

# Institutionen för systemteknik

## Department of Electrical Engineering

**Examensarbete**

## **Diagnosis of a Truck Engine using Nonlinear Filtering Techniques**

Examensarbete utfört i Fordonssystem  
vid Tekniska högskolan i Linköping  
av

**Fredrik Nilsson**

LITH-ISY-EX--07/3982--SE

Linköping 2007



**Linköpings universitet**  
**TEKNISKA HÖGSKOLAN**



# Diagnosis of a Truck Engine using Nonlinear Filtering Techniques

Examensarbete utfört i Fordonssystem  
vid Tekniska högskolan i Linköping  
av

**Fredrik Nilsson**


LITH-ISY-EX--07/3982--SE

Handledare: **Gustaf Hendeby**  
ISY, Linköpings universitet  
**Erik Frisk**  
ISY, Linköpings universitet  
**Erik Höckerdal**  
SCANIA

Examinator: **Erik Frisk**  
ISY, Linköpings universitet

Linköping, 25 May, 2007



	<b>Avdelning, Institution</b> Division, Department  Division of Vehicular Systems Department of Electrical Engineering Linköpings universitet SE-581 83 Linköping, Sweden		<b>Datum</b> Date  2007-05-25
	<b>Språk</b> Language  <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English  <input type="checkbox"/> _____	<b>Rapporttyp</b> Report category  <input type="checkbox"/> Licentiatavhandling <input checked="" type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input type="checkbox"/> Övrig rapport <input type="checkbox"/> _____	<b>ISBN</b> _____ <b>ISRN</b> LITH-ISKY-EX--07/3982--SE <b>Serietitel och serienummer ISSN</b> Title of series, numbering _____
<b>URL för elektronisk version</b> <a href="http://www.vehicular.isy.liu.se">http://www.vehicular.isy.liu.se</a> <a href="http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-8959">http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-8959</a>			
<b>Titel</b> Title   <b>Författare</b> Fredrik Nilsson Author	Diagnos av lastbilmotor med användning av tekniker för olinjär filtrering Diagnosis of a Truck Engine using Nonlinear Filtering Techniques		
<b>Sammanfattning</b> Abstract  <p>Scania CV AB is a large manufacturer of heavy duty trucks that, with an increasingly stricter emission legislation, have a rising demand for an effective On Board Diagnosis (OBD) system. One idea for improving the OBD system is to employ a model for the construction of an observer based diagnosis system. The proposal in this report is, because of a nonlinear model, to use a nonlinear filtering method for improving the needed state estimates. Two nonlinear filters are tested, the Particle Filter (PF) and the Extended Kalman Filter (EKF). The primary objective is to evaluate the use of the PF for Fault Detection and Isolation (FDI), and to compare the result against the use of the EKF.</p> <p>With the information provided by the PF and the EKF, two residual based diagnosis systems and two likelihood based diagnosis systems are created. The results with the PF and the EKF are evaluated for both types of systems using real measurement data. It is shown that the four systems give approximately equal results for FDI with the exception that using the PF is more computational demanding than using the EKF. There are however some indications that the PF, due to the nonlinearities, could offer more if enough CPU time is available.</p>			
<b>Nyckelord</b> Keywords    particle filter, diagnosis, extended Kalman filter, likelihood, cusum			



# Abstract

Scania CV AB is a large manufacturer of heavy duty trucks that, with an increasingly stricter emission legislation, have a rising demand for an effective On Board Diagnosis (OBD) system. One idea for improving the OBD system is to employ a model for the construction of an observer based diagnosis system. The proposal in this report is, because of a nonlinear model, to use a nonlinear filtering method for improving the needed state estimates. Two nonlinear filters are tested, the Particle Filter (PF) and the Extended Kalman Filter (EKF). The primary objective is to evaluate the use of the PF for Fault Detection and Isolation (FDI), and to compare the result against the use of the EKF.

With the information provided by the PF and the EKF, two residual based diagnosis systems and two likelihood based diagnosis systems are created. The results with the PF and the EKF are evaluated for both types of systems using real measurement data. It is shown that the four systems give approximately equal results for FDI with the exception that using the PF is more computational demanding than using the EKF. There are however some indications that the PF, due to the nonlinearities, could offer more if enough CPU time is available.





# Acknowledgments

First of all I would like to thank Scania for their contributions and especially my supervisor Erik Höckerdal for his support and the department of NED for their interest in this work.

Further I would like to thank my supervisors at Linköpings universitet, Gustaf Hendeby at the division of Automatic Control for his help and expertise with the complicated particle filter, and Erik Frisk at the division of Vehicular Systems for his comments and ideas for improvements concerning this report.

I would also like to thank the entire crew at the division of Vehicular Systems for helping me and answer my many, and not always, bright questions. The other Master's thesis students at Vehicular Systems also deserves my compliments for providing company, fun breaks and lunches.

*Fredrik Nilsson*  
Linköping 2007



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Objectives . . . . .	2
1.2	Contributions . . . . .	3
1.3	Limitations . . . . .	3
<b>2</b>	<b>Nonlinear Filtering</b>	<b>5</b>
2.1	Particle Filter . . . . .	6
2.1.1	PDF estimation . . . . .	6
2.1.2	Resampling . . . . .	8
2.1.3	Bootstrap Algorithm . . . . .	8
2.1.4	Likelihood . . . . .	9
2.1.5	State Estimates . . . . .	9
2.1.6	Improvements . . . . .	10
2.2	Extended Kalman Filter . . . . .	10
2.2.1	Kalman equations . . . . .	11
2.2.2	Likelihood . . . . .	11
2.2.3	Linearization . . . . .	11
2.3	Comparison . . . . .	12
<b>3</b>	<b>Model Based Diagnosis</b>	<b>13</b>
3.1	Fault Isolation . . . . .	15
3.2	Thresholds . . . . .	16
3.2.1	Adaptive Thresholds . . . . .	17
3.3	Test Quantities . . . . .	17
3.3.1	Residual Based . . . . .	17
3.3.2	Likelihood Based . . . . .	18
3.3.3	Cusum Test . . . . .	18
3.4	Power Function . . . . .	19
<b>4</b>	<b>Engine Models</b>	<b>21</b>
4.1	Engine with VGT and EGR . . . . .	22
4.2	Time Discretization . . . . .	22
4.3	Model Overview . . . . .	23
4.3.1	6-state Model . . . . .	23
4.3.2	3-state Model . . . . .	24

4.3.3	Comparison . . . . .	24
4.3.4	3-state Model Equations . . . . .	26
<b>5</b>	<b>Diagnosis Systems Design</b>	<b>33</b>
5.1	Considered Faults . . . . .	33
5.2	Diagnosis Systems . . . . .	34
5.2.1	Residual Based . . . . .	35
5.2.2	Likelihood Based . . . . .	36
5.2.3	Tuning the Cusum Parameters . . . . .	37
5.2.4	Discussion . . . . .	37
5.3	Modeling Faults . . . . .	38
5.3.1	Particle Filter . . . . .	38
5.3.2	Extended Kalman Filter . . . . .	38
5.3.3	Discussion . . . . .	39
5.4	Tuning the Filters . . . . .	39
5.4.1	Particle Filter . . . . .	40
5.4.2	Extended Kalman Filter . . . . .	42
5.4.3	Discussion . . . . .	43
5.5	Compensating for Model Errors . . . . .	43
5.5.1	Estimating the Model Errors . . . . .	44
5.5.2	Adaptive Thresholds . . . . .	45
5.5.3	Adaptive Noise Distribution . . . . .	46
5.5.4	Discussion . . . . .	46
<b>6</b>	<b>Diagnosis Systems Evaluation</b>	<b>47</b>
6.1	Fault Detection . . . . .	47
6.1.1	Power Functions . . . . .	52
6.2	Fault Isolation . . . . .	54
6.3	Performance Enhancements with an Adaptive Threshold . . . . .	58
6.4	Estimation Problems for High Engine Speed . . . . .	59
6.5	Performance Limiting Factors . . . . .	59
6.6	Summary and Discussion . . . . .	60
<b>7</b>	<b>Conclusions</b>	<b>61</b>
	<b>References</b>	<b>63</b>
<b>A</b>	<b>Appendix: Cusum Parameters</b>	<b>65</b>

# Chapter 1

## Introduction

This Master's thesis is performed in a collaboration with Scania CV AB in Södertälje, the division of Vehicular Systems and the division of Automatic Control at the department of Electrical Engineering at Linköpings universitet.

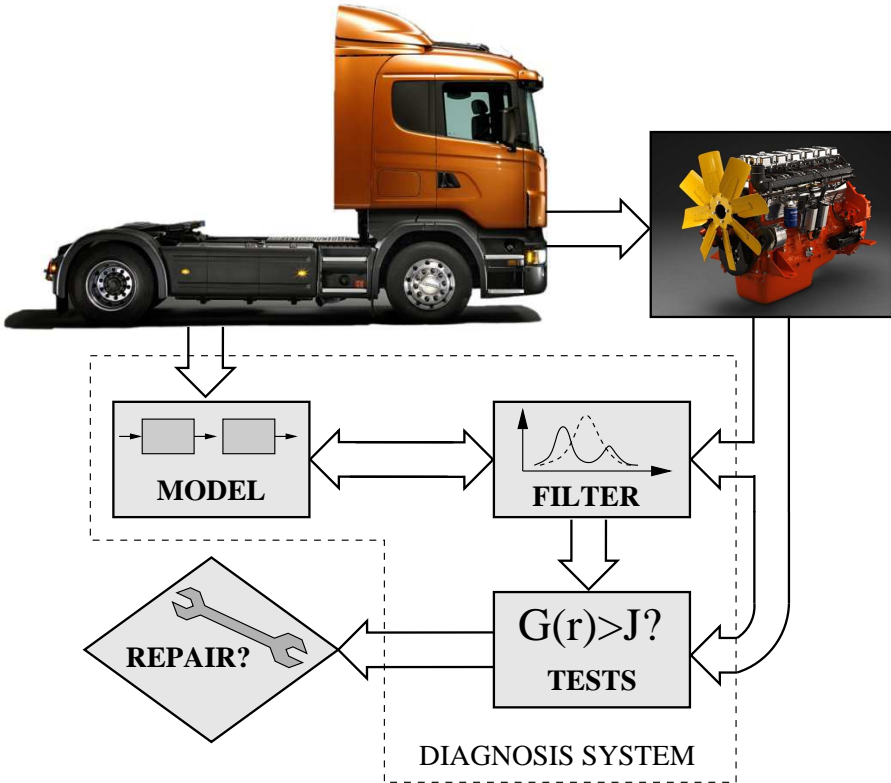
Scania is a large manufacturer of heavy duty trucks, that together with other truck manufacturers meet an increasingly stricter emission legislation. Scania needs to comply and their demand for an effective On Board Diagnosis (OBD) system is therefore rising. The OBD system's main purpose is to detect faults, e.g. faults in actuators and sensors, leading to emissions beyond legislated levels. Even small faults not severely affecting the emissions are of interest to detect in an early stage. The information from detection and isolation of an arbitrary fault can be used during regular maintenance, or for fast repair and replacement of faulty components.

One idea for improving the OBD system is to employ a model for the construction of an observer based diagnosis system. In this thesis, the model is of a Scania diesel truck engine with Exhaust Gas Recirculation (EGR) and a Variable Geometry Turbine (VGT). Inputs and measurements from the engine are compared with the estimates from the observer and a statement of the system condition is made.

The proposal in this thesis is, because of a nonlinear model, that a nonlinear filter method is used as an observer for improving the model estimates. Due to the nonlinearities in the model, the use of a nonlinear filter hopefully provides an advantage for Fault Detection and Isolation (FDI) compared to other methods.

Two nonlinear filters are tested, one is the particle filter and the other is the more commonly used extended Kalman filter. The particle filter mentioned here is not to confuse with a particle filter used for the removal of particles in the exhausts and the extended Kalman filter is simply the ordinary Kalman filter together with a linearization of the nonlinear model.

An overview of a diagnosis system that is monitoring a truck engine is presented in Figure 1.1. The diagnosis systems constructed in this thesis have the same structure as shown in the figure.



**Figure 1.1.** This figure gives an overview of a diagnosis system monitoring a process which in this case is the engine. The arrows represent data flows between the different subsystems. The two arrows out from the truck is input signals (control signals) to the engine which are used by the model to predict the state. The state prediction and the sensor values are used by the filter to make a state estimate (the filter also provides other information) which is used by the tests for monitoring the system. In case of a fault, the diagnosis system alarm and appropriate actions should be taken.

## 1.1 Objectives

The main objective is to, by using the particle filter and the extended Kalman filter, construct a diagnosis systems for a Scania diesel truck engine and evaluate the properties for FDI. Using a particle filter as the primary method, different approaches for the diagnosis problem are tried.

Construction of a diagnosis system with as good performance as possible, that is still easy to implement with a short execution time, is considered as a secondary objective.

Absolute performance improvements, if there are any compared to the methods used in the OBD system today are not presented in this thesis.

## 1.2 Contributions

The contributions in this thesis are summarized to:

- Modifications of a model of a Scania truck engine for the application of the particle filter and the extended Kalman filter.
- The construction of a particle filter applicable to the engine model with performance good enough for the diagnostics purposes.
- The construction of an extended Kalman filter applicable to the engine model for comparison with the particle filter.
- The design of tests with good performance of finding faults in the engine.

## 1.3 Limitations

Some factors that confines the scope of this Master's thesis:

- The engine model is pre-constructed for another purpose, no modifications of the model are made for the use in this thesis.
- There are a limited availability of measurement data for the diagnosis systems evaluation.





# Chapter 2

## Nonlinear Filtering

This chapter is an introduction to two methods for nonlinear filtering, the Particle Filter (PF) and the Extended Kalman Filter (EKF). The focus in this Master's thesis lies on the PF, but the result is compared to that of the EKF and therefore the theory of the EKF is also included.

The filters introduced in this section are applicable to discrete systems with nonlinear dynamics and a nonlinear measurement, described by the functions  $f(x_k, u_k)$  and  $h(x_k, u_k)$  where  $x_k$  is the state variable and  $u_k$  is the system input signal at time  $k$ . Further let  $y_k$  denote the measurement. The system is then assumed to be in the form

$$x_{k+1} = f(x_k, u_k) + v_k, \quad (2.1a)$$

$$y_k = h(x_k, u_k) + n_k, \quad (2.1b)$$

where the variables  $v_k$  and  $n_k$  are the system and measurement noise with known distributions. The initial distribution of the state variable also have to be known.

Both the PF and the EKF are Bayesian filters. In Bayesian filtering the solution lies in calculating the state Probability Density Function (PDF), for every iteration. Using the expressions

$$p(x_k | Y_{k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | Y_{k-1}) dx_{k-1}, \quad (2.2a)$$

$$p(x_k | Y_k) = \frac{p(y_k | x_k) p(x_k | Y_{k-1})}{p(y_k | Y_{k-1})}, \quad (2.2b)$$

where the filter solution is the PDF  $p(x_k | Y_k)$ , which is the probability density of  $x_k$  given all previous values of the measurements, i.e.  $Y_k = \{y_i\}_{i=1}^k$ . The expressions (2.2a) and (2.2b) are often referred to as the time update and the measurement update, see [1] for more information.

It is not always possible to analytically calculate these expressions for any kind of distributions. The EKF approximates the solutions when the distributions are Gaussian and the PF approximates the solutions for any kind of distributions. The PDF solution from (2.2b) contains everything needed for state estimation which is the main task for the filters in this thesis.

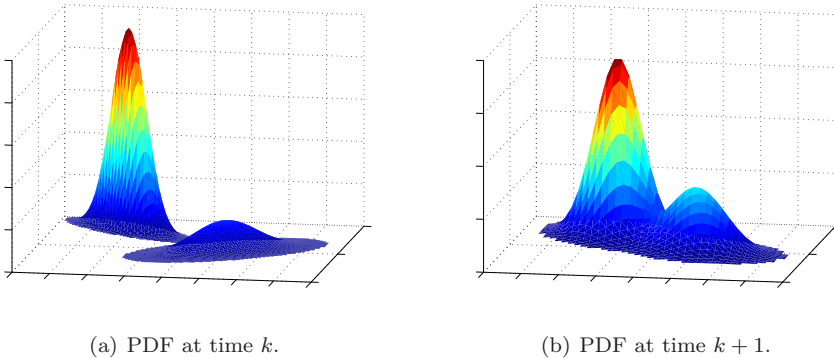
## 2.1 Particle Filter

There are several types and variations of PFs. A general method will be presented as the Bootstrap filter, as well as various methods for improving the Bootstrap algorithm for the application in this thesis. The Bootstrap filter was introduced in [7] and is a PF easy to apply to a given system. For a more extensive description of PF methods, see [6], [4] and [1].

The key idea in the Bootstrap filter, as well as in any PF, is to represent the required PDF by a set of samples. The samples are associated with weights which represent how important these samples are. Each sample, with its respective weight, is referred to as a particle.

The particles are in each time step as the iteration goes on, due to model errors and noise, likely to drift away from the real state. This results in small weights which is bad because the weights represent the significance of the PDF estimate. This is called the degeneracy problem and can be solved by removing particles with low weight and duplicating those with high weight. An example of how a particle cloud is affected by the degeneracy can be observed in Figure 2.2.

One of the strengths of the PF, is that the discrete PDF representation has no difficulties with non-Gaussian distributions. Consider a two dimensional system with the PDF according to Figure 2.1. This distribution is bimodal and therefore, clearly not Gaussian. This distribution would be impossible to represent with a KF/EKF and information about the states, in that case, would be lost.



**Figure 2.1.** A PF estimation of the state PDF at two time steps for a two dimensional system with a bimodal state distribution.

### 2.1.1 PDF estimation

The PF procedure for the PDF estimation can be divided into three steps, Initiation, Prediction and Update. A fourth step can be added to counter the effect of degeneracy. The PDF is through the entire estimation procedure represented by a set of particles, each particle consisting of a state vector and one weight.

### Initiation

The initiation stage of the filter is done by drawing  $N$  samples,  $x_0^{(i)*}$  where  $i = 1, \dots, N$ , from a known distribution of the state, i.e.  $p(x_0)$ . The initial weight for each sample is  $1/N$ .

### Prediction

The particles  $x_{k-1}^{(i)*}$  are propagated through the system (2.1a). Note that the distribution of  $v_k$ ,  $p(v_k)$ , has to be known, or at least samples from the distribution have to be available. The new set of particles  $x_k^{(i)}$  represents  $p(x_k|Y_{k-1})$  which is the approximation of (2.2a).

### Update

Each particle,  $x_k^{(i)}$ , are compared with the obtained measurement  $y_k$  through the observation (2.1b). The comparison gives the estimate of the conditional PDF  $p(y_k|x_k) = p(y_k - h(x_k, u_k))$ , defined by the statistics of the measurement noise  $p(n_k)$ . Estimating the conditional PDF correspond to computing

$$p(y_k|x_k) = \int \delta(y_k - h(x_k, u_k) - n_k)p(n_k) dx_k \quad (2.3)$$

for each particle, where  $\delta$  is the dirac function. A weight depending on the estimate of (2.3) and the prior weight

$$w_k^{(i)} \propto p(y_k|x_k^{(i)})w_{k-1}^{(i)} \quad (2.4)$$

can be set to each particle, see [1]. If the prior weight for each particle is  $1/N$  then the weight could be computed as [5]

$$w_k^{(i)} = p(y_k|x_k^{(i)})\frac{1}{N}. \quad (2.5)$$

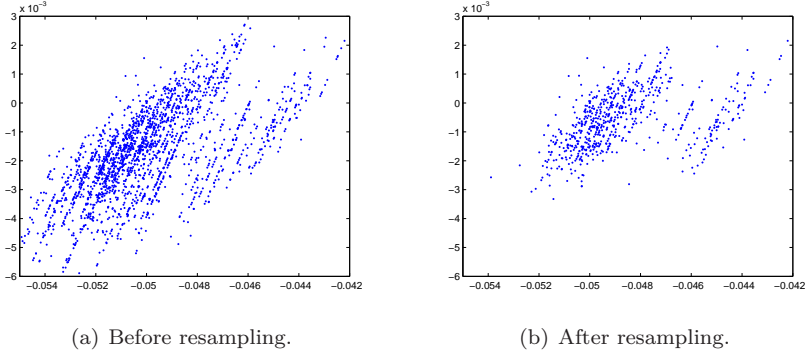
The wanted state PDF can now be approximated with

$$p(x_k|Y_k) \approx \sum_{i=1}^N \bar{w}_k^{(i)} \delta(x_k - x_k^{(i)}) \quad (2.6)$$

where  $\bar{w}_k^{(i)}$  is the normalized weight

$$\bar{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=1}^N w_k^{(i)}}. \quad (2.7)$$

The approximation in (2.6) can be shown to approach the true PDF as  $N \rightarrow \infty$ , see [1].



