Dipl.-Ing. Mathias Jesussek

Fault detection and force estimation for a railway vehicle system

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der technischen Wissenschaften

eingereicht an der

Technischen Universität Graz

Betreuerin

Univ.-Prof. Dr.-Ing. habil. Katrin Ellermann

Institut für Mechanik

Graz, Juli 2014

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Dissertation identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral dissertation.

Datum / Date

Unterschrift / Signature

Abstract

Reliability and dependability in complex mechanical systems can be improved by fault detection and isolation methods (FDI) as well as methods for force estimation. These techniques are key elements for maintenance on demand, which can decrease service costs and time significantly. This work addresses FDI and force estimation for a railway vehicle: The mechanical model is described as a multibody system, which is excited randomly due to track irregularities. Numerous parameters, like masses, spring- and damper-characteristics, influence the dynamics of the vehicle. Often, the exact values of the parameters are unknown and might even change over time. Some of these changes are considered as critical with respect to the operation of the system and they require immediate maintenance. The aim of this work is to detect faults in the suspension system of the vehicle and to estimate wheel rail contact forces as well as forces in the suspension system. A Kalman filter is used for the state estimation. In order to detect and isolate faults, several Kalman filters are used, each of them is specific for a fault in the system. A statement of the condition of the railway vehicle is then given by comparing the data of the system under investigation and the data estimated with the Kalman filters. The estimated states are also used to solve the inverse problem in the force calculation. A full scale train model with a nonlinear wheel rail contact model serves as an example for the described techniques. Numerical results for different test cases are presented. The analysis shows that for the given system it is possible not only to detect a failure in the suspension system from the system's dynamic response, but also to distinguish clearly between different possible causes for the changes in the dynamical behavior. The wheel rail forces as well as the suspension forces can be estimated accurately.

Kurzfassung

Die Zuverlässigkeit in komplexen mechanischen Systemen kann durch Fehlerdetektion und Isolationsverfahren sowie durch die Schätzung der Kräfte im System verbessert werden. Diese Techniken sind Schlüsselelemente für eine zustandsorientierte Instandhaltung, welche Servicekosten und -zeit signifikant verringern können. Das Ziel dieser Arbeit ist die Fehlererkennung und Kraftschätzung im Schienenfahrzeug. Das Schienenfahrzeug wird als Mehrkörpersystem modelliert und die Anregung des Systems resultiert aus Gleislagefehlern. Verschiedene Parameter, wie Massen, Federn und Dämpfer beeinflussen die Zugdynamik. Oft sind die genauen Werte der Parameter unbekannt und könnten sich sogar im Laufe der Zeit ändern. Einige dieser Veränderungen sind kritisch im Hinblick auf den Betrieb des Systems und erfordern eine sofortige Wartung. Das Ziel dieser Arbeit ist es, Fehler in Dämpfern des Fahrzeuges zu erkennen und Kräfte im Kontakt zwischen Rad und Schiene sowie in Dämpferelementen zu schätzen. Um die Dynamik des Schienenfahrzeuges zu rekonstruieren, werden mittels eines Kalman Filters die Zustände des Systems geschätzt. Zur Erkennung und Isolation der Fehler werden mehrere Kalman Filter verwendet, die jeweils spezifisch für einen Fehler sind. Durch Vergleich der Daten des zu untersuchenden Systems mit den per Kalman Filter geschätzten Werten, wird auf das Auftreten von Fehlern zurück geschlossen.

Die inverse Berechnung der im System auftretenden Kräfte basiert ebenfalls auf dem Kalman Filter. Der Kalman Filter wird verwendet um die Zustände des Systems zu schätzen. Anschließend muss der Zusammenhang zwischen dem Zustand und den auftretenden Kräften ermittelt werden, durch diesen Zusammenhang können dann die Kräfte geschätzt werden.

Acknowledgments

First of all, I would like to thank my research supervisor professor Katrin Ellermann, head of the Institute of Mechanics at Graz University of Technology, for supervising this thesis. Thank you very much for the guidance, enthusiastic encouragement and constructive feedback.

I also want to thank Siemens AG Österreich for the cooperation with the Institute for Mechanics and for providing the railway vehicle data.

Especially I would like to express my deep gratitude to Gerald Grabner at the bogie competence center in Graz for the constructive feedback and for the many stimulating ideas he has given me.

Finally, I would like to thank my family and my girlfriend Hannah Volk for supporting me.

Contents

Nomenclature

Functions

$E[]$	expectation
$F()$	transfer function
$f_1(), f_2()$	geometry of the left and right wheel
$f_c()$	spring force
$f_d()$	damper force
$F_X()$	cumulative distribution function
$f_X()$	probability density function
$g_1(), g_2()$	geometry of the left and right rail
$P()$	probability likelihood
$S_{cl}()$	power spectral density of cross-level track irregularities
$S_h()$	power spectral density of horizontal track irregularities
$S_v()$	power spectral density of vertical track irregularities
$S_X()$	power spectral density
$Var()$	variance
δ_k	Kronecker delta function

Scalars

A, B	constants depending on the radii of curvatures
a, b	the contact ellipse semi-axes
$\mathcal{A}, \mathcal{B}$	event in terms of probability
A_i, B_i, C_i	constants depending on the radii of curvatures
A_j, B_j, C_j	constants depending on the radii of curvatures
A_m, B_m, C_m, D_m	constants for Hertz calculation
A_n, B_n, C_n, D_n	constants for Hertz calculation
b_{h0}, b_{v0}, b_{cl2}	coefficients for track irregularities
C_h	normal damping coefficient
c_{ji}	creepage coefficients
C_{xy}	cross-covariance

d	distance between rail and wheel contact point
e	scalar residual
$fault$	sum of fault indication values
F_d	normal damping force
F_h	Hertz normal load
f_{Hertz}	Hertz contact force
F_n	normal load
F_x, F_y	creep forces
G	modulus of rigidity
G_r	modulus of rigidity of rail
G_w	modulus of rigidity of wheel
ℓ_0	unstrained spring length
K_1, K_2	constants for Hertz calculation depending on material properties
K_{Hertz}	Hertz coefficient
m	mass of a body
m_H, n_H	constants for Hertz calculation
M_Z	creep moment
p	pressure in the contact point
p_0	maximal pressure in the contact point
Q_F	vertical wheel force
$R_{i1}, R_{i2}, R_{j1}, R_{j2}$	radii of curvature
R_{xy}	cross-correlation
R_{xx}	auto-correlation
s	arc length
s^r	surface parameter of the rail
s^w	surface parameter of the wheel
t	time
V	magnitude of the wheel velocity
v_x^r, v_y^r	velocity of the contact point at the rail
v_x^w, v_y^w	velocity of the contact point at the wheel
X	random variable
$\hat{X}$	expected value of a random variable
Y	random variable
Y_F	lateral wheel force
$\hat{Y}$	expected value of random variable
y_t	lateral displacement
z_t	vertical displacement
$\gamma_x, \gamma_y, \omega_z$	longitudinal, lateral and spin creepage
Δt	Sample time

δ	penetration for Hertz normal Force
δ_F	contact angle at the contact point
ϵ	fault indication values
θ	pitch-angle
μ	friction coefficient
ν	Poisson's ratio
ρ_{XY}	correlation coefficient
σ	standard deviation
τ	time shift
ϕ	roll-angle
Ψ	yaw-angle

Matrices and vectors

$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$	system, input, output and feed-through matrices
$\mathbf{A}_{rot}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$	rotational matrices
$\mathbf{A}_s$, $\mathbf{B}_s$, $\mathbf{C}_s$, $\mathbf{D}_s$	system, input, output and feed-through matrices of the train model
$\mathbf{A}_{ffi}$, $\mathbf{A}_{ff}$, $\mathbf{A}_{FF}$	system matrices for form filter
$\boldsymbol{\alpha}$	Euler angles
$\mathbf{B}_{ffi}$, $\mathbf{B}_{ff}$, $\mathbf{B}_{FF}$	input matrices for form filter
$\mathbf{C}_{ffi}$, $\mathbf{C}_{ff}$, $\mathbf{C}_{FF}$	output matrices for form filter
$\mathbf{C}_F$	forces output matrices
$\mathbf{D}_{ffi}$, $\mathbf{D}_{ff}$, $\mathbf{D}_{FF}$	feed-through matrices for form filter
$\mathbf{D}_F$	force feed-through matrices
$\boldsymbol{F}$	forces
$\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{L}$	Discrete system, input, output and feed-through matrix
$\mathbf{F}_c$	spring force
$\mathbf{F}_d$	damper force
$\mathbf{I}$	identity matrix
$\mathbf{J}$	moment of inertia tensor
$\mathbf{Jac}$	Jacobian matrix
$\mathbf{K}$	observer or Kalman filter gain matrix
$\boldsymbol{\ell}$	vector between two points
$\mathbf{M}$	moment
$\mathbf{n}^r$	normal vector of the rail surface
$\mathbf{n}^w$	normal vector of the wheel surface
$\mathbf{O}$	observability matrix
$\mathbf{P}$	state residual covariance matrix

$\mathbf{p}$	position of contact point
$\hat{\mathbf{P}}_k^-$	a-priori covariance matrix
$\hat{\mathbf{P}}_k^+$	a-posteriori covariance matrix
$\mathbf{p}^r$	contact point at the rail
$\mathbf{p}^w$	contact point at the wheel
$\mathbf{Q}$	disturbance input covariance matrix
$\mathbf{R}$	measurement error covariance matrix
$\boldsymbol{R}$	global position vector of the origin of the body
$\boldsymbol{r}$	global position vector
$\boldsymbol{R}^r$	global position vector of the origin of the rail
$\boldsymbol{R}^w$	global position vector of the origin of the wheel
$\mathbf{r}_x$	state vector residual
$\mathbf{r}_y$	measurement residual
$\mathbf{t}^r$	tangential vector of the rail surface
$\mathbf{t}^w$	tangential vector of the wheel surface
$\mathbf{u}$	input signal
$\mathbf{u}^r$	contact point position
$\mathbf{u}_k$	discrete input signal
$\mathbf{u}^w$	contact point position
$\mathbf{v}$	measurement noise vector
$\mathbf{w}$	process noise vector
$\mathbf{x}$	state vector
$\dot{\mathbf{x}}$	time derivative of state vector
$\hat{\mathbf{x}}$	estimated state vector
$\dot{\hat{\mathbf{x}}}$	time derivative of estimated state vector
$\mathbf{x}_{ff}$, $\mathbf{x}_{ffi}$, $\mathbf{x}_{FF}$	state vector of form filter
$\mathbf{x}_k$	discrete state vector
$\hat{\mathbf{x}}_k$	discrete estimated state vector
$\hat{\mathbf{x}}_k^-$	a priori estimated state vector
$\hat{\mathbf{x}}_k^+$	a posteriori estimated state vector
$\mathbf{x}_s$	state vector of railway system
$\mathbf{y}$	output signal, measurement vector
$\hat{\mathbf{y}}$	estimated measurement vector
$\mathbf{y}_{ff}$, $\mathbf{y}_{ffi}$, $\mathbf{y}_{FF}$	output vector of form filter
$\mathbf{y}_k$	discrete measurement vector
$\hat{\mathbf{y}}_k$	discrete estimated measurement vector
$\mathbf{y}_s$	measurement vector of train model
$\boldsymbol{\omega}$	angular velocity

1 Introduction

The capacity of transport as well as the number of passengers is growing in the railway industry. At the same time, the pressure to reduce service costs rises. The effort to reduce service costs and time starts during the planning phase of the vehicle and the rail, and it continues during the operation phase. During the operation phase an enormous cost factor results from the maintenance of the railway track as well as the railway vehicle. Thus, decreasing maintenance costs is a major focus in the railway industry.

Reliability and dependability in complex mechanical systems can be improved by fault detection and isolation methods (FDI) as well as force estimation techniques. These techniques, FDI and force estimation, are key elements for maintenance on demand, which could decrease service costs and time significantly.

Further, an accurate and easy way to estimate forces in the system improves the Asset Management. In the case of the railway vehicle, forces occurring in the wheel rail contact and the suspension system are of special interest.

This work addresses two major topics: The first topic is the detection and isolation of faults in the suspension system, the second topic is the estimation of forces in the wheel rail contact point as well as in the suspension system.

In both topics, the dynamics of the railway vehicle is modeled by a multibody system, which is excited randomly due to track irregularities. The Velaro RUS serves as an example. Various parameters, like masses, spring- and damper-characteristics, influence the dynamics of the vehicle. Often, the exact values of the parameters are unknown and might even change over time. Some of these changes are considered as critical with respect to the operation of the system and they require immediate maintenance. In both topics the railway vehicle dynamics is captured with a Kalman filter. In order to improve the state estimation process, the characteristics of the track irregularities are considered in the fault detection process. The characteristics of the track irregularities are modeled with form filters. The form filters are based on power spectral density functions of the assumed random track irregularities and are included in the system's equations.

Thus, the aim of the first topic of this work is to detect and isolate faults in the suspension system. The detection and isolation of faults is divided into two steps: First, changes in the system, which are caused by a fault, are identified (fault

detection). Second, the location of the faults are determined (fault isolation).

To detect and isolate faults, noisy acceleration measurements of the railway vehicle are used. With this acceleration measurements multiple state estimations are performed. Each estimation is based on a different model of the railway vehicle, one model represents the fault free case and the other models represent the faulty cases. Out of all available models, the one with the smallest estimation error is selected. Kalman filters are used to perform the state estimations. A full scale train model with nonlinear wheel rail contact and nonlinear suspension forces serves as an example for the described techniques.

Important components in railway suspension systems are the anti-yaw damper, the secondary lateral, and vertical damper. The anti-yaw damper is critical for running stability, the other two influence mainly the ride comfort. In this study, the presented FDI procedure is used in order to detect faults in any of these dampers. Furthermore, it is possible to distinguish between a large number of different fault scenarios: It is possible to isolate a single fault at a single suspension element as well as to detect and isolate multiple failures in the complex suspension system. Numerical results for different test cases are presented and the robustness of the proposed method is analyzed.

The aim of the second topic is to estimate the lateral and vertical wheel rail contact forces and suspension forces in the railway vehicle system. The use of indirect methods for the estimation of dynamic forces acting on mechanical systems is very interesting as a direct measurement of these forces is usually very difficult and sometimes impossible. The wheel-rail system is a major part of the railway vehicle and has to fulfill several functions, such as carrying, guiding and transmitting power. Thus, forces occurring in the wheel-rail contact influence the running safety, the track loading and the ride characteristics of the vehicle. Currently, wheel-rail contact forces are determined using a wheelset equipped with strain gauges. In this method the contact forces are calculated by evaluating the measured strains. These custom made wheelsets are only operated during specific train rides, for example to test new trains or tracks. On the other hand, the use of these strain gauges is too expensive for normal operation. Therefore, there is demand for alternative ways to calculate the wheel-rail forces. In addition, the forces resulting from the suspension system provide information, which can be used for maintenance purposes and strength calculation. The obtained force data serves to design and develop optimized railway vehicles and tracks. Further, the evolution of the forces over time can be analyzed and the correlation between different components of the train and track can be evaluated.

The estimation of the forces in the railway vehicle system is an inverse problem, which includes several difficulties. The inverse problem is often not unique and

badly scaled. Similar to the first topic, a linear Kalman filter is used for the state estimation based on given acceleration data and specific track characteristics. The relation between the state vector and the various forces is determined. This relation is used and the different forces are calculated based on the estimated states.

In order to test the FDI method for a variety of different suspension faults, a detailed model of the Velaro RUS with nonlinear suspensions as well as nonlinear wheel rail contact forces is used. This model is used to simulate all test scenarios.

1.1 Review of literature

A number of recent publications address the detection and isolation of faults and force estimation in railway vehicle systems.

Parameter estimation using a Rao-Blackwellized particle filter and Extended Kalman filter [1, 2, 3] gives good results for linear and nonlinear suspension systems using a two dimensional linear railway vehicle model. A multiple-model algorithm for the detection of faults is given in [4, 5], a two dimensional half train serves as a model. FDI methods for the handling of damping coefficients are described in [6, 7]: Depending on the sign of the relative damper velocity, the coefficients switch between two distinct values. Suspension parameter estimation in the frequency domain is presented in [8]. Model-less monitoring methods for railway vehicles are demonstrated in [9, 10, 11]. A comparative study on fault detection methods of urban rail vehicle suspension systems is presented in [12]: The fault detection methods are applied to a three dimensional vehicle model, in which all four primary suspension springs or all four secondary suspension dampers fail at the same time. A Hybrid Extended Kalman filter for fault detection in nonlinear suspension elements for a half train model is given in [13]. In [14] a distributed fault detection method is proposed for light rail vehicles. A vertical train model is used for the fault detection. In [15], a fault detection method is introduced which uses a consensus principle component analysis. The method is model less and it is shown that it is possible to detect faults in the primary spring and damper when all primary springs or dampers were faulty. It is not shown that it is possible to distinguish between different dampers and suspensions. In [16] a primary conceptual overview of processing possibilities for condition monitoring systems is given.

The estimate of wheel rail contact forces based on measured wheel disc strains is presented in [17]. Three different numerical algorithms are presented. An inverse method for dynamic load estimation using measurements of system responses is presented in [18]. In [19] and [20] an inverse dynamics method is introduced in order to identify lateral wheel-rail forces. These methods use the Thikonov regular-

ization and the Bellman principle of optimality to minimize the objective function for the estimation of applied forces due to excitation. In [21] the vertical dynamic interaction between train and track is estimated by using an extended state space vector approach in conjunction with a complex model superposition for the track. In [22] vertical wheel rail contact forces at high-frequency are estimated based on [21]. Measured results from two field test campaigns are used to validate the vehicle–track interaction model. Four vehicle models and two visco-elastic track models are compared.

Further, there are some investigations to estimate the track irregularities. In [23] inverse linear parametric models are used to identify rail track irregularities. A Kalman filter is used in [24] to estimate the track geometry from carbody vibration.

Beyond the fault detection and force estimation there is a trend to use the Kalman filter to estimate several entities of the train. In [25] an unscented Kalman filter is used to estimate the friction coefficient of the wheel rail contact. A multi-Kalman filter approach for estimation of wheel rail contact conditions is presented in [26]. A nonlinear model of lateral and yaw dynamics of a wheelset is used in the presented work. In [27] the Kalman filter is applied for control strategies for active steering of bogie-based railway vehicles.

The strategy developed in this work analyses a railway vehicle based on acceleration data. Compared to most work published in this area, a highly detailed three dimensional train model is used. Because of the accurate train model and the database of Kalman filters, the fault detection as well as the force estimation can be used to give a clear indication of the condition of a multitude of different components of the train model.

1.2 Outline of the work

In the first chapter, the motivation for this work and a literature review of research related to fault detection and force estimation in the railway systems are given.

The next chapter presents the theoretical background of linear systems, signals and probability theory, which are needed for this work. Basic characteristics of linear systems are derived and the general form of a state observer is introduced. The section on probability theory provides the information which is needed for the Kalman filter.

Chapter 3 addresses the train model, which is used for simulation and testing. The first section covers the wheel rail contact model, which mainly influences the train dynamics. The second section describes the components of the railway vehicle and how the equations of motion are derived.

Chapter 4 presents the procedure of fault detection and isolation and force estimation. The Kalman filter is introduced, conditions for the state estimations and tuning parameters are explained. The introduction to Kalman filters is followed by the proposed method for fault detection and isolation. Starting with the characteristics of the estimation error, the Kalman filter is extended for the fault detection and isolation in the suspension system. The results are shown and a stability analysis of the proposed methods is performed. The chapter ends with a method to estimate wheel rail contact forces [28] as well as suspension forces. The estimation of both forces is an inverse problem, thus the concepts of system inversion are introduced and the difficulties, which could occur, are worked out. It is presented how the Kalman filter is integrated in this inversion task. At the end, the results of the force estimation and the relevance for safety and design optimization are shown.

Chapter 5 presents a summary of the obtained results. Both topics, the fault detection as well as the force estimation are discussed.

2 Theory

This section covers theory and fundamental methods, which are extensively used in this work. First, basic concepts of linear systems and signals are provided, which are important to capture the train dynamics in an efficient way. Then, probability theory is covered, which is used to analyze the random excitation of the train model and to define the Kalman filter for the state estimation of the train.

2.1 Linear systems and signals

Dynamical systems can be divided into linear and nonlinear systems. The theory of linear systems is general and widely developed, whereas for the nonlinear case, many concepts and theories for special issues exist. The concepts of linear system theory are used in this work to perform the model based fault detection as well as the force estimation. Thus, some basic theory of linear systems and signals is reviewed in this chapter. More investigations can be found in [29], [30] and [31].

2.1.1 Linear systems

Figure 2.1 shows the block diagram of a dynamical system, which transforms an input signal $\mathbf{u}(t)$ into the output signal $\mathbf{y}(t)$. If the output of the system is pro-

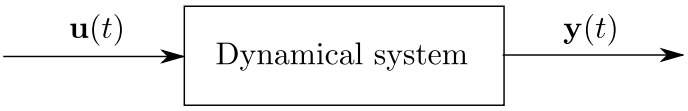


Figure 2.1: Dynamical system

portional to the input, the system is called linear. A linear function satisfies two properties, the additivity and the homogeneity property

$$\text{Additivity} : f(a+b) = f(a) + f(b), \tag{2.1}$$

$$\text{Homogeneity} : \lambda g(a) = g(\lambda a), \text{ for all } \lambda. \tag{2.2}$$

Applying equations (2.1) and (2.2) to a linear dynamical system, the results are

$$\mathbf{u}_1(t) \mapsto \mathbf{y}_1(t) \quad \text{and} \quad \mathbf{u}_2(t) \mapsto \mathbf{y}_2(t), \tag{2.3}$$

$$\mathbf{u}(t) = a\mathbf{u}_1(t) + b\mathbf{u}_2(t) \quad \mapsto \quad \mathbf{y} = a\mathbf{y}_1(t) + b\mathbf{y}_2(t), \tag{2.4}$$

where the arrow " $\mapsto$ " denotes, that the system maps the input $\mathbf{u}(t)$ onto the output $\mathbf{y}(t)$.

If a system is linear and in addition the system's behavior does not change over time, the system is called a linear time invariant system (LTI). A time invariant system must satisfy

$$\mathbf{u}_1(t) = \mathbf{u}_2(t-\tau) \quad \mapsto \quad \mathbf{y}_1(t) = \mathbf{y}_2(t-\tau), \tag{2.5}$$

where τ is any time shift. Many methods to analyze dynamical systems are based on linear models. Thus, for nonlinear systems, suitable methods are needed for linearisation.

2.1.2 State-space representation

A linear and time-invariant system, shown in Figure 2.2, can be expressed by a set of linear differential equations describing the physical characteristics of the system. This set of differential equations models the influence of the system input $\mathbf{u}(t)$ on the system output $\mathbf{y}(t)$. A so-called state space representation is given, if the set of

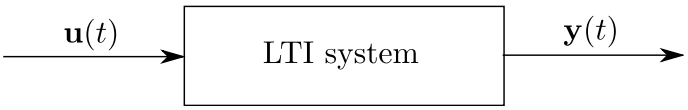


Figure 2.2: LTI system

differential equations is transformed into a set of first order differential equations. In addition, if the dynamical system is a LTI system the differential equation can be written in matrix form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \tag{2.6}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}, \tag{2.7}$$

where $\mathbf{x}$ is the n-dimensional state vector with its time derivative $\dot{\mathbf{x}}$. The m inputs of the system are combined in the input vector $\mathbf{u}$, the l outputs of the system are combined in the output vector $\mathbf{y}$. The matrix $\mathbf{A}$ defines the system matrix, the matrix $\mathbf{B}$ is the input matrix and $\mathbf{C}$ is the output matrix. The matrix $\mathbf{D}$ is called the feed-through matrix. It is zero, if there is no direct feed-through in the system.

The system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are constant in the time-invariant case. Thus, by using the state space representation, a system with multiple inputs and outputs can be modeled and analyzed in a compact way.

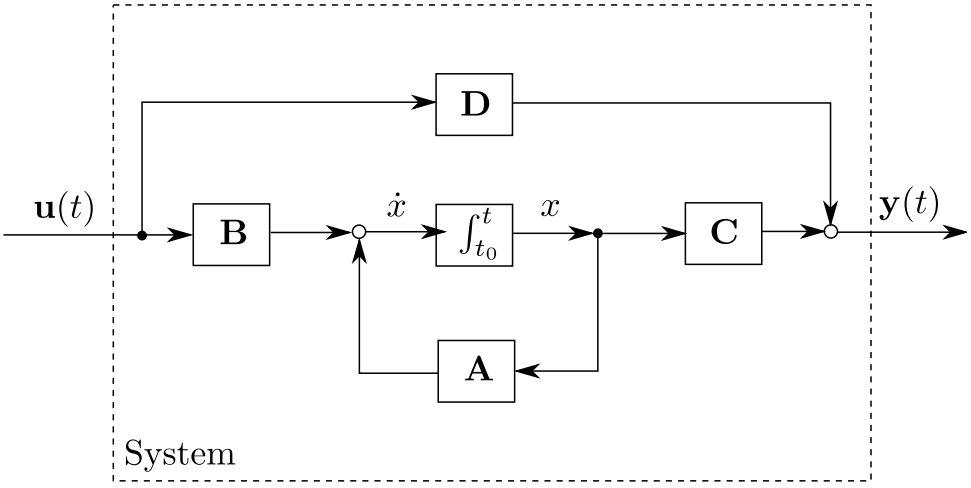


Figure 2.3: State space system

A block diagram of the state-space model given by equations (2.6) and (2.7) is shown in Figure 2.3. The diagram shows the principal parts of the system and their connections.

2.1.3 Observability

The fault detection as well as the wheel rail force estimation use many concepts of control theory. The states of the system are in general not directly accessible, thus a controller needs a state estimation in order to perform state feedback control. In this work, the state estimation is needed to reconstruct the system dynamics and to determine faults or forces in the system. The concepts of controllability and observability are related to each other, they are called dual problems. A system is called controllable, if it is possible to find a control input that takes the system from any initial point to any final point in any given time. A system is called observable, if the values of its state variables can be uniquely determined from its input and output signals. Figure 2.4 shows the concept of a closed loop state estimator. The estimated state is denoted by $\hat{\mathbf{x}}$ and the estimated measurement is denoted by $\hat{\mathbf{y}}$. The state observer uses the measured output signal $\mathbf{y}(t)$ to compute the output estimation error $\hat{\mathbf{y}}(t) - \mathbf{y}(t)$, which is multiplied by a gain matrix $\mathbf{K}$ and added to the estimator. The estimator dynamics is described by

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{KC}\left(\hat{\mathbf{x}}(t) - \mathbf{x}(t)\right) + \mathbf{Bu}(t) \tag{2.8}$$

$$\dot{\hat{\mathbf{x}}}(t) = (\mathbf{A} + \mathbf{KC})\hat{\mathbf{x}}(t) - \mathbf{KCx}(t) + \mathbf{Bu}(t) \tag{2.9}$$

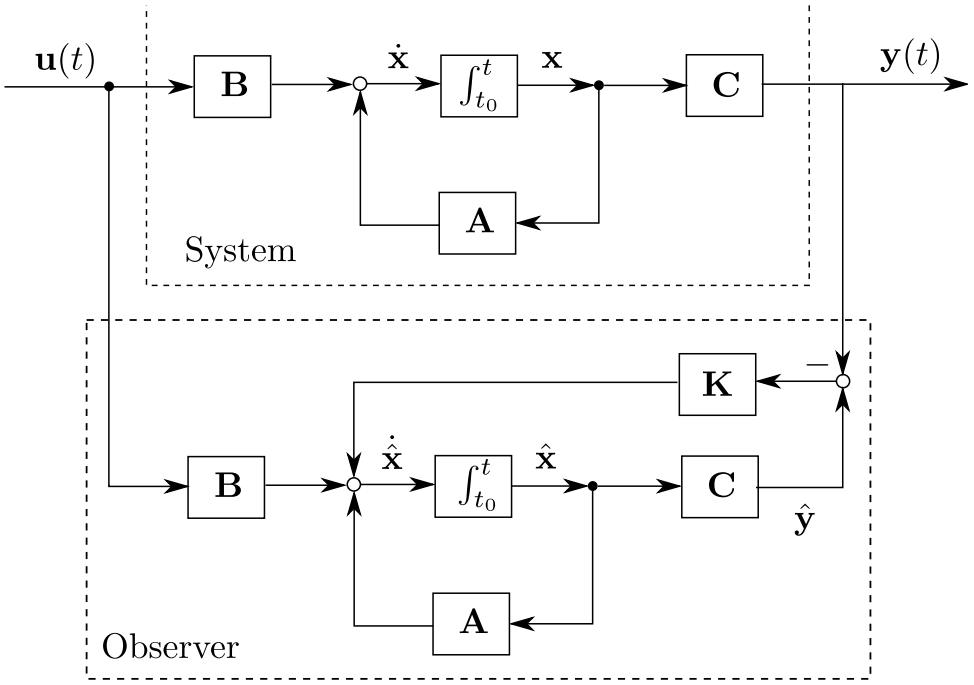


Figure 2.4: State estimator

for the given state space model

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{2.10}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}. \tag{2.11}$$

The error dynamics of the estimator is

$$\dot{\mathbf{x}}(t) - \dot{\hat{\mathbf{x}}}(t) = \mathbf{A}(\mathbf{x}(t) - \hat{\mathbf{x}}(t)) + \mathbf{K}\mathbf{C}(\mathbf{x}(t) - \hat{\mathbf{x}}(t)) \tag{2.12}$$

$$\dot{\tilde{\mathbf{x}}} = (\mathbf{A} + \mathbf{K}\mathbf{C})\tilde{\mathbf{x}}. \tag{2.13}$$

Thus, the eigenvalues of $\mathbf{A} + \mathbf{K}\mathbf{C}$ determine the stability of the estimation process. The matrices $\mathbf{A}$ and $\mathbf{C}$ result from the mechanical system, the gain matrix $\mathbf{K}$ can be chosen freely to obtain desired eigenvalues.

If it is possible to find a gain matrix $\mathbf{K}$ to obtain any set of desired observer eigenvalues, the so-called observability matrix

$$\mathbf{O}(\mathbf{C}, \mathbf{A}) = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{n-1} \end{bmatrix} \tag{2.14}$$

has full rank. A weaker form of observability is detectability. If $\mathbf{O}(\mathbf{C}, \mathbf{A})$ does not

have full rank, but a gain matrix $\mathbf{K}$ exists such that all eigenvalues of $\mathbf{A}+\mathbf{KC}$ are in the left half plan, i.e they have a negative real part, the system is detectable. Thus, the train model must be at least detectable in order to reconstruct the state.

In reality, the system is corrupted with process noise $\mathbf{w}$ and measurement noise $\mathbf{v}$, which is shown in Figure 2.5.

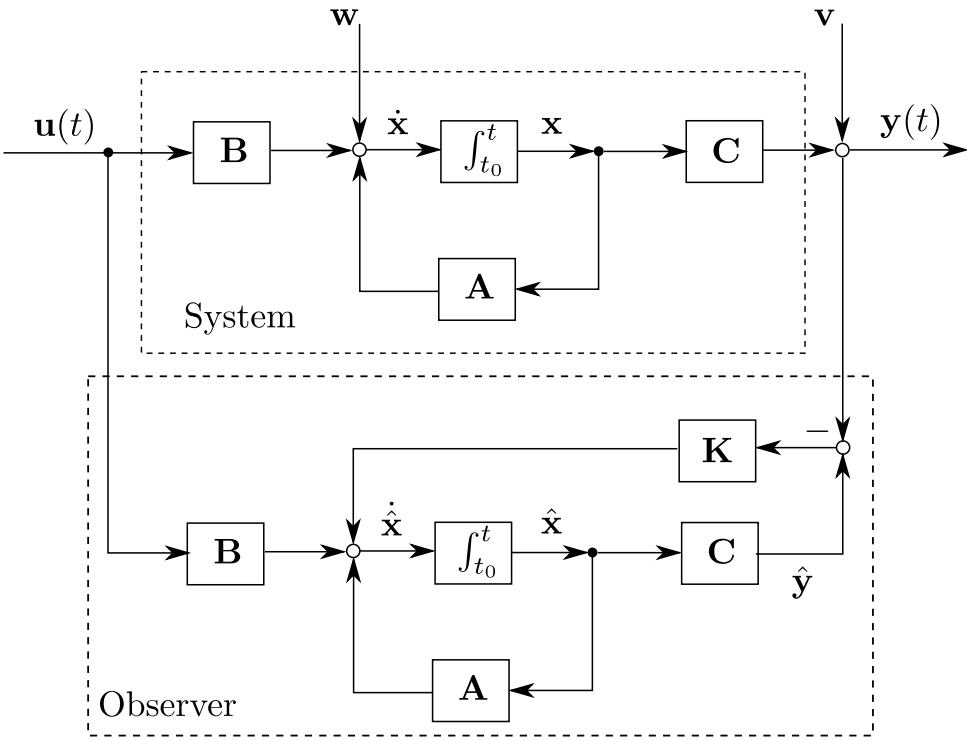


Figure 2.5: State estimator with unknown input

Both noise terms influence the estimation process and the estimator dynamics is described by

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{KC}\left(\hat{\mathbf{x}}(t) - \mathbf{x}(t)\right) - \mathbf{Kv}(t) + \mathbf{Bu}(t). \tag{2.15}$$

A special case of a state estimator is the so called Kalman filter, this filter requires white process and measurement noise. The Kalman filter is explained in more detail in section 4.1 and is used extensively in the fault detection and force estimation methods.

2.1.4 Discretization

In order to estimate the states of the train model, acceleration measurements are used. The measurements are sampled at discrete instances of time, thus, to perform a state estimation, the train model is discretized.

