

Linköping University Post Print

Observer design and model augmentation for bias compensation with a truck engine application

Erik Höckerdal, Erik Frisk and Lars Eriksson

N.B.: When citing this work, cite the original article.

Original Publication:

Erik Höckerdal, Erik Frisk and Lars Eriksson, Observer design and model augmentation for bias compensation with a truck engine application, 2009, CONTROL ENGINEERING PRACTICE, (17), 3, 408-417.

<http://dx.doi.org/10.1016/j.conengprac.2008.09.004>

Copyright: Elsevier Science B.V., Amsterdam.

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-17160>

Related articles available from Vehicular Systems' Research Group, Linköping University:

<http://www.fs.isy.liu.se/Publications/>

Observer Design and Model Augmentation for Bias Compensation With a Truck Engine Application

Erik Höckerdal^{a,b}, Erik Frisk^a, and Lars Eriksson^a

^a*Department of Electrical Engineering, Linköping University, Sweden, {hockerdal,frisk,larer}@isy.liu.se*

^b*Scania CV AB, Södertälje, Sweden, erik.hockerdal@scania.com*

Abstract

A systematic design method for reducing bias in observers is developed. The method utilizes an observable default model of the system together with measurement data from the real system and estimates a model augmentation. The augmented model is then used to design an observer which reduces the estimation bias compared to an observer based on the default model. Three main results are a characterization of possible augmentations from observability perspectives, a parameterization of the augmentations from the method, and a robustness analysis of the proposed augmentation estimation method. The method is applied to a truck engine where the resulting augmented observer reduces the estimation bias by 50% in a European Transient Cycle.

Key words: bias compensation, EKF, non-linear, observer

1. Introduction

In many application areas there are quantities that are important for control and diagnostics but that are not measured due to for example difficulties with the measurement methods or high costs of the sensors. This has made estimation an important and active research area, which is especially true in the automotive area where cost is important, see (Lino et al., 2008; García-Nieto et al., 2008; Andersson and Eriksson, 2004) for some examples.

In all model-based control or diagnosis systems, the performance of the system is directly dependent on the accuracy of the model. In addition, modeling is time consuming and, even if much time is spent on physical modeling, there will always be errors in the model. This is especially true if there are constraints on the model complexity, as is the case in most real time systems. Another scenario is that a model developed for some purpose, e.g. control, exists but needs improvements before it can be used for other purposes, for example diagnosis.

In many applications, like for example engine control and engine diagnosis, it is crucial to have unbiased estimates. In model based diagnosis, the true system is often monitored by comparing measured signals to estimated signals. If the magnitude of the difference, the residual, is above a certain limit a decision that something is wrong is made. In engine control, one objective is to maximize torque output while

keeping the emissions below legislated levels and the fuel consumption as low as possible. For diesel engines this is especially hard since the control system does normally not have any feedback information from a λ - or NO_x -sensor and have to rely on estimated signals instead (Wang, 2008). In both cases, biased estimates impairs the performance.

The objective of this work is to develop a systematic method for reducing estimation bias in observers without involving further modeling efforts. This work is an extension of preliminary results in (Höckerdal et al., 2008b) and the main extensions are a theoretical characterization of all solutions and additional method evaluations, including a robustness analysis with respect to measurement noise and model uncertainty.

The method utilizes an observable model and measurement data from the true system. The given model, referred to as the default model, and the measured inputs and outputs from the true system are used to estimate a suitable model augmentation. Then, the augmented model is used to design an observer that is shown to give estimates with reduced bias compared to an observer based on the default model. Three approaches for estimating a bias compensating augmentation are developed and evaluated with respect to measurement noise and model errors. Key results are a theoretical characterization of all possible augmentations from observability perspectives and a parametrization of the estimated augmentations. Finally the method is evalu-

ated on a non-linear diesel engine model with experimental data from an engine test cell.

2. Problem formulation

Previous experience at Scania CV AB of state estimation based on an existing state-space model of a truck engine reveals that the model captures dynamic behavior reasonably well but suffers from stationary errors (Höckerdal et al., 2008a). Designing an observer based on this model results in biased estimates. How to reduce the bias in a systematic manner is the topic of this paper.

The starting point is an existing model, referred to as the default model, that is provided in state-space form

$$\dot{x} = f(x, u) \quad (1a)$$

$$y = h(x), \quad (1b)$$

where x is the state-vector, u the known control inputs, y the measurement vector, and f and h are non-linear functions.

The objective is to find a systematic way to design an observer that gives an unbiased estimate of either the complete state x or a function of the state $z = g(x)$. This should be done even though the default model is subject to significant bias errors. A direct approach to compensate for constant, or slowly varying, biases is to augment the default model with bias variables q as

$$\dot{x} = \tilde{f}(x, u, q) \quad (2a)$$

$$\dot{q} = 0 \quad (2b)$$

$$y = \tilde{h}(x, q), \quad (2c)$$

and design the observer using this augmented model. If the augmentation captures the true modeling errors and the augmented system is observable, the observer estimates is made unbiased. An obvious question is then how to introduce the bias variable q in the model equations. One way could be through process knowledge, which have been successfully applied in (Andersson and Eriksson, 2004; Tseng and Cheng, 1999). However, in this paper an estimation procedure based on available measurement data is proposed.

Besides the natural restriction, that the augmented model (2) is observable, it is also desirable not to introduce more extra bias states than necessary. It is therefore desirable to find a bias vector q with as low dimension as possible that manages to reduce the bias. Another reason for finding a low dimensional bias is that, since the model often is a first-principles physical model, bias in multiple states may be explained by one underlying bias affecting all these states. For example, bias in two pressures can originate from a bias in the mass flow between the two volumes or an incorrect modeling of energy conservation can give rise to bias in several states connected to the energy. However, the bias is necessarily not the same in the entire operating region of the system and may vary between operating points. This is part of the reason for introducing

the bias as new states, rather than just a parameter, which allows some tracking ability of the bias.

In model (1) there are two natural ways to introduce biases, in the dynamic equation (1a) or in the measurement equation (1b). In the truck engine application the sensors, intake and exhaust manifold pressures and turbine speed, are considered more reliable than the model and the bias augmentation is therefore introduced in the dynamic equations according to

$$\dot{x} = f(x - A_q q, u) \quad (3a)$$

$$\dot{q} = 0 \quad (3b)$$

$$y = h(x), \quad (3c)$$

where a stationary point of the system is moved by $A_q q$. The matrix A_q is thus a description of how the underlying bias variable q influences the stationary value of the state variable x . The model (3) will be referred to as the augmented model. It is worth mentioning that although the result in this paper focuses on biases in the dynamic equation, it is straightforward to modify the approach to also cover sensor biases.

