TU Dortmund University
Department of Statistics
Chair of Statistics with Applications
in Engineering Sciences

Master's thesis

Comparison of state-space models via K-sign depth

Author:
Anna Dustert

Supervisors:
Prof. Dr. Christine Müller
Dr. Mirko Jakubzik

22nd July 2026

Contents

1 Introduction

Electricity grids are the backbone of modern society. They ensure a reliable supply of energy to private households, industry and critical infrastructure. The dynamics and complexity of these networks are constantly increasing due to the growing integration of renewable energies and rising electricity demand. In particular, the feed-in of fluctuating sources such as wind and solar energy leads to significant fluctuations in network operations. So-called smart grids, intelligent measurement and control systems, enable near real-time monitoring of power networks (Cavus 2025). These developments make precise modelling of network behaviour possible and are necessary to ensure the stability, efficiency and security for the supply of the power networks.

This ongoing digitalisation is generating huge amounts of data containing important information about the status and consumption of the networks. This data is not free from uncertainty, because consumption fluctuations and external influences can have a significant impact on system behaviour (Wang et al. 2025). As a result, traditional statistical models reach their limits in such dynamic and multidimensional scenarios. This means that more robust methods for modelling and analysing electrical power networks, which take into account both the temporal dynamics and the uncertainties of the data, are becoming increasingly important.

For modelling the dynamics of these complex systems, such as electrical power networks, state-space models have proven to be a very effective methodology. These allow the description of system states and their development over time. In addition, these models are flexible enough to take into account a wide variety of input variables and disturbances.

The aim of this thesis is to investigate the use of K-sign depth as a criteria for comparing different state-space models. First, a simulation study is conducted to analyse the behaviour of the models under different assumptions. This includes the evaluation of residuals for correctly specified models, the impact of parameter misspecification, and the robustness against outliers.

In a second step, different K-sign depth methods are applied to different state-space models fitted to electrical power grid data. The models are compared based on their residual behaviour to assess their ability to represent the underlying dynamics of the system. The results provide insight into whether depth-based measures can serve as a robust and objective tool for evaluating state-space models.

To address this research question, chapter 2 first provides an explanation of the problem and a detailed description of the purpose of the report. After that, the data

structure and characteristics of the electrical power grid data used in the application are described.

Chapter 3 'Statistical methods', presents the statistical methods used throughout this thesis. The concepts of state-space models, the extended Kalman filter, and K-sign depth are introduced and described in detail. Additionally, the implementation of the applied methods in the software environment `R` (R Core Team 2025) is explained.

Chapter 4 presents the simulation study. The behaviour of the state-space models and their residuals is analysed under different scenarios, including correct model specification, parameter misspecification, and the presence of outliers.

The Application of the methods to real electrical power grid data is done in chapter 5. Several state-space models with different assumptions regarding state dynamics and seasonal effects are estimated and compared using K-sign depth measures.

Finally, the main findings of the thesis are summarised and discussed in Chapter 6. A list of symbols is given behind this. Additional tables and graphics are listed in the appendix.

2 Problem statement

This chapter begins with an explanation of the problem addressed in this thesis and a precise definition of its objectives. This is followed by a detailed description of the data used in this report.

2.1 Problem statement

Our electricity supply systems are becoming increasingly complex, and so are the statistical models used to describe them. The models must be able to adapt dynamically to the feed-in from renewable energy sources and distributed generation, as well as to changes in load profiles. Due to this complexity of power supply systems, state-space models are particularly well suited to this modelling task because of their general flexibility in modelling dynamics and latent state variables.

Finding a suitable model for these systems is not easy, given the wide range of possible modelling approaches. Models may vary considerably in terms of their structure, parametrisation or level of complexity. Consequently, the central question in any modelling exercise is which model best describes the underlying system relationships. Choosing the selection criterion for assessing the models quality is also

not straightforward. Classical selection criteria, such as likelihood-based methods, are frequently used; however, they often require certain distribution assumptions to be met or are sensitive to outliers or model deviations.

Due to these assumptions, which often cannot be met, the use of robust, non-parametric comparison methods is becoming increasingly important. K-sign depth offers such a promising approach. It enables observations to be evaluated in terms of their centrality within a distribution without the need to satisfy restrictive model assumptions. This makes this method particularly suitable for complex time-series data, such as that typically found in electricity supply systems.