If the state space model from equations (2.6) and (2.7) is discretized with a sample time Δt, the discretized system results to

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{G}\mathbf{u}_k, \tag{2.16}$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{L}\mathbf{u}_k \tag{2.17}$$

with

$$\mathbf{F} = e^{\mathbf{A}\Delta t}, \qquad \mathbf{G} = \mathbf{F}(\mathbf{I} - e^{-\mathbf{A}\Delta t})\mathbf{A}^{-1}\mathbf{B}, \qquad \mathbf{H} = \mathbf{C}, \qquad \mathbf{L} = \mathbf{D}. \tag{2.18}$$

Many methods exist to discretize a system numerically. Since the railway model is a very stiff problem, in this work an implicit Euler method is used. Using the implicit Euler method, equation (2.6) can be approximated as

$$\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\Delta t} \approx \mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{u}_k \tag{2.19}$$

solving this equation for $\mathbf{x}_{k+1}$ results to

$$\begin{aligned} \mathbf{x}_{k+1} - \mathbf{x}_k &\approx \Delta t\mathbf{A}\mathbf{x}_{k+1} + \Delta t\mathbf{B}\mathbf{u}_k \\ \mathbf{x}_{k+1} - \Delta t\mathbf{A}\mathbf{x}_{k+1} &\approx \mathbf{x}_n + \Delta t\mathbf{B}\mathbf{u}_k \\ (\mathbf{I} - \Delta t\mathbf{A})\mathbf{x}_{k+1} &\approx \mathbf{x}_n + \Delta t\mathbf{B}\mathbf{u}_k \\ \mathbf{x}_{k+1} &\approx (\mathbf{I} - \Delta t\mathbf{A})^{-1}\mathbf{x}_n + (\mathbf{I} - \Delta t\mathbf{A})^{-1}\Delta t\mathbf{B}\mathbf{u}_k \end{aligned} \tag{2.20}$$

and finally

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_n + \mathbf{G}\mathbf{u}_k, \tag{2.21}$$

with $\mathbf{F} = (\mathbf{I} - \Delta t\mathbf{A})^{-1}$ and $\mathbf{G} = (\mathbf{I} - \Delta t\mathbf{A})^{-1}\Delta t\mathbf{B}$.

With the implicit Euler method it is possible to derive a stable discrete time state space model for the railway vehicle. The state space model is stable, if all eigenvalues of $\mathbf{F}$ lie within the unit circle in the complex plane or rather if all eigenvalues have an absolute magnitude smaller than one.

2.2 Probability theory

For both topics of this work, the fault detection procedure as well as the force estimation process, noisy measurement signals are used to determine the railway vehicle dynamics. In order to get as much information as possible from these signals, it is important to know the characteristics of the noise and how it influences the system behaviour or rather the estimation process. In this section basic concepts are reviewed of how to handle this uncertainty and it gives an introduction to probability theory and random variables. Probability theory plays a major role in defining the Kalman filter and thus for the fault detection and force estimation.

2.2.1 Probability

The probability $P(\mathcal{A})$ is the likelihood that a given event $\mathcal{A}$ will occur. An event is the outcome of a process, called the experiment. The set of all possible events is called sample space of the experiment and is denoted by Ω. The number of all possible outcomes can be finite or infinite. The probability of $P(\mathcal{A})$ is a numerical measure and the number is between 0 and 1, the number 0 indicates impossibility and the number 1 indicates certainty.

According to [32], there are the following probability axioms

1. (Non-negativity) $P(\mathcal{A}) \geq 0$, for every event $\mathcal{A}$.

2. (Additivity) If $\mathcal{A}$ and $\mathcal{B}$ are two disjoint events, then the probability of their union satisfies

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}). \tag{2.22}$$

   Furthermore, if the sample space has an infinite number of elements and $\mathcal{A}_1, \mathcal{A}_2, ...$ is a sequence of disjoint events, then the probability of their union satisfies

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup ...) = P(\mathcal{A}_1) + P(\mathcal{A}_2)... \tag{2.23}$$

3. (Normalization) The probability of the entire sample space Ω is equal to 1, that is, $P(\Omega) = 1$.

2.2.2 Conditional probability and independence

Conditional probability is an important concept, which is used by the Kalman filter to perform a state estimation. It is used to propagate the state estimation through

time and to increase the accuracy from time step to time step.

In probability theory, a conditional probability gives information about the outcome of an event, given that another event has occurred. If the event $\mathcal{B}$ has occurred, the conditional probability of event $\mathcal{A}$ is defined as

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A},\mathcal{B})}{P(\mathcal{B})} \tag{2.24}$$

with the assumption that $P(\mathcal{B}) > 0$. $P(\mathcal{A},\mathcal{B})$ is the joint probability of $\mathcal{A}$ and $\mathcal{B}$.

If an event $\mathcal{A}$ is not affected by an event $\mathcal{B}$, the events are independent. In probability theory, independent events have the joint probability

$$P(\mathcal{A},\mathcal{B}) = P(\mathcal{A})P(\mathcal{B}). \tag{2.25}$$

If events $\mathcal{A}$ and $\mathcal{B}$ are independent, the conditional probability of $P(\mathcal{A}|\mathcal{B})$ is $P(\mathcal{B})$ or vice versa the conditional probability of $P(\mathcal{B}|\mathcal{A})$ is $P(\mathcal{A})$. The Kalman filter uses this concept in order to quantify the likelihood of a state under the condition that specific states were estimated in the past.

2.2.3 Random variables

According to [33], a random variable is defined as a function mapping a set of experimental outcomes to a set of real numbers. The outcome of a particular experiment is not a random variable. The random variable X exists independently of any of its realizations, it will always be random and will never be equal to a specific value. A random variable can be either continuous or discrete.

One way to describe a random variable is its cumulative distribution function (cdf)

$$F_X(x) = P(X \leq x). \tag{2.26}$$

The cdf has the following properties

$$\begin{aligned} F_X(x) &\in [0,1], \\ F_X(-\infty) &= 0, \\ F_X(\infty) &= 1, \\ F_X(a) &\leq F_X(b) \text{ if } a \leq b, \\ P(a < X \leq b) &= F_X(b) - F_X(a). \end{aligned} \tag{2.27}$$

The derivative of the cdf is defined as the probability density function (pdf)

$$f_X(x) = \frac{dF_X(x)}{dx}, \tag{2.28}$$

with the properties

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(z)dz, \\ f_X(-\infty) &\geq 0, \\ \int_{-\infty}^{\infty} f_X(x)dx &= 1, \\ P(a < X \leq b) &= F_X(x_2) - F_X(x_1) \\ &= \int_a^b f_X(x)dx. \end{aligned} \tag{2.29}$$

Input noise is often assumed to be a Gaussian random variable, which is essential when defining the Kalman filter. A random variable is called Gaussian or normal distributed if its pdf is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\hat{x})^2}{2\sigma^2}\right). \tag{2.30}$$

The values $\hat{x}$ and σ in the pdf are the expected value and standard deviation of the Gaussian random variable.

A pdf of a Gaussian random variable with a mean of zero and a variance of one is given in Figure 2.6.

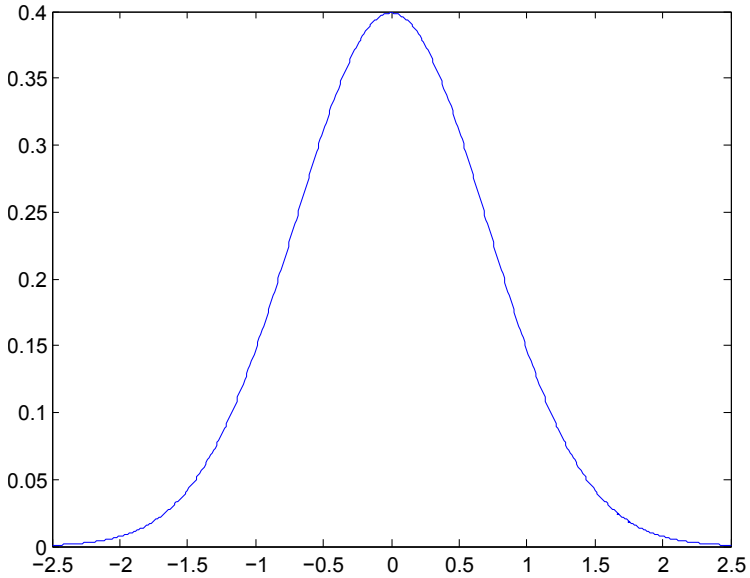


Figure 2.6: Gaussian function

The expected value for a discrete random variable X is

$$\hat{X} = E[X] = \sum_{i=1}^{m} x_i P(X = x_i). \tag{2.31}$$

For a continuous random variable X the expectation E(X) is defined as:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{2.32}$$

The variance of X

$$Var(X) = E[(X - E(X))^2] \tag{2.33}$$

is the expected value of the squared deviance of the random variable X to its expected value. The variance is calculated for the discrete and continuous case as follows:

$$Var(X) = \sum_i (x_i - \hat{x})^2 P(x_i) \qquad \text{for } X \text{ discrete,} \tag{2.34}$$

$$Var(X) = \int_{-\infty}^{\infty} (x - \hat{x})^2 f_X(x) dx \qquad \text{for } X \text{ continuous.} \tag{2.35}$$

The standard deviation is the square root of the variance

$$\sigma = \sqrt{Var(X)}. \tag{2.36}$$

The expected value of a random variable X is also called the first central moment, and the variance the second central moment.

Measures to quantify the dependence of two random variables X and Y are the cross-covariance C_{XY} and the cross-correlation R_{XY}. The cross-covariance is defined as

$$\begin{aligned} C_{XY} &= E[(X - \hat{X})(Y - \hat{Y})] \\ &= E[XY] - \hat{X}\hat{Y}. \end{aligned} \tag{2.37}$$

The cross-correlation is defined as

$$R_{XY} = E[XY]. \tag{2.38}$$

The cross-correlation coefficient is defined as

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y}. \tag{2.39}$$

Two random variables are linearly independent, if $\rho_{XY} = 0$.

By considering a n-dimensional random variable X and a m-dimensional random variable Y, the cross-covariance is defined as

$$C_{XY} = E[XY^T] = \begin{bmatrix} E[X_1Y_1] & \cdots & E[X_1Y_m] \\ \vdots & & \vdots \\ E[X_nY_1] & \cdots & E[X_nY_m] \end{bmatrix}. \tag{2.40}$$

The auto-correlation defines the correlation of a signal with itself and is given by

$$R_X = E[XX^T] = \begin{bmatrix} E[X_1X_1] & \cdots & E[X_1X_n] \\ \vdots & & \vdots \\ E[X_nX_1] & \cdots & E[X_nX_n] \end{bmatrix}. \tag{2.41}$$

2.2.4 Stochastic processes

According to [33] a stochastic process $X(t)$ is a mathematical model of a probabilistic experiment that evolves in time and generates a sequence of numerical values. Each numerical value is modeled by a random variable, so a stochastic process is simply a sequence of random variables. In general, the distribution function of a random process changes with time

$$F_X(x,t) = P(X(t) \leq x). \tag{2.42}$$

A special case is the stationary random process, which has a constant distribution function. The cross-correlation and auto-correlation function can also be applied to a stochastic processes. Given the values $X(t_1)$ and $X(t_2)$ or $Y(t_2)$, taken at time t_1 and t_2 the auto-correlation of a stochastic process is defined as

$$R_X(t_1,t_2) = E[X(t_1)X(t_2)^T], \tag{2.43}$$

and the cross-correlation as

$$R_{XY}(t_1,t_2) = E[X(t_1)Y(t_2)^T]. \tag{2.44}$$

If the auto-correlation $R_X(t_1,t_2)$ of a stochastic process is zero for all $t_1 \neq t_2$ then $X(t)$ is called white noise. Otherwise, X(t) is called colored noise. Random processes are often defined by the power spectral density, which is defined using the Fourier transformation of $X(t)$.

$$X(\omega) = \int_{-\infty}^{\infty} = X(t)e^{-j\omega t}dt. \tag{2.45}$$

Separating $X(\omega)$ in its real and imaginary parts

$$X(\omega) = X_R(\omega) + jX_I(\omega) \tag{2.46}$$

and its complex conjugate

$$X^*(\omega) = X_R(\omega) - jX_I(\omega), \tag{2.47}$$

the real-valued, scalar power spectral density function $S_X(\omega)$ results from

$$\begin{aligned} |X(\omega)|^2 &= X(\omega)X^*(\omega) \\ &= X_R(\omega) + X_I(\omega) \\ &= S_X(\omega). \end{aligned} \tag{2.48}$$

Another way to describe the power spectral density is given by the Fourier-transformation of the autocorrelation function:

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau. \tag{2.49}$$

For a discrete-time random process the power spectral density function is defined as

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X(k) e^{-j\omega k} \quad \omega \in [-\pi, \pi]. \tag{2.50}$$

A discrete-time random process X(k) is white noise if

$$R_X = \sigma^2 \delta_k, \tag{2.51}$$

where δ_k is the Kronecker delta function,

$$\delta_k = \begin{cases} 0, & \text{if } k \neq 0 \\ 1, & \text{if } k = 0 \end{cases}, \tag{2.52}$$

i.e. the auto-correlation for a white-noise signal is zero, except for a time-shift $\tau = 0$.

3 Train model

This section gives a detailed description of the train model and the equations of motion. In order to derive the equations of motion of a train model, it is important to have an accurate wheel rail contact model. The wheel rail contact determines the performance of the railway vehicle system. The contact model is described at the beginning of this section. Next, this section covers the components of the train model. The various bodies of the railway vehicle are presented and the structure of the suspension system is introduced. In the end, the equations of motion for the railway vehicle are given.

3.1 Wheel-Rail Contact

The wheel-rail-system has to perform several tasks: First, supporting the weight leads to vertical forces in the contact point. Then, lateral forces guide the vehicle along the track and third, forces tangential to the rail allow for acceleration or deceleration of the vehicle. The forces, which are needed to perform all these tasks are transferred in the contact area between wheel and rail. This contact area has the size of about $1.5cm^2$ and is influencing the whole railway dynamics.

For the FDI process, a reliable model of the railroad vehicle system is needed. A reliable model needs to include the wheel-rail-contact model. The formulation of the contact problem is a complex task and has been the subject of several investigations, which presented different solutions. In general, studies are based on one of two approaches: on the one side, there is the constraint approach and on the other side the elastic approach. In the first case, the wheel and rail surfaces are assumed to remain rigid and no separation or penetration are allowed. Kinematic algebraic constraint equations are used to describe the contact between wheel and rail. In the second case the wheel and rail surfaces are deformed in the contact region. The normal force between the wheel and rail is obtained by using force models depending on the intersection of the two surfaces and thus, separation and penetration are allowed.

The computation of the wheel-rail contact forces can be divided into three steps. The first step refers to the wheel and rail geometry and the determination of the con-

tact points with the constraint or elastic approach. In the second step, the creepages are determined, which measure the relative velocities between the wheel and the rail at the contact points. In the third step, the resulting contact forces are determined. In order to solve the wheel-rail contact problem, an accurate representation of the geometry of the wheel and the rail surfaces is required. To calculate the contact points and forces that act on the wheel and the rail, the radii of curvature and the tangent and normal vectors to the surfaces are needed in each point.

3.1.1 Wheel and rail geometry

The rail and wheel geometry differs from country to country and is given in engineering standards. The geometry of the wheel and the rail has a significant influence on the driving stability of a vehicle.

When the railway vehicle is running on a straight track with small unavoidable deviations, the task of the profile is to keep the wheelset in central position. Only at large deviations or when running through a curve, the wheel flanges of the wheelset must take the guiding function. Further, the increase in travel speed over a critical point can lead to instabilities, after small perturbation the wheel will not come back to the central position and, in the worst case build-up increasingly. Figure 3.1 shows an example for the wheel and rail profile.

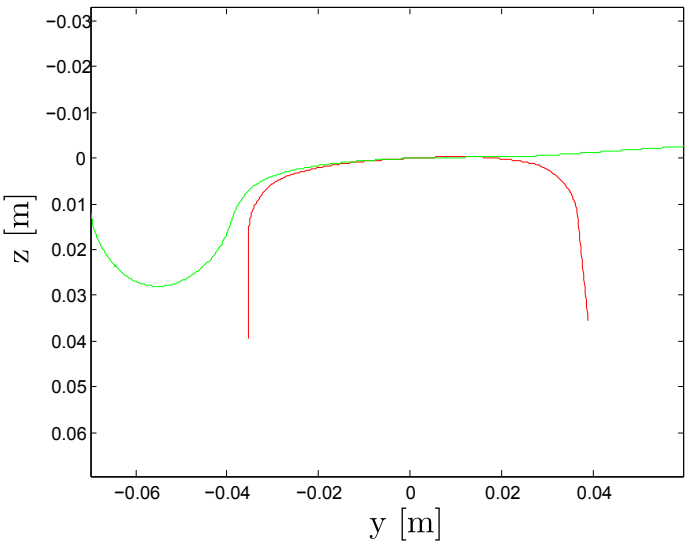


Figure 3.1: Wheel and rail geometry

Constantly acting forces during rolling, braking, accelerating and driving through curves create a permanent wear. On the one hand, this results in surface abrasion of the track, and on the other hand a surface abrasion of the wheel. In this work, the original wheel rail profile is used.

The task of the first step in computing the wheel rail contact forces is to find the contact point between wheel and rail. It is assumed that the yaw motion of the wheel is small, thus the problem is considered in two dimensions.

In the constraint approach, no indentation of the wheel and rail is allowed. When considering the two dimensional case the wheel has one degree of freedom. All parameters, which need to be calculated, are expressed in terms of the lateral wheel movement. The lateral movement is described by the y-coordinate. The vertical position $z(y)$, the roll angle $\phi(y)$ and $dz(y)/dy$ are of interest. In order to calculate the contact point for a given lateral position y, the Newton-Raphson method is used. The contact points at the wheel and rail surfaces are described by

$$\mathbf{p}_1^w = \boldsymbol{R}^w + \mathbf{A}_{rot}^w \bar{\mathbf{u}}_1^w, \tag{3.1}$$

$$\mathbf{p}_2^w = \boldsymbol{R}^w + \mathbf{A}_{rot}^w \bar{\mathbf{u}}_2^w, \tag{3.2}$$

$$\mathbf{p}_1^r = \boldsymbol{R}^r + \mathbf{A}_{rot}^r \bar{\mathbf{u}}_1^r, \tag{3.3}$$

$$\mathbf{p}_2^r = \boldsymbol{R}^r + \mathbf{A}_{rot}^r \bar{\mathbf{u}}_2^r \tag{3.4}$$

which is shown in Figure 3.2. The vector $\boldsymbol{R}$ describes the position of the body fixed coordinate system, $\bar{\mathbf{u}}$ is the vector from the body fixed coordinate system to the contact point and $\mathbf{A}_{rot}$ is the rotational matrix. The upper index stands for the

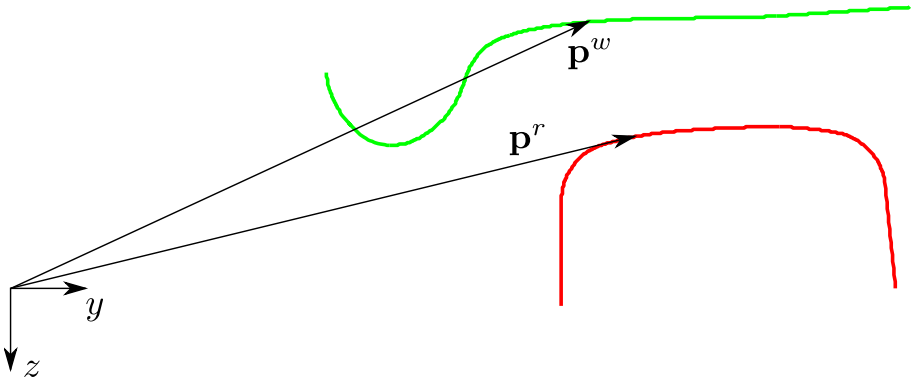


Figure 3.2: Wheel and rail contact point

contact point at the wheel w or at the rail r. The lower index determines the side of the train, 1 stands for the right, 2 for the left side.

Each wheel has one contact point, if the following equations are true

$$\mathbf{p}_1^w - \mathbf{p}_1^r = 0, \tag{3.5}$$

$$\mathbf{p}_2^w - \mathbf{p}_2^r = 0, \tag{3.6}$$

$$\mathbf{t}_1^w \cdot \mathbf{n}_1^r = 0, \tag{3.7}$$

$$\mathbf{t}_2^w \cdot \mathbf{n}_2^r = 0, \tag{3.8}$$

where $\mathbf{n}^r$ is the normal vector of the rail at the contact point and $\mathbf{t}^w$ the tangential vector of the rail surface at the contact point shown in Figure 3.3. The condition, that the contact point of the wheel is equal to the contact point of the rail, is satisfied by equations (3.5) and (3.6). Further, the two surfaces of the bodies must have the same tangent planes at the contact point, thus the tangential vector of the wheel must be normal to the normal vector of the rail. This is given by equations (3.7) and (3.8).

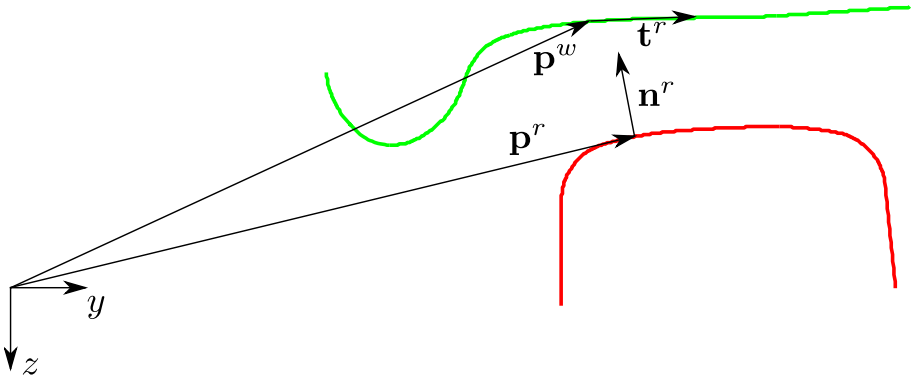


Figure 3.3: Tangential and normal vector at the contact point

In the two dimensional case, the rotational matrix results to

$$\mathbf{A}_{rot} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}. \tag{3.9}$$

The vector $\bar{\mathbf{u}}$ describes the vectors from the body fixed coordinate system to the contact point in coordinates of the body fixed coordinate system.

$$\bar{\mathbf{u}}_1^w = \begin{bmatrix} 0 \\ s_1^w \\ f_1(s_1^w) \end{bmatrix} \quad \bar{\mathbf{u}}_2^w = \begin{bmatrix} 0 \\ s_2^w \\ f_2(s_2^w) \end{bmatrix} \quad \bar{\mathbf{u}}_1^r = \begin{bmatrix} 0 \\ s_1^r \\ g_1(s_1^r) \end{bmatrix} \quad \bar{\mathbf{u}}_2^r = \begin{bmatrix} 0 \\ s_2^r \\ g_2(s_2^r) \end{bmatrix}. \tag{3.10}$$

The functions f and g describe the geometry of the wheel and rail, respectively, and s^w and s^r are the lateral surface parameters, which represent the independent variable for the wheel and rail surface. The tangential vectors of the wheels and the normal vectors of the rails result to

$$\bar{\mathbf{t}}_1^w = \frac{1}{\sqrt{1+(f_1')^2}} \begin{bmatrix} 0 \\ 1 \\ f_1' \end{bmatrix} \quad \bar{\mathbf{t}}_2^w = \frac{1}{\sqrt{1+(f_2')^2}} \begin{bmatrix} 0 \\ 1 \\ f_2' \end{bmatrix} \tag{3.11}$$

$$\bar{\mathbf{n}}_1^r = \frac{1}{\sqrt{1+(g_1')^2}} \begin{bmatrix} 0 \\ g_1' \\ -1 \end{bmatrix} \quad \bar{\mathbf{n}}_2^r = \frac{1}{\sqrt{1+(g_2')^2}} \begin{bmatrix} 0 \\ g_2' \\ -1 \end{bmatrix}. \tag{3.12}$$

This leaves four independent unknown parameters ϕ, s_1^w, s_2^w and r_z, where r_z is the vertical component of the body fixed coordinate system of the wheelset. The four equations to solve this problem are

$$p_{1_z}^w - p_{1_z}^r = 0, \tag{3.13}$$
$$p_{2_z}^w - p_{2_z}^r = 0, \tag{3.14}$$
$$[t_{1_y}^w \cos\phi - t_{1_z}^w \sin\phi] n_{1_y}^r + [t_{1_y}^w \sin\phi + t_{1_z}^w \cos\phi] n_{1_z}^r = 0, \tag{3.15}$$
$$[t_{2_y}^w \cos\phi - t_{2_z}^w \sin\phi] n_{2_y}^r + [t_{1_y}^w \sin\phi + t_{1_z}^w \cos\phi] n_{2_z}^r = 0. \tag{3.16}$$

Inserting equations (3.11) and (3.12) into equations (3.13)-(3.16) and expanding yields

$$r_z + s_1^w \sin\phi + f_1(s_1^w)\cos\phi - g_1(s_1^r) = h_1 = 0,$$
$$r_z + s_2^w \sin\phi + f_2(s_2^w)\cos\phi - g_2(s_2^r) = h_2 = 0,$$
$$\frac{1}{\sqrt{1+f_1'(s_1^w)^2}} \frac{1}{\sqrt{1+g_1'(s_1^r)^2}} ([\cos\phi - f_1'(s_1^w)\sin\phi] g_1(s_1^r)' - [\sin\phi + f_1'(s_1^w)\cos\phi]) = h_3 = 0,$$
$$\frac{1}{\sqrt{1+f_2'(s_2^w)^2}} \frac{1}{\sqrt{1+g_2'(s_2^r)^2}} ([\cos\phi - f_2'(s_2^w)\sin\phi] g_2(s_2^r)' - [\sin\phi + f_2'(s_2^w)\cos\phi]) = h_4 = 0,$$

with

$$s_1^r = r_y + s_1^w \cos\phi - f_1(s_1^w)\sin\phi, \tag{3.17}$$
$$s_2^r = r_y + s_2^w \cos\phi - f_2(s_2^w)\sin\phi. \tag{3.18}$$

Starting from an initial guess, the contact point is calculated numerically with the Newton Raphson method

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{Jac}^{-1}\mathbf{h}(\mathbf{x}_n) \tag{3.19}$$

where **Jac** is the Jacobian matrix defined as

$$\mathbf{Jac} = \begin{bmatrix} \frac{\partial h_1}{\partial r_z} & \frac{\partial h_1}{\partial \phi} & \frac{\partial h_1}{\partial s_1^w} & \frac{\partial h_1}{\partial s_2^w} \\ \frac{\partial h_2}{\partial r_z} & \frac{\partial h_2}{\partial \phi} & \frac{\partial h_2}{\partial s_1^w} & \frac{\partial h_2}{\partial s_2^w} \\ \frac{\partial h_3}{\partial r_z} & \frac{\partial h_3}{\partial \phi} & \frac{\partial h_3}{\partial s_1^w} & \frac{\partial h_3}{\partial s_2^w} \\ \frac{\partial h_4}{\partial r_z} & \frac{\partial h_4}{\partial \phi} & \frac{\partial h_4}{\partial s_1^w} & \frac{\partial h_4}{\partial s_2^w} \end{bmatrix}, \tag{3.20}$$

and $\mathbf{x} = \begin{bmatrix} \phi & s_1^w & s_2^w & r_z \end{bmatrix}^T$.

In the elastic approach, the wheelsets are considered as six dimensional. At the point of contact, the two touching bodies deform. The normal contact forces are calculated using Hertz contact theory. There are two advantages of the elastic approach over the constraint approach: Firstly, it allows the separation of the wheel and the rail, thus track irregularities can be applied more easily. Secondly, the management of multiple contact points is simplified. Because of these advantages, the elastic approach is adopted in this work.

The procedure of finding the contact point is based on the so-called DIFF method [34]. This method is based on the idea, that the contact points minimize the difference between the wheel surface and the rail surface in the vertical direction.

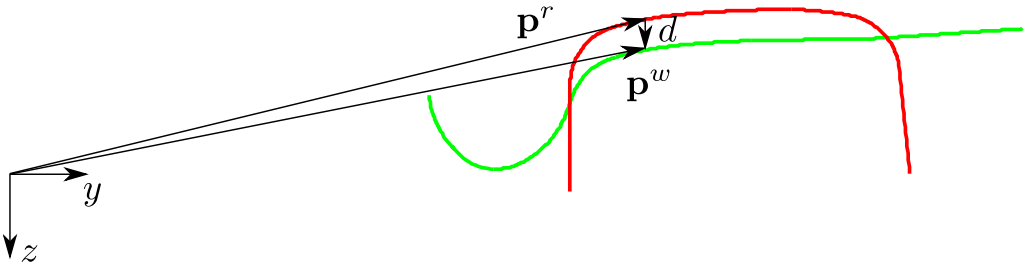


Figure 3.4: Wheel and rail intersection

Given a contact point at the wheel $\mathbf{p}^w$ the contact point at the rail is calculated as the intersection between the rail surface and a line parallel to the z-axis passing through the wheel contact point $\mathbf{p}^w$. With the assumption, that the yaw motion of the wheel is small, the contact point of the rail depends on the y-position of the contact point of the wheel. Thus, the difference between the rail contact point and the wheel contact point leads to

$$d = p_z^r(y_w) - p_z^w(y_w) \tag{3.21}$$

and is shown in Figure 3.4. Of all possible sets of contact points, the set of contact points which minimizes the difference d is used. The smallest value of d is denoted by δ and is used to calculate the normal force according to the theory of Hertz. If there is more than one local minimum with negative sign, all these points are considered as contact point. In this way it is possible to deal with multiple contact points.

3.1.2 Hertz theory

After determining the contact point, the contact force normal to the contact plane is calculated. This is done with the Hertz contact force model

$$f_{Hertz} = K_{Hertz}\delta^{\frac{3}{2}}, \tag{3.22}$$

where δ represents the relative compression of the touching bodies and K_{Hertz} is based on the studies of Hertz.

To calculate the normal force with Hertz theory, some assumptions are made. On the one side, these assumptions concern the geometry of the two touching bodies at the contact point, on the other side, the assumptions also refer to the material properties. Depending on the shape of the two bodies in the area of contact, the problem can be conformal or non-conformal. If the contacting shape fits even closely together, conformal contact occurs. If the two bodies touch at a line or at a point, non-conformal contact occurs.

For the non-conformal contact problem, Heinrich Hertz presented a solution to calculate the contact area and pressure distribution. According to [35] the assumptions used in Hertz theory are:

1. The surfaces of the bodies are continuous and non-conformal.
2. The strains are small.
3. The stress resulting from the contact force vanishes at a distance far from the contact area.
4. The surfaces are frictionless.
5. The bodies are elastic, and no plastic deformation occurs in the contact area.

For the wheel-rail contact case, this is found to give a good approximation. The Hertz theory assumes that the area of contact is elliptical. To describe the two elliptical surfaces in the area of contact, two axes z_i and z_j, which are normal to the tangent, are attached to the two touching bodies i and j, respectively. The origins of both axes are at the contact point $\mathbf{p}$ and are considered as positive towards the interior of the bodies. In the area of the contact point $\mathbf{p}$, the elliptical shape of the surface of each body is described by the equations

$$z_i = A_i x_i^2 + B_i y_i^2 + C_i x_i y_i + ... \tag{3.23}$$

$$z_j = A_j x_j^2 + B_j y_j^2 + C_j x_j y_j + ... \tag{3.24}$$

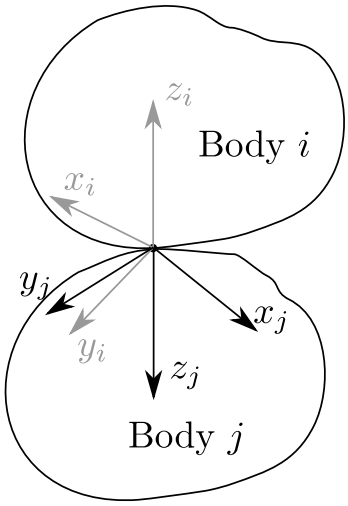


Figure 3.5: Two bodies in contact

where A_i, B_i, C_i, A_j, B_j and C_j are constants depending on the radii of the curvature of the two bodies at the contact point. By introducing a new coordinate system the distance between the two surfaces near the origin is defined as

$$h = z_i + z_j = Ax^2 + By^2 + Cxy. \tag{3.25}$$

By choosing the orientation of the x and y axes in that way, that xy vanishes, the upper equation leads to

$$h = z_i + z_j = Ax^2 + By^2. \tag{3.26}$$

The constants A and B again depend on the geometric shape of the two surfaces near the contact point $\mathbf{p}$. A and B can be expressed as

$$A + B = \frac{1}{2}\left(\frac{1}{R_{i1}} + \frac{1}{R_{i2}} + \frac{1}{R_{j1}} + \frac{1}{R_{j2}}\right) \tag{3.27}$$

$$B - A = \frac{1}{2}\sqrt{\left(\frac{1}{R_{i1}} - \frac{1}{R_{i2}}\right)^2 + \left(\frac{1}{R_{j1}} - \frac{1}{R_{j2}}\right)^2 + 2\left(\frac{1}{R_{i1}} - \frac{1}{R_{i2}}\right)\left(\frac{1}{R_{j1}} - \frac{1}{R_{j2}}\right)\cos 2\Psi}, \tag{3.28}$$

where R_{k1} and R_{k2} are the principle radii of curvature of the surfaces of bodies j and i at the origin.

If the center of curvature lies within the body, the curvature is assumed to be

positive. Further, an auxiliary angle θ is defined as

$$\cos(\theta) = \frac{B - A}{B + A}. \tag{3.29}$$

Knowing the radii of curvature at the contact point $\mathbf{p}$, equations (3.27) and (3.28) can be solved. The auxiliary angle θ is calculated from equation (3.29). The principal radii of curvature are calculated from $\partial^2 z_k / \partial {x_k}^2 = 1/R_{k1}$ and $\partial^2 z_k / \partial {y_k}^2 = 1/R_{k2}$.

Since the contact area is assumed to be small compared to the dimensions of the two bodies, one can consider the contact area of the two bodies as semi-infinite. The contact pressure is assumed to satisfy the following requirements for the equilibrium of the two bodies [34, 35]:

1. The total applied force F_n must be equal to the total resisting force generated by the vertical component of the pressure p in the contact area, that is,

$$F_n = \int\int p \; dxdy. \tag{3.30}$$

2. The components of displacement vanish at infinity, therefore, the displacement at a distance away from the contact region can be neglected.

3. The normal stresses outside the contact region are assumed to be zero.

4. The normal stresses acting on the two bodies are in balance within the contact region.

5. The shear stresses τ_{xz} and τ_{yz} along the surfaces of the bodies are zeros.

If the pressure p is a quadratic function of x and y, these conditions can be satisfied. The pressure distribution in the contact area is then given by

$$p = p_0 \sqrt{1 - \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2}, \tag{3.31}$$

where a and b are the lengths of the ellipse semi-axes and p_0 is the maximal pressure.