2.1. Problem and paper outline

Based on the discussion above, the problem studied in the sections to follow can now be stated as: Given an observable default model (1) and available measurement data, find a low order bias augmented model (3) and design an observer that estimates x with reduced bias compared to using the default model. The observer should also be implementable in an Engine Control Unit (ECU).

To solve the problems, some issues need to be addressed. First, which matrices A_q are possible at all? All are not possible since it is required that the augmented system is observable and a characterization of possible augmentations is derived in Section 4. Among these possible bias augmentations, which should be used? Section 5 describes three approaches for how to estimate a, for bias compensation, suitable low order A_q based on measurement data.

Section 6 presents two examples of the proposed estimator design methodology applied to a Scania diesel engine using simulated and real measurement data respectively.

3. Discretization and Linearization

As a first step, the nonlinear augmented model (3) is transformed to a linearized time discrete model. A reason for the discretization is the demand on the implementation, which will be done in the ECU as a time discrete system. Here, a simple Euler forward discretization with step size T_s seconds is used. Note that observability does not depend on the choice of discretization method, since as long as T_s is chosen small enough the results are valid also for, e.g. zero-order-hold (Kalman et al., 1963).

One objective of the paper is to find a suitable A_q such that (3) is locally observable and to be able to use simple

observability conditions, the observability analysis is here performed on a linearization of the non-linear model (3). Of course, non-linear observability is not guaranteed from observability of the linearization. Nevertheless, observability of a linearization in a stationary point is a sufficient condition for local observability of the non-linear system, see Theorem 6.4 in (Lee and Markus, 1968). Even though observability is not strictly guaranteed, e.g. in transient mode when moving between operating points, the referred result gives theoretical support for using the linearized system in the observability analysis. Thus, when analyzing (3) the following model will be used

$$\begin{pmatrix} x_{t+1} \\ q_{t+1} \end{pmatrix} = \begin{pmatrix} I + T_s A & -T_s A A_q \\ 0 & I \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix} + \begin{pmatrix} T_s B \\ 0 \end{pmatrix} u_t \quad (4a)$$

$$y_t = \begin{pmatrix} C & 0 \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix}, \quad (4b)$$

where

$$A = \left. \frac{\partial f}{\partial x} \right|_{\substack{x=x_0 \\ u=u_0}}, \quad B = \left. \frac{\partial f}{\partial u} \right|_{\substack{x=x_0 \\ u=u_0}}, \quad \text{and} \quad C = \left. \frac{\partial h}{\partial x} \right|_{\substack{x=x_0 \\ u=u_0}}.$$

In the following, $I + T_s A$ is substituted for F to increase readability and (4) becomes

$$\begin{pmatrix} x_{t+1} \\ q_{t+1} \end{pmatrix} = \begin{pmatrix} F & -(F - I)A_q \\ 0 & I \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix} + \begin{pmatrix} T_s B \\ 0 \end{pmatrix} u_t \quad (5a)$$

$$y_t = \begin{pmatrix} C & 0 \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix}. \quad (5b)$$

4. Possible augmentations

Augmenting a model with more states may affect the observability of the model. Since the purpose of the augmented model is to use it for estimation, observability has to be maintained also after the augmentation. To find which augmentations that are possible an observability investigation of the augmented model is performed. The aim is to derive a necessary and sufficient condition on A_q such that the augmented model is observable. The observability criterion used in the analysis is known as the Popov-Belevitch-Hautus (PBH)-test (Kailath, 1980).

Similar results can be found in (Bembenek et al., 1998), which also includes a discussion regarding the observability results, similar to the short discussion in the end of this section.

Theorem 4.1 *A pair (C, F) is observable if and only if*

$$\begin{pmatrix} C \\ \lambda I - F \end{pmatrix}$$

has full column rank $\forall \lambda \in \mathbb{C}$.

Now, using Theorem 4.1 and the assumption that the default model is observable the main result of this section can be formulated as

Theorem 4.2 *Assume that (C, F) in (5) is an observable pair then the augmented system (5) is observable if and only if*

$$\text{Ker}((F - I)(A_q \ N_C)) = \{0\},$$

where the columns of N_C span $\text{Ker } C$.

Note that this is equivalent to

$$(F - I)(A_q \ N_C),$$

having full column rank.

PROOF. See Appendix B. □

This means that the space spanned by the columns in A_q can lie neither in $\text{Ker } C$ nor in $\text{Ker}(F - I)$ for the augmented model to be observable. These interpretations of the rank condition can be understood by analyzing the two requirements separately. First, the requirement that A_q can not lie in $\text{Ker } C$ is easily seen by studying a linear example.

Example 1. Starting with a linear model with a stationary bias

$$\begin{aligned} x_{t+1} &= Fx_t - (F - I)A_q q_t \\ q_{t+1} &= q_t \\ y_t &= Cx_t \end{aligned} \quad (6a)$$

and performing a change of variables, $z_t = x_t - A_q q_t$, gives

$$\begin{aligned} z_{t+1} &= Fx_t - (F - I)A_q q_t - A_q q_t = Fz_t \\ q_{t+1} &= q_t \\ y_t &= Cz_t + CA_q q_t, \end{aligned} \quad (6b)$$

which shows that columns of A_q in $\text{Ker } C$ are not observable. ◇

Second, a non-empty $\text{Ker}(F - I)$ implies that the system contains pure integrators, and a bias in $\text{Ker}(F - I)$ is not distinguishable from an unknown initialization of the integrator and is therefore not observable.

A closer look at the requirement that $(F - I)(A_q \ N_C)$ has to have full column rank conveys some other interesting results. First, assuming full column rank of $(F - I)$, it is easily seen that the number of augmented states n_q never can exceed the number of linearly independent measurement signals n_y since

$$\begin{aligned} \text{rank}(F - I)(A_q \ N_C) &= \text{rank}(A_q \ N_C) \\ &\leq \text{rank } A_q + \text{rank } N_C = n_q + n_x - n_y \leq n_x, \end{aligned} \quad (7)$$

i.e. $n_q \leq n_y$. Second, again imagine that $(F - I)$ has full rank which means that the model does not have any pure integrators, then the full column rank condition on $(F - I)(A_q \ N_C)$ reduces to requiring full column rank of $(A_q \ N_C)$ or, equivalently, full column rank of the product CA_q . Now if C has one or several zero columns, then CA_q will not contain any information from those rows in A_q corresponding to zero columns in C . That is, those rows in A_q

that correspond to zero columns in C will not contribute to the observability, see the following example.

