Neural Network Models for Virtual Sensing of NO_x Emissions in Automotive Diesel Engines with Least Square-Based Adaptation

Ivan Arsie*[a], Andrea Cricchio*, Matteo De Cesare, Francesco Lazzarini*, Cesare Pianese*, Marco Sorrentino***

** Department of Industrial Engineering, University of Salerno,*
via Giovanni Paolo II, 132, Fisciano (SA) 84084, Italy
*** Magneti Marelli Powertrain,*
via del Timavo, 33, Bologna, 40134, Italy

Abstract: To meet current Diesel engine pollutant legislation, it is important to manage after-treatment devices. The paper describes the development of Neural Network based virtual sensors used to estimate NO_x emissions at the exhaust of automotive Diesel engines. Suitable identification methodologies and experimental tests were developed with the aim of meeting the conflicting needs of feasible on-board implementation and satisfactory prediction accuracy. In addition, since the prediction of control-oriented models is typically affected by engine aging and production spread as well as components drift, least square technique features were exploited in order to overcome these issues by adapting the virtual sensor output. The NO_x adaptive virtual sensor was tested via comparison with experimental data, measured at the engine test bench on a turbocharged common-rail automotive Diesel engine. Furthermore, besides model validation, the experimental measurements were modified to simulate a sensor drift in order to enable full assessment of the proposed LS-based algorithm adaptation capabilities.

Keywords: Modeling, supervision, control and diagnosis of automotive systems; Engine modelling and control; Automotive system identification and modeling; NOx emissions; Neural Networks; Least Square adaptation.

[a] Corresponding author. E-mail address: iarsie@unisa.it
Tel. +39 089964080; fax +39 089968781

1. INTRODUCTION

To meet current regulations for NO_x emissions in Diesel engines, complex after-treatment devices have been introduced, such as the selective catalytic reduction (SCR) system and the lean NO_x trap (LNT). In particular, the SCR management requires NO_x evaluation upstream and downstream of the catalyst, for both control and diagnosis purposes, in order to evaluate the right amount of urea to be injected and to avoid ammonia slip[1] (Willems et al., 2007; De Cesare et al., 2009). The physical sensors that are usually applied for these issues simultaneously provide a measurement of the relative air-to-fuel ratio (lambda) and NO_x concentration and have a major role in the SCR management (Blanco et al., 2014). Nevertheless, for those applications where the physical sensor is either too expensive or impractical, virtual sensors (VS) can be applied instead. In such cases, a NO_x sensor is usually located at the SCR outlet, to observe the occurrence of NO_x and ammonia slip, while feed-forward control of the urea injection is performed using VS to estimate NO_x at the SCR inlet.

Virtual sensors are based on real-time models with fast dynamic response and offline prediction capabilities (Johri et al., 2014). They are a good solution to reduce the hardware complexity of the engine management system (EMS) and can be useful especially in cases where direct measurement of the signal is not possible or whenever the available sensor might not ensure the required sensing characteristics.

Actually modelling of engine exhaust emissions is a complex task to be accomplished due to the involvement of thermo-chemical and fluid-dynamic processes. Therefore, on- board computational applications (e.g. control and diagnostics algorithms) must be fast enough to ensure the required performance. A successful attempt to develop a steady-state first principles-based NO_x virtual sensor was made by Brand et al. (2007). Current state of the art mainly relies on static maps or polynomial models where the process knowledge is stored in a large experimental database, composed of both steady-state and transient data. Tscahnz et al. (2010) proposed a polynomial model based on a steady- state map coupled with a dynamic model to account for the emissions deviation during transients. Del Re et al. (2005) proposed a global formulation for NO_x emissions of a Diesel engine based on system identification tests. A genetic programming method was proposed by Alberer et al. (2005), based on a mixed approach by combining both first principle and experimental data. Although the results were very promising, the main drawback lies in the method complexity.

[1] The ammonia slip is an undesired process in SCR operation: it corresponds to the emission of ammonia (NH_3) out of the SCR that takes place when the amount of urea injected into the SCR is higher than that reacting with the available NOx and O_2.

Apart from maps and polynomial regression models, neural networks have proved to be valid candidates to fulfill the requirements of virtual sensors. Examples of neural network based virtual sensors for engine emissions were proposed by Atkinson et al. (1998), Traver et al. (1999). The authors themselves have proposed the use of recurrent neural networks (RNNs) to estimate NO_x emissions during engine transients in case of both SI (Arsie et al., 2010) and Diesel engines (De Cesare et al., 2011, Arsie et al., 2013a).

In addition to the previous studies, this paper investigates methodologies to confer adapative features to a NO_x virtual sensor, placed upstream of the SCR. The virtual sensor receives as input operating and control variables detected by the EMS. In addition, the intake O_2 concentration estimated by an observer developed by the authors in a previous work (Arsie et al. 2013b) is also fed to the VS, as shown in Figure 1.

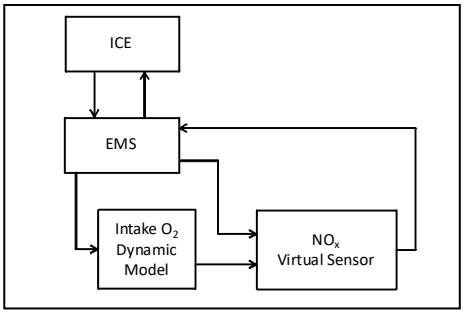

Fig. 1. Modeling structure for the on-board estimation of NO_x – The inputs are fed by the EMS and the O_{2_intake} observer.

It can be expected that the VS accuracy will be affected by variations of the internal combustion engine (ICE) operation due to variable environmental conditions and driving cycles as well as plant aging and production tolerances. Therefore, in order to ensure satisfactory VS accuracy during the entire ICE lifetime, the adaptation of the parameters is a key issue (Pavkovic et al., 2006).

The adaptive algorithms that are usually applied to update system parameters are the LS (least square), the NLMS (normalized least mean square), the RLS (recursive least square) and the Kalman filter (KF). Among these, the LS has probably become the most popular for its robustness, good tracking capabilities and simplicity in stationary environment. On the other hand, the RLS is the most suited for non-stationary environment with high convergence speed but at the cost of higher complexity. Therefore, a trade-off is required with respect to convergence speed and computational complexity and the NLMS provides the best compromise to these issues (Thenua et al., 2010).

The concept of adaptive control in the ICE and virtual sensors fields has not been much dealt with in literature, except in a few studies. Isermann et al. (2001) focused on adaptation algorithms for grid-based look-up tables for multi-dimensional nonlinear modeling of NO_x emissions in Diesel engines and adaptive feedforward control of spark advance in

gasoline engines. Guardiola et al. (2013) proposed the use of adaptive look-up tables applied to NO_x estimation in Diesel engines. Further applications of adaptive control have been investigated for other engine components, such as electronic throttle (Pavkovic et al., 2006) and lithium-ion battery for hybrid vehicles (Liu et al., 2012).

The paper is organized as follows. After an overview of NO_x formation mechanism and SCR management, the experimental setup and the RNN model formulation are described. Afterwards the adaptive control concepts and their application to the RNN virtual sensor are presented and the results of the model validation vs. experimental data are discussed.

