

Development of recurrent neural networks for virtual sensing of NOx emissions in internal combustion engines

Ivan Arsie, Cesare Pianese, Marco Sorrentino

Department of Mechanical Engineering, University of Salerno

Copyright © 2009 Society of Automotive Engineers, Inc.

ABSTRACT

The paper focuses on the experimental identification and validation of recurrent neural networks (RNN) for virtual sensing of NO emissions in internal combustion engines (ICE). Suited training procedures and experimental tests are proposed to improve RNN precision and generalization in predicting NO formation dynamics. The reference Spark Ignition (SI) engine was tested by means of an integrated system of hardware and software tools for engine test automation and control strategies prototyping. A fast response analyzer was used to measure NO emissions at the exhaust valve. The accuracy of the developed RNN model is assessed by comparing simulated and experimental trajectories for a wide range of operating scenarios. The results evidence that RNN-based virtual NO sensor will offer significant opportunities for implementing on-board feedforward and feedback control strategies aimed at improving the performance of after-treatment devices.

INTRODUCTION

Automotive engines and control systems are more and more sophisticated due to increasingly restrictive environmental regulations. Particularly in Diesel and GDI engines, complex after-treatment devices, such as SCR and DPF, have been introduced to meet the imposed NOx and particulate emission limits. Furthermore, on-board diagnostics for both Diesel and spark ignition (SI) engines is becoming tighter and tighter; in fact engine faults could lead to performance degradation, which may cause higher emissions. More recently, multi-fuel powertrains control units are required to identify the actual fuel composition and to adapt the strategies accordingly.

To cope with the new issues associated with novel complex hardware and to improve both powertrain performance and after-treatment efficiency, engine design methods and control and diagnostic strategies have to be revised. A key role in achieving enhanced performance is played by on-board implementation of dedicated models, either as predictors or embedded in

model-based controllers or in diagnostics schemes. Virtual sensors (VSs) can be used as substitute of real sensors providing useful information about the actual process. In this scenario, virtual sensors are an interesting opportunity in achieving more challenging control and diagnostics targets.

In the last decade virtual sensors have been developed for many applications e.g. aerospace, biotechnological systems, environmental monitoring systems, to mention just few cases (Goodwin, 1999). VSs can be especially useful when: i) the direct measurements are impossible to perform, such as highly hazardous chemical or nuclear plants, (Sevilla and Pulido, 1998); ii) the available sensor might not guarantee the required sensing characteristics (i.e. accuracy, dynamic performance); iii) the sensor is too expensive or iv) does not fit within the sensing location. It is also evident that VSs can be used as feedback signal generators and their use provides opportunities to implement innovative control schemes and diagnostic. Actually, VSs must be designed in such a way as to simulate time-dependent processes and guarantee accuracy, stability and short computational time. In this work it will be shown that the proposed virtual sensor also allows a limited recourse to experiments, thus reducing costs and developing time.

Internal combustion engines are complex systems to be controlled with great precision and monitored continuously; moreover, their dynamic behavior requires both high sampling rates and rapid actuation. Therefore, on-board computational applications (e.g. models, controllers, monitoring and diagnostics algorithms) must be fast enough to ensure the required performance. VSs may support the development of new control and diagnostic concepts in achieving more challenging engine targets. Emission monitoring and control are the most complex tasks to be accomplished in an engine; this is mainly due to the complexity of simulating a process that entails advanced thermo-chemistry and fluid-dynamics within models and controllers. The state of the art mainly relies on burdensome applications where the process knowledge is stored in a large experimental database implemented on maps. A successful attempt to develop a steady-state first

principles-based NO virtual sensor was made by Brand et al. (2007). However, to fulfill the requirements of virtual sensor, neural networks seem to be valid candidates. Some examples of neural network-based virtual sensors for emissions simulation were proposed by Atkinson et al. (1998), Ayeb et al. (2005) and Subramaniam et al. (2008). A genetic programming approach was proposed by Alberer et al. (2005), who proposed a mixed approach by combining both first principles and experimental data. Although the results were very promising, the main drawback lies in the method complexity.

The authors have proposed the use of recurrent neural networks (RNNs) to simulate time varying air-to-fuel ratio (AFR) for SI engines (Arsie et al., 2002) that is suitable for virtual sensing for both control and diagnostics (Arsie et al., 2006; 2009). Starting from this application the same approach is presented to simulate the dynamic behavior of NO emission during SI engine transient maneuvers. In the following the main physical principles of NO formation are presented together with the developed RNN model. Also is highlighted the impact of the experimental analysis to provide measured data set informative enough to guarantee the proper model accuracy.