Resulting from the semi-ellipsoidal pressure distribution, the total normal load F_n is given by

$$F_n = \frac{2}{3} p_0 \pi a b. \tag{3.32}$$

Using equations (3.31) and (3.32), one obtains [34, 35, 36]

$$p = \frac{3F_n}{2\pi a b} \sqrt{1 - \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2}. \tag{3.33}$$

With the two constants K_1 and K_2

$$K_1 = \frac{1 - {\nu_i}^2}{\pi E_i}, \tag{3.34}$$

$$K_2 = \frac{1 - {\nu_j}^2}{\pi E_j}, \tag{3.35}$$

which depend on the material properties of the two bodies, the contact ellipse semi-axes can be calculated from

$$a = m_H \left(\frac{3\pi F_n (K_1 + K_2)}{4(A + B)} \right)^{1/3}, \tag{3.36}$$

$$b = n_H \left(\frac{3\pi F_n (K_1 + K_2)}{4(A + B)} \right)^{1/3}. \tag{3.37}$$

The coefficients m_H and n_H in equations (3.36) and (3.37) are needed to calculate the semi-axes a and b. The coefficients m_H and n_H are given by Hertz in Table 3.1 as functions of the angular parameter θ, for the values of θ between 0 deg and 180 deg [34].

To get m_H and n_H for an arbitrary value of the angle θ, the entries of the table must be interpolated.

To increase the calculation speed, a closed-form expression for the coefficients m_H and n_H as functions of θ according to [34] is given by

$$m_H = A_m \tan(\theta - \pi/2) + \frac{B_m}{\theta^{C_m} + D_m}, \tag{3.38}$$

$$n_H = \frac{1}{A_n \tan(\theta - \pi/2) + 1} + B_n \theta^{C_m} + D_n \sin(\theta). \tag{3.39}$$

Here, the value of θ is given in radians, and the coefficients A_k, B_k, C_k, and D_k ($k = m_H, n_H$) are given in Table 3.2 [37].

Finally, the normal force F_h according to Hertz can be calculated by

$$F_h = K_{Hertz} \delta^{3/2} = \frac{4\beta}{3(K_1 + K_2)\sqrt{A + B}} \delta^{3/2}, \tag{3.40}$$

where β is a constant and is given in Table 3.3 [35, 38].

For the computer implementation, the following steps are performed: The radii of curvature R_{i1}, R_{i2}, R_{j1} and R_{j2} of the two bodies near the contact point must be known. With the radii of curvature the constants A and B are evaluated from equations (3.27) and (3.28). With A and B the auxiliary angle is computed. After determining the constants K_1 and K_2 from equations (3.34) and (3.35) the semi-axes a and b are calculated from the equations (3.36) and (3.37) and Table 3.1.

Table 3.1: Hertz coefficients m and n

θ (deg)	m_H	n_H	θ (deg)	m_H	n_H
0.5	61.4	0.1018	40	2.136	0.567
1	36.89	0.1314	45	1.926	0.604
1.5	27.48	0.1522	50	1.754	0.641
2	22.26	0.1691	55	1.611	0.678
3	16.50	0.1964	60	1.486	0.717
4	13.31	0.2188	65	1.378	0.759
6	9.790	0.2552	70	1.284	0.802
8	7.860	0.2850	75	1.202	0.846
10	6.604	0.3112	80	1.128	0.893
20	3.813	0.4125	85	1.061	0.944
30	2.731	0.4930	90	1.0	1.0
35	2.397	0.5300			

Table 3.2: Coefficients used for the closed-form functions m and n

Coeff.	Value	Coeff.	Value
A_m	-1.086419052477	A_n	-0.773444080706
B_m	-0.106496432832	B_n	0.256695354565
C_m	1.350000000000	C_n	0.200000000000
D_m	1.057885958251	D_n	-0.280958376499

Table 3.3: Hertz coefficient β for Hertz force

A/B	β
1.0	0.3180
0.7041	0.3215
0.4903	0.3322
0.3333	0.3505
0.2174	0.3819
0.1325	0.4300
0.0718	0.5132
0.0311	0.6662
0.0076	1.1450

According to [39], a damping force is added to the Hertzian component. For a given intersection δ, the damping force is proportional to the velocity of indentation in the contact point. The expression of the normal force F_n is then given by

$$F_n = F_h + F_d = K_{Hertz}\delta^{3/2} + C_h\dot{\delta}|\delta|. \tag{3.41}$$

The velocity of indentation $\dot{\delta}$ is calculated as the dot product of the relative velocity vector of the contact points and the normal vector to the surface at the contact point. To guarantee that the contact force is zero when the indentation is zero, the factor $|\delta|$ is added. The value C_h is a damping coefficient.

The normal forces are added to the dynamic equations of motion as generalized applied forces.

3.1.3 Creep forces

The relative motion of two touching bodies can be the result of rolling and sliding motion. Tangential forces occur, when the two bodies have different velocities at the contact point and different angular velocities. When two elastic bodies are in contact, some points of the contact area slip, while other points stick. Figure 3.6 shows the first two-dimensional approximation of the contact area according to Carter [40].

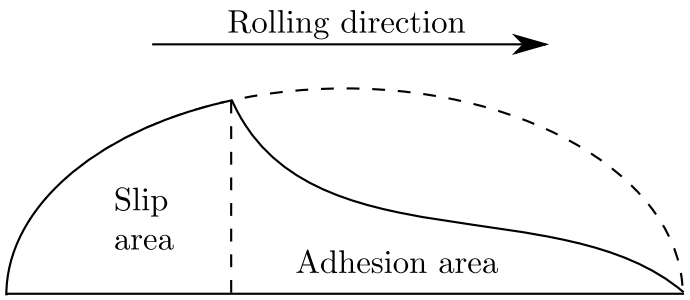


Figure 3.6: Stick slip according to Carter [40]

The contact area can be divided in two parts, one stick area and one slip area. This mix of elastic deformation and local slipping is known as creepage. For very small creepage, Kalker suggested, that the area of slip is very small and can be neglected [41]. Thus, the area of adhesion is equal to the area of contact. To calculate the tangential forces, the normal force as well as the creepage is needed. The creepage is split into a longitudinal, a lateral and a spin part and is calculated according to [42] with

$$\begin{aligned}\gamma_x &= \frac{v_x^w - v_x^r}{V},\\ \gamma_y &= \frac{v_y^w - v_y^r}{V},\\ \omega_z &= \frac{\Omega_z^w - \Omega_z^r}{V},\end{aligned} \tag{3.42}$$

where V is the magnitude of the wheel velocity and v^w and v^r are the velocities at the contact point of the wheel and the rail. The values Ω_z^w and Ω_z^r are the different angular velocities of the contact point at the wheel and the rail.

With the normal force and the creepages, the creep forces are calculated according to the linear theory of Kalker. As stated in [42] the creep forces and moments are determined from

$$\begin{aligned}F_x &= -Gabc_{11}\gamma_x,\\ F_y &= -Gabc_{22}\gamma_y - G(ab)^{1.5}c_{23}\omega_z,\\ M_z &= -G(ab)^{1.5}c_{32}\gamma_y - G(ab)^2c_{33}\omega_z,\end{aligned} \tag{3.43}$$

where G is the modulus of rigidity and a is the contact ellipse semi-axis in the rolling direction and b is the contact ellipse semi-axis dimension in the lateral direction. Kalker calculated the modulus of rigidity G with the modulus of rigidity of the wheel G_w and the rail G_r

$$G = \frac{1}{2}\left(\frac{1}{G_w} + \frac{1}{G_r}\right). \tag{3.44}$$

The creep forces as well as the normal force at the wheel are shown in Figure 3.7.

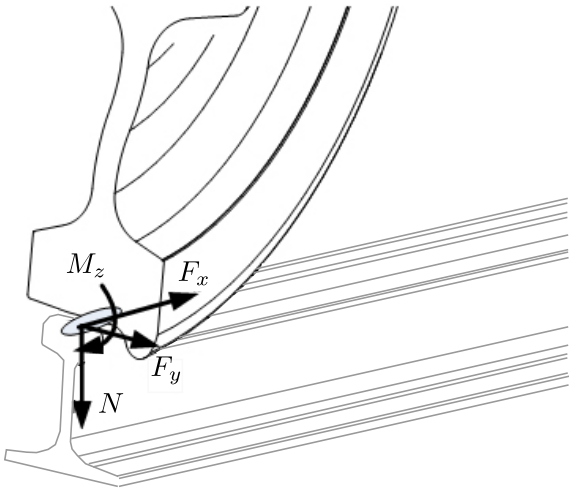


Figure 3.7: Wheel rail force [43]

The creepage coefficients c_{ji}, that depend only on Poisson's ratio ν and the ratio of the semi-axis of the contact ellipse, are given in Table 3.4 [35] at which $c_{32} = -c_{23}$.

Table 3.4: Kalker's creepage and spin coefficients

	c_{11}			c_{22}			c_{23}			c_{33}		
g	ν=0	0.25	0.5	ν=0	0.25	0.5	ν=0	0.25	0.5	ν=0	0.25	0.5
a/b												
0.1	2.51	3.31	4.85	2.51	2.52	2.53	0.334	0.473	0.731	6.42	8.28	11.7
0.2	2.59	3.37	4.81	2.59	2.63	2.66	0.483	0.603	0.809	3.46	4.27	5.66
0.3	2.68	3.44	4.8	2.68	2.75	2.81	0.607	0.715	0.889	2.49	2.96	3.72
0.4	2.78	3.53	4.82	2.78	2.88	2.98	0.720	0.823	0.977	2.02	2.32	2.77
0.5	2.88	3.62	4.83	2.88	3.01	3.14	0.827	0.929	1.07	1.74	1.93	2.22
0.6	2.98	6.72	4.91	2.98	3.14	33.1	0.930	1.03	1.18	1.56	1.68	1.86
0.7	3.09	3.81	4.97	3.09	3.28	3.48	1.03	1.14	1.29	1.43	1.50	1.60
0.8	3.19	3.91	5.05	3.19	3.41	3.65	1.13	1.25	1.40	1.34	1.37	1.42
0.9	3.29	4.01	5.12	3.29	3.54	3.82	1.23	1.36	1.51	1.27	1.27	1.27
b/a												
1.0	3.4	4.12	5.2	3.40	3.67	3.98	1.33	1.47	1.63	1.21	1.19	1.16
0.9	3.51	4.22	5.3	3.51	3.81	4.16	1.44	1.57	1.77	1.16	1.11	1.06
0.8	3.65	4.36	5.42	3.65	3.99	4.39	1.58	1.75	1.94	1.10	1.04	0.954
0.7	3.82	4.54	5.58	3.82	4.21	4.67	1.76	1.95	2.18	1.05	0.965	0.852
0.6	4.06	4.78	5.8	4.06	4.50	5.04	2.01	2.23	2.50	1.01	0.892	0.751
0.5	4.37	5.10	6.11	4.37	4.90	5.56	2.35	2.62	2.96	0.958	0.819	0.650
0.4	4.84	5.57	6.57	4.84	5.48	6.31	2.88	3.24	3.70	0.912	0.747	0.549
0.3	5.57	6.34	7.34	5.57	6.40	7.51	3.79	4.32	5.01	0.868	0.674	0.446
0.2	6.96	7.78	8.82	6.96	8.14	9.79	5.72	6.63	7.89	0.828	0.601	0.341
0.1	10.7	11.7	12.9	10.7	12.8	16.0	12.2	14.6	18.0	0.795	0.526	0.228

Equivalent to the normal forces, the creep forces are added to the dynamic equations of motion as generalized applied forces.

3.1.4 Rail irregularities

The vehicle does not travel along a perfect path, neither on a straight track nor on a curved track. Track irregularities act as disturbances, that stimulate the entire vehicle to vibrate and are noticed by passengers. The excitation of a railway vehicle results mainly from geometric irregularities of the rails or the wheels. By assuming that the wheels are perfectly shaped, the only irregularities are given by the rail track. The rail track irregularities are considered as random and occur in different directions. The random track irregularities are commonly characterized by their power spectral densities (PSDs). These spectral characteristics are obtained from exemplary track measurements and depend on the specific track, which is described more detailed in [44].

In this work, the train model is excited randomly due to horizontal, vertical and cross level track irregularities, as shown in Figures 3.8 and 3.9.

Track gauge irregularities are not considered since they are changing the wheel-

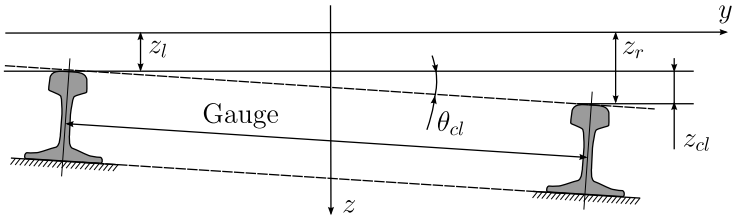


Figure 3.8: Crosslevel track irregularities

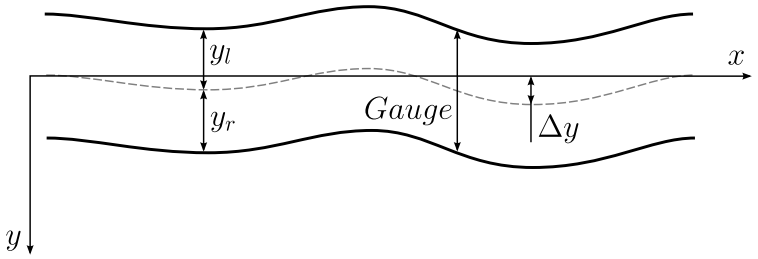


Figure 3.9: Horizontal track irregularities

rail contact geometry, thus they do not influence input forces directly but rather influence the characteristic of the system. The track is assumed in this work to be rigid and immovable

The power spectral densities (PSDs) for horizontal, vertical and cross-level track irregularities are considered as defined in ERRI B176 [46]. According to ERRI B176, the polynomials of the PSDs for track irregularities are

$$\begin{aligned} S_h(\Omega) &= \frac{b_{h0}}{a_{h0} + a_{h2}\Omega^2 + \Omega^4} \\ &= \frac{b_{h0}}{0.00028855 + 0.6803895\Omega^2 + \Omega^4}, \\ S_v(\Omega) &= \frac{b_{v0}}{a_{v0} + a_{v2}\Omega^2 + \Omega^4} \\ &= \frac{b_{v0}}{0.00028855 + 0.6803895\Omega^2 + \Omega^4}, \\ S_{cl}(\Omega) &= \frac{b_{cl2}\Omega^2}{a_{cl0} + a_{cl2}\Omega^2 + a_{cl4}\Omega^4 + \Omega^6} \\ &= \frac{b_{cl2}\Omega^2}{5.535659 \cdot 10^{-5} + 0.1308172\Omega^2 + 0.8722335\Omega^4 + \Omega^6} \quad . \end{aligned} \tag{3.45}$$

where Ω is the spatial frequency in $\frac{rad}{m}$. The units of the PSDs are $\frac{m^2}{rad/m}$ in lateral and vertical direction and $\frac{rad^2}{rad/m}$ in cross-level direction. For a vehicle traveling with the velocity V the angular frequency ω in $\frac{rad}{s}$ is defined by $\omega = V\Omega$. Figure 3.10

shows the PSD for the low vertical and Figure 3.11 shows the PSD for the low horizontal track irregularities. Figure 3.12 shows the PSD for the low cross-level track irregularity.

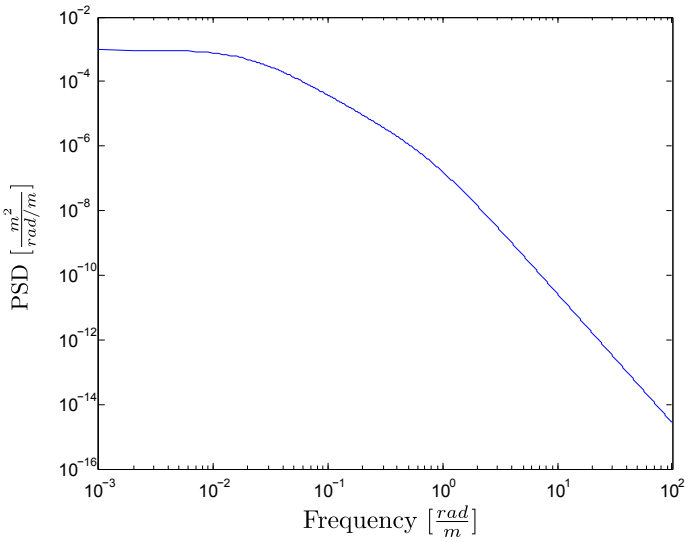


Figure 3.10: Vertical track irregularities

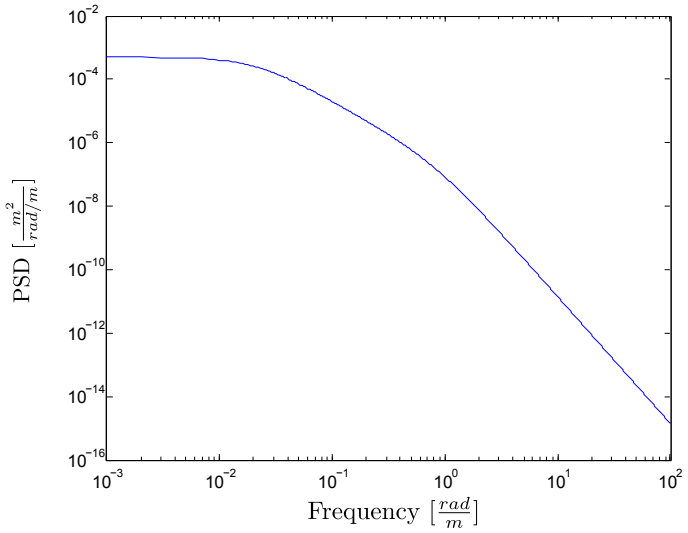


Figure 3.11: Horizontal track irregularities

The information of the spectral characteristics of the unknown input are used to increase the accuracy of the fault detection method. This is achieved by extending the state space model of the train with a form filter containing the information about the track irregularities.

The coefficients of the form filter are calculated from the PSD. According to [47] a PSD $S_X(\omega)$ can be formulated with the PSD S_0 of white noise as following:

$$S_X(\omega) = S_0 \cdot |F(j\omega)|^2 \tag{3.46}$$

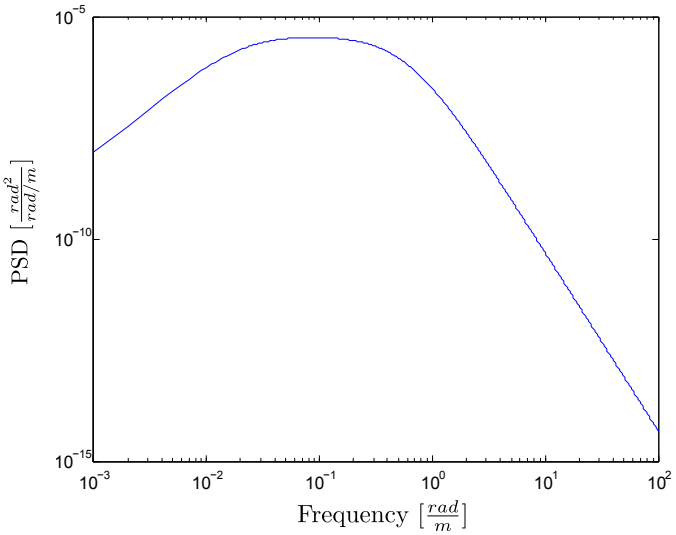


Figure 3.12: Cross-level track irregularities

where $F(j\omega)$ is the transfer function of the form filter. The transfer function $F(j\omega)$ of the form filter can be derived through factorization of the PSD polynomial since

$$|F(j\omega)|^2 = F(j\omega) \cdot F^*(j\omega) \tag{3.47}$$

where $F^*(j\omega)$ is the complex conjugate of the transfer function $F(j\omega)$. The transformation with $s = j\omega$ leads to

$$S_X(s) = S_0 F(-s)F(s). \tag{3.48}$$

Substituting the spatial frequency Ω with $s = j\omega$ and $\Omega = \frac{s}{jV}$ and factorizing the PSD polynomials of the track irregularities, equation (3.45) are rewritten as

$$\begin{aligned}
S_h(s) =& \frac{b_{h0}}{0.00028855 + 0.6803895(s/V)^2 + (s/V)^4} \\
=& \left(\frac{\sqrt{b_{h0}}V^2}{s^2 + 0.8452Vs + 0.01698676V^2}\right)\left(\frac{\sqrt{b_{h0}}V^2}{s^2 - 0.8452Vs + 0.01698676V^2}\right), \\
S_v(s) =& \frac{b_{v0}}{0.00028855 + 0.6803895(s/V)^2 + (s/V)^4} \\
=& \left(\frac{\sqrt{b_{v0}}V^2}{s^2 + 0.8452Vs + 0.01698676V^2}\right)\left(\frac{\sqrt{b_{v0}}V^2}{s^2 - 0.8452Vs + 0.01698676V^2}\right), \\
S_{cl}(s) =& \frac{b_{cl2}(s/V)^2}{5.535659 \cdot 10^{-5} + 0.1308172(s/V)^2 + 0.8722335(s/V)^4 + (s/V)^6} \\
=& \left(\frac{\sqrt{b_{cl2}}sV^2}{0.00744V^3 + 0.387184V^2s + 1.2832Vs^2 + s^3}\right)\ldots \\
& \left(\frac{\sqrt{b_{cl2}}sV^2}{0.00744V^3 - 0.387184V^2s + 1.2832Vs^2 - s^3}\right) \quad .
\end{aligned} \tag{3.49}$$

Thus, the transfer functions for the horizontal $F_h(s)$, vertical $F_v(s)$ and cross-level $F_{cl}(s)$ track irregularities result to

$$\begin{aligned} F_h(s) =& \frac{\sqrt{b_{h0}}V^2}{s^2 + 0.8452Vs + 0.01698676V^2}, \\ F_v(s) =& \frac{\sqrt{b_{v0}}V^2}{s^2 + 0.8452Vs + 0.01698676V^2}, \\ F_{cl}(s) =& \frac{\sqrt{b_{cl2}}sV^2}{s^3 + 1.2832Vs^2 + 0.387184V^2s + 0.00744V^3}. \end{aligned} \tag{3.50}$$

The values of b_{h0}, b_{v0} and b_{cl2} are given in Table 3.5 for low and high track irregularities.

Using the transfer functions $F_h(s)$, $F_v(s)$ and $F_{cl}(s)$ from above, the form filter is given as state space model [45]

$$\dot{\mathbf{x}}_{ffi} = \mathbf{A}_{ffi}\mathbf{x}_{ffi} + \mathbf{B}_{ffi}\mathbf{w}_{ffi}, \tag{3.51}$$

$$\mathbf{y}_{ffi} = \mathbf{C}_{ffi}\mathbf{x}_{ffi}, \tag{3.52}$$

with the matrices

$$\mathbf{A}_{ffh} = \begin{bmatrix} 0 & 1 \\ -0.01698676V^2 & -0.8452V \end{bmatrix}, \ \mathbf{B}_{ffh} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{3.53}$$

$$\mathbf{C}_{ffh} = \begin{bmatrix} \sqrt{b_{h0}}V^2 & 0 \\ 0 & \sqrt{b_{h0}}V^2 \end{bmatrix}, \tag{3.54}$$

$$\mathbf{A}_{ffv} = \begin{bmatrix} 0 & 1 \\ -0.01698676V^2 & -0.8452V \end{bmatrix}, \ \mathbf{B}_{ffv} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{3.55}$$

$$C_{ffv} = \begin{bmatrix} \sqrt{b_{v0}}V^2 & 0 \\ 0 & \sqrt{b_{v0}}V^2 \end{bmatrix}, \tag{3.56}$$

Table 3.5: Values of b_{h0}, b_{v0} and b_{cl2}

	b_{h0}	b_{v0}	b_{cl2}
low	$1.440846 \cdot 10^{-7}$	$2.741619 \cdot 10^{-7}$	$4.87399 \cdot 10^{-7}$
high	$4.164787 \cdot 10^{-7}$	$7.343623 \cdot 10^{-7}$	$1.305533 \cdot 10^{-6}$

$$\mathbf{A}_{ffcl} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.00744V^3 & -0.387184V^2 & -1.2832V \end{bmatrix}, \ \mathbf{B}_{ffcl} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \tag{3.57}$$

$$\mathbf{C}_{ffcl} = \begin{bmatrix} 0 & \sqrt{b_0}V^2 & 0 \\ 0 & 0 & \sqrt{b_0}V^2 \end{bmatrix} \quad . \tag{3.58}$$

Combining all three track irregularities in one set of equations the resulting matrices are

$$\mathbf{A}_{ff} = \begin{bmatrix} \mathbf{A}_{ffh} & 0 & 0 \\ 0 & \mathbf{A}_{ffv} & 0 \\ 0 & 0 & \mathbf{A}_{ffcl} \end{bmatrix}, \ \mathbf{B}_{ff} = \begin{bmatrix} \mathbf{B}_{ffh} & 0 & 0 \\ 0 & \mathbf{B}_{ffv} & 0 \\ 0 & 0 & \mathbf{B}_{ffcl} \end{bmatrix}, \tag{3.59}$$

$$\mathbf{C}_{ff} = \begin{bmatrix} \mathbf{C}_{ffh} & 0 & 0 \\ 0 & \mathbf{C}_{ffv} & 0 \\ 0 & 0 & \mathbf{C}_{ffcl} \end{bmatrix}. \tag{3.60}$$

The matrix $\mathbf{A}_{ff}$ has the dimension 7×7, matrix $\mathbf{B}_{ff}$ 7×3 and matrix $\mathbf{C}_{ff}$ 6×7. The excitation of the wheelsets is assumed to be independent, i.e. any correlation due to the wheels running on the same track is neglected. By the assumption given above, the state space model for the form filter including all three track irregularities for all four wheelsets is given by:

$$\dot{\mathbf{x}}_{FF} = \mathbf{A}_{FF}\mathbf{x}_{FF} + \mathbf{B}_{FF}\mathbf{w}_{FF}, \tag{3.61}$$

$$\mathbf{y}_{FF} = \mathbf{C}_{FF}\mathbf{x}_{FF}, \tag{3.62}$$

with the matrices

$$\mathbf{A}_{FF} = \begin{bmatrix} \mathbf{A}_{ff} & 0 & 0 & 0 \\ 0 & \mathbf{A}_{ff} & 0 & 0 \\ 0 & 0 & \mathbf{A}_{ff} & 0 \\ 0 & 0 & 0 & \mathbf{A}_{ff} \end{bmatrix}, \ \mathbf{B}_{FF} = \begin{bmatrix} \mathbf{B}_{ff} & 0 & 0 & 0 \\ 0 & \mathbf{B}_{ff} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{ff} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{ff} \end{bmatrix}, \tag{3.63}$$

$$\mathbf{C}_{FF} = \begin{bmatrix} \mathbf{C}_{ff} & 0 & 0 & 0 \\ 0 & \mathbf{C}_{ff} & 0 & 0 \\ 0 & 0 & \mathbf{C}_{ff} & 0 \\ 0 & 0 & 0 & \mathbf{C}_{ff} \end{bmatrix}. \tag{3.64}$$

3.2 Multibody simulation of railway vehicle systems

The applied fault detection and isolation procedure, as well as the force estimation requires knowledge of the dynamic behavior of the railway vehicle. In order to describe the dynamic behavior of a railway vehicle, the system must be described as a mechanical model. Modern mechanical systems, like the railway vehicle, are often very complex and consist of many components interconnected by joints and force elements such as springs, dampers, and actuators. In order to derive the mechanical model, the positions of these different force elements at the bodies are needed, as well as the characteristics of the elements.

The accuracy of the train model plays a major role for the fault detection procedure and is indispensable for a precise simulation process. The drawback of a complex train model lies in the high calculation efforts. Thus, the complexity of the train model must be a reasonable compromise between accuracy and calculation time.

Flexible multibody models are more accurate approximations of the system, especially at high frequencies, but result in a raise of computational task. The train model for the fault detection and force estimation procedure is considered as a ridged body model and thus high frequencies are neglected. The force elements connecting the bodies are springs and dampers with different linear and nonlinear characteristics, and the nonlinear wheel rail contact force element acting on the wheelsets.

This section covers the elements of the train and how the well known methods of multibody dynamics are used to describe the general structure of the train model.

3.2.1 Train Model

To test the fault detection and isolation method, the Velaro Rus , shown in Figure 3.13, is used as an example.

The Siemens Velaro is a family of high speed trains and operates for example in Germany, Spain and Russia. The Velaro RUS is built for Russia and entered passenger service at the end of 2009. The train serves on the line between Saint Petersburg and Moscow.

To simulate the Velaro Rus, a mechanical model must be defined which characterizes the train with all its relevant components. The model, shown in Figures 3.14 and 3.15, consists of a car body, two bogies, four wheelsets and two motors. The bogie is the running gear of the Velaro which carries the car body. It consists among other things of the bogie frame, wheels, springs and brakes. There are basically two different types of bogies, on the one hand, there are powered bogies, and on the other hand trailer bogies. At the power bogies of the Velaro trains, engines

Figure 3.13: Velaro RUS

are mounted, which drive the train. The trailer bogies have no drive. To ensure safety and comfort, the train is equipped with different types of suspensions and dampers. The design of the bogie is essential for the stable running of the train on the track. The sine wave, which is resulting during running, increases its frequency with increasing vehicle speed, and eventually leads to instability. This movement is damped by the anti-yaw damper between bogie and car body and, even at high speeds, the stability of the vehicle is guaranteed.

The wheelsets are connected to the bogies by the so-called primary suspension. The primary suspension compensates the unevenness of the track and increases the ride comfort together with the secondary springs. The secondary suspension connects the bogies and the car body. Another suspension is to be found between the motor and the bogie. As stated above, the anti-yaw dampers are a major part of the secondary suspension system, they are used on vehicles with a maximum speed over 160 km/h and thus also in high-speed trains such as the Velaro Rus. The shock absorbers are mounted horizontally between the car body and bogie to attenuate the sinusoidal running.

The blue circles in Figure 3.15 indicate the positions, where the suspension elements are fixed to the bodies. The red lines between the blue circles denote suspension elements.

Car body, bogie and motor motion are characterized by six degrees of freedom each. By assuming constant running speed and constant rotational motion of the

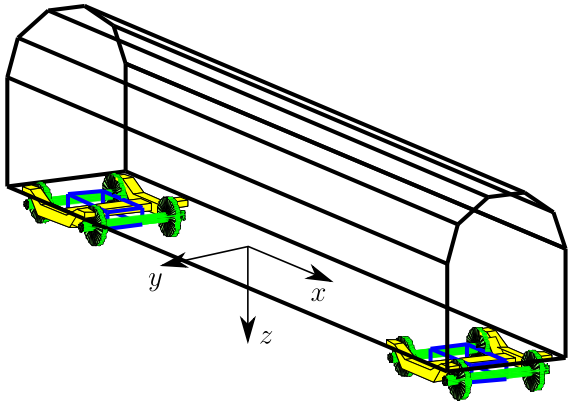


Figure 3.14: Matlab train model

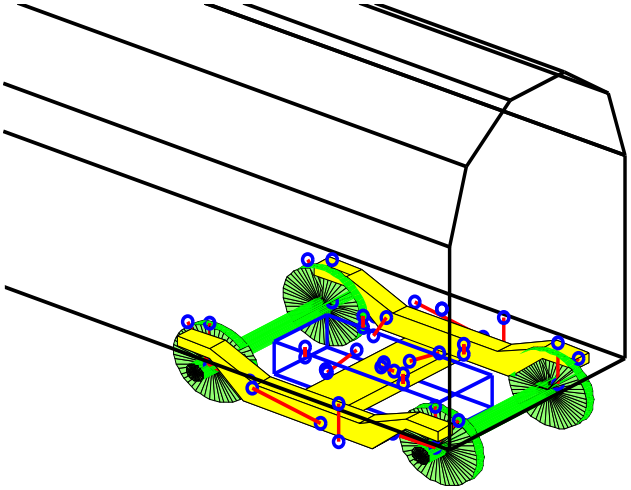


Figure 3.15: Matlab train model enlarged

wheelset around the y-axis, the wheelset motion is considered with four degrees of freedom. A wheelset consists of an axis and two wheels rigidly coupled. Since the selected wagon is a motor coach, the wheelsets carry not only brakes but also a gear box.

The aim of this work is, inter alia, to detect faults in the anti-yaw damper, the secondary vertical damper and the secondary lateral damper. The first is important for ride quality, while the second and third influence the running stability. Figures 3.16 and 3.17 show the suspension structure of the bogies. The positions of the three dampers, which are used in the fault detection, are marked.

Some of the components of the suspension system of the train have nonlinear characteristics, especially those chosen for the fault detection and isolation, the anti-yaw damper, the secondary vertical damper and the secondary lateral damper.