2. NO_x FORMATION AND CONTROL

2.1 NO_x Formation

Besides particulate matter, nitrogen oxides are the most critical pollutants produced by Diesel engines. They are composed of nitric oxide NO and nitrogen dioxide NO_2, collectively referred as NO_x.

NO_x formation is affected by three different mechanisms: prompt, fuel and thermal (Heywood, 1988). Prompt NO_x is generated in the flame reaction zone, while fuel NO_x is due to the presence of nitrogen based compounds in the fuel. The thermal mechanism is the most important in Diesel engines and takes place in the burned gas region where high temperatures enhance the reaction of N and O_2 from air producing NO_x. In case of automotive engines thermal NO_x, which are mostly composed of Nitrogen Monoxide (NO), amounts to 90% of overall NO_x emissions. The principal reactions governing NO formation from molecular nitrogen and oxygen during combustion are usually modelled by the Zeldovich mechanism, which considers three reactions with seven species as main responsible for NO production (Heywood, 1988; Ramos, 1989).

The process presents a strong, non-linear dependence with respect to both in-cylinder temperature and oxygen concentration. Thus, either precise measurements or estimates of these variables are needed to achieve satisfactory model accuracy. Other operating and control variables with the largest effect on NO_x emissions are: engine speed, boost pressure, air mass flow, injected fuel mass, injection timing, rail pressure, air-to-fuel ratio (AFR) and exhaust gas recirculation (EGR) valve position.

As the number of input variables increases, the complexity of the RNN model and the computational burden related to model identification and simulation increases accordingly. However, the whole set of mentioned variables contains redundant information and therefore, in practice, only a selection of them would be needed as model inputs. Particularly for the current application, the signals of EGR valve position, boost pressure and AFR were comprehended by the estimation of the intake oxygen concentration performed by an observer (Arsie et al., 2013b). This is because the estimation gives direct information on both the oxygen content and the residual mass fraction in the combustion chamber. The rail pressure signal was also not considered, being directly correlated to engine speed and injected fuel mass.

This leaves the following five variables as model inputs:

- engine speed;
- oxygen concentration in the intake manifold;
- air mass flow;
- injected fuel mass;
- injection timing.

It is worth noting that, apart from the intake O_2 concentration, all the mentioned variables are usually measured on production vehicles, thus no additional sensors are required for the virtual sensor application.

2.2 SCR System

Selective catalytic reduction (SCR) of NO_x with ammonia (NH_3) is the most selective and active reaction for the removal of NO_x in the presence of excess oxygen, producing N_2 and water

$$\begin{aligned} &4NH_3 + 4NO + O_2 \rightarrow 4N_2 + 6H_2O \\ &2NH_3 + NO + NO_2 \rightarrow 2N_2 + 3H_2O \\ &4NH_3 + 3NO_2 \rightarrow 3.5N_2 + 6H_2O \end{aligned} \quad (1)$$

Ammonia can be generated on board via rapid hydrolysis of urea (Walker et al., 2003). Urea from a storage tank is injected into the engine exhaust stream. In order to achieve a high efficiency for the reduction reactions, it is important to guarantee a stoichiometric ratio between nitrogen oxides and ammonia. A deficiency of ammonia leads to an incomplete reduction of NO_x, whilst an excess of ammonia leads to deposits formation that could cause clogging of the pipe or ammonia slip. Therefore, for the calculation of the reducing agent injection rate, it becomes very important to measure or predict the NO_x concentration upstream of the SCR.

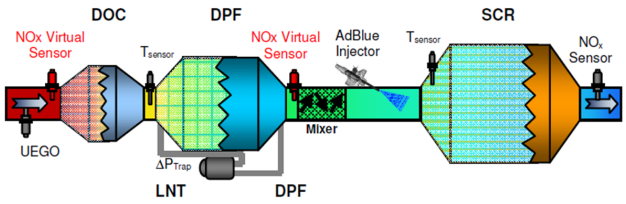

Fig. 2. Possible after-treatment layout for an EURO 6 Common rail Diesel engine, composed of Diesel Oxidation Catalyst (DOC), Diesel Particulate Filter (DPF), LNT and SCR, plus temperature and NO_x sensors (De Cesare et al., 2009).

Figure 2 shows a typical SCR system scheme for an EURO 6 Common rail Diesel engine. Usually, two NO_x sensors are used together with the catalyst: the first one, located upstream of the SCR device, detects NO_x concentration and is used to dose the right amount of urea needed for the chemical reaction. This measurement can be replaced by a prediction of NO_x performed via static maps, as proposed in several automotive applications, or via NO_x virtual sensors as proposed in the present work. The second NO_x sensor is placed downstream of the catalyst. It provides the measurement of the effective amount of nitrogen oxides released to the atmosphere, in order to perform a control on the reduction reaction.

The downstream sensor will even be used to update the parameters of the virtual sensor model. In fact, when the urea injection is disabled, no chemical reactions take place in the SCR and the NO_x concentration measured downstream of the SCR correspond to the engine NO_x emissions. These data will be used to modify and adapt the virtual sensor's parameters, in order to predict the current NO_x concentration, including the effect of engine aging and production spread, ambient condition variations and possible drift of sensors and/or actuators.

3. EXPERIMENTAL SETUP

The experimental data used for the identification and the validation of RNN virtual sensor were measured on a Fiat 1.3 litre, 4 cylinders, EURO 5 Diesel engine, equipped with Common-rail injection system, high pressure EGR and variable geometry turbocharger (VGT), whose main technical data are described in Table 1.

Table 1. Engine technical data

Cylinders	4 in line
Displaced volume	1248 cc
Valves per cylinder	4
Max. Power	70 kW @ 4000 rpm
Max. Torque	210 Nm @ 1750 rpm
Fuel Injection System	Common Rail Solenoid Injectors
Compression Ratio	16.8:1

The experimental data were collected both in steady state and transient conditions. The tests were carried out on a dynamic test bench, with an AVL-APA ELIN Motoring AC Engine Dynamometer controlled by an AVL PUMA system (Figure 3), and on vehicle test rig. In both cases, the engine was equipped with an air flow meter upstream of the compressor, pressure and temperature sensors in the intake manifold and a linear lambda sensor (UHEGO) downstream of the turbine. The locations of the standard engine sensors, whose measurements are used as model inputs, are shown in Figure 4.

In addition, a production vehicle grade heated electro-ceramic sensor (Continental-SNS15 based on the amperometric principle) was used to measure the nitrogen oxides concentration. The sensor was located downstream of the turbine in order to ensure a response time compatible with the dynamics of the emissions phenomenon and to avoid an irregular distribution of the exhaust gas flow (De Cesare et al., 2011). The features of the Continental NO_x sensor were assessed in a former paper (De Cesare et al., 2011) by comparing its dynamic performance with those of a laboratory grade Horiba MEXA-720 fast response sensor. The analysis evidenced that locating the sensor downstream of the turbine results in comparable measured signals, since the gas flow dynamics in the exhaust manifold is comparable to the time constant of the Continental sensor.

