15th Triennial World Congress, Barcelona, Spain

NEURAL NETWORK APPLICATIONS FOR MODEL BASED FAULT DETECTION WITH PARITY EQUATIONS

A. Schwarte [*,1], R. Isermann [*,1]

[*] *Darmstadt University of Technology*
Institute of Automatic Control
Landgfraf-Georg-Str. 4, D-64283 Darmstadt
Phone: +49 6151 162214, Fax: +49 6151 293445
ASchwarte@IAT.TU-Darmstadt.de

Abstract: The rising complexity of modern automotive engines with an increasing number of actuators and sensors to minimise emissions and fuel consumption and to maximise engine driveability require a detailed supervision for fault detection and on-board diagnosis. The European Community Directive 98/69/EC requires on-board diagnosis for spark ignition engines and will require it for diesel engines as of January 2003, mainly to prevent excessive emissions. Beside this regulation it is also in the interest of the automobile manufactures to establish capable diagnosis systems for maintenance, repair and the benefit of their customers. This paper will describe applications of neural networks for modelling complex fluid- and thermodynamics with unknown physical model structure. Reference models, which describe the fault free process, are set up and identified with the special neural network LOLIMOT (Local-Linear-Model-Tree). Fault detection algorithms, which employ the method of parity equations, were successfully implemented and tested in real time with a 2 litre diesel engine and a Rapid Control Prototyping System. Measurements of online fault detection are shown for several built-in faults in the intake system of this diesel engine. *Copyright © 2002 IFAC*

Keywords: Neural network models, fault detection, modelling, identification, real time.

1. INTRODUCTION

This paper focuses on neural network applications for model based fault detection with parity equations applied on a modern DI diesel engine. For automotive mass production applications suitable models have to be found, which permit a statement about the faultless or faulty system state with the strained information of only a few sensors. Starting from physical considerations or a detailed physical model simplified substitution models can be derived and their parameters then have to be obtained by identification methods.

But physical models of combustion engines are characterised by complex non-linear fluid- and thermodynamics so that even simplified physical models are often to complex for implementation in engine control units. Another drawback of physical models is that the may rest upon unmeasured signals or even simplified physical model structures are not known at all. Therefore neural networks have proven to be a powerful tool to model strong non-linearities with unknown physical model structure. By this way black box modelling with neural networks comprises both automatic model structure generation and identification of its parameters. Another favourably usage of neural networks is the hybrid modelling of processes with only partly known physical structure. There the combination of modelling the dominant characteristics physically and modelling non-linear phenomena and secondary effects with neural networks results in an overall good performance of the hybrid model. Several neural network models of the turbo charged diesel engine have been identified and were successfully implemented in real time model based fault detection algorithms on a dSPACE Rapid Control Prototyping System (Schwarte *et al*.,

[1]The authors are indebted to the Forschungsvereinigung Verbrennungskraftmaschinen (FVV), Frankfurt/Germany and the Arbeitsgemeinschaft industrieller Forschungs-anstalten (AiF) for the financial support of this research.

The investigations are accompanied by a committee under the chairmanship of Dipl.-Ing. B. Neumann, Adam Opel AG. Their support is also greatly acknowledged.

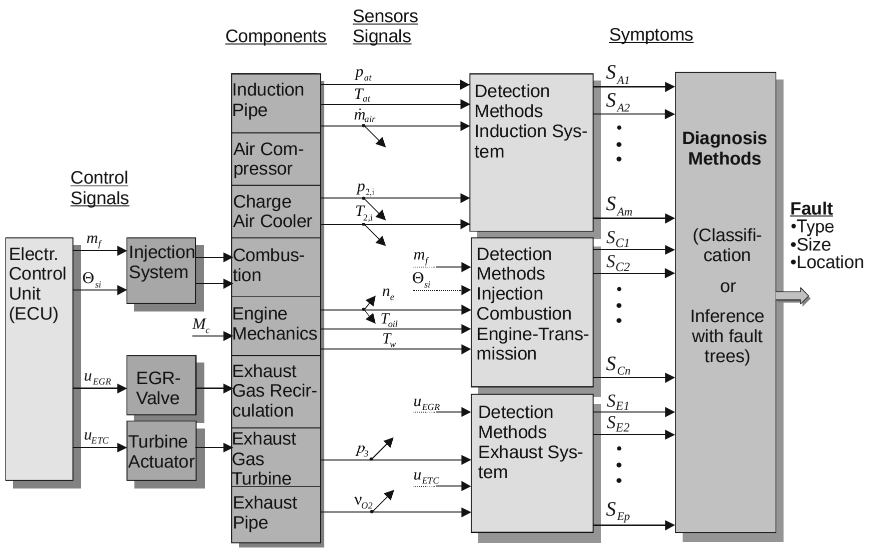

Fig. 1. Concept of unitised model-based fault detection of the complete combastion engine