**Figure 2.2.** Both plot (a) and (b) display a particle cloud with  $N = 2000$  particles for the same system at time  $k$ . Each particle in (a) are represented as a dot, and has been set an individual weight according to (2.4). After the resampling step, plot (b), the weight of each particle has been set to  $1/N$  and therefore several dots represent more than one particle.

### 2.1.2 Resampling

This stage is used to avoid the degeneracy phenomenon mentioned in Section 2.1. The problem is that the particles are likely to drift away from the real state and therefore approximating the PDF with particles of low weight. The weights represent the significance in the estimation and are desired to be as high as possible.

The solution to this problem is to remove particles with low weights and duplicate those with high weight. This can be done by resampling  $N$  particles  $x_k^{(i)*}$  according to the rule

$$\Pr(x_k^{(j)*} = x_k^{(i)}) = \bar{w}_k^{(i)}. \quad (2.8)$$

Statistically, the particles with high weight will be selected many times. This way of resampling in every iteration is called Sampling Importance Resampling (SIR) and will reset the normalized weight of each particle to  $1/N$ . The particles  $\{x_k^{(i)}\}_{i=1}^N$  with high weight before the resampling stage, are now represented many times in  $\{x_k^{(i)*}\}_{i=1}^N$  instead.

An example of a particle cloud before and after the resampling procedure can be seen in Figure 2.2.

### 2.1.3 Bootstrap Algorithm

A summary of the PF method described in Sections 2.1.1 and 2.1.2 is here presented as the Bootstrap algorithm.

1. INITIATION

$k := 1$ , draw  $N$  samples  $x_0^{(i)*}$  from the initial PDF  $p(x_0)$ .

2. PREDICTION

Draw  $N$  independent samples of the noise  $v_{k-1}$  according to  $p(v_{k-1})$  and

compute  $x_k^{(i)} = f(x_{k-1}^{(i)*}, u_{k-1}) + v_{k-1}^{(i)}$ .

### 3. MEASUREMENT UPDATE

Given  $y_k$ , compute the weight  $w_k^{(i)} = p(y_k | x_k^{(i)}) \frac{1}{N}$  and the normalized weight  $\bar{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=1}^N w_k^{(i)}}$  for each sample.

### 4. RESAMPLING

Draw  $N$  samples  $x_k^{(j)*}$ , replacing the old set of particles, with the probability  $\Pr(x_k^{(j)*} = x_k^{(i)}) = \bar{w}_k^{(i)}$  to draw a certain particle.

5.  $k := k + 1 \rightarrow$  step 2.

## 2.1.4 Likelihood

The PDF  $p(y_k | Y_{k-1})$  can be referred to as the likelihood and is defined as the probability of  $y_k$  given  $Y_{k-1}$ ,

$$p(y_k | Y_{k-1}) = \int p(y_k | x_k) p(x_k | Y_{k-1}) dx_k. \quad (2.9)$$

The likelihood can be found in expression (2.2b) as the denominator but does not have to be calculated explicitly for the state PDF estimation. The likelihood can be approximated as

$$p(y_k | Y_{k-1}) \approx \sum_{i=1}^N w_k^{(i)}. \quad (2.10)$$

This quantity is in the diagnostics section used for hypothesis testing. For a proof of approximation (2.10), see [5].

## 2.1.5 State Estimates

A couple of different methods for estimating the state values, given the PDF estimate, are

$$\hat{x}_k = \sum_{i=0}^N w_k^{(i)} x_k^{(i)}, \quad (2.11a)$$

$$\hat{x}_k = \frac{1}{N} \sum_{i=0}^N x_k^{(i)*}, \quad (2.11b)$$

$$\hat{x}_k = x_k^{(i)} \quad \text{where } i = \arg \max_i w_k^{(i)}. \quad (2.11c)$$

The first two are quite similar, both are mean value estimations with the difference that (2.11b) is done after the resampling stage and forsakes some particles with low weights in favor of saving CPU time. The third one, (2.11c), which uses the value of the particle with the highest weight, tends to be useful when the distribution is multimodal. If (2.11a) or (2.11b) is used on a bimodal distribution (see Figure 2.1), the estimate will end up between the peaks and lose its significance.

### 2.1.6 Improvements

There are a number of possible modifications that can be made to improve the performance of the PF, but only a few of them will be briefly explained here. There are e.g. several different resampling algorithms with different properties, that are for different systems more or less appropriate. Other methods than the SIR method will not be presented in this thesis, see [1] for more information.

#### Importance density

One possible way to improve the importance sampling, i.e. the way the weights are chosen, is to use an importance density,  $q(\cdot)$ . The modified computation of the weights with an importance density are computed as

$$w_k^{(i)} \propto \frac{p(y_k | x_k^{(i)}) p(x_k^{(i)} | x_{k-1}^{(i)})}{q(x_k^{(i)} | x_{k-1}^{(i)}, y_k)} w_{k-1}^{(i)}. \quad (2.12)$$

The importance density represent where the particles are believed to end up after the prediction step. This density gives the opportunity to move the particles closer to the measurement, and in this way, increase the total weight of the estimate, i.e. the likelihood. The difficulties here lies in choosing  $q(\cdot)$ . If  $q(\cdot)$  is chosen well, fewer particles are needed for the same quality of the estimate. For more information how to use and choose the importance density, see [1], [6] and [4].

#### Roughening

There is a possibility that the particles may collapse to a single value, this would happen relatively quick if there were no system noise. The system noise spreads the particles when they are propagated through (2.1a) in the prediction step. If the particles represent the same state value before the prediction, and there is no system noise, the particles would again end up representing the same state value. This problem is likely to occur after the resampling if there are few particles chosen to represent the PDF.

A solution to this problem is the roughening procedure, which is a simple procedure that will jitter the resampled values. The jitter effect will boost the number of particles close to the real state which gives a better estimate of the PDF, see [7].

## 2.2 Extended Kalman Filter

The EKF is basically the ordinary Kalman Filter (KF) for a linearization of a nonlinear system. The theory for the KF extends only to linear systems and the KF equations will be presented in this section along with a method for linearization.

The KF was introduced in [11] and is an optimal filter for a linear system with Gaussian noise. If the system or measurement noise has another distribution, the KF is still the optimal unbiased *linear* filter. For a formal proof, see for instance

[8]. For a more general description of the KF and the EKF that extends beyond the scope of this thesis, see [11], [10] and [8].

### 2.2.1 Kalman equations

Consider the linear system

$$x_{k+1} = A_k x_k + B_k u_k + v_k, \quad (2.13a)$$

$$y_k = C_k x_k + n_k, \quad (2.13b)$$

where  $v_k$  and  $n_k$  are Gaussian noise with covariance matrices  $Q_k$  and  $R_k$ , respectively. The time update and measurement update equations for the KF solution of (2.2a) and (2.2b) for the linear system (2.13) are

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + B_k u_k \quad (2.14a)$$

$$P_{k+1|k} = A_k P_{k|k} A_k^T + Q_k \quad (2.14b)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - C_k \hat{x}_{k|k-1}) \quad (2.14c)$$

$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1} \quad (2.14d)$$

$$K_k = P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + R_k)^{-1}. \quad (2.14e)$$

The Kalman equations, (2.14), solves (2.2a) and (2.2b) for a linear system with the assumption that  $v_k$  and  $n_k$  are Gaussian. For a formal proof of (2.14), see [8].

### 2.2.2 Likelihood

The likelihood for the EKF is defined in the same way as for the PF, i.e. as (2.9). The Kalman solution of the likelihood is

$$p(y_k | Y_{k-1}) = N(C \hat{x}_{k|k}, C P_{k|k} C^T + R_k). \quad (2.15)$$

### 2.2.3 Linearization

For the KF equations to be applicable to a system, the system has to be linear in the form described by (2.13). If the system is nonlinear and described according to (2.1), the system has to be linearized. The most straightforward way to linearize the system is by using a Taylor expansion around the linearization point  $(x^*, u^*)$ . The Taylor expansion for  $f(x_k, u_k)$ , neglecting second order terms and higher, is

$$f(x_k, u_k) \approx f(x^*, u^*) + \frac{df(x^*, u^*)}{dx_k} (x_k - x^*) + \frac{df(x^*, u^*)}{du_k} (u_k - u^*). \quad (2.16)$$

For a system with dimension  $i$  where the dynamics of each state is represented by a function  $f_i(x_k, u_k)$ , the linearization in every time step leads to the system matrix

$$A_k = \begin{pmatrix} \frac{df_{k,1}}{dx_{k,1}} & \cdots & \frac{df_{k,1}}{dx_{k,i}} \\ \vdots & \ddots & \vdots \\ \frac{df_{k,i}}{dx_{k,1}} & \cdots & \frac{df_{k,i}}{dx_{k,i}} \end{pmatrix}. \quad (2.17)$$

Further let  $j$  be the number of input signals. The dynamics between the input and the states is then described by the matrix

$$B_k = \begin{pmatrix} \frac{df_{k,1}}{du_{k,1}} & \cdots & \frac{df_{k,1}}{du_{k,j}} \\ \vdots & \ddots & \vdots \\ \frac{df_{k,i}}{du_{k,1}} & \cdots & \frac{df_{k,i}}{du_{k,j}} \end{pmatrix}. \quad (2.18)$$

With the variable change,  $z = x - x^*$ ,  $\tilde{u} = u - u^*$ , the matrices in (2.17) and (2.18) are valid for the following system

$$z_{k+1} = A_k z_k + B_k \tilde{u}_k \quad (2.19)$$

around the linearization point  $(x^*, u^*)$ . The  $C$  matrix is, for the measurement  $\zeta$  in the observation equation  $\zeta = Cz$ , obtained in the same way for the nonlinear function  $h(x_k, u_k)$ <sup>1</sup> in (2.1b).

## 2.3 Comparison

The strength in using a PF for state estimation, is not only its ability to handle nonlinear systems. In comparison to the EKF, the PF has no problem with handling non-Gaussian system and measurement noise. A downside with the PF is that the performance depends on the number of particles used. Too many particles will increase the need of computational power and too few will give a bad PDF estimate.

If the linearization does not explain the dynamics well enough the EKF will give a bad estimate of the PDF. This could also happen if the system or measurement noise is non-Gaussian, but the KF is not very sensitive to non-Gaussian noise as long as the noise is unimodal. A Gaussian approximation of a bimodal distribution, e.g as in Figure 2.1, would not be very accurate.

With the assumption that an infinite number of particles can be used, the performance of the PF are only matched by the EKF, or KF, if the system is linear and the noise is Gaussian. When a limited number of particles is used with a linear system with Gaussian noise, the EKF, or KF, will outperform the PF. In the last case, there is nothing to gain by using the PF.

---

<sup>1</sup>The observation function  $h(\cdot)$  for the model considered in this thesis does not depend on  $u$ . Therefore is it only needed to linearize  $h(\cdot)$  in respect to  $x$ .

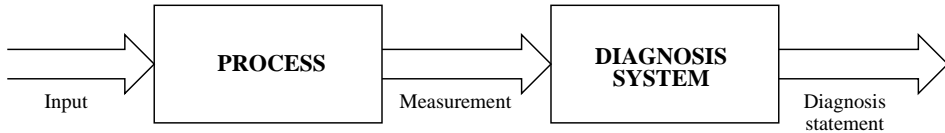


# Chapter 3

## Model Based Diagnosis

This chapter is a short introduction to the concepts of model based diagnosis used in this Master's thesis. For a more extensive description, see Model Based Diagnosis of Technical Processes by Nyberg and Frisk [12] which contains most of the material presented here. For some other sources of information about diagnosis in general, FDI and nonlinear approaches, see [13] and [3].

The purpose of a diagnosis system is to find faults in a process. And if possible, also to identify the fault, i.e. make a diagnosis. The principle of a basic diagnosis system can be seen in Figure 3.1.



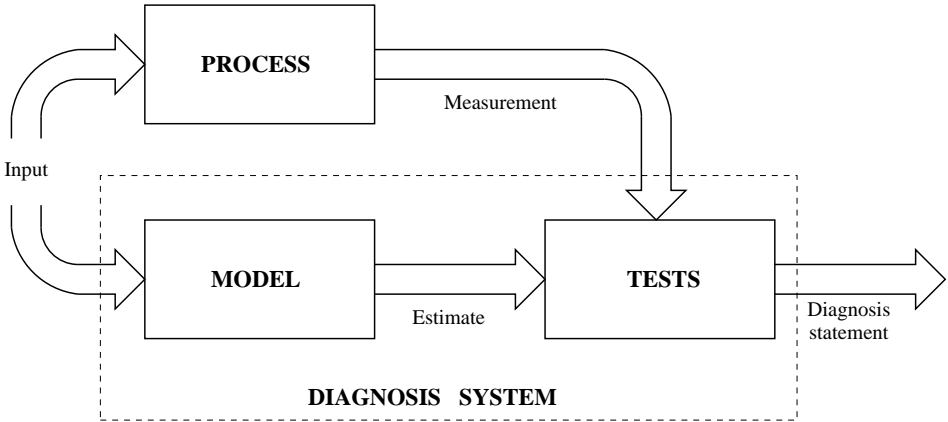
**Figure 3.1.** A sketch over a basic diagnosis system applied to a process.

**Definition 3.1 (Fault)** *A fault is a not permitted deviation of at least one characteristic property or variable of the system from acceptable/usual/standard/nominal behavior.*

**Definition 3.2 (Diagnosis)** *A conclusion of what fault or combination of faults that can explain the process behavior is said to be a diagnosis.*

If there exists a fault in a process and that fault is successfully located by the diagnosis system, appropriate actions could be taken, e.g. if a sensor is faulty, that sensor could be replaced. Or if a control system (if there is any) controlling the process has feedback from the diagnosis systems, it could compensate for that fault directly.

A model based diagnosis system uses a model of the process along with measurements from the actual process to make its diagnosis statement, see Figure 3.2.



**Figure 3.2.** A sketch over the principals of a basic model based diagnosis system applied to a process.

For a diagnosis system to be able to make a diagnosis statement, there has to be redundancy in the system, i.e. there has to be at least two different ways to observe a variable, either directly or indirectly. Typically one from a model and the other from a sensor.

---

### Example 3.1: Model based diagnosis system

---

Consider a simple process described by the discrete model

$$x_{k+1} = u_k \quad (3.1)$$

and the measurement  $y_k$  from a sensor measuring  $x_k$ , i.e.  $y_k = x_k$ . The signal  $u_k$  is the input to an actuator. Using the redundancy available, a comparison between  $y_k$  and  $u_k$  can be made according to

$$T_k = y_k - u_{k-1}, \quad (3.2)$$

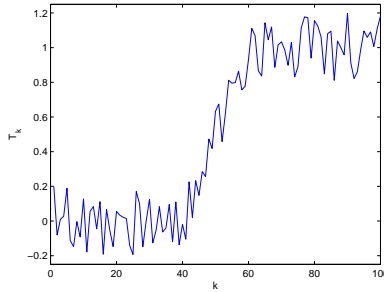
where  $T_k$  should be close to zero when the system is fault free. In the presence of a fault, in the sensor measuring  $x_k$ , or in the actuator  $u_k$ ,  $T_k$  will differ from zero. If the system is fault free up to time  $k = 40$  and faulty after time  $k = 40$ ,  $T_k$  could e.g. be observed as in Figure 3.3. In this example it is impossible to decide whether the sensor or the actuator is faulty after time  $k$ , i.e. the diagnosis statement is, a fault in either  $u_k$  or  $y_k$ .

---

The function  $T_k$  in (3.2) is called a test quantity and is defined as

$$T_k = T(y_k, u_k), \quad (3.3)$$

and typically returns a scalar value. A test quantity is said to respond when it is large/low enough and crosses a specified limit (threshold).



**Figure 3.3.** A test quantity,  $T_k$ , for a system subjected to a fault at time  $k = 40$ . The small deviations is sensor noise.

An important issue for a diagnosis system is the ability to isolate a fault, i.e. to be able to determine which fault that has caused the behavior of the faulty system. Isolation is not possible for the system in Example 3.1 because there is only one test quantity and two possible faults and no information of the system behavior during faulty conditions. If isolation is wanted for a system with many possible faults, it is often required (not always) that several test quantities have to be created. An overview of a more complex diagnosis system, with many test quantities and faults, can be obtained with the help of a decision structure.

---

**Example 3.2: Decision structure**

---

Let  $F_i$  denote a certain fault  $i$ , then consider a system with, say three faults and three test quantities  $T_i$ , i.e.  $i = \{1, 2, 3\}$ . Further let the test quantity  $T_1$  be sensitive to all faults,  $T_2$  to  $F_2$ , and  $T_3$  sensitive to  $F_2$  and  $F_3$ . The decision structure for the diagnosis system described is then

	$F_1$	$F_2$	$F_3$	
$T_1$	X	X	X	(3.4)
$T_2$	0	X	0	
$T_3$	0	X	X	

where an X marks that  $T_i$  is sensitive to fault  $F_j$ . A zero in row  $i$ , column  $j$  means that  $T_i$  will never respond to  $F_j$ , i.e. the test quantity will never be affected by fault  $F_j$ . Sensitivity mentioned here means that a test quantity has a chance of responding when the system is subjected to a fault and not that it always responds.

---

### 3.1 Fault Isolation

Observe the decision structure (3.4) in Example 3.2. If that diagnosis system is subjected to, for instance fault  $F_1$ , the system is not able to isolate which fault has

occurred. That is because the only test quantity sensitive to  $F_1$ , is also sensitive to faults  $F_2$  and  $F_3$ . When  $T_1$  responds to fault  $F_1$  the system can only draw the conclusion that any of the faults has occurred. A decision structure where all faults are isolable from each other can look like

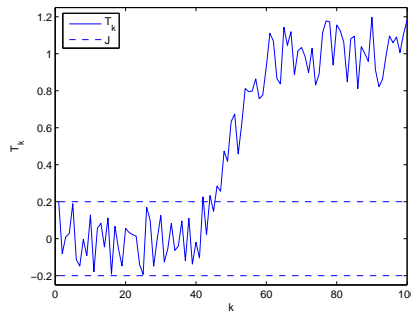
$$\begin{array}{c|ccc}
 & F_1 & F_2 & F_3 \\
 \hline
 T_1 & X & X & 0 \\
 T_2 & 0 & X & X \\
 T_3 & X & 0 & X
 \end{array} \tag{3.5}$$

where it is easy to see that if fault  $F_1$  occur, test quantity  $T_1$  and  $T_3$  might respond. And if they do respond, the only reasonable explanation will be that  $F_1$  is the cause. This conclusion is made by looking in (3.5) and seeing that the only common factor for  $T_1$  and  $T_3$  is  $F_1$ . The zeros in a decision structure are decoupled faults. Decoupling of faults, i.e. to make the diagnosis system insensitive to a specific fault, is important to do for obtaining a wanted decision structure and in that way, a diagnosis systems able of isolating each fault.

Sometimes it can be necessary to have more test quantities than faults to achieve a diagnosis system able of isolating each fault. More test quantities than faults can also be used to increase the chance of detecting the faults.