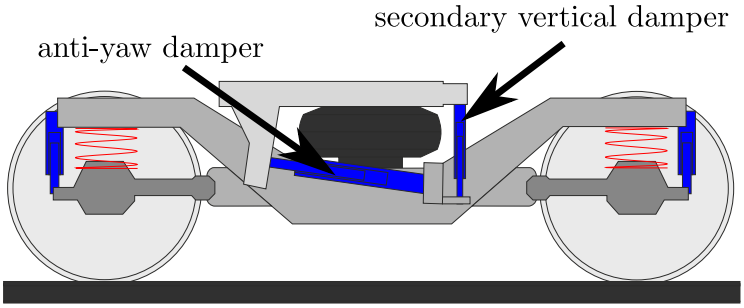


Figure 3.16: Bogie: Side view

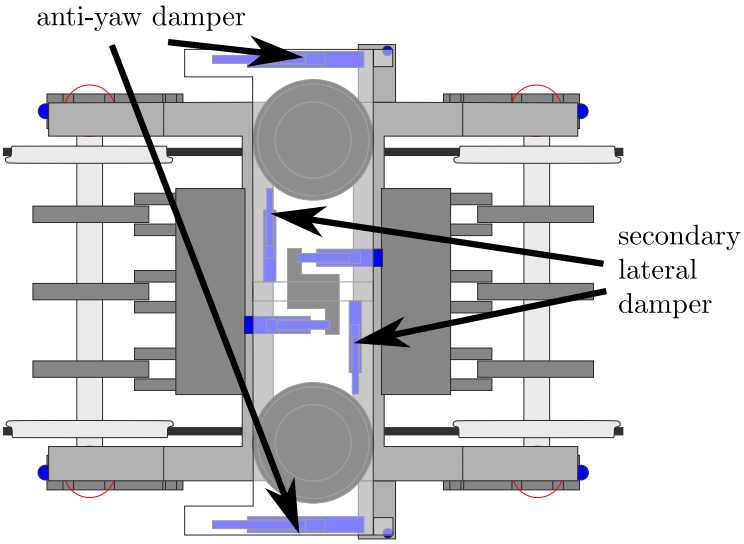


Figure 3.17: Bogie: Top view

The nonlinear characteristics of the anti-yaw damper are shown in Figure 3.18. The

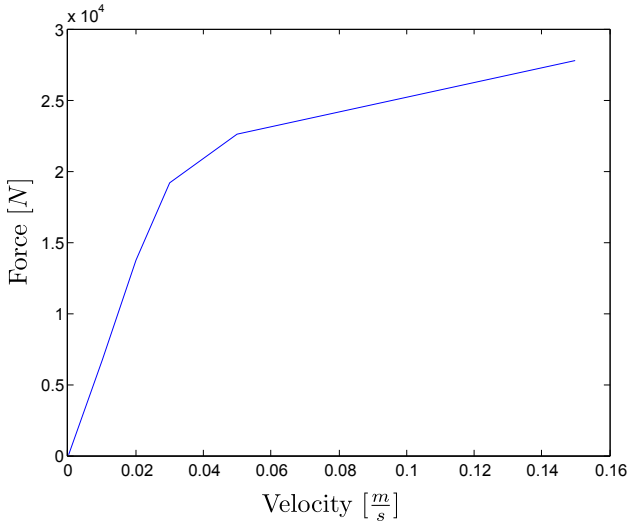


Figure 3.18: Anti-yaw damper

nonlinear characteristics of the anti-yaw damper give a good damper response at low

velocities while avoiding damages at high velocities. The nonlinear characteristics of the secondary vertical and secondary lateral damper are given in Figures 3.19 and 3.20.

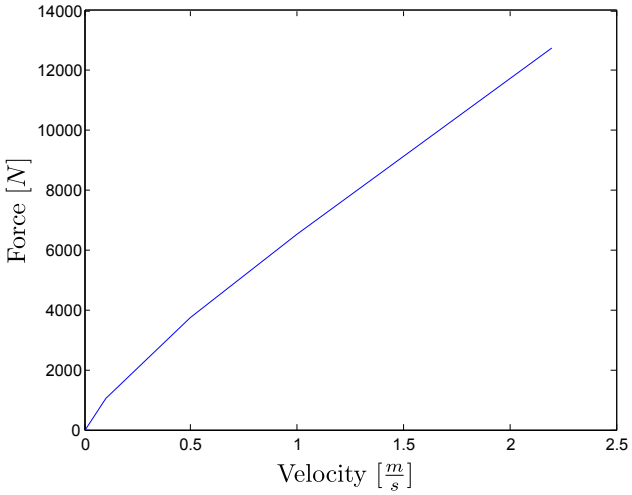


Figure 3.19: Vertical damper

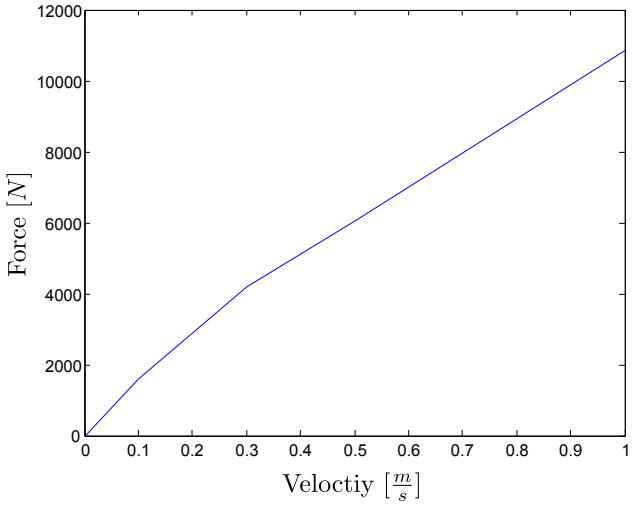


Figure 3.20: Lateral damper

Compared to the anti-yaw damper, the nonlinearity of the secondary vertical and secondary lateral damper is small.

3.2.2 Multibody Dynamics

After defining the components of the mechanical model of the railway vehicle, the equations of motion are formulated. The bodies of the train are considered as ridged bodies, and thus the displacement of each body is composed of translations and rotations. The bodies are connected by force elements, each force element is connected with two bodies and fixed at a specific point at the bodies, respectively.

To uniquely determine the position of the different bodies of the train, the coordinate system shown in Figure 3.21 is used. In case of the train model, ϕ describes the roll motion, θ the pitch motion and ψ the yaw motion. The train dynamics are described in a coordinate system, which is moving along the railway track. The displacement of a body can then be described using the six so-called trajectory coordinates

1. the arc length coordinate s,
2. the lateral displacement y_t relative to the trajectory,
3. the vertical displacement z_t relative to the trajectory,
4. the roll-angle ϕ about x_t,
5. the pitch-angle θ about y_t and
6. the yaw-angle Ψ about z_t.

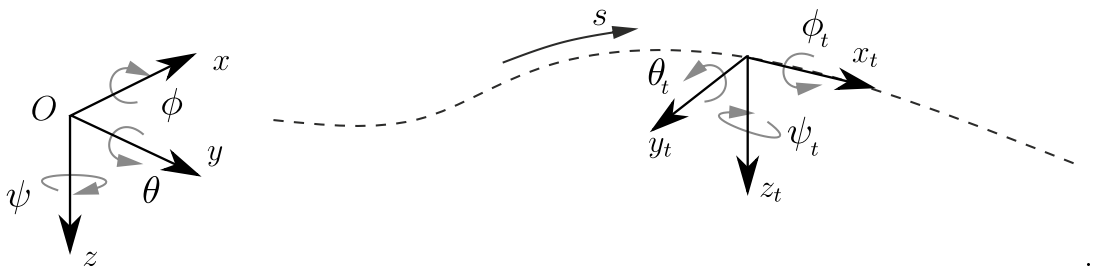


Figure 3.21: Coordinate system

The Newton-Euler equations are used to derive the equations of motion. The origin of the body coordinate system is attached to the center of mass of the body. The Newton-Euler equations of motion for a rigid single body are given in matrix form as follows [35]

$$\begin{bmatrix} m\mathbf{I} & 0 \\ 0 & \mathbf{J} \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{R}} \\ \dot{\bar{\boldsymbol{\omega}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F} \\ \bar{\mathbf{M}} - \bar{\boldsymbol{\omega}} \times (\mathbf{J}\bar{\boldsymbol{\omega}}) \end{bmatrix}, \tag{3.65}$$

where m is the mass of the rigid body, $\mathbf{I}$ is 3×3 identity matrix, $\mathbf{J}$ is the inertia tensor, $\mathbf{F}$ is the resultant of the external forces and $\bar{\mathbf{M}}$ is the resultant of the external moments defined in the body coordinate system. The vector $\bar{\boldsymbol{\omega}}$ results from

$$\bar{\boldsymbol{\omega}} = \bar{\mathbf{G}}\dot{\boldsymbol{\alpha}} \tag{3.66}$$

with

$$\bar{\boldsymbol{G}} = \begin{bmatrix} -\cos\phi\sin\theta & \cos\theta & 0 \\ \sin\phi & 0 & 1 \\ \cos\phi\cos\theta & \sin\theta & 0 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \psi \\ \phi \\ \theta \end{bmatrix}. \tag{3.67}$$

The moment $\bar{\mathbf{M}}$ is calculated by

$$\mathbf{M} = \mathbf{A}_{rot}\bar{\mathbf{M}} \tag{3.68}$$

and

$$\mathbf{M} = \mathbf{u}_i \times \boldsymbol{F} = (\mathbf{A}_{rot}\bar{\mathbf{u}}) \times \boldsymbol{F}. \tag{3.69}$$

The matrix $\mathbf{A}_{rot}$ is a 3×3 rotation matrix, which defines the orientation of the axes of the body coordinate system with respect to the global coordinate system. The Euler angles are used to describe the rotational matrix. In this method three independent parameters define the orientation of the body in space. The rotational matrix $\mathbf{A}_{rot}$ is the product of three simple rotations $\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3$. The first rotation is about the z-axis by the angle ψ

$$\mathbf{A}_1 = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.70}$$

the second rotation is about the x-axis by the angle ϕ

$$\mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}, \tag{3.71}$$

and the third rotation is about the y-axis by the angle θ

$$\mathbf{A}_3 = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}. \tag{3.72}$$

Using these three rotations, the orientation of a body is described by the product of all three matrices

$$\mathbf{A}_{rot} = \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3, \tag{3.73}$$

which after multiplication gives

$$\mathbf{A}_{rot} = \begin{bmatrix} \cos\psi\cos\theta - \sin\psi\sin\phi\sin\theta & -\sin\psi\cos\phi & \cos\psi\sin\theta + \sin\psi\sin\phi\cos\theta \\ \sin\psi\cos\theta + \cos\psi\sin\phi\sin\theta & \cos\psi\cos\phi & \sin\psi\sin\theta - \cos\psi\sin\phi\cos\theta \\ -\cos\phi\sin\theta & \sin\phi & \cos\phi\cos\theta \end{bmatrix}. \tag{3.74}$$

The global position vector of an arbitrary point on the rigid body is written as

$$\boldsymbol{r} = \boldsymbol{R} + \mathbf{A}_{rot}\bar{\mathbf{u}}, \tag{3.75}$$

where $\boldsymbol{R}$ is the global position vector of the origin of the body coordinate system defined as

$$\boldsymbol{R} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}. \tag{3.76}$$

The vector $\bar{\mathbf{u}}$ is the position vector of the arbitrary point on the body with respect to the origin of the body coordinate system and is defined as

$$\bar{\mathbf{u}} = \begin{bmatrix} \bar{u}_x \\ \bar{u}_y \\ \bar{u}_z \end{bmatrix}. \tag{3.77}$$

Equations (3.65) to (3.77) are to be evaluated for each of the nine bodies. Different forces and moments act on individual components: They are due to springs and dampers, linking the components, or due to the irregular track, which causes an external excitation.

If the force elements are springs and dampers, the force is calculated from the relative motion between the specific points. To calculate the relative motion, the global position vector of the specific points is needed.

Two bodies are connected with a suspension at the points denoted by $\mathbf{r}_j$ and $\mathbf{r}_i$. Defining

$$\boldsymbol{\ell} = \mathbf{r}_j - \mathbf{r}_i, \tag{3.78}$$

the spring force f_c and the damper force f_d is expressed as functions of $\boldsymbol{\ell}$ and its derivative:

$$f_c = f_c(||\boldsymbol{\ell}||_2 - \ell_0), \quad f_d = f_d(\frac{\boldsymbol{\ell}^T\dot{\boldsymbol{\ell}}}{||\boldsymbol{\ell}||_2}), \tag{3.79}$$

where l_0 is the unstressed spring length. The force vectors acting on the points $\mathbf{r}_j$ and $\mathbf{r}_i$ are

$$\mathbf{F}_c = \pm f_c \frac{\boldsymbol{\ell}}{||\boldsymbol{\ell}||_2}, \quad \mathbf{F}_d = \pm f_d \frac{\boldsymbol{\ell}}{||\boldsymbol{\ell}||_2}. \tag{3.80}$$

With the above equations, the train model is simulated. Figure 3.22 shows an example of the lateral track position and the lateral movement of the first wheelset for the nonlinear train model.

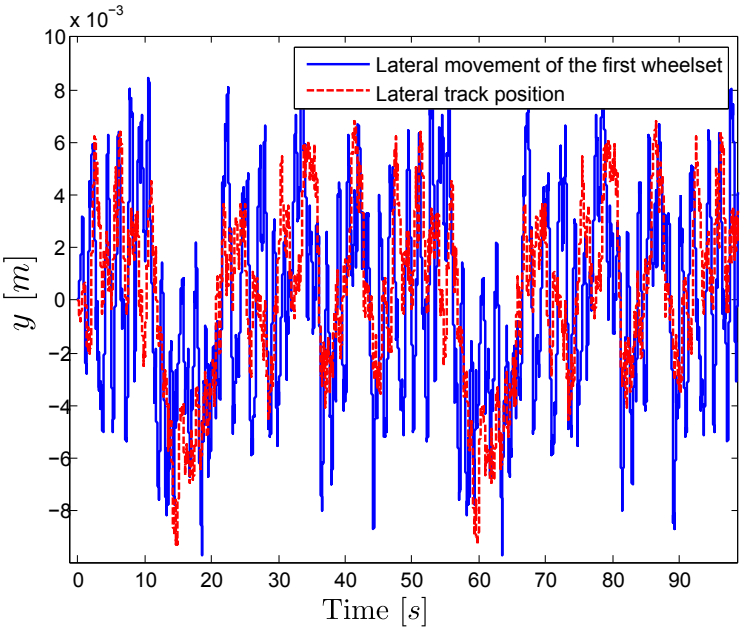


Figure 3.22: Lateral track and wheelset position

3.2.3 Simulation of acceleration sensors

For the reconstruction of the system dynamics, the vehicle model is prepared with acceleration sensors. The measurements of these acceleration data are used to perform a state estimation with the Kalman filter. The railway vehicle model is prepared with 14 acceleration sensors, each wheelset has two sensors measuring the acceleration in all three directions. Each bogie has two sensors measuring the acceleration of all three directions. The carbody has two sensors, one at the front one at the rear measuring the acceleration in the y-direction. It is assumed, that the measurement noise of all sensors is white with a given standard deviation. The acceleration for an arbitrary point at a rigid body is given by

$$\ddot{\boldsymbol{r}} = \ddot{\boldsymbol{R}} + \dot{\boldsymbol{\omega}} \times \mathbf{u} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{u}),$$

for the linear case this results to

$$\ddot{\boldsymbol{r}} = \ddot{\boldsymbol{R}} + \dot{\boldsymbol{\omega}} \times \mathbf{u},$$

or

$$\begin{aligned}
\ddot{r}_x &= \ddot{x} + \ddot{\theta} u_z - \ddot{\psi} u_y, \\
\ddot{r}_y &= \ddot{y} + \ddot{\psi} u_x - \ddot{\phi} u_z, \\
\ddot{r}_z &= \ddot{z} + \ddot{\phi} u_y - \ddot{\theta} u_x.
\end{aligned}$$

Figure 3.24 shows the vertical and Figure 3.23 the lateral acceleration data of the simulated sensor placed at the front wheelset at the left side.

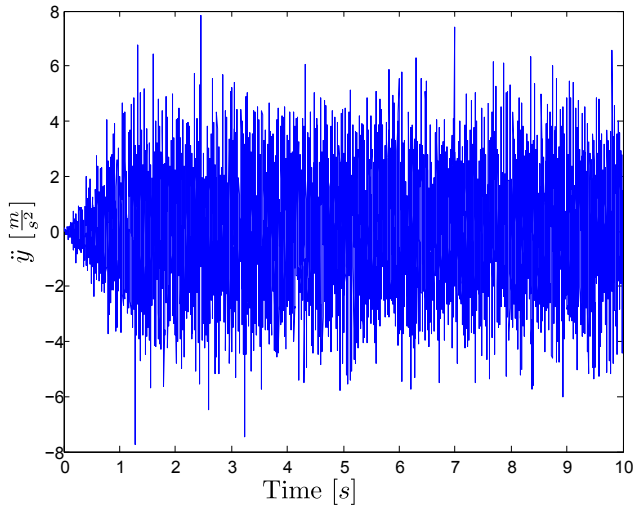


Figure 3.23: Lateral measurements wheelset

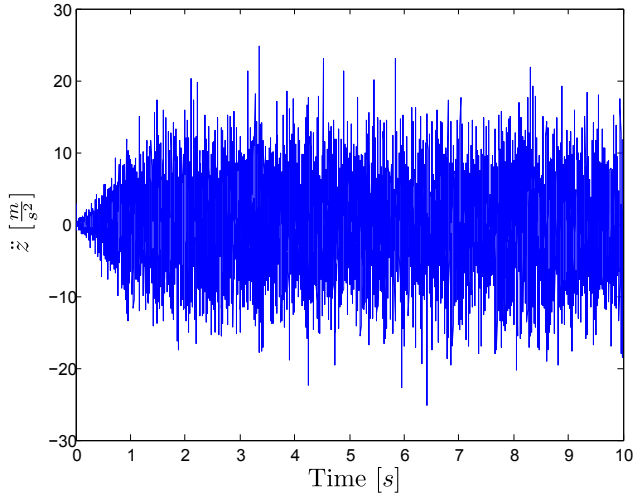


Figure 3.24: Vertical measurements wheelset

3.2.4 Linearization and state-space formulation

The train model is simulated with the nonlinear wheel rail contact model and nonlinear suspension systems. For the fault detection procedure, the system must be linearized about the operating point because a linear Kalman filter is used to perform a state estimation. The linear Newton-Euler equations of motion for the train model are given by

$$\begin{aligned} \dot{\mathbf{x}}_s &= \mathbf{A}_s\mathbf{x}_s + \mathbf{B}_s\mathbf{w}_s \\ \mathbf{y}_s &= \mathbf{C}_s\mathbf{x}_s + \mathbf{D}_s\mathbf{w}_s, \end{aligned} \tag{3.81}$$

where $\mathbf{A}_s$, $\mathbf{B}_s$,$\mathbf{C}_s$ and $\mathbf{D}_s$ indicate the system, input, output and feed-through matrices of the linearized train model. The state vector $\mathbf{x}_s$ includes all states of the train model. Since the train consists of five bodies with six degrees of freedom, and four bodies with four degrees of freedom, there are 46 degrees of freedom and 92 components in the state vector. The vector $\mathbf{y}_s$ is the measurement vector, which includes the outputs of the acceleration sensors. The two sensors placed at the coach, measure the y-direction, the other 12 sensors measure all three directions. Thus the output vector $\mathbf{y}_s$ consists of 38 values.

Since the equations of motion are nonlinear, the equations are linearized to obtain the form of equation (3.81). The train model consists of complex nonlinear components, which are sometimes only defined piecewise. For example the wheel rail profile is given piecewise and thus the contact force as well. Therefore, it is difficult to linearize the system analytically and a numerical linearisation is performed. Finite differences are used for the linearisation about the operation point x_0. Given a function u, a second order approximation

$$u' \approx \frac{u(x_0 + h) - u(x_0 - h)}{2h} \tag{3.82}$$

is used for the first derivative.

3.2.5 Train model and form filter

In order to increase the ability to correctly detect and isolate faults in the system, the train model is combined with the model of the track irregularities. The system containing the train model and the form filter for the track irregularities is written for the linear case as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{w}_{FF}, \\ \mathbf{y} &= \mathbf{C}\mathbf{x}, \end{aligned} \tag{3.83}$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s\mathbf{C}_{FF} \\ \mathbf{0} & \mathbf{A}_{FF} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_{FF} \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} \mathbf{C}_s & \mathbf{D}_s\mathbf{C}_{FF} \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_{FF} \end{bmatrix}. \quad (3.84)$$

In this way it is possible to capture both, the railway vehicle and the track dynamic in one system and the estimation process is improved significantly. If the track irregularities are described with other power spectra, the form filters must be adjusted accordingly. Friedrich for example, defined power spectral density functions in [44] which can be used alternatively.

The new matrix $\mathbf{A}$ has the dimension $(m_s + 4 \cdot 7) \times (m_s + 4 \cdot 7)$, matrix $\mathbf{B}$ $(m_s + 4 \cdot 7) \times (4 \cdot 3)$ and matrix $\mathbf{C}$ $n_s \times (m_s + 4 \cdot 7)$, where m_s is the number of states of the train model and n_s is the number of measurement outputs of the train model, in this work $m_s = 92$ and $n_s = 38$. Figure 3.25 shows the block diagram of the combined system.

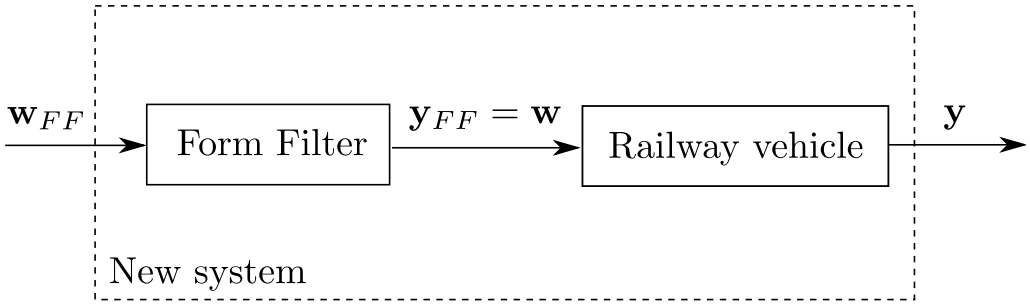


Figure 3.25: Form filter combined with railway vehicle

The input of the system $\mathbf{w}_{FF}$ is white noise, by passing through the form filter, the track specific excitation function is generated. This signal then acts as an input to the rail vehicle. The output of specific points of the railway vehicle $\mathbf{y}$ is the measurement signal, i.e. the acceleration.

4 Fault detection and force estimation

In this section, fault detection and force estimation, are presented. For the fault detection as well as the force estimation, the Kalman filter plays a major role. Thus, at the beginning the Kalman filter is introduced. Afterwards, it is explained how the Kalman filter is expanded to observe errors in the estimation process and how several expanded Kalman filters are used to detect and isolate faults in the suspension system. The last section describes the estimation of wheel rail forces as well as suspension forces.

4.1 Kalman filter

In order to detect and isolate faults in the train model as well as to estimate wheel rail forces or suspension forces, it is necessary to capture the dynamic states of the train model. The only available information to perform this estimation comes from acceleration sensors fixed at the train. The Kalman filter is used to estimate all states of the train from these acceleration signals.

The Kalman filter, also known as linear quadratic estimator, solves the problem of estimating the states of a linear dynamic system perturbed by white noise. The state estimation uses measurements, which are linearly related to the states and corrupted by white noise. The purpose of the Kalman filter is to minimize the spread of the estimate-error probability density [33].

The Kalman filter is applied for the control of complex dynamic systems such as continuous manufacturing processes, aircraft, ships, or spacecraft. A detailed description of the Kalman filter can be found in [33, 48, 49].

4.1.1 Propagation of states and covariance

The Kalman filter propagates the expected value and the covariance of the state through time. It takes into account system dynamics and inputs as well as it incorporates measurements and measurement error statistics.

In this chapter, the equations to calculate the propagation of the expected value of the state and covariance are given based on a discrete mathematical description of a dynamic system. These equations are fundamental for the state estimation procedure with the Kalman filter.

Having the following linear discrete-time system:

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \tag{4.1}$$

where $\mathbf{u}_k$ is a known input and $\mathbf{w}_k$ is Gaussian zero-mean white noise with covariance $\mathbf{Q}_k$, the expected value of $\mathbf{x}_k$ results to

$$\begin{aligned} \hat{\mathbf{x}}_k &= E(\mathbf{x}_k) \\ &= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1}. \end{aligned} \tag{4.2}$$

By using equations (4.1) and (4.2) the covariance of $\mathbf{x}_k$ changes with time in the following way

$$\begin{aligned} (\mathbf{x}_k - \hat{\mathbf{x}}_k)(...)^T &= (\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} - \hat{\mathbf{x}}_k)(...)^T \\ &= [\mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \mathbf{w}_{k-1}][...]^T \\ &= \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^T\mathbf{F}_{k-1}^T + \mathbf{w}_{k-1}\mathbf{w}_{k-1}^T + \\ &\quad \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})\mathbf{w}_{k-1}^T + \mathbf{w}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})\mathbf{F}_{k-1}^T. \end{aligned} \tag{4.3}$$

Therefore, the covariance of $\mathbf{x}_k$ is obtained as the expected value of the above expression. Since $(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})$ is uncorrelated with $\mathbf{w}_{k-1}$, it results

$$\begin{aligned} \mathbf{P}_k &= E[(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})(...)^T] \\ &= \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \end{aligned} \tag{4.4}$$

where $\mathbf{Q}$ is the disturbance input covariance matrix. This equation is called a discrete time Lyapunov equation, or a Stein equation [50, 33].

Equation (4.1) shows, that the process noise is directly entering the system dynamics. Often, the process noise is multiplied by a matrix, before it enters the system dynamics. That is,

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\tilde{\mathbf{w}}_{k-1}, \quad \tilde{\mathbf{w}}_k \sim (0, \tilde{\mathbf{Q}}_k). \tag{4.5}$$

This can be put into the conventional form of equation (4.1) by two steps. First,

the rightmost term of equation (4.5) has a covariance given by

$$\begin{aligned} E[(\mathbf{L}_{k-1}\tilde{\mathbf{w}}_{k-1})(\tilde{\mathbf{w}}_{k-1}\mathbf{L}_{k-1})^T] &= \mathbf{L}_{k-1}E[\tilde{\mathbf{w}}_{k-1}\tilde{\mathbf{w}}_{k-1}^T]\mathbf{L}_{k-1}^T \\ &= \mathbf{L}_{k-1}\tilde{\mathbf{Q}}_{k-1}\mathbf{L}_{k-1}^T, \end{aligned} \tag{4.6}$$

therefore, equation (4.5) is equivalent to the equation

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{w}_k \sim (0, \mathbf{L}_k\tilde{\mathbf{Q}}_k\mathbf{L}_k^T). \tag{4.7}$$

The same type of transformation is made with noisy measurement equations. That is, the measurement equation

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{L}_k\tilde{\mathbf{v}}_k, \quad \tilde{\mathbf{v}}_k \sim (0, \tilde{\mathbf{R}}_k) \tag{4.8}$$

is equivalent to the measurement equation

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim (0, \mathbf{L}_k\tilde{\mathbf{R}}_k\mathbf{L}_k^T) \tag{4.9}$$

where $\mathbf{L}_k\tilde{\mathbf{R}}_k\mathbf{L}_k^T = \mathbf{R}$ is the measurement error covariance matrix.

4.1.2 Discrete time Kalman filter

The Kalman filter operates by propagating the expected value and covariance of the state through time. The goal is to minimize the steady-state error covariance by constructing a state estimate $\hat{\mathbf{x}}$. Figure 4.1 shows the block diagram of a system with an integrated Kalman filter.

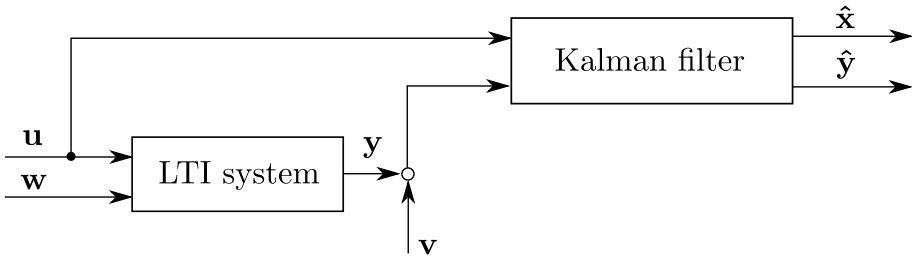


Figure 4.1: Kalman filter

Since, as for the estimation of the state $\mathbf{x}$ the expected state is used, the expression expected state and estimated state are used equivalently.

The discrete time state space model for this system is given by

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \tag{4.10}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \tag{4.11}$$

with the state vector $\mathbf{x}$, the known input $\mathbf{u}$ and the output $\mathbf{y}$. The noise processes $\mathbf{w}_k$ and $\mathbf{v}_k$ are white, zero-mean, uncorrelated, and have known covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively:

$$\begin{aligned} \mathbf{w}_k &=\sim (0, \mathbf{Q}_k) \\ \mathbf{v}_k &=\sim (0, \mathbf{R}_k) \\ E[\mathbf{w}_k \mathbf{w}_k^T] &= \mathbf{Q}_k \delta_{k-j} \\ E[\mathbf{v}_k \mathbf{v}_k^T] &= \mathbf{R}_k \delta_{k-j} \\ E[\mathbf{v}_k \mathbf{w}_k^T] &= 0 \end{aligned} \tag{4.12}$$

The aim is to estimate the state $\mathbf{x}_k$ based on the knowledge of the system dynamics and the availability of the noisy measurements $\mathbf{y}_k$ and the input $\mathbf{u}_k$. Depending on the availability of the measurement $\mathbf{y}_k$, a distinction in the notation for an estimated state $\hat{\mathbf{x}}$ is made as follows: If all the measurements up to and including time k are used to estimate $\mathbf{x}_k$, an a posteriori estimate is performed, which is denoted as $\hat{\mathbf{x}}_k^+$. The "+" superscript denotes that the estimate is a posteriori

$$\hat{\mathbf{x}}_k^+ = E[\mathbf{x}_k | \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k] = \text{ a posteriori estimation.} \tag{4.13}$$

If all the measurements before (but not including) time k are used to estimate $\mathbf{x}_k$, an a priori estimate is performed, which is denoted as $\hat{\mathbf{x}}_k^-$. The "-" superscript denotes that the estimate is a priori

$$\hat{\mathbf{x}}_k^- = E[\mathbf{x}_k | \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{k-1}] = \text{ a priori estimation.} \tag{4.14}$$

Since the a posteriori estimation uses more information, it is considered that the a posteriori estimation is better than the a priori estimation.

$$\hat{\mathbf{x}}_k^- = \text{estimate of } \mathbf{x}_k \text{ without the measurement at time } k, \tag{4.15}$$

$$\hat{\mathbf{x}}_k^+ = \text{estimate of } \mathbf{x}_k \text{ including the measurement at time } k. \tag{4.16}$$

To measure the statistic relationship between real and estimated states, the covariance of the estimation error $\mathbf{x}_k - \hat{\mathbf{x}}_k$ is calculated and denoted by

$$\mathbf{P}_k^- = E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T], \tag{4.17}$$

$$\mathbf{P}_k^+ = E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^+)(\mathbf{x}_k - \hat{\mathbf{x}}_k^+)^T]. \tag{4.18}$$

To start the estimation process, an initialization has to be performed by declaring an initial estimate $\hat{\mathbf{x}}_0$ of the initial state $\mathbf{x}_0 = \mathbf{x}(t = t_0 = 0)$ and an initial estimation

of the covariance of the estimation error $\mathbf{P}_0$,

$$\hat{\mathbf{x}}_0^+ = E(\mathbf{x}_0), \tag{4.19}$$

$$\mathbf{P}_0^+ = E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0^-)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^-)^T] \tag{4.20}$$

$$= E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)^T]. \tag{4.21}$$

The matrix $\mathbf{P}_0$ represents the uncertainty in the initial estimate of $\mathbf{x}_0$. In case of a perfectly known initial state $\mathbf{P}_0 = 0$, if there is no information of the initial state $\mathbf{P}_0 = \infty$.

The Kalman filter algorithm is divided in two steps, the first step is the time update and the second step is the measurement update.

The estimation process starts with $\hat{\mathbf{x}}_0^+$, the estimate of the initial state $\mathbf{x}_0$. By using equation (4.2) the expected value of $\mathbf{x}$ propagates with

$$\hat{\mathbf{x}}_1^- = \mathbf{F}_0\hat{\mathbf{x}}_0^+ + \mathbf{G}_0\mathbf{u}_0. \tag{4.22}$$

This equation can be extended to obtain a more general equation, which is the time update equation for $\hat{\mathbf{x}}$

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1}^+ + \mathbf{G}_{k-i}\mathbf{u}_{k-1}. \tag{4.23}$$

After computing the time update of $\hat{\mathbf{x}}$, the time update of $\mathbf{P}$ is computed. Following equation (4.4) the time update for the initial guess of $\mathbf{P}_0$ as well as the general case, the time update from $\mathbf{P}_k^-$ to $\mathbf{P}_k^+$ can be computed from

$$\mathbf{P}_1^- = \mathbf{F}_0\mathbf{P}_0^+\mathbf{F}_0^T + \mathbf{Q}_0 \tag{4.24}$$

or in general

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}. \tag{4.25}$$

After computing the time update of the state $\mathbf{x}$ and the error covariance $\mathbf{P}$ the estimation is improved with the measurement update. The measurement update uses the information of the actual measurement.

The measurement $\mathbf{y}_k$ changes the estimate of a constant $\mathbf{x}$ according to [33] as

follows:

$$\begin{aligned}
\mathbf{K}_k =& \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\
=& \mathbf{P}_k^+ \mathbf{H}_k^T \mathbf{R}_k^{-1} \\
\hat{\mathbf{x}}_k^+ =& \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \\
\mathbf{P}_k^+ =& (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T \\
& ((P_k^-)^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \\
& (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-,
\end{aligned} \tag{4.26}$$

where $\mathbf{K}_k$ is called the Kalman filter gain. Again the vector $\hat{\mathbf{x}}_k^-$ and the matrix $\mathbf{P}_k^-$ define the estimated state and its error covariance matrix before the measurement $\mathbf{y}_k$ is available, and $\hat{\mathbf{x}}_k^+$ and $\mathbf{P}_k^+$ are the estimated state and its error covariance matrix after the measurement $\mathbf{y}_k$ is available.

Combining all equations for the estimation process the discrete-time Kalman filter can be defined. The state estimation performed with the Kalman filter can be done with the following steps [33]:

1. The dynamic system is given by the equations:

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{G}_{k-1} \mathbf{u}_{k-1} + \mathbf{w}_{k-1} \\
\mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \\
E[\mathbf{w}_k \mathbf{w}_k^T] &= \mathbf{Q}_k \delta_{k-j} \\
E[\mathbf{v}_k \mathbf{v}_k^T] &= \mathbf{R}_k \delta_{k-j} \\
E[\mathbf{v}_k \mathbf{w}_k^T] &= 0
\end{aligned} \tag{4.27}$$

2. The Kalman filter is initialized as:

$$\begin{aligned}
\hat{\mathbf{x}}_0^+ &= E(\mathbf{x}_0) \\
\mathbf{P}_0^+ &= E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)^T]
\end{aligned} \tag{4.28}$$

3. The Kalman filter is given by the equations, which are computed for each time step $k = 1, 2, ..$:

$$\begin{aligned}
\hat{\mathbf{x}}_k^- &= \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1}^+ + \mathbf{G}_{k-1} \mathbf{u}_{k-1} \\
\mathbf{P}_k^- &= \mathbf{F}_{k-1} \mathbf{P}_{k-1}^+ \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} \\
\mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\
\hat{\mathbf{x}}_k^+ &= \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \\
\mathbf{P}_k^+ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-
\end{aligned} \tag{4.29}$$

As mentioned above the Kalman filter minimizes the squared estimation error. The Kalman gain matrix is influenced by the noise characteristics of the process

and measurement noise. Both characteristics are not exactly known in reality and thus the assumed matrices $\mathbf{Q}$ and $\mathbf{R}$ can be used as tuning parameters to increase robustness and/or accuracy.

