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

Control of future low Temperature Combustion Technologies with nonlinear Model based Predictive Control based on Neural Networks

Kai Hoffmann*, Dieter Seebach†, Prof. Stefan Pischinger†, Prof. Dirk Abel*

** Institute of Automatic Control IRT, RWTH Aachen University, 52056 Aachen, Germany (Tel: +49-241-8027507; e-mail: k.hoffmann@irt.rwth-aachen.de)*
† Institute for Combustion Engines VKA, RWTH Aachen University, 52056 Aachen, Germany (e-mail: office@vka.rwth-aachen.de)

Abstract: The combustion in future engines will work with a very high amount of recirculated exhaust gas in part load conditions to enable a low peak combustion temperature. This combustion suffers from instabilities of the process and a highly nonlinear behaviour. The paper presents the use of neural nets for observing the engine. A nonlinear model without feedback of measurements is linearised online and combined with an extended Kalman filter. This observer is compared to a neural net with observer structure by application to two different valve timing strategies. The more promising observer is combined with a model based predictive controller with a quadratic cost function. Its analytic solution is compared with quadratic programming for respecting constraints in the prediction for improving the control error.

1. INTRODUCTION

The need for a reliable energy supply characterizes the everyday life in private households, industrial processes and transportation. The awareness of environmental pollution has rapidly grown in the whole world within the recent years. Most industrialized countries agreed in the Kyoto Protocol to rapidly reduce their over all emissions. It entered into force on February 16th 2005. A big part of a nation's emission is caused by traffic. Therefore the legislature in most countries enforced tighter emission laws for automobiles and transportation in general. E.g. the European parliament decided the Euro 6 legislation in December 2006. According to the stricter emission specifications the research on combustion systems focuses apart from a general improvement of efficiency and carbon dioxide reduction on the reduction of the pollutants nitric oxides (NOx), soot, carbon monoxides and hydrocarbons, respectively. Therefore the development of new combustion processes is conditioned by the trade-offs between those emissions, which make the simultaneous reduction to contrary tasks. With the introduction of premixed lean combustion instead of the former diffusion controlled combustion especially the NOx-emissions could be reduced considerably. These combustion processes with a high homogenisation and exhaust gas recirculation (EGR) are known as HCCI (Homogeneous Charge Compression Ignition) or CAI (Controlled Auto Ignition).

As a part of a major project, the Collaborative Research Centre 686 "Model based control of homogenized low-temperature Combustion", a controller for combustion engines will be set up. Diesel and gasoline are the main focus of actual research, as their supply system is already highly developed and spread. The latter will be addressed in this paper. The low-temperature combustion of a gasoline-fuel will be named CAI in the following, but many other acronyms exist in literature. This combustion only can be realised with a modified valve train, which can offer a higher variability than the common fix valve train in usual series-engines. For the CAI-combustion the cylinder load is compressed until it self-ignites nearly simultaneously in several centres of the compression volume without the use of a spark. The nearly simultaneous reaction leads to a very steep pressure rise and therefore to a high thermodynamical efficiency. By recirculating exhaust gas with a high thermal heat capacity the peak temperature can be kept in a low range and a reduction of NOx by 90-99% can be realised. The fuel consumption drops due to the increased efficiency up to 15% (Lang *et al.*, 2005). However, instabilities arise in form of spatial and temporally distributed areas of ignition, which are sensitively depended on the temperature and the reaction's kinetics. The self-ignition is influenced by the thermal attributes of the cylinder load and its stratification, no longer by a spark or only the injection timing. As CAI works with a high amount of residual gas, the combustion of one cycle depends on the exhaust of the previous one. The start of combustion will be advanced with a higher compression end temperature. High pressure gradients and therefore inadmissible noise results are the consequence. The auto ignition may even misfire or at least be retarded with too low temperature at the top dead centre after the compression stroke. The only way to address this problem is a controller stabilizing the combustion while taking constraints into account, e.g. the limited acceptable frame for the actuators. Therefore a controller is needed, which can respect those constraints.

