**Proceedings of the 17th World Congress**
**The International Federation of Automatic Control**
**Seoul, Korea, July 6-11, 2008**

# Grey-Box Control Oriented Emissions Models

**Markus Hirsch* Daniel Alberer** Luigi del Re****

**LCM- Linz Center of Mechatronics, Linz, Austria*
*(Tel: 0043-732-2468-1842; email: markus.hirsch@lcm.at)*
***Institute for Design and Control of Mechatronical Systems,*
*Johannes Kepler University, Linz, Austria (email: {daniel.alberer, luigi.delre}@jku.at)*

**Abstract:** Further improvements of emission control will require reliable estimation of emissions in real time. While many progresses are being done in terms of physical sensors, there is a wide agreement that virtual sensors and more in general real time emission models will play a central role in the next steps. While there is a deep understanding of the physics of the regulated pollutants, most general emission models tend to be too complex and poorly parametrized to be used on-line, while most data based models tend to be either insufficiently precise or of limited scope. To avoid this problem, this paper proposes a combined approach in which static maps are identified numerically, but the effect of dominant factors, like cylinder-head temperature and air path dynamics, is included on the basis of physical assumptions. Differently from most models developed for sensors, this approach is based on pure engine control unit (*ECU*) data, *i.e.* can be used for the computation of optimal control laws. As the paper shows, this strategy is able to provide not only real time estimation of $NO_x$ as a function of the *ECU* outputs, but also of particulate matter (*PM*).

## 1. INTRODUCTION

The high mobility requirements have well known negative side effects, in particular in terms of environmental damage and air pollution by exhaust gas, as carbon dioxide ($CO_2$), carbon monoxide (*CO*), residual hydrocarbons (*HC*), particulate matter (*PM*) and nitrogen oxides ($NO_x$, essentially a mixture of *NO* and $NO_2$). Accordingly, vehicles with combustion engines may be put on the market only after passing standardized emission tests and require special after-treatment systems. In the case of Diesel engines, *CO* and *HC* limits can be achieved rather easily with oxidation catalysts, while $NO_x$ as well as *PM* limits are much more demanding. Although these limits can be met by Diesel particulate filters (*DPF*) and selective catalytic reduction (*SCR*), a minimization of the engine row emissions is important to optimize or even avoid these exhaust aftertreatment devices.

The increasing number of possible control actions in the combustion process of modern Diesel engines allows a much better optimization of the combustion, but complicates the locating process for the optimal set points. The challenge here is finding models that represent this highly nonlinear process in a way suitable for optimization. Models of combustion processes have been developed for a long time. As shown in (Patterson *et al*. 1994; Li *et al*. 1995; Wang *et al*. 1999; Jung *et al*. 2001), these models usually define different interacting zones in terms of chemical compositions and operation conditions in the combustion chamber. These models are often included in finite-element programs and can be very useful in finding optimal chamber geometrics by numerical analysis of the combustion process. Though theoretical models are very helpful in this case, their ability to simulate exhaust emissions accurately is limited.

Therefore, for on-line use, usually identified models are used instead. The variety of these models is huge, as the numbers of publications in the last years show. In principle they can be divided in two different groups: on the one hand models (Ouenou-Gamo *et al*. 1998; Schilling *et al.* 2006) whose structure is defined by basic chemical and physical equations (see: Heywood 1988; Warnatz *et al*. 1999) and only parameters are adapted and identified respectively to represent the emissions of a specific engine. In general, these models can represent global behaviors well, but are often overstrained to simulate local, engine dependent, behaviors. Another issue for restriction at real engine applications can be the use of characteristic values like the heat release rate, often not available, as input variables available in the *ECU*.