4.1.3 Steady state Kalman filter

If the state space model is linear and time invariant, a steady state Kalman filter can be used to perform the estimation. The advantage of the steady state Kalman filter is, that the Kalman gain matrix $\mathbf{K}$ must only be computed once and can be used for the whole estimation process. This decreases the computational effort significantly. The drawback is, that it is not possible to change the system matrices during the operation process and that the initial condition is not treated in the right way. The steady state solution for the continuous time Kalman filter can be derived from the differential Riccati equation which is given by

$$\dot{\mathbf{P}} = -\mathbf{P}\mathbf{C}\mathbf{R}^{-1}\mathbf{C}\mathbf{P} + \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{Q}. \tag{4.30}$$

In the case of a LTI system $\mathbf{A}$, $\mathbf{C}$, $\mathbf{Q}$ and $\mathbf{R}$ are constant and $\mathbf{P}$ can reach a steady state solution, when $\dot{\mathbf{P}}$ reaches zero. This implies that

$$-\mathbf{P}\mathbf{C}\mathbf{R}^{-1}\mathbf{C}\mathbf{P} + \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{Q} = 0, \tag{4.31}$$

which is the so-called algebraic Riccati equation (ARE). Assuming $\mathbf{Q} \geq 0$ and $\mathbf{R} > 0$ and defining $\mathbf{G}$ as any matrix such that $\mathbf{G}\mathbf{G}^T = \mathbf{Q}$, the Kalman filter gain is computed as

$$\mathbf{K} = \mathbf{P}\mathbf{C}^T\mathbf{R}^{-1}. \tag{4.32}$$

The state estimation performed with the steady state Kalman filter is then given by

$$\dot{\hat{\mathbf{x}}} = (\mathbf{A} - \mathbf{K}\mathbf{C})\hat{\mathbf{x}} + \mathbf{K}\mathbf{y}. \tag{4.33}$$

To solve the algebraic Riccati equation and to guarantee that the solution is stable, the problem is formulated using the so-called linear matrix inequalities. Efficient solvers exist for linear matrix inequalities, which treat the problem in a reliable and stable way, supposing a stable solution exists.

4.1.4 State estimation of the railway vehicle

The Kalman filter is used to obtain information of the states of the railway vehicle. Based on the available acceleration sensors, the state is estimated. Since the train

model has no known input $\mathbf{u}$, the state prediction step of the Kalman filter reduces to

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1}^+. \tag{4.34}$$

Further, the train model is considered as time invariant. Thus, state, input and output matrix are constant.

Figures 4.2 and 4.3 exemplarily show the results for the estimation of the lateral and vertical position of the first wheelset. The lateral state is estimated reliably. The estimation of the vertical state is performed well after a short integration phase.

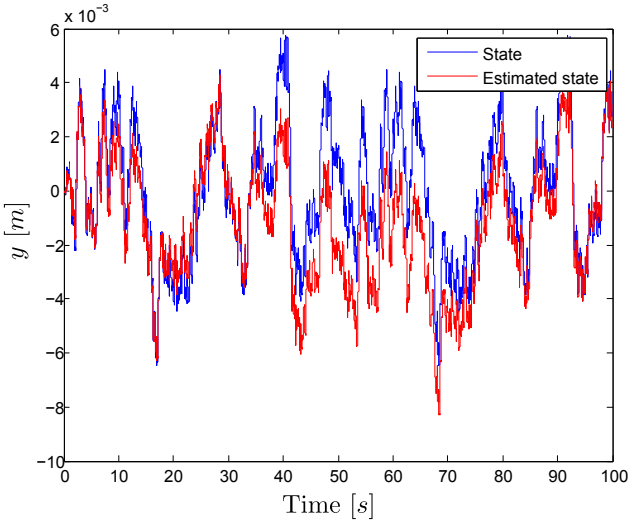


Figure 4.2: Estimation of lateral position of wheelset

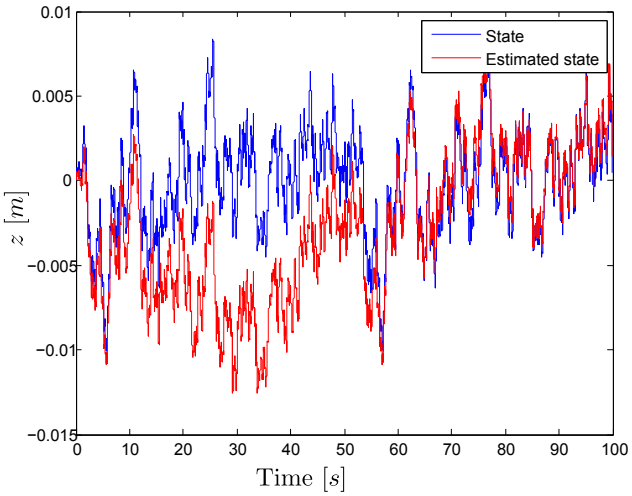


Figure 4.3: Estimation of vertical position of wheelset

Figures 4.4 and 4.5 show the results for the estimation of the lateral and vertical velocity of the first wheelset. Both states are estimated correctly and are even better

estimated than the position. This results from the integration error from velocity to position.

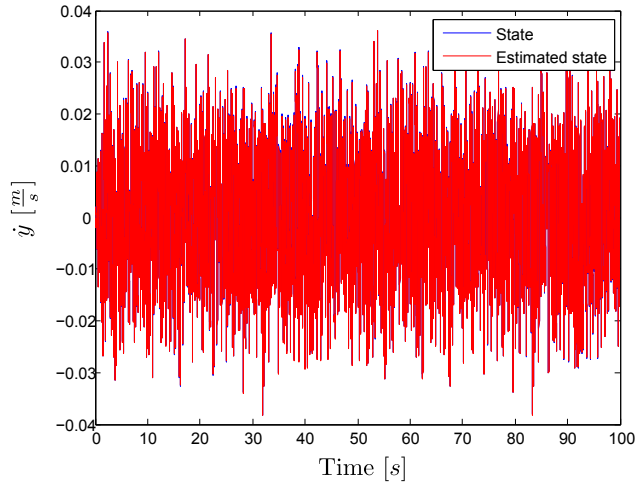


Figure 4.4: Estimation of lateral velocity of wheelset

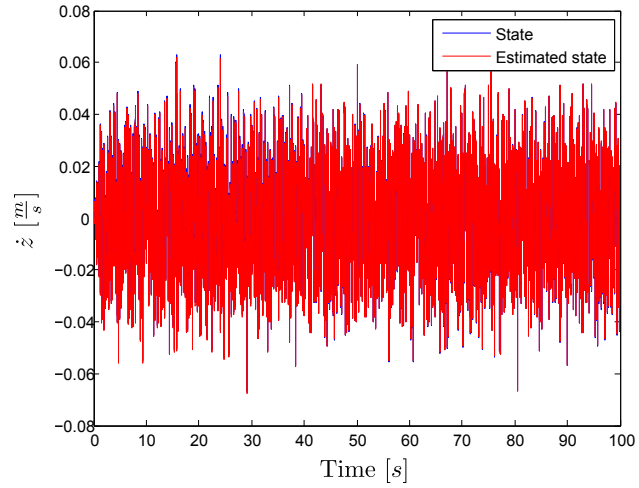


Figure 4.5: Estimation of vertical velocity of wheelset

Figures 4.6 and 4.7 show the results for the estimation of the angular yaw position and the angular yaw velocity of the first bogie. Both figures show qualitatively the same results as the lateral and vertical wheelset motion. The angular yaw position is estimated correctly with a small error. The angular yaw velocity is estimated nearly without an error.

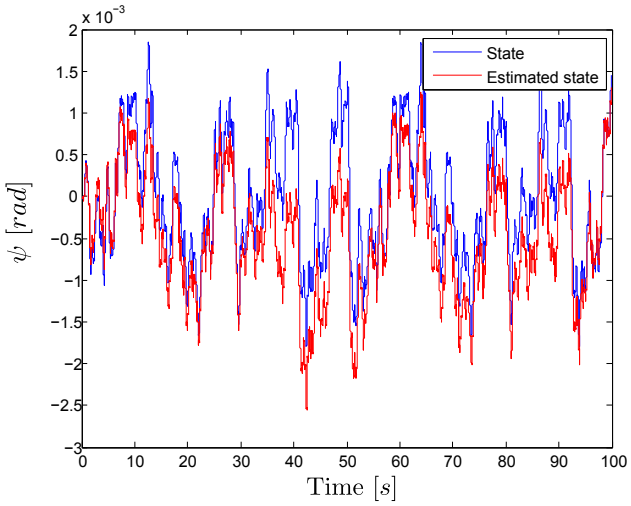


Figure 4.6: Estimation the angular yaw position of the first bogie

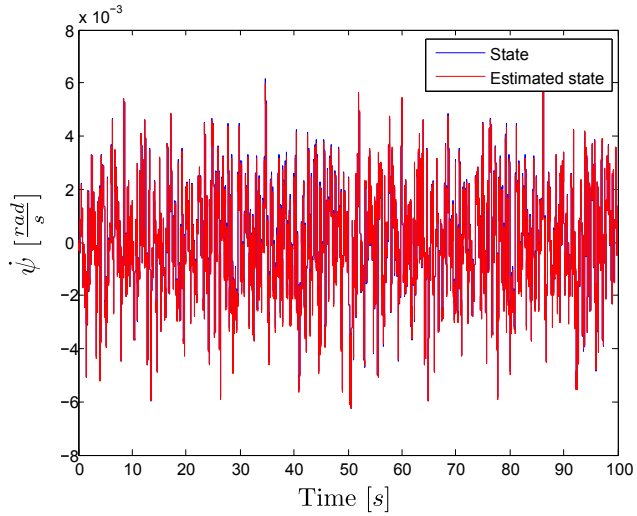


Figure 4.7: Estimation the angular yaw velocity of the first bogie

The state estimation of the other states is equivalent to the estimations shown in the figures. Thus, with the given acceleration signals, it is possible to estimate the states of the train.

4.2 Fault detection and isolation

Increasing demands for higher system performance and product quality on the one hand and more cost efficiency on the other hand made fault detection and isolation more and more interesting for the railway industry. Fault detection and isolation is the procedure of monitoring a system, identifying when a fault has occurred, and clarifying the type of fault and its location. There are two approaches to detect and isolate faults. On the one hand, there is the model based FDI procedure and on the other hand there is the signal processing based FDI. In the case of the model based FDI process, a model of the system is used to monitor the system. In the case of the signal processing based FDI process, a mathematical algorithm is used to analyze the signals and to make a statement about the health of the system. Thus, no model of the mechanical system is required. In this work, the FDI process is performed with a model based approach. A reliable train model and the Kalman filter discussed in the last section are essential parts of the fault detection procedure. In the case of the train model, the FDI process is used to detect and isolate faults in the suspension system. The anti-yaw damper, the secondary lateral, and vertical damper are exemplarily used to show the results of the proposed FDI method.

4.2.1 Estimation error

For the fault detection and isolation procedure, the state observation is performed with different systems. One of the systems represents the fault free case, the other systems represent the different faults, which should be detected. Thus, if n different faults should be detected and isolated, the observation process must be performed for $n+1$ systems

$$\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{w}, i), \qquad \mathbf{y}(k) = \mathbf{h}_i(\mathbf{x}(k), \mathbf{v}(k), i) \quad i = 0 : n. \tag{4.35}$$

In this work, $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{w}, i)$ is the differential equation describing the train dynamics combined with the form filter, and $\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{v}_k, i)$ describes the measurement output. All $n+1$ systems are linearized at the operating point

$$\mathbf{A}(i) = \left.\frac{\partial \mathbf{g}(i)}{\partial \mathbf{x}}\right|_{\mathbf{x}_0} \quad \mathbf{C}(i) = \left.\frac{\partial \mathbf{h}(i)}{\partial \mathbf{x}}\right|_{\mathbf{x}_0} \qquad i = 0 : n \tag{4.36}$$

and discretized. The discretized matrix of $\mathbf{A}(i)$ is denoted by $\mathbf{F}(i)$ and the discretized output matrix $\mathbf{C}(i)$ is denoted by $\mathbf{H}(i)$. Having $n+1$ systems, each system is used to compute an estimation of the state, denoted by $\hat{\mathbf{x}}(i)$. Using the estimation $\hat{\mathbf{x}}(i)$,

an estimation of the output is constructed

$$\hat{\mathbf{y}}_k(i) = \mathbf{H}(i)\hat{\mathbf{x}}_k(i). \tag{4.37}$$

To quantify the estimation error of each system, the state residual

$$\mathbf{r}_{x_k}(i) = \mathbf{x}_k - \hat{\mathbf{x}}_k(i) \tag{4.38}$$

and the measurement residual

$$\mathbf{r}_{y_k}(i) = \mathbf{y}_k - \hat{\mathbf{y}}_k(i) \tag{4.39}$$

are defined. Both residuals give information about the performance of the estimation process. Since the state vector $\mathbf{x}$ is not available, only the measured residual can be used to clarify the state estimation performance. To compare the performance of the different systems, the variance of the residual is calculated

$$Var[\mathbf{r}_{y_k}(i)] = \mathbf{r}_{y_k}(i)^T\mathbf{r}_{y_k}(i). \tag{4.40}$$

The lower the variance of the measurement residual, the better is the state estimation and the better the system is suited for the state estimation. Using the Kalman filter [33] to perform the state estimation, the covariance matrix of the measured residuals is calculated by

$$\begin{aligned} E[\mathbf{r}_{y_k}(i)\mathbf{r}_{y_k}(i)^T] &= E[(\mathbf{y}_k - \mathbf{H}(i)\hat{\mathbf{x}}_k^-(i))(\mathbf{y}_k - \mathbf{H}(i)\hat{\mathbf{x}}_k^-(i))^T] \\ &= \mathbf{H}(i)E[\mathbf{r}_{x_k}(i)\mathbf{r}_{x_k}(i)^T]\mathbf{H}(i)^T + E[\mathbf{v}_k\mathbf{v}_k^T] \\ &= \mathbf{H}(i)\mathbf{P}_k^-(i)\mathbf{H}(i)^T + \mathbf{R}. \end{aligned} \tag{4.41}$$

This matrix is used to attach specific weights to the measurement residuals, which increase the ability to choose the right system. The resulting scalar value is denoted by $e(i)$

$$e_k(i) = \mathbf{r}_{y_k}(i)^T(\mathbf{H}(i)\mathbf{P}_k^-(i)\mathbf{H}(i)^T + \mathbf{R})^{-1}\mathbf{r}_{y_k}(i). \tag{4.42}$$

For each system $i = 0 : n$, $e_k(i)$ is calculated at each instant of time k.

4.2.2 Kalman filter based fault detection

For all $n + 1$ systems, the state is estimated with the corresponding Kalman filter. In order to perform a state estimation, the different Kalman filters use the available measurement values. In each sample point, the measurement residual $\mathbf{r}_{y_k}(i) = \mathbf{y}_k - \mathbf{H}(i)\hat{\mathbf{x}}_k^-(i)$ and the fault indication value $e_k(i)$ are calculated.

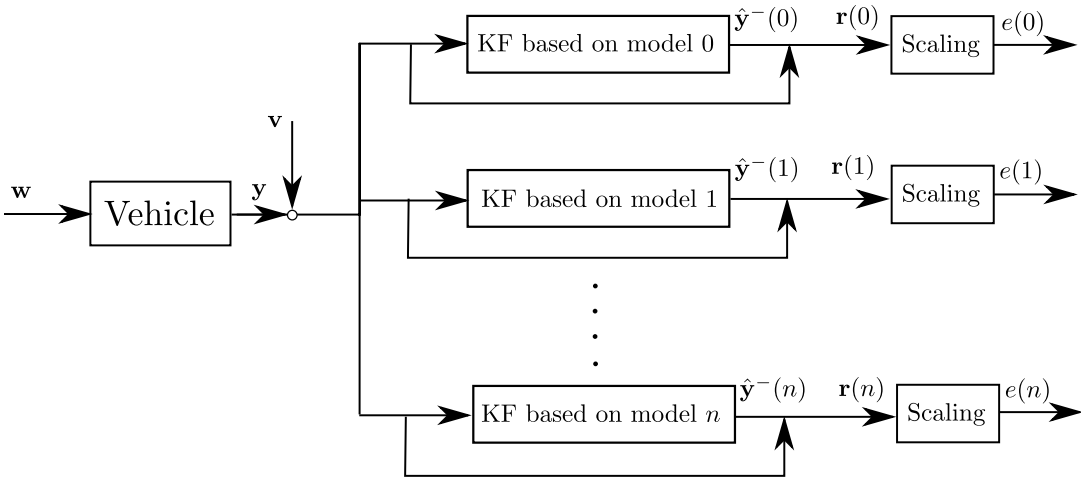


Figure 4.8: Multiple Kalman filters

The different linearized and discretized systems are given by

$$\mathbf{x}_k = \mathbf{F}(i)\mathbf{x}_{k-1} + \mathbf{w}_k \tag{4.43}$$

$$\mathbf{y}_k = \mathbf{H}(i)\mathbf{x}_k + \mathbf{v}_k, \tag{4.44}$$

with white, zero mean and uncorrelated process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$. The process noise and the measurement noise have known covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively,

$$\mathbf{w}_k \sim (\mathbf{0}, \mathbf{Q}), \tag{4.45}$$

$$\mathbf{v}_k \sim (\mathbf{0}, \mathbf{R}). \tag{4.46}$$

The different Kalman filters are initialized with the expected initial condition $\hat{\mathbf{x}}_0^+$ and the uncertainty of the initial estimation $\mathbf{P}_0^+$. The vector $\hat{\mathbf{x}}_0^+$ and the matrix $\mathbf{P}_0^+$ are considered to be the same for all $n+1$ systems

$$\hat{\mathbf{x}}_0^+(i) = \mathbf{E}[\mathbf{x}_0] \tag{4.47}$$

$$\mathbf{P}_0^+(i) = \mathbf{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)^T]. \tag{4.48}$$

The state estimation and the residual generation using the Kalman filter is performed with

$$
\begin{aligned}
&\textit{Prediction}: \\
\hat{\mathbf{x}}_k^-(i) =& \mathbf{F}(i)\hat{\mathbf{x}}_{k-1}^+(i) \\
\mathbf{P}_k^-(i) =& \mathbf{F}(i)\mathbf{P}_{k-1}^+(i)\mathbf{F}(i) + \mathbf{Q} \\
&\textit{Correction}: \\
\mathbf{K}_k(i) =& \mathbf{P}_k^-(i)\mathbf{H}(i)^T(\mathbf{H}(i)\mathbf{P}_k^-(i)\mathbf{H}(i)^T + \mathbf{R})^{-1} \\
\hat{\mathbf{x}}_k^+(i) =& \hat{\mathbf{x}}_k^-(i) + \mathbf{K}_k(i)(\mathbf{y}_k - \mathbf{H}(i)\hat{\mathbf{x}}_k^-(i)) \\
\mathbf{P}_k^+(i) =& \mathbf{P}_k^-(i) - \mathbf{K}_k(i)\mathbf{H}(i)\mathbf{P}_k(i)^- \\
&\textit{Residual}: \\
\mathbf{r}_{y_k}(i) =& \mathbf{y}_k - \mathbf{H}(i)\hat{\mathbf{x}}_k^-(i) \\
e_k(i) =& \mathbf{r}_{y_k}(i)^T(\mathbf{H}(i)\mathbf{P}_k^-(i)\mathbf{H}(i)^T + \mathbf{R})^{-1}\mathbf{r}_{y_k}(i).
\end{aligned}
\tag{4.49}
$$

The estimated states are used to calculate the measurement residual. With the measurement residual the fault indication e is calculated.

In order to decouple the fault detection alarm from the size of the track disturbance and unknown parameters, the values of $e(1)$ to $e(n)$ are always compared with the fault free case $e(0)$. The fault detection alarm is calculated by subtracting the fault indication value of the error free system from each fault indication value of the faulty systems

$$
\begin{aligned}
\epsilon_k(1) =& e_k(1) - e_k(0) \\
\epsilon_k(2) =& e_k(2) - e_k(0) \\
&\vdots \\
\epsilon_k(n) =& e_k(n) - e_k(0).
\end{aligned}
\tag{4.50}
$$

The index $i = 0$ stands for the fault free system. If there is a fault at the ith suspension element the estimation with the ith system should have the smallest estimation error. Thus, the filter that gives the "best" state estimates identifies the actual system.

Since the output $\mathbf{y}$ is corrupted by measurement noise, the residual e of a single point does not have sufficient information. Therefore, ϵ is averaged over a test period $T = 1 : m$

$$
\begin{aligned}
fault_1 =& \frac{1}{m}\sum_{k=1}^{m}\epsilon_k(1) \\
&\vdots \\
fault_n =& \frac{1}{m}\sum_{k=1}^{m}\epsilon_k(n).
\end{aligned}
\tag{4.51}
$$

If no fault occurs $fault_1$ to $fault_n$ should be positive. If there is a fault at the ith position, $fault_i$ should be negative to detect the fault, further it should be the smallest value to isolate the fault from the other possible faults.

4.2.3 Results of fault detection and isolation

The fault detection and isolation design procedure explained in the last section is used to detect faults in a full scale train model. The Velaro RUS serves as an example using parameters provided by the vehicle manufacturer. The track irregularities are generated numerically based on the PSDs. The FDI procedure is used to detect faults in all four secondary vertical dampers, in all four anti-yaw dampers and in all four secondary lateral dampers. Two of the dampers connect the front bogie to the carbody and two of them connect the rear bogie to the carbody, respectively. For the simulation process, the full scale train model is simulated with nonlinear wheel rail dynamics and nonlinear suspension system. This model is used to generate the signals of the acceleration sensors. For the estimation process, the train dynamic is linearized around the operation point.

Six different test cases were considered to test the FDI method:

1. a fault free train,
2. a fault at the front left secondary vertical damper,
3. a fault at the front left anti-yaw damper,
4. a fault at the front left secondary lateral damper,
5. faults at the front right secondary vertical, anti-yaw and secondary lateral damper,
6. faults at the front left, front right, rear left and rear right secondary vertical damper.

In order to test the robustness of the FDI procedure against track uncertainties, the gauge is modified in the simulation process, whereas the model for the FDI method uses the unchanged track gauge. In one case the gauge is increased by a constant value, in the other one it is decreased by a constant value. Further the FDI method is tested for a case where the track irregularities are twice as high as expected.

The measurements of the acceleration sensors were created with the nonlinear train model. To generate the measurements of the faulty systems, the corresponding damping coefficient is decreased by 30 %. The linearized test systems for the

estimation process consider a damper coefficient decrease of 20 % for the different fault cases. The test period is 100 seconds.

Figure 4.9 shows the results for the fault free case. All 12 fault indication values $fault_i$ have positive values.

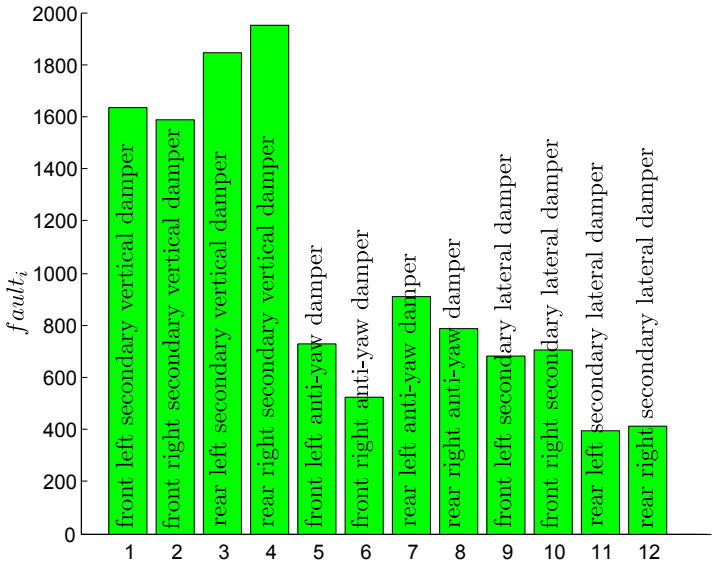


Figure 4.9: Fault free train

Thus, if all fault indication values are positive, the system is accurately classified as a fault free system. Figure 4.10 gives the results for the proposed FDI procedure for a fault at the front left secondary vertical damper. The results show that the indicator for the front left vertical damper has the smallest value. These results indicate a fault in the front left vertical damper. It can also be seen that the detection value for the front right vertical damper has negative values. This is due to the very similar effect of both dampers on the overall dynamics. For reasons of safety, both dampers, the left and the right anti-yaw damper, should be checked for faults.

Figure 4.11 shows the results for the FDI procedure for a fault at the front left anti-yaw damper. The results show again the accuracy of the algorithm. The value of the front left anti-yaw damper is the smallest. The second smallest value results from the fault detection value of the right front anti-yaw damper, which is caused by the close dynamic relation.

In Figure 4.12, the results for the FDI procedure for a fault at the front left secondary lateral damper is shown. The fault indicator for this damper has the smallest value and the fault detection value for the front right secondary lateral damper has the second smallest value. Again, this, is due to the fact, that both dampers have a similar influence on the system behavior. In case of the lateral damper, it is not possible to distinguish between faults at the front left or front

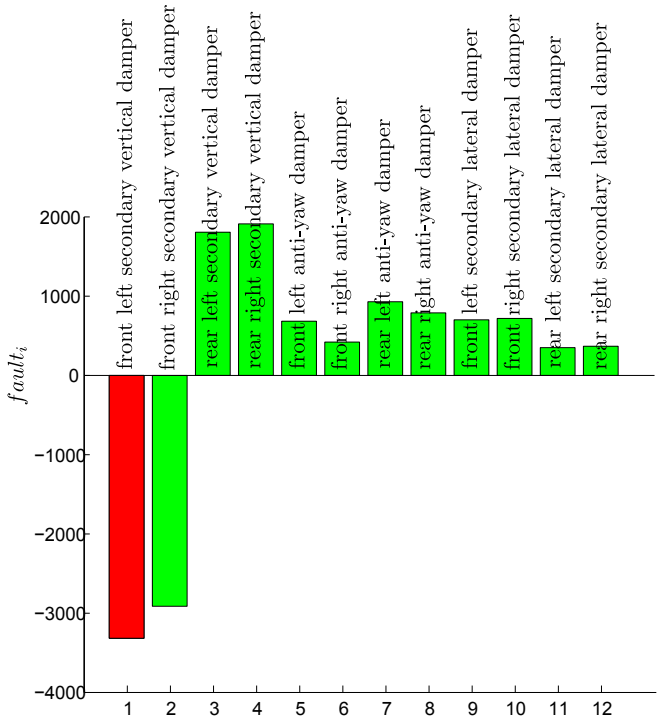


Figure 4.10: Fault at the front left vertical damper

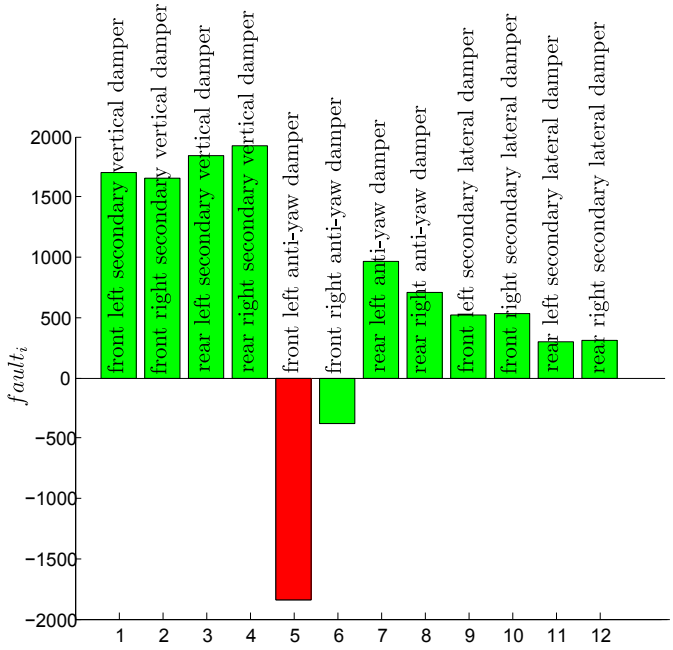


Figure 4.11: Fault at the front left anti-yaw damper

right side.

Figure 4.13 shows the results for a fault at the front right secondary vertical, anti-yaw and secondary lateral damper. The fault detection procedure is working well. All three dampers with a fault have a negative indication value.

Figure 4.14 shows the results for a fault at the front left, front right, rear left and

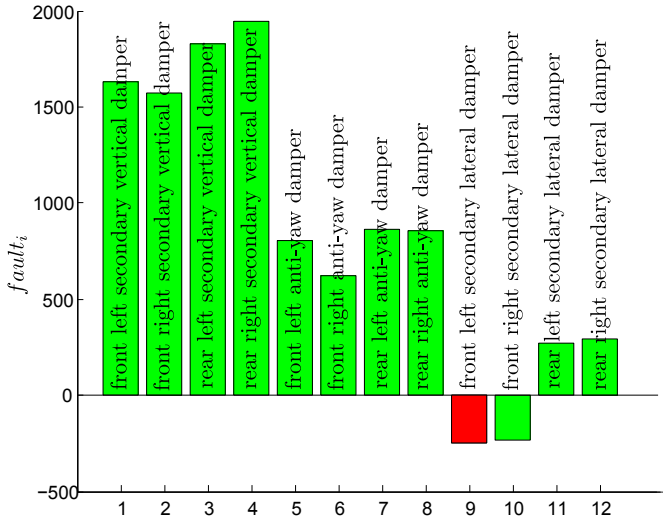


Figure 4.12: Fault at the front left lateral damper

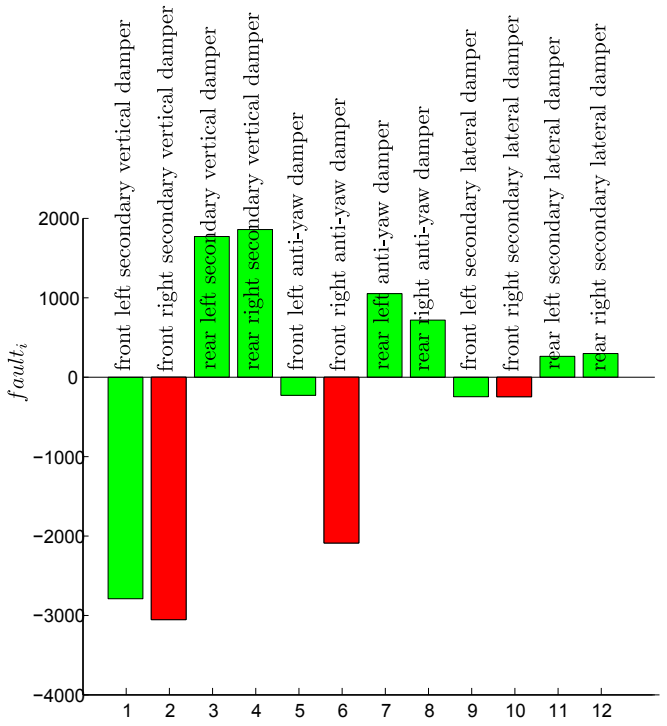


Figure 4.13: Multiple fault 1

rear right secondary vertical damper. All four faulty dampers have a negative fault value which correctly indicates the faults.

For all six test cases the FDI procedure gives an accurate fault indication. In the five faulty test cases the detection as well as the isolation of the specific fault is possible in most cases. Thus, it is possible to perform an accurate fault detection and isolation under these conditions.

In a second step, to investigate the robustness of the FDI method against uncertainties, parameters are varied. The track gauge is modified in the simulation

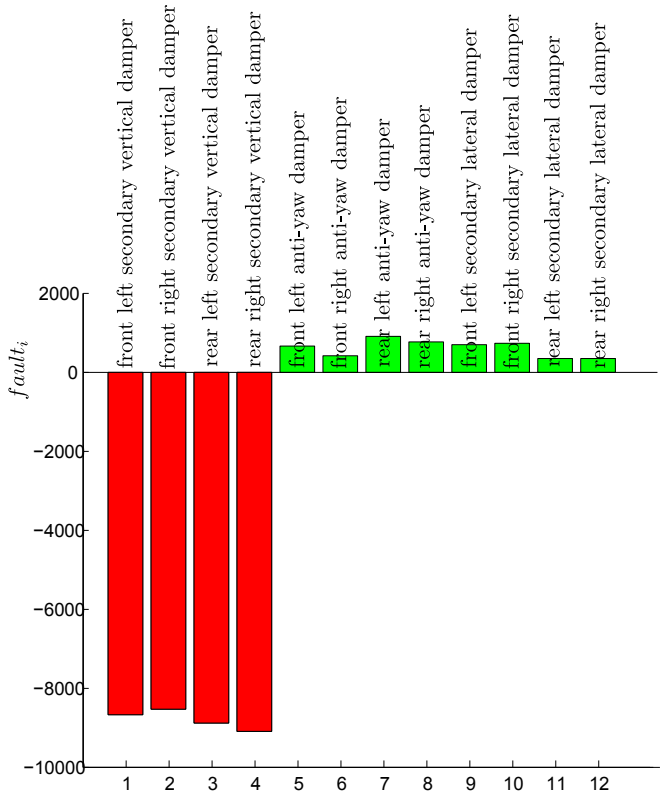


Figure 4.14: Multiple fault 2

process, whereas the FDI algorithm is not altered and uses the original track gauge. Figures 4.15 and 4.16 show the results for a track gauge reduction of $4mm$. Figure 4.15 shows the fault free case. Here, the indicator is accurate. Figure 4.16 shows the results for a fault in the front left anti-yaw damper, the FDI procedure accurately detects the fault.