## 3.2 Thresholds

To make the decision that a fault has occurred simply based on when a test quantity differs from zero is in reality not a good idea. Consider the test quantity presented in Figure 3.3, due to sensor noise the test quantity differs from zero even when a fault has not occurred. This problem is solved using thresholds and the rule that a fault has occurred when the test quantity exceeds a threshold. The same test quantity as in Figure 3.3 can be seen in Figure 3.4 with two thresholds denoted  $J$ . A difficult part when thresholding a test quantity is to decide the size of the threshold, if it is set too high it will not respond to small faults, and if it is

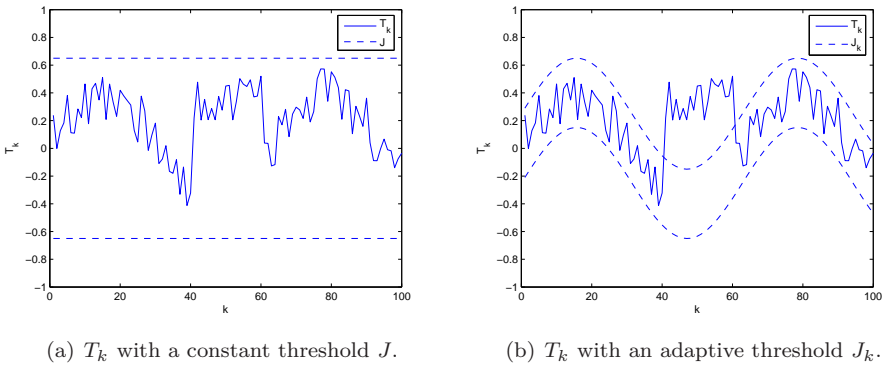


**Figure 3.4.** A test quantity  $T_k$  with a lower and an upper threshold for a system subjected to a fault at time  $k = 40$ . The small deviations is sensor noise.

too small it will generate false alarms, i.e. respond even when the system is fault free.

### 3.2.1 Adaptive Thresholds

If a constant threshold is used according to Figure 3.4 the test could lose detection performance when applied to a test quantity based on a model with large model errors, see Figure 3.5(a). This problem can be solved, if the model error is somewhat systematic, by the use of an adaptive threshold, i.e. a threshold  $J_k$  that is a function of the model error. The threshold adapts to the disturbances and follows the test quantity in in the fault free case. The threshold should not adapt itself due to a fault. See Figure 3.5 for an example of a diagnosis system with adaptive thresholds.



**Figure 3.5.** Consider a system with model errors subjected to a fault between time  $k = 40$  and  $k = 60$ , both reflected in test quantity  $T_k$ . In Plot (a) the fault is not detected at all and in Plot (b) the test quantity clearly exceeds the threshold and the fault is therefore detected.

## 3.3 Test Quantities

There are many different ways to construct a test quantity, this section only provides information about the type of test quantities used in this thesis.

### 3.3.1 Residual Based

Let  $y_{k,i}$  denote the measurement of state  $i$  at time  $k$  and  $\hat{y}_{k,i}$  the estimate from a model. Then one way to make a test quantity, as done in Example 3.1, is simply to use the residual

$$T_k = y_{k,i} - \hat{y}_{k,i}. \quad (3.6)$$

A residual is typically around zero when in the fault free case.

Sometimes it is wise to let the test quantity be a function or filter of the residual,

$$T = G(y_{k,i} - \hat{y}_{k,i}). \quad (3.7)$$

In Example 3.1 where the deviations are noise, it is appropriate to use a low pass filter. With the use of a low pass filter, the threshold could be lowered to increase the possibility to detect small faults.

### 3.3.2 Likelihood Based

For the presentation of the likelihood based test an explanation of the hypothesis concept is needed. Let a hypothesis  $H_i$  be defined by  $H_i^0$  and  $H_i^1$  as

$$H_i^0 : F \in S_i \quad (3.8)$$

$$H_i^1 : F \notin S_i \quad (3.9)$$

where  $S_i$  is a set of faults. The null hypothesis  $H_i^0$  can only be rejected, not accepted. If it is rejected then  $H_i^1$  is accepted, e.g. if a residual based test quantity  $T_i$  has been made based on the hypothesis  $H_i$  and responds, it has responded due to a fault that is not in the set  $S_i$  and the null hypothesis is then rejected.

When statistical information is available (based on a hypothesis) in the form of the likelihood defined in (2.10), a likelihood ratio test can be used. With the notation  $L(H_i) = p(y_k|Y_{k-1}, H_i)$ , given that  $p(y_k|Y_{k-1})$  is based on a model with the hypothesis  $H_i$ , i.e. the model includes the set of faults  $S_i$ , a test quantity can be written as

$$T = \log \left( \frac{L(H_i)}{L(H_j)} \right). \quad (3.10)$$

The test quantity (3.10) is referred to as a  $\log^1$ -likelihood ratio test, which will respond with positive values indicating that it is more likely that  $H_i$  is true and with negative values in the favor of  $H_j$ .

In comparison to a residual based test, the likelihood based test must have information from two models, one including the faults in  $S_i$  and the other including the faults in  $S_j$ .

As in using a residual based test, it can be wise to filter the log-likelihood ratio test. Uncertainties in the models, which is based on a hypothesis, can be inexact and lead to noisy behavior. In the case when the model contains feedback from the process, measurement noise also affects the likelihood.

### 3.3.3 Cusum Test

The cusum test is in this thesis based on either a residual or a log-likelihood ratio, and corresponds to the filter  $G(\cdot)$  in (3.7). For more information about the cusum test than presented here, see [8].

---

<sup>1</sup>The logarithm is used simply based on some issues regarding the realization. An ordinary likelihood test could be used with similar performance.

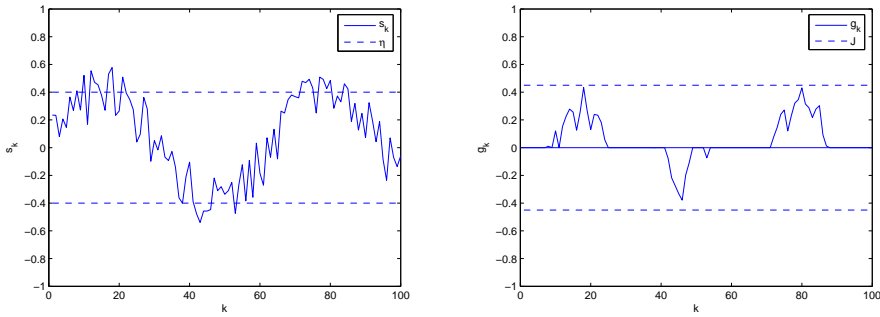
A one-sided cusum test is constructed for an arbitrary signal  $s_k$  (in this thesis,  $s_k$  is equal to a residual or a log-likelihood ratio) according to

$$g_k = g_{k-1} + s_k - \eta, \quad (3.11a)$$

$$g_k = 0, \text{ if } g_k < 0, \quad (3.11b)$$

and alarm when the sum  $g_k$  exceeds a threshold  $J$ . For a residual based test quantity the cusum test is two-sided, i.e. another one-sided test, for negative faults, is used in parallel.

The parameter  $\eta$  is called a drift parameter and is a constant used to compensate for model errors and is used similarly to a constant threshold, see Section 3.2. The sum  $g_k$  works as a low pass filter and the threshold  $J$  must be high enough so the test does not alarm in the fault free case. An example on how  $\eta$  and  $J$  are tuned in a two-sided cusum test, for a fault free signal  $s_k$  with model errors, can be seen in Figure 3.6.



(a)  $s_k$  with a constant drifting parameter  $\eta$ .

(b)  $g_k$  with a constant threshold  $J$ .

**Figure 3.6.** Consider a fault free system not in a stationary point with model errors and noise. In Plot (a) the drifting parameter is set to compensate for the largest model errors in the signal  $s_k$  and in Plot (b), the sum  $g_k$  is thresholded with  $J$  to avoid alarms.

## 3.4 Power Function

If the null hypothesis  $H_i^0$  is rejected when it in reality is true, it is called a false alarm, i.e. a test quantity  $T_i$  has crossed a threshold even though a fault has *not* occurred. In the other case of when a fault has occurred and a test quantity  $T_i$  does not respond, it is called missed detection.

The probability of false alarm and missed detection is used to evaluate tests. The probability of detection often rises with the size of the fault and are wanted as low as possible when a fault has not occurred. The probabilities of false alarm and missed detection can be described by the power function defined as the probability to reject  $H^0$  given a specific value on  $\theta$

$$\beta(\theta) = \Pr(\text{reject } H^0 \mid \theta). \quad (3.12)$$

---

**Example 3.3: Power Function**


---

Consider a residual based test quantity  $T_1$  based on the hypothesis  $H_1$  according to

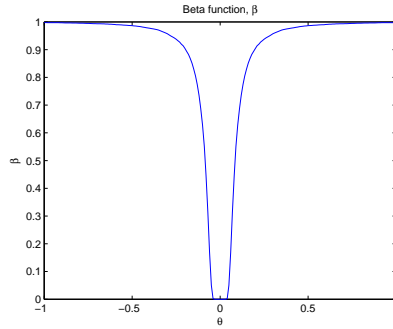
$$H_1^0 : \theta = \theta_0 \quad \text{No fault, } F = NF$$

$$H_1^1 : \theta \neq \theta_0 \quad \text{Faulty, } F = F1$$

with an upper threshold  $J_1$  and a lower threshold  $J_2$ . The parameter  $\theta$  is an arbitrary parameter in the system that can deviate from its nominal value  $\theta_0$ . For this system, the power function (3.12) corresponds to

$$\beta(\theta) = \Pr(J_1 < T, J_2 > T \mid \theta), \quad (3.13)$$

and depended on the system (that is not presented here), a typical power function is presented in Figure 3.7.



**Figure 3.7.** A typical power function  $\beta(\theta)$  where the nominal value for  $\theta$  in the null hypothesis is zero.

---



# Chapter 4

## Engine Models

Two available models of a Scania truck engine are considered in this chapter. One of them is from Scania [9], implemented in SIMULINK. The other one is developed by Johan Wahlström at Vehicular systems at Linköpings universitet [14] and is available as a MATLAB script. The models will be evaluated against each other based on the accuracy of the state prediction and the execution time.

For the application of the PF and the EKF to be possible, the models have to be available as scripts, i.e. not as a SIMULINK scheme. For evaluation of the two models, the model from Scania is converted into a MATLAB script. The data used for the evaluation is fault free, and the models are compared with the application of the filters in mind.

With an accurate model that gives a good state prediction, the dependence on the filters is decreased and estimation and diagnosis performance are gained. Therefore the model accuracy is important, but so is the execution time, shorter execution time can be used for e.g. increasing the number of particles and help in practical issues regarding simulation and implementation.

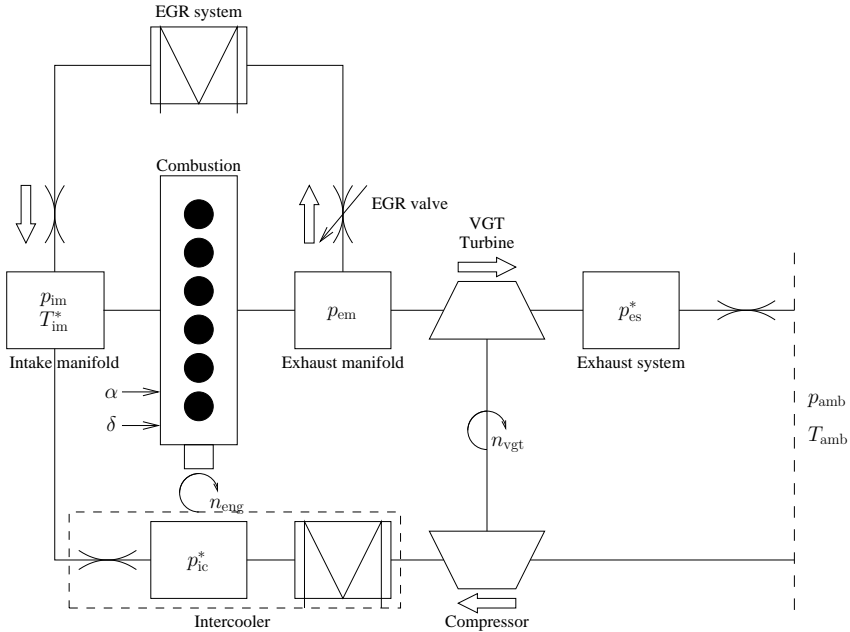
Both models represent a continuous system with a nonlinear function denoted  $g(x(t), u(t))$ , where  $x(t)$  is the state,  $u(t)$  is the input signal and  $t$  is the time. The system can be written in the form

$$\dot{x}(t) = g(x(t), u(t)), \tag{4.1a}$$

$$y(t) = Cx(t), \tag{4.1b}$$

where  $y(t)$  is the measurement which relates to states through the constant matrix  $C$ . To be able to apply the filters to the models, the models have to be in the form of (2.1), i.e. the models have to be discretized in time.

Section 4.1 gives a short overview of which engine the models represent and the method used for the time discretization of the models is presented in Section 4.2. The comparison and evaluation, as well as the selection of which model best suited for the diagnosis systems construction in Chapter 5, takes place in Section 4.3. The last section, Section 4.3.4, gives more details about the selected model and the modeled properties.



**Figure 4.1.** A sketch of a Scania diesel engine with VGT and EGR. The parameters in the boxes are the modeled dynamics, a \* marks that the parameter is *not* modeled in the 3-state model explained in Section 4.3, and generally, just some of the engine properties are modeled and therefore presented here.  $A \Rightarrow$  marks an one-way gas flow and a short description of the parameters in the sketch can be found in Table 4.1.

## 4.1 Engine with VGT and EGR

The studied Scania truck engine is a diesel engine with a variable geometry turbocharger (VGT) and exhaust gas recirculation (EGR). A sketch of the modeled properties in the engine are presented in Figure 4.1 and a short description of the parameters used can be found in Table 4.1. The EGR system is for reducing the  $NO_x$  emissions by increasing the fraction of recirculated emission in the intake manifold and the VGT is simply a fuel efficient way to increase engine power. Both the EGR system and the VGT introduce complexity to the system resulting in the feedbacks (loops) seen in Figure 4.1, which generally affects the diagnosis system performance for the systems constructed in Chapter 5 negatively.

## 4.2 Time Discretization

The easiest way to discretize the system defined by (4.1) is to approximate the derivative with the Euler forward method as

$$\dot{x}(t) \approx \frac{x_{k+1} - x_k}{T_d}, \quad (4.2)$$

**Table 4.1.** Engine parameters important for the representation of the models

Parameter	Description
$p_{\text{amb}}$	Ambient pressure
$p_{\text{im}}$	Intake manifold pressure
$p_{\text{em}}$	Exhaust manifold pressure
$p_{\text{ic}}$	Intercooler pressure
$p_{\text{es}}$	Exhaust system pressure
$T_{\text{amb}}$	Ambient temperature
$T_{\text{im}}$	Intake manifold temperature
$n_{\text{eng}}$	Rotational engine speed
$n_{\text{vgt}}$	Rotational compressor speed
$u_{\text{egr}}$	EGR control signal
$u_{\text{vgt}}$	VGT control signal
$u_{\delta}$	Injected amount of fuel
$u_{\alpha}$	Injection angle

where the step length  $T_d$  is the time between time  $k$  and  $k+1$ . The discrete system will then be

$$x_{k+1} = x_k + T_d g(x_k, u_k), \quad (4.3a)$$

$$y_k = C x_k. \quad (4.3b)$$

There are several other methods for discretization but Euler forward (4.2), with an appropriate step length  $T_d$ , is sufficient for the application in this thesis.

## 4.3 Model Overview

This section gives a short presentation of the two available models, along with a comparison based on the issues for the application in this thesis. For more information about the models, see [9] and [14].

### 4.3.1 6-state Model

The model made by Scania is implemented in SIMULINK and is a continuous state space model with six states represented by the state variable

$$x = (p_{\text{im}}, T_{\text{im}}, p_{\text{em}}, p_{\text{ic}}, n_{\text{vgt}}, p_{\text{es}})^T, \quad (4.4)$$

the input variable

$$u = (n_{\text{eng}}, u_{\text{egr}}, u_{\text{vgt}}, u_{\delta}, u_{\alpha})^T \quad (4.5)$$

and the measurement

$$y = (p_{\text{im}}, T_{\text{im}}, p_{\text{em}}, n_{\text{trb}})^T. \quad (4.6)$$

The nonlinear function  $g(x(t), u(t))$  in (4.1a), for the 6-state model, contains several maps (look-up-tables) of complex engine functions where the maps only represent some of the nonlinearities in the engine. Other nonlinearities are e.g. saturations, min/max functions.

All the modeled properties of the engine are represented in Figure 4.1.

### 4.3.2 3-state Model

The model developed by Johan Wahlström at Vehicular systems at Linköpings universitet is implemented as a Matlab script and is a continuous state space model with three states represented by the state variable

$$x = (p_{\text{im}}, p_{\text{em}}, n_{\text{vgt}})^{\text{T}}, \quad (4.7)$$

the input variable

$$u = (n_{\text{eng}}, u_{\text{egr}}, u_{\text{vgt}}, u_{\delta})^{\text{T}} \quad (4.8)$$

and the measurement

$$y = (p_{\text{im}}, p_{\text{em}}, n_{\text{vgt}})^{\text{T}}. \quad (4.9)$$

This model is like the 6-state model based on physical relations but the parameters is tuned for the states to match measurement data from an ETC<sup>1</sup> cycle of the Scania truck engine. The model is with its three states not representing all the details in Figure 4.1.

The 3-state model, apart from the three states, also include the relation between the control signals  $u_{\text{egr}}$  and  $u_{\text{vgt}}$ , and their actuators  $\tilde{u}_{\text{egr}}$  and  $\tilde{u}_{\text{vgt}}$ . These relations are modeled as first order dynamics

$$\dot{\tilde{u}}_{\text{egr}} = \frac{1}{\tau_{\text{egr}}}(u_{\text{egr}} - \tilde{u}_{\text{egr}}), \quad (4.10)$$

$$\dot{\tilde{u}}_{\text{vgt}} = \frac{1}{\tau_{\text{vgt}}}(u_{\text{vgt}} - \tilde{u}_{\text{vgt}}), \quad (4.11)$$

where  $\tau_{\text{egr}}$  and  $\tau_{\text{vgt}}$  are time constants. Due to the simple dynamics, where  $\tilde{u}_{\text{egr}}$  and  $\tilde{u}_{\text{vgt}}$  are the only variables in the model that are affected by the inputs  $u_{\text{egr}}$  and  $u_{\text{vgt}}$  respectively, they are not considered as state variables. Therefore  $\tilde{u}_{\text{egr}}$  and  $\tilde{u}_{\text{vgt}}$  are viewed as inputs directly affecting the system, instead of  $u_{\text{egr}}$  and  $u_{\text{vgt}}$ . The result is that modeled dynamics are pushed outside the model, and the states are instead considered as inputs.

### 4.3.3 Comparison

The main difference between the two models is the way they model the engine properties. The parameters in the 6-state model represent actual physical properties in the engine, for instance is the parameter representing the intake manifold

---

<sup>1</sup>European Transients Cycle, a test cycle for emission certification of heavy-duty diesel engines.

volume  $V_{\text{im}}$  set to match the actual intake manifold volume in the engine. In the 3-state model, the system has been reduced and does not represent all the dynamics in the engine and a couple of the parameters has been altered as compensation, e.g. the parameter  $V_{\text{im}}$  also exist but do not represent the actual volume in the intake manifold.

The execution time of each discretized model depends on both the complexity of the model and the step length  $T_d$  used in the discretization. A too large  $T_d$  can make the system unstable. For information on Euler forward and stability, see [2].

The input and measurement data used for the evaluation is collected with a sampling time  $T_s = 0.01$  s. Trying with  $T_d = T_s$  for the 3-state model shows no sign of divergence problems in the predictions and the execution time is acceptable. For the 6-state model it is necessary to use the maximum step length of  $T_d = 0.002$  to avoid divergence.

When the same step length  $T_d = 0.002$  s is used for both models, the 6-state model shows bad performance in execution time compared to the 3-state model. Most of the time is used by look-up-tables functions interpolating in the maps. Here it is important to note that the implementation in SIMULINK outperforms the implementation in MATLAB greatly, both using an Euler solver. SIMULINK uses different functions for the interpolation which shortens the execution time. The SIMULINK functions are sadly of no use for the PF application.

It is more difficult to compare the state predictions (model accuracy) between the models. The 3-state model gives a more accurate state prediction during transients, and both models give approximately equal results during stationarity with. The states  $p_{\text{im}}$ ,  $p_{\text{em}}$  and  $n_{\text{vgt}}$  are shown in Figure 4.2 for a  $20T_s$  prediction with  $T_s = 0.002$ . When changing  $T_s$  to 0.01 for the 3-state model, the result becomes worse but still comparable to the result of the 6-state model with  $T_s = 0.002$ . In the figure, it is seen that the 3-state model gives better state prediction when using the same  $T_s = 0.002$ .