On the other hand, so called black-box models assume no specific model structure. Due the highly nonlinear behavior of an engine and the huge interaction of the different inputs, standard linear identification algorithms like FIR, ARX or OE are even locally overstrained. For the identification of nonlinear systems many different possibilities exist (Ljung 1999, Nelles 2001), each of them having advantages and disadvantages. Very common algorithms to describe engine emissions are artificial neural networks (*ANN*) (Hafner *et al*. 2000; Desantes *et al*. 2002; Ayeb *et al*. 2005; Galindo *et al*. 2005). Models applying genetic algorithms (del Re *et al*. 2005; Alberer *et al*. 2005) are also used and brought good results for the standard operating mode (under closed loop action of the *ECU*).

This restriction – to closed loop operation under the action of the *ECU* – is critical, as it strongly reduces the dimension of the model space. Indeed, an *ECU* tends to produce fixed relationships between control variables (for instance, for a single speed, temperature and demanded torque, there is one single combination of several control variables, like rail pressure, pre-injection time and amount, main injection time and amount, etc. which have been optimized for the specific operation). While this generally increases the precision of the model derived under closed loop operation for the closed loop condition, it makes it mostly worthless for the general operation and therefore for optimization, for which a sufficiently general model is needed.

This work is concerned in deriving such an open loop model for a Diesel engine. As the engine cannot be operated without an *ECU*, and some combinations of control inputs could even lead to an immediate damage of the engine, a real open loop model identification is not possible, but the problem can be reduced using tools both at the measurement and at the interpretation level: the model is designed using a combination of a comprehensive steady state model (obtained modifying the set points of the controller so to cover the whole range of combinations of control inputs) and statistical tools are used to reduce the correlation present in data and thus improve the problem condition.

Additionally the model uses first-principle dynamic extensions to account for the effects of the temperature and of transport phenomena in the air path. The resulting model is able to represent with a sufficient precision both the $NO_x$ and *PM* values over the whole engine operating range using only values available in the ECU.

## 2. MODEL DESIGN

### *2.1 General issues*

The design of a control oriented emission model involves addressing several aspects, in particular the strong nonlinearity of the engine map, the strong correlation between many measurements and the essentially open problem of the right formula structure of an $NO_x$ and especially of a *PM* estimator. These problems are solved in our case in the following way:

(a) a data set "as rich as possible" under open loop conditions is determined experimentally

(b) variable selection methods are used to determine the input channels with the largest independent impact on emissions (resulting in table 1)

(c) the whole engine map is subdivided in local regions (as shown in fig. 1) and the outputs of the global model $y_{global}$ (emission values for $NO_x$ and *PM*) are computed switching between outputs of local models $y_{r,}$ using a linear combination in overlapping regions (1).

**Table 1. Input Variables**

| Input varible | Unit | Description |
|---|---|---|
| *qMI* | mg/Inj | fuel mass of main injection |
| *phiMI* | Deg | crank angle position of main injection before top dead centre |
| *qPI* | mg/Inj | fuel mass of pilot injection |
| *tPI* | µs | time of pilot injection before main injection |
| *pRail* | bar | fuel pressure in common rail |
| *N* | rpm | engine speed |
| *MAP* | mbar | manifold air pressure |
| *MAF* | kg/h | manifold fresh air mass flow |
| *Toil* | °C | engine oil temperature |

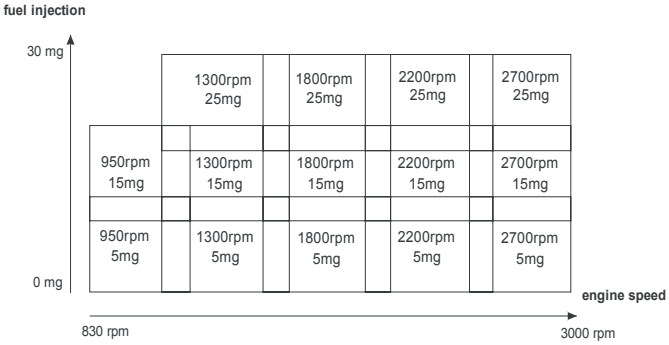


Fig. 1. Switching between local models defines the output of the global model

$$y_{glob} = \sum_{r=1}^{M} w_r(N, qMI) \cdot y_r$$
$$\sum_{r=1}^{M} w_r(N, qMI) = 1 \qquad (1)$$