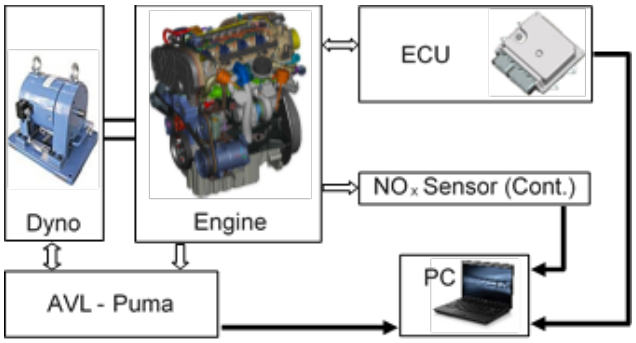

Fig. 3. Layout of the experimental plant on the engine test bench.

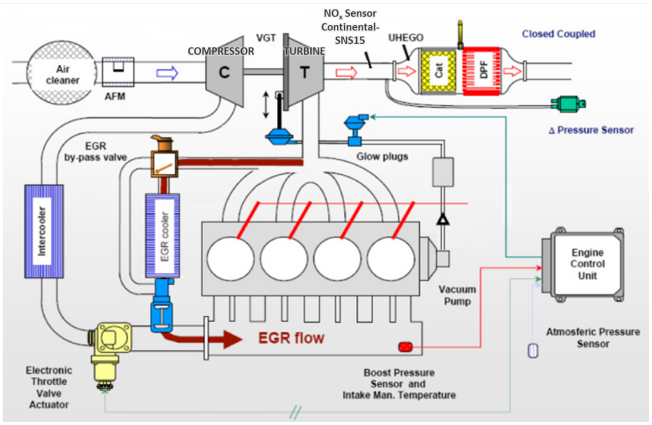

Fig. 4. Engine layout.

Several experimental data sets, corresponding to different driving transients, were considered to select a training data set for RNN model identification and seven different test data sets for RNN model validation. The training transient was designed in such a way as to cover the most significant engine working conditions, in order to guarantee the model generalization to reproduce any transient. It is the result of a process of combining an initial transient maneuver and a part of the New European Driving Cycle (NEDC) (610 ÷ 770 s) corresponding to the urban driving cycle (Arsie et al., 2013a).

The test transients were designed in such a way as to reproduce generic driving conditions and to cover a wide area of engine operation. In addition, the NEDC was also assumed as test transient for virtual sensor validation. It is worth noting that the first 500 seconds of the NEDC refer to engine

cold operation, before reaching steady thermal state. Since the SCR light-off temperature is about 200 °C, when the engine operates in cold conditions (e.g. during warm-up), the urea injection is disabled and the NOx estimation is not of interest for SCR management. Consequently, the experimental and simulation analyses were performed only in warmed-up engine operation and the NEDC data corresponding to engine cold operation were neglected. The resulting transient will be referred as HOT NEDC in the following (Arsie et al., 2013a).

3.1 Delay removal from measured data

In a turbocharged Diesel engine, several measurement and physical delays can occur and should be taken into account. In this case, the different location of the various sensors used during bench tests (see Figure 5) leads to a time delay between model inputs and outputs that represents a critical issue during RNN training (Arsie et al., 2013 a).

Considering the most significant variables, the following delays are expected:

1. Delay between pressure/air mass and lambda sensor: $\Delta t_{m_{air}\,2\lambda} = \Delta t_{p_{intake}\,2\lambda}$

2. Delay between fuel mass/start of injection and lambda sensor: $\Delta t_{m_f\,2\lambda} = \Delta t_{SOI\,2\lambda}$

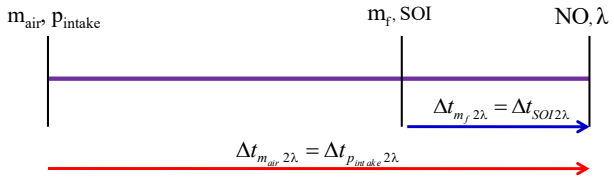

Fig. 5. Qualitative representation of the main input variables delays.

In order to have comparable data as inputs to the RNN, it becomes necessary to remove these delays. Therefore two curve fits were developed from the time delays measured during the experiments, to model their dependence on engine speed

$$\Delta t_{m_{air}\,2\lambda} = a \cdot rpm^{-b} - c \qquad (2)$$

$$\Delta t_{m_f\,2\lambda} = \mathrm{d} \cdot \mathrm{e}^{-\mathrm{f}\cdot\mathrm{rpm}} - \mathrm{g} \qquad (3)$$

where *a*, *b*, *c*, *d*, *f* and *g* are model parameters to be identified.

The time delay models are used to back-shift (anticipate in time) the experimental trajectories of the lambda and NO_x signals by the amount of time determined by eq. (2). In addition, the difference between eq. (2) and eq. (3) was used to synchronize m_f and SOI with the other input variables. It is worth mentioning here that the NO_x signal was also anticipated in time to avoid including the NO_x measuring delay in the training set. This simplifies the learning process by removing such delay from the dynamic process to be learned by the RNN model (Arsie et al., 2013a).

4. RNN WITH O_2 DESIGN

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, 1998). For the current application a nonlinear dynamic output error model (NOE) based on an external RNN was selected (Nørgaard et al., 2000; Arsie et al., 2010). This model structure exhibits a proper trade-off between prediction accuracy and computational/memory burden. Compared to lower dimensional look-up table, the RNN reduces the memory requirement due to the lower number of parameters to be stored in the EMS. A 2D look-up table with equal number of parameters without further correction factors, would not present comparable accuracy and, most importantly, would not be able to account for transient behavior and emissions dynamics. Furthermore, the experimental effort requested for the RNN training is significantly smaller compared to look-up table identification and this is a very important issue for control system design. In fact, one engine transient can be assumed as RNN training data set against the hundreds of steady state engine operating conditions needed to set-up a 2D look-up table. On the other hand, the computational burden of RNN would be higher than the simple interpolation of a 2D lookup table, taking into account the normalization, calculation, functions interpolation and de-normalization processes. Nevertheless, even so, the simulation time is significantly faster than the real-time and the increased computational burden is still feasible with current EMS, based on multi-core micro-processors. In conclusion, the application of a NOE based on RNN allows *i*) modelling the non linear formation mechanism, *ii*) enhancing dynamic features to improve model accuracy during engine transient operation and *iii*) reducing the experimental burden requested for model identification.

The reduction of experimental burden is particularly important in case of on-board virtual sensors update. The parameter adaptation can be applied to the output of any model. However, the combination of adaptation and RNN is expected to be more accurate and to converge in a reduced number of iterations. This allows reducing the time window of the adaptation process and, in turn, the SCR deactivation with reduced impact on tailpipe emissions.

The NO_x virtual sensor includes within itself a real-time intake O_2 dynamic model based on a mean value approach (Arsie et al., 2013b). The model was developed and experimentally validated on a similar engine to the one considered in the present work (Tab. I). The simulation results for the NEDC demonstrated the physical meaning and reliability of the estimated O_2 concentration in the intake manifold. Thus, in the current work the oxygen concentration

values predicted by the intake O_2 estimator are fed as input to the RNN model.