The group of researchers of the Collaborative Research Centre 686 will create a model based predictive controller (MPC) which uses a physically based model for this task. This will be reduced from detailed models, e.g. CFD, to a form running in real-time on the prototyping-hardware, which will calculate the controller. E.g. (Shaver *et. al.*) have demonstrated a physically based closed loop controller to control the peak pressure of a propane HCCI-combustion. At the start of our project no model at all is available for the model-based controller. Therefore this has to be substituted by an identified model allowing its extension by new found physical knowledge later in the project. As soon as a physical model of a part-aspect of the whole combustion process is available, this knowledge can be implemented. The part model's output *v(t)* can be either the total input to an identified model like an artificial neural net (ANN), case 2 in (Fig. 1), or a part of the inputs to the latter, case 4 in (Fig. 1). The third possibility is to set up two completely separate, but parallel models, one physical and one identified, each independently calculating one part of the output, case 3 in (Fig. 1). The identified model has to represent the highly nonlinear behaviour of the low temperature combustion in the engine and provide the opportunity for online use in a controller. E.g. (Colin *et al.,* 2007) applied a linearised neural predictive controller to reduce throttling losses of a downsized turbo charged SI engine. We also choose a special ANN for control which can be linearised to a discrete state space form, see (Hoffmann *et al.,* 2007) and section 3. This allows the easy superposition of a physical with the identified model with a physical part model in the linear and nonlinear case, see (Fig. 1).

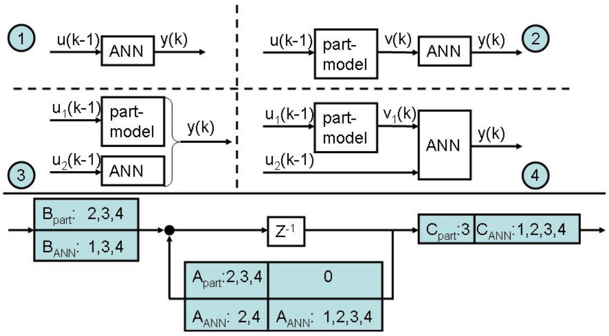

Fig. 1: Discrete time (non-)linear state space model superposition. The numbers indicate in which cases the linearised state space matrices do have an entry. Otherwise the lot is filled with the zero matrix of the correct size.

2 ENGINE SETUP AND ACTUATORS

The research is based on a single cylinder research engine, which is run by the Institute for Combustion Engines, RWTH Aachen University, while the controller will be set up by the Institute of Automatic Control, RWTH Aachen University. The motored engine is equipped with direct injection and a fully variable electromechanical valve train (EMVT). Further engine details are shown in (Table 1). The piston's shape allows an opening of the valves at any desired point of time.

Table 1 Setup research single cylinder engine

Bore	Stroke	Con-rod length	Displacement	Compression-ratio
84 mm	90 mm	159 mm	0.499 dm^3	13 [-]

The combustion process is controlled by two actuators, namely the valve timing for controlling the exhaust gas recirculation and the fuel injection for controlling as well the engine's load as the homogenisation process.

The fully variable valve train offers a high degree of freedom for the recirculation of the exhaust gas to the next combustion cycle which is reduced to the two most promising techniques.

The first option is to hold the exhaust from a previous cycle partly inside the combustion chamber (Combustion Chamber Recirculation CCR). This can be achieved by closing the exhaust valves before top dead centre of the gas exchange (TDC_{GE}). An amount of exhaust is trapped inside the cylinder and compressed until top dead centre is reached and released afterwards. Subsequently the necessary amount of air is aspired during a short lift of the intake valve.

A second possibility is to push out the exhaust, keep the valves opened over TDC_{GE} and reaspirate it from the exhaust port (Exhaust Port Recirculation EPR). Afterwards the needed fresh air is drawn from the intake port while the intake valves are opened. (Fig. 2) shows the direction of change for reducing the amount of residual gas for both valve timing strategies. Shown is the lift of the intake- and exhaust valve (IV / EV) over crank angel from lower dead centre of the high pressure cycle (LDC_{HP}) over top dead centre of the gas exchange (TDC_{GE}) till lower dead centre of the gas exchange (LDC_{GE}).

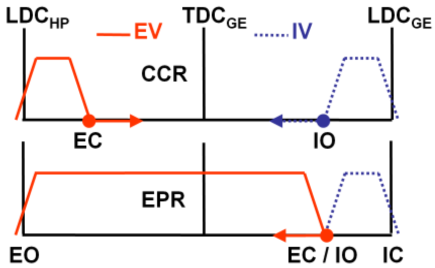

Fig. 2: Valve timing strategies for internal exhaust recirculation for Combustion Chamber Recirculation CCR (top) and Exhaust Port Recirculation EPR (bottom) with the direction of change for reducing residual gas.

With these two valve-timing strategies the degree of freedom for actuating the valve train can be reduced to one. With CCR the events exhaust valve opening (EO) and intake valve closing (IC) are fix, at least for one engine speed, and exhaust valve closing (EC) and intake valve opening (IO) are set symmetrically to the top dead centre position shifted with a little tradeoff. With EPR the same valve events depend on each other as exhaust valve closing and intake valve opening fall together or at least have a constant overlap for one engine speed.