Example 2. Illustration of possible augmentations for a default model without pure integrators and

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Let \times denote a non-zero element, then some possible augmentations are

$$A_q^1 = \begin{pmatrix} \times & 0 \\ 0 & \times \\ 0 & 0 \end{pmatrix}, \text{ and } A_q^2 = \begin{pmatrix} 0 & \times \\ \times & 0 \\ 0 & 0 \end{pmatrix}$$

since

$$CA_q^1 = \begin{pmatrix} \times & 0 \\ 0 & \times \end{pmatrix}, \text{ and } CA_q^2 = \begin{pmatrix} 0 & \times \\ \times & 0 \end{pmatrix},$$

which have full column rank. While an augmentation

$$A_q^3 = \begin{pmatrix} \times & 0 \\ 0 & 0 \\ 0 & \times \end{pmatrix} \text{ is not possible since } CA_q^3 = \begin{pmatrix} \times & 0 \\ 0 & 0 \end{pmatrix}$$

does not have full column rank. \diamond

5. Augmentation estimation

Now that all possible model augmentations have been characterized, the next question is how to find a suitable augmentation, using measured data from the real system, that fulfills the requirements derived in Section 4. The proposed augmentation estimation procedure is divided into two steps, i) from measured data estimate samples of the bias and ii) compute a basis for the bias samples. Three approaches for how to conduct the first step are developed. In the second step a low order augmentation is computed by performing an Singular Value Decomposition (SVD) on selected samples of the bias found in step one.

5.1. Bias estimation

The first step in the estimation of a low order model augmentation deals with estimating the bias, i.e. collect samples of the bias $\beta_t = A_q q_t$. The first approach is quite simple and its main purpose is to illustrate the basic ideas for the estimation of bias samples, whereas the second and third approach are applicable to more general systems. Since the method aims at reducing bias in stationary operating points only stationary behavior and data is studied.

5.1.1. Approach 1

The first approach utilizes the discretized linearization directly and the assumptions that all states are measured,

i.e. C has full column rank, and that the system does not have any pure integrators, i.e. $(I - F_t)$ has full column rank. The linearized and time discretized augmented model is

$$\begin{aligned} x_{t+1} &= F_t x_t + (I - F_t) A_q q_t + T_s B_t u_t \\ y_t &= C_t x_t. \end{aligned}$$

Due to the full column rank assumptions on C and $(I - F_t)$ it is possible to invert the measurement equation and insert the resulting x in the dynamic equation. This gives that

$$\beta_t = A_q q_t = (I - F_t)^{-1} (C_{t+1}^\dagger y_{t+1} - F_t C_t^\dagger y_t - T_s B_t u_t),$$

where \dagger denotes the pseudo inverse.

This approach has three evident flaws, it requires a full column rank C and $(I - F_t)$ and, since no filtering of the measurements is involved, it is sensitive to low Signal to Noise Ratio (SNR).

Therefore two other approaches are proposed for estimating bias samples. Common for both these approaches are that they utilize the residuals from an observer and the assumption that the true bias enters the model according to Equation (3). The fact that they are based on observers makes them less sensitive to low SNR and imply that they do not require full column rank C to work. The first employ an observer based on the default model and the bias samples are computed by inverting the observer system. The second employ a fully augmented model fulfilling the observability requirements developed in Section 4.

5.1.2. Approach 2

The second approach uses the residuals originating from an observer based on the default model. Here, the observer is an Extended Kalman Filter (EKF) (Kailath et al., 2000), where the noise covariance matrices Q and R are design parameters tuned by the user. Of course, other observer designs are equally possible but here an EKF is used. Let K_t be the EKF feedback gain then the estimation error becomes,

$$\begin{aligned} e_{t+1} &= x_{t+1} - \hat{x}_{t+1|t+1} \\ &= F_t x_t + (I - F_t) A_q q + T_s B_t u_t - \\ &\quad (F_t \hat{x}_{t|t} + T_s B_t u_t + K_t (y_{t+1} - C_t F_t \hat{x}_{t|t} - C_t T_s B_t u_t)) \\ &= \{y_{t+1} = C_t F_t x_t + C_t (I - F_t) A_q q + C_t T_s B_t u_t\} \\ &= (F_t - K_t C_t F_t) e_t + (I - K_t C_t) (I - F_t) A_q q. \end{aligned} \quad (9)$$

Equation (9) can not be used directly since the state estimation error is not known. Therefore, the output error

$$r_t = y_t - \hat{y}_{t|t} = C_t (x_t - \hat{x}_{t|t}) = C_t e_t, \quad (10)$$

is used for estimating the bias.

As previously stated, solely stationary parts of the residuals are involved in the bias estimation. It would be possible to use also dynamic parts of the residuals and a dynamic inverse. The reason for not utilizing these here is to prevent dynamic estimation errors from affecting the estimation of the constant or slowly varying bias.

Now, utilizing that only stationary data is considered, (9) and (10) can be combined resulting in

$$\begin{aligned}
r_{\text{stat}} &= C_{\text{stat}} e_{\text{stat}} \\
&= C_{\text{stat}} (I - F_{\text{stat}} + K_{\text{stat}} C_{\text{stat}} F_{\text{stat}})^{-1} \times \\
&\quad (I - K_{\text{stat}} C_{\text{stat}}) (I - F_{\text{stat}}) A_q q_{\text{stat}}
\end{aligned}$$

and the bias can be estimated as

$$\beta_t = A_q q_t = (C_{\text{stat}} (I - F_{\text{stat}} + K_{\text{stat}} C_{\text{stat}} F_{\text{stat}})^{-1} \times (I - K_{\text{stat}} C_{\text{stat}}) (I - F_{\text{stat}}))^{\dagger} r_t. \quad (11)$$

5.1.3. Approach 3

An alternative to Approach 2 for finding β_t is to augment the default model with as many extra states as possible. According to Theorem 4.2 the requirement on A_q is that, $(F - I)(A_q N_C)$ has to have full column rank. This means that A_q can have a maximum of n_y columns. These columns have to be linearly independent of the columns of N_C and can not lie in $\text{Ker}(F - I)$. One way to construct such an augmentation is to use C^{\dagger} and leave out those columns that become zero when multiplied by $(F - I)$ from the left. Then run the observer based on the augmented model, estimating both \hat{x} and \hat{q} , and assemble $\beta_t = C^{\dagger} \hat{q}_t$.