(d) Compared to the uncertain and slow dynamics (delay time and low pass filter) of the emission measurement devices, the fast injection dynamics can be neglected. Dynamics of the air path - which are much slower than those of the injection system - could be excluded as the output values of this system manifold air flow (*MAF*) and boost pressure (*MAP*), both measured by the standard *ECU*, were taken as inputs for the emission model. As the air mass flow sensor is not mounted directly on the manifold but in the upstream of the turbo charger, a second order filter was included here to approximate the time lag (based on the results of (Vierlinger 2005)). The resulting transfer function from the measured value *MAF* to the real manifold air mass flow *MAF** can be written in the frequency domain as follows:

$$MAF^* = \frac{1}{T^2 s^2 + 2T\xi s + 1} MAF \qquad (2)$$

Here $T$ defines the time constant of the system $\xi$ the damping and $s$ the complex Laplace parameter.

*2.2 Local models*

The local models describing the global model are defined by in parameters linear regression functions. Eq. (3) shows the basic structure of these functions: the logarithmic engine output ($NO_x$ and opacity respectively) is defined by a mean value and a deviation depending on the standard deviation $\sqrt{\sigma_{uu}}$ multiplied by a sum of input functions $f_i(\Delta u)$ weighted by regressors $\theta_i$. The logarithm is used to achieve constant relative prediction errors over a larger output range (error homoscedasticity) and therefore accurate results for lower emissions levels too. Both the output and the input variables have been standardized by subtraction of the mean values ($\overline{y}$, $\overline{u}$) and division by the square root of their variance ($\sigma_{yy}$, $\sigma_{uu}$).

$$\ln(\hat{y}) = \overline{y} + \sqrt{\sigma_{yy}} \cdot \Delta y$$
$$\Delta y = \sum_{i=1}^{p} \theta_i \cdot f_i(\Delta u) \tag{3}$$
$$\Delta u = \frac{u - \overline{u}}{\sqrt{\sigma_{uu}}}$$

$\Delta u$ represents the vector of standardized input variables and $p$ the number of regressors. Several approaches for the generation of possible new input variables have been tested. In conclusion, a polynomial attempt of second order is sufficient and therefore (3) can be rewritten in matrix notation as follows:

$$\Delta y = \theta^T \cdot \Delta\tilde{u} \tag{4}$$
$$\Delta\tilde{u} = \left[\Delta u_1 \quad \cdots \quad \Delta u_p \quad \Delta u_1\Delta u_1 \quad \Delta u_1\Delta u_2 \quad \cdots \quad \Delta u_p\Delta u_p\right]^T$$

A criterion $J$ to assess the quality of a given estimated model is given by the sum of the square residuals:

$$J = \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \tag{5}$$

*2.3 Variable selections models*

The estimation problem with the cost function defined by eq. (5) can be solved by a standard least square method, i.e.

$$\hat{\theta} = \left(\phi^T\phi\right)^{-1}\phi^T Y \tag{6}$$

where $Y$ represents the data vector of the target variable and $\phi$ the data matrix of the input vectors (3), albeit without the convergence properties of the standard ARX due to the correlation of the inputs. This problem can be reduced either by well known regularization methods, or even better by choosing the most representative and therefore least correlated description basis. The procedure follows the basic ideas of orthogonal forward selection (Henning 2004), and essentially consists in looking for the sensitivity of the estimation result to each candidate regressor. Therefore, the normalized standard error of each regressor $SE(\hat{\theta}_i)$ was calculated according to (7). For robustness, these values had to be smaller then a defined level which was decreased iteratively until validation error on a not in the identification data included test data set raised significantly. Otherwise, the input channel was rejected and a new identification was done.