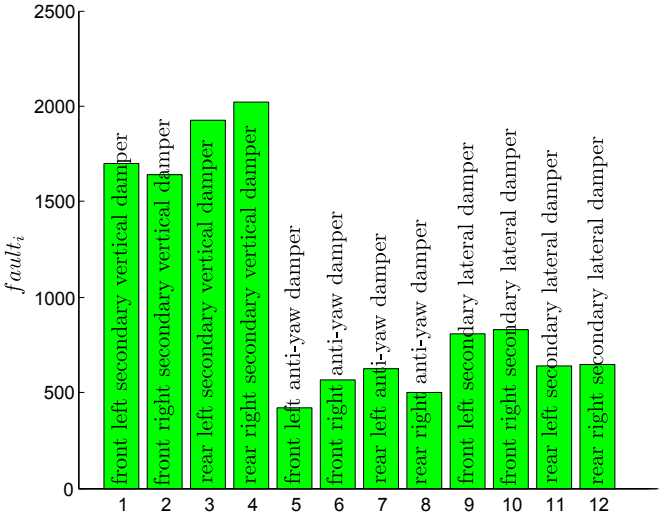


Figure 4.15: Fault free train with gauge reduction

Figures 4.17 and 4.18 show the results for a track gauge increase of $4mm$. The

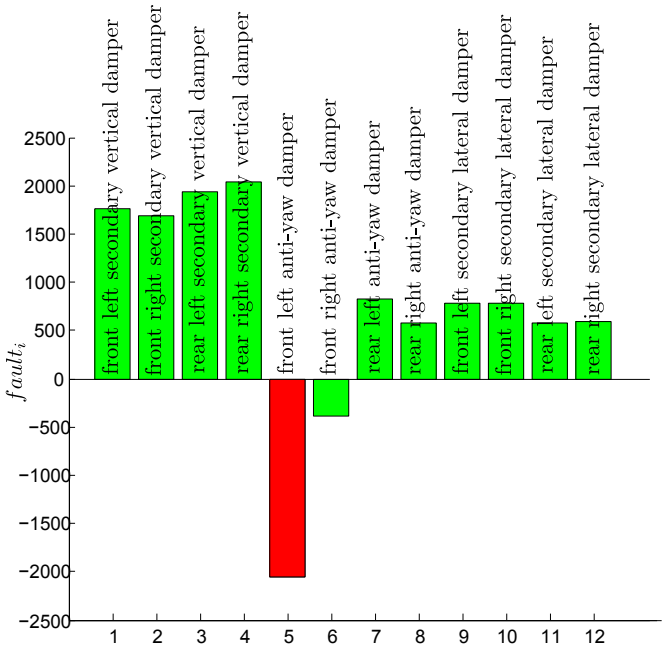


Figure 4.16: Fault at the front left anti-yaw damper with gauge reduction

results for the fault free case, shown in Figure 4.17, as well as for the case with a fault in the front left anti-yaw damper, shown in Figure 4.18, are detected correctly.

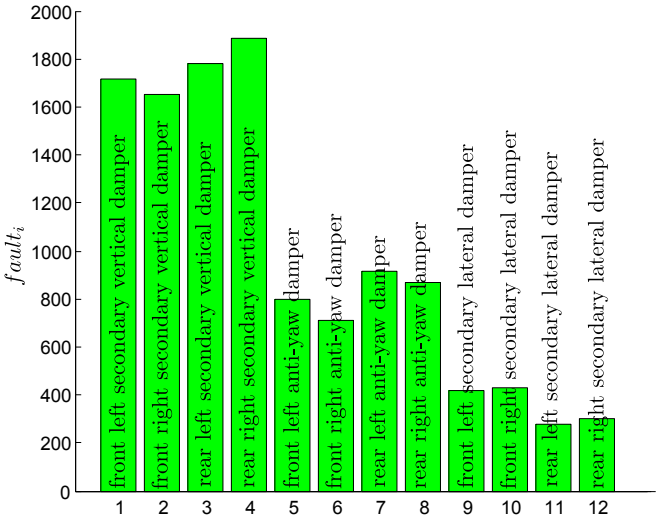


Figure 4.17: Fault free train with gauge increase

Thus, after variations of the track gauge, it is still possible to detect and isolate faults.

Figures 4.19 and 4.20 show the test cases where the track irregularities are twice as high as expected. Figure 4.19 shows the fault free case and Figure 4.20 shows the results for the test case with a fault at the front left anti-yaw damper.

Both results are qualitatively equal to the case with no track excitation increase and thus detect both test cases correctly.

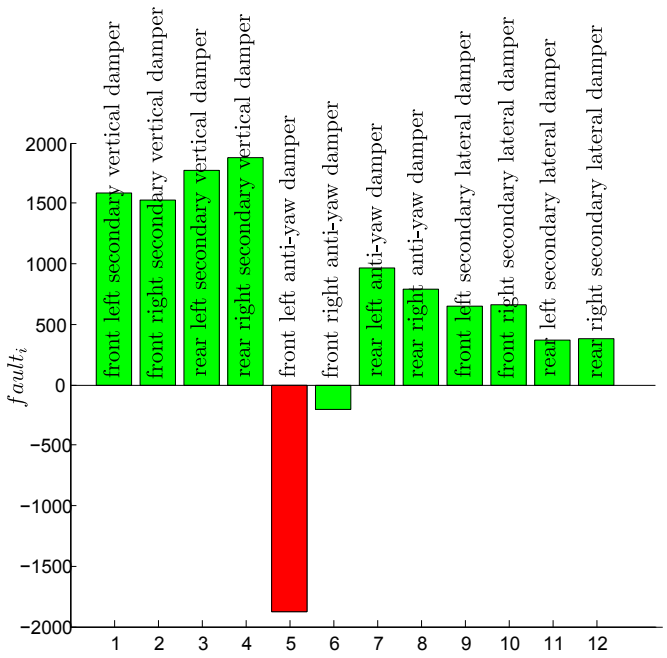


Figure 4.18: Fault at the front left anti-yaw damper with gauge increase

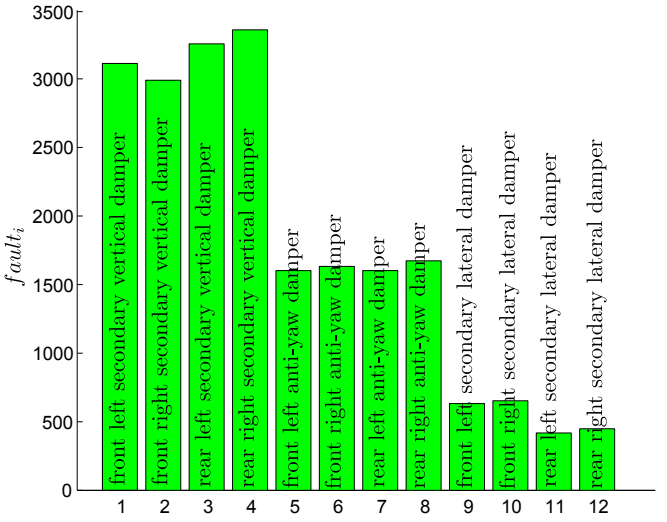


Figure 4.19: Fault at the front left anti-yaw damper with gauge increase

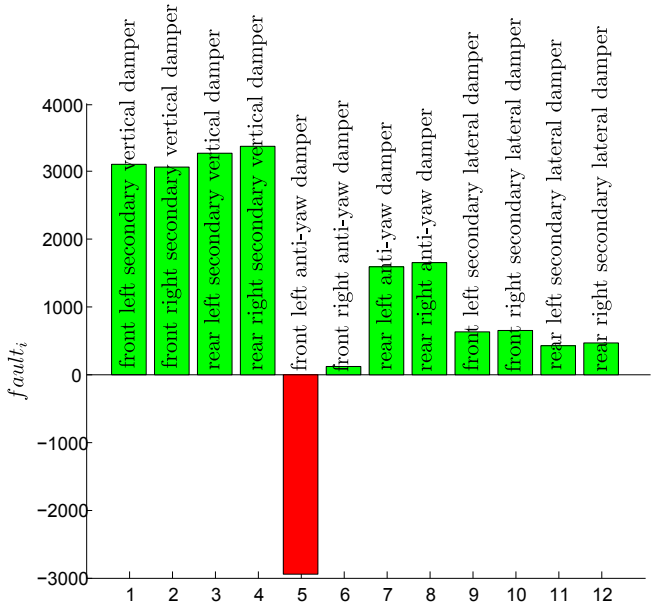


Figure 4.20: FDI results plotted over time

4.2.4 Stability analysis

Until now, the investigation to test the FDI method has considered single test cases. Since the results of the method depends, among others, on stochastic parameters, the fluctuation range of the results is tested. To determine the robustness of the proposed method, an extensive stochastic analysis is performed. Five different test cases are used for the robustness analysis.

1. a variation of the track excitation,
2. a variation of the velocity,
3. a variation of the mass of the coach,
4. a variation of the gauge size.
5. a variation of the track excitation strength.

In the first test case, the train is simulated 50 times. For each simulation, the track excitation is changed. Figure 4.21 shows the results for the fault free case and Figure 4.22 for the case with a fault at the front left anti-yaw damper. The height of the bar indicates the mean value of all 50 simulations, the upper line the maximum value and the lower line the minimum value. It can be seen that for all 50 simulations the fault free train as well as the fault at the anti-yaw damper is indicated correctly. The results for all 12 damper faults are shown in the Appendix A.

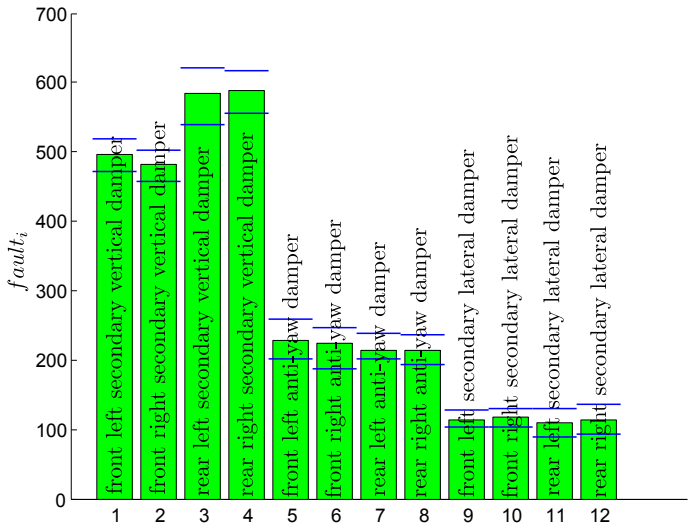


Figure 4.21: Fault free train with different excitations

For the second robustness test, the mass of the coach is varied from minus 20% to plus 20% in 1% steps. The mass is only changed for the simulation process, for

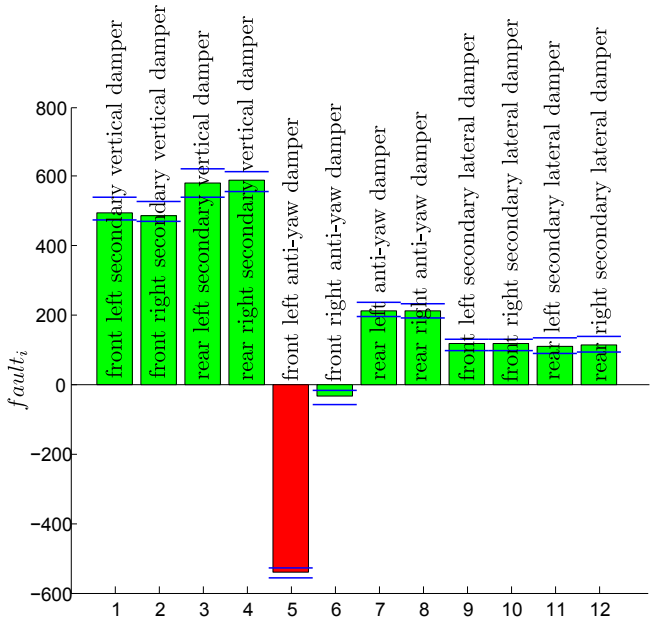


Figure 4.22: Fault at the front left anti-yaw damper with different excitations

the fault detection procedure the mass is kept constant. Figure 4.23 shows the fault free train and Figure 4.24 a fault at the anti-yaw damper. Both cases were indicated correctly. The results for all 12 damper faults are shown in the Appendix B.

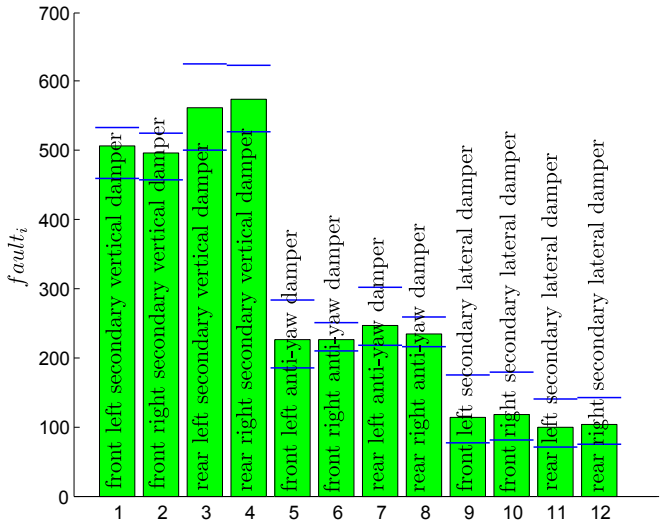


Figure 4.23: Fault free train with different weights for the coach

For the third robustness test, the track gauge is varied form minus 5 mm to plus 5 mm in 1 mm steps, for the fault detection procedure, the gauge is kept constant again. Figures 4.25 and 4.26 show the results for the fault free case and a fault at the anti-yaw damper. Both cases were indicated correctly.

For the next robustness test, the ride velocity is varied from 150 km/h to 300 km/h in 10 km/h steps. Figures 4.27 and 4.28 show the results for the fault free

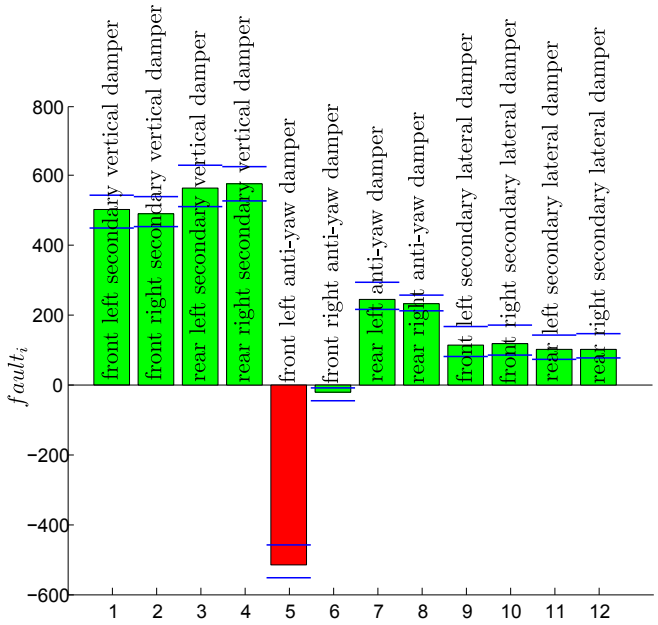


Figure 4.24: Fault at the front left anti-yaw damper with different weights for the coach

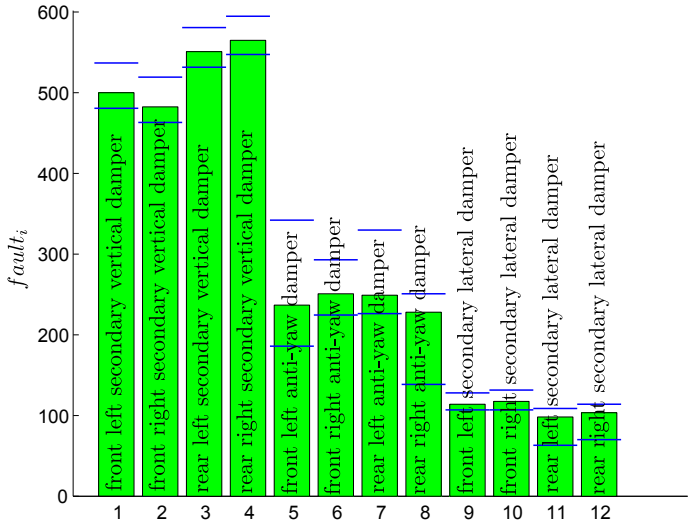


Figure 4.25: Fault free train with different gauge size

case and the case with a fault at the front left anti-yaw damper. The results for all 12 damper faults are shown in the Appendix C.

For the last robustness test, the track excitation is multiplied by a factor between 0.5 and 3 in 0.1 steps. Figures 4.29 and 4.30 show the results for the fault free case and the case with a fault at the front left anti-yaw damper.

Figure 4.31 shows the correlation between the different fault indications. The real fault is given on the x-axis, the fault detection is marked on the y-axis. A blue box indicates a positive fault indication value and therefore no fault, a red box indicates a negative fault indication value and thus a fault. All faults are indicated correctly

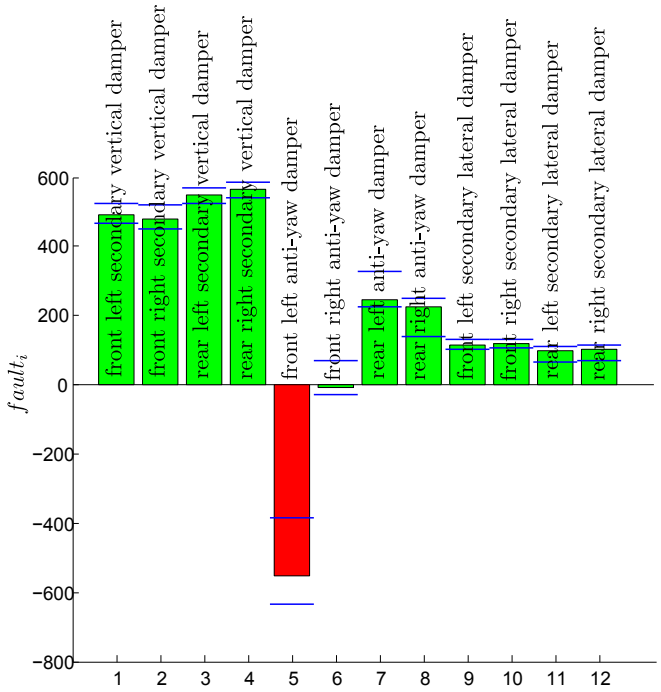


Figure 4.26: Fault at the front left anti-yaw damper with different gauge size

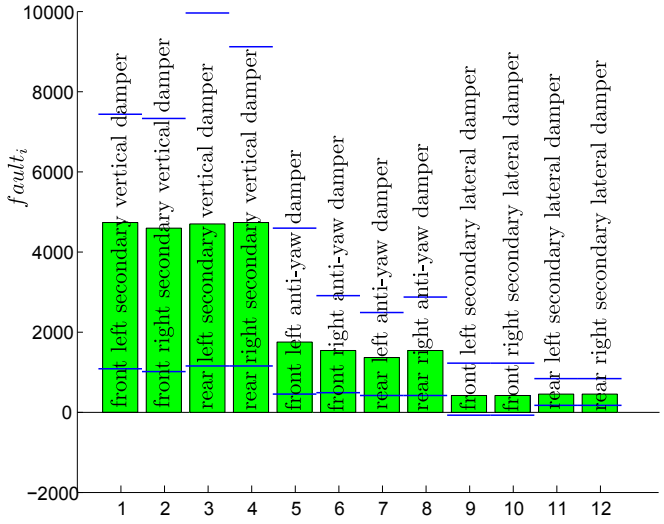


Figure 4.27: Fault free train with different forward velocities

and there are only few false detections. For all three types of suspensions, it is not possible to distinguish between a fault at the left and a fault at the right side of a suspension of the same type and the same bogie. Except for this case, all other faults are indicated correctly.

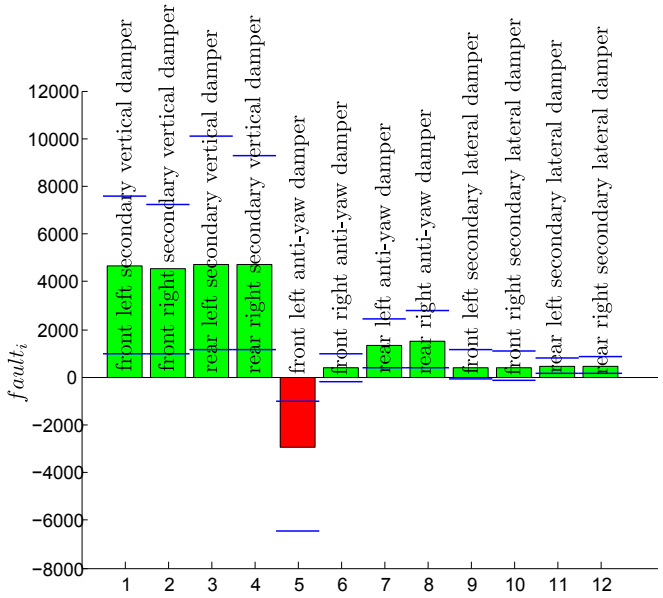


Figure 4.28: Fault at the front left anti-yaw damper with different forward velocities

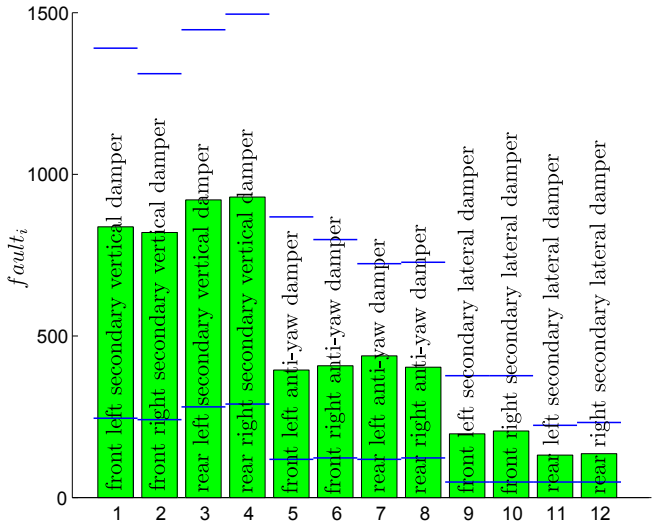


Figure 4.29: Fault free train with different forward velocities

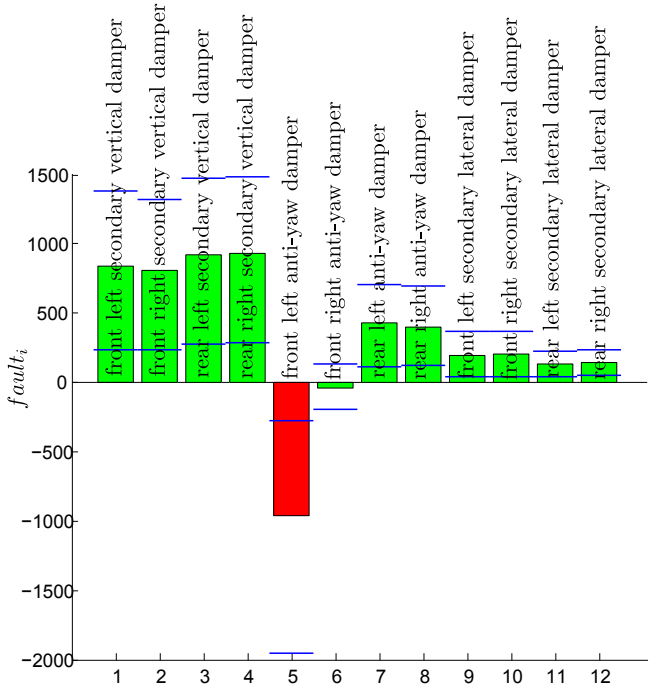


Figure 4.30: Fault at the front left anti-yaw damper with different forward velocities

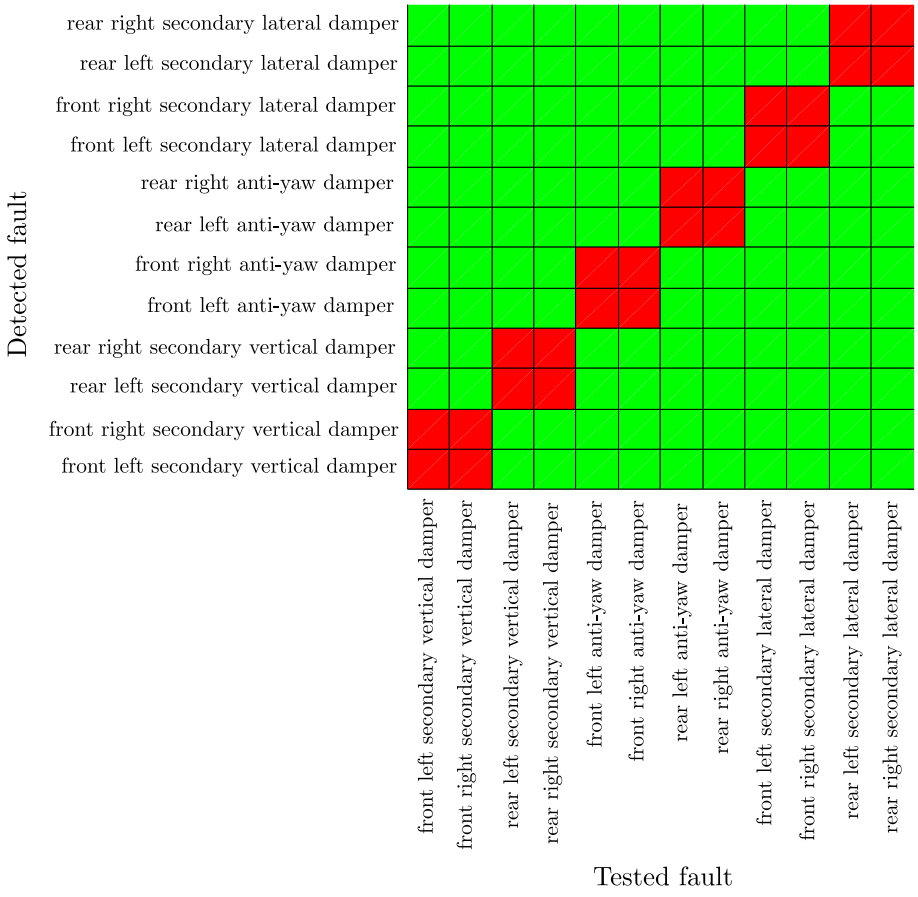


Figure 4.31: Fault correlation

4.3 Force estimation

The knowledge of the wheel rail forces provides information of the health of the track and the railway vehicle as well as information about ride conditions. Currently, wheel-rail contact forces are determined based on wheelsets equipped with strain gauges. In this method the contact forces are calculated by evaluating the measured strains. These custom made wheelsets are only used for specific and individual railway vehicle rides, for example to test new railway vehicles or tracks. Because of the high effort, this method is not used during regular operation. Therefore, a great demand of alternative ways to calculate the wheel-rail forces exists.

In recent years, several methods were published, which use acceleration sensors to determine the wheel-rail contact forces. In this section, the method described in [28] is used to determine the wheel-rail forces. A Kalman filter based approach is introduced to determine the wheel rail contact forces. The procedure to calculate the wheel-rail forces can be seen as an inverse problem. The forward problem refers to the determination of the acceleration values of the vehicle based on forces applied to the train. In the inverse problem, the forces are determined based on acceleration data. Further, the inverse force estimation procedure is extended to estimate the forces, which occur in the suspension system. Therefor the lateral and vertical secondary damper and the anti-yaw damper are chosen. The estimation of suspension forces brings benefits for maintenance purposes and strength calculations. The estimated forces can be used to optimize the compound of train–track and the suspension system.

At the beginning of this section, basics of inverse problem theory are given, followed by the procedure to calculate the wheel rail forces as well as the forces which occur in the suspension system.

4.3.1 Matrix inversion

Calculating the input of a system by using the output values of a system is called an inverse problem. Inverse problems come along with several difficulties as there might exist more than one solution. Depending on the forward problem, three different types are considered. For a linear system they are:

Type 1 A linear system is considered as determined if the number of equations is equal to the number of unknowns, in this case there exists either a unique solution or no solution.

Type 2 A linear system is considered as underdetermined, if there are fewer equations than unknowns, in this case there exist infinitely many solutions or no solution.

Type 3 A linear system is considered as overdetermined if there are more equations than unknowns, in this case there is a unique solution or no solution.

The second type is the one, which brings the most difficulties. If there is a solution, the only statement which could be given is, that the solution must be in a specific space. For the third type, if there is no solution, a unique approximation can be found which is as close as possible to the exact solution related to a specific norm.

For a system of linear equations, the mathematical description of a forward and an inverse problem can be given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad \text{forward problem} \tag{4.52}$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \qquad \text{inverse problem} \tag{4.53}$$

with the state vector $\mathbf{x}$ the output vector $\mathbf{y}$ and the output matrix $\mathbf{A}$. The matrix $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, in order for $\mathbf{A}^{-1}$ to exist, the matrix $\mathbf{A}$ must be square and it has to be non-singular. If the matrix $\mathbf{A}$ is invertible, the first type is addressed. If the dimension of $\mathbf{y}$ and $\mathbf{x}$ are not equal or the matrix $\mathbf{A}$ is singular, the inverse of $\mathbf{A}$ can not be calculated. However, to solve this problem, the so-called pseudoinverse matrix or generalized matrix is defined. Pseudoinverse or generalized matrices have similar characteristics as the regular inverse of a matrix and denoted by $\mathbf{A}^{+}$. Depending on the dimensions of $\mathbf{y}$ and $\mathbf{x}$ two different cases are considered. The case $r < n$ refers to the second of the three types denoted above, thus there are fewer equations than unknowns and the inverse solution may be overdetermined, the case $r > n$ refers to the third type, thus there are more equations than unknowns.

To cover the two different cases, the rank of a matrix is introduced. The rank of a matrix is defined as the maximum number of linearly independent rows or columns, thus the rank of a nonsingular $(n \times n)$ matrix is n. The rank of a $(r \times n)$ matrix can not be greater than the smaller dimension. The rank of the product of two multiplied matrices can be no greater than the smaller rank of the original matrices. To define the pseudoinverse the product $\mathbf{A}\mathbf{A}^T$ is used. If $\mathbf{A}$ is a $(r \times n)$ matrix, thus the dimension of $\mathbf{A}^T\mathbf{A}$ is $(n \times n)$ and of $\mathbf{A}\mathbf{A}^T$ is $(r \times r)$. Both products use the same matrix $\mathbf{A}$ and thus both products have a rank which is equal or less r or n, depending on which is the smaller one. For the case $r > n$, $\mathbf{A}^T\mathbf{A}$ may be invertible, depending on the rank of $\mathbf{A}$, but $\mathbf{A}\mathbf{A}^T$ is definitely not. For the overdetermined case, both sides are first multiplied with $\mathbf{A}^T$ and then with $(\mathbf{A}^T\mathbf{A})^{-1}$

$$\begin{aligned} \mathbf{A}^T\mathbf{A}\mathbf{x} &= \mathbf{A}^T\mathbf{y}, \\ \mathbf{x} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}, \\ \mathbf{x} &= \mathbf{A}^{+}\mathbf{y}. \end{aligned} \tag{4.54}$$

The overdetermined case has redundant information. If the vector $\mathbf{y}$ is consistent and error free, equation (4.54) provides the exact value of $\mathbf{x}$ with more computations than necessary. If the vector $\mathbf{y}$ is not consistent or contains errors, equation (4.54) minimizes the mean-square error in the estimation of $\mathbf{x}$.

For the underdetermined case, by considering that

$$\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{I} \tag{4.55}$$

equation (4.52) can be written as

$$\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y} \tag{4.56}$$

or

$$\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y} \tag{4.57}$$

$$\mathbf{x} = \mathbf{A}^+\mathbf{y}. \tag{4.58}$$

In contrast to the overdetermined case, the underdetermined case does not provide enough information for a unique solution of $\mathbf{x}$, but out of all possible solutions it does provide the solution that minimizes the quadratic norm of $\mathbf{x}$.

4.3.2 Inverse problem

For the train model, the relation between the wheel rail force and the sensor outputs is given by the state space model

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{w} \tag{4.59}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}, \tag{4.60}$$

where $\mathbf{x}$ is the state vector, $\mathbf{w}$ the input vector and $\mathbf{y}$ the output vector and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the system matrices. The forward problem is to calculate the measurements $\mathbf{y}$ for given track excitation $\mathbf{w}$. The inverse problem is to calculate $\mathbf{w}$ for given $\mathbf{y}$. Compared to the case given in equation (4.52) the relation between the input and output can not be solved by a simple algebraic equation, further, the actual state of the dynamical system must be considered.

The problem can be divided into two steps: The first step is to calculate the states of the system by given measurements

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

this is a problem of the second type, there exist either no solution or infinitely many solutions. If the problem is solved with the pseudoinverse, the solution is the one which minimizes the quadratic norm of $\mathbf{x}$. In this case, no information about the previous state and the dynamic characteristic of the system is used. To involve the dynamic characteristic in the state estimation process, a state observer is used. The Kalman filter is applied, as stated above, to estimate the state vectors of the system.

Having information about the state, the inverse problem

$$\mathbf{w} = \mathbf{B}^{+}(\dot{\mathbf{x}} - \mathbf{A}\mathbf{x})$$

can be solved. This is a problem of the third type and a unique solution can be found by solving the linear least square problem.

4.3.3 Force calculation

In the case of the state space model of the railway vehicle given in section 3, the input vector $\mathbf{w}$ describes the track position. If the information about the force at the wheel is needed, new output matrices must be introduced. The forces of all wheelsets are summarized in the vector $\mathbf{F}$. The first step is to calculate the influence of the input $\mathbf{w}$ and the state vector $\mathbf{x}$ to the force vector $\mathbf{F}$. This is simply done by using the nonlinear equation describing the train model and performing a variation of the state and input vector. The linear relation between state vector, input vector and the wheel forces can then be given by

$$\mathbf{F} = \mathbf{C}_F\mathbf{x}_s + \mathbf{D}_F\mathbf{w}.$$

For each wheelset, the vertical and horizontal forces shown in Figure 4.32 are estimated.

The same procedure is performed to estimate an arbitrary suspension force. New output matrices are defined, which describe the influence of the input $\mathbf{w}$ and the state vector $\mathbf{x}$ on the suspension force.