An advantage with this approach is that no inversions as those in (11) are needed. A disadvantage though is that since a fully augmented model is used the order of the observer might be unnecessarily high.

5.2. Augmentation computation

As stated in the problem formulation in Section 2, the bias is necessarily not the same in the entire operating region of the system. This makes it important to collect samples of the bias from stationary operating points selected such that the entire operating region is covered. From the first step of the proposed procedure, bias samples are collected according to this. Based on the discussion of only a few underlying biases affecting several states in Section 2, the task of step two is to find a low order basis spanning the space in which these bias samples are located.

To start with bias samples from N stationary operating points are assembled

$$\bar{\beta}^{n_x \times N} = \begin{pmatrix} \beta^1 & \dots & \beta^N \end{pmatrix},$$

Then the SVD of $\bar{\beta}$ is computed,

$$\bar{\beta} = U \Sigma V^*,$$

where U contains orthogonal vectors spanning the space in which the bias moves and Σ the corresponding singular values. The singular values in Σ are ordered in non-increasing order which means that the far left columns of U , corresponding to large singular values, represent the most dominating directions along which the bias moves. Therefore the dimension of q can be found by comparing the singular values in Σ , and picking the most significant ones. Then the corresponding columns of U are used to assemble \hat{A}_q .

This way of computing an augmentation from bias samples is optimal with respect to the Frobenius norm.

5.3. Properties of the estimated augmentation

According to the discussion in the end of Section 4, the properties of C place restrictions on which A_q 's that are possible to find. The conclusion of that discussion is that rows in A_q corresponding to zero columns in C become zero in the estimation step. However, a more thorough analysis of the three bias estimation approaches shows that more can be said.

Theorem 5.1 *Assume that the observer gain, K , is chosen such that the observer is strictly stable and does not have any poles in the origin. Then, in absence of noise, the bias samples are spanned by the rows of C and can thereby be written as*

$$\beta_t = C^T \Gamma.$$

PROOF. See Appendix B. □

Note that Theorem 5.1 holds for the pseudo inverse and is not generally true for an arbitrary left inverse.

As a consequence, the observer based on an estimated augmentation may not be able to reduce the bias in the estimates to acceptable levels. This problem can be circumvented in, for example one of the two following ways. The first is for an engineer to design an A_q not possible to find through estimation, for example through knowledge of the underlying physics. The second is to, during the design phase, add extra sensors to the true system to acquire a full column rank C which enables estimation of all rows in A_q . When utilizing this possibility one must be cautious and check the observability of the augmented system that in the end will not rely on the additional sensors used for estimating \hat{A}_q . That is, check the column rank of $(F - I)(\hat{A}_q N_C)$, and in case of column rank deficiency remove those columns in \hat{A}_q causing rank deficiency. Since SVD is used, the columns in \hat{A}_q are arranged in non-increasing significance order which makes it appropriate to remove the columns in \hat{A}_q starting from the right to get an augmentation that is observable.

The example below illustrates the remarks regarding the effects that properties of C have on the augmentation estimation.

Example 3. Consider a true system with

$$F = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix}, \text{ and } C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \end{pmatrix}$$

and a true bias,

$$A_q = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Then the estimate of A_q , according to Theorem 5.1, will be

$$\hat{A}_q = \begin{pmatrix} 1 \\ 2 \times 7/5 \\ 1 \times 7/5 \end{pmatrix},$$

where the factor 7/5 comes from minimizing

$$\left\| A_q - C^T \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} \right\| = (1 - \Gamma_1)^2 + (2 - 2\Gamma_2)^2 + (3 - \Gamma_2)^2$$

with respect to Γ_1 and Γ_2 .

That is, because of the structure of C the top element in \hat{A}_q will be correct while the bottom two elements will not. \diamond

5.4. Approach evaluation

Two main approaches, approaches 2 and 3, for estimating the bias have been proposed. It is important to understand how these approaches perform under varying operating conditions and model uncertainty. Therefore, the approaches are evaluated with respect to robustness against model errors and robustness to changes in noise levels. This is done by introducing noise and modeling errors in a non-linear simulation model of a Scania diesel engine with exhaust gas recirculation (EGR) and variable geometry turbine (VGT), and performing Monte Carlo simulations. In the simulations, a one-dimensional q is also introduced, i.e. A_q is a vector with three elements.

Modeling errors can be introduced in many ways and it is difficult to obtain a comprehensive evaluation of robustness properties of a non-linear method. Therefore, a more pragmatic approach is adopted. First, model errors are introduced by manipulating physical constants in the simulation model and thus making the simulation model, that generates the observations, different from the default model used for designing the observer. Another way model errors are introduced is by pre-multiplying the vector field f in (1a) by a slowly varying sinusoid, i.e. the simulation is done with $\tilde{f}(x, u)$ defined as $\tilde{f}(x, u) = (1 + \gamma \sin(\Lambda t))f(x, u)$, where Λ is the model error frequency, and γ is a small number varied between 0.1 and 0.5. Doing Monte Carlo simulations with such model errors introduced reveal that both approaches react similarly to the model errors with respect to degraded performance in bias estimation and variance in the estimation. No certain conclusion can be drawn concerning which approach is more robust against modeling errors and the overall picture is that both approaches have similar graceful performance degradation with increased modeling errors.

Examining the effect of measurement noise is done by introducing white Gaussian noise with different noise levels in the simulation model and estimating the effect on the augmentation estimation by computing the variance in the β^i 's. In Figure 1 the effect of increased measurement-noise level on the variance in the estimated β^i 's is shown. It is seen that Approach 3 is significantly less sensitive to measurement noise and therefore preferable when estimating an augmentation.