RECURRENT NEURAL NETWORKS FOR NOX PREDICTION

RNN ARCHITECTURE

The RNNs are derived from the static multi layer perceptron feed forward (MLPFF) networks by adding feedback connections among the neurons, thus introducing a dynamic effect in the computational system. Therefore, due to the non-linear mapping features of the MLPFFs, the RNNs are suitable for black-box non-linear dynamic modeling (Patterson, 1995; Haykin, 1999).

Depending upon the feedback typology, which can either involve all the neurons or only those located in the output and input layers, RNNs are classified into global, local or external recurrent neural networks (Haykin, 1999). The latter is the class used in this work with one hidden layer of nodes. The sketch of Fig. 1 gives the schematic representation of the following general form:

$$\hat{y}(t+1, \theta) = F[\hat{y}(t, \theta), \dots, \hat{y}(t-i+1, \theta), u(t), \dots, u(t-j+1), x(t), \dots, x(t-j+1)] \quad (1)$$

where $\hat{y}$ and u are the output and input variables, respectively, while θ is the vector of parameters (weights). The indices i and j are set once the lag spaces of both external input x and feedback variable $\hat{y}$ have been fixed. In Fig. 1 the lag space values are set equal to 2. The terms in square brackets of Eq. (1) compose the RNN regressors vector, which is representative of the past process dynamics. F is the non-linear mapping

operation performed by the neural network. The RNN described above is known in the literature as nonlinear output error model (NOE) (Nørgaard et al., 2000).

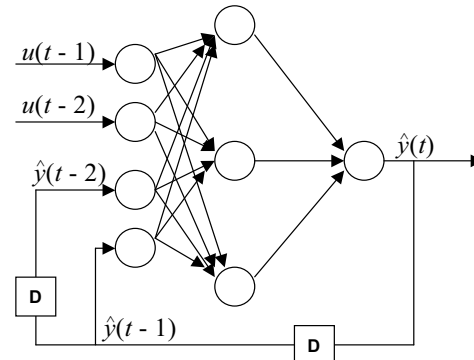


Fig. 1 – Structure of NOE RNN with one input variable, one output, one hidden layer and two output delays D .

NO FORMATION

Three mechanisms are generally considered for NO emissions from Internal Combustion Engines: prompt, fuel and thermal (Heywood, 1988). Prompt NO generates in the flame reaction zone, fuel NO is due to the presence of nitrogen based compounds in the fuel. Thermal NO formation is enhanced by high temperature and lean mixtures and takes place in the burned gas region. In case of automotive engines, both gasoline and Diesel fueled, thermal NO accounts for more than 90% of overall NO emissions.

The thermal NO formation process is usually modeled by means of the well known extended Zeldovich mechanism, that considers three reactions with seven species as main responsables for NO production (Heywood, 1988; Ramos, 1989; Arsie et al., 1998). More detailed models have been proposed, as the super extended Zeldovich mechanism by Miller et al. (1998), which accounts for 13 species and up to 67 reactions and can lead to a significant improvement of model accuracy.

Being controlled by chemical kinetic reactions, NO formation is a non-linear function of the burned gas temperature and O_2 concentration. Therefore the operating and control variables affecting NO emissions to be considered as model inputs are engine speed (n), engine load, spark advance (θ_s), air-fuel-ratio (AFR) and residual fraction. However, for the current application neither internal nor external exhaust gas recirculation (EGR) were considered as model inputs since the test engine was not equipped with neither EGR nor VVA systems. Furthermore, for SI engine application manifold pressure can be considered in place of throttle opening to express the dependence on engine load. Therefore the output and input variables of the virtual sensor are the following:

$$\hat{y} = NO$$

$$u = [n, p_{man}, \theta_s, AFR]$$

and the RNN model resulting from Eq. (1) will be:

$$\begin{aligned} NO(t, \theta) = F[& NO(t-1, \theta), \dots, NO(t-i, \theta), AFR(t-1), \\ & \dots, AFR(t-j), n(t-1), \dots, n(t-j), p_m(t-1), \\ & \dots, p_m(t-j), \theta_s(t-1), \dots, \theta_s(t-j)] \end{aligned} \quad (2)$$

It is worth noting that the output feedbacks in the regressors vector (i.e. NO in Eq. 2) are simulated by the network itself. Furthermore, the AFR signals can be both measured by a UEGO sensor or predicted by an AFR virtual sensor (Arsie et al. 2006, 2009). This way the RNN Model (RNNM) does not require any measurements of AFR nor NO as feedback to perform the online estimation.