Figure 6 shows an exemplary schematic of an NOE model structure, obtained in case the current output $y(t)$ is evaluated as a function of j past inputs and i output feedbacks values that are fed back as additional network inputs. As previously described, engine speed (n), air mass flow ($\dot{m}_a$), fuel mass (m_f), Start of Injection (SOI) and intake oxygen concentration (O_2) are considered as input variables for the RNN model, while the output variable corresponds to the NO_x concentration measured by the Continental sensor. The resulting RNN model is then expressed by the following formulation

$$\begin{aligned} N\hat{O}_x(t,\theta) = F\big[&N\hat{O}_x(t-1,\theta),...,N\hat{O}_x(t-i,\theta),\\ &n(t),..,n(t-j+1),\dot{m}_a(t),..,\dot{m}_a(t-j+1),\\ &m_f(t),...,m_f(t-j+1),SOI(t),...,\\ &SOI(t-j+1),O_2(t),...,O_2(t-j+1)\big] \end{aligned} \quad (4)$$

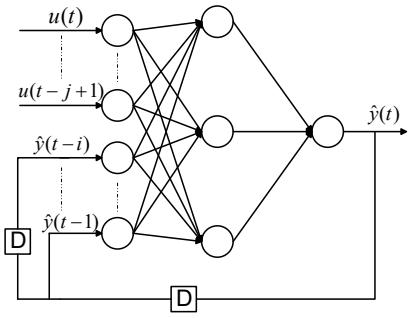

Fig. 6. General Schematization of an NOE RNN with one hidden layer, input lag space = i and output lag space = j.

The methodology applied for the RNN training consisted of the following steps:

- Parametric analysis and selection of the most suitable neural network structure;
- Deterministic analysis to set RNN weights initial conditions.

The parametric analysis was aimed at selecting the RNN structure that better meets the conflicting needs of accurate prediction and reduced computational time. This latter is one of the most critical requirements for the on-board implementation, as it has to be faster than the real-time. The number of past inputs and past outputs was varied between 1 and 3, whereas hidden neurons were varied between 2 and 18. The maximum number of training epochs was varied in the range [60; 500], with a step of 20.

Upon complete execution of the parametric analysis, several recurrent neural network configurations were trained. Nevertheless only one of them was selected according to the best results achieved. To this purpose, two indexes were considered: the root mean squared error (RMSE) and the relative integral error index (IE) defined as

$$IE = \frac{\left|\int_t^{t_{end}} N\hat{O}_x(t)dt - \int_t^{t_{end}} NO_x(t)dt\right|}{\int_t^{t_{end}} NO_x(t)dt} \quad (5)$$

The initial conditions can affect the RNN performance and the high non-linearity may lead to multiple solutions. Therefore a deterministic analysis was carried out by training the RNN with 1024 sets of initial conditions. Moreover, the 'Early stopping' method (Haykin, 1998), based on the RMSE minimization on the NEDC cycle, was applied as a training termination criterion.
Afterwards, by analyzing the frequency distributions of RMSE and IE, the RNN with the greatest densification around the minimum error was selected (i.e. H=12, i=2, j=1). Table 2 lists the performance of the selected RNN for two different test transients. For further details on the training and test transients used in the present work, the reader is addressed to a previous authors' contribution (Arsie et al., 2013a).

Table 2. Results of the initial conditions analysis for the selected RNN on HOT NEDC and Test transient 1

	H	i	j	Iter	RMSE [ppm]	IE
HOT NEDC	12	2	1	40	39.06	0.086
TEST1				40	101.89	0.099

5. ADAPTIVITY

An adaptive algorithm has to be conceived in such a way so as to change its behavior based on information available at the current time. This might be information from computational resources available or the history of data recently measured.
Many algorithms are adaptive or have adaptive variants, which usually mean that the algorithm parameters are automatically adjusted according to statistics about the optimization thus far.

Among these, LS is the most widely-used adaptive algorithm due to its computational simplicity and robust adaptation properties (Ljung, 1999). Considerable efforts have been made in developing even more efficient algorithms, like the recursive least squares (Guo et al., 1993). As compared to the least squares algorithm, the RLS offers a superior convergence rate, especially for highly correlated input signals. The LS algorithm can be seen as an approximation of the RLS algorithm, because it would represent a weighted average of the experimental data for the whole investigated time period (Paleologu et al., 2008).

For the current application, the LS method was selected. The motivation behind such a choice lies in the unavailability of the updating data set for the entire engine working time, e.g. during the whole driving cycle. Indeed, the specific application dealt with in this paper entails stopping SCR

functionality, as already explained at the end of section 2, so that suitable transient data can be acquired at the engine exhaust. This will ensure that non-treated NO_x emissions become available for subsequent LS-based adaptation of virtual sensor parameters, while avoiding the release of excessive direct NO_x engine tailpipe emissions.

If RLS is used during the above-mentioned short SCR-disabled time-window, only the last values of LS parameters (see Eq. 7) (i.e. estimated by RLS at the last time instant of the adaption task) would be available, as a consequence of the abrupt interruption of RNN adaptation. Since the NO_x dynamics reproduced by the RNN is very fast, and also due to the increased modeling uncertainties characterizing transients triggered by abrupt engine maneuvers, in many cases the last available LS parameters vector could result in unreliable adaptation during subsequent virtual sensing tasks. On the other hand, an average estimation of LS parameters through simple LS was proven to be much more robust and effective against the above-mentioned issues.
By referring to the most prominent contribution available in the literature on LS-based model adaptation (Ljung (1999)), the following least square estimate was adopted to perform online adaptation of the generic vector of model parameters (i.e. $\hat{\vartheta}$):

$$\hat{\vartheta} = \arg\min(V(\vartheta, N)) = \arg\min(\frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}(y(t) - \varphi^T$$
$$= \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) \quad (6)$$

where V(ϑ,N) is the criterion function measuring model validity, whereas N, $\varphi(t)$ and y(t) represent the adaptation discrete time horizon, the vector of x_i inputs and the available measured trajectory of the output variable, respectively.

5.1 Application to RNN NO_x virtual sensor

The work is focused on the problem of real-time NO_x estimation. The approach consists of modeling NO_x emissions by using an adaptive model based on an NOE recurrent neural network, as described in the previous section. The model parameters are adjusted online by the LS updating method while the engine is running but the SCR is switched off. During this phase, the parameters evolve, correcting RNN estimation drift. Afterwards, as soon as the SCR switches back on, the final stored parameters are kept constant to predict next NO_x values.

Least squares adaptation is applied to NOE output. In principle, least squares algorithm could update all network parameters, i.e. weights and biases, but this could lead to unacceptable computational burden, especially in view of on-board application. Indeed, the NOE network structure consists of 157 parameters, so in terms of computational burden it could be too onerous. Thus, in this study, a simplification was made to reduce the adaptation task complexity. Particularly, the following coefficients were introduced to enable online adaptation of NOE outputs, namely a gain k_0 and an offset k_1, thus resulting in the following LS parameters vector ϑ:

$$\vartheta = [k_0, k_1] \quad (7)$$

$$\hat{NO}_x = k_0 \cdot NO_x + k_1 \quad (8)$$

where NO_x is the output provided by the neural network (thus representing the only variable x_i the input vector φ consists of (see eq. (6)) and $\hat{\hat{NO}}_x$ is the NO_x prediction after LS-based adaptation.