2001). The test engine is an Opel 2litre, 4 cylinder, 16 valve tu rbo charged DI Diesel engine with a power of 74 kW and a torque of 205Nm. The engine employs exhaust gas recirculation and a variable swirl of the inlet gas for emission reduction. The concept for model based fault detection and diagnosis of the complete engine is shown in Fig.1. The engine is partitioned in three major subsystems: Induction system, injection, combustion and engine transmission system as well the exhaust gas system. Fig.1 shows the information flow. The actuators are driven by the electronic control unit and act on the different components of the combustion engine. There a selection of some state variables is measured. For each major subsystem, fault detection methods are developed to detect faults in the shown components and indicate them with symptoms.

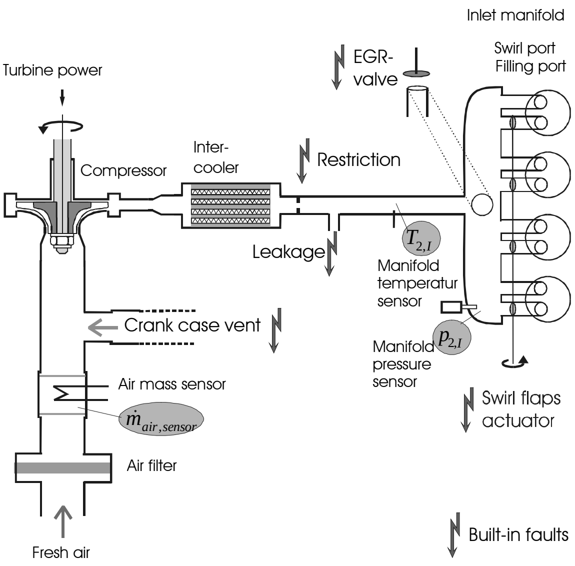

Fig. 2. Air path of the intake system.

In a further step symptoms are processed with diagnosis methods to characterise faults according to their type, size and location. Further on in this paper the fault detection in the intake system will be addressed.

2. INTAKE SYSTEM

The intake system of the considered Opel DTI Engine is representative for a modern turbo charged passenger car diesel engine. Fig. 2 shows the air path of the diesel engine. According to the flow direction of the fresh air the components air filter, air mass flow sensor, compressor, intercooler and inlet manifold are flown through. Between air mass flow sensor and compressor the blow-by is lead back by a small vacuum. Recirculated exhaust gas is added to the fresh air by the exhaust gas recirculation (EGR) valve in the inlet manifold. Each cylinder is filled by a swirl port and a filling port. By throttling the filling port of each cylinder with the swirl flaps actuator a variable swirl can be adjusted. Further components of the intake system are the pneumatic membrane actuators to control the swirl flaps and the EGR valve as well as the electro-pneumatic converters which convert a pulse width modulated (PWM) signal of the electronic engine control unit to the control pressure for the pneumatic actuators. By cutting the intake system from its neighboring subsystems, namely environment, engine block and exhaust gas system, input and output variables emerge according to Fig. 3. Only a few input and output variables are measured and can be used for fault detection.

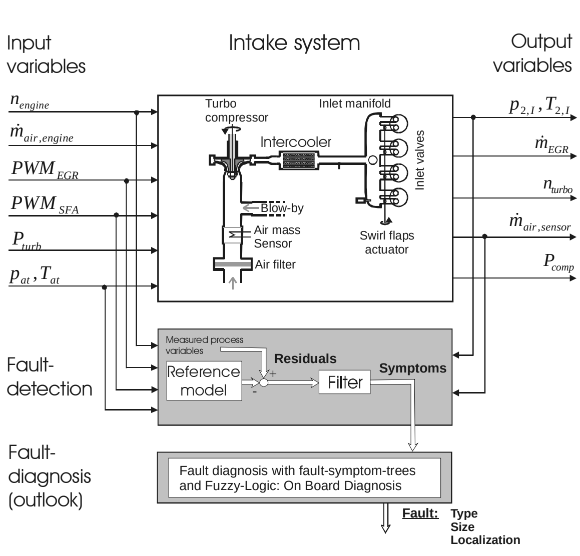

Fig. 3. Fault diagnosis structure of the intake system.