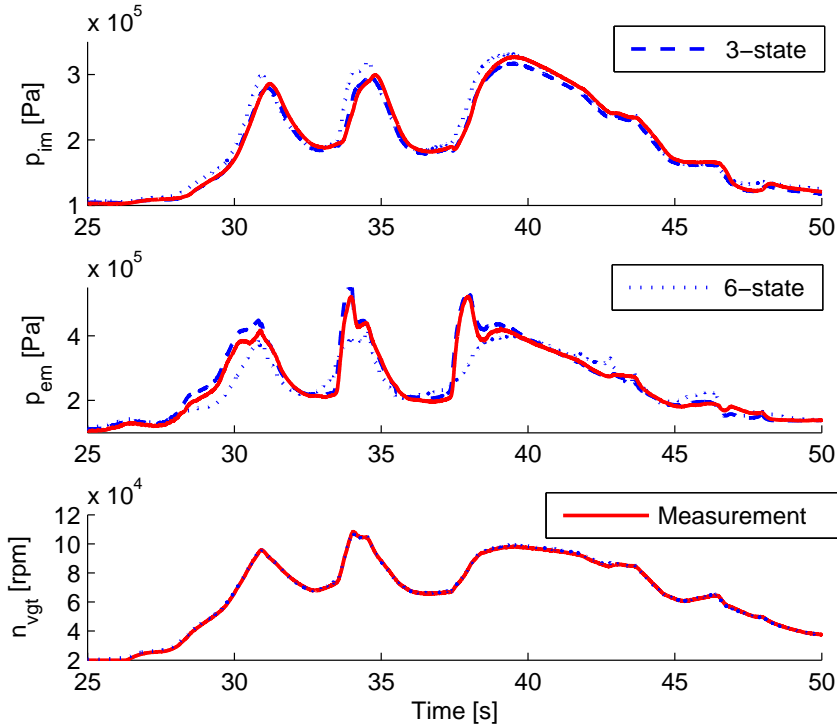
Another important issue regarding the execution time is when the PF is applied to the model. The PF performance will depend on the number of particles used, and the number of particles needed is highly dependent on the dimension of the state vector in the model. The 3-state model having only three states compared to that of six, gives that a smaller number of particles can be used for the same performance. The total execution time for one time step is proportional to the number of particles used times the execution time of the model.

Of the two available models, the 3-state model is chosen to be used in the construction of the diagnosis systems. This decision is based on the issues discussed in this section and are summarized to:

- The 3-state model has half the number of states than the 6-state model has, which should lower the number of particles needed substantially.
- The actual execution time of the model is much faster, which directly decreases the computational power needed.
- The discretization step length  $T_d$  can be set five times higher leading to five times less simulation time is needed. The execution time of the model is

wanted as low as possible because the actual simulation time for the evaluation should not practically interfere with the diagnosis systems construction.

- The 3-state model has generally more accurate state predictions, which is maybe the most important argument for using it.



**Figure 4.2.** The states  $p_{im}$ ,  $p_{em}$  and  $n_{vgt}$  during a part of an ETC cycle for the 6-state model (dotted) and the 3-state mode (dashed). The states are presented for a  $20T_s$  prediction with  $T_s = 0.002$  and plotted with real measurement data (line).

#### 4.3.4 3-state Model Equations

For a better understanding of the model complexity, the model equations for the 3-state model are presented in this section. Explanations for the origin of the equations are given but very briefly and with no motivations. For a more extensive description of the equations, see [14].

The model is slightly modified compared to the original in [14], but the modifications are quite small<sup>2</sup> and the equations presented here are the same as those

<sup>2</sup>A few equations has been excluded because they do not contribute to the state modeling that is important in this study.

**Table 4.2.** Equation parameters used for the representation of the 3-state model.

Parameter	Description	Unit
$W_c, W_{egr}, W_{ei}, W_{eo}, W_t, W_f$	Massflow	kg/s
$A_{egr}, A_{vgt}$	Area	m <sup>2</sup>
$p_{im}, p_{em}, p_{amb}$	Pressure	Pa
$P_t, P_c$	Power	W
$T_{im}, T_{em}, T_{amb}$	Temperature	K
$V_{im}, V_{em}, V_d$	Volume	m <sup>3</sup>
$\eta_{ig}, \eta_{igch}, \eta_m, \eta_{tm}, \eta_c$	Efficiency	-
$n_{cyl}$	Number of cylinders	-
$n_e, n_t$	Rotational speed	rpm
$R_a, R_e$	Gas constant	J/(kgK)
$R_t, R_c$	Radius	m
$a, c$	Constants	-
$x$	Fraction	-
$u_{egr}, u_{vgt}$	EGR, VGT input	%
$u_\delta$	Injected fuel	mg/cycle
$q_{HV}$	Heating value	J/kg
$q_{in}$	Energy in to the cylinders	J
$\gamma_a, \gamma_e, \gamma_{cyl}$	Specific heat capacity ratio	-
$r_c$	Compression ratio	-
$\Pi_{egr}, \Pi_e, \Pi_t, \Pi_c$	Pressure ratio	-
$M_e, M_{ig}, M_p, M_{fric}$	Torque	Nm
$\omega_t$	Turbine angular speed	rad/s
$J_t$	Turbine Inertia	kgm <sup>2</sup>
$\tau_{egr}, \tau_{vgt}$	Time constant	s
BSR	Blade speed ratio	-
pow $\pi$	Exponent	-
$\Phi_c$	Volumetric flow coefficient	-
$\Psi_{egr}, \Psi_c$	Energy transfer coefficient	-

used in the original model. The most important parameters used for the presentation of the equations are briefly explained in Table 4.2.

Several equations contain complex nonlinearities and the use of a demanding nonlinear estimation method like the PF will hopefully be justified.

## Manifolds

The dynamics in the intake manifold and the exhaust manifold are represented by their respective pressure,  $p_{im}$  and  $p_{em}$ , which constitute two of the three states in the 3-state model. The models are based on a standard isothermal model based

on mass conservation together with a constant intake manifold temperature  $T_{im}$ .

$$\dot{p}_{im} = \frac{R_a T_{im}}{V_{im}} (W_c + W_{egr} - W_{ei})$$

$$\dot{p}_{em} = \frac{R_e T_{em}}{V_{em}} (W_{eo} - W_t - W_{egr})$$

## Cylinders

The following equations represent the mass flow in/out of the cylinders which can be observed in Figure 4.1 as the line in/out of the combustion block. The mass flow into the engine is denoted  $W_{ei}$  and is a function of the engine speed, the intake manifold pressure and volumetric efficiency  $\eta_{vol}$ . The volumetric efficiency is a measurement on how effective the cylinders can be filled with air.

$$W_{ei} = \frac{\eta_{vol} p_{im} n_e V_d}{120 R_a T_{im}}$$

$$\eta_{vol} = c_{vol1} \sqrt{p_{im}} + c_{vol2} \sqrt{n_e} + c_{vol3}$$

$$W_f = \frac{10^{-6}}{120} u_\delta n_e n_{cyl}$$

$$W_{eo} = W_f + W_{ei}.$$

The exhaust manifold temperature  $T_{em}$  is a function of the cylinder out temperature  $T_e$  according to

$$T_{em} = T_{amb} + (T_e - T_{amb}) e^{-\frac{h_{tot} \pi d_{pipe} l_{pipe} n_{pipe}}{W_{eo}} c_{pe}}.$$

The cylinder output temperature  $T_e$  is modeled based on a ideal Seliger cycle together with a couple of other equations they represent the exhaust manifold temperature  $T_{em}$ . These equations are nonlinear and dependent on each other and therefore solved numerically using fixed point iteration with the initial values  $x_{r,0}$  and  $T_{1,0}$ .

$$q_{in,j+1} = \frac{W_f q_{HV}}{W_{ei} + W_f} (1 - x_{r,j})$$

$$x_{p,j+1} = 1 + \frac{q_{in,j+1} x_{cv}}{c_{va} T_{1,j} r_c^{\gamma_a - 1}}$$

$$x_{v,j+1} = 1 + \frac{q_{in,j+1} (1 - x_{cv})}{c_{pa} \left( \frac{q_{in,j+1} x_{cv}}{c_{va}} + T_{1,j} r_c^{\gamma_a - 1} \right)}$$

$$x_{r,j+1} = \frac{\Pi_e^{1/\gamma_a} x_{p,j+1}}{r_c x_{v,j+1}}$$

$$T_{e,j+1} = \eta_{sc} \Pi_e^{1-1/\gamma_a} r_c^{1-\gamma_a} x_{p,j+1}^{1/\gamma_a - 1} \left( q_{in,j+1} \left( \frac{1 - x_{cv}}{c_{pa}} + \frac{x_{cv}}{c_{va}} \right) + T_{1,j} r_c^{\gamma_a - 1} \right)$$

$$T_{1,j+1} = x_{r,j+1} T_{e,j+1} + (1 - x_{r,j+1}) T_{im}.$$



The cylinder torque, i.e. the engine torque  $M_e$ , is modeled with three components where the torque  $M_{ig}$  is the gross indicated work that is coupled to the energy in the fuel. The torque loss are due to the friction torque  $M_{fric}$  and the pumping losses  $M_p$ , which is a function of the pressure differences. The engine torque is computed according to

$$\begin{aligned}
 M_e &= M_{ig} - M_p - M_{fric} \\
 M_p &= \frac{V_d}{4\pi} (p_{em} - p_{im}) \\
 M_{ig} &= \frac{u_\delta 10^{-6} n_{cy1} q_{HV} \eta_{ig}}{4\pi} \\
 \eta_{ig} &= \eta_{igch} \left( 1 - \frac{1}{r^{\gamma_{cy1}-1}} \right) \\
 M_{fric} &= \frac{V_d}{4\pi} 10^5 (c_{fric1} n_{eratio}^2 + c_{fric2} n_{eratio} + c_{fric3}) \\
 n_{eratio} &= \frac{n_e}{100}.
 \end{aligned}$$

### EGR valve

The mass flow through the EGR valve is modeled as a simplification of a compressible one way flow where some of the parameters representing physical properties has been replaced with tuning parameters. With the EGR cooler equations and  $\tilde{u}_{egr}$  not considered as a state, the following equations are presented

$$\begin{aligned}
 W_{egr} &= \frac{A_{egr} p_{em} \Psi_{egr}}{\sqrt{T_{em} R_e}} \\
 \Psi_{egr} &= 1 - \left( \frac{1 - \Pi_{egr}}{1 - \Pi_{egropt}} - 1 \right) \\
 \Pi_{egr} &= \begin{cases} \Pi_{egropt} & \text{if } \frac{p_{im}}{p_{em}} < \Pi_{egropt} \\ \frac{p_{im}}{p_{em}} & \text{if } \Pi_{egropt} \leq \frac{p_{im}}{p_{em}} \leq 1 \\ 1 & \text{if } 1 < \frac{p_{im}}{p_{em}} \end{cases} \\
 A_{egr} &= A_{egrmax} f_{egr}(\tilde{u}_{egr}) \\
 f_{egr}(\tilde{u}_{egr}) &= \begin{cases} c_{egr1} \tilde{u}_{egr}^2 + c_{egr2} \tilde{u}_{egr} + c_{egr3} & \text{if } \tilde{u}_{egr} \leq -\frac{c_{egr2}}{2c_{egr1}} \\ c_{egr3} - \frac{c_{egr2}^2}{4c_{egr1}} & \text{if } \tilde{u}_{egr} > -\frac{c_{egr2}}{2c_{egr1}} \end{cases} \\
 \dot{\tilde{u}}_{egr} &= \frac{1}{\tau_{egr}} (u_{egr} - \tilde{u}_{egr}).
 \end{aligned}$$

### Turbo

The third state is the turbine speed,  $\omega_t = n_{vgt}$ , which is modeled using Newton's second law of motion as

$$\dot{\omega}_t = \frac{P_t \eta_m - P_c}{J_t \omega_t},$$

where  $\eta_t$  is the turbine efficiency which is one of the nonlinear parameters modeled using a map in the 6-state model. The turbine efficiency are in the 3-state model represented by the following equations

$$\begin{aligned}
 P_t \eta_m &= \eta_{tm} W_t c_{pe} T_{em} \left(1 - \Pi^{1-1/\gamma_e}\right) \\
 \Pi_t &= \frac{p_{amb}}{p_{em}} \\
 \eta_{tm} &= \eta_{tm,max} - c_m (\text{BSR} - \text{BSR}_{opt})^2 \\
 \text{BSR} &= \frac{r_t \omega_t}{\sqrt{2 c_{pe} T_{em} (1 - \Pi^{1-1/\gamma_e})}} \\
 c_m &= c_{m1} (\omega_t - c_{m2})^{c_{m3}}.
 \end{aligned}$$

The turbine mass flow model equations with  $\tilde{u}_{vgt}$  not considered as a state are given by

$$\begin{aligned}
 W_t &= \frac{A_{vgtmax} p_{em} f_{\Pi_t}(\Pi_t) f_{vgt}(\tilde{u}_{vgt})}{\sqrt{T_{em}}} \\
 f_{\Pi_t}(\Pi_t) &= \sqrt{1 - \Pi_t^{K_t}} \\
 f_{vgt}(\tilde{u}_{vgt}) &= c_{f2} + c_{f1} \sqrt{1 - \left(\frac{\tilde{u}_{vgt} - c_{vgt2}}{c_{vgt1}}\right)^2} \\
 \dot{\tilde{u}}_{vgt} &= \frac{1}{\tau_{vgt}} (u_{vgt} - \tilde{u}_{vgt})
 \end{aligned}$$

and are also one of the functions represented by a map in the 6-state model. The compressor efficiency  $\eta_c$  are computed similarly to the turbine efficiency as

$$\begin{aligned}
 P_c &= \frac{W_c c_{pa} T_{amb}}{\eta_c} \left(\Pi_c^{1-1/\gamma_a} - 1\right) \\
 \Pi_c &= \frac{p_{im}}{p_{amb}} \\
 \eta_c &= \eta_{cmax} - X^T Q_c X \\
 X &= \begin{bmatrix} W_c - W_{copt} \\ \pi_c - \pi_{copt} \end{bmatrix} \\
 \pi_c &= (\Pi_c - 1)^{\text{pow}_\pi} \\
 Q_c &= \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}.
 \end{aligned}$$

Last, the mass flow through the compressor  $W_c$  (which is mapped in the 6-state model), is computed as

$$W_c = \frac{p_{\text{amb}} \pi R_c^3 \omega_t}{R_a T_{\text{amb}}} \Phi_c$$

$$\Phi_c = \sqrt{\frac{1 - c_{\Psi 1} (\Psi_a - c_{\Psi 2})^2}{c_{\Phi 1}}} + c_{\Phi 2}$$

$$\Psi_c = \frac{2c_{\text{pa}} T_{\text{amb}} (\Pi_c^{1-1/\gamma_a} - 1)}{R_c^2 \omega_t^2}$$

$$c_{\Psi 1} = c_{\omega \Psi 1} \omega_t^2 + c_{\omega \Psi 2} \omega_t + c_{\omega \Psi 3}$$

$$c_{\Phi 1} = c_{\omega \Phi 1} \omega_t^2 + c_{\omega \Phi 2} \omega_t + c_{\omega \Phi 3}.$$



# Chapter 5

## Diagnosis Systems Design

This chapter presents how to apply the PF and the EKF to a model of a truck engine to perform diagnosis. The 3-state model is chosen from Chapter 4 and the PF and the EKF is tuned in Section 5.4 for that specific model. The information provided by the filters are used in Section 5.2 to design four different diagnosis systems, two based on the PF and two based on the EKF, that are capable of detecting and isolating each of the considered faults. See Figure 1.1 for a system overview. Some of the constructed tests used by the diagnosis systems require the engine model to be extended with models of the considered faults. The fault models are implemented differently depending on which signal the fault affects, this is discussed in Section 5.3. This chapter is concluded with a couple of additional methods for improving the diagnosis systems performance.

The tests and the diagnosis systems are evaluated with fault free and faulty data from an actual truck engine. The evaluation and discussion take place in Chapter 6.

### 5.1 Considered Faults

In this thesis six faults are considered for isolation, three sensor and three actuator faults. All other possible faults and deviations from the nominal system behavior that do not disappear in the model error are still detectable, but not isolable. This depends on the system architecture, see Figure 4.1, where the three states are observable from each of the sensors because of the feedbacks due to the EGR and VGT. The stated system observability is merely based on observing the system architecture and no formal proof of observability is made.

There is no knowledge about the behavior of the sensors and actuators under faulty conditions. There are therefore many possible types of faults that can be used to evaluate the tests, but they are restricted to gain faults because the evaluation of all types faults would be very time consuming. Gain faults are chosen over additive faults because Scania already has methods for detecting and dealing with additive sensor faults.

Gain faults are here represented by a parameter  $\theta$ , affecting an arbitrary measurement or actuator signal  $s$  according to

$$s_{\text{faulty}} = (1 + \theta)s. \quad (5.1)$$

The faults considered for isolation are denoted as in Table 5.1 and presented along the physical signal they affect (defined in Table 4.1).

**Table 5.1.** The name of the considered faults along with a short explanation and which parameter is affected by the fault.

Fault name	Affect signal	Description
$FS_{\text{pim}}$	$p_{\text{im}}$	Fault in intake manifold pressure sensor
$FS_{\text{pem}}$	$p_{\text{em}}$	Fault in Exhaust manifold pressure sensor
$FS_{\text{vgt}}$	$n_{\text{vgt}}$	Fault in compressor speed sensor
$FU_{\text{egr}}$	$u_{\text{egr}}$	Faulty EGR actuator
$FU_{\text{vgt}}$	$u_{\text{vgt}}$	Faulty VGT actuator
$FU_{\delta}$	$u_{\delta}$	Faulty amount of fuel injected

None of the considered faults are tested on a actual engine for generating new data which is desirable if the possibility existed. For the evaluation of the diagnosis systems in Chapter 6, the faults in Table 5.1 are set to affect the already existing measurement and input data as gain faults according to (5.1).

## 5.2 Diagnosis Systems

In this section two different types of test quantities are constructed, resulting in two independent diagnosis systems. All test quantities are constructed for both the PF and the EKF, resulting in four systems in total. One type of test quantity is residual based and the other is likelihood based and they are constructed based on the same hypotheses. The two resulting types of diagnosis systems, one residual based and one likelihood based, are both able of isolating all faults in Table 5.1.

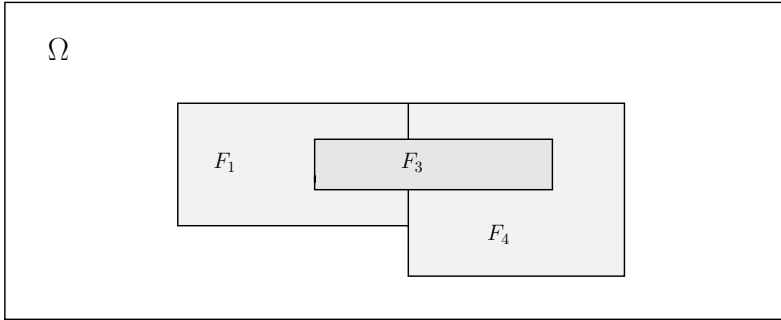
Test quantities for both the residual and likelihood based systems are constructed based on the following hypotheses

$$\begin{array}{ll}
 H_0^0 : F \in S_0 = \{NF\} & H_0^1 : F \notin S_0 \\
 H_1^0 : F \in S_1 = \{NF, FS_{\text{pim}}\} & H_1^1 : F \notin S_1 \\
 H_2^0 : F \in S_2 = \{NF, FS_{\text{pem}}\} & H_2^1 : F \notin S_2 \\
 H_3^0 : F \in S_3 = \{NF, FS_{\text{vgt}}\} & H_3^1 : F \notin S_3 \\
 H_4^0 : F \in S_4 = \{NF, FU_{\text{egr}}\} & H_4^1 : F \notin S_4 \\
 H_5^0 : F \in S_5 = \{NF, FU_{\text{vgt}}\} & H_5^1 : F \notin S_5 \\
 H_6^0 : F \in S_6 = \{NF, FU_{\delta}\} & H_6^1 : F \notin S_6
 \end{array} \quad (5.2)$$

where the hypothesis  $H_0$  represent the 3-state model without fault models. The faults in the other hypotheses 1–6 are modeled according to Section 5.3 or decoupled in another way, e.g. by a disconnected sensor. The thresholds used for each

test quantity are set separately based on the model errors and noise in the fault free case.