The scheme represented in Figure 7 shows the recurrent neural network with least squares update, which results into an adaptive NO_x virtual sensor. The input signals are processed by the RNN that estimates the NO_x concentration in the engine exhaust emissions. In order to enable LS output adaptation, it is necessary to compare this estimation to NO_x sensor measurement at the same time instants. The observed error is used to evaluate the LS criterion (see eq. 6), which in turn enables the parameters vector update. Finally $\hat{\hat{NO}}_x$ will be the updated prediction of the nitrogen oxides concentration.

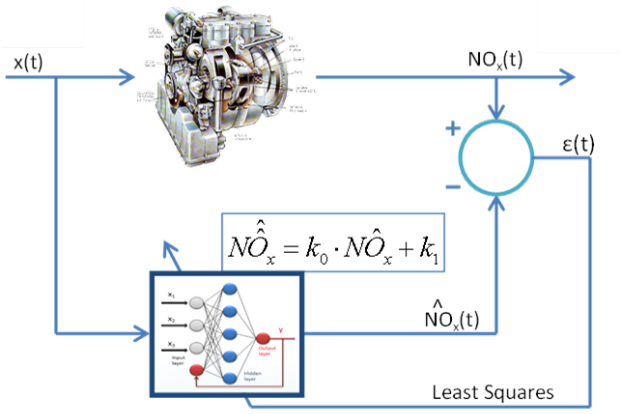

Fig. 7. Scheme of the RNN adaptive virtual sensor, obtained by suitably exploiting NO_x sensor measurement during the LS adaptation phase (i.e. SCR temporarily disabled).

6. RESULTS

The NO_x adaptive virtual sensor was tested via comparison with real experimental data, which were suitably treated in order to enable full assessment of the proposed LS-based algorithm adaptation capabilities.

As mentioned in the introduction section, on-line model adaptation is needed to deal with several real-life behaviors that may occur, such as engine aging, production variation, change in driving style, move from urban to extra-urban

route, failure of some engine component, etc. Nevertheless, experimental data on engine test-bench are generally measured in normal operation of the system composed by engine, SCR, sensors, etc. Therefore, in order to simulate some of the above mentioned realistic cases, the experimental measurements were modified, in such a way as to emulate either a sensor malfunctioning or unexpected changes in input variables values.

The first modification consists in a drift applied to the NO_x sensor signal. This results in an RNN estimation error with respect to the new (i.e. drifted signal) experimental NO_x emissions. Therefore, the least squares method may be the solution to recover this difference between the simulated and measured values. It is worth mentioning here that sensor drift may also be interpreted as an effective NO_x decrease due to either realistic faults or engine operations modifications.

The test procedure follows three steps, which are intended to reproduce a realistic situation:

a. Fault/failure;

b. update;

c. validation.

The fault simulation was performed by imposing a drift equal to -30% on the NO_x sensor output provided throughout the available experimental transients. Such a drift may be considered as a residual between the expected engine NO_x emissions, from the given set of model input variables, and sensor output. In this sense the residual may be the indication of a fault in any sensor/actuator involved in the model input/output path. As an example, if one of the injectors fails to provide the theoretical amount of fuel defined by the ECU, the measured engine NO_x emissions (i.e. sensor output) will be lower than those estimated by the model. Due to the exponential relationship between NO_x kinetics and in-cylinder temperature, a fault in the fuel injection process can easily result in a NO_x residual of 30 %. This analysis could be extended to other model input variables, such as oxygen concentration and air flow rate, whose wrong detection, due to flow meter fault or air leakage in the intake manifold, would result in a similar residual between current and expected NO_x emissions.

In the adaptation time frame, the LS algorithm updates the RNN estimation, as described in section 5.1. The effectiveness of the proposed adaptation methodology was first verified by selecting the training data-set (Arsie et al., 2013a) as parameters update data set, i.e. the one adopted to apply the LS-based adaptation methodology. For the sake of space, only the improvements results achieved by the LS adapted RNN virtual sensor (hereinafter referred to as Least squares in all figures) on the HOT NEDC are illustrated in the following.

Figure 8 shows the prediction accuracy of the non-adapted RNN along the HOT NEDC transient, via comparison between measured (i.e. with -30% drift) and predicted NO_x emissions. For this transient the RMSE and integral error index IE equal 66 ppm and 55.28 %, respectively, as indicated in Table 3.

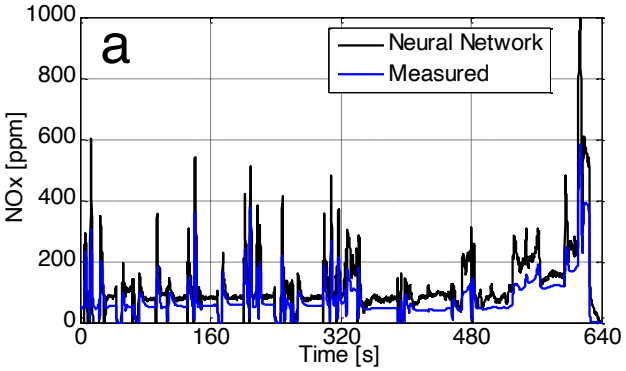

Fig. 8. Comparison between NO_x emissions measured (i.e. with -30 % drift) and estimated with non-adapted RNN along the HOT NEDC, during the failure phase (a).

The comparison between measured and LS adapted RNN simulation is shown in the lower graph of Figure 9, for the parameters update data set (i.e. the same as the original RNN training data set). The higher graph illustrates the evolution of the parameters over the entire updating phase. The final parameters values, used for the subsequent LS validation phase, are k_0 = 0.701 and k_1 = 3.282, respectively. As expected, the LS modified parameters evolve with time, in such a way as to suitably account for sensor drift with respect to normal operation. This behavior in turn results in improved NO_x prediction accuracy, as evidenced in Table 3 and when comparing Figure 10 and Figure 8. Particularly, Figure 10 shows that $N\hat{O}_x$ values (i.e. red-dashed line) are always much closer to the measured trajectory as compared to $N\hat{O}_x$ (i.e. continuous black line), with a prediction error almost always below 15 % that corresponds to a satisfactory accuracy for the intended application.

It is worth noting in Figure 9 that the adaption process, which takes place when the SCR is disabled, almost converges in about 160 s. This time duration corresponds to approximately 15 % of the duration of the NEDC. Considering the relatively short time that the SCR is disabled and the high NO_x reduction efficiency achievable by a suitably managed SCR, it is expected that the LS adaption process will not have any trouble to comply with emissions regulations. It is also worth remarking that LS, when compared to RLS, optimizes parameters taking in consideration all the data available in the "update time interval", thus resulting in a more accurate prediction in the validation phase. This happens since the values of adapted parameters converge as time goes by. On the other hand, RLS will deliver, at the end of the adaptation phase, the last updated parameters values, which may be affected by unpredictable issues, such as: unexpected loss of steady-state accuracy of the virtual sensor; occurrence of

occasional/instantaneous measurement error and missing experimental data.

The above-commented results indicate that if the vehicle is going to be driven along a route consisting of:

a) one HOT-NEDC cycle with non-adapted RNN virtual sensor (see Figure 8),

b) one training-set cycle, during which LS adaptation of RNN is enabled (and SCR disabled),

c) and one HOT-NEDC cycle (i.e. LS validation phase) with via LS adapted RNN virtual sensor,

the proposed methodology ensures reducing RMSE error from 66 [ppm] down to 26 [ppm], which corresponds to an improvement of accuracy of 60 % passing from phase a) to phase c) through the updating phase b).