3. REQUIREMENTS FOR MODEL BASED FAULT DETECTION

For practical applications the model based fault detection methods have to be developed with respect to following aspects:

1. Detection of small faults.
2. Robustness.
3. Hardware requirements.
4. Application effort.
5. Transparency and interpretability.
6. Validity in a wide operating range.

Detection of small faults and robustness are very contradictory requirements and often the compromise has to be in favour of robustness. In automotive applications the hardware requires fault detection algorithms with restricted computational effort. The application effort should also be small and without (many) iteration loops. Transparency and interpretability are important for the transfer of developed methods to applications in automotive control units. Neural networks often have the disadvantage of bad transparency and interpretability. Therefore the approach in this paper is to use neural networks for identification in the first step and to generate look-up tables for real time fault detection in the second step. The validity of fault detection methods in the operating range is often narrowed by conditions like steady state operation, certain engine speed or load, etc. because process models can gain significant accuracy by limiting their validity range and therefore enhance detection of small faults.

4. MODELING AND IDENTIFICATION

4.1 Modelling of non-linear engine pumping

The engine pumping, describing the air mass flow into the engine, was modelled with a hybrid physical and neural network model. It is a mean value model of one working cycle neglecting the periodic working principle. The physical model part models the engine pumping corresponding to an ideal positive-displacement pump and describes the so-called theoretical air mass flow into the engine:

$$\dot{m}_{air,th} = \frac{1}{2} \cdot n_{eng} \cdot V_D \cdot \frac{p_{2,I}}{R \cdot T_{2,I}} \qquad (1)$$

All other non-linear induction phenomena are combined and described with the operating point dependent coefficient volumetric efficiency (Heywood, J. B. 1988). This coefficient is the ratio of real air mass flow and theoretical air mass flow:

$$\eta_V = \frac{\dot{m}_{air,eng}}{\dot{m}_{air,th}} = \frac{\dot{m}_{air,eng}}{\frac{1}{2} \cdot n_{eng} \cdot V_D \cdot \frac{p_{2,I}}{R \cdot T_{2,I}}} \qquad (2)$$

The non-linear phenomena like charge heating, backflow, flow friction, ram effect, etc. are very hard to model physically and are therefore modelled and identified with a neural network in dependency of engine speed and boost air density. The modelled air mass flow is as follows:

$$\dot{m}_{air,eng} \simeq f_{\eta_V}\left(n_{eng}, \rho_{2,I}\right) \cdot \frac{1}{2} \cdot n_{eng} \cdot V_D \cdot \frac{p_{2,I}}{R \cdot T_{2,I}} \qquad (3)$$

4.2 Modelling air mass flow oscillation

The air mass flow oscillation is stimulated by the periodic the suck in of each cylinder. In the time domain the frequencies of the air mass flow oscillation are proportional to the engine speed. But in the angle domain these frequencies are constant. The main (angle-)frequency of the air mass flow oscillation has a constant 180°CA (crank angle) period for four cylinder two-stroke engines. Measurements have shown that an approximation with a mean value and one harmonic describes the real mass flow sufficiently. Therefore, a signal model was set up with operating point dependent amplitude and phase of the mass flow oscillation.

$$\begin{aligned} &\dot{m}_{air,sensor}(\alpha) \\ &= \bar{\dot{m}}_{air,snsor} + A_{\dot{m}_{air,sensor}} \cdot \cos\left(2\pi \cdot \frac{\alpha}{180°CA} + \varphi_{\dot{m}_{air,sensor}}\right) \end{aligned} \qquad (4)$$

$$A_{\dot{m}_{air,sensor}} \simeq f_{A_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right) \qquad (5)$$

$$\varphi_{\dot{m}_{air,sensor}} \simeq f_{\varphi_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right) \qquad (6)$$

Fig.4 shows the air mass flow for an exemplary operation point with a one harmonic approximation.