RNN TRAINING METHODOLOGIES

Network training is performed by minimizing a cost function based on the mean squared error (MSE). A weight regularization term is added to improve model generalization:

$$E(\theta) = \frac{1}{2N} \sum_{t=1}^N (\hat{y}(t | \theta) - y(t))^2 + \frac{1}{2N} \cdot \theta^T \cdot \alpha \cdot \theta \quad (3)$$

where α is a scalar and represents the weight decay. The above function minimization can be conducted in either a batch or an adaptive fashion. The former is usually preferred at the initial development stage, whereas the latter is greatly beneficial to online RNN implementation as it allows to adapt network weights to the exogenous variations of the controlled/simulated system.

A proper RNN design can be ensured by appropriately addressing the following issues (Arsie et al., 2006): *i*) generate a training data set extensive enough to guarantee acceptable generalization of the knowledge retained in the training examples, *ii*) select the proper stopping criteria to prevent overtraining and *iii*) define the network structure with the minimum number of weights.

As far as the impact of point *i*) concerns the current application, NO dynamics can be well learned by the RNN estimator once the trajectories of engine variables (i.e. n , p_m , AFR and θ_s) described in the training-set are informative enough. This means that the training experiments on the test-bench have to be performed in such a way as to cover most of the engine working domain. Furthermore a proper description of both low- and high-frequency dynamics is necessary. Thus the required experimental profile is obtained by alternating steady operations of the engine with both smooth and sharp acceleration/deceleration maneuvers.

Point *ii*) can be approached by introducing the early-stopping method as stopping criterion (Nørgaard, 2000). This technique consists in interrupting the search for the minimum, once the MSE computed on a data-set different from the training one stops decreasing. Finally, point *iii*) is addressed according to a previous paper (Arsie et al., 2006), in which a trial and error analysis was performed to select the optimal network architecture in terms of hidden nodes and lag space. Although some theories about MLPFF network sizing are addressed in the specific literature (Kolmogorov, 1965), finding the best architecture for recurrent neural network is a more challenging task due to the presence of feedback connections and past input values (Nørgaard et al., 2000). Therefore, the recourse to an heuristic error-based approach is suitable to satisfactorily fulfill point *iii*).

EXPERIMENTAL SET-UP

The RNN Model (RNNM) for NO prediction was trained and tested against transient data sets measured on the engine test bench at University of Salerno.

The experiments were carried out on an SI engine, 4 cylinders, 1.2 liters, with multi-point injection. The test bench is equipped with a Borghi & Saveri (FE-200S) eddy current dynamometer. A data acquisition system, based on National Instruments cards (PCI MIO 16E-1) and Sample & Hold Amplifier (SC-2040), measures engine variables with a sampling frequency up to 10 kHz. An AVL gravimetric balance is used to measure fuel consumption in steady-state condition and to calibrate the injector flow rate. The engine control system was replaced by a dSPACE® MicroAutobox (MABX) system and a power conditioning unit. Such a system allows to control all the engine tasks and to customize the control algorithms. To guarantee the controllability and reproducibility of the transient maneuvers, both throttle valve and engine speed are controlled by an AVL PUMA engine automation tool (see Fig. 2).

A fast response chemiluminescence analyzer Cambustion CLD 500 was used to measure NO emissions while the exhaust AFR was sensed by an ETAS Lambda Meter LA4, equipped with a Bosch LSU 4.2 UEGO sensor. Both the instruments were placed immediately after the exhaust valve of the first cylinder (see Fig. 3). This choice allows removing the dynamic effects induced by gas transport and mixing phenomena occurring in the exhaust pipes. Also non predictable effects generated by cylinder-to-cylinder unbalance due to uneven processes such as air breathing, thermal state and fuel injection can be neglected. Therefore, the time shift between spark timing and NO measurement mostly accounts for combustion and exhaust valve phasing.

The offset of the time delay between the model inputs and output is a critical point for RNN training. For the current application, apart from the NO signal, an