The load can mainly be influenced by the injected amount of fuel per cycle or the energizing duration (ED) respectively, which means the opening or actuating time of the fuel direct injector. The grade of homogenisation of the injected fuel in the gas volume consisting of exhaust and air depends on the end of energizing event (EE) and influences as well the load as the exhaust gas.

Apparently the events end of energizing (EE), energizing duration (ED) and exhaust valve Closing (EC) and intake valve Opening (IO) are the main actuating variables for a constant revolution speed.

The values chosen to control are the indicated mean effective pressure *imep* and the efficiency of the combustion. A characteristic value for the latter in general and for control purposes in particular is the fuel mass fraction burned *MFB* in general and especially the 50% conversion point. It should always be kept five to ten degrees after top dead centre as presented in (Chiang *et al.*, 2007). This value so far is not calculated in the custom-made ECU based on a dSPACE MicroAutoBox, which is coevally developed. An easy to detect alternative is given by the location of the peak pressure *apmax*. For best performance in one operating point a specific location of the maximum pressure exists. Unlike the 50% *MFB* it is not nearly constant for best efficiency. Its optimal location changes slightly with the load but is a suitable measure for internal efficiency.

3 OBSERVER DESIGN

The aim of the research presented suggests a form for the identified model similar to a nonlinear state space, as the controller shall be a model based predictive controller (MPC), which is the only controller able to respect constraints in the algorithm. Additionally the identified model should be adoptable to new physical knowledge.

The model based predictive controller demands a linear state space model in every calculation step. The model therefore should have a form which can easily be linearised for the use internally in the MPC. A neural net is chosen which represents the nonlinear process in a sufficient way and has certain attributes of a linear state space. Different researchers have shown that MLP nets can approximate any continuous function to any desired accuracy, e.g. (Hornick *et al.*, 1989). Therefore the chosen model form is Neural Network Statespace Innovations Form (NNSSIF) following (Nørgaard 2000). This multilayer perceptron net (MLP) with one hidden and one output layer has hyperbolic tangent and linear activation functions in the hidden and the output layer, respectively, see (1).

$$\hat{y}_i(t) = g_i(\varphi,\theta) = \left(\sum_{j=1}^{n_h} W_{i,j} \tanh\left(\sum_{l=1}^{n_\varphi} w_{j,l}\varphi_l + w_{j,0} \right) + W_{i,0} \right) \quad (1)$$

Here $\varphi_l(t)$ is the regression vector containing the input values to the net while the vector θ contains the weights and biases $W_{i,j}$ and $w_{i,j}$ or $W_{i,0}$ and $w_{i,0}$, respectively. These are optimized during the offline training of the net. With a MIMO system for each of the systems outputs $\hat{y}_i$ one separate equation (1) is solved. The NNSSIF net is a special form of an MLP net whose outputs are part of the system's states. The actual states are built by adding retarding states to retarded states as described in (Hoffmann *et al.*, 2007). These become part of the regression vector, which also contains the inputs to the system and the deviation of the model from the process' output. The output of the process is calculated by multiplying the states with a fix matrix **C**, see (Fig. 3).

As shown in (Hoffmann *et al.,* 2007) this net already provides a structure of which the linearised form is a discrete time state space model including an observer matrix **K**, which originates from the feedback of the model's deviation $\varepsilon(t)$ from the measured process output. This, of course, is a form of observer. But it does not show integrating behaviour and therefore still leaves a deviation from the observed values, when other data than the training set is provided.

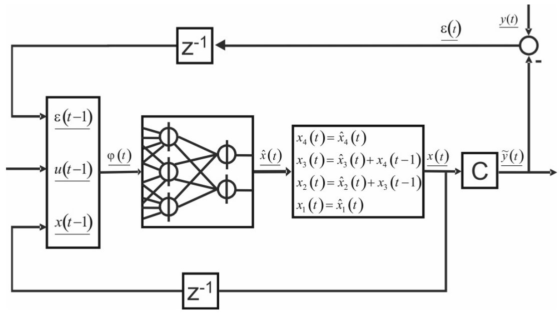

Fig. 3 Schema of an NNSSIF net with error feedback $\varepsilon(t)$.