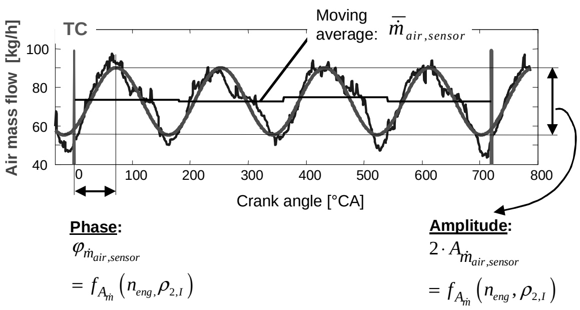

Fig.4. Example of air mass oscillation with its characteristics at $1200 min^{-1}$

4.3 Modelling boost pressure oscillation

Like with the air mass flow oscillation the boost pressure oscillation is also stimulated by the periodic induction of each cylinder and its signal model is similar to the one before:

$$\begin{aligned} &p_{2,I}(\alpha) \\ &= \bar{p}_{2,I} + A_{p_{2,I}} \cdot \cos\left(2\pi \cdot \frac{\alpha}{180°CA} + \varphi_{p_{2,I}}\right) \end{aligned} \qquad (7)$$

$$A_{p_{2,I}} \simeq f_{A_p}\left(n_{eng}, \rho_{2,I}\right) \qquad (8)$$

$$\varphi_{p_{2,I}} \simeq f_{\varphi_p}\left(n_{eng}, \rho_{2,I}\right) \qquad (9)$$

4.4 Identification of reference models with neural networks

In this paper neural networks are employed for modelling parts of the intake systems with little knowledge of the inner physical model structure. The special neural network LOLIMOT, which was developed by Oliver Nelles at Institute of Automatic Control (Nelles, O., 1999, 2001), uses a structure of local linear models. The local linear models are identified with an orthogonal divided input space. The validity of each model is 100% in its centre and decrease towards its neighbour models, so that the sum of model validities at each point is 100%. By this way there is a smooth transition between the local models and the overall model is steady differentiable. The approach with local linear models leads to fast NN training properties. LOLIMOT comprises both automatic model structure generation and the identification of model parameters. The model structure is generated by the LOLIMOT algorithm in an iterative way adapting to the variable complexity, respectively non-linearity, of the identified system. Fig.5 shows as an example the input space and the local linear model structure of the neural network which models the volumetric efficiency of the engine pumping. The ‘x’ mark the centre the different local model. The lines separate the local models by their dominant validity.

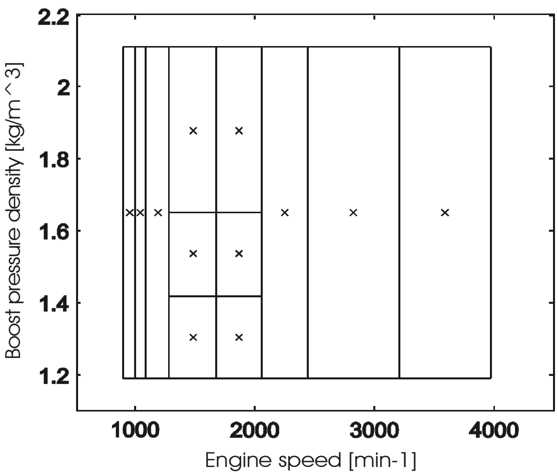

Fig.5. Structure of local linear models with their centres ‘x’ and range of dominant validity

For the fault free description of the intake system 5 static reference models were identified, which describe the volumetric efficiency, the amplitude of air mass flow oscillation, the phase of air mass flow oscillation, the amplitude of boost pressure oscillation and the phase of boost pressure oscillation. The reference models were identified for a closed EGR valve and opened swirl flaps actuator with a quasi stationary identification cycle, Fig.6. The identification cycle should preferable stimulate only very low frequencies and should evenly distribute data points over the complete input space. The distribution of data points is shown in Fig.7. However, high boost pressure cannot be achieved at low engine speed and therefore the input space region with low engine speed and medium to high boost density is empty. Fig.8 shows the identified reference model for the volumetric efficiency compared with the data points.