2.2 Purpose of the report

This thesis focuses on comparing various state-space models and investigating whether, and if so which, depth values are particularly well-suited to evaluating their performance. It analyses the extent to which the K-sign depth can be used as a suitable and robust criteria for comparing state-space models. In addition to this, the question of which depth values are better suited to determining model quality is also a key focus.

To investigate these issues, both simulated datasets and data from a real power supply network are examined. The real data from a power supply network ensures a direct practical relevance and enables the method to be tested under realistic conditions. Furthermore, as the method of depth values itself is to be evaluated, simulated data is used for this purpose. The simulated data offers the advantage that the underlying model structures are known and individual influencing factors can be specifically varied. Therefore, it is possible to systematically investigate how different model assumptions and parameter specifications affect the resulting depth values. In addition the consequences of misspecified parameters and the effect of outliers on the models and the depth values are analysed.

Another component of this work is the analysis of various state-space modes. Therefore, different models are considered for the simulation study. After conducting the simulation study and learning more about the parameters, different state-space models are fit to the data from a real power supply network. The purpose of this application is to find out which models are best suited for a real life application and how well depth values perform on a real life setting.

The results of this work are intended to help better understand the applicability of K-sign depth in the context of state-space models and to evaluate the practical application of both the state-space models and the depth values to this kind of data.

2.3 Data origin and structure

The data used in this work contains time-dependent active power values P and reactive power values Q at nodes of an electrical power supply network. The data is based on the SimBench dataset, specifically the network *1-LV-rural1–1-no_sw*, which is part of a publicly available benchmark collection of realistic power system models (Meinecke et al. 2020). SimBench provides standardised network and time series data for the development, evaluation, and comparison of methods in power systems research (Meinecke et al. 2020, Sec. 2, pp. 3–5).

The active power P describes the proportion of electrical power that is actually converted into usable energy. This power is directly linked to energy consumption or energy feed-in. Active power is therefore a key variable for the energy balance of the network. The unit of active power is the watt and is usually expressed in kilowatts or megawatts in network models (Meinecke et al. 2020, Sec. 3.1.4, p. 12 & Sec. 4.3.1, Tab. 4.29, p. 103).

The reactive power Q describes the proportion of power required for the periodic build-up of electric and magnetic fields in operating equipment. This power is essential for stable grid operation, as it influences the voltage level and voltage distribution in the grid. The unit of this quantity is given in volt-amperes reactive (Meinecke et al. 2020, Sec. 3.1.4, p. 12 & Sec. 4.3.1, Tab. 4.29, p. 103).

The power data is defined in relation to nodes. This means that for each network node, a pair consisting of active power $P_n(t)$ and reactive power $Q_n(t)$ is available at any given time. These variables describe the power fed into or withdrawn from the network at the respective node. Positive values of P and Q correspond to power consumption by consumers, while negative values indicate power being fed into the grid (Meinecke et al. 2020, Sec. 4.1, pp. 76–78).

Overall, the power was observed at fourteen nodes. The nodes are arranged in the network as shown in figure 1. The data is a time series. Over a period of one year, 2016, the active and reactive power at each node was measured every fifteen minutes (Meinecke et al. 2020, Sec. 3.5, pp. 31–37). The physical power values are normalised to a defined nominal power of the respective network (Meinecke et al. 2020, Sec. 4.1, pp. 76–79).

The per-unit representation reduces the numerical range of the values and facilitates both numerical stability and comparability between different network sizes and voltage levels. A per-unit value of $P = 1.0$ corresponds to the respective reference power, while smaller or larger values represent proportional powers (Meinecke et al. 2020, Sec. 4.1, p. 78).

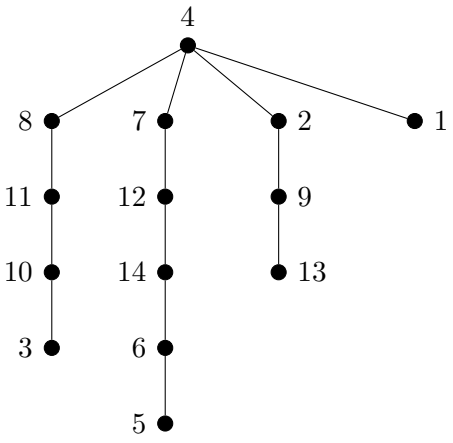

Figure 1: Schematic representation of the electrical network.