It is worth noting that the computational time to perform in off-line the test procedure composed by the three steps a), b) and c) was about 40 s, on an Intel Core i7 2.30 GHz, against an overall transient duration of 1760 s, resulting in a computational time that is almost 2 orders of magnitude faster than real time. These results, compared to the features of current EMS, that are based on multi-core micro-processors with large memory capacity, prove that the additional computational burden and memory usage related to the RNN and the batch-wise adaptation process are not expected to be so troublesome for the on-board implementation.

Figure 11 focuses on the comparison between the simulated and measured NO_x concentration in a time window corresponding to the Extra Urban Driving Cycle (EUDC), where a greater improvement occurs.

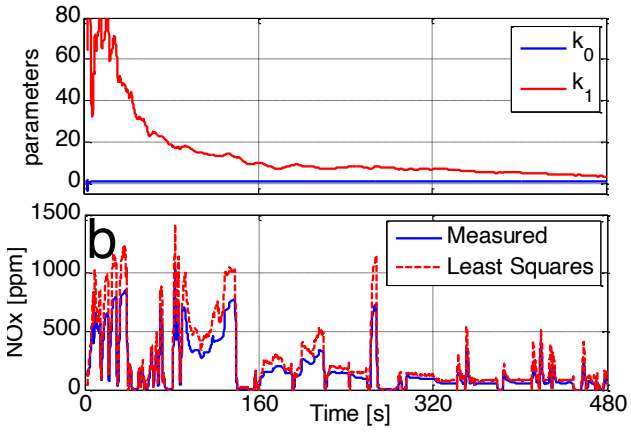

Fig. 9. Variability of the parameters k_0 and k_1 (above) and comparison between NO_x emissions measured and predicted by least squares (below) along the Training data set, during the adaptation phase (b).

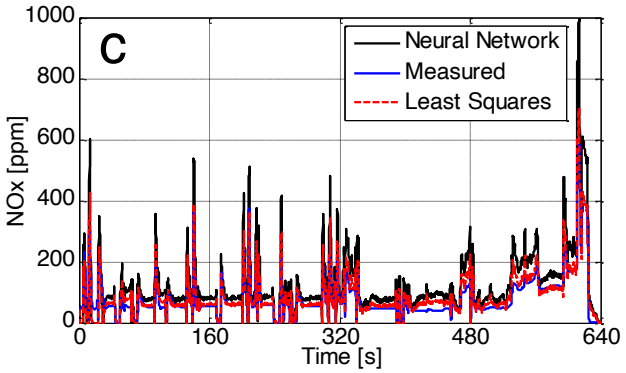

Fig. 10. Comparison between experimental drifted NO_x concentration and RNN and least squares predicted trajectories along the HOT NEDC, during the LS validation phase (c).

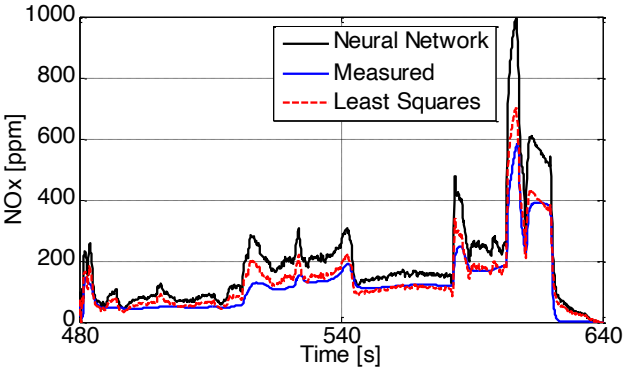

Fig. 11. Comparison between experimental drifted NO_x concentration and RNN and least squares predicted trajectories along the HOT NEDC, focusing on a tip-in transient in the time window 480 ÷ 640 [s] (c, close window).

Table 3. RNN and LS model prediction errors on the NO_x emissions for HOT NEDC transient.

	RMSE [ppm]	IE [%]
RNN	66.000	55.277
RNN with LS adaptation	26.069	12.796

The above described LS verification procedure demonstrates the value of the LS adaptation. The failure (i.e. a) and validation (i.e. c) phases were performed by referring to the same transient profile in order to highlight the benefits achievable from the parameters update.

Different performance of RNN output adaptation can be verified by varying the transient data sets to be adopted for the updating (i.e. b) and validation (i.e. c) phases, respectively. For this reason, a comprehensive analysis of all available test transient combinations was carried out. Thus 56 simulations were performed, resulting from the combination of the seven transient profiles introduced in Arsie et al. (2013a) for the two phases considered. Table 4 summarizes the results yielded by the simulation analyses. The

advantages obtainable with the Least Squares method can be assessed via eq. (9), through which the percent error variation detailed in Table 4 is obtained. On the right hand side of eq. (9) the pedices LS and RNN stand for via LS adapted and non adapted RNN virtual sensor. It is also worth noting that positive and negative values of $RMSE_{comparison}$ refer to worsened and improved RNN accuracy.

$$RMSE_{Comparison} = \frac{RMSE_{LS} - RMSE_{RNN}}{RMSE_{RNN}} \quad (9)$$

Table 4. $RMSE_{comparison}$ values yielded on output by the verification procedure, which was specifically set-up to assess LS adaptation effectiveness depending on updating and validation data set selection.

RMSE COMPARISON		VALIDATION DATA SET								
		Train	Test 2	Test 3	Test 1	NEDC	Test 5	Test 6	Test 7	Average
PARAMETERS UPDATE DATA SET	Train	-0,752	-0,691	-0,464	-0,617	-0,605	-0,626	-0,494	-0,561	-0,601
	Test 2	-0,299	-0,577	-0,209	-0,521	0,389	-0,289	-0,409	-0,430	-0,293
	Test 3	-0,472	-0,604	-0,298	-0,626	-0,033	-0,476	-0,523	-0,542	-0,447
	Test 1	-0,543	-0,534	-0,303	-0,722	-0,600	-0,702	-0,591	-0,660	-0,582
	NEDC	-0,539	-0,522	-0,305	-0,720	-0,678	-0,729	-0,581	-0,668	-0,593
	Test 5	-0,556	-0,514	-0,319	-0,718	-0,679	-0,727	-0,576	-0,663	-0,594
	Test 6	-0,213	-0,274	-0,050	-0,604	-0,176	-0,514	-0,562	-0,577	-0,371
	Test 7	-0,251	-0,302	-0,081	-0,627	-0,253	-0,557	-0,581	-0,601	-0,406

The application of the least squares algorithm to RNN adaptation enhances virtual sensor accuracy in almost all cases, regardless of the chosen profile for parameters updating, thus confirming the appropriateness of selecting the linear model (see Eq. 8) for RNN adaptation. The last column of Table 4 is the average of the $RMSE_{comparison}$ for all transient validation profiles; it particularly highlights that the Train profile guarantees the best results if it is used as parameters update data set. Moreover, it appears how the selection of the parameters update data set mostly influences the extent, up to which the drift can be recovered via the proposed adaptation technique. Overall, the parametric analyses detailed in Table 4 confirm that adaptation provides improvement in most cases independently from driving habits. According to the RMSE comparison between pre and post adaptation RNN, the most accurate simulation is achieved when HOT NEDC cycle is used for sensor adaptation and the Test profile 5 is the validation data set. The worst case occurs in correspondence of the weakest parameters update data-set (i.e. Test 2), thus suggesting that, eventually, a pattern recognition task could be profitably combined with the proposed LS-based adaptation method to improve prediction accuracy.