The state space model can also be implemented in an extended Kalman filter to improve the NNSSIF net. Here the error feedback is neglected, as this is replaced by the extended Kalman filter. Additionally the observer is implemented with an output disturbance observer for zero tracking error following (Rossiter, 2003). To show the improvement by the extended Kalman filter, an NNSSIF net was trained for both valve-strategies, EPR and CCR. This net is used to observe an identified process model with superposed noise for verification. The used 5x5 engine simulation-model is a second NNSSIF without feedback of $\varepsilon(t)$. It is also trained and validated on separate step response measurements like the observing net. Inputs are EE, ED, EC, IO as described in section 2 and the revolution speed. Outputs are *imep*, *apmax,* the maximum pressure, its maximum gradient and the location of the latter. Only the actuating variables are used for the test, the other inputs remain constant. The noise is white Gaussian on the model's outputs and its intensity is the same as the measurement has shown. Both, the observer and the simulation-model are of second order which is evaluated in (Bengtsson *et al.*, 2006) for the control of a HCCI combustion engine with linear MPC.

As first actuating variable the energizing duration ED is chosen. This value corresponds to the time the injector is held open and therefore the injection of fuel. That means the increase of load as well as the retardation of the location of the peak pressure. For evaluation the test is accomplished with data different from the training data set of the NNSSIF net observer for both valve timing strategies as in (Fig. 2). The model based predictive controller to develop will base on a prediction of the future engine behaviour, which in turn depends on a good estimate of the current states. The controlled values *imep* and *apmax* are part of the states and are taken as a measure for the quality of the observance. It is essential not only that the observer offers a good result, but also that it is available for all valve timing strategies under research. Therefore (Fig. 4) and (Fig. 5) show the comparison of the NNSSIF net with extended Kalman filter NNSSIF observer for both valve timing strategies Combustion Chamber Recirculation CCR and Exhaust Port Recirculation EPR. Applying the extended Kalman filter permits the improvement of the observer for both engine operation modes. This can easily be seen from the accumulated standardised square errors. The simulation results in (Fig. 4) and (Fig. 5) demonstrate that the extended Kalman filter (red dashed lines) is able to observe the noisy model output (green solid lines) better than the original NNSSIF net (blue solid lines) can, although the operating point of the engine and therefore the operating conditions

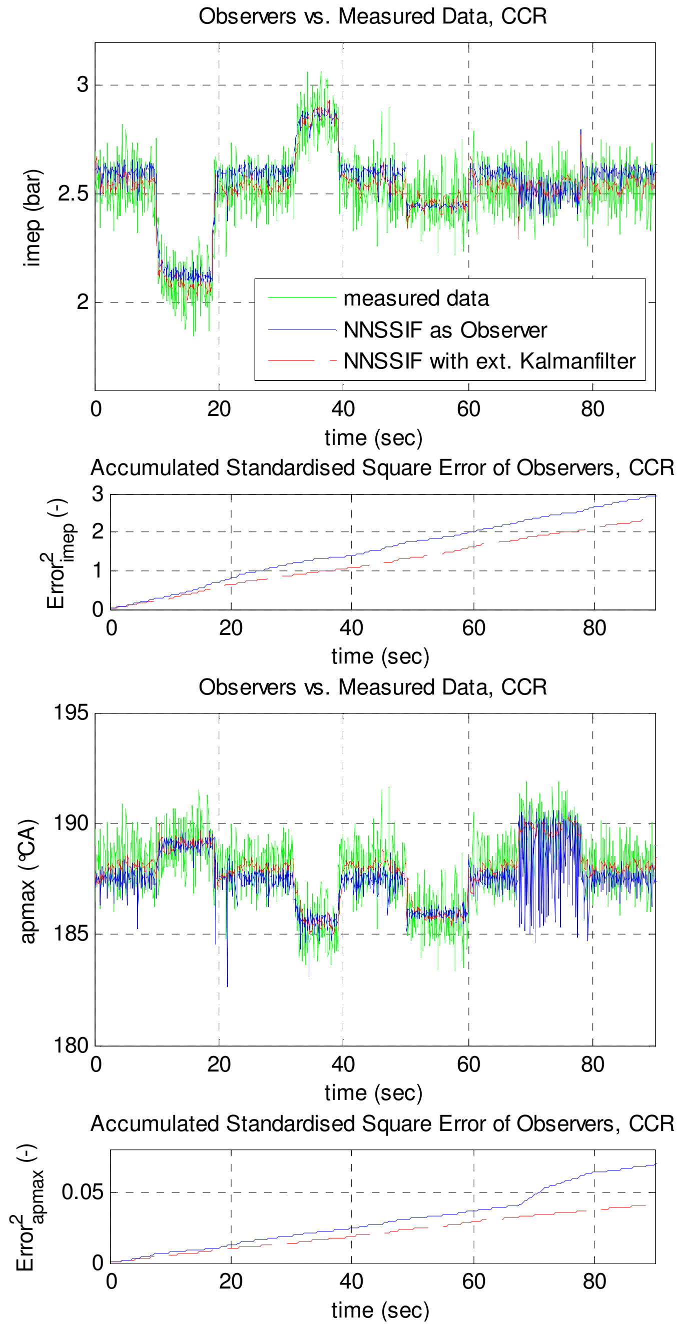

Fig. 4: Comparison of an NNSSIF observer with trained Error feedback to one with an integrating Kalman filter, engine operated in CCR-mode