3 Statistical methods

This chapter introduces the statistical methods used in this work. First, a general state-space model with load profiles is defined. Afterwards, the Kalman filter for a nonlinear link function is introduced. To evaluate these models, the Normalized Estimation Error Squared and Normalized Innovation Squared are explained. Next, the Bonferroni correction, which will be used for the tests, is defined as well as the used depth methods and their corresponding tests. Following that, the Shapiro-Wilk test is explained. Lastly, the `R` packages used for the implementation are listed.

3.1 General state-space model

The starting point is the formulation of a state-space model. This chapter is based on the paper by Müller (2026), the contents of which are reproduced here in an adapted form. Let $(V_t)_{t\in[0,T]}$ denote the $2N$-dimensional state process, where each state vector collects the real and imaginary parts of the grid voltages

$$V_t = \Big[e_1, \quad f_1, \quad e_2, \quad f_2, \quad \dots \quad e_N, \quad f_N\Big]^\top.$$

Here, e_n and f_n correspond to the real and imaginary components of the voltage at node n, with the slack node excluded from the state representation.

The observation process $(Z_t)_{t\in[0,T]}$ is also $2N$-dimensional and consists of the active and reactive power injections at each node of the grid,

$$Z_t = \Big[P_1, \quad Q_1, \quad \dots \quad P_N, \quad Q_N\Big]^\top.$$

Model equations.
The temporal evolution of the state and its relation to the measurements is captured by the following general state-space system (Müller 2026):

$$V_t = \Theta_{v,t} V_{t-1} + \Theta_{w,t} w_t + e_t^v,$$
$$Z_t = h_t(V_t) + e_t^z.$$

Here:

- $\Theta_{v,t} \in \mathbb{R}^{2N \times 2N}$ describes the autoregressive effect of past states,
- $\Theta_{w,t} \in \mathbb{R}^{2N \times 2N}$ accounts for the effect of the explanatory variables,
- $w_t \in \mathbb{R}^N$ represents a vector of explanatory (exogenous) variables given by load profiles, such as load forecasts or weather conditions,
- e_t^v and e_t^z are independent and identically distributed Gaussian noise terms, independent of past states and observations,
- $h_t : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ is a nonlinear function linking the latent voltages to the observed powers.

3.2 Load profiles

Seasonal load profiles with sinusoidal modelling.
To model seasonal effects, smoothed load profiles are constructed using a sinusoidal basis function. Let $d(t)$ be the calendar day and $w(t)$ the day of the week at time t. For each hour h and each node n, the following regression model is assumed:

$$P_{n,d,h} = \mathbf{x}(d)^\top \beta^P_{n,w(d),h} + e^P_{n,d,h},$$
$$Q_{n,d,h} = \mathbf{x}(d)^\top \beta^Q_{n,w(d),h} + e^Q_{n,d,h},$$

with

$$\mathbf{x}(d) = \begin{pmatrix} 1 \\ \sin\left(\frac{d-d_0}{365} \cdot 2\pi\right) \end{pmatrix},$$

where d_0 represents the beginning of spring (e.g., March 21, then $d_0 = 80$).

The regression parameters are estimated using least squares:

$$\hat{\beta}_{n,w,h} = (\mathbf{X}_w^\top \mathbf{X}_w)^{-1} \mathbf{X}_w^\top \mathbf{Y}_{n,w,h}.$$

The estimated smoothed load profiles are then

$$\hat{P}_{n,d,h} = \mathbf{x}(d)^\top \hat{\beta}^P_{n,w(d),h},$$
$$\hat{Q}_{n,d,h} = \mathbf{x}(d)^\top \hat{\beta}^Q_{n,w(d),h}.$$

The resulting smoothed power vector is

$$\hat{Z}_t = (\hat{P}_{1,d(t),h(t)}, \hat{Q}_{1,d(t),h(t)}, \ldots, \hat{P}_{N,d(t),h(t)}, \hat{Q}_{N,d(t),h(t)})^\top.$$

Derivation of the link function.
A direct inversion of the nonlinear function h is generally not unique. Therefore, a first-order Taylor approximation is used. For small changes, the following approximation holds:

$$h(V_t) \approx h(V_{t-1}) + H(V_{t-1})(V_t - V_{t-1}),$$

where

$$H(V_{t-1}) = \left.\frac{\partial h(v)}{\partial v}\right|_{v=V_{t-1}}$$

denotes the Jacobian matrix of h.