As reported in Table 4, when the HOT NEDC is selected as LS validation data set, the best result is obtained with the Test 5 as updating profile. Figure 12 and Table 5 show simulated and experimental trajectories and error values achieved by non-adapted and adapted RNN virtual sensor, respectively. As expected from the above commented parametric analysis on $RMSE_{comparison}$ (see eq. 9), Test profile 5 can rightly be considered as the best parameters update data set for the NEDC profile, as it allows considerably reducing the estimation errors after LS-based adaptation (see error improvement going from Table 3 to Table 5). Indeed, the maximum improvement achievable is 67.9 % and 99.4 % for RMSE and IE, respectively. This clear and excellent result is even more evident in the graph shown in Figure 13, where the NO_x trajectory predicted by the LS adapted RNN almost completely overlaps the experimental drifted one.

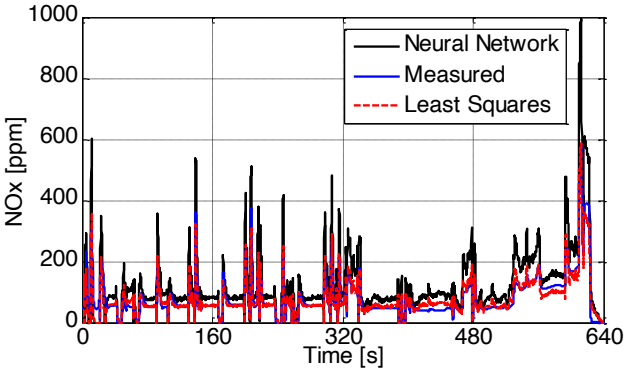

Fig. 12. Comparison between experimental drifted NO_x concentration and RNN and least squares predicted trajectories along the HOT-NEDC, with the Test profile 5 as parameters update data set.

Table 5. RNN and LS model prediction errors on HOT NEDC transient, by using the Test profile 5 for parameters update.

	RMSE [ppm]	IE [%]
RNN	65.993	55.370
LS	21.187	0.347

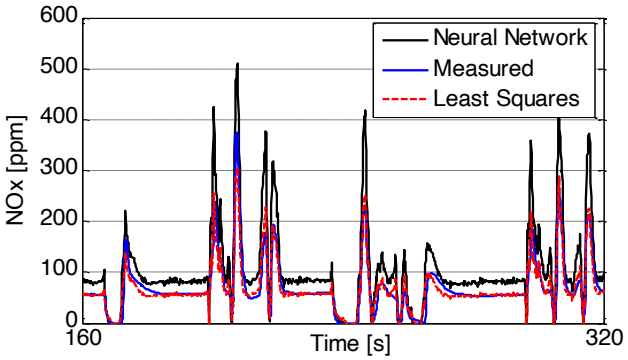

Fig. 13. Comparison between experimental drifted NO_x concentration and RNN and least squares predicted trajectories along the HOT-NEDC, focusing on a tip-in transient in the time window 160 ÷ 320 [s], with the Test profile 5 as parameters update data set.

Figure 14 focuses on the best case among all simulations. It shows the comparison between the RNN and LS prediction profiles and the measured drifted NO_x concentration along the Test profile 5 after the updating phase carried out on the HOT NEDC. The achieved improvements are 72.9 % and 99.4 % for RMSE and IE, respectively. This transient profile presents many 'aggressive' maneuvers and most real working conditions, thereby strengthening the validity of the procedure.

The important role of the updating test-set is denoted in the Table 4. The engine operation domains investigated along the

stitched training data set (see Section 3), HOT NEDC, Test 5 profile and Test 2 profile are shown in Figure 15. The diagram highlights the engine working domain, thus providing useful information to assess the correct selection of both training and parameters update data sets. Particularly, Test Profile 2 exhibits the lower coverage extent as compared to the other driving cycles, thus providing an exhaustive physical explanation to the worst value shown in Table 4.

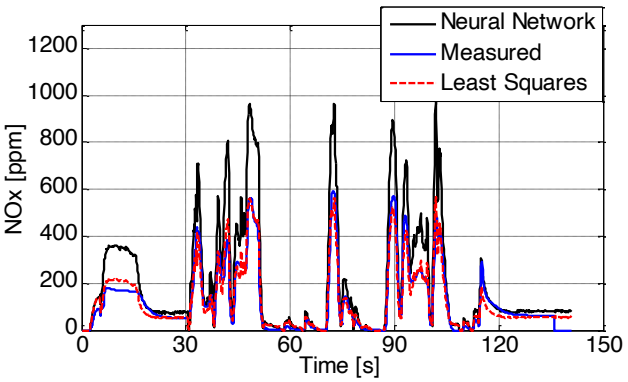

Fig. 14. Comparison between experimental drifted NO_x concentration and RNN and least squares predicted trajectories along the Test profile 5, with the HOT-NEDC as parameters update data set.

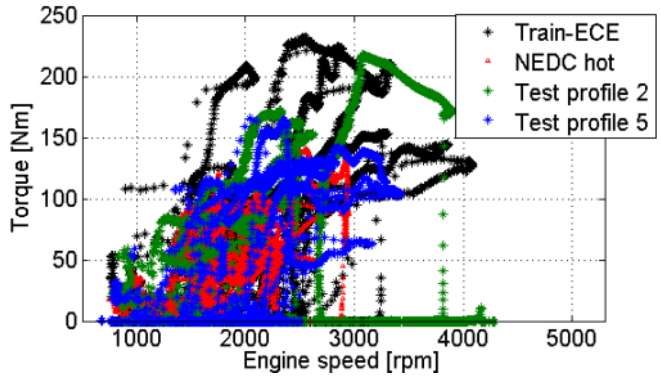

Fig. 15. Engine operation domains investigated along Stitched Training data set (black marker), HOT NEDC (red marker), Test profile 2 (green marker) and Test profile 5 (blue marker).

Figure 16 shows the frequency distributions of both Integral error (above) and RMSE (below) indexes, as yielded on output by both via LS adapted and non-adapted RNN virtual sensor when simulating all transients available. Each bar represents the sum of the times (occurrences) that the relative error index occurs as result of the various simulations along the NEDC. The LS adaptation allows considerably improving the prediction of NO_x concentration. The achieved accuracy is in line with the requirements for on-board application of a virtual sensor.