If the state is estimated with the Kalman filter, the only unknown is $\mathbf{w}$. By introducing the form filter of the track irregularities, given in Section 3.1.4, the output vector $\mathbf{w}$ is part of the state vector and the system is solved by

$$\mathbf{F} = \begin{bmatrix} \mathbf{C}_F & \mathbf{D}_F \end{bmatrix} \mathbf{x}.$$

Thus, to calculate the wheel rail contact force or any suspension force, the estimated state $\hat{\mathbf{x}}_s$ and the estimated input $\hat{\mathbf{w}}$ of the railway vehicle system are needed.

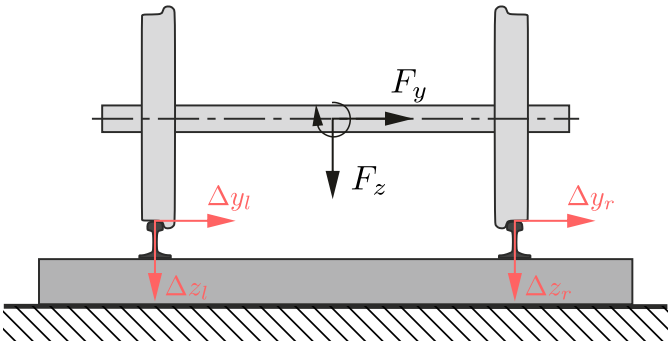


Figure 4.32: front left secondary vertical damper

According to [28], the procedure of determining the wheel-rail forces is divided into five steps:

1. Simulation of the railway vehicle system and generation of measurement data
2. Extension of the linearized system with a form filter
3. State and input estimation with a Kalman filter on basis of measurement data
4. Force calculation with estimated states and inputs
5. Comparison of estimated and simulated results for the input forces

If real measurements are available the simulated measurements are simply replaced by the real measurements.

4.3.4 Wheel rail contact force estimation

With the method presented above, the wheel rail contact forces are estimated. In this section, the estimated forces are compared with the simulated force. For this purpose, the train is simulated with a speed of $160\frac{km}{h}$ and excited with track irregularities.

A comparison of estimated and simulated signals for the resultant vertical contact forces for the first wheelset is shown in Figure 4.33 and an enlargement of the diagram is given in Figure 4.34. The real and the estimated contact forces show a significant correlation.

Figure 4.35 shows the results for the estimation of the lateral contact forces of the first wheel set. An enlargement of the diagram is given in Figure 4.36. There is,

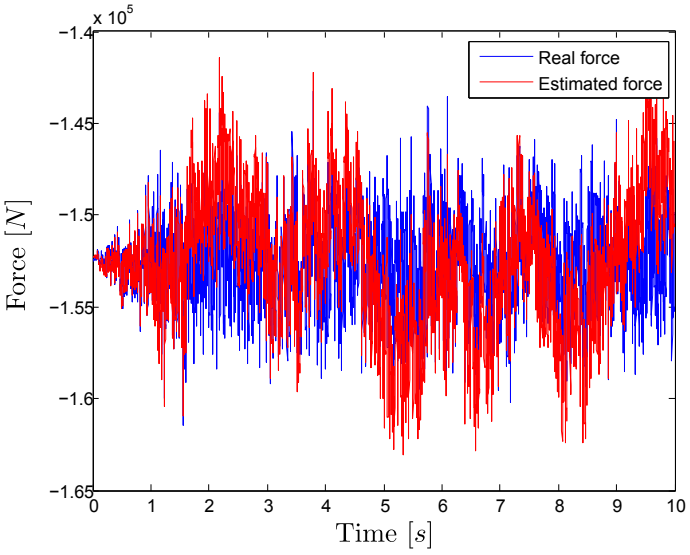


Figure 4.33: Vertical contact force first wheel

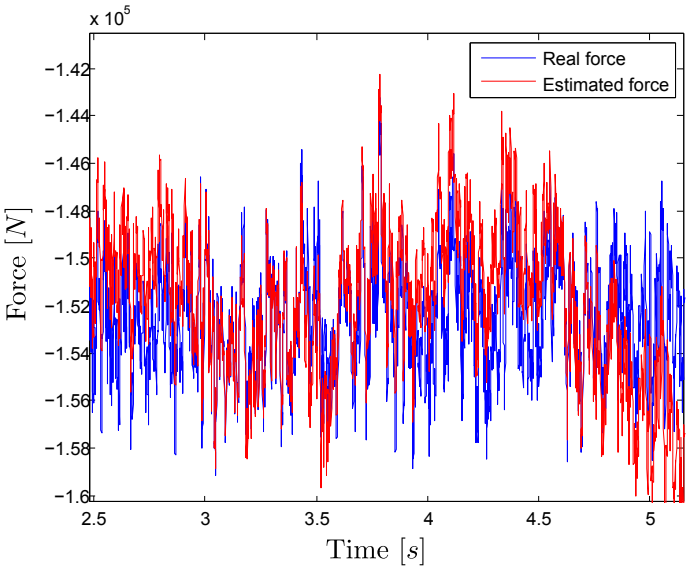


Figure 4.34: Vertical contact force first wheel enlarged

again, a high correlation between the estimated and the simulated force. Figures 4.37 and 4.38 show the results for the estimation of the vertical and lateral contact forces of the first wheelset at the left wheel.

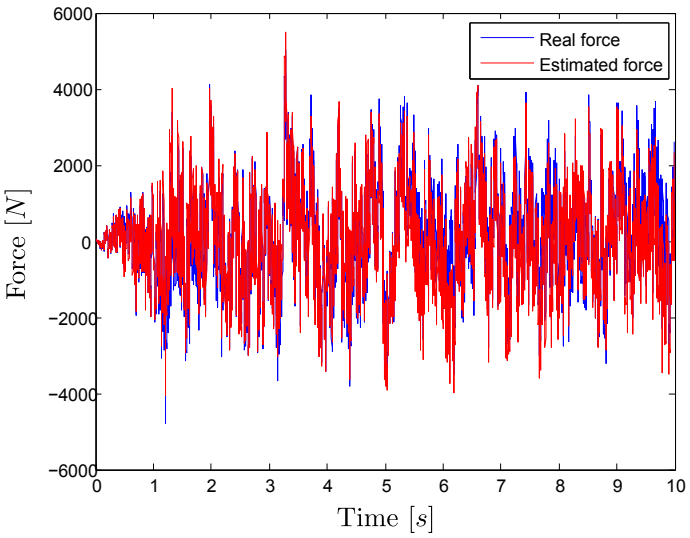


Figure 4.35: Lateral contact force first wheel

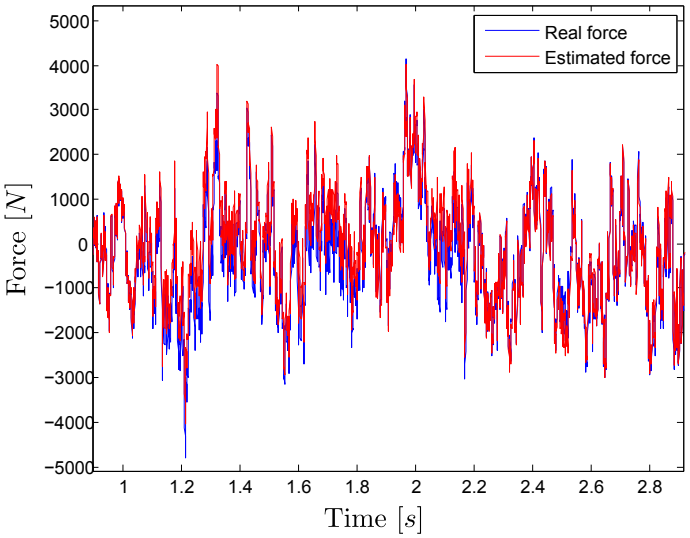


Figure 4.36: Lateral contact force first wheel enlarged

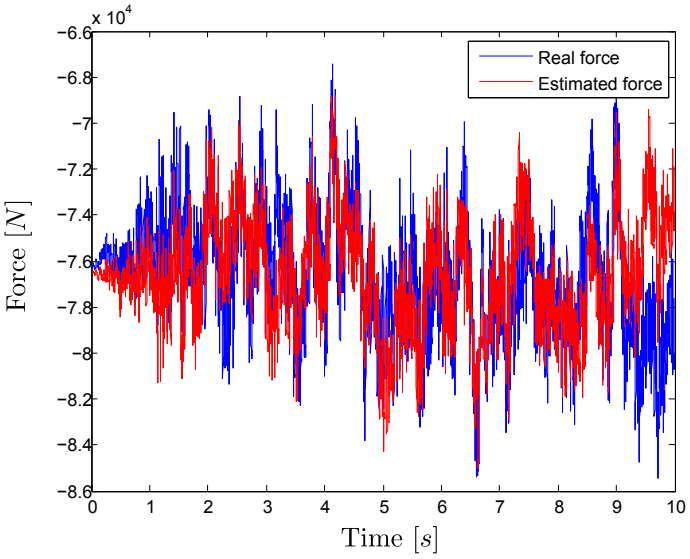


Figure 4.37: Vertical contact force first wheel

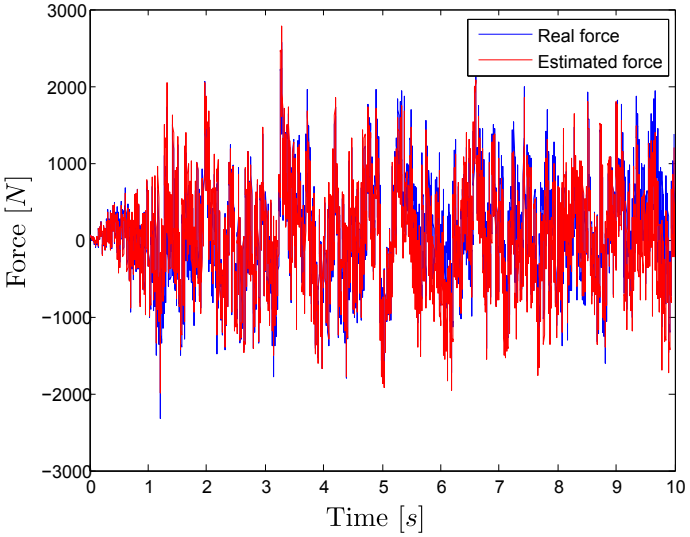


Figure 4.38: Lateral contact force first wheel

The lateral and vertical forces of a single wheel play a major role in the design of the railway vehicle [51]. For example, Nadal proposed a single-wheel limit criterion to guarantee flange climb safety. This criterion is based on the ratio of the vertical wheel force Q_F and the lateral wheel force Y_F of a single wheel shown in Figure 4.39.

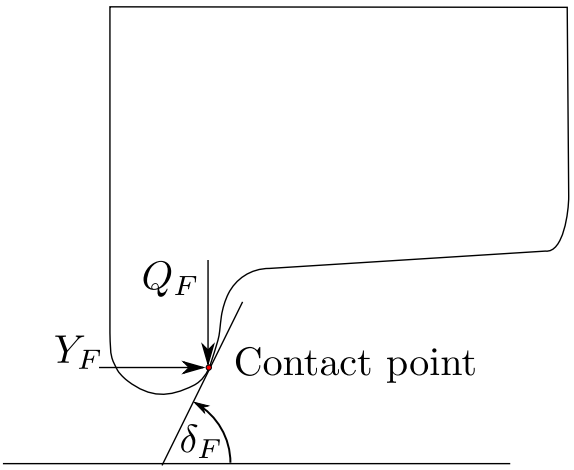


Figure 4.39: Flange contact

The critical ratio is given by [52]

$$\frac{Y_F}{Q_F} = \frac{\tan\delta_F - \mu}{1 + \mu\tan\delta_F}, \tag{4.61}$$

where δ_F is the contact angle. By using the maximum contact angle, this equation gives the minimum wheel Y_F/Q_F ratio at which the flange climb derailment occurs for given friction coefficient μ. Thus the estimated lateral and vertical forces of a single wheel can be used to perform safety analysis based on acceleration measurements.

Equivalent to the safety calculations of the wheel, the estimated forces are used to quantify the load condition applied to the rail. Lateral track shifts are of major interest. These are caused by repeated lateral axle loads. The track can shift under large lateral forces and since both, speed and load, of the railway vehicle increase, it is important to check the load condition. Large lateral forces can induce a gauge widening that can lead to a wheel-rail separation, as shown in Figure 4.40 or a rail rollover shown in Figure 4.41.

The estimated lateral wheel rail forces can be used to verify the load applied to the rail and prevent such occurrences.

Further, the estimated wheel rail forces can be used to perform wear calculation for example like proposed in [53].

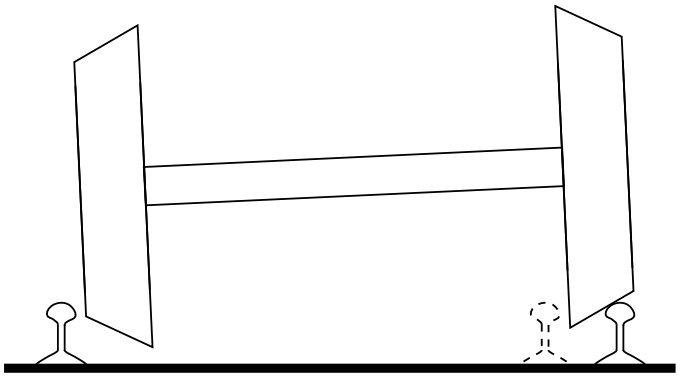


Figure 4.40: Separation of rail

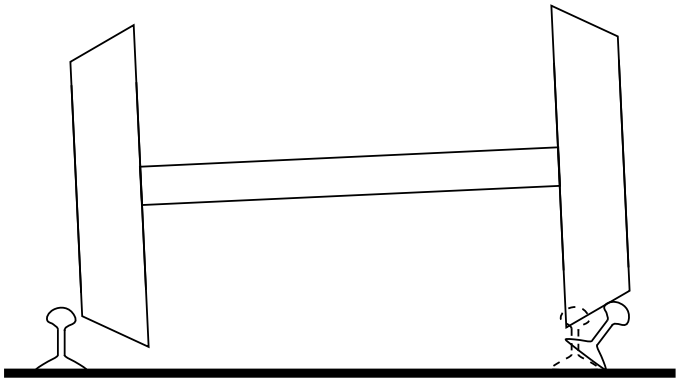


Figure 4.41: Rollover of rail

4.3.5 Suspension force estimation

Most components of railway vehicles, including bogies or wheelsets, are elements where dynamic stresses are decisive in design calculations. Material fatigue of the bogie or the wheel components takes place due to dynamic loads, which is superimposed on the static loads from weight and loading. A precise information of the dynamic loads during the life cycle of a component is therefore important for optimal component design or failure analysis. The consideration of dynamic load components in design calculations as carried out to date, however, are almost entirely greatly simplified through dynamic plus-factors [54]. The information about the dynamic load can be provided by estimating the suspension forces. The obtained results of the suspension force estimation are shown in this section.

Figure 4.42 shows the results for the estimation of the front left secondary vertical damper. Figure 4.43 shows the results for the estimation of the front left anti-yaw damper and Figure 4.44 shows the results for the estimation of the front left secondary lateral damper.

It can be seen, that all three suspension forces can be estimated accurately.

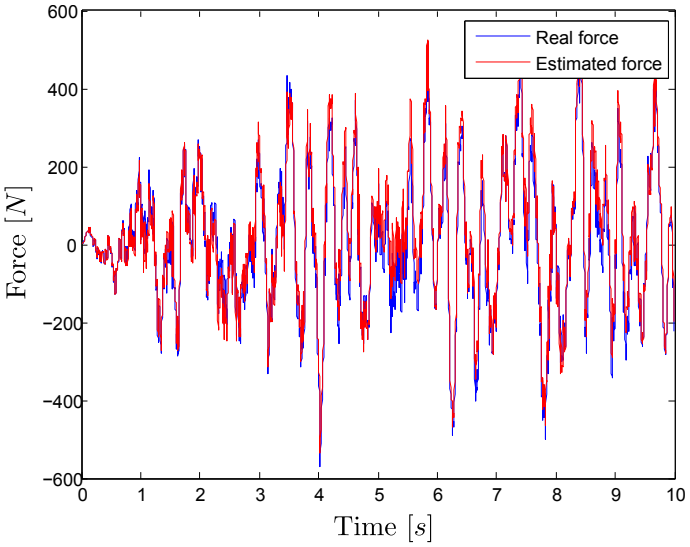


Figure 4.42: Front left secondary vertical damper

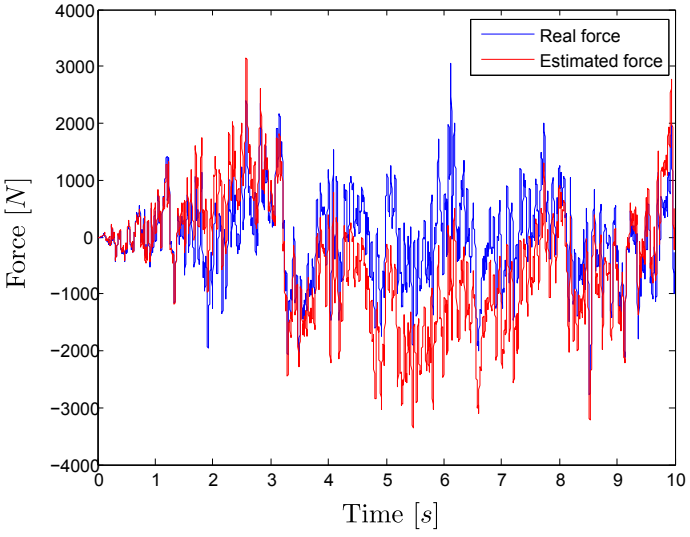


Figure 4.43: Front left anti yaw-damper

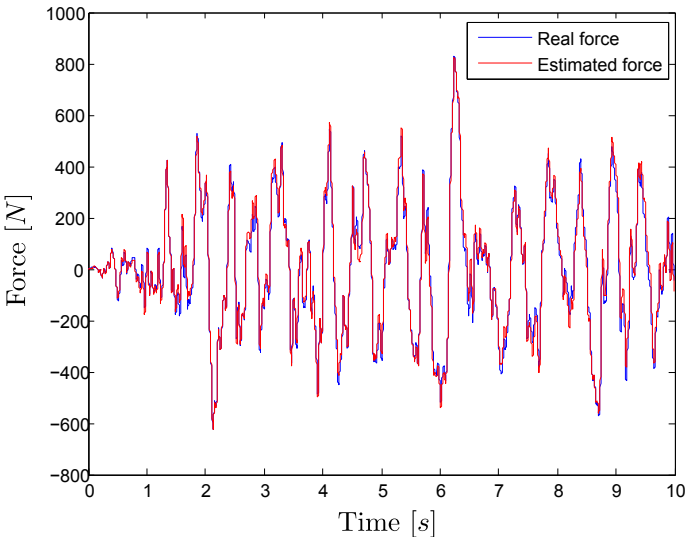


Figure 4.44: Front left secondary lateral damper

5 Conclusion

Increasing system reliability and dependability while decreasing maintenance costs is a major topic in the railway industry. On-line fault detection and isolation offers advantages, when early detection of faults and wear is crucial. Further fault detection and isolation together with the ability to estimate forces in the railway system brings benefits for maintenance, strength calculation and asset management. Important forces which are needed for such purposes, occur in the wheel rail contact and the suspension system of the railway vehicle.

The review of literature in the first chapter has shown, that there is demand for an improve in fault detection and force estimation. Many concepts were presented but no procedure was found, which allows for the detection and estimation of a variety of faults and forces with such an accurate railway vehicle model and track model as provided in this work.

The task of this work is to analyze a railway vehicle model based on acceleration data measurements. The train dynamics are captured with a state observer and an algorithm is introduced to detect and isolate faults in the suspension system and to estimate forces in the wheel rail contact as well as in the suspension system. The characteristics of the track irregularities are used to increase the accuracy of the proposed method. The PSDs given in ERRI B176 are taken to describe the characteristic of the track. A form filter is designed to model the random track irregularities corresponding to the PSD, which improves the results of the force estimation. The railway vehicle system is combined with the form filter, and a Kalman filter is used to perform a state estimation of the whole system.

In the first topic of this work, expanded multiple model Kalman filters were used to detect and isolate faults in a suspension system of a full scale railway vehicle model. Important components in railway suspension systems are the anti-yaw damper, the secondary lateral, and vertical damper. The anti-yaw damper is critical for running stability, the other two influence mainly the ride comfort. In this study, the presented FDI procedure is used in order to detect faults in any of these dampers. As test case, a model of the Velaro RUS is used to detect faults. With the presented method, faults in the vertical, lateral and anti-yaw damper can be distinguished sensitively and reliably. It is even possible to isolate faults with nearly the same influence to

the overall dynamics. Further single as well as multiple faults can be isolated.

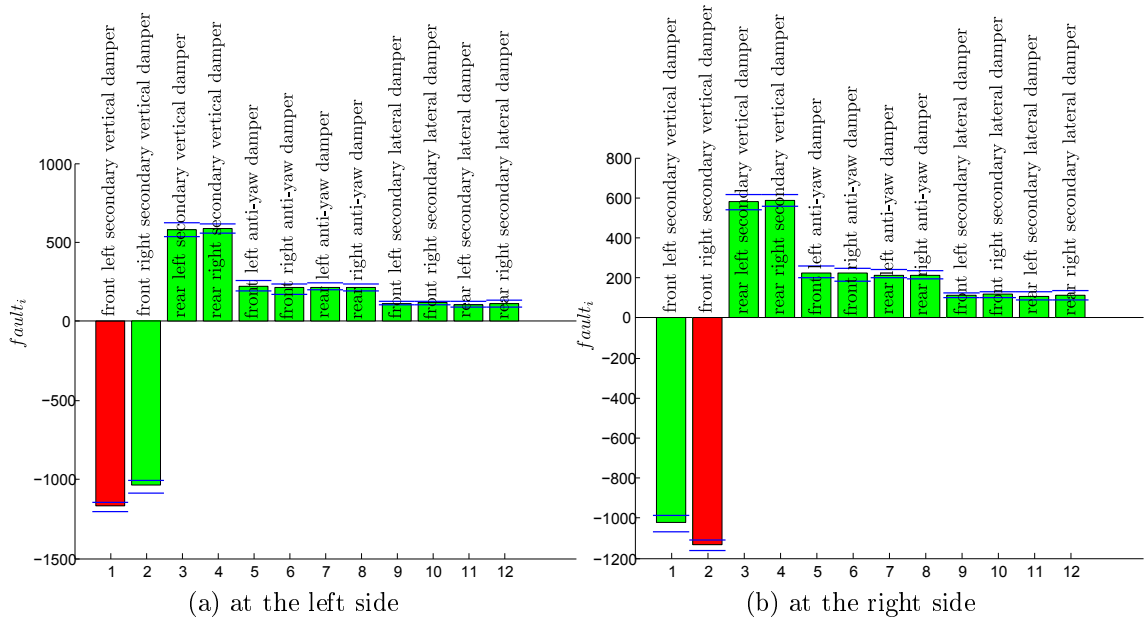
The robustness of the FDI procedure is tested by a variety of modifications. The random values to create the track irregularities are changed 50 times to quantify the deviation of the FDI method. Further, the travel velocity is changed between 150 km/h and 300 km/h in 10 km/h steps. The mass of the coach is modified between minus and plus 20%. The gauge size is changed between minus 5 mm and plus 5 mm in 1 mm steps and at last the excitation amplitude is changed between 0.5 and 3 times of the regular excitation. Compared to the other changes, the change of velocity is included in the FDI model. For the other test cases the FDI model holds the original values. All test cases show the robustness of the proposed method and are working well and reliably. Because of the robustness of the FDI method, the author believes that the proposed method could be used to detect and isolate faults in real scenarios.

In the second topic the Kalman filter is used for the inverse determination of lateral and vertical wheel-rail contact forces and suspension forces. The wheel rail contact is a major part of a railway vehicle system and information about these forces is important for wear and comfort analysis. Currently, wheel-rail contact forces are determined based on wheelsets equipped with strain gauge. As in the first topic, the method in this work only uses data from acceleration sensors to estimate the forces. A Kalman filter is used to estimate the states of the system, whereas the system combines the railway vehicle and the form filter with the characteristics of the track irregularities. With the states of the railway vehicle and the track geometry, the forces are estimated. The results show that it is possible to estimate the lateral and vertical wheel rail contact forces. Methods are shown, which can use the estimated wheel rail contact forces for a variety of safety calculations. Further, the estimation of the forces in the secondary vertical damper, the secondary lateral damper and the anti yaw-damper is possible. This estimation can be used for strength analysis.

In both cases, the estimation of the wheel rail contact forces as well as the estimation of the suspension forces, the real and estimated forces highly correlate. Thus, the proposed methods for both, fault detection and forces estimation can be used to increase the safety and decrease maintenance costs.

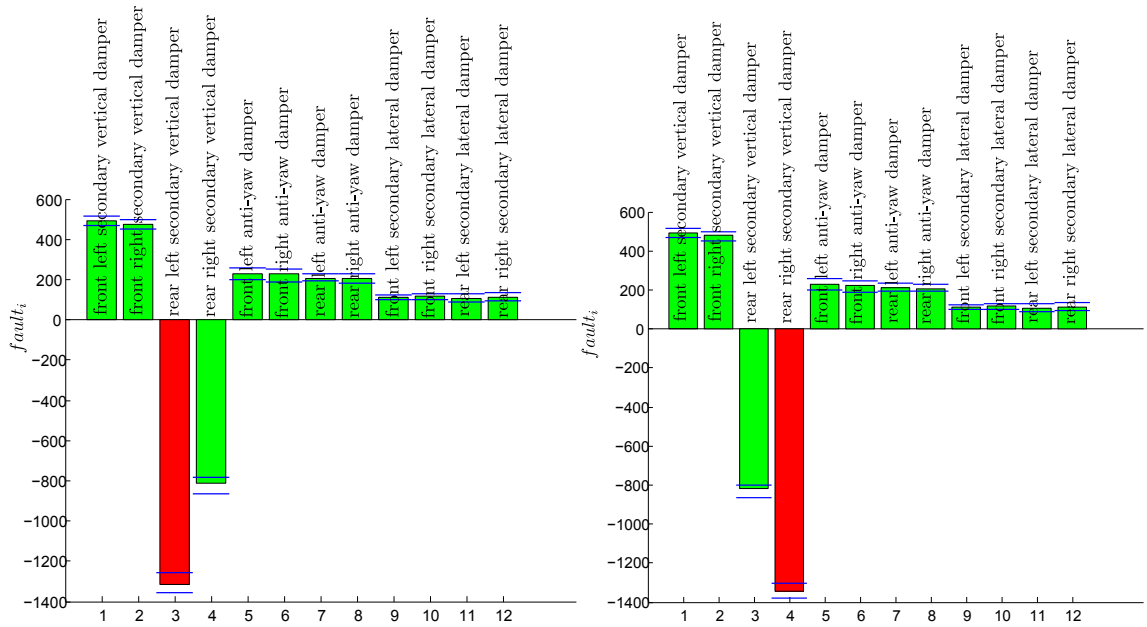
The focus of this work was to develop a reliable strategy to analyze a railway vehicle based on acceleration sensors. The proposed methods have been tested by simulating a railway vehicle with nonlinearities. The methods are based on a linearisation of this model and show good results.

A Results for all suspension faults



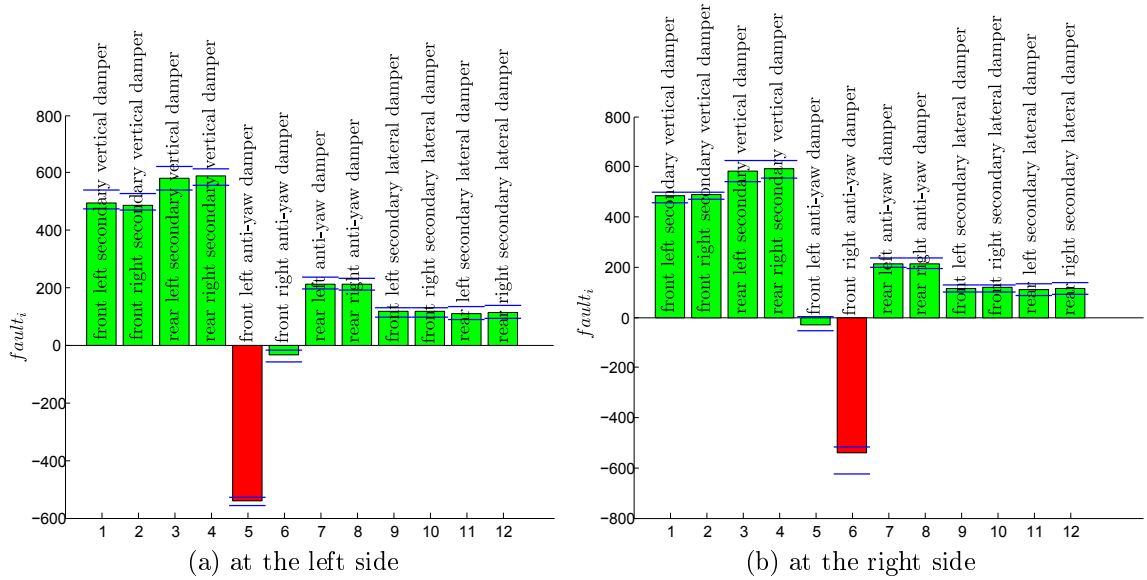
(a) at the left side (b) at the right side

Figure A.1: Front secondary vertical damper fault



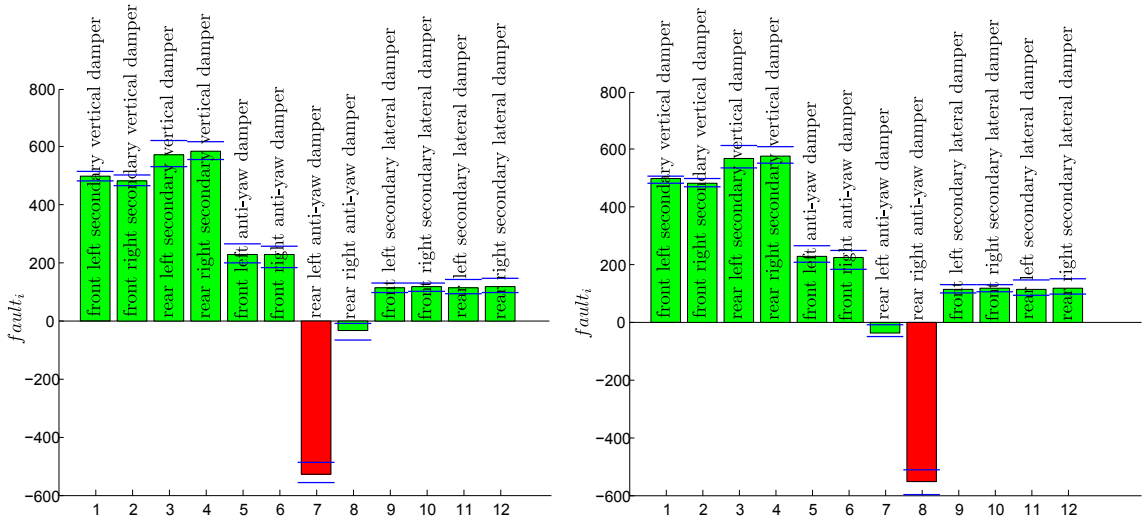
(a) left side (b) right side

Figure A.2: Rear secondary vertical damper fault



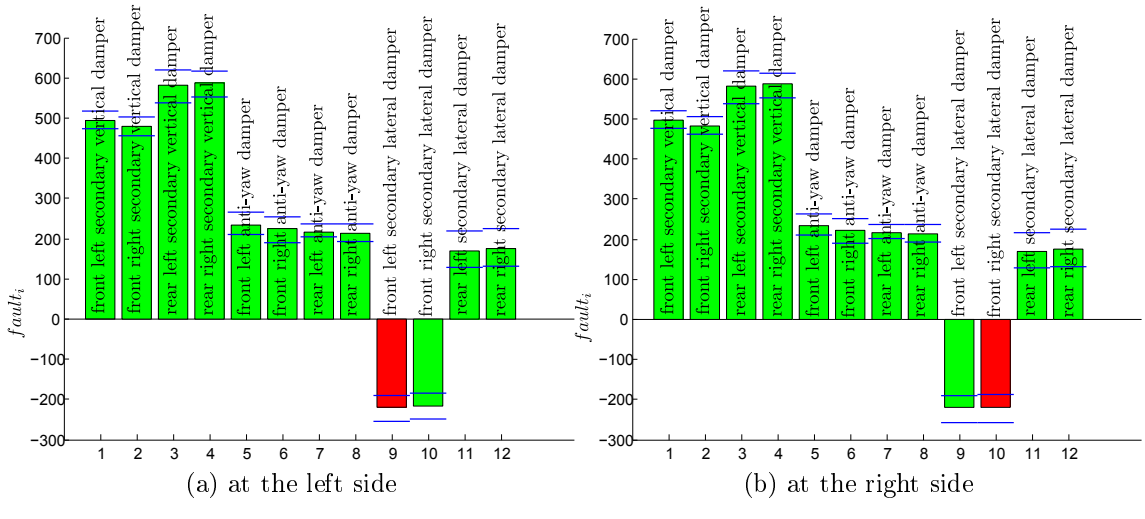
(a) at the left side (b) at the right side

Figure A.3: Front anti-yaw damper fault



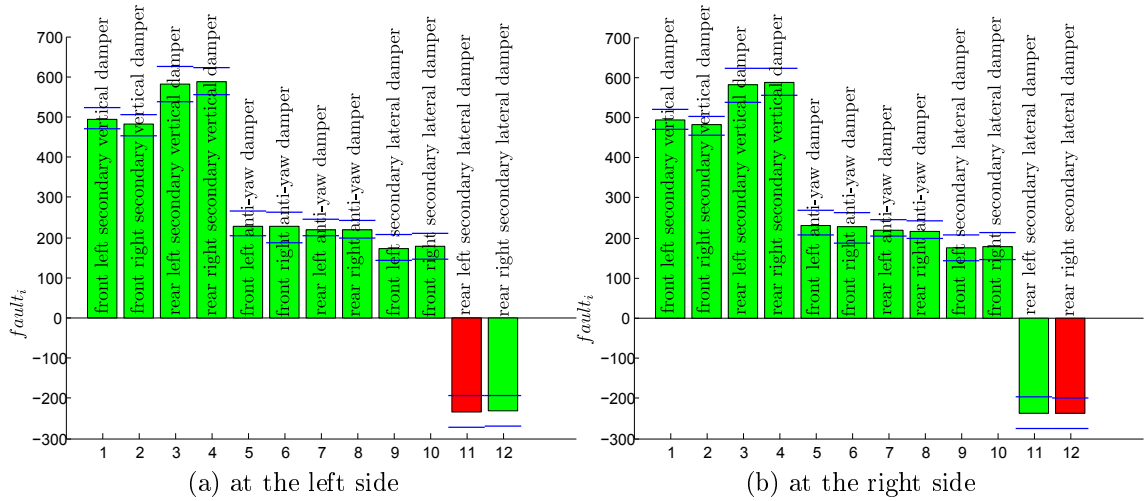
(a) at the left side (b) at the right side