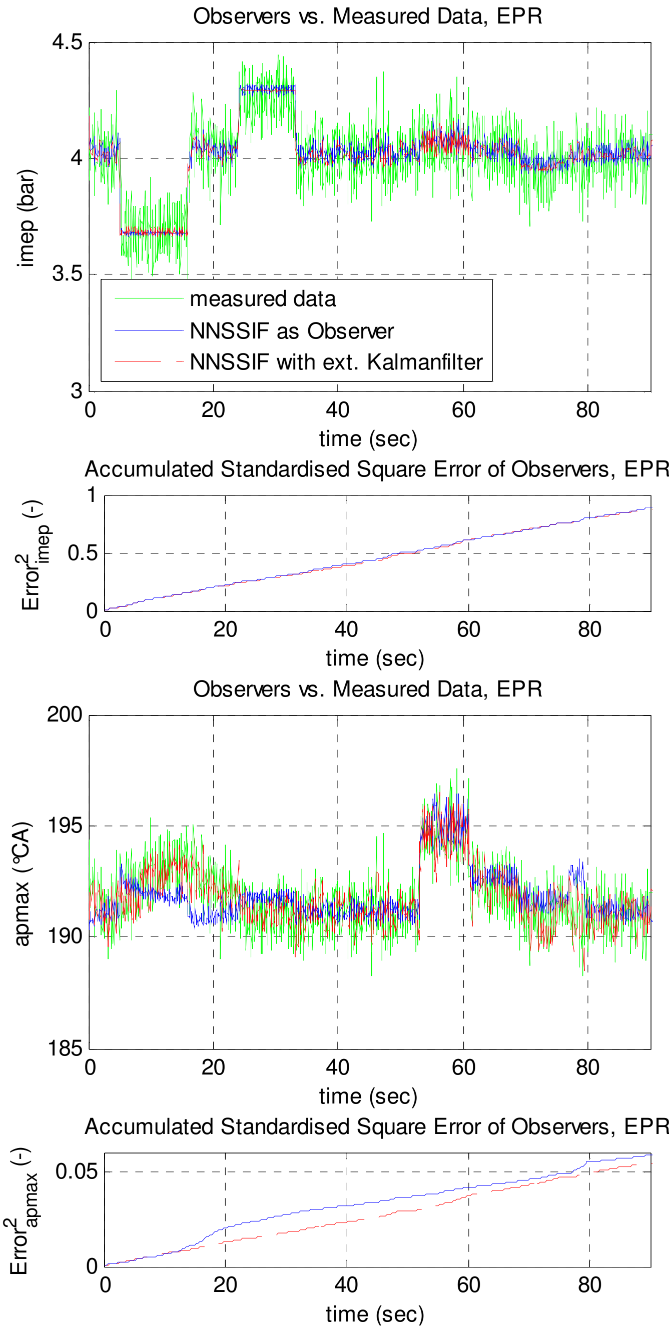

Fig. 5: Comparison of an NNSSIF observer with trained error feedback to one with an integrating Kalman filter, engine operated in EPR-mode

for the observers are close to the training data set of the neural net. The engine model is excited with similar actuating values as in the training set. Therefore the deviation between both observers cannot be caused by extrapolating effects of the neural net. The accumulated square error is standardised over the corresponding measurement value. The engine model is actuated first by a de-/increase sequence in energizing duration and second in exhaust valve closing. Note that this means for CCR an in-/decrease, for EPR a de-/increase sequence in recirculated gas. For both valve timing strategies the degree of freedom for the actuation of intake and exhaust valve is reduced to one according to section 2. The deviation of the estimate of the indicated mean effective pressure *imep* is well for both observers in both cases regarded, CCR and EPR. The example for CCR shows a higher slope in the deviation error for *imep* with the NNSSIF observer than with the extended Kalman filter. The EPR case shows only a minimally better performance of the extended Kalman filter for *imep*. With the location of the maximum pressure *apmax* the extended Kalman filter NNSSIF observer proofs its better performance, especially with rather instable operating conditions for the engine: E.g. with CCR the reduction of residual gas by shifting exhaust valve closed late at 70 seconds leads to rather instable conditions. With EPR the load de- and increase at seven and seventeen seconds respectively show lead to a high deviation. Here the accumulated error increases significantly.