$$SE\left(\hat{\theta}_k\right) = \frac{\sqrt{\Sigma_{kk}}}{\hat{\theta}_k}$$
$$\Sigma = \hat{\sigma}^2\left[\phi^T\phi\right]^{-1} \tag{7}$$
$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2$$

*2.4 First principle extensions*

High combustion chamber wall temperatures have an increasing influence on $NO_x$ by the Zeldovich-formation (Warnatz *et al*. 1999) and a decreasing one on soot. These temperatures, which are normally not known, can be approximated by the engine oil temperature $T_{oil}$, also a variable of a standard *ECU*. Actually, this value is a dynamic output of the engine system and therefore not settable but describable only by quite complex definitions of the engine which are depending on much more parameters than those used here for the emission model. The very low excitation of $T_{oil}$ at the local identification areas causes huge standard errors *SE* in the regressors of the channels containing this value. Due to the selection criterion described before these channels were rejected and temperature dependence would disappear. There could be many approaches to tackle this problem, but a simple and efficient possibility to include effects of temperature proven to be multiplying the output of the nonlinear model by a scaling factor representing temperature differences between the main temperature of the identification data and current oil temperature. The weighting function for this temperature dependence was – on the basis of physical considerations (Warnatz *et al*. 1999) - an exponential function. The growth factor which is positive for $NO_x$ and negative for *PM*, is defined by separate warm-up tests under constant speed and load tests (Fig. 3).

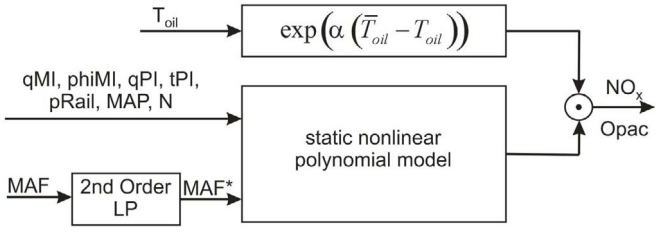


Fig. 2. Structure of the local models

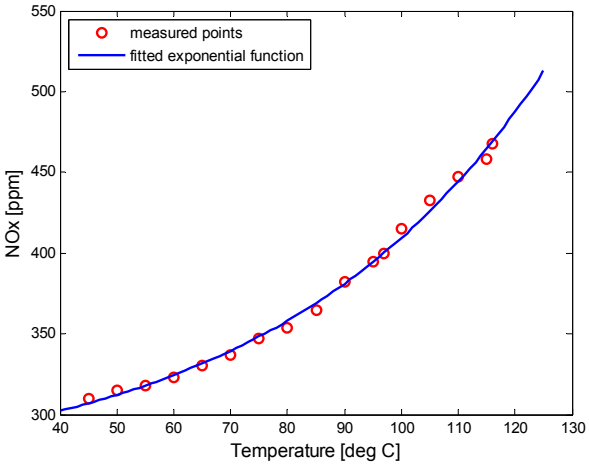


Fig. 3. Measured influence of oil temperature on emissions under constant speed and load.

## 4. EXPERIMENTAL FACILITIES AND IDENTIFICATION

The engine used to conduct the experiments was a production 2 litre 4 cylinder BMW Diesel engine on an AVL dynamical test bench. The engine parameters *qMI*, *phiMI*, *qPI, tPI*, *pRail*, positions of the exhaust gas recirculation valve (*EGR*), and that of the variable geometry turbocharger (*VGT*), defining *MAF* and *MAP*, have been specified directly and independently. Therefore the *ECU* was used to reach the operating point, but was partly out of authority during measurements. Also engine speed *N*, controlled by the torque of the test bench dynamometer, was changed separately and uncorrelated to the other inputs. Emission analyses were done by the HORIBA-MEXA 7000 for $NO_x$ and the opacimeter AVL439 as an indicator for *PM* respectively.