Figure A.4: Rear anti-yaw damper fault



(a) at the left side (b) at the right side

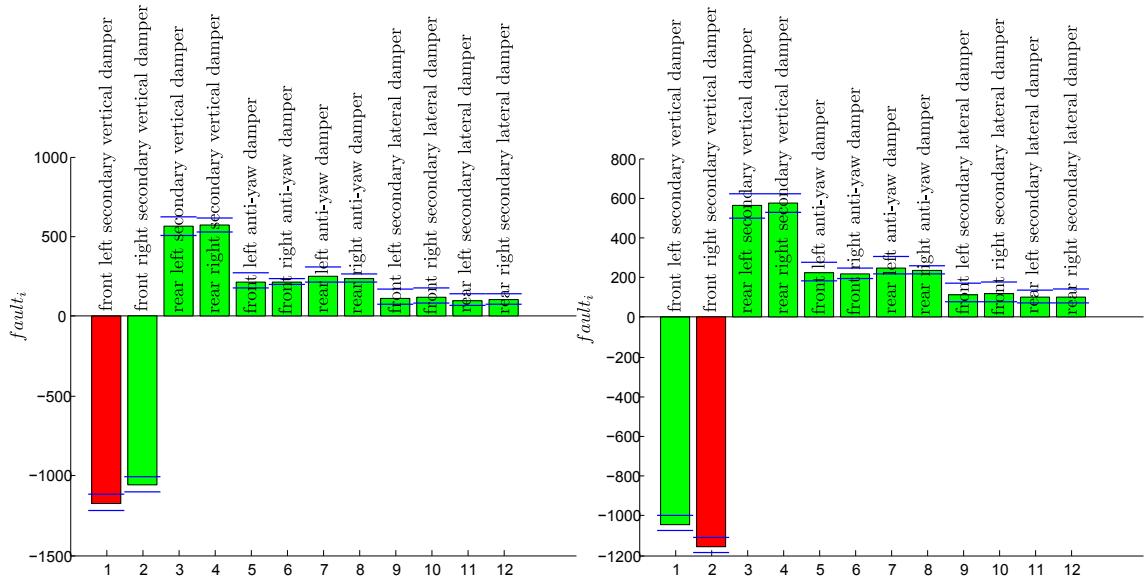
Figure A.5: Front secondary lateral damper fault



(a) at the left side (b) at the right side

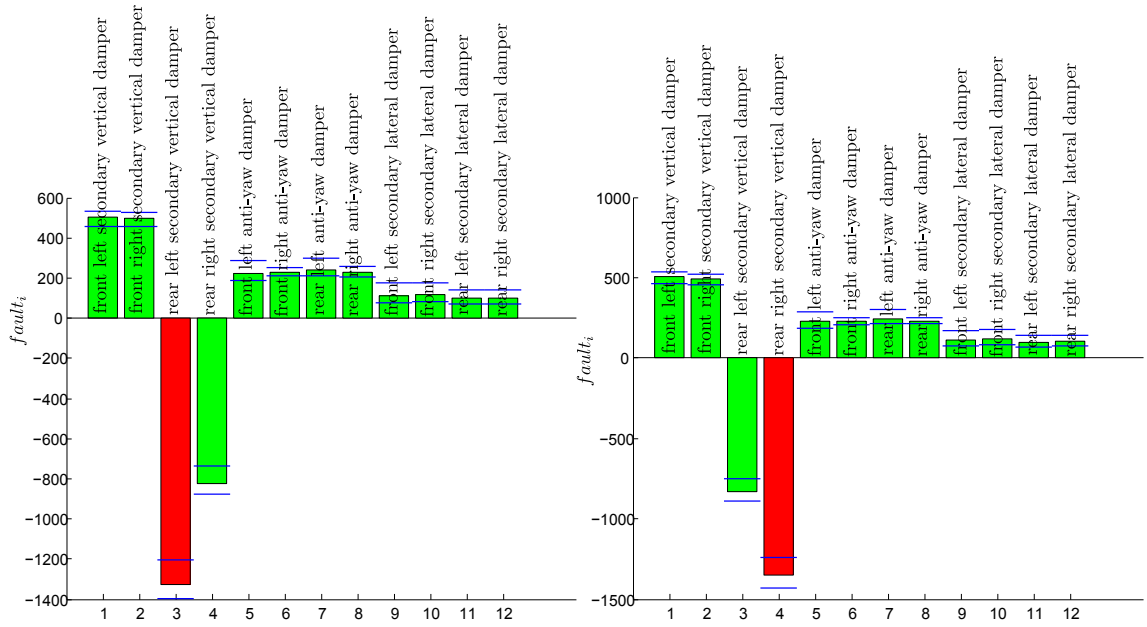
Figure A.6: Rear secondary lateral damper fault

B Results for all suspension faults for different masses



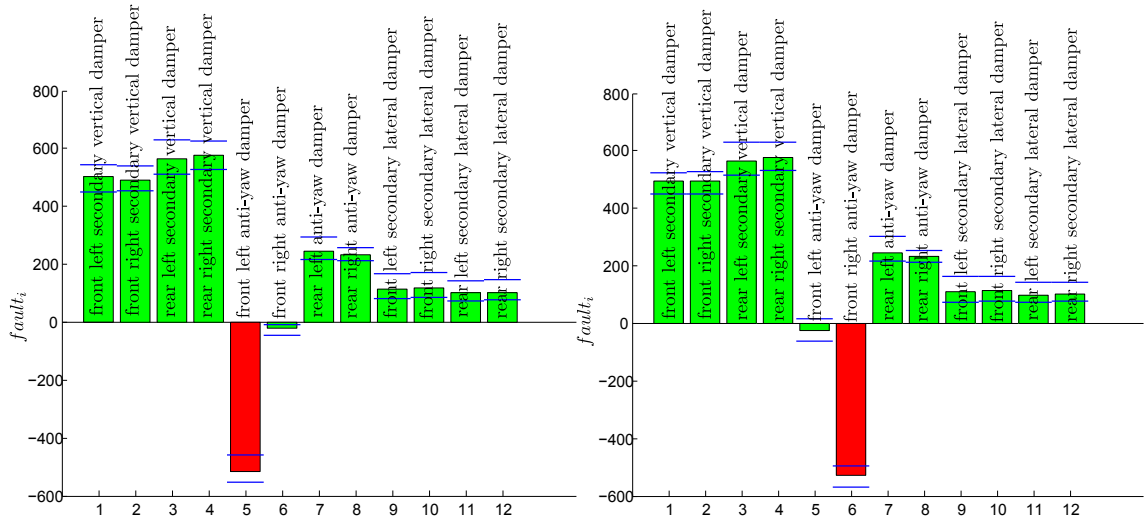
(a) at the left side (b) at the right side

Figure B.1: Front secondary vertical damper fault



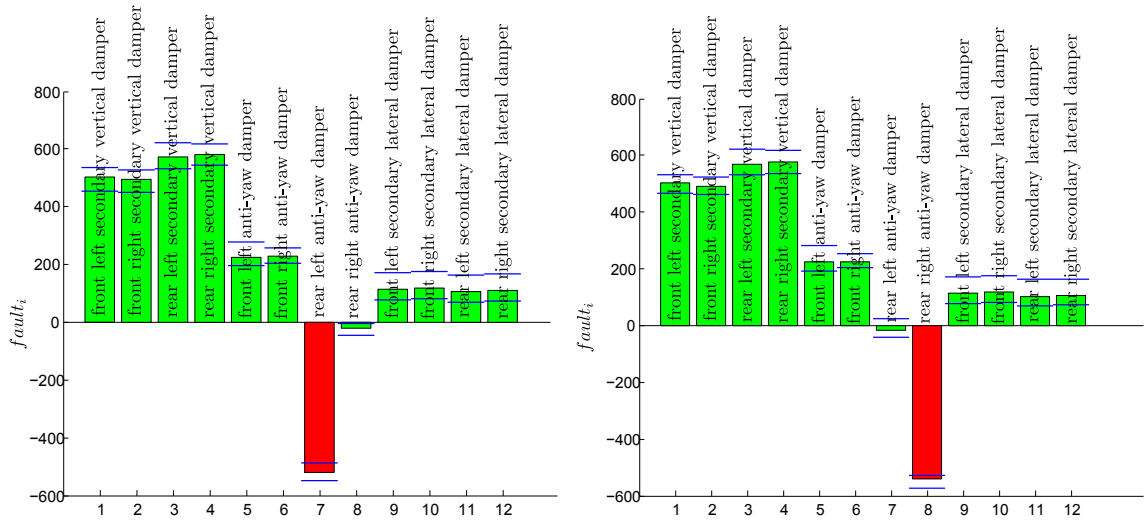
(a) left side (b) right side

Figure B.2: Rear secondary vertical damper fault



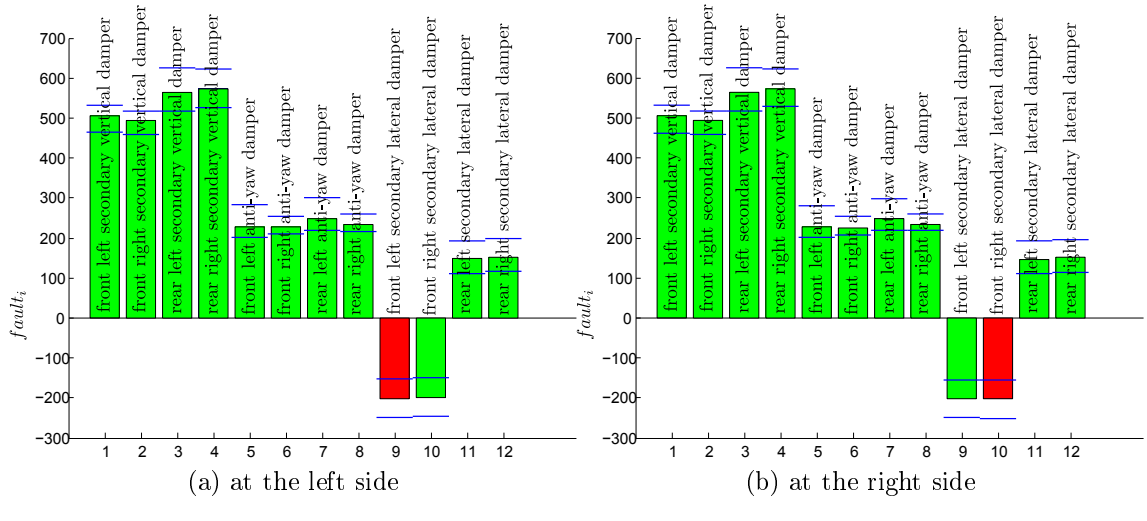
(a) at the left side (b) at the right side

Figure B.3: Front anti-yaw damper fault



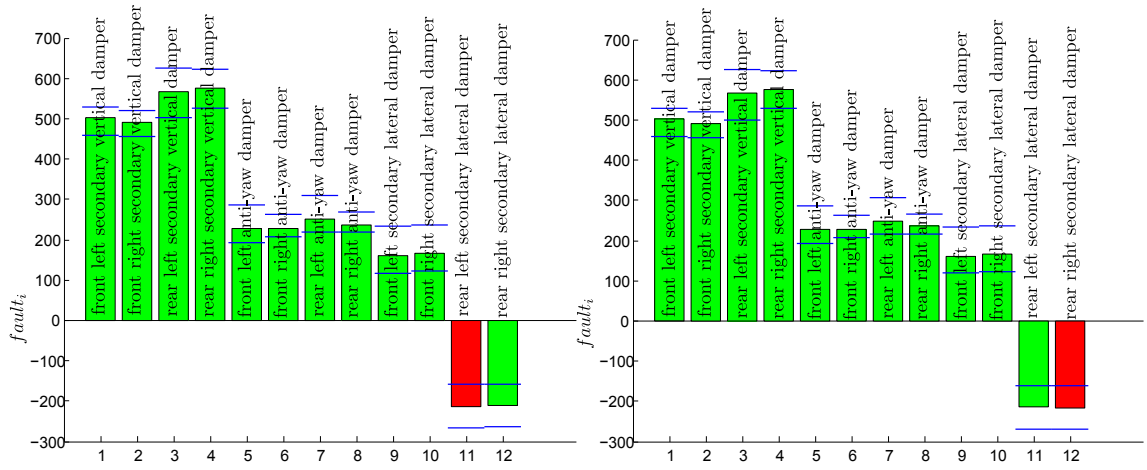
(a) at the left side (b) at the right side

Figure B.4: Rear anti-yaw damper fault



(a) at the left side (b) at the right side

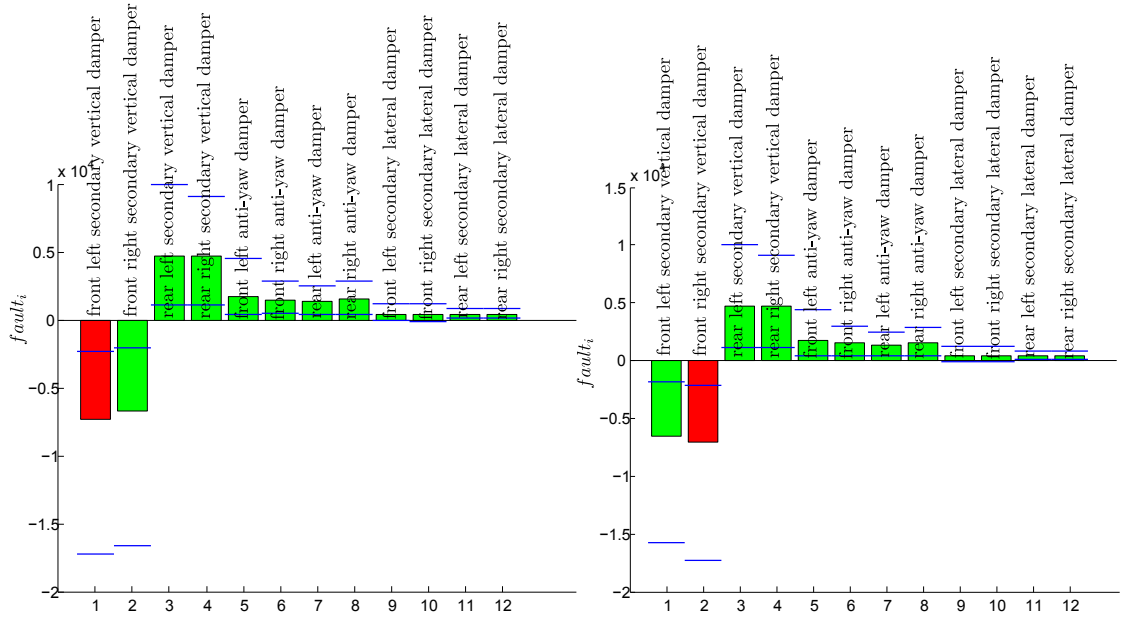
Figure B.5: Front secondary lateral damper fault



(a) at the left side (b) at the right side

Figure B.6: Rear secondary lateral damper fault

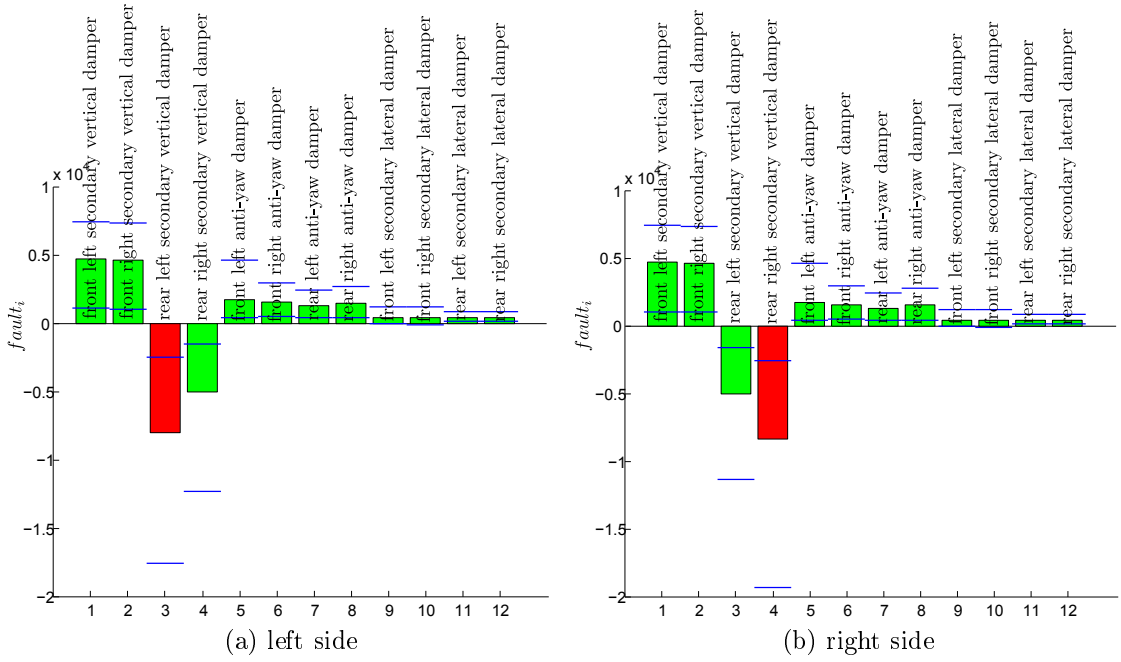
C Results for all suspension faults for different velocities



(a) at the left side

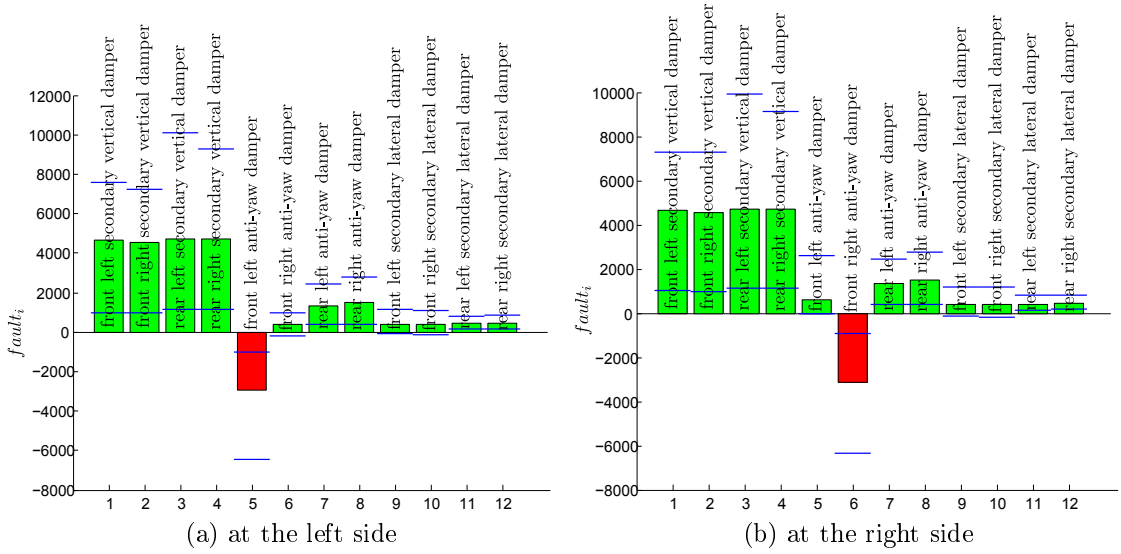
(b) at the right side

Figure C.1: Front secondary vertical damper fault



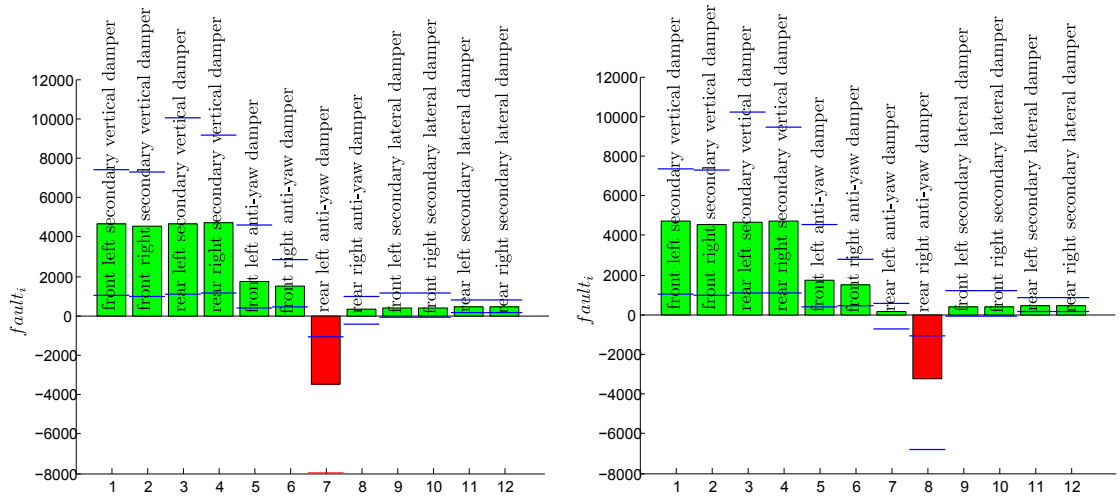
(a) left side (b) right side

Figure C.2: Rear secondary vertical damper fault



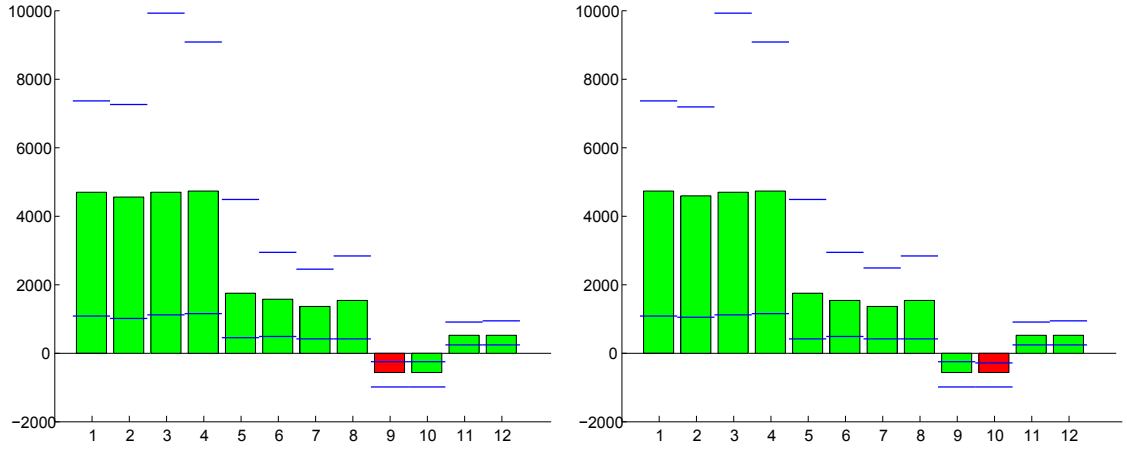
(a) at the left side (b) at the right side

Figure C.3: Front anti-yaw damper fault



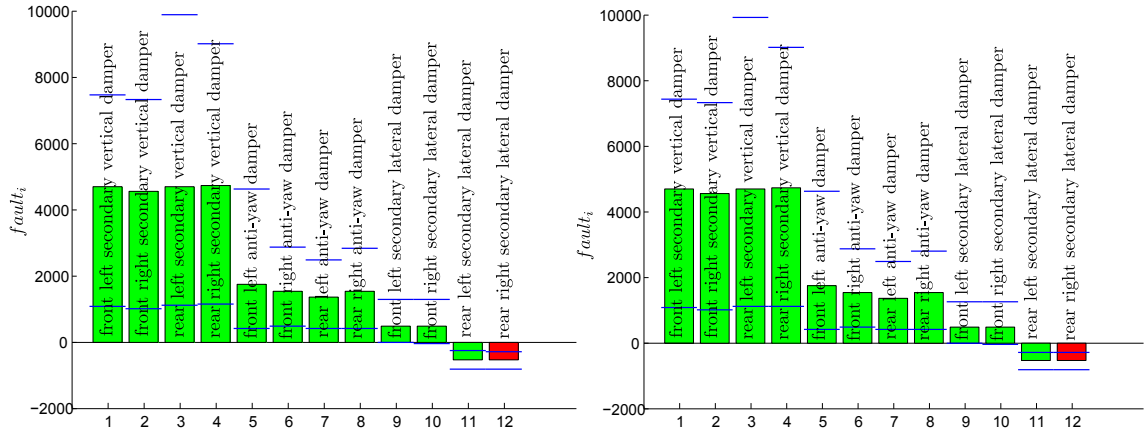
(a) at the left side (b) at the right side

Figure C.4: Rear anti-yaw damper fault



(a) at the left side (b) at the right side

Figure C.5: Front secondary lateral damper fault



(a) at the left side (b) at the right side

Figure C.6: Rear secondary lateral damper fault

List of Figures

List of Tables

Bibliography

[1] P. Li, R. Goodall, P. Weston, C. S. Ling, C. Goodman and C. Roberts. *Estimation of railway vehicle suspension parameters for condition monitoring*, Control Engineering Practice 15 (1) (2007), p. 43–55.

[2] P. Li, R. Goodall and V. Kadirkamanathan. *Estimation of parameters in a linear state space model using a Rao-Blackwellised particle filter*, Control Theory and Application 151 (6) (2004), p. 727–738.

[3] P. Li and R. Goodall. *Model-based Condition Monitoring for Railway vehicle Systems*, Control University of Bath, (2004).

[4] Y. Hayashi, T. Kojima, H. Tsunashima and Y. Marumo. *Real Time Fault Detection of Railway Vehicles and Tracks*, Railway Condition Monitoring (2006), p. 20–25.

[5] Y. Hayashi, H. Tsunashima and Y. Marumo. *Fault Detection of Railway Vehicles Using Multiple Model Approach*, SICE-ICASE (2006), p. 2812–2817.

[6] M. Börner, H. Straky, T. Weispfenning and R. Isermann. *Model based fault detection of vehicle suspension and hydraulic brake systems*, Mechatronics 12 (8) (2002), p. 999–1010.

[7] P. Metallidis, G. Verros, S. Natsiavas and C. Papadimitriou. *Fault Detection and Optimal Sensor Location in Vehicle Suspension*, Journal of Vibration and Control 9 (3-4) (2003), p. 337–359.

[8] A. N. Thite, S. Banvidi, T. Ibicek and L. Bennett. *Suspension parameter estimation in the frequency domain using a matrix inversion approach*, Vehicle System Dynamics 49 (12) (2011), p. 1803–1822.

[9] R. M. Martinod, G. R. Betancura and L. F. C. Heredia. *Identification of the technical state of suspension elements in railway systems*, Vehicle System Dynamics 50 (7) (2012), p. 1121–1135.

[10] T. X. Mei and X. J. Ding. *A model-less technique for the fault detection of rail vehicle suspensions*, Vehicle System Dynamics 46 (1) (2008), p. 277–287.

[11] T. X. Mei and X. J. Ding. *Condition monitoring of rail vehicle suspensions based on changes in system dynamic interactions*, Vehicle System Dynamics 47 (9) (2009), p. 1167–1181.

[12] X. Wei, L. Jia and H. Liu. *A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements*, Vehicle System Dynamics 51 (5) (2013), p. 700–720.

[13] M. Jesussek and K. Ellermann. *Fault detection and isolation for a nonlinear railway vehicle suspension with a Hybrid Extended Kalman filter*, Vehicle System Dynamics 51 (10) (2013), p. 1489-1501.

[14] X. Wei, S. Lin and H. Liu. *Distributed Fault Detection Observer for Rail Vehicle Suspension Systems*, Control and Decision Conference (2012), p. 3396-3401.

[15] X. Wei, Y. Guo, L. Jia and H. Liu. *Fault detection of rail vehicle suspension system based on CPCA*, Control and Fault-Tolerant Systems (2013), p. 700 - 705.

[16] R. M. Goodall and C. Roberts. *Concepts and Techniques for Railway Condition Monitoring*, Railway Condition Monitoring (2006), p. 90 - 95 .

[17] H. Ronasia and J. C. O. Nielsen. *Inverse identification of wheel–rail contact forces based on observation of wheel disc strains: an evaluation of three numerical algorithms*, Vehicle System Dynamics 51 (1) (2013), p 74-90

[18] T. Uhl. *The inverse identification problem and its technical application*, Archive of Applied Mechanics 77 (5) (2007), p. 325-337.

[19] T. Zhu, S. Xiao, G. Yang, W. Ma and Z. Zhang. *An inverse dynamics method for railway vehicle systems*, Transport 29 (1) (2014), p 107-114

[20] Z. Tao, X. Shoune, Y. Guangwu, M. Weihua and Z. Zhixin. *The inverse identification theory and application to high-speed trains*, Advanced Science Letters 19 (6) (2013), p. 1582-1586.

[21] J. C. O. Nielsen and A. Igeland. *Vertical dynamic interaction between train and track influence of wheel and track imperfections*, Journal of Sound and Vibration 187 (5) (1995), p. 825–839.

[22] J.C.O. Nielsen. *High-frequency vertical wheel–rail contact forces—Validation of a prediction model by field testing*, Wear 265 (9-10) (2008), p. 1465–1471.

[23] P. Czop, K. Mendrok and T. Uhl. *Application of inverse linear parametric models in the identification of rail track irregularities*, Archive of Applied Mechanics 81 (11) (2011), p. 1541–1554.

[24] H. Tsunashima, Y. Naganuma and T. Kobayashi. *Track geometry estimation from carbody vibration*, Vehicle System Dynamics 51 (1) (2014), p. 207–219.

[25] Y. Zhao, B. Liang and S. Iwnicki. *Friction coefficient estimation using an unscented Kalman filter*, Vehicle System Dynamics 52 (1) (2014), p. 220-234.

[26] I. Hussain and T. X. Mei *Multi Kalman Filtering Approach for Estimation of Wheel-Rail Contact Conditions*, Control 2010 (2010), p. 1–6

[27] J. Pérez, J.M. Busturia and R.M. Goodall. *Control strategies for active steering of bogie-based railway vehicles*, Control Engineering Practice 10 (9) (2002), p. 1005–1012

[28] T. David, M. Jesussek and K. Ellermann. *An approach for the estimation of wheel-rail contact forces*, Multibody System Dynamics

[29] J. Lunze. *Regelungstechnik 1*, Springer (2013)

[30] J. Lunze. *Regelungstechnik 2*, Springer (2012)

[31] Da-Wei Gu. *Robust Control Design with MATLAB*, Springer (2013)

[32] D. P. Bertsekas and J. N. Tsitsiklis. *Introduction to probability Second Edition*, Massachusetts Institute of Technology. Athena Scientific (2008)

[33] S. Dan. *Optimal State Estimation*, Wiley (2006).

[34] M. Malvezzi, E. Meli, S. Falomi and A. Rindi. *Determination of wheel–rail contact points with semianalytic methods*, Multibody System Dynamics 20 (4) (2008), p. 327–358.

[35] A. A. Shabana, K. E. Zaazaa, and H. Sugiyama. *Railroad Vehicle Dynamics: A Computational Approach*, Crc Pr Inc (2007),

[36] H. Hertz. *Über die berührung fester elastische Körper und über die Härte*, Verhandlungen des Vereins zur Beförderung des Gewerbefleisses (1882), p. 449-463

[37] A. A. Shabana and J. R. Sany. *An augmented formulation for mechanical systems with non-generalized coordinates: application to rigid body contact problems*, Nonlinear Dynamics 24 (2) (2001),pp. 183–204

[38] W. Goldsmi. *The Theory and Physical Behaviour of Colliding Solids*, Dover Publications Inc (2001)

[39] A. A. Shabana, K. E. Zaazaa, J. L. Escalona and J. R. Sanyc. *Development of elastic force model for wheel/rail contact problems*, Journal of Sound and Vibration 269 (1-2) (2004), p. 295–325.

[40] F.W Carter. *On the action of a locomotive driving wheel*, Proceedings of the Royal Society of London (1926), p. 151–157

[41] J. J. Kalker. *Survey of Wheel—Rail Rolling Contact Theory*, Vehicle System Dynamics 8 (4) (1979), p. 317-358

[42] S. Iwnicki. *Simulation of wheel–rail contact forces*, Journal of Sound and Vibration 26 (10) (2003), p. 887–900.

[43] C. Kang, H. Kim, M. Kim and B. Goo *Simulations on Creep Forces Acting on the Wheel of a Rolling Stock*, International Conference on Control, Automation and Systems (2008), p.7–12

[44] F. Frederich. *Die Gleislage - aus fahrzeugtechnischer Sicht*, ZEV-Glas Ann. 108 , (1984), pp. 355–362.

[45] D. H. Wang and W. H. Liao. *Semi-active suspension systems for railway vehicles using magnetorheological dampers. Part I: system integration and modelling*, Vehicle System Dynamics 47 (11) (2009), p. 1305–1325.

[46] *Simpack*, Simpack Documentation (11) (2012), pp. 1618.

[47] H. Schlitt. *Systemtheorie für stochastische Prozesse*, Springer Berlin Heidelberg (1992).

[48] R. F. Stengel. *Optimal control and Estimation*, Dover Pubn Inc (1994).

[49] M. S. Grewal and A. P. Andrews. *Kalman Filtering Theory and Practice Using Matlab*, John Wiley & Sons (2008).

[50] P. Stein. *Some general theorems on iterants*, Journal of Research of the National Bureau of Standards 48 (1) (1952), p. 82–83.

[51] L. Fendrich and W. Fengler. *Handbuch Eisenbahninfrastruktur*, Springer Vieweg (2014).

[52] S. Iwnicki. *Handbook of Railway Vehicle Dynamics*, Crc Pr Inc (2006).

[53] T. Telliskivi, *Simulation of wear in a rolling–sliding contact by a semi-Winkler model and the Archard's wear law*, Wear 256 (7-8) (2004), p. 817-831.

[54] K. Knothe and S. Stichel, *Schienenfahrzeugdynamik*, Springer (2003).