4 CLOSED LOOP SIMULATION WITH MPC

The comparison of the observers showed the advantage of the NNSSIF-observer with extended Kalman filter. Therefore it is used in the implementation in a simulated closed control loop. The engine model is the same nonlinear second order model as described in section 3. Used actuators are the energizing duration ED and the valve timings exhaust valve closing EC or intake valve opening IO respectively. The end of energizing and the revolution speed are kept constant. Again the model's outputs are noisy as before. The observer focuses on the controlled variables indicated mean effective pressure *imep* and location of the maximum pressure *apmax*. In the closed control loop the observer linearises the nonlinear NNSSIF system in (Fig. 3) as described in (Hoffmann *et al.*, 2007) to a linear discrete time state space. The matrices **C** and **D** are constant and therefore become parameters to the controller. The matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ in contrast change with every time step and therefore are an input to the controller. The controller also receives the estimated states $\hat{\boldsymbol{x}}_k$ from the observer, the measured system's output *y(k)* and the predicted setpoint w_k. The letter is assumed constant over the predicted future.

The MPC is based on the optimization of the quadratic cost function (2). The two terms regard the deviation between predicted setpoint w_k and predicted system output $\hat{y}_k$ in a timeframe from lower to upper prediction horizon, from N_1 to N_2. The change of the actuating variable Δu_k is considered from the actual time step to the control horizon N_u. Both are quadratically weighted with the matrices $\mathbf{\Lambda}$ or $\mathbf{\Gamma}$ respectively:

$$J = \sum_{j=N_1}^{N_2}\left(\left(w_{k+j}-\hat{y}_{k+j}\right)^T \mathbf{\Gamma}\left(w_{k+j}-\hat{y}_{k+j}\right)\right)+\dots \quad \textbf{(2)}$$
$$\sum_{j=1}^{N_u}\left(\Delta u_{k+j}^T \mathbf{\Lambda} \Delta u_{k+j}\right)$$

$$\hat{Y}_k = Y(k) + \mathbf{F}_k\hat{x}_k - \mathbf{F}_{k-1}\hat{x}_{k-1} + \mathbf{H}_k\Delta U_k \quad (3)$$

$$\mathbf{F}_k = \begin{pmatrix} \mathbf{CA}_k^{N1} \\ \vdots \\ \mathbf{CA}_k^{N2} \end{pmatrix}$$

$$\mathbf{H}_k = \begin{pmatrix} \mathbf{CA}_k^{N1-1}\mathbf{B}_k & \cdots & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{CA}_k^{N2-2}\mathbf{B}_k & \ddots & \ddots & \sum_{i=0}^{N2-Nu-1}\mathbf{CA}_k^i\mathbf{B}_k \\ \mathbf{CA}_k^{N2-1}\mathbf{B}_k & \mathbf{CA}_k^{N2-2}\mathbf{B}_k & \cdots & \sum_{i=0}^{N2-Nu}\mathbf{CA}_k^i\mathbf{B}_k \end{pmatrix}$$

Notice that the system is realigned in (3). The prediction of the free system response is calculated at time step *k* by the multiplication of the actual matrix $\mathbf{F}_k$ with the actually estimated state $\hat{\boldsymbol{x}}_k$. To realign the prediction to the system's output, not the free system response itself, but only its deviation from the previous prediction is added for every predicted time step to the actual measured system output *Y(k)*. The predicted influence of the actuating values is calculated by the last term in (3). From (2) and (3) the problem can be reformulated to a quadratic term, which can be minimized analytically. This leads to the optimal solution (4). The term obviously only depends on the prediction of the setpoint and the system's predicted output $\hat{Y}_k$. Here the vector $\Delta U_{opt,k}$ holds (N_u*dim_u) entries, for each predicted control step one set of actuating variables. Of these only the first set is applied.

$$\Delta U_{opt,k} = \tfrac{1}{2}\mathbf{H}_{quad,k}^{-1}\mathbf{G}_{quad,k}$$
$$\mathbf{G}_{quad,k} = 2\mathbf{H}_k\mathbf{\Gamma}\left(W_k - \hat{Y}_k\right), \mathbf{H}_{quad,k} = \mathbf{H}_k^T\mathbf{\Gamma}\mathbf{H}_k + \mathbf{\Lambda} \quad (4)$$