For each of the local models, measurements were done. Here, emissions response on uncorrelated steps of the input variables mentioned above were measured. As already mentioned delay time varies heavily for different operating points on the global working range (Fig. 1). In this case the description with local models again shows an advantage, namely adaptation of the delay time. After correction of this, the static polynomial model (4) and the parameters of the transfer function (2) have been identified iteratively to minimize the cost function *J*. Furthermore, elimination of single input channels of $\Delta\tilde{u}$ was done iteratively by decreasing the limit of maximal allowed normalized standard error until residuals on a defined data interval, only used for validation, began to increase significantly. Fig. 4 and 5 show identification and validation for $NO_x$ and opacity at local levels.

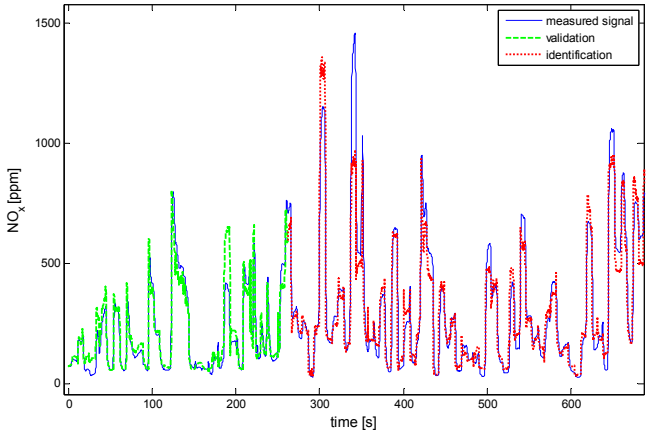


Fig. 4. Local model for $NO_x$ for the working point 950 rpm and 15 mg/cycle of main injection

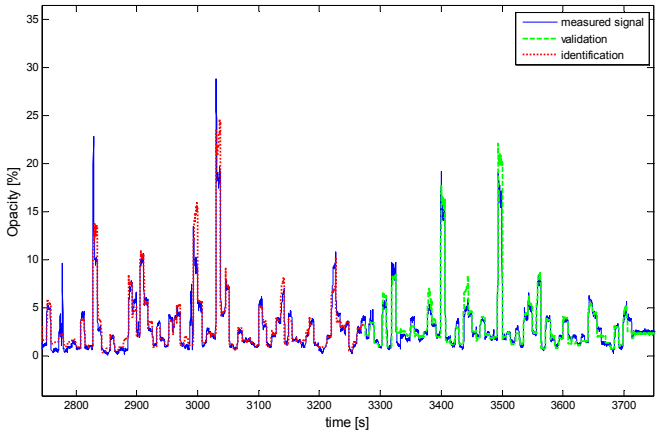


Fig. 5. Local model for opacity (representing PM) for the working point 2200 rpm and 15 mg/cycle of main injection.

## 6. RESULTS

Validation of the global model was done by using the identified global model as a virtual sensor during the *NEDC* standard driving cycle. Though no data of this cycle has been included in the identification process at all, results of good quality could be reached as Figures 6 and 7 show. Here is to mention that delay times of the measured signals have been compensated in order to enable a fair comparison.

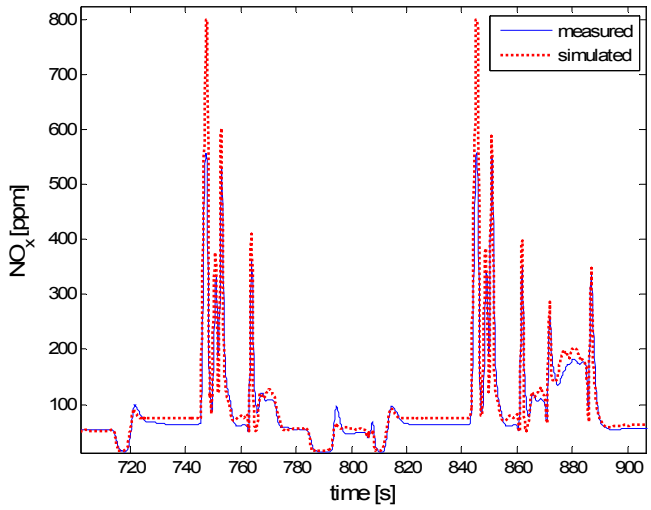


Fig. 6. Comparison between measurements and simulation result for $NO_x$ (selection of the *NEDC*)