This results in the input variable

$$w_t := H(V_{t-1})^{-1}(\hat{Z}_t - \hat{Z}_{t-1}).$$

This definition allows structured load information to be incorporated into the state model without having to perform an explicit inversion of h.

3.3 Kalman filter for a nonlinear link function

In modern power systems, the accurate estimation of internal states such as grid voltages is of fundamental importance for both operational security and economic efficiency. Since measurements are often noisy, incomplete, and nonlinear in relation to the underlying states, advanced statistical filtering methods are required. Among these, the family of Kalman filters provides a systematic approach to recursively estimate the latent state of a dynamic system given a sequence of observations.

To estimate the hidden states a nonlinear Kalman filter is applied. The formulas in this subchapter are taken from the book *Dynamic Models with R* by Petris, Petrone, and Campagnoli 2009, pp. 53-57 and the paper *Optimal Designs for the Kalman Filter Applied to Electrical Power Distribution Grid* by Müller 2026. The

approach relies on linearization of the nonlinear measurement function $h(v)$ around the predicted state. The noise assumptions are:

$$e_t^v \sim \mathcal{N}(0, \Sigma_t^v), \quad V_0 \sim \mathcal{N}(c_0, C_0),$$

with $\Sigma_t^v = \sigma_v^2 \mathbb{I}_{2N\times 2N}$ and $\Sigma_t^z = \sigma_z^2 \mathbb{I}_{2N\times 2N}$. The initial observation is given by $z_0 = h(c_0)$.

Prediction step. The one-step-ahead prediction for the latent state follows:

$$V_t | z_{1:t-1} \sim \mathcal{N}_{2N}(a_t, R_t),$$

with

$$\begin{aligned} a_t &= \mathbb{E}[V_t | z_{1:t-1}] = \Theta_{v,t} c_{t-1} + \Theta_{w,t} w_t, \\ R_t &= \mathrm{Cov}(V_t | z_{1:t-1}) = \Theta_{v,t} C_{t-1} \Theta_{v,t}^\top + \Sigma_t^v. \end{aligned}$$

Observation prediction. The predicted distribution of the observations is

$$Z_t | z_{1:t-1} \sim \mathcal{N}_{2N}(b_t, S_t).$$

Since h is nonlinear, we approximate using a first-order Taylor expansion:

$$\begin{aligned} b_t &= \mathbb{E}[Z_t | z_{1:t-1}] \approx h_t(a_t), \\ H_t &= \left. \frac{\partial h_t}{\partial v} \right|_{v=a_t}, \\ S_t &= H_t R_t H_t^\top + \Sigma_t^z, \end{aligned}$$

where H_t denotes the Jacobian of h_t.

Update step. The update incorporates the new measurement z_t to refine the state estimate:

$$\begin{aligned} K_t &= R_t H_t^\top S_t^{-1}, \quad \text{(Kalman gain)} \\ c_t &= a_t + K_t (z_t - b_t), \\ C_t &= (\mathbb{I}_{2N\times 2N} - K_t H_t) R_t. \end{aligned}$$

The Kalman gain balances model predictions with the new information provided by the observations.

Nonlinear function and its derivative.
The function $h(v)$ maps voltages to power injections. For each node $n = 1, \ldots, N$, the complex power is

$$S_n = v_n \cdot \sum_{k=1}^{N+1} a_{nk}^* \cdot v_k^*,$$

where $v_n = e_n + jf_n$ is the complex voltage, a_{nk} denotes the admittance between nodes n and k, and the summation includes the slack node (Müller 2026).

The observation function is defined as

$$h(x) = \begin{bmatrix} Re(S_1) \\ Im(S_1) \\ \vdots \\ Re(S_N) \\ Im(S_N) \end{bmatrix} \in \mathbb{R}^{2N}.$$

Real-Valued Representation.
To avoid complex numbers in the filter equations, the model is expressed in real-valued form (Müller 2026). The multiplication by v_n is represented by the matrix

$$U_n = \begin{bmatrix} e_n & f_n \\ -f_n & e_n \end{bmatrix}, \quad \overline{U}_n = \begin{bmatrix} e_n & f_n \\ f_n & -e_n \end{bmatrix}.$$