Enforcing constraints on the actuating variable or its derivate is only possible by saturating the optimal solution when using the analytic solution (4). Constraints on the controlled variables are not possible at all. The direct consideration of constraints can be realized by using an optimisation routine which minimizes (2) subject to boundaries on ΔU. Such a quadratic solver (QP) is described for example in (Schittkowski 1986). All constraints on the actuating variable and its derivate as well as constraints on the controlled variable can be reformulated in dependency on the optimization variable ΔU analogous to (Maciejowski 2002). To evaluate the improvement of the implementation of a quadratic solver only the actuating variables are bounded as shown in (Fig. 6), as this is also possible with the analytic solution. The benefit therefore only arises from the QP controller's feature to regard constraints in the optimization. The controller was set up with an upper prediction horizon N_2 of 20 simulation steps with a lower prediction horizon N_1 of 1. The prediction horizon N_u was set to 5. $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are the same for both controllers. The actuators seem to be quite fast, see (Fig. 6). But the actuation is only necessary every second revolution, as we operate a four stroke engine. The electromechanical valve train offers the needed response time, as it is actuated once per cycle, see section 2. The visual benchmark is hindered by the noise applied to the system's output, but it demonstrates the controllers' ability to cope with the disturbance,

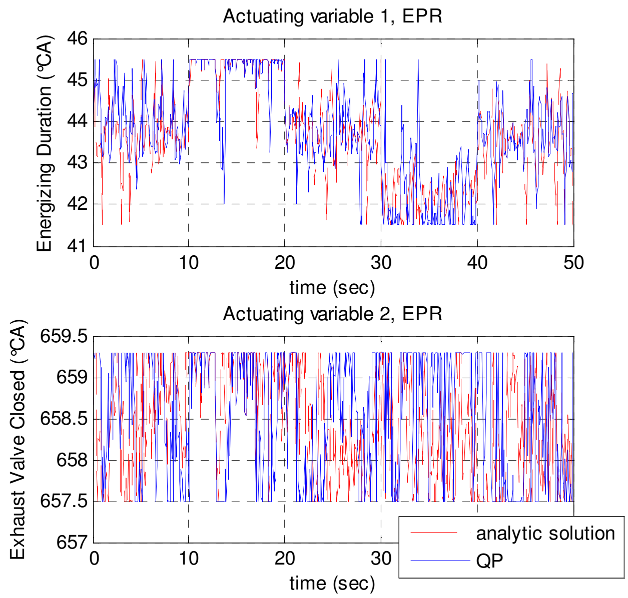

Fig. 6: Bounded actuating variables for the comparison test of QP and analytic solver shown in Fig. 7