Observing the hypotheses in (5.2), one can see that no null hypothesis include more than one fault. To have a test quantity, insensitive to more than one fault, more than one fault has to be modeled (decoupled). A disadvantage with having several modeled faults integrated with the engine model, is that they together may be able to explain another non-modeled fault as illustrated in Figure 5.1. An advantage is that fault isolation becomes easier when only a few test quantities have to respond.



**Figure 5.1.** A system with the set of  $N$  possible faults represented by  $\Omega$  as the behavior of  $\{NF, F_1, \dots, F_N\}$ , where  $NF$  is the no fault state. Fault  $F_1$  and  $F_4$  together, can explain the behavior of fault  $F_3$ . Therefore if  $F_1$  and  $F_4$  are modeled, the system will also be insensitive to  $F_3$ .

### 5.2.1 Residual Based

The residual based test quantities, denoted  $T_i^R$  where  $i = \{0, \dots, 6\}$ , are based on the hypotheses in (5.2) with the same index. The diagnosis system in this section has the following decision structure:

	$FS_{pim}$	$FS_{pem}$	$FS_{vgt}$	$FU_{egr}$	$FU_{vgt}$	$FU_{\delta}$
$T_0^R$	X	X	X	X	X	X
$T_1^R$	0	X	X	X	X	X
$T_2^R$	X	0	X	X	X	X
$T_3^R$	X	X	0	X	X	X
$T_4^R$	X	X	X	0	X	X
$T_5^R$	X	X	X	X	0	X
$T_6^R$	X	X	X	X	X	0

The test quantity  $T_0$  is based on the 3-state model without any fault models, therefore it is sensitive to any measurement deviation, i.e. sensitive to all faults in Table 5.1. The test quantity  $T_0$  is not contributing to the isolation and are simply used for detection of an arbitrary fault. Test quantities  $T_4^R$ ,  $T_5^R$  and  $T_6^R$  have fault models constructed according to Section 5.3. Instead of modeling the sensor faults

in test quantities  $T_1^R$ ,  $T_2^R$  and  $T_3^R$ , one sensor for each test quantity, is disconnected from the system, leaving the system decoupled to that sensor whatever it does. Test quantities 1–3 are still based on the hypotheses in (5.2). An advantage with this type of decoupling compared to modeling faults is that the problem seen in Figure 5.1 can not occur. A disadvantage is that the system loses one measurement signal used for state estimation. This alone does not need to be a problem but if the model is poor, all measurement signals are needed for obtaining a good estimate.

The diagnosis statement from test quantities 0, and 4–6 are each made by using the residuals

$$r_{\text{pim}} = y_{\text{pim}} - \hat{y}_{\text{pim}} = y_{\text{pim}} - \hat{x}_{\text{pim}}, \quad (5.4a)$$

$$r_{\text{pem}} = y_{\text{pem}} - \hat{y}_{\text{pem}} = y_{\text{pem}} - \hat{x}_{\text{pem}}, \quad (5.4b)$$

$$r_{\text{vgt}} = y_{\text{vgt}} - \hat{y}_{\text{vgt}} = y_{\text{vgt}} - \hat{x}_{\text{vgt}}, \quad (5.4c)$$

where  $y$  is a sensor measurement and  $\hat{y} = \hat{x}$  its estimate. The test quantities are said to respond whenever one of the three residuals, filtered with  $G^R(\cdot)$ , exceeds one of its thresholds  $J^R$ . In the fault free case, or in the case where the fault test quantities 4–6 are insensitive to happens, the filtered residuals should be between the thresholds as

$$J_{\text{lower, pim}}^R < G_{\text{pim}}^R(r_{\text{pim}}) < J_{\text{upper, pim}}^R, \quad (5.5a)$$

$$J_{\text{lower, pem}}^R < G_{\text{pem}}^R(r_{\text{pem}}) < J_{\text{upper, pem}}^R, \quad (5.5b)$$

$$J_{\text{lower, vgt}}^R < G_{\text{vgt}}^R(r_{\text{vgt}}) < J_{\text{upper, vgt}}^R. \quad (5.5c)$$

For each test quantity 1–3, only two of the residuals are used whereas the third, the one based on the decoupled sensor, are sensitive to all faults and therefore not contributing to the isolation. Each of the test quantities in this section together with the filter  $G^R(\cdot)$  and the thresholds  $J$  denotes a cusum test as described in Section 3.3.3.

## 5.2.2 Likelihood Based

The likelihood based test quantities are denoted  $T_i^L$  where  $i = \{1, \dots, 6\}$  and together they represent a likelihood based diagnosis system. The test quantities compares the likelihood for each hypothesis  $H_i$  against the hypothesis  $H_0$  according to

$$T_i^L = G_i^L \left( \log \left( \frac{L(H_i)}{L(H_0)} \right) \right) > J_{i, \text{upper}}^L, \quad (5.6)$$

where the  $G_i^L(\cdot)$  is a filter. The likelihoods are obtained directly from the PF or the EKF according to (2.10) and (2.15), respectively. If one of the test quantities  $T_i^L$  respond, it means that it is more likely that hypothesis  $H_i$  is true than  $H_0$ . If more than one test quantity respond, the likelihoods of the responding test quantities are compared between each other to decide the *most* likely hypothesis. The likelihoods are generated as

$$\begin{array}{c|cccccc} & T_1^L & T_2^L & T_3^L & T_4^L & T_5^L & T_6^L \\ \hline F & L(H_1) & L(H_2) & L(H_3) & L(H_4) & L(H_5) & L(H_6) \end{array} \quad (5.7)$$



where  $F$  can be any fault or the no fault state  $NF$ . If e.g.  $F = FS_{\text{pim}}$  leads to that  $T_1^L$  and  $T_3^L$  respond, then  $L(H_1)$  is compared with  $L(H_3)$  to decide the most likely hypothesis. A decision structure can not be presented because the test quantities can not be guaranteed to be insensitive to the non-modeled faults which is due to the problem illustrated in Figure 5.1.

As well as for the likelihood based diagnosis system, the test quantities in this section together with the filter  $G^L(\cdot)$  and the threshold  $J$  denotes a cusum test as described in Section 3.3.3.

### 5.2.3 Tuning the Cusum Parameters

For the residual based tests, the drifting parameter  $\eta$  and the threshold  $J$  are set based on fault free data and have to be set separately for each test used by the PF or the EKF. The model errors increase with increasing state values, and if  $\eta$  is set to the largest model error during stationarity, the tests will have better performance when the system is in those states. States with large errors will thereby contribute to the detection and  $\eta$  is then reflecting the size of largest positive and negative model errors during stationarity. Ways to compensate for the performance loss in states without those errors are presented in Section 5.5.

The threshold  $J$  is used to compensate for noise and non-stationary model errors that are often obtained during transients. Larger errors during transients in the fault free case leads to a larger  $J$  and therefore the ability to detect small faults during a limited time window decreases. This last effect is, especially when using the PF, observed in the power function, see Section 6.1.1. The upper and lower threshold  $J_{\text{upper}}$  and  $J_{\text{lower}}$ , respectively, are set equal in magnitude, i.e.  $|J_{\text{upper}}| = |J_{\text{lower}}|$ , which may not give the best performance. This is done anyway for robustness because of the limited availability of real fault free measurement data.

The drifting parameter  $\eta$  and the threshold  $J$  are for the likelihood based systems are set based on a similar discussion.

The parameters heavily affects the performance of the tests and they are set on the same basis for both the PF based diagnosis systems and the EKF diagnosis systems, see Section 3.3.3 for more information about how the parameters are tuned. The values of the cusum parameters are presented Appendix A.

### 5.2.4 Discussion

When constructing the tests in this chapter, the assumption of single faults is made, i.e. that it is not possible for more than one fault to occur at the same time. This is, under the condition that the faults are independent and based on information from Scania that the considered faults are not likely to occur, a reasonable assumption to make. It is of course a problem if more than one fault actually occurs and the same problem also exist for non-modeled faults.

For making the best possible diagnosis system, the best residual based tests along with the best likelihood tests can be combined. This is however not done, due to lack of time and because the focus lies on evaluating the properties of the

PF against the EKF for FDI. This is also why there are only two types of test quantities used. Other tests than the cusum test can also be used, but it is a time consuming task to find the best suited test for the application in this thesis, and there is really no need for this because the cusum test has the desired properties. The desired properties for a test quantity in this thesis are the possibility to compensate for model errors and noise which in the cusum test corresponds to the parameter  $\eta$  and the sum  $g_k$  with the threshold  $j$ , respectively.

## 5.3 Modeling Faults

Decoupling faults can be done by e.g. modeling the fault dynamics, which is done for several test quantities as a stochastic process according to

$$\theta_{k+1} = \theta_k + \nu_k, \quad (5.8)$$

where  $\nu_k$  is a stochastic variable with a known Gaussian distribution with mean 0 and variance  $D$ . The fault  $\theta$  is set to affect the signal subjected to the fault, as a gain fault according to (5.1).

### 5.3.1 Particle Filter

The faults  $FS_{\text{pim}}$ ,  $FS_{\text{pem}}$ ,  $FS_{\text{vgt}}$  and  $FU_{\delta}$  are modeled as in (5.8). The variance  $D$  of  $\nu_k$  is preferably set large for adaption to fast changes in the fault. If the variance is set too large, the particles will be spread across a large space and the PDF estimate will lose its significance. The variance  $D$  is for the fault models set to

$$\frac{D}{D} \left| \begin{array}{cccc} FS_{\text{pim}} & FS_{\text{pem}} & FS_{\text{vgt}} & FS_{\delta} \\ 4.1 \cdot 10^{-5} & 7.2 \cdot 10^{-4} & 5.1 \cdot 10^{-5} & 1 \cdot 10^{-2} \end{array} \right. \quad (5.9)$$

The faults  $FU_{\text{egr}}$  and  $FU_{\text{vgt}}$  affects the control signal to the EGR and the VGT which has the limited range  $[0, 100]$ . This gives the opportunity to spread the particles uniformly in that interval with no extra state that (5.8) otherwise implies. Doing this, the signals  $u_{\text{egr}}$  and  $u_{\text{vgt}}$  will be completely decoupled and any fault affecting these signals as well.

### 5.3.2 Extended Kalman Filter

The faults  $FS_{\text{pim}}$ ,  $FS_{\text{pem}}$ ,  $FS_{\text{vgt}}$  and  $FU_{\delta}$  are modeled as described in Section 5.3.1 with the variances according to (5.9), and with the exception that (5.1) has to be linearized.

The idea of modeling  $FU_{\text{egr}}$  and  $FU_{\text{vgt}}$  are also the same as in Section 5.3.1 except that a uniform distribution can not be implemented together with the KF equations. Instead, with the fault model

$$\theta_{k+1} = \nu_k, \quad (5.10)$$

a large variance  $D$  is used. Paying no attention to the previous state  $\theta_k$ ,  $\theta_{k+1}$  can still end up outside the range of  $[0, 100]$ . KF estimates of  $\hat{\theta}_k$  outside the interval

are truncated up/down to 0/100 to not exceed the physical boundaries the 3-state model is based on, i.e. the signals  $u_{\text{egr}}$  and  $u_{\text{vgt}}$  are physically confined to the values in  $[1, 100]$ . Using this large variance for a normal distribution is motivated with that all input values in the interval  $[0, 100]$  should have the same probability, and therefore be decoupled. The variance  $D$  are in fault models of  $FU_{\text{egr}}$  and  $FU_{\text{vgt}}$  set to

$$\frac{D}{\left| \begin{array}{cc} FU_{\text{egr}} & FU_{\text{vgt}} \\ 1 \cdot 10^3 & 1 \cdot 10^3 \end{array} \right|} \quad (5.11)$$

which are relatively large compared to the values in (5.9). With the variances in (5.11), the same effect as using the uniform distribution with the PF is obtained.

### 5.3.3 Discussion

The fault models are of the same type as the faults used for the evaluation of the diagnosis systems in Chapter 6. This may not be appropriate when in Section 5.1, saying that there is no knowledge of the actual behavior of a faulty sensor. But when setting the fault variance  $D$ , it is set as high as possible and therefore giving the ability to compensate for other types of faults affecting the system. The fault model (5.8), with a high variance could also handle constant bias faults.

When using a stochastic fault model, as described in Section 5.3, on a model with errors, the errors are to some extent compensated for, i.e. the problem illustrated in Figure 5.1 is imminent. This is because the modeled fault is a free variable in the system, and the filter estimates this variable to best fit the measurement. In fact, due to the EGR and VGT, all faults considered in this thesis can be explained to some extent by another modeled fault. The problem just stated indicates the effectiveness of excluding a sensor for decoupling when possible.

Stochastic modeling in this thesis either increases the number of states or directly affects the system signal as noise. This leads to that the number of particles needed by the PF to maintain the quality of the PDF estimate increases. The result of the modeling is discussed more in detail during the evaluation of the diagnosis systems in Chapter 6.

## 5.4 Tuning the Filters

This section explains how the parameters in the PF and the EKF are adjusted for the engine model. The Bootstrap PF introduced in Chapter 2 does not need many modifications to work with the used 3-state model. There are however a number of parameters that need to be tuned:

- Initial distribution of the states,  $p(x_0)$ .
- The distribution of the system noise,  $p(v_k)$ .
- The distribution of the measurement noise,  $p(n_k)$ .
- The number of particles used,  $N$ .

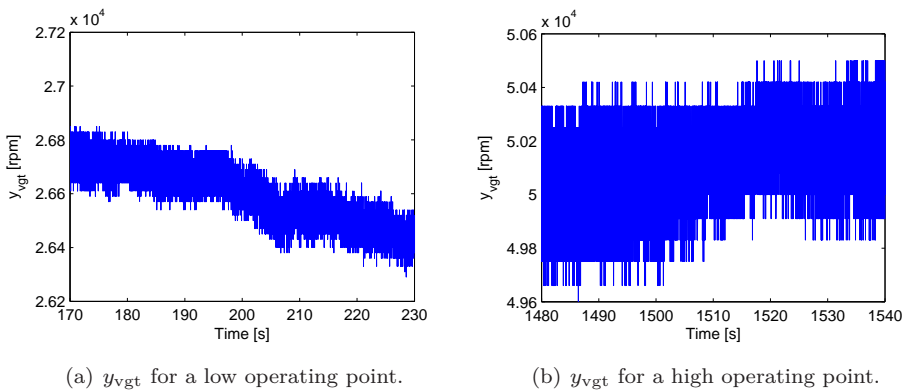
These parameters, except  $N$ , are also needed for the EKF but have the restriction that the distributions have to be Gaussian.

### 5.4.1 Particle Filter

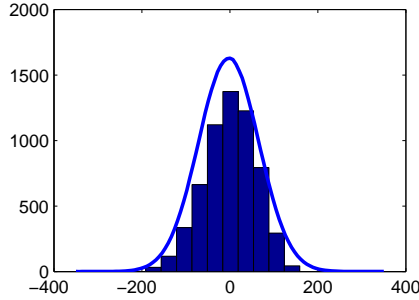
The initial distribution of the state  $p(x_0)$  only represent the knowledge of the state in the first time step and therefore the type of distribution is not of great importance, but it still has to be set in the vicinity of the real state for not losing the estimates completely. The distribution is out of simplicity set to be Gaussian,  $N(m_0, \Pi_0)$ , where  $m_0$  is the initial value of the measurement and  $\Pi_0$  is the covariance matrix with elements of the same size as in the system noise covariance matrix  $Q$ .

It is difficult to approximate the distribution of the system noise and a straightforward way is to assume that the noise has a Gaussian distribution with mean 0 and a constant covariance matrix with only diagonal elements, which is done here. This assumption may cause that unnecessary many particles are needed for obtaining a good PDF estimate. Spreading the particles over a wide space leads to less belief in the model and in that way, decreasing the diagnosis systems performance. The covariance matrix  $Q$  reflects the size of the model errors and therefore it has to be adequately large to handle those. The model errors are highly dependent on the system state which in general, for the 3-state model, gives more inaccurate estimates for larger values of  $x_{\text{pim}}$ ,  $x_{\text{pem}}$  and  $x_{\text{vgt}}$ . This indicates that it would be appropriate to use a  $p(v_k) = N(m(x_k), Q)$  or  $p(v_k) = N(0, Q(x_k))$  dependent on the looks of the model errors. There are though problems to decide the magnitude of  $m(x_k)$  and  $Q(x_k)$  for different operating points and for simplicity a constant  $Q$  and  $m$  are used, but this is still discussed further in Section 5.5.3.

The measurement noise, i.e. the sensor noise, are often considered to be Gaussian when the sensors work correctly. There are indications from measurement data that the variance changes with the state as seen in Figure 5.2 where the measurement  $y_{\text{vgt}}$  is presented from two different operating points. Still, a constant



**Figure 5.2.** The measurement  $y_{\text{vgt}}$  is in plot (a) shown for a low operating point and in plot (b) for a high operating point.



**Figure 5.3.** The same measurement data used in Figure 5.2(a) from the measurement signal  $y_{\text{vgt}}$ , with the linear trend removed, is here plotted as a histogram with an example of a Gaussian distribution of the same variance.

covariance matrix  $R$  is used for simplicity. No information about the sensor properties from the manufacturer is available and therefore the distribution  $p(n_k)$  is set to be Gaussian with mean 0 and constant covariance  $R$ . In Figure 5.3 the same measurement data used in Figure 5.2(a) are here used to support the assumption of a Gaussian distribution.

The elements in  $Q$  and  $R$  are, in lack of better knowledge about the actual noise and errors in the system, in a way considered as design parameters, i.e. parameters to be tuned for satisfactory performance. A couple of assumptions about the sensors can be made on good grounds, e.g. the sensor measuring the compressor speed uses a highly accurate method and therefore the variance should be set low. The sensors measuring the pressure are of the same type and knowledge of the variance could be obtained from measurement data.

Considering the 3-state model accuracy and the above discussion, the elements in  $Q$  are set as low as possible still obtaining a reasonable accurate state estimate. The distributions are based on above arguments set to

$$p(x_0) = N(m_0, \Pi_0), \quad p(v_k) = N(0, Q), \quad p(n_k) = N(0, R), \quad (5.12)$$

where

$$Q = \begin{pmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_3 \end{pmatrix}, \quad R = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_3 \end{pmatrix}, \quad (5.13)$$

and  $\Pi_0 = Q$ . The elements  $Q_1$  and  $Q_2$  are set lower than  $R_1$  and  $R_2$ , respectively. Because of the accuracy of the sensor measuring the turbine speed, element  $R_3$  is lower than  $Q_3$ . This gives more belief in the dynamic equations that represent the states  $x_{\text{pim}}$  and  $x_{\text{pem}}$  than their respective measurement signal, which is good for the diagnosis systems performance because of its reliance to the model. With the

numerical values inserted, the matrices in (5.13) are presented as

$$Q = \begin{pmatrix} 4 \cdot 10^6 & 0 & 0 \\ 0 & 1.6 \cdot 10^7 & 0 \\ 0 & 0 & 1 \cdot 10^4 \end{pmatrix}, R = \begin{pmatrix} 2.5 \cdot 10^7 & 0 & 0 \\ 0 & 2.5 \cdot 10^7 & 0 \\ 0 & 0 & 1 \cdot 10^2 \end{pmatrix}. \quad (5.14)$$

The number of particles  $N$  is important for the PF performance and should be set as high as possible. With too few particles the representation of the system noise, the state PDF and the likelihood will be bad. The number of particles is set to

$$N = 200, \quad (5.15)$$

so that the execution of the filtered model can be done real time on a ordinary computer<sup>1</sup>. This however, may not give the wanted performance and is a compromise due to the execution time.

### 5.4.2 Extended Kalman Filter

The distribution of the initial state along with the assumed distributions for the system and measurement noise, is set equal to those for the PF in Section 5.4.1. This is done based on the same motivations as there, but also for the reason that the PF and EKF should be evaluated under the same conditions. This may not be the best case for getting good performance out of the EKF but makes the comparison between the two filters easier. The FDI performance when using the EKF are here neglected for easier evaluation of the differences.