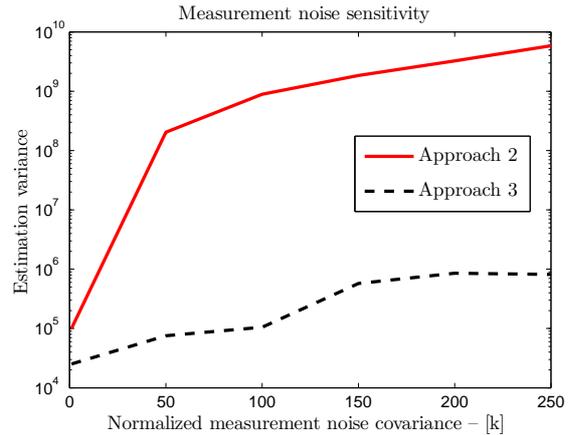


Fig. 1. Estimation variance for bias estimation approaches 2 and 3 with different measurement noise $- k \cdot \mathcal{N}(0, R)$.

5.5. Method summary

The procedure can be summarized in three steps.

Step 1 - Linearize and discretize the model if necessary. Normally, the default model is a non-linear time continuous model such as (1) and has to be linearized and discretized.

Step 2 - Find an appropriate augmentation, A_q , and compile an augmented model (4). Here the designer has a choice, either to estimate an augmentation from measured data, introduce an augmentation found in some other way, or to combine an estimated augmentation with one found through system knowledge.

The estimation procedure contains two steps, i) estimation of bias samples utilizing one of the three approaches presented in Section 5.1, ii) compute a basis for the bias samples using SVD.

With good knowledge of the system, the designer might have some idea of what is causing the bias in the estimates and can choose an appropriate A_q .

To combine an augmentation found through process knowledge with one found through estimation can be desirable if some model deficiencies are known but does not manage to achieve satisfactory bias reduction. In this case the estimation approach can be applied to the, by the engineer, partly augmented model to find an additional augmentation that captures the remaining dominating bias.

Step 3 - Design an observer based on the augmented model (3) and the A_q found in Step 2. In this paper, an Extended Kalman Filter is used but any non-linear observer design methodology is possible.

6. Experimental evaluation

To evaluate the method experiments are performed using a non-linear model of a heavy-duty truck engine. The experiments consist of a simulation study of the non-linear model, and evaluation of the method on measurement data

from an engine test cell.

The non-linear model of the Diesel engine has three states: p_{im} , p_{em} , and n_{trb} , that represent intake and exhaust manifold pressures, and turbine speed respectively. See Appendix A for more information about the engine and model. In the second experiment, real data from the engine is used together with the engine model to illustrate the properties of the proposed approach in a real application. In both experiments the stationary parts of the data, used in the augmentation estimation, are separated out through visual inspection and estimation Approach 3 is chosen to estimate the bias.

6.1. Evaluation using simulated data

The objective of the first experiment is to illustrate how the approach, which is based on linearization procedures, performs when fed with data from a non-linear simulation model. Thus, synthetic data is created where known biases are introduced in the simulation. The method is then applied to show how biases in non-linear systems can be estimated.

The introduced bias is represented by a matrix

$$A_q = \begin{pmatrix} 1 & -2 \\ 2 & 1 \\ 0 & 0.2 \end{pmatrix},$$

and two slowly varying biases q_1 and q_2 . This A_q means that there are two independent biases affecting the model states which varies between approximately 0 and 10% of the state values. The default system has linear measurement equations where $y_1 = p_{im}$ and $y_2 = n_{trb}$. However, according to the discussion in Section 5.3, an augmentation as the one introduced in this example can not be estimated without a direct connection between p_{em} and y . Therefore the measurement equation is extended with an extra sensor for p_{em} for the augmentation estimation. Note that this extra sensor is not used for feedback neither in the observer based on the default model nor in the observer based on the augmented observer. This reflects the situation that a lab environment or development system may be equipped with extra sensors to achieve a better augmentation estimation.

The observer based on the default model is referred to as the default observer while the observer based on the augmented model is referred to as the augmented observer. Both observers only use the p_{im} and p_{em} measurements. To make the simulation more realistic, white system and measurement noise are added in the creation of the synthetic data.

Using the simulated data and the default model, the augmentation estimation results in

$$\Sigma \approx \begin{pmatrix} 5.0259 & 0 & 0 \\ 0 & 4.8669 & 0 \\ 0 & 0 & 0.0024 \end{pmatrix} 10^5,$$

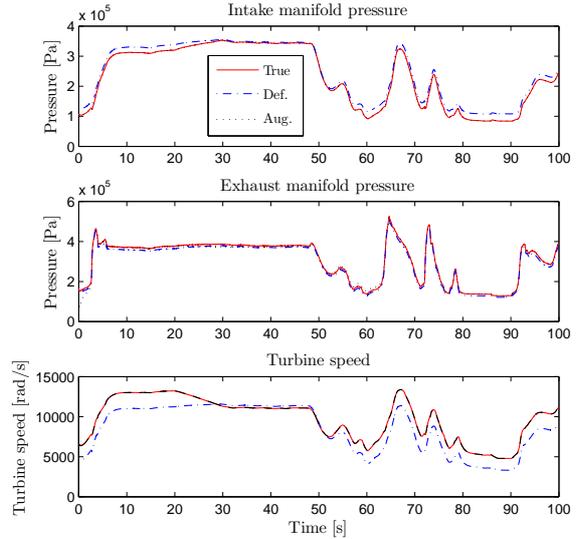


Fig. 2. True states and estimated states using default and augmented observer in the simulation study.

and

$$U \approx \begin{pmatrix} -0.8295 & -0.5527 & 0.0800 \\ 0.5515 & -0.8233 & -0.0388 \\ 0.0881 & 0.0123 & 0.9960 \end{pmatrix},$$

where Σ indicates that there are two slowly varying biases present. Hence, \hat{A}_q is estimated using the first two columns of U .

At a first look \hat{A}_q does not appear similar to A_q . However, the crucial fact is that the columns of \hat{A}_q and A_q span, approximately, the same space. A closer look reveals that the elements in the bottom row is significantly smaller than the other elements, and that the factor between row one and two are approximately 2. That is, the only thing that differs between A_q and \hat{A}_q is a scaling.

The objective was not only to estimate the bias, but rather to obtain an observer that compensated for the model bias. Thus, an observer is created using EKF methodology for a model augmented according to the estimated \hat{A}_q . The performance is compared to the default observer. The state estimates are presented in Figure 2 together with the true states. It is easily seen that the augmented observer estimates p_{im} and n_{trb} better than the default observer. To obtain a better view on observer performance, the estimation errors are plotted in Figure 3. Here it is clear that all three state estimates become better with the augmented observer than with the default observer.