For the admittance coefficients, we define

$$\overline{A}_{nk} = \begin{bmatrix} Re(\overline{a}_{nk}) & Im(\overline{a}_{nk}) \\ -Im(\overline{a}_{nk}) & Re(\overline{a}_{nk}) \end{bmatrix} = \begin{bmatrix} g_{nk} & b_{nk} \\ -b_{nk} & g_{nk} \end{bmatrix},$$

where $a_{nk} = g_{nk} + jb_{nk}$. Using these matrices, matrix $\overline{A}$ is formed as follows:

$$\overline{A} := (\overline{A}_{nk})_{n,k=1,\ldots,N} \in \mathbb{R}^{2N \times 2N}.$$

The slack node enters as a fixed-voltage contribution, collected in additional blocks

$$\overline{A}_{n,\text{slack}} = \begin{bmatrix} Re(\overline{a}_{n,N+1}) & Im(\overline{a}_{n,N+1}) \\ -Im(\overline{a}_{n,N+1}) & Re(\overline{a}_{n,N+1}) \end{bmatrix}.$$

The Jacobian matrix.
The Jacobian matrix of link function $h(v)$ captures the sensitivity of the power injections with respect to the voltage components (Müller 2026):

$$
\begin{aligned}
H(v) &= \frac{\partial h}{\partial v} \\
&= \begin{bmatrix}
\overline{U}_1\,\overline{A}_{11}^{\top} + \sum_{k=1}^{N+1} U_k\,\overline{A}_{1k} & \overline{U}_1\,\overline{A}_{12}^{\top} & \cdots & \overline{U}_1\,\overline{A}_{1N}^{\top} \\
\overline{U}_2\,\overline{A}_{12}^{\top} & \overline{U}_2\,\overline{A}_{22}^{\top} + \sum_{k=1}^{N+1} U_k\,\overline{A}_{2k} & \cdots & \overline{U}_2\,\overline{A}_{2N}^{\top} \\
\vdots & \vdots & \ddots & \vdots \\
\overline{U}_N\,\overline{A}_{1N}^{\top} & \overline{U}_N\,\overline{A}_{2N}^{\top} & \cdots & \overline{U}_N\,\overline{A}_{NN}^{\top} + \sum_{k=1}^{N+1} U_k\,\overline{A}_{Nk}
\end{bmatrix} \\
&= \operatorname{Diag}(\overline{U}_1, \ldots, \overline{U}_N)\,\overline{A}^{\top} + \operatorname{Diag}(U_1, \ldots, U_N)\,\overline{A} + \operatorname{Diag}(\overline{A}_{1(N+1)}, \ldots, \overline{A}_{N(N+1)}).
\end{aligned}
$$

Filter log-likelihood and its numerical implementation.
The log-likelihood of a state-space model evaluated via the Kalman filter can be written using the prediction error decomposition (Durbin and Koopman 2012, p. 170 f.). Let r_t denote the one-step-ahead prediction error (innovation) and S_t its covariance matrix. Ignoring the additive constant term $\frac{m}{2}\log(2\pi)$, which does not depend on the model parameters, the negative log-likelihood is

$$
-\ell(\theta) = \frac{1}{2}\sum_{t=1}^{T}\left(\log\det(S_t) + r_t^{\top} S_t^{-1} r_t\right).
$$

To reduce the computational effort of this formula, $\det(S_t)$ and S_t^{-1} are calculated using the Cholesky decomposition

$$
S_t = L_t L_t^{\top},
$$

where L_t is a lower triangular matrix (Horn and Johnson 2012, p. 441). The covariance matrices are positive definite, which means that the prerequisite for Cholesky decomposition is fulfilled (tested by the query `chol(S_t)` in the code). This decomposition allows both terms of the likelihood contribution to be computed in a numerically stable way.

Since $\det(S_t) = \prod_{i=1}^{m}(L_{t,ii})^2$, the logarithm of the determinant simplifies to

$$\log\det(S_t) = 2\sum_{i=1}^{m}\log(L_{t,ii}).$$

The Mahalanobis distance $r_t^\top S_t^{-1} r_t$ can be computed without explicitly inverting S_t. With the help from the Cholesky factorization it can be simplified with

$$S_t^{-1} = (L_t^\top)^{-1} L_t^{-1},$$

to

$$r_t^\top S_t^{-1} r_t = (L_t^{-1} r_t)^\top (L_t^{-1} r_t) = \|y_t\|^2,$$

where y_t solves the triangular linear system

$$L_t y_t = r_t.$$

Combining both parts yields the single-step contribution to the negative log-likelihood:

$$-\ell(\theta) = \frac{1}{2}\left(\log\det(S_t) + r_t^\top S_t^{-1} r_t\right).$$

This formulation is mathematically equivalent to the expression in Durbin and Koopman 2012 p. 170-171, but is substantially more numerically stable because it avoids explicit matrix inversion and exploits the properties of the Cholesky factorization.

