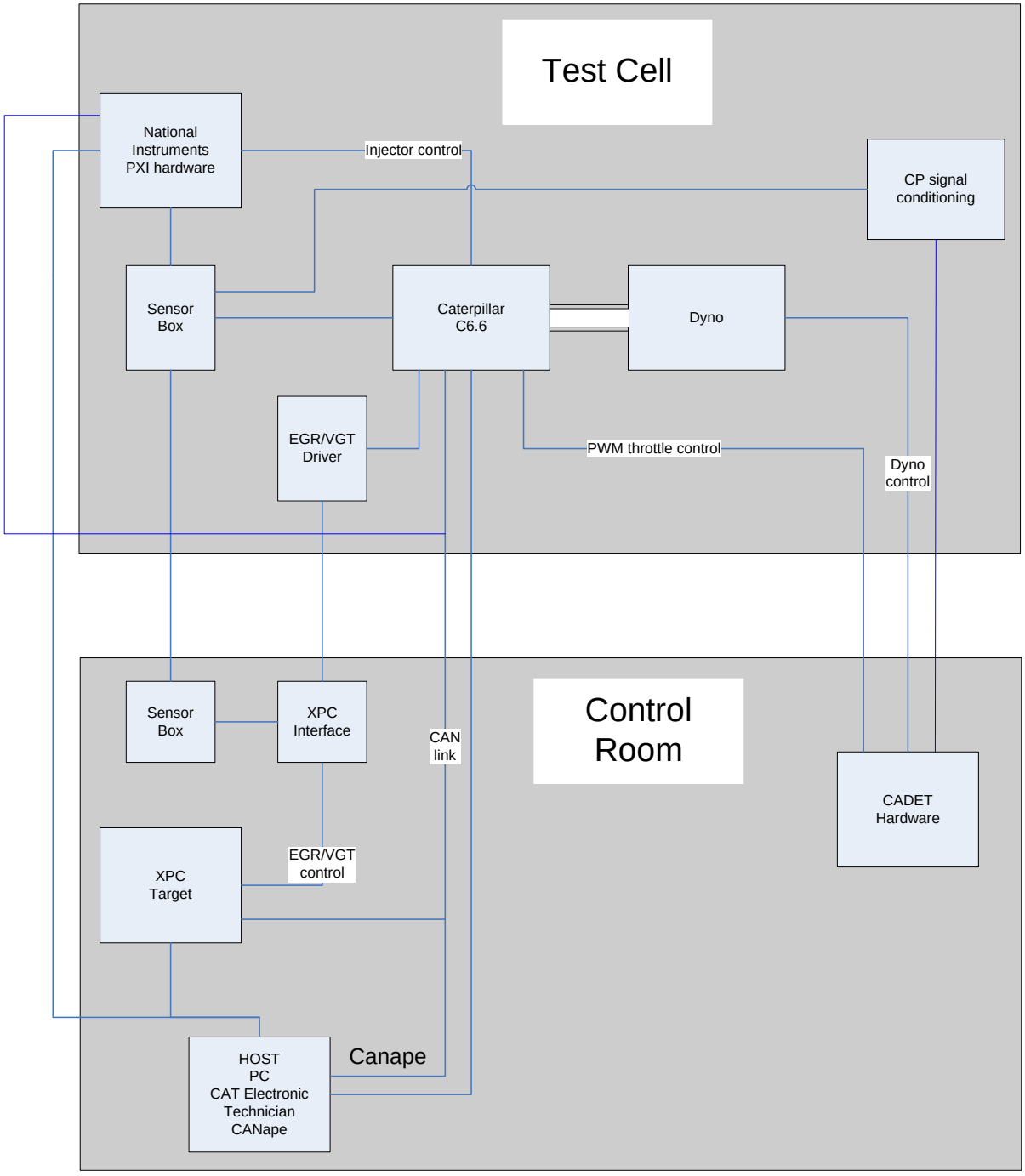


Figure 1: Schematic graph of the engine and operation station