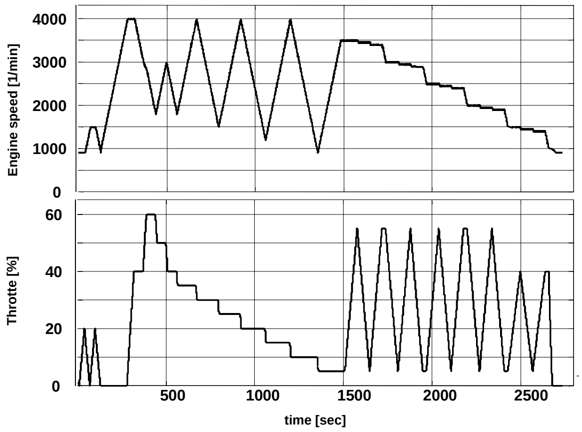

Fig.6. Identification cycle

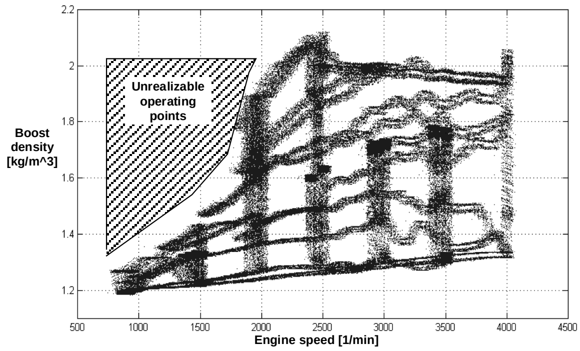

Fig.7. Input space with the distribution of data points

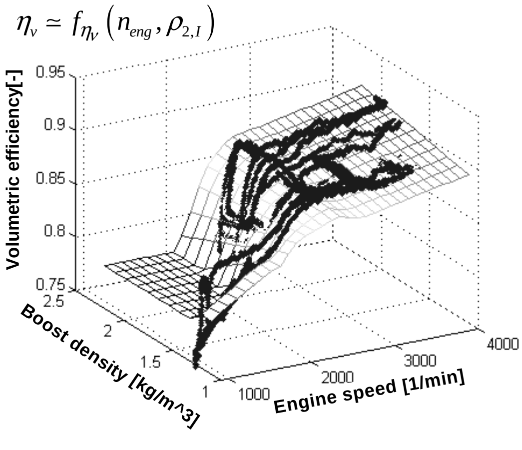

Fig.8. Identification result; reference model of volumetric efficiency compared with data points

LOLIMOT identifies the model with a continues mathematical function. However in the online application look-up tables are employed for the reference models in order to reduce computational effort and to be compatible to the general representation of non-linear models in automotive applications. This approach provides a great advantage opposite to the general grating measurement regarding flexibility, complete measurement and the effort to measure exactly certain operating points. With the only once identified neural network any desired grating of the look-up table in the electronic engine control can be applied afterwards. Fig.9 shows the identified reference model for the amplitude of the air mass flow oscillation. At an engine speed of about 2000min^{-1} there is the highest amplitude at the resonance frequency of the air path pipe. Fig.10 then shows the corresponding identification result and the generated reference model of the phase of the air mass flow oscillation. The phase is mainly a linear function of the engine speed which is reasonable. It was proceeded likewise with the identification of the boost pressure oscillation. The two reference models for amplitude and phase of the boost pressure oscillation are similar to the air mass flow oscillation and are not shown here.

$$A_{\dot{m}_{air,sensor}} \simeq f_{A_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right)$$

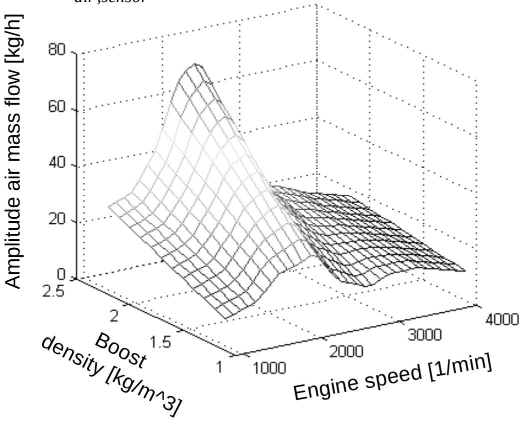

Fig.9. Identification result; reference model amplitude of air mass flow oscillation

$$\varphi_{\dot{m}_{air,sensor}} \simeq f_{\varphi_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right)$$

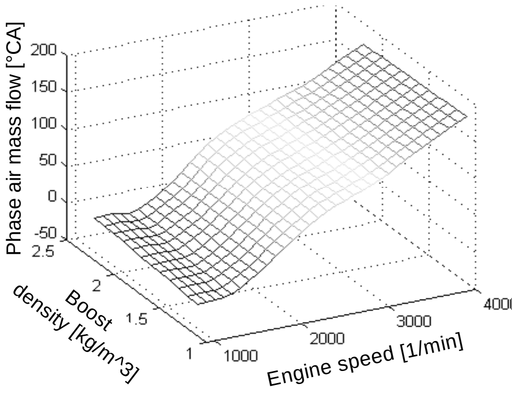

Fig.10. Identification result; reference model phase of air mass flow oscillation