If the linearization is done according to Section 2.2.3, it is necessary to calculate the matrices (2.17) and (2.18). This can either be done analytically or by using a numerical method. The preferred way to do this would be analytically which gives an exact calculation of the matrices and also saves CPU time. Because of the complex model equations, presented in Section 4.3.4, this is however not done. Instead various different numerical methods were tested, including Euler forward presented in Section 4.2. There were no clear performance enhancements with some of the more complex methods and Euler forward are therefore used.

The linearization together with the time discretization of the model require that Euler forward is used twice. The following calculations are made for computing the system matrix  $A_k$  at every time step,

$$\begin{aligned} \dot{x}(t) &= g(x(t), u(t)) \Rightarrow \text{/Euler Forward/} \Rightarrow \\ x_{k+1} &= T_d g(x_k, u_k) + x_k := f(x_k, u_k) \Rightarrow \\ A_k &= \frac{\partial f(x_k, u_k)}{\partial x_k} = T_d \frac{\partial g(x_k, u_k)}{\partial x_k} + \frac{\partial x_k}{\partial x_k} \end{aligned}$$

leading to

$$A_k = T_d \frac{\partial g(x_k, u_k)}{\partial x_k} + I, \quad (5.16)$$

---

<sup>1</sup>2 GHz Centrino with 1 GB memory.

where  $I$  is the unit matrix. The Euler forward method is here used (again) to approximate the derivative according to

$$\frac{\partial g(x_{k,i}, u_k)}{\partial x_{k,i}} \approx \frac{g(x_{k,i} + he_i, u_k) - g(x_{k,i}, u_k)}{h}, \quad (5.17)$$

where  $e_i = \{(1, 0, 0)^T, (0, 1, 0)^T, (0, 0, 1)^T\}$  and the step length  $h$  was set to 0.001. Other step lengths was tried with no apparent performance enhancements.

The matrix  $B_k$  is in this case not needed for the KF equations, because the prediction of the state  $\hat{x}_{k+1|k}$  in (2.14a) is obtained directly by using (4.3a).

### 5.4.3 Discussion

Using only Gaussian distributions, the PF loses some of its theoretical advantage against the EKF. Another setback for the PF is the limitation of the number of particles used. With the number of particles chosen, which is set for as short execution time as possible without losing the significance of the PDF estimate, the EKF execution still runs faster. This is because the PF runs the 3-state model  $N$  times and the EKF only does that (due to the linearization) 4–5 times, dependent if there is a fault modeled. The execution time for the PF based systems,  $t_{\text{PF}}$ , and the execution time of the EKF based systems,  $t_{\text{EKF}}$  are dependent on the 3-state model execution time  $t_{\text{model}}$  according to

$$t_{\text{PF}} : Nt_{\text{model}}, \quad \text{and} \quad t_{\text{EKF}} : (n + 1)t_{\text{model}} \quad (5.18)$$

where  $n$  is the number of states. This leads to that the only performance advantages for the PF, if there is any, is due to the nonlinearities. The nonlinearities are though expected to be a matter for the estimation performance of the filters, see the model equations in Section 4.3.4.

## 5.5 Compensating for Model Errors

In Section 5.4 it is mentioned that there are model errors affecting both the filter performance and the diagnostics. The model errors are highly dependent on the state and on the engine speed, and therefore it is relatively easy to approximate the model errors as a function of the state and the engine speed.

Two methods of dealing with this problem is described in this section, but only one method for one test quantity is actually implemented. The implementation is done with an adaptive threshold for the test quantity  $T_0^R$  presented in Section 5.2. This is because the methods described in this section are time consuming to implement and the FDI performance is considered as a secondary objective.

### 5.5.1 Estimating the Model Errors

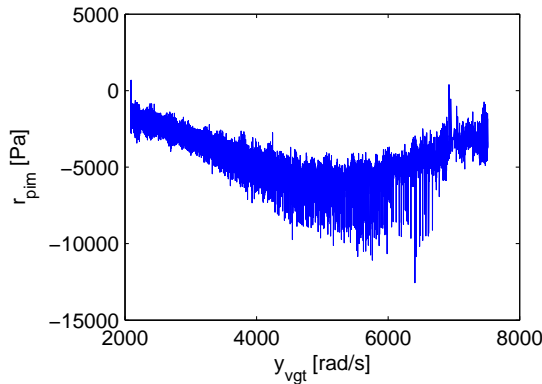
Let the model errors be defined as the residuals in (5.4) according to

$$\Delta x_{\text{pim}} = r_{\text{pim}}, \quad (5.19a)$$

$$\Delta x_{\text{pem}} = r_{\text{pem}}, \quad (5.19b)$$

$$\Delta x_{\text{vgt}} = r_{\text{vgt}}. \quad (5.19c)$$

The model error dependence on the state and the engine speed, is seen by observing the model error for a number of simulations. The model error  $\Delta x_{\text{pim}}$  is in Figure 5.4 plotted against  $y_{\text{vgt}}$  for  $n_{\text{eng}} = 800$  rpm. The mean error is clearly dependent on the VGT speed.



**Figure 5.4.** The model error  $\Delta x_{\text{pim}}$  plotted against  $y_{\text{vgt}}$  with  $n_{\text{eng}} = 800$  rpm for approximately  $3.2 \cdot 10^5$  samples.

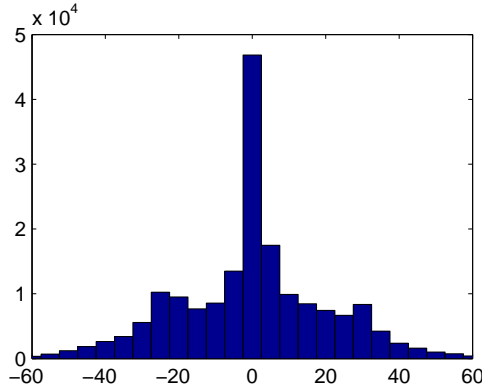
The least square method is used to approximate the mean value of the model error as a function of the state. The data used for the approximation is only from a couple of different engine speeds, where many points from the entire range of stationary operating points are used. The only state that has been used in the approximation is  $x_{\text{vgt}}$  which is the VGT speed. This is because  $\hat{x}_{\text{vgt}}$  is the most accurate state estimate. Other state estimates can also be used but are not expected improve the model error estimates much in comparison to the added complexity.

In fault free simulations,  $\Delta x_{\text{vgt}}$  mostly look like noise and the piecewise mean value, over a time window, is always close to zero. In Figure 5.5,  $\Delta x_{\text{vgt}}$  is presented as a histogram for approximately  $1.8 \cdot 10^5$  samples from an ETC cycle. Therefore the model errors estimations are only done for  $\Delta x_{\text{pim}}$  and  $\Delta x_{\text{pem}}$ . The approximated mean model errors  $\Delta \tilde{x}$  are for each available engine speed value  $r$  computed as

$$\Delta_r \tilde{x}_{\text{pim}} \approx a_{r0} + a_{r1} \hat{x}_{\text{vgt}} + a_{r2} \hat{x}_{\text{vgt}}^2 + \dots + a_{r9} \hat{x}_{\text{vgt}}^9, \quad (5.20a)$$

$$\Delta_r \tilde{x}_{\text{pem}} \approx b_{r0} + b_{r1} \hat{x}_{\text{vgt}} + b_{r2} \hat{x}_{\text{vgt}}^2 + \dots + b_{r9} \hat{x}_{\text{vgt}}^9, \quad (5.20b)$$





**Figure 5.5.** A histogram of  $\Delta x_{\text{vgt}}$  for approximately  $1.8 \cdot 10^5$  samples from an ETC cycle.

where  $\{a_{ri}\}_{i=0}^9$  and  $\{b_{ri}\}_{i=0}^9$  are different constants for each available speed  $r = \{600, 800, 1100, 1300, 1500\}$ . The model error  $\Delta\tilde{x}$  is set to a linear interpolation between the five functions  $\Delta_r\tilde{x}$  for the engine speeds not represented by  $r$ .

### 5.5.2 Adaptive Thresholds

In Section 3.3.3, the drifting parameter  $\eta$  used to compensate for model errors is set to a constant. If this is done for a system with model errors as seen in Figure 3.5, the diagnosis system could lose detection performance.

For a two-sided cusum test the drifting parameters  $\eta_{\text{upper}}$  and  $\eta_{\text{lower}}$  can be set as a function of the mean model error  $\Delta\tilde{x}$  according to

$$\eta_{\text{upper}} = \Delta\tilde{x} + \lambda_{\text{upper}}, \tag{5.21a}$$

$$\eta_{\text{lower}} = \Delta\tilde{x} - \lambda_{\text{lower}}. \tag{5.21b}$$

Here  $\lambda$  could be chosen as a constant or a function. Either way,  $\lambda$  should represent the uncertainty in the model error approximation and are set based on how much  $\Delta x$  deviate from  $\Delta\tilde{x}$ , i.e. based on the variance of  $\Delta x - \Delta\tilde{x}$ . For the implementation of the adaptive threshold,  $\lambda$  is chosen to be a function of the estimated state  $\hat{x}_{\text{vgt}}$  as

$$\lambda_{\text{upper}}(\hat{x}_{\text{vgt}}) = c_{1,\text{upper}}\hat{x}_{\text{vgt}} + c_{2,\text{upper}}, \tag{5.22a}$$

$$\lambda_{\text{lower}}(\hat{x}_{\text{vgt}}) = c_{1,\text{lower}}\hat{x}_{\text{vgt}} + c_{2,\text{lower}}. \tag{5.22b}$$

Setting  $\lambda$  as a function is done ad hoc and the constants  $c$  in the linear function are set entirely based on observations from simulations, e.g. as from the observation seen in Figure 5.4.

### 5.5.3 Adaptive Noise Distribution

Instead of compensating for the model errors when making the diagnostic tests, the errors could be compensated for already when making the PF and the EKF. It is already said in Section 5.4.1 that the size of the elements in  $Q$  should reflect the model errors, but a constant covariance matrix might not always turn out well, e.g. if the model errors have a constant bias, it would be better to give an offset to the distribution rather than increasing the magnitude of the separate elements in  $Q$ . This method implies that the system noise distribution  $p(v_k) = N(0, Q)$  instead should be defined as

$$p(v_k) = N(m, Q), \quad (5.23)$$

where the mean value  $m$  could be set to  $m = \Delta\tilde{x}$ . If the problem is not a offset and instead an changing variance, the separate elements in  $Q$  could be set as function of the model error.

### 5.5.4 Discussion

The implementation of the methods described is only done for an adaptive threshold accordingly, and for the test quantity  $T_0^R$  presented in Section 5.2 using the PF. Using a residual based test rather than a likelihood based test for the implementation is because the likelihood tests use the ratio between two models that are both affected by the model errors, and therefore the model errors should not be as clear as in the residual tests.

The model errors can be almost linear in the states for lose  $n_{\text{eng}}$ , and therefore, instead of making a tenth order least square approximation a smaller order can be used. The performance will decrease lose, but will still increase compared to the use of a constant  $\eta$ . A tenth order is used in first place just to be sure that the nonlinear tendencies are captured in the approximation.

There is also the possibility to use an adaptive  $J_k$  that is used with the sum  $g_k$  in the cusum test. This could give lose performance enhancements because the noise variance also increases with large state values. This increase in variance is however not as legible as the constant model errors that  $\eta$  compensates for.

An adaptive noise distribution also solves another problem with the PF. The number of particles used is not always enough to represent  $Q$  when the model errors is large, particles do not always end up in the tails of the distribution  $Q$  and therefore loses the state PDF estimate.

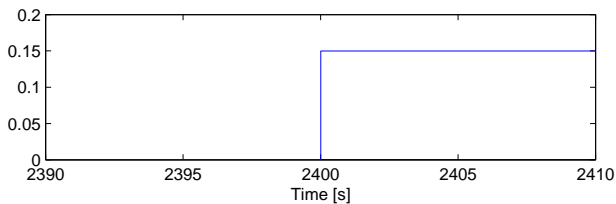
# Chapter 6

## Diagnosis Systems Evaluation

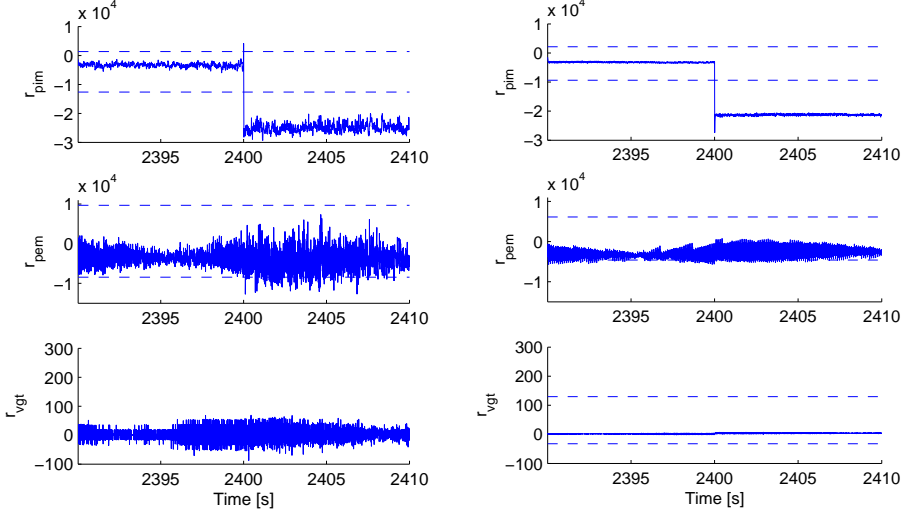
The tests and diagnosis systems constructed in Section 5.2 are evaluated with real fault free measurement data and faulty measurement data with the faults defined in Section 5.1. The systems are compared and discussed based on their properties for FDI. The discussion focuses on the difference between the residual based tests and the likelihood based tests as well as the difference between the use of the PF and the EKF.

### 6.1 Fault Detection

In this section, some of the differences between PF based and EKF based tests are presented. Remember that it is stated in Section 5.4 that both the PF and the EKF have the covariance matrices  $Q$  and  $R$  of same magnitude. Lets first consider the residual based diagnosis systems, and further, let the systems be subjected to a gain fault  $FS_{pim}$  according to Figure 6.1. The PF and EKF diagnosis systems residuals for test quantity  $T_0^R$  during the fault  $FS_{pim}$  are seen in Figure 6.2. The residuals are plotted along with their drifting parameters. By observing Figure 6.2 one can see a couple of differences. First, the variance of the residuals are larger



**Figure 6.1.** A gain fault that has magnitude  $\theta = 0$  up to time 2400, and after that time, a magnitude of  $\theta = 0.15$ .



(a) Using the PF based diagnosis system.

(b) Using the EKF based diagnosis system.

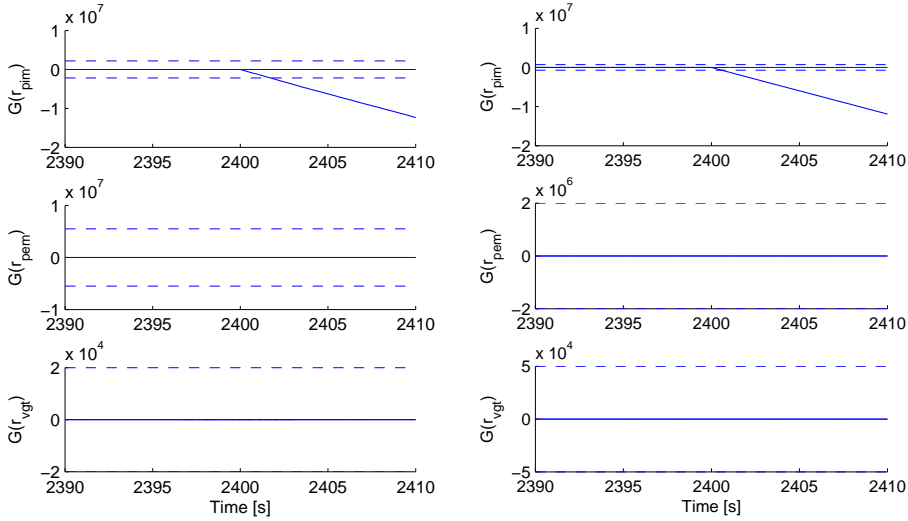
**Figure 6.2.** Plot (a) and (b) shows the residuals in test quantity  $T_0^R$  when the PF based and EKF based diagnosis system in subjected to a gain fault,  $FS_{pim}$  of size  $\theta = 0.15$  at time 2400. The residuals are plotted with its drifting parameters  $\eta_{upper}$  and  $\eta_{lower}$  marked as a dashed line.

when using the PF which is an effect of using a limited number of particles for the estimation. Second, the EKF residual  $r_{pim}$  reacts weaker than the respective PF based residual which is an effect of the EKF linearization. Due to the linearization, some of the fault effect manifest itself in the other residuals which is not seen clearly in Figure 6.2(b), but however clear for larger faults. Because of the model error differences between the two systems (not explicitly presented here), the parameters  $\eta_{pim}$  and  $\eta_{pem}$  are set larger for the PF based system. In this case when the PF reacts stronger, it leads to that the PF based system and the EKF based system detects faults seen in  $r_{pim}$  and  $r_{pem}$  of approximately the same magnitude.

It is however possible to set the drifting parameter  $\eta_{vgt}$  for the PF closer to zero in comparison with the same parameter for the EKF based system. This is a general effect for all the tests and indicates that the PF based diagnosis system is better at detecting small turbine related faults, e.g. a small  $FS_{vgt}$ .

Just because the PF based system and the EKF based system can detect the fault  $FS_{pim}$  of the same magnitude, the detection time, i.e. the time from the occurrence of the fault to the time of detection, is not the same. Observing the the sums  $g_k$  from the cusum test with the respective thresholds  $J$  in Figure 6.3, one can see that  $g_k$  increases in magnitude with approximately the same speed but the detection time is longer for the PF based system mainly due to the lower value of  $J_{lower,pim}$ .

When subjecting the PF and EKF based system to the sensor fault  $FS_{pem}$ ,

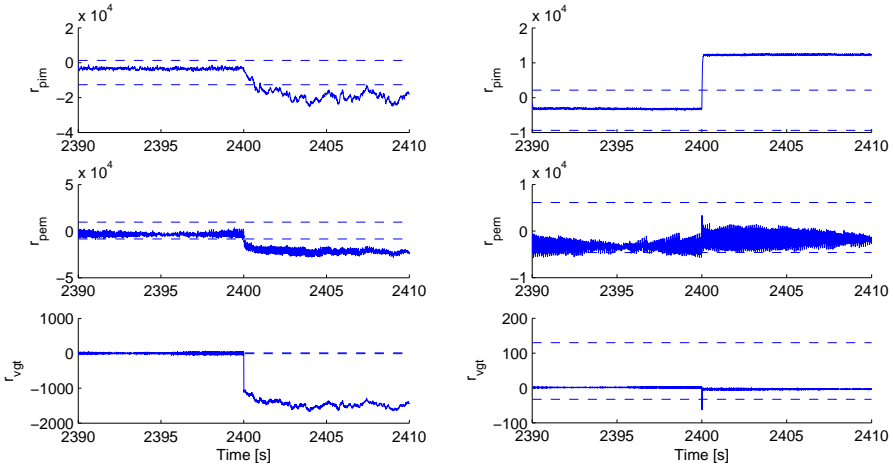


(a) The sum  $g_k$ , using the PF based diagnosis system. (b) The sum  $g_k$ , using the EKF based diagnosis system.

**Figure 6.3.** Plot (a) and (b) shows the sum  $g_k$  from the cusum test  $G^R(\cdot)$  when the PF based and EKF based diagnosis system in subjected to a gain fault,  $FS_{pim}$  of size  $\theta = 0.15$  at time 2400. The sums are plotted with its threshold  $J$  marked as a dashed line. A fault is said to be detected when  $g_k$  crosses  $J$ .

the result is similar to those presented for  $FS_{pim}$ . The appearance in the residuals in test quantity  $T_0$  is also similar during the faults  $FU_{egr}$ ,  $FS_{pem}$  and  $FS_{pim}$  for the two systems. But when subjecting the system to fault  $FS_{vgt}$  according to Figure 6.1, still considering  $T_0$ , an interesting observation is made which is presented in Figure 6.4. In Figure 6.4 it is seen that all residuals using the PF crosses its  $\eta$  and for the EKF, the fault is only detectable in  $r_{pim}$ . This result for the EKF is understandable because that  $Q_3 > R_3$ , leading to that  $y_{vgt}$  is often considered to be correct even under faulty conditions which gives that the fault may be visible in another way. Due to the strong connection between the states  $x_{pim}$  and  $x_{vgt}$  (see Chapter 4), the residual  $r_{pim}$  is the most affected. Note that  $Q_1 < R_1$ .