The conclusion of this small simulation example is that the approach managed to get a good enough estimate of a bias in a non-linear model to improve the state estimates.

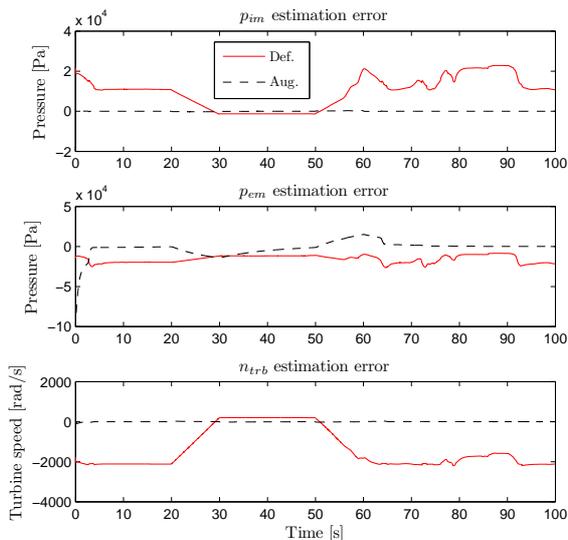


Fig. 3. Estimation errors using default and augmented observer in the simulation study.

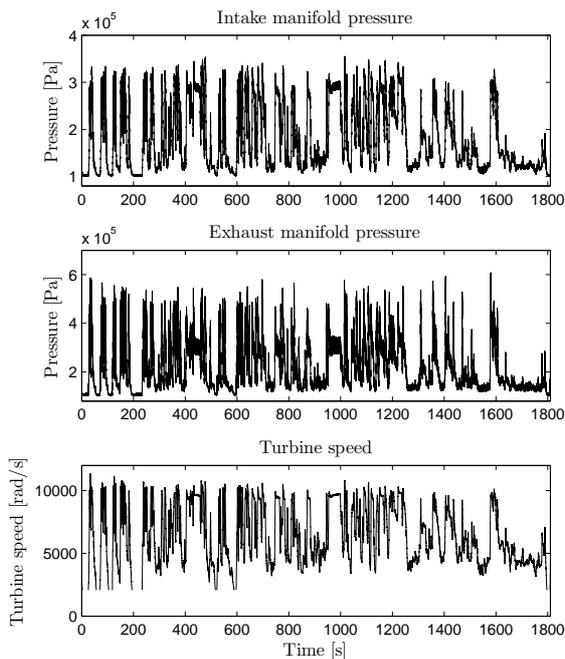


Fig. 4. Measurements of p_{im} , p_{em} , and n_{trb} from the ETC used in the experimental evaluation. Note that turbine speeds below approximately 2100 [rad/s] are missing. This is due to the limited measurement range of the turbine speed sensor.

6.2. Two experimental evaluations

The experimental data described in Appendix A is used to evaluate the augmentation estimation and observer performance. The true states are approximated by non-causal, zero-phase, low-pass filtered measurements, where the filter has a cut off frequency of 2 Hz, see Figure 4. Note that

Table 1

Data from observers None – default observer, H^\dagger – fully augmented observer, and \hat{A}_q – observer using reduced dimension augmentation found using augmentation estimation approach 3. All observers use feedback from all states.

	Max abs. error			Mean error		
	None	H^\dagger	\hat{A}_q	None	H^\dagger	\hat{A}_q
p_{im} [Pa]	5459	6840	6599	-985	11	37
p_{em} [Pa]	14411	14277	14278	443	86	132
n_{trb} [rad/s]	0.8	0.7	0.6	0.005	-0.003	-0.007

parts of the turbine speed data is missing, which is due to the fact that the measuring range of the turbine speed sensor is limited, speeds below 20000 [rpm], or approximately 2100 [rad/s] can not be measured.

Based on the measurement data, an augmentation is estimated using data from two stationary operating points in the European transient cycle (ETC) of about 1000 samples each. All states are measured and the augmentation estimation results in

$$\Sigma \approx 10^5 \begin{pmatrix} 5.3230 & 0 & 0 \\ 0 & 0.3739 & 0 \\ 0 & 0 & 0.0044 \end{pmatrix},$$

and

$$U \approx \begin{pmatrix} -0.2610 & 0.9650 & -0.0249 \\ -0.9648 & -0.2671 & -0.0274 \\ -0.0329 & 0.0169 & 0.9993 \end{pmatrix}, \quad (12)$$

where Σ indicates that there is one dominant slowly varying bias present. Hence, \hat{A}_q is selected to be only the first column of U .

6.2.1. Reduced augmentation order

In this system it is possible to augment the system with three extra states and still have an observable system if all states are measured. One interesting question is if the proposed method that estimates a lower dimension augmentation can still capture most of the bias. Therefore, three observers are designed: the default observer, a fully augmented observer, and a one dimensional augmentation observer.

The aim of this comparison is thus to conclude whether the proposed method works, and is performed by analyzing the estimation errors from the three observers. The resulting Probability Density Functions (PDF) of the estimation errors are shown in Figure 5 and mean and maximum absolute errors for the entire ETC are presented in Table 1. From the data it is clear that the default observer has a bias and that the augmented observers reduce the bias. Now comparing the two augmented observers it is seen that the observer with only a one dimensional augmentation delivers close to the same reduction in bias as the fully augmented observer. This is a clear illustration that the method succeeds in finding the dominant bias in the model.

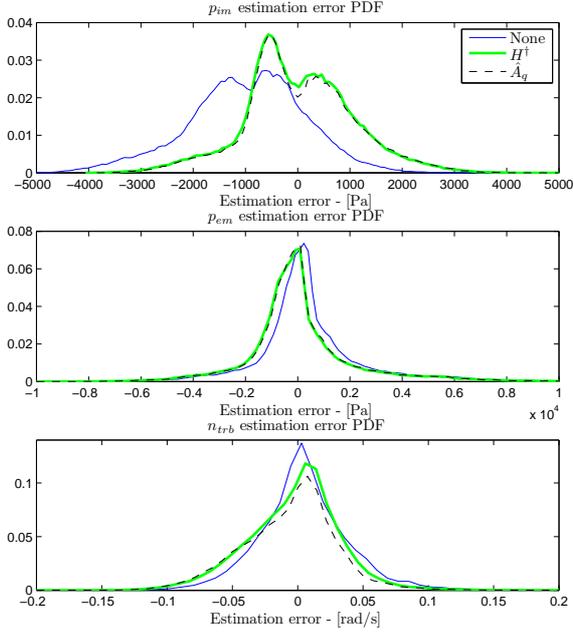


Fig. 5. Probability density functions for three observers: None – default observer, H^\dagger – observer augmented with three states, and \hat{A}_q – observer augmented with one state and the estimated \hat{A}_q .