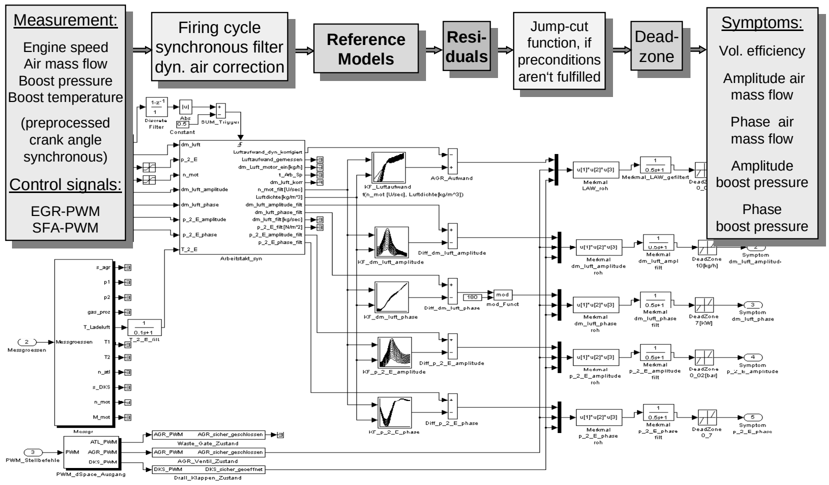

Fig. 11. Implementation of real time model based fault detection on a dSPACE Rapid Control Prototyping System with MATLAB/SIMULINK

5. PARITY EQUATIONS AND RESIDUALS

Based on the five identified reference models one is now able to use the analytical redundancy between the model outputs and the measured values to set up five independent parity equations. Therefore five residuals can be calculated as follows:

$$r_{\eta_V} = f_{\eta_V}\left(n_{eng}, \rho_{2,I}\right) - \eta_v \tag{10}$$

$$r_{A_{\dot{m}}} = A_{\dot{m}_{air,sensor}} - f_{A_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right) \tag{11}$$

$$r_{\varphi_{\dot{m}}} = \varphi_{\dot{m}_{air,sensor}} - f_{\varphi_{\dot{m}}}\left(n_{eng}, \rho_{2,I}\right) \tag{12}$$

$$r_{A_p} = A_{p_{2,I}} - f_{A_p}\left(n_{eng}, \rho_{2,I}\right) \tag{13}$$

$$r_{\varphi_p} = \varphi_{p_{2,I}} - f_{\varphi_p}\left(n_{eng}, \rho_{2,I}\right) \tag{14}$$

6. REAL TIME FAULT DETECTION

The signal processing and the model based fault detection algorithms are implemented on a Rapid Control Prototyping System with MATLAB/-SIMULINK according to Fig.11. The measured signals are crank angle synchronously preprocessed for a better signal to noise ratio. Then firing cycle synchronous algorithms calculate the characteristic values. These values are compared to the output of the reference models and residuals are calculated. The residuals are further processed and the symptoms result. Finally the symptoms indicate faults by deviating from zero. The results of real time fault detection shall be presented in Fig.12 for an exemplary operating point. Several faults were temporarily built in at the intake system. The fault detection thresholds are marked by dotted lines. The reference models, volumetric efficiency, amplitude air mass flow oscillation, amplitude boost pressure oscillation, show the expected behaviour. In the fault free case these residuals are almost zero. The reference models phase air flow and phase boost pressure have wider range in the fault free case of approximately 4°CA. Therefore, the first three symptoms are best suited for fault detection. The first fault example in Fig.12 shows the case of undesired closed filling ports. The residuals amplitude boost pressure oscillation and amplitude air mass flow oscillation clearly trespass the thresholds and with the corresponding symptoms this fault is clearly detected without position sensor. The second fault example is an undesired opened EGR valve. The residual volumetric efficiency responses intensively and it is obvious that smaller EGR faults are detected as well. Additionally, the residuals amplitude boost pressure oscillation and amplitude air flow oscillation show a strong deflection, too. The third example shows different size leakages in the air path between intercooler and engine. The bigger the leakage diameter, the stronger is the deflection of the residual volumetric efficiency. The leakages with 4mm, 5mm and 7mm in diameter are well detected. As a further fault example the crank case vent pipe was removed. This is equivalent to a leakage between air mass flow sensor and compressor. Additional air which is not measured, is sucked into the air path. With the symptom volumetric efficiency this fault is clearly detected as well. Note that the symptom volumetric efficiency responses opposite to the leakage between intercooler and engine. The last fault example presented depicts a short-time restriction between intercooler and engine. This fault is detected by the symptoms amplitude boost pressure oscillation and amplitude air mass flow oscillation. The demonstrated faults were detected very fast in just a few tenth part of a second. Fig.14 depicts the fault-symptom-correlation in a summary.