3.4 Normalized Estimation Error Squared and Normalized Innovation Squared

To evaluate the statistical consistency of a Kalman filter, the Normalized Estimation Error Squared (NEES) and the Normalized Innovation Squared (NIS) are commonly used metrics. Both measures test whether the estimated covariance correctly reflects the true estimation uncertainty.

Normalized Estimation Error Squared

The NEES evaluates the consistency of the state estimate by comparing the estimation error with the predicted state covariance. It is primarily used in simulation studies, where the true system state is available, to assess whether the filter uncertainty is statistically consistent.

Let $\hat{v}(k)$ denote the state estimate and $v(k)$ the true state. The estimation error is defined as

$$\tilde{v}(k) = v(k) - \hat{v}(k).$$

The NEES is then given by

$$\epsilon_v(k) = \tilde{v}(k)^\top P^{-1}(k)\tilde{v}(k),$$

where $P(k)$ is the state estimation error covariance matrix (Bar-Shalom, Li, and Kirubarajan 2001, p. 234).

Under the assumption of a correctly specified linear Gaussian model, $\epsilon_v(k)$ follows a chi-square distribution with n_v degrees of freedom:

$$\epsilon_v(k) \sim \chi^2(n_v),$$

where n_v is the dimension of the state vector (Bar-Shalom, Li, and Kirubarajan 2001, p. 234).

To check whether a state estimator provides a consistent result, the expectation of $\epsilon(k)$ can be used

$$\mathbb{E}(\epsilon(k)) = n_v.$$

The NEES is evaluated using a two-sided hypothesis test. The null hypothesis states that the estimated covariance is consistent with the true estimation error,

$$\begin{aligned} H_0: \quad & \text{The estimation error } \hat{v}(k) \\ & \text{is consistent with the filter covariance matrix } P(k). \end{aligned}$$

The null hypothesis is accepted if

$$b_1 \leq \epsilon_v(k) \leq b_2,$$

where b_1 and b_2 denote the lower and upper confidence bounds of the χ^2 distribution. For a significance level α, the confidence interval is computed as

$$r_1 = F^{-1}\left(\frac{\alpha}{2}, n_v\right), \qquad r_2 = F^{-1}\left(1 - \frac{\alpha}{2}, n_v\right),$$

where $F^{-1}(\cdot, n_v)$ is the inverse cumulative distribution function of the chi-square distribution with n_v degrees of freedom (Bar-Shalom, Li, and Kirubarajan 2001, p. 235).

In practice, the NEES is typically evaluated by plotting $\epsilon_v(k)$ together with the corresponding confidence interval. Values outside the interval indicate that the filter is statistically inconsistent. A NEES above the upper bound suggests that the estimation covariance is underestimated, whereas values below the lower bound indicate an overly conservative covariance estimate.

Normalized Innovation Squared

The NIS evaluates the consistency of the innovation by comparing the measurement residual with the predicted innovation covariance. Since it only requires measurement data, it can be applied to both simulated and real-world systems where the true state is unknown.

Let $z(k)$ be the measurement and $\hat{z}(k)$ the predicted measurement. The innovation is defined as

$$\nu(k) = z(k) - \hat{z}(k).$$

The innovation covariance is

$$S(k) = H(k)P(k|k-1)H^\top(k) + R(k),$$

where $H(k)$ is the measurement matrix and $R(k)$ the measurement noise covariance (Bar-Shalom, Li, and Kirubarajan 2001, p. 236).

The NIS is defined as

$$\epsilon_\nu(k) = \nu(k)^\top S^{-1}(k)\nu(k).$$

Under correct modelling assumptions, the NIS follows a chi-square distribution with n_z degrees of freedom:

$$\epsilon_\nu(k) \sim \chi^2(n_z),$$

where n_z is the dimension of the measurement vector (Bar-Shalom, Li, and Kirubarajan 2001, p. 236).

The NIS is used to determine whether the innovation is statistically consistent with the predicted measurement uncertainty. The null hypothesis assumes that the innovation follows the expected probability distribution defined by the innovation covariance,

H_0 : The measurement residual $\nu(k)$
is consistent with the innovation covariance matrix $S(k)$.