This behavior just stated for the EKF is not the same for the PF, but explainable. The PF representation of  $Q_3$  is due to the finite number of particles and the small magnitude of  $Q_3$ , quite limited in space. The particles ending up too far away from the measurement give numerical problems in computing the weights, and therefore the track of  $x_{vgt}$  is lost. If the track of  $x_{vgt}$  is lost,  $\hat{x}_{vgt}$  will only depend on the modeled dynamics and the measurements  $y_{pim}$  and  $y_{pem}$ , which results in the effect seen in Figure 6.4. With more particles, the representation of  $Q_3$  should be better. Using 1000 particles instead of 200 still gives this effect but the increased  $N$  is at least clearly seen in the decreasing variance for the state



(a) Using the PF based diagnosis system.

(b) Using the EKF based diagnosis system.

**Figure 6.4.** Plot (a) and (b) shows the residuals in test quantity  $T_0^R$  when the PF based and EKF based diagnosis system is subjected to a gain fault,  $FS_{vgt}$  of size  $\theta = 0.15$  at time 2400. The residuals is plotted with its drifting parameters  $\eta_{upper}$  and  $\eta_{lower}$  marked as a dashed line.

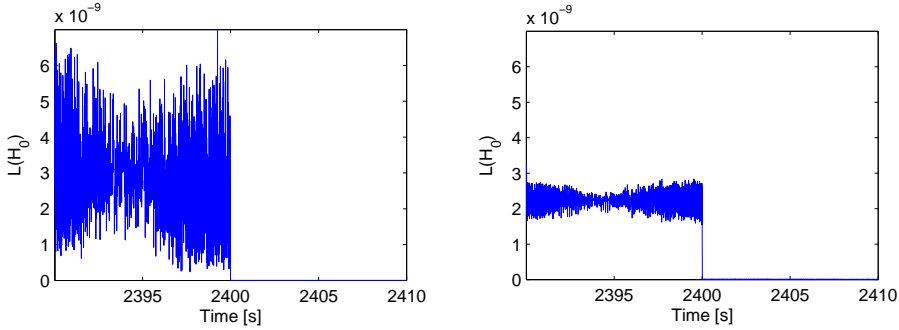
estimates. The EKF still has lower variance in its estimates though, and the question of how many particles that is needed for approximately equal variance, which relates to the representation of the noise distribution, is not answered here. Using a larger number of particles leading to a decreasing variance can of course be used to lower, especially  $J$ , which decreases the detection time.

This effect (just discussed) is due to the limited number of particles used and can explain the stronger reaction seen in Figure 6.2 if it is not due to the EKF linearization, which is not bad in that particular case. The detection time and therefore the robustness of the PF based system is however decreased when the number of particles is low, which is noticed already when setting the cusum parameters.

An interesting note can be made from observing Figure 6.2, especially from Plot (a) with the PF, where almost the entire fault manifest itself in  $r_{pim}$ . This leads to that the fault can be approximated with the residual. The fault size is  $\theta = 0.15$  leading to an absolute value of  $3 \cdot 10^5$ , and the residual jumps  $2 \cdot 10^5$  which is a decent approximation. These observation are not explored further in this thesis.

In general, the residual based diagnosis systems respond quite similar with the largest difference that there are problems for the PF due to the variance in the estimates. Therefore,  $J$  is able to be set lower when using the EKF which result in faster detection.

Now considering the likelihood based systems subjected to the gain fault  $FS_{pim}$  according to Figure 6.1. Observing the likelihood  $L(H_0)$  in Figure 6.5 it is easy

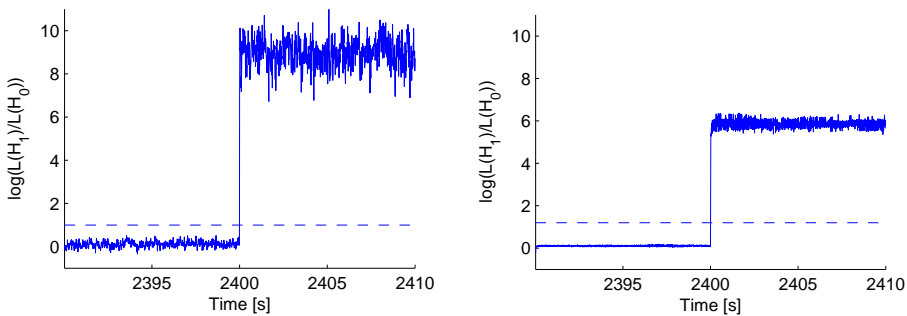


(a) Using the PF based diagnosis system. (b) Using the EKF based diagnosis system.

**Figure 6.5.** Plot (a) and (b) shows the likelihoods  $L(H_0)$  when the PF based and EKF based diagnosis system in subjected to a gain fault,  $FS_{pim}$  of size  $\theta = 0.15$  at time 2400.

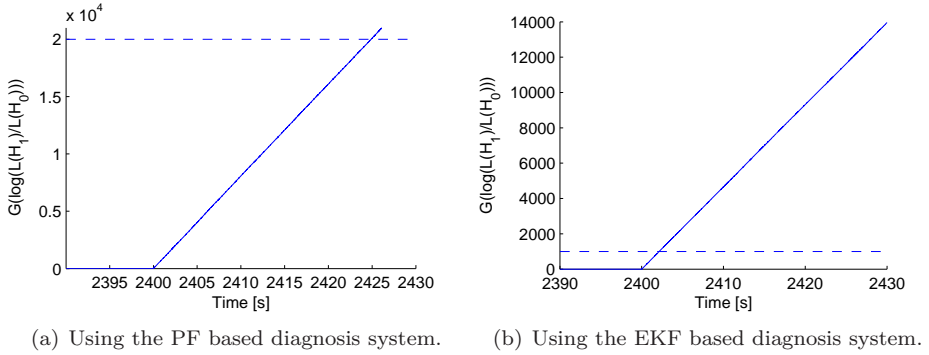
to see that the larger variance observed for the residual based system using the PF, due to limited number of particles, is also seen here. This gives a hint that the likelihood based test quantities have a shorter detection time when using the EKF because of the threshold  $J$ , which can be set low compared to when using the PF. Using more particles decrease the likelihood variance and  $J$  can therefore be lowered. Observing the likelihood closer during the fault free period, one can see that mean likelihood for the time 2390-2400 seems to be higher for the PF based system than for the EKF based system. This indicates that using the PF provides an advantage due to the nonlinearities in the system compared to use of the EKF (the likelihood represent the significance in the estimation).

The log-likelihood ratio  $\log(L(H_1)/L(H_0))$ , used by the test quantity  $T_1^L$ , for fault  $FS_{pim}$  is presented with the drifting parameter  $\eta$  in Figure 6.6. Observing the



(a) Using the PF based diagnosis system. (b) Using the EKF based diagnosis system.

**Figure 6.6.** Plot (a) and (b) shows the log-likelihood ratio  $L(H_1/H_0)$  using the PF and the EKF based diagnosis system, respectively. The systems are subjected to a gain fault,  $FS_{pim}$ , of size  $\theta = 0.15$  at time 2400. The dashed line is the systems respective drifting parameter.



**Figure 6.7.** Plot (a) and (b) shows the sum  $g_k$  for the log-likelihood ratio using when the PF based and EKF based diagnosis system in subjected to a gain fault,  $FS_{\text{pim}}$  of size  $\theta = 0.15$  at time 2400. The sums are plotted with its threshold  $J$  marked as a dashed line. A fault is said to be detected when  $g_k$  crosses  $J$ .

figure it is seen that the likelihood during the fault free period is slightly above zero indicating that hypothesis  $H_1$  is more likely than hypothesis  $H_0$ . When modeling a fault as discussed in Chapter 5.3, the fault model will compensate for the model errors and therefore give this result. Here it is also indicated, as in the residual plots during the fault  $FS_{\text{vgt}}$ , that the PF based system can detect smaller faults. The threshold  $J$  is set lower for the likelihood based system using the EKF, which therefore gives a shorter detection time as seen in Figure 6.7.

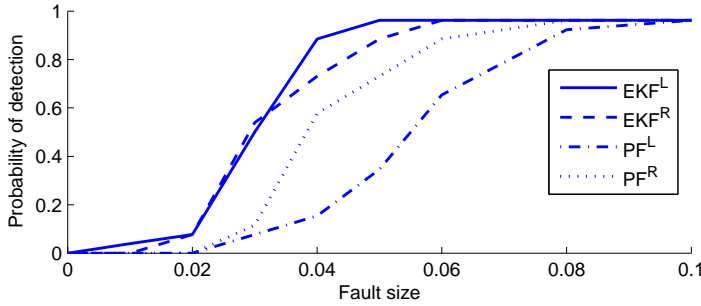
In Figure 6.7 it is seen, with the current settings of the parameters, that it will take a long time for the PF based system to make the statement that a fault has occurred. The observation of the likelihood based system gives in general similar differences between the use of the PF and the EKF, as when observing the residual based system.

### 6.1.1 Power Functions

Here another measurement is used to determine fault detectability. Using an ETC cycle divided into 27 parts, each part about 30–100 seconds long, and using the diagnosis systems on each part separately, a power function (defined in Section 3.4) can be made. This power function are used to determine the detectability of faults during an ETC cycle and should only be used as a comparison between the tests constructed in this thesis.

The power functions for the tests used in the previous section, i.e. test quantities  $T_0^R$  and  $T_1^L$  for the fault  $FS_{\text{pim}}$  using both the PF and the EKF, are seen in Figure 6.8. Observing the figure, it is clear that the diagnostic tests using the EKF provides better detection then using the PF for this type of measurement with an ETC cycle. This is explained by what was said in the previous section, that the detection time is longer for the PF based tests, which leads to that the PF based tests do not have time to react using these short intervals. In general, this is an effect of the noisy estimates using the PF which affects the thresholds.





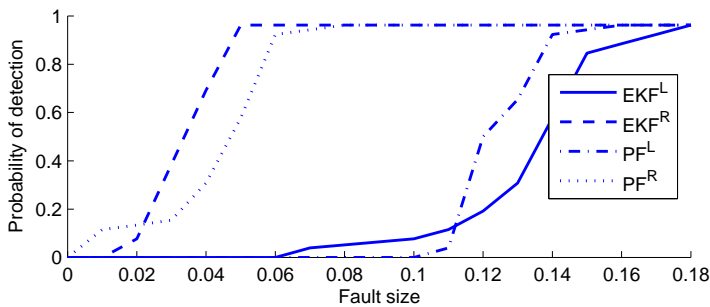
**Figure 6.8.** The power function for the test quantities  $T_0^R$  and  $T_1^L$  using both the PF and the EKF for the positive fault  $FS_{pim}$ . The fault detection probability has a resolution of  $1/27 \approx 3.7\%$ .

The thresholds is then affecting the detection time and therefore the probability of missed detection during small faults. The tests must react in the time 30–100 seconds, which is less likely for small faults and a large  $J$ .

In the previous section the residual based systems are also subjected to the fault  $FS_{vgt}$  and the power functions for test quantities  $T_0^R$  and  $T_1^L$  for both the PF and EKF are presented in Figure 6.9.

In the figure it is seen that the PF based tests may able to detect a smaller fault than the EKF based tests. This possibility to detect a smaller fault using the PF, stated in the previous section, is not seen in Figure 6.8 and depends on the longer detection time for the PF with this type of measurement.

Observing the likelihood tests in Figure 6.9, the effect of a fault model explaining another non-modeled fault is seen. Test quantity  $T_1^L$  includes the model of  $FS_{pim}$  and still responds to fault  $FS_{vgt}$ . How this affects the isolation is discussed in the next section.



**Figure 6.9.** The power function for the test quantities  $T_0^R$  and  $T_1^L$  using both the PF and the EKF for the positive fault  $FS_{vgt}$ . The fault detection probability has a resolution of  $1/27 \approx 3.7\%$ .

## 6.2 Fault Isolation

For isolation of faults using the residual based diagnosis systems, at least five of the test quantities 1–6 has to respond. If *less* than five tests respond, there will be a conclusion that more than one specific fault can be responsible for the system behavior. The same problem when isolation is not possible, happens for the likelihood based diagnosis system when *more* than one test quantity responds. Using the likelihood based systems in the case where multiple tests respond, which is shown in Section 6.1 to be a common problem, the likelihood comparison

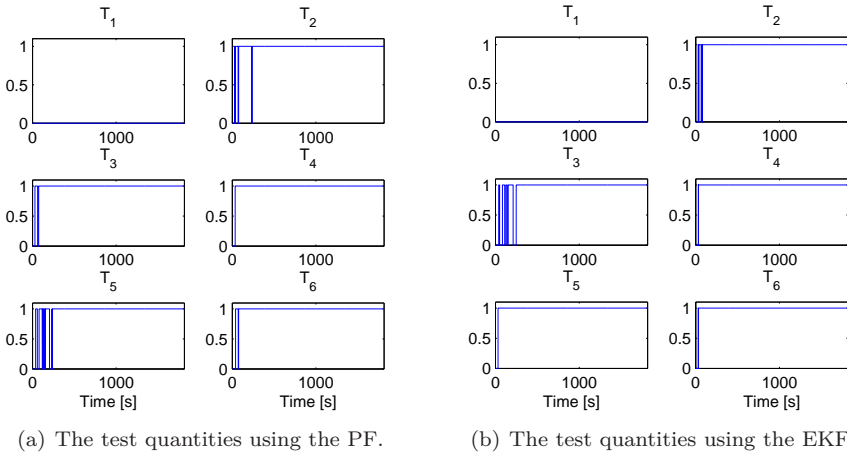
$$\frac{L(H_i)}{\sum_{j \in S} L(H_j)}, \quad (6.1)$$

can be useful. The set  $S$  are the set of test quantities that has responded. Using (6.1), a decision which fault corresponding to the more likely hypothesis  $i$  can be made.

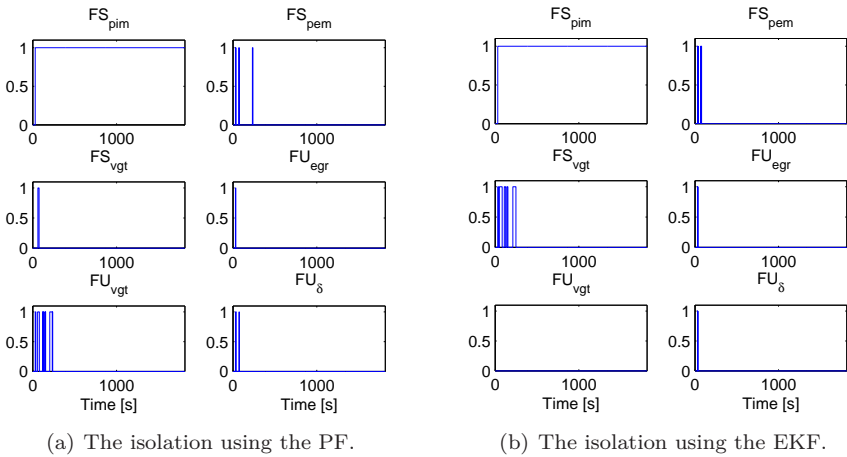
Consider the residual based diagnosis systems for both the PF the EKF, and let the systems be subjected to the fault  $FS_{\text{pim}}$  for  $\theta = 0.1$  during an entire ETC cycle. The result of which test quantities that respond are presented in Figure 6.10. With the decision structure (5.3) in mind, the isolation result is presented in Figure 6.11. Observing Figure 6.10 and Figure 6.11, one can see that the result for the PF based system and the EKF based system are quite similar. Overall for other faults (not presented here) the PF diagnosis system has slightly better isolation, but in general the systems have the same detection capabilities for a fault whose size is detectable for both systems.

Now consider the isolation for the likelihood based diagnosis system for both the PF and the EKF. Which test quantities that respond are presented in Figure 6.12 and which responding test quantity that has the highest likelihood, computed according to (6.1), can be seen in Figure 6.13. The first test quantity,  $T_1^L$ , for both the PF based and the EKF based system is plotted separately in Figure 6.14 for a better view.

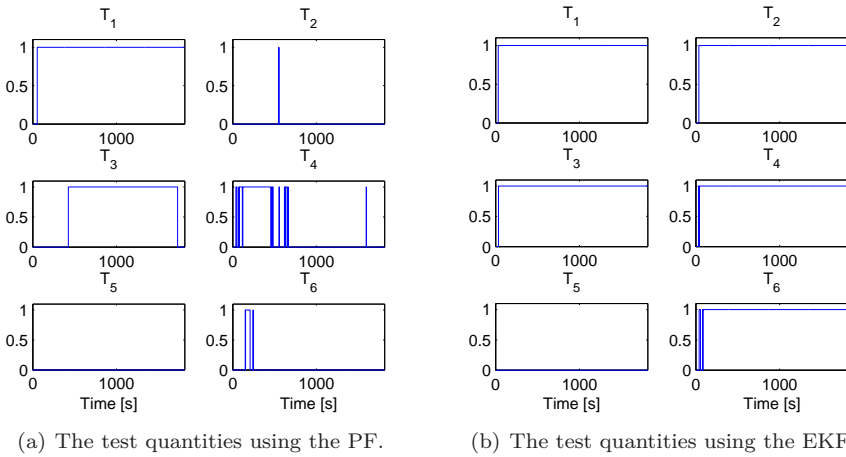
The likelihood based test quantities, except  $T_5^L$ , almost all reacts to the fault  $FS_{\text{pim}}$  which indicates that the fault modeled in hypothesis  $H_5$  can not explain fault  $FS_{\text{pim}}$  well. According to Figure 6.12 this seems to be more clear for the test quantities using the EKF which is explained by size of the cusum parameters. The drifting parameters and the thresholds can be set to compensate for this effect, but the diagnosis system are then losing detection performance (detection of small faults are going to be harder) which is seen in the power functions in the previous chapter for the PF based tests. This is however only a problem for small faults, i.e. faults of those sizes that can be explained by another modeled fault and large faults are generally harder for another modeled fault to compensate for. This problem can to some extent be evaded because the likelihood of each responded test can be compared as in Figure 6.13 and an idea of which hypothesis is more likely is obtained.



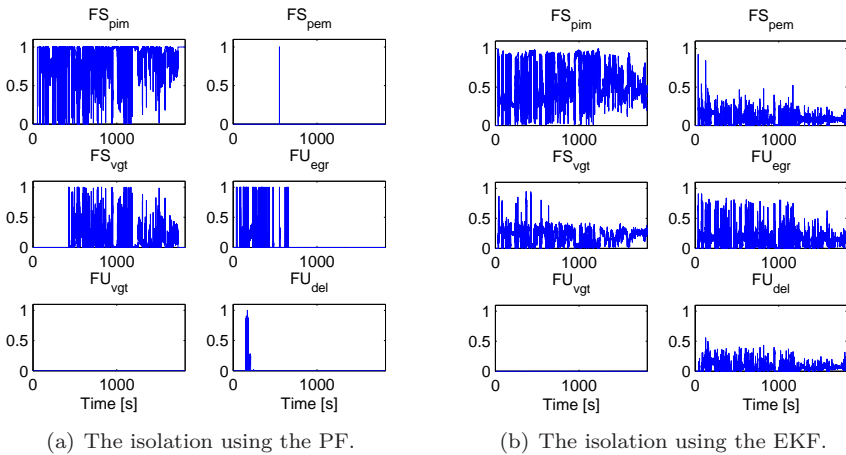
**Figure 6.10.** The test quantities  $T_i^R$  where  $i = \{1, \dots, 6\}$  for the PF in plot (a) and for the EKF in plot (b), that has responded to a gain fault  $FS_{pim}$  of size  $\theta = 0.10$  during an ETC cycle. A one means that the test quantity is responding and a zero that is not responding. The diagnosis systems is subjected to the fault during the entire cycle.