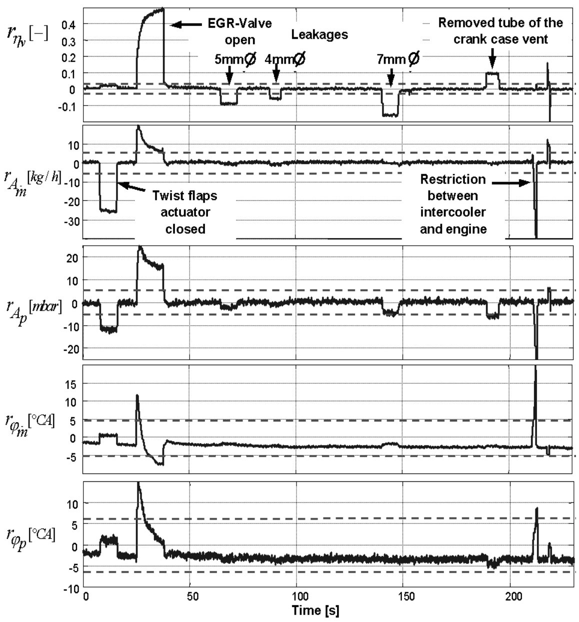

Fig.12. Residuals in dependency of faults (online), 2000min^{-1}, 130Nm, $p_{2,I}$ =1.5bar, air flow 165kg/h

Fault	Symptoms				
	s_{η_V}	$s_{A_{\dot m}}$	s_{A_p}	$s_{\varphi_{\dot m}}$	s_{φ_p}
Crank case vent: pipe is open at junction	+	o	o	o	o
Leakage between intercooler and engine	-	o	o	o	o
Restriction between intercooler and engine	o	-	-	+	+
Swirl flaps actuator, filling port is closed	o	-	-	o	o
EGR-valve: stuck at open	++	-	-	o	o
Leaky EGR-valve	+	o	o	o	o

Legend:

++ Symptom responds intense positive
+ Symptom responds positive
- Symptom responds negative
o Symptom does not respond

s_{η_V} Symptom volumetric efficiency
$s_{A_{\dot m}}$ Symptom amplitude air mass flow oscillation
s_{A_p} Symptom amplitude boost pressure oscillation
$s_{\varphi_{\dot m}}$ Symptom phase air mass flow oscillation
s_{φ_p} Symptom phase boost pressure oscillation

Fig.14. Fault-symptom-table of the intake system

For each fault it is listed how the described symptoms respond. In doing so it is differentiated between intense positive, positive, negative and no response. It results a symptom pattern for each fault, so that all faults can be diagnosed.

6. CONCLUSIONS

The results in this paper show that the application of neural networks for model based fault detection with parity equations offers great advantages. The identification of non-linear models is faster, more flexible and covers more data points than the usual identification with grating measurements. The advantage will even be bigger if one is forced to identify non-linear models with more dimensions. The special neural network LOLIMOT can also easily be extended to dynamic non-linear models. In final applications look-up tables can be generated with any desired grating form the only once identified continues mathematical model functions. This approach provides great flexibility and sustains the compatibility to the usual non-linear model representation in automotive applications.

7. REFERENCES

Heywood, J. B. (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, Inc.

Nelles, O. (1999), *Nonlinear System Identification with Local Linear Neuro-Fuzzy Models*, Shaker Verlag, Aachen.

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

Schwarte, A., Kimmich, F. and Isermann, R. (2001), *Modellgestützte präventive Diagnosemethoden (Fehlerfrüherkennung) für Dieselmotoren*, FVV-Abschlussbericht, Heft 708, Forschungsvereinigung Verbrennungskraftmaschinen e.V., Frankfurt am Main.