Table 2

Static data from observers None – default observer, \bar{A}_q – observer using augmentation estimated using only measurements of p_{im} , and n_{trb} , and \hat{A}_q – observer using augmentation estimated using measurements of all states. All observers use feedback from p_{im} , and p_{em} only.

	Max abs. error			Mean error		
	None	\bar{A}_q	\hat{A}_q	None	\bar{A}_q	\hat{A}_q
p_{im} [Pa]	4191	3650	3641	-622	-80	-176
p_{em} [Pa]	58758	58197	51322	6810	6328	-678
n_{trb} [rad/s]	0.1	0.1	0.1	0.02	0.004	0.006

Table 3

Dynamic data from observers None – default observer, \bar{A}_q – observer using augmentation estimated using only measurements of p_{im} , and n_{trb} , and \hat{A}_q – observer using augmentation estimated using measurements of all states. All observers use feedback from p_{im} , and p_{em} only.

	Max abs. error			Mean error		
	None	\bar{A}_q	\hat{A}_q	None	\bar{A}_q	\hat{A}_q
p_{im} [Pa]	5748	6828	6316	-533	2	-34
p_{em} [Pa]	180279	177982	174486	16604	16479	8922
n_{trb} [rad/s]	0.9	0.6	0.5	0.02	0.0007	-0.001

6.2.2. Benefits of additional sensor during design

Another interesting question is what can be achieved by allowing extra sensors, compared to what is used in the final observer, while estimating an augmentation. The application chosen is to estimate the exhaust manifold pressure with reduced bias compared to a default observer without having a sensor measuring it. That is, design an observer

for p_{em} using feedback from p_{im} and n_{trb} . The analysis is performed by comparing the estimates from two observers; one based on the augmentation

$$\hat{A}_q = \begin{pmatrix} -0.2610 & -0.9648 & -0.0329 \end{pmatrix}^T$$

estimated using measurements of p_{im} , p_{em} , and n_{trb} , i.e. column one in (12), and another based on an augmentation

$$\bar{A}_q = \begin{pmatrix} -0.9864 & 0 & -0.1644 \end{pmatrix}^T$$

estimated using measurements of p_{im} , and n_{trb} only.

The two augmented observers are compared to the default observer and the results are shown in Figure 6 and Table 2 and 3. Figure 6 shows the probability density function for the estimation errors for the default observer, the observer based on the model augmented with \hat{A}_q , and the observer based on the model augmented with \bar{A}_q . It is seen that both augmented observers reduce the mean of the bias for p_{im} and n_{trb} compared to the default observer and that the observer based on the model augmented with \hat{A}_q significantly reduces also the bias in p_{em} . Table 2 and 3 show the mean and maximum absolute estimation errors for selected stationary parts of an ETC and for the entire ETC respectively. In both tables it is obvious that the observer based on a model augmented with \hat{A}_q significantly reduces the estimation bias. The mean error is reduced by approximately 50% during an entire ETC and by approximately 90% for selected stationary parts, while the maximum absolute errors are almost unaffected. These, quite large, differences in the different measures are all explained by the fact that the suggested method reduces stationary bias and, as can be seen in Figure 4 the ETC is a rather dynamic sequence and the maximum absolute errors occur in transients.

7. Conclusions

A method for bias compensation in observers is developed. The idea is to, based on measurement data, compute a low dimension augmentation of the model that describes the most significant model biases. This augmented model is used to design an augmented observer that results in a state estimate with reduced bias. Three main results are a characterization of possible augmentations from observability perspectives, a parameterization of the augmentations from the method, and a robustness analysis of the proposed augmentation estimation method.

The method is successfully applied to a diesel engine with VGT and EGR, using a non-linear default model and measurement data from an engine in a test cell. It is shown that an augmentation according to the suggested augmentation procedure reduces the mean estimation error, that is the bias, by approximately 50% in an ETC.

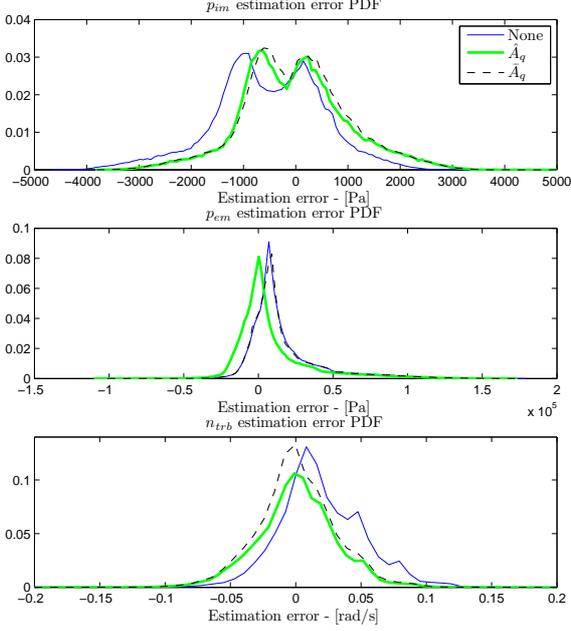


Fig. 6. Probability density functions for default and augmented observers applied to real measurement data using feedback from p_{im} and n_{trb} . The two augmented observers are Red. – augmentation estimated measuring p_{im} and n_{trb} and Full – augmentation estimated measuring p_{im} , p_{em} , and n_{trb} respectively.

Appendix A. Engine model and data

The model, on which the method is applied, is a third order non-linear state space model of a six cylinder Scania diesel engine with VGT and EGR. The model states are intake manifold pressure, p_{im} , and exhaust manifold pressure, p_{em} , and turbine speed, n_{trb} . The inputs are injected amount of fuel, engine speed, VGT and EGR positions. It is based on a model developed in (Wahlström and Eriksson, 2006) but slightly simplified. The simplifications are that the states for the EGR mass fraction and actuator dynamics are removed.

The data is collected in an engine test cell at Scania CV AB in Södertälje, Sweden. The data is from a six cylinder Scania diesel engine with VGT and EGR and was collected during an ETC. The sensor signals used are; intake and exhaust manifold pressures, turbine speed, and engine speed, and actuator signals used are; VGT and EGR positions, and injected amount of fuel. All these signals are available on a standard engine, i.e. no extra lab sensors were used, and collected with a sampling rate of 100 Hz.