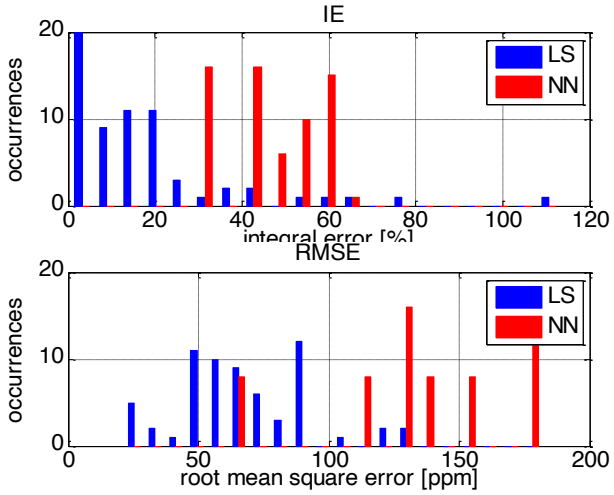

Fig. 16. Frequency distributions of IE (above) and RMSE (below) indexes for transient NO_x trajectories simulated with both LS adapted and non-adapted RNN.

7. CONCLUSIONS

In this paper, least square technique features were exploited to adapt the output of a Recurrent Neural Network-based NO_x virtual sensor, in such a way as to account for the effects due to engine aging and production spread as well as components drift. Appropriate justification to the selection of batch LS in place of continuous recursive least square adaptation was provided.

The LS adapted RNN virtual sensor was extensively tested in a simulation environment. The results of such an analysis confirmed the efficacy of the selected adaptation methodology for the real world application on-board of automotive Diesel engines. The statistical analysis, specifically carried-out on estimation error related indexes, indicated that both timing and duration of batch LS adaptation should be carefully selected and scheduled, especially considering that the selective catalytic reduction device must be disabled during this task.

The results of the off-line analysis also evidenced that the additional computational burden and memory usage related to RNN estimation and adaptation process are expected to be compatible with current EMS features. Furthermore, the adaptation process takes a relatively short time to converge. Consequently, a minor impact is expected on the amount of NOx emitted along the driving cycles to be referred to when complying with emissions legislation.

Future work will focus first of all on testing the proposed algorithm on-board. Afterwards, techniques that are suitable to solve some practical issues (e.g. SCR disabling time) will have to be investigated, including the adoption of NLMS for online adaptation.

ACKNOWLEDGEMENTS

This work was funded by Magneti Marelli Powertrain and University of Salerno.

REFERENCES

Alberer, D., del Re, L., Winkler, S., Langthaler, P. (2005). Virtual Sensor Design of Particulate and Nitric Oxide Emissions in a DI Diesel Engine. SAE Technical Paper 2005-24-063.

Arsie, I., Pianese, C., and Sorrentino, M. (2010). Development of recurrent neural networks for virtual sensing of NO_x emissions in internal combustion engines. *SAE Int. J. of Fuels and Lubricants,* 2 (2): 354-361.

Arsie, I., Cricchio, A., De Cesare, M., Pianese, C., Sorrentino, M. (2013a). A Methodology to Enhance Design and On-Board Application of Neural Network Models for Virtual Sensing of NO_x Emissions in Automotive Diesel Engines. SAE Technical Paper 2013-24-0138.

Arsie, I., Cricchio, A., Pianese, C., and De Cesare, M. (2013b). Real-Time Estimation of Intake O_2 Concentration in Turbocharged Common-Rail Diesel Engines. *SAE Int. J. of Engines* 6 (1): 237-245.

Atkinson, C.M., Long, T. W., Hanzevack, E.L. (1998). Virtual Sensing: A Neural Network based Intelligent Performance and Emissions Prediction System for On-Board Diagnostics and Engine Control. SAE Technical Paper 980516.

Blanco Rodriguez, D. (2014). *Modelling and observation of exhaust gas concentrations for diesel engine control*. *Chapter 2*. Springer.

Brand, D., Onder, C., Guzzella, L. (2007). Virtual NO Sensor for Spark-Ignition Engines". *Int. J. of Engine Research* (8): 221-240.

De Cesare, M., Covassin, F., Preziuso, M., Serra, G. (2009). A Mean Value Model of the Exhaust System with SCR for an Automotive Diesel Engine. SAE Technical Paper 2009-24-0131.

De Cesare, M., Covassin, F., (2011). Neural Network Based Models for Virtual NO_x Sensing of Compression Ignition Engines. SAE Technical Paper 2011-24-0157.

Del Re, L., Langthaler, P., Furtmueller, C., Winkler, S. and Affenzeller, M. (2005). NO_x virtual sensor based on structure identification and global optimization. SAE Technical Paper 2005-01-0050.

Guardiola, C., Pla, B., Blanco-Rodriguez, D., Eriksson, L. (2013). A computationally efficient Kalman filter based estimator for updating look-up tables applied to NO_x estimation in diesel engines. *Control Engineering Practice* 21 (11): 1455- 1468, doi: 10.1016/j.conengprac.2013.06.015.

Guo, L., Ljung, L., Priourei, P. (1993). Performance analysis of the forgetting factor RLS algorithm. *Int. J. of Adaptive Control and Signal Processing* 7: 525-537.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. *Chapter 4*. Prentice Hall.

Heywood, J.B. (1998). *Internal Combustion Engine Fundamentals. Chapter 11*. MC Graw Hill.

Isermann, R., Müller, N. (2001). Modelling and Adaptive Control of Combustion Engines with Fast Neural Networks. Proc. of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems, Tenerife, Spain.

Johri, R., Filipi, Z. (2014). Neuro-fuzzy model tree approach to virtual sensing of transient diesel soot and NO_x emissions. *Int. J. of Engine Research* 15 (8): 918-927.

Liu, G., Ouyang, M., Lu, L., Xu, L, Li, J. (2012). Online Monitoring of Lithium-ion Battery Aging Effects by Internal Resistance Estimation in Electric Vehicles. Proc. of 31st Chinese Control Conference.

Ljung, L. (1999). *System Identification – Theory for the user. Chapter 7*. Prentice-Hall.

Nørgaard, M., Ravn, O., Poulsen, N.L., Hansen, L.K. (2000). *Neural Networks for Modelling and Control of Dynamic Systems*. Springer-Verlag.

Paleologu, C., Benesty, J., Ciochina, S. (2008). A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification. *IEEE Signal Processing Letters* 15: 597-600.

Pavkovic, D., Deur, J., Jansz, M., Peric, N. (2006). Adaptive control of automotive electronic throttle. *Control Engineering Practice* 14: 121–136.

Ramos, J.I., 1989, *Internal Combustion Engine Modeling*. *Chapter 4*. Hemisphere Publishing Corporation.

Thenua, K.R., Agarwal, S.K. (2010). Simulation and performance analysis of adaptive filter in noise cancellation. *Int. J. of Engineering Science and Technology* 2 (9): 4373-4378.

Traver, M.L., Atkinson, C.M., Atkinson R.J. (1999). Neural Netwok-Based Diesel Engine Emissions Prediction Using In-Cylinder Combustion Pressure. SAE Technical Paper 1999-01-1532.

Tschanz, F., Amstutz, A., Onder, C.H. and Guzzella, L. (2010). A real-time soot model for emission control of a diesel engine. Proc. of the 6th IFAC Symposium on Advances in Automotive Control: 222-227.

Walker, A., Allansson, R., Blakeman, P., Lavenius, M. et al. (2003). The Development and Performance of the Compact SCRTrap System: A 4-Way Diesel Emission Control System. SAE Technical Paper 2003-01-0778.

Willems, F., Cloudt, R., van den Eijnden, E. et al. (2007). Is closed-loop SCR control required to meet future emission targets? SAE Technical Paper 2007-01-1574.