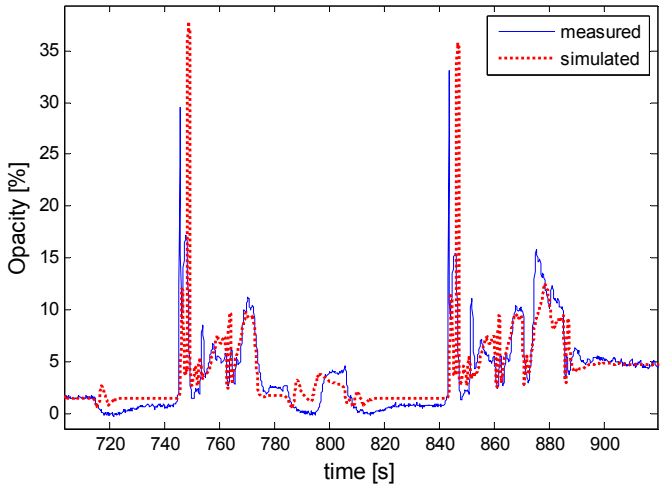


Fig. 7. Comparison between measurements and simulation result for opacity representing *PM* (selection of the *NEDC*)

In table 2 two characteristic values for identification evaluation the FIT-value and the mean square error (MSE) (8), are listed for global validation and local validation (mean values).

**Table 2. Model Errors**

| Output | mean local FIT | mean local MSE | global FIT | global MSE |
|---|---|---|---|---|
| $NO_x$ | 88% | 34 ppm | 84% | 35ppm |
| *Opac* | 75% | 3.2 % opc | 52% | 5.7 % opc |

$$FIT = \max\left(1 - \frac{\sqrt{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}}{\sqrt{\sum_{k=1}^{n}(y_k - \bar{y})^2}}, 0\right) \qquad (8)$$

$$MSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}$$

Though especially the model for $NO_x$ shows good results, difficulties can be seen in the high peaks where the virtual sensor displays even higher values. These and also other inaccuracies can be explained easily by the fact that here the local models have to be extrapolated. Fast and high fuel reduction rates from high levels of injections where high values of *MAF* and especially *MAP* occur, cause fast changes to models of lower mean injection. *MAF* and especially *MAP* values are not decreasing that fast which means that the local model of low load has to be extrapolated in a wide range as in identification data neither points containing such high values were included nor were possible. Polynomial functions of low order show much better extrapolation abilities than for example *ANN*-models but here even these are overstrained. To handle these effects more accurate, suitable regions of transient identification cycles could be included into the identification process of the local models.

These effects are apparently influencing the opacity model much more. Though the validation at local models is quite good, global validation seems to have room for improvement. It is to mention here that especially soot model identification was very difficult at low speed levels with higher injection. Especially in regions where rail pressure is limited by low engine speed, many effects are caused by local flame instabilities which are hardly describable.

General speaking, identification of models at low load was more difficult than doing this at higher load areas. This can be explained by the fact that variation of inputs was hold constant for all models and therefore for instance a 3 mg/cycle fuel mass change at a mean injection level of 5 mg/cycle has relatively much more influence than doing this at a mean level of 25 mg/cycle.