Appendix B. Proof of Theorems 4.2 and 5.1

Theorem 4.2 *Assume that (C, F) in (5) is an observable pair then the augmented system (5) is observable if and only if*

$$\text{Ker}\left(\left(F - I\right)\begin{pmatrix} A_q & N_C \end{pmatrix}\right) = \{0\},$$

where the columns of N_C span $\text{Ker } C$.

PROOF. From Theorem 4.1 it follows that the augmented model (5) is observable if and only if $x = 0$, $q = 0$ is the only solution to

$$Cx = 0 \quad (\text{B.1a})$$

$$(\lambda I - F)x + (F - I)A_q q = 0 \quad (\text{B.1b})$$

$$(\lambda I - I)q = 0 \quad (\text{B.1c})$$

for all $\lambda \in \mathbb{C}$. For $\lambda \neq 1$ it is immediate from (B.1c) that $q = 0$. Then the assumption that (C, F) is an observable pair together with (B.1a), (B.1b), and Theorem 4.1 gives that $x = 0$. Thus, only $\lambda = 1$ needs to be investigated further.

For $\lambda = 1$ in (B.1) the augmented model is observable if and only if $x = 0$, $q = 0$ is the only solution to

$$(F - I)(x - A_q q) = 0$$

$$Cx = 0.$$

Let the columns of N_C be a basis for $\text{Ker } C$, then $x = N_C \xi$ for some ξ and observability is equivalent to $q = 0$, $\xi = 0$ being the only solution to the equation

$$(F - I)(N_C \xi - A_q q) = 0.$$

This is equivalent to that the matrix

$$(F - I)\begin{pmatrix} A_q & N_C \end{pmatrix}$$

has full column rank which ends the proof. \square

Theorem 5.1 *Assume that the observer gain, K , is chosen such that the observer is strictly stable and does not have any poles in the origin. Then, in absence of noise, the bias samples are spanned by the rows of C and can thereby be written as*

$$\beta_t = C^T \Gamma.$$

PROOF. Since Approach 1 only is applicable if C has full column rank and due to the augmentation, C^\dagger , used in Approach 3 the theorem automatically holds for these cases. It is therefore sufficient to prove the result for Approach 2.

Now, starting with the output error and rewriting it

$$\begin{aligned} r_t &= C \overbrace{(I - F + KCF)}^W (I - KC)(I - F)\beta_t \\ &= CW^{-1}(W - KC)A_q q = (I - CW^{-1}K)C\beta_t, \end{aligned} \quad (\text{B.2})$$

where the assumption that K is chosen such that the observer system, $(I - F + KCF)$, is strictly stable and does not have any eigenvalues equal to zero which assures that W^{-1} exists, is used. Then, using the pseudo inverse, (B.2) can be written as

$$C\beta_t = (I - CW^{-1}K)^\dagger r_t = \bar{r}_t. \quad (\text{B.3})$$

A unique solution to (B.3) is received by computing the minimum square solution with least Euclidean norm. Writing

$$\beta_t = \beta_t^o + \beta_t^\perp, \quad (\text{B.4})$$

where

$$\beta_t^o \in (\text{Ker } C)^\perp = \text{span}\{C^T\} \quad (\text{B.5})$$

and

$$\beta_t^\perp \in \text{Ker } C, \quad (\text{B.6})$$

the solution with least Euclidean norm is the solution with $\beta_t^\perp = 0$, i.e.

$$\beta_t = \beta_t^o = C^T \Gamma \quad (\text{B.7})$$

which concludes the proof. \square

Tech. Rep. LiTH-R-2747, Department of Electrical Engineering, Linköpings Universitet, SE-581 83 Linköping, Sweden.

Wang, J., December 2008. Air fraction estimation for multiple combustion mode diesel engines with dual-loop egr systems. *Control Engineering Practice* 16 (12), 1479–1486.

References

- Andersson, P., Eriksson, L., 2004. Cylinder air charge estimator in turbocharged SI-engines. In: *Electronic Engine Controls*. No. 2004-01-1366 in SAE Technical paper series SP-1822.
- Bembenek, C., Chmielewski, T. A. J., Kalata, P. R., June 1998. Observability conditions for biased linear time invariant systems. In: *Proceedings of the American Control Conference*. Philadelphia, Pennsylvania, pp. 1180–1184.
- García-Nieto, S., Martínez, M., Blasco, X., Sanchis, J., December 2008. Nonlinear predictive control based on local model networks for air management in diesel engines. *Control Engineering Practice* 16 (12), 1399–1413.
- Höckerdal, E., Eriksson, L., Frisk, E., 2008a. Air mass-flow measurement and estimation in diesel engines equipped with EGR and VGT. In: *Electronic Engine Controls*. No. 2008-01-0992 in SAE Technical paper series SP-2159. SAE World Congress, Detroit, USA.
- Höckerdal, E., Frisk, E., Eriksson, L., 2008b. Observer design and model augmentation for bias compensation applied to an engine. *IFAC World Congress*. Seoul, Korea.
- Kailath, T., 1980. *Linear Systems*. Prentice-Hall, Inc, Englewood Cliffs, New Jersey 07632.
- Kailath, T., Sayed, A. H., Hassibi, B., 2000. *Linear Estimation*, 2nd Edition. Prentice-Hall, Inc, Upper Saddle River, New Jersey 07458.
- Kalman, R., Ho, B. L., Narendra, K., 1963. Controllability of linear dynamical systems. *Contributions to Differential Equations* 1.
- Lee, E. B., Markus, L., 1968. *Foundations of Optimal Control Theory*. John Wiley & Sons, Inc, New York.
- Lino, P., Maione, B., Amorese, C., October 2008. Modelling and predictive control of a new injection system for compressed natural gas engines. *Control Engineering Practice* 16 (10), 1216–1230.
- Tseng, T.-C., Cheng, W. K., 1999. An adaptive air/fuel ratio controller for si engine throttle transients. In: *Electronic Engine Controls*. No. 1999-01-0552 in SAE Technical paper series SP-1419. SAE International Congress and Exposition, Detroit, USA.
- Wahlström, J., Eriksson, L., 2006. Modeling of a diesel engine with VGT and EGR including oxygen mass fraction.