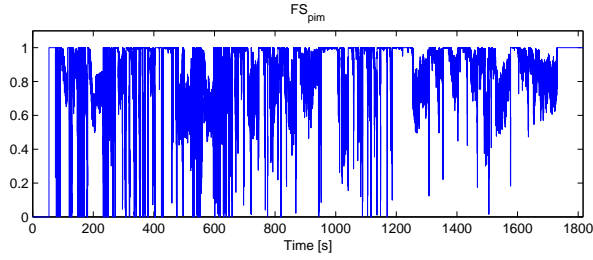
**Figure 6.11.** With the isolation structure (5.3) the isolation, in plot (a) for the PF and in plot (b) for the EKF, are presented. A one means that the fault could be responsible for the system behavior. The isolation is presented for the test quantities in Figure 6.10 for the gain fault  $FS_{pim}$  of size  $\theta = 0.10$ .



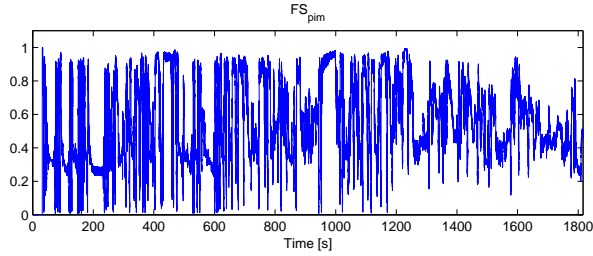
**Figure 6.12.** The likelihood based test quantities  $T_i^L$  where  $i = \{1, \dots, 6\}$  for the PF in plot (a) and for the EKF in plot (b), that has responded to a gain fault  $FS_{pim}$  of size  $\theta = 0.10$  during an ETC cycle. A one means that the test quantity is responding and a zero that is not responding. The diagnosis systems is subjected to the fault during the entire cycle.



**Figure 6.13.** With the isolation structure (5.7) the isolation, in plot (a) for the PF and in plot (b) for the EKF, are presented. The plotted value 0–1 means that the fault is 0–100 % responsible for the system behavior. The isolation is presented for the test quantities in Figure 6.12 for the gain fault  $FS_{pim}$  of size  $\theta = 0.10$ .



(a) Isolation of  $FS_{pim}$  for the PF.



(b) Isolation of  $FS_{pim}$  for the EKF.

**Figure 6.14.** A zoomed view of the isolation plots  $FS_{pim}$  presented in Figure 6.13 for the PF in plot (a) and for the EKF in plot (b).

Computing the mean of the values in Figure 6.13 with the formula (6.1), a general idea of which fault that is more likely over time is obtained, presented in Table 6.1. The values in Table 6.1 are in favor of the PF, but the probability of fault  $FS_{pim} \rightarrow 100\%$  as  $\theta$  increases for both systems. The values favor the PF based system due to the larger drifting parameters and the larger thresholds.

The problem of a fault model explaining another fault, is not affecting the likelihood based systems in the same way as it affects the residual based systems. Using the residuals, a fault model capable of partially explaining another fault only gives the system worse detection performance. But if the thresholds are set tight in the likelihood case, it will be easy for the test quantities to respond. This is because when a fault model can explain another fault, the likelihood of the respective hypothesis  $H_i$  will be greater than the hypothesis  $H_0$ .

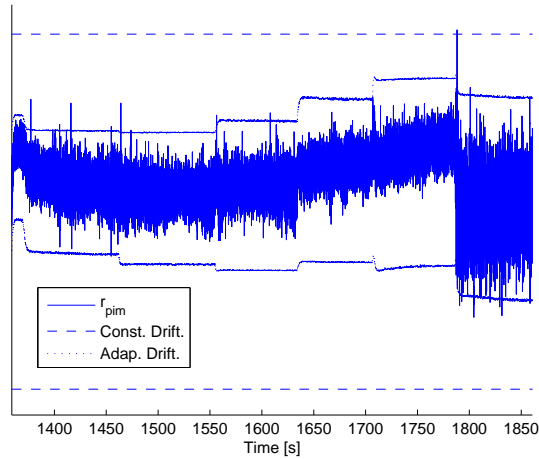
It is noted, by looking in Table 6.1, that both systems are able to isolate  $FS_{pim}$  as the most likely cause for the system behavior.

**Table 6.1.** The mean values of the likelihoods for the test quantities presented in Figure 6.13 in percentage. Both filters estimate the probability for each fault according to (6.1).

	$FS_{pim}$	$FS_{pem}$	$FS_{vgt}$	$FU_{egr}$	$FU_{vgt}$	$FU_{\delta}$
PF	87	0	9	5	0	0
EKF	57	7	19	12	0	5

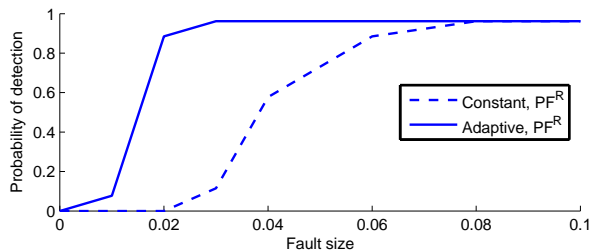
### 6.3 Performance Enhancements with an Adaptive Threshold

An adaptive threshold is constructed according to Section 5.5.2 for test quantity  $T_0^R$  using the PF and the effect is illustrated in Figure 6.15. Since making a diagnosis system with as good performance as possible is considered as a secondary objective, this adaptive threshold is only implemented for one test quantity to prove that it works. Observing Figure 6.15 it is not hard to see that using an



**Figure 6.15.** The residual  $r_{pim}$  for test quantity  $T_0^R$  plotted with an adaptive drift  $\eta_k$  parameter and a constant drift parameter  $\eta$ .

adaptive drifting parameter, the diagnosis systems performance can be increased. The power function, using the same data as explained in Section 6.1.1, for  $T_0^R$  with an adaptive and a constant drifting parameter is presented in Figure 6.16. The performance improvement is clear.



**Figure 6.16.** The power function for test quantity  $T_0^R$  using the PF with both a constant and an adaptive drifting parameter for the positive fault  $FS_{pim}$ . The fault detection probability has a resolution of  $1/27 \approx 3.7\%$ .

## 6.4 Estimation Problems for High Engine Speed

When evaluating the systems for very high state values, including high engine speed, the PF and the EKF lose estimation performance. This means that the model errors are imminent and for the PF in some cases, devastating, i.e. the PF loses the state estimate totally. This is because the 3-state model is tuned for an ETC cycle which does not contain those high speed revolutions.

This problem happens even for lower engine speed when a fault is modeled using the PF. Here the limitation due to the number of particles is seen. Expanding the system with an extra state usually requires more particles to maintain the state estimate. Generally, these problems are seen using the EKF also but not so severe.

It is important to say that the parameters set for the diagnostic tests in this thesis, is not based on information for these data containing high engine speed, if so, the diagnosis performance would not be as good as the result presented in this chapter.

## 6.5 Performance Limiting Factors

A couple of factors that effect the presented result are here summarized to:

- The 3-state model is tuned based on data from an ETC cycle made in year 2004, and the data that the diagnosis system are evaluated for in this thesis are from 2006. The problem here is that the engine used by Scania in their trucks may have changed during this time, and therefore the 3-state model may not be as accurate as it could be. The model can of course be re-parameterized for the available data from 2006, but this is not really in scope of this thesis but it would hopefully increase the diagnosis systems performance.
- All the data used are made in engine test cells<sup>1</sup>, and therefore if the diagnosis systems are used in a truck, their performance may not be the same as presented here.
- There are no data obtained especially for the diagnosis systems in this thesis. The data used for the setting of the parameters in the PF and EKF are from an engine test cell. There are a total of seven data sets, where one of them is an ETC cycle and the others are from stationary states for six different constant engine rpm values.
- The EKF is tuned with same parameters (when possible) as the PF, which may not be appropriate, and therefore EKF diagnosis systems may have performance enhancements to gain.

---

<sup>1</sup>The engine is separately run indoors without the actual truck. Several parameters affecting the engine mounted inside a truck, e.g. the exhaust system, are not seen here.

## 6.6 Summary and Discussion

First of all, there are only a couple of tests presented in this section and the results, e.g. power functions, can have worse performance for other test quantities, but all faults are detectable in some manner during an ETC cycle for less than 20 % gain faults. Faults  $FU_{\text{vgt}}$  and  $FU_{\text{egr}}$  are hardest to detect.

Fault isolation is hard for both residual based systems for small faults, and somewhat more complicated for the likelihood based systems, especially using the EKF. Isolation for large faults works rather well for all systems.

Due to the larger  $J$  for the PF based systems, the power functions indicate better performance for using the EKF. This is however not fully correct because the intervals used when making the power functions might for some faults be too short for the PF. The general effect of this is that the EKF based systems has a shorter detection time but the PF might still detect smaller faults.

When observing the likelihoods, it is indicated that the use of the PF has advantages compared to the EKF due to the nonlinearities in the system. This is clear when using more particles which is needed for the same robustness that the EKF provides.

Finally, the performance result presented in this chapter, is not for the entire state space. If the time were at hand, the 3-state model should have been re-parameterized and hopefully with performance improvements.



# Chapter 7

## Conclusions

The main objective in this Master's thesis is to, by using the Particle Filter (PF) and the Extended Kalman Filter (EKF), construct a diagnosis system for a Scania diesel truck engine and evaluate the properties for Fault Detection and Isolation (FDI).

Using the PF as the primary method and the using EKF for comparison, two residual based system and two likelihood based system are created. Each system able to detect and in some way isolate the considered faults. Some of the faults are harder to detect and/or isolate than others, but they are so for both the PF systems and the EKF systems.

The FDI performance of the PF systems and the EKF systems compared to each other are too similar to justify the use of the more computer power demanding PF. The PF systems with 200 particles need approximately 20 times more CPU time than the EKF systems, and this can not be considered to be enough for the same robustness and accuracy that the EKF provides. However, some stronger detection capabilities are seen using the PF due to the nonlinearities in the system and the method is recognized to be useful if CPU time is not a problem. When using only 200 particles the EKF has a shorter detection time in almost all cases.

The general FDI performance of the PF systems and the EKF systems are quite good but there are also more to gain from improvements such as using an adaptive threshold, tuning the engine model and using a more advanced PF algorithm than the Bootstrap algorithm. Improvements when using the EKF can also be expected when tuning the EKF parameters independently of the PF parameters, for the result presented in this report, the same parameter values are used in both filters.

The use of the likelihood based tests/systems are, compared to the use of the residuals, shown to be an useful and a competitive method. If one diagnosis system is implemented in a real On Board Diagnosis (OBD) system, both the residuals and the likelihoods can be made from the same filtered model, and therefore the residuals and the likelihoods should preferably be combined for a better result.

The EKF based systems, and especially the PF based systems, will be difficult to install in the Engine Control Units (ECU) used today mainly due to the execution time of the model. But for a possible installation, the execution time of

the diagnosis systems should be evaluated more closely, e.g. the implementation in another software language than MATLAB can have a considerable positive effect on the execution time. Exploring the possibilities to use a model with less CPU expensive computations is of course also an option. Another suggestion that can improve the FDI performance is the use of a more specific fault model than the used stochastic model. This, however implies that there has to be knowledge of the fault behavior available.

Finally, the most important statements concerning the PF based diagnosis systems and the EKF based diagnosis systems are summarized:

- The EKF systems has the shorter detection time in almost all cases when using 200 particles for the PF systems.
- The PF systems can detect smaller faults than the EKF systems, even with only 200 particles.
- The use of the PF provides some improvements due to the system nonlinearities compared to the use of the EKF.
- More than 200 particles is needed for the PF if the same robustness as with the EKF should be obtained.
- The PF system approximately needs 20 times more CPU time than the EKF systems which make a possible implementation difficult for the PF systems.
- The EKF systems are also CPU demanding due to the complexity of the engine model.

# References

- [1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Signal Processing*, 50(2), 2002.
- [2] U. M. Asher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. Society for Industrial and Applied Mathematics, 1998. ISBN 0-89871-412-5.
- [3] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes*. PTR Prentice-Hall, Inc, 1993. ISBN 0-13-126780-9.
- [4] N. Bergman. *Recursive Bayesian Estimation*. Department of Electrical Engineering, Linköpings universitet, 1999. ISBN 91-7219-473-1.
- [5] Y. Boers and H. Driesen. A particle-filter-based detection scheme. *IEEE Signal Processing Lett.*, 10(10), 2003.
- [6] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001. ISBN 0-387-95146-6.
- [7] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140(2), 1993.
- [8] F. Gustafsson. *Adaptive Filtering and Change Detection*. Wiley, 2000. ISBN 0-471-49287-6.
- [9] A. Jerhammar and E. Höckerdal. *Gas flow observer for a Scania Diesel Engine with VGT and EGR*. Department of Electrical Engineering, Linköpings universitet, 2006. Master's thesis.
- [10] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, Inc, 2000. ISBN 0-13-022464-2.
- [11] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineering — Journal BasicEngineering*, 82(Series D):35–45, Mar. 1960.
- [12] M. Nyberg and E. Frisk. *Model Based Diagnosis of Technical Processes*. Department of Electrical Engineering, Linköpings universitet, 2006. Course material for TSFS06 at Linköpings Univeristet.

- [13] R. J. Patton, P. M. Frank, and R. N. Clark, editors. *Issues of Fault Diagnosis for Dynamic Systems*. Springer, 2000.
- [14] J. Wahlström. *Control of EGR and VGT for emission control and pumping work minimization in diesel engines*. Department of Electrical Engineering, Linköpings universitet, 2006. ISBN 91-85643-83-1.

# Appendix A

## Appendix: Cusum Parameters

Table A.1 and Table A.2 contains the cusum parameters for the likelihood based diagnosis systems and the residual based diagnosis systems, respectively.

**Table A.1.** The cusum parameters for the likelihood based diagnosis systems.

Filter	Ratio	$J_{\text{upper}}$	$\eta_{\text{upper}}$
PF			
$T_1^L$	$L(H_1)/L(H_0)$	$2.0 \cdot 10^4$	1.0
$T_2^L$	$L(H_2)/L(H_0)$	$1.2 \cdot 10^4$	3.0
$T_3^L$	$L(H_3)/L(H_0)$	$2.5 \cdot 10^4$	2.0
$T_4^L$	$L(H_4)/L(H_0)$	$1.5 \cdot 10^4$	4.0
$T_5^L$	$L(H_5)/L(H_0)$	$1.7 \cdot 10^4$	4.0
$T_6^L$	$L(H_6)/L(H_0)$	$1.5 \cdot 10^4$	4.0
EKF			
$T_1^L$	$L(H_1)/L(H_0)$	$1.0 \cdot 10^3$	1.2
$T_2^L$	$L(H_2)/L(H_0)$	$1.0 \cdot 10^3$	0.5
$T_3^L$	$L(H_3)/L(H_0)$	$0.8 \cdot 10^3$	1.0
$T_4^L$	$L(H_4)/L(H_0)$	$1.0 \cdot 10^3$	0.7
$T_5^L$	$L(H_5)/L(H_0)$	$0.9 \cdot 10^3$	0.5
$T_6^L$	$L(H_6)/L(H_0)$	$0.6 \cdot 10^3$	0.5

**Table A.2.** The cusum parameters for the residual based diagnosis systems.

Filter	Residual	$J_{\text{upper}}$	$J_{\text{lower}}$	$\eta_{\text{upper}}$	$\eta_{\text{lower}}$
PF					
$T_0^{\text{R}}$	$r_{\text{pim}}$	$2.2 \cdot 10^6$	$-2.2 \cdot 10^6$	1349.1	-12645
	$r_{\text{pem}}$	$5.5 \cdot 10^6$	$-5.5 \cdot 10^6$	9672.4	-8434.1
	$r_{\text{vgt}}$	$2.0 \cdot 10^4$	$-2.0 \cdot 10^4$	14.230	-6.6504
$T_1^{\text{R}}$	$r_{\text{pem}}$	$6.0 \cdot 10^6$	$-6.0 \cdot 10^6$	8889.2	-9288.7
	$r_{\text{vgt}}$	$2.0 \cdot 10^4$	$-2.0 \cdot 10^4$	14.516	-5.3759
$T_2^{\text{R}}$	$r_{\text{pim}}$	$1.0 \cdot 10^6$	$-1.0 \cdot 10^6$	1808.3	-12711
	$r_{\text{vgt}}$	$6.0 \cdot 10^4$	$-6.0 \cdot 10^4$	25.872	-50.721
$T_3^{\text{R}}$	$r_{\text{pim}}$	$3.5 \cdot 10^6$	$-3.5 \cdot 10^6$	3753.2	-7953.5
	$r_{\text{pem}}$	$2.0 \cdot 10^6$	$-2.0 \cdot 10^6$	6570.5	-4108.8
$T_4^{\text{R}}$	$r_{\text{pim}}$	$3.0 \cdot 10^6$	$-3.0 \cdot 10^6$	2080.7	-9071.6
	$r_{\text{pem}}$	$6.0 \cdot 10^6$	$-6.0 \cdot 10^6$	4946.9	-8004.7
	$r_{\text{vgt}}$	$2.0 \cdot 10^4$	$-2.0 \cdot 10^4$	12.935	-5.1246
$T_4^{\text{R}}$	$r_{\text{pim}}$	$2.2 \cdot 10^6$	$-2.2 \cdot 10^6$	1380.8	-16985
	$r_{\text{pem}}$	$6.0 \cdot 10^6$	$-6.0 \cdot 10^6$	4181.3	-4078.9
	$r_{\text{vgt}}$	$4.0 \cdot 10^4$	$-4.0 \cdot 10^4$	30.762	-4.0309
$T_4^{\text{R}}$	$r_{\text{pim}}$	$5.5 \cdot 10^6$	$-5.5 \cdot 10^6$	1380.1	-12296
	$r_{\text{pem}}$	$1.7 \cdot 10^7$	$-1.7 \cdot 10^7$	20464	-2498.8
	$r_{\text{vgt}}$	$1.7 \cdot 10^5$	$-1.7 \cdot 10^5$	167.82	-7.4334
EKF					
$T_0^{\text{R}}$	$r_{\text{pim}}$	$7.0 \cdot 10^5$	$-7.0 \cdot 10^5$	2136.2	-9383.3
	$r_{\text{pem}}$	$2.0 \cdot 10^6$	$-2.0 \cdot 10^6$	6129.5	-4595.4
	$r_{\text{vgt}}$	$5.0 \cdot 10^4$	$-5.0 \cdot 10^4$	129.83	-32.505
$T_1^{\text{R}}$	$r_{\text{pem}}$	$2.0 \cdot 10^6$	$-2.0 \cdot 10^6$	5515.3	-4877.6
	$r_{\text{vgt}}$	$6.0 \cdot 10^4$	$-6.0 \cdot 10^4$	115.77	-53.565
$T_2^{\text{R}}$	$r_{\text{pim}}$	$6.0 \cdot 10^5$	$-6.0 \cdot 10^5$	2464.3	-8907.3
	$r_{\text{vgt}}$	$3.0 \cdot 10^4$	$-3.0 \cdot 10^4$	99.385	-27.651
$T_3^{\text{R}}$	$r_{\text{pim}}$	$3.0 \cdot 10^6$	$-3.0 \cdot 10^6$	2869.4	-4741.2
	$r_{\text{pem}}$	$1.6 \cdot 10^6$	$-1.6 \cdot 10^6$	6202.2	-4123.1
$T_4^{\text{R}}$	$r_{\text{pim}}$	$1.5 \cdot 10^6$	$-1.5 \cdot 10^6$	1872.5	-6590.2
	$r_{\text{pem}}$	$1.3 \cdot 10^6$	$-1.3 \cdot 10^6$	3597.5	-4621.6
	$r_{\text{vgt}}$	$5.0 \cdot 10^4$	$-5.0 \cdot 10^4$	131.21	-32.635
$T_5^{\text{R}}$	$r_{\text{pim}}$	$7.0 \cdot 10^5$	$-7.0 \cdot 10^5$	2140.8	-9059.3
	$r_{\text{pem}}$	$3.0 \cdot 10^5$	$-3.0 \cdot 10^5$	1331.1	-72.114
	$r_{\text{vgt}}$	$2.5 \cdot 10^4$	$-2.5 \cdot 10^4$	85.162	-32.340
$T_6^{\text{R}}$	$r_{\text{pim}}$	$1.2 \cdot 10^6$	$-1.2 \cdot 10^6$	2137.4	-9259.0
	$r_{\text{pem}}$	$3.0 \cdot 10^6$	$-3.0 \cdot 10^6$	1892.4	14420
	$r_{\text{vgt}}$	$4.0 \cdot 10^4$	$-4.0 \cdot 10^4$	78.789	-32.134