## 6. CONCLUSIONS

The main message of this work is that our method is able to provide a sufficient good model. The global structure of switching between local models shows advantages as different dynamics as well as sensitivities can be handled easily. In combination with the variable selection criterion for the local models, this structure shows a wide range of validity as well as high accuracy. Though improvements especially for the *PM* model are possible, their necessity has to be checked by using these for control optimization. Finally, it could be shown that accurate engine emission models containing only *ECU* values are possible and therefore predestined as basis for control applications.

## 7. ACKNOWLEDGEMENT

This work was supported by the Linz Center of Competence in Mechatronics (LCM) under grant LCM-001

## REFERENCES

Alberer, D., L. del Re, S. Winkler, P. Langthaler (2005) Virtual Sensor Design of Particulate and Nitric Oxide in a DI Diesel Engine, *ICE 2005*, paper Nr: 2005-24-063.

Ayeb, M., H. J. Theuerkauf and Thomas Winsel (2005) SI Engine Emissions Model Based on Dynamic Neural Networks and D-Optimality, *SAE 2005*, Paper Nr. 2005-01-0019 .

Del Re, L., P. Langthaler, C. Furmüller, S. Winkler, M. Affenzeller (2005) NOx Virtual Sensor Based on Structure Identification and Global Optimization, *SAE 2005*, Paper Nr. 2005-01-0050.

Desantes, J. M., J. J. López, J. M. García and L. Hernández (2002) Application of Neural Networks for Prediction and Optimization of Exhaust Emissions in a H.D. Diesel Engine, *SAE 2002*, Paper Nr. 2002-01-1144.

Galindo, J., J. M. Luján, J. R. Serrano, L. Hernández (2005), Combustion simulation of turbocharger HSDI Diesel engines during transient operation using neural networks, *Applied Thermal Engineering 25*, 877-898.

Hafner, M., M. Schüller, O. Nelles, R. Isermann (2000) Fast neural networks for diesel engine control design, *Control Engineering Practice 8*, 1211-1221.

Hennig C. (2004). Modellwahl und Variablenselektion in der Statistik [Online]. Available: http://www.math.uni-hamburg.de/home/hennig/lehre/mskript1.pdf (September 10, 2007

Heywood, J. B. (1988). *Internal Combustion Engine Fundamentals*, McGrwa-Hill series in mechanical engineering, Singapore.

Jung, D. and D. Assanis (2001). Multi-Zone DI Diesel Spray Combustion Model for Cycle Simulation Studies of Engine Performance and Emissions, *SAE 2001*, Paper Nr. 2001-01-1246 .

Li, X. and J. S. Wallace (1995). A Phenomenological Model for Soot Formation and Oxidation in Direct-Injection Diesel Engines, *SAE 1995*, Paper Nr. 952428.

Ljung, L. (1999) *System Identification 2nd edition* Prentice Hall PTR, United States of America.

Nelles, O. (2001) Nonlinear System Identification, Springer-Verlag, Berlin Heidelberg New York.

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

Patterson, M. A., S.-C. Kong, G. J. Hampson and R. D. Reitz (1994). Modeling the Effects of Fuel Injection Characteristics on Diesel Engine Soot and NOx Emissions, *SAE 1994*, Paper Nr. 940523.

Schilling, A., A. Amstutz, C. H. Onder and L. Guzzella (2006) A Real-Time Model for the Prediction of the NOx Emissions in DI Diesel Engines, *Proceedings of the 2006 IEEE International Conference on Control Applications.*

Vierlinger, H. (2005) *Verbesserung eines Frischluft-massenmodells mit Hilfe von Umgebungs-größen an einem modernen Dieselmotor*, Diploma Thesis – Johannes Kepler University, Linz, Austria, Institute for Design and Control of Mechatronical Systems.

Wang, G., G. Li, Y. Liu, L. Chen, X. Zhang and J. Lu (1999), A Developed Model for Emission Prediction, *SAE 1999*, Paper Nr. 1999-01-0233.

Warnatz, J., U. Maas and R.W. Dibble (1999), *Combustion second Edition* Springer-Verlag, Berlin Heidelberg.