The hypothesis is accepted if the NIS value lies within the confidence interval

$$b_1 \leq \epsilon_\nu(k) \leq b_2,$$

where the bounds are obtained from the inverse cumulative distribution function of the chi-square distribution,

$$b_1 = F^{-1}\left(\frac{\alpha}{2}, n_z\right), \qquad b_2 = F^{-1}\left(1 - \frac{\alpha}{2}, n_z\right),$$

with significance level α and measurement dimension n_z (Bar-Shalom, Li, and Kirubarajan 2001, p. 236).

The test is typically performed by comparing the NIS values with the corresponding confidence interval over time. Persistent violations of the confidence bounds indicate that the assumed measurement uncertainty does not adequately describe the observed innovations, suggesting an inconsistent filter or inaccurate measurement noise model.

For the NIS application, the measurement vector has dimension $n_z = 6$. Therefore, the expected value of the NIS is 6. For a significance level of $\alpha = 0.05$, the corresponding 95% confidence interval is approximately [1.24, 14.45]. These values are used later to evaluate the considered state-space models.

3.5 Bonferroni correction

If several hypothesis tests are performed simultaneously, the probability of incorrectly rejecting at least one true null hypothesis increases. Whilst in a single test the significance level α controls the probability of a type I error, this probability can rise significantly when multiple tests are performed. Of particular interest here is the so-called family-wise error rate (FWER), i.e. the probability of committing at least one type I error within a family of tests (Goeman and Solari 2014).

The Bonferroni correction is a simple method for controlling the FWER. Its theoretical basis is the Bonferroni inequality. Let $E_1, \ldots, E_m$ be the events that a type I error occurs in each of the m tests carried out. If each individual test is conducted at a significance level of α/m, the following applies:

$$\mathbb{P}\left(\bigcup_{i=1}^{m} E_i\right) \leq \sum_{i=1}^{m} \mathbb{P}(E_i) \leq \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha.$$

The Bonferroni correction therefore controls the global significance level independently of any potential dependencies between the individual tests.

In practice, the correction is applied by using the adjusted significance level

$$\alpha_{\text{Bonf}} = \frac{\alpha}{m}$$

for each of the m tests. The Bonferroni correction is characterised by its simplicity and its strict control of the Type I error α (Goeman and Solari 2014).

In this study, the Bonferroni correction is used to control the global significance level when conducting multiple tests simultaneously due to the multidimensional nature of the data.

In the following tests, a quantile level of $\alpha = 0.05$ is used. In addition, an adjusted significance level is considered using the Bonferroni correction. This is $0.05/6 = 0.0083$ for univariate tests and $0.05/3 = 0.0167$ for bivariate tests.

3.6 Sign-based depth

When analysing the adequacy of a model through univariate residuals $r_1, \ldots, r_N \in \mathbb{R}$, one possibility is to evaluate how often their signs alternate in subsets of fixed length K. This idea was formalised by Malcherczyk, Leckey, and Müller 2021 in the form of the K-sign depth. The basic idea behind K-sign depth is that the signs are examined in tuples to check whether they alternate. The various K-sign depth values are defined and explained below. The text also explains how these depth values can be calculated in `R` (R Core Team 2025). The `SignDepth` package is primarily used for these calculations in `R` (Nagy 2026).

3.6.1 Full K-sign depth

The full K-sign depth is defined as the proportion of K-triples with alternating signs:

$$d_K^{u,F}(r_1, \ldots, r_N) = \binom{N}{K}^{-1} \sum_{1 \le n_1 < \cdots < n_K \le N} \left[\prod_{k=1}^{K} \mathbb{1}\Big((-1)^k r_{n_k} > 0\Big) + \prod_{k=1}^{K} \mathbb{1}\Big((-1)^k r_{n_k} < 0\Big) \right],$$

where $\mathbb{1}(\cdot)$ denotes the indicator function (Malcherczyk, Leckey, and Müller 2021).

For $K \geq 2$, it is defined as the proportion of all $\binom{N}{K}$ possible K-tuples of indices for which the signs of the corresponding residuals alternate. The two products correspond to the two possible alternating sign patterns: starting with a positive or starting with a negative residual.