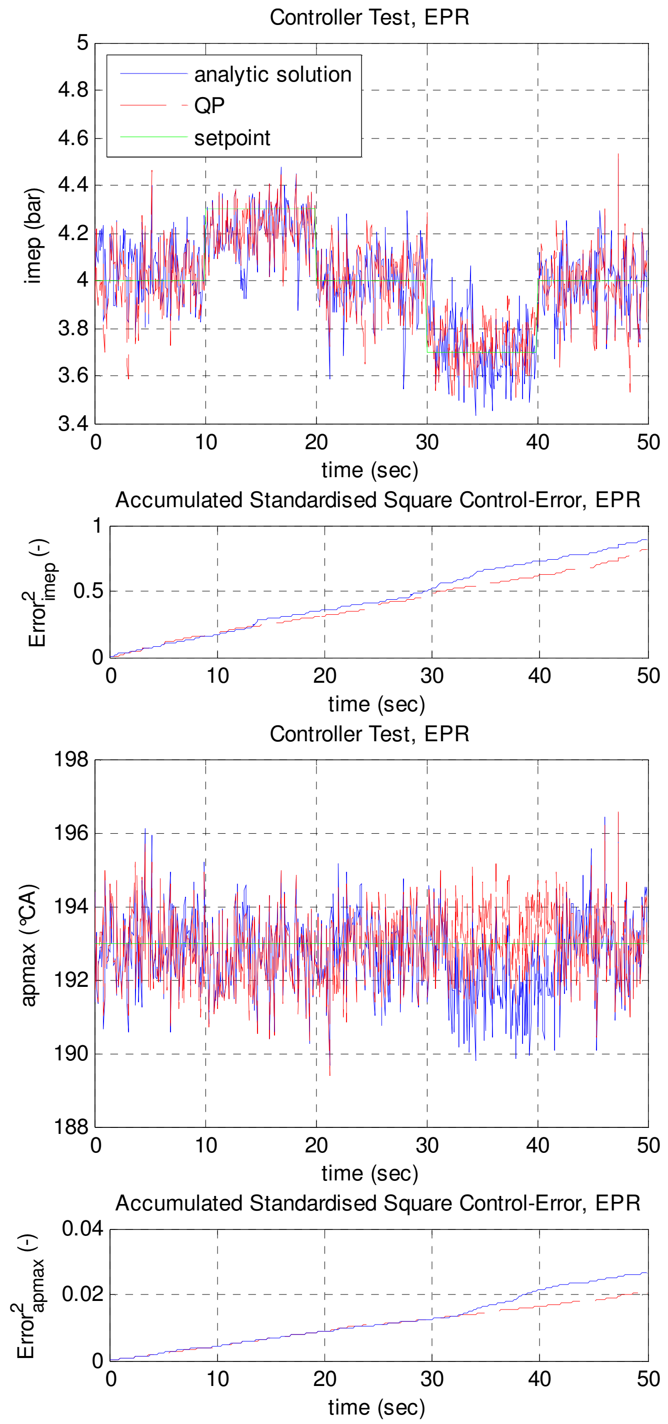

Fig. 7: Model based predictive controllers in closed loop simulation, setpoint step in imep with EPR mode, comparison of quadratic programming and analytic solution

see (Fig. 7). A better measure for the performance is the accumulated square error standardized over the setpoint. Especially the step to a lower load at 30 seconds leads to a better performance of the QP solver especially in following the setpoint for *apmax*. As remarked before the step towards low loads conduces to a less stable engine operation for EPR. The analytic controller is not capable of keeping the location of the maximum pressure *apmax* constant between 30 and 40 seconds simulation time, while the quadratic controller is. The shown test only holds for the actuating variable, not for its derivate. The implementation of constraints on the controlled variable is only possible for the QP controller using (3) to make the constraints dependent on the optimization variable ΔU.

5 CONCLUSIONS

A nonlinear model based predictive controller combined with a neural net observer has been presented in this paper. The benefit of the combination of an extended Kalman filter with the neural net observer has been shown for different operation strategies for a CAI single cylinder engine. By combination of the neural extended Kalman filter with two different MPCs their performance could be evaluated. This benchmark proved the advantage of respecting constraints in the prediction by using quadratic programming solvers. Further research has to show the convenience and closed loop-stability of the proposed concept by application to the real engine.

5.1 Acknowledgement

The presented work is part of the Collaborative Research Centre 686 “Model based control of homogenized low-temperature Combustion” supported by the Deutsche Forschungsgemeinschaft, which is kindly appreciated.

REFERENCES

Bengtsson, J., P. Strandh, R. Johansson, P. Tunestål, and B. Johansson (2006). Hybrid control of homogeneous charge compression ignition (HCCI) engine dynamics *International Journal of Control*, **79**, pp. 422-448.

Colin, G., G. Bloch, A. Charlet (2007), Exact and Linearised Neural Predictive Control: A Turbocharged SI Engine Example, *Journal of Dynamic Systems, Measurement and Control*, **129**, pp. 527-533

Chiang, C., A. G. Stefanopoulou and M. Jancović (2007) Nonlinear Observer-Based Control of Load Transitions in Homogeneous Charge Compression Ignition Engines, *IEEE Transactions on Control System Technology*, **15**(3), pp. 438-448

Hoffmann, K., D. Seebach, S. Pischinger and D. Abel, (2007). Neural Networks for Controlling future low Temperature Combustion Technologies, *3rd IFAC Advanced Fuzzy and Neural Network Workshop, Proceedings,* Valenciennes, France

Hornick, K., M. Stinchcombe and H. White (1989). Multilayer Feedforward Networks are Universal Approximators, *Neural Networks*, **2**, pp.359-366.

Lang, O., W. Salber, J. Hahn, S. Pischinger, K. Hortmann, C. Bücker (2005). Thermodynamical and Mechanical Approach Towards a Variable Valve Train for the Controlled Auto Ignition Combustion Process, *SAE Paper* 2005-01-0762

Maciejowski, J.(2002), *Predictive Control with Constraints*, Prentice Hall, London

Nørgaard, M., O. Ravn, N.K. Poulen and L.K. Hansen, (2000). *Neural Networks for Modelling and Control*, Springer, London

Rossiter, J. A. (2003). *Model-based Predictive Control: A Practical Approach*, CRC Press, Boca Raton

Shaver G., J. Gerdes and M. Roelle, (2004), Physics-based closed-loop control of phasing, peak pressure and work output in HCCI eninges utilizing variable valve actuation, *Proc. American Control Conference*, pp. 150–155, Boston, Massachusetts

Schittkowski, K. (1986), QLD: *A FORTRAN Code for Quadratic Programming, User'sGuide*, Mathematisches Institut, Universität Bayreuth, Germany