This is computed via the `KSign`-function of the package `SignDepth` with `r` as the sign vector of the residuals, `K` as the parameter of the depth and `N` as the number of residuals:

```
sum(KSign(r, k = K)[K,]) / choose(N,K)
```

The quantile levels were determined by simulation and are listed in table 1. To approximate the quantiles, $M = 500000$ simulations were performed, each consisting of $N = 1\,000$ independent random samples.

Quantile level	$K = 3$	$K = 4$
0.8333%	0.2477	0.1233
1%	0.2478	0.1234
2%	0.2482	0.1237
5%	0.2488	0.1241
10%	0.2492	0.1244

Table 1: Simulated quantiles of the K-sign depth for $K = 3$ and $K = 4$ (Simulations: $M = 500\,000$, Sample size within a single simulation: $N = 1\,000$).

Standardised full K-sign depth

To obtain the standardised full K-sign depth, the full K-sign depth is centred and scaled (Malcherczyk, Leckey, and Müller 2021):

$$D_K^{u,F}(r_1, \ldots, r_N) = N \left(d_K^{u,F}(r_1, \ldots, r_N) - \left(\frac{1}{2}\right)^{K-1} \right).$$

This depth value can be calculated in `R` via

```
N * (sum(KSign(r, k=K)[K,]) / choose(N,K) - (0.5)^(K-1))
```

Table 2 lists the simulated quantiles of the standardised K-sign depth. The quantiles given are based on $M = 500\,000$ independent simulation runs, each consisting of $N = 1\,000$ observations.

Quantile level	$K = 3$	$K = 4$
0.8333%	−2.3065	−1.7199
1%	−2.1957	−1.6398
2%	−1.7769	−1.3306
5%	−1.2196	−0.9234
10%	−0.7984	−0.6190

Table 2: Simulated quantiles of the standardised K-sign depth for $K = 3$ and $K = 4$ (Simulations: $M = 500\,000$, Sample size within a single simulation: $N = 1\,000$).

3.6.2 Simplified K-sign depth

Checking all $\binom{N}{K}$ subsets can be computationally demanding for larger data sets. To address this, Kustosz, Müller, and Wendler 2016 proposed a simplified variant

which restricts attention to consecutive blocks of length K. The resulting simplified K-sign depth is given by

$$d_K^{u,S}(r_1,\ldots,r_N) = \frac{1}{N-K+1}\sum_{n=1}^{N-K+1}\left[\prod_{k=1}^{K}\mathbb{1}\Big((-1)^k r_{n+k-1}>0\Big)+\prod_{k=1}^{K}\mathbb{1}\Big((-1)^k r_{n+k-1}<0\Big)\right].$$

Since this version only evaluates sliding windows of length K, it is far more efficient for long residual sequences, while still capturing the essential alternation behaviour that motivates the full depth.

This depth value is looked at later on for $K = 3$ and $K = 4$. The simplified 3-sign depth can be calculated as follows:

```
simple_3depth <- function(residuals){
  N <- length(residuals)
  pattern_count <- sum(sapply(1:(N-2), function(i){
    block <- sign(residuals[i:(i+2)])
    # check (+,-,+) or (-,+,-)
    all(block == c(1,-1,1)) || all(block == c(-1,1,-1))
  }))
  return(pattern_count / (N - 2))
}
```

The following function calculates the simplified 4-sign depth:

```
simple_4depth <- function(residuals){
  N <- length(residuals)
  pattern_count <- sum(sapply(1:(N-3), function(i){
    block <- sign(residuals[i:(i+3)])
    # check (+,-,+,-) or (-,+,-,+)
    all(block == c(1,-1,1,-1)) || all(block == c(-1,1,-1,1))
  }))
  return(pattern_count / (N - 3))
}
```

The simulated quantiles of this depth value can be found in table 3. The quantiles were determined on the basis of $M = 500\,000$ simulations with a sample size of $N = 1\,000$.

Quantile level	$K = 3$	$K = 4$
0.8333%	0.2084	0.0903
1%	0.2094	0.0913
2%	0.2144	0.0943
5%	0.2214	0.1003
10%	0.2275	0.1053

Table 3: Simulated quantiles of the simplified 3- and 4-sign depth (Simulations: $M = 500\,000$, Sample size within a single simulation: $N = 1\,000$).

Normalized simplified K-sign depth

The normalized version of the simplified K-sign depth is introduced as well in Kustosz, Müller, and Wendler 2016 and is given by

$$D_K^{u,S}(r_1, \ldots, r_N) = \sqrt{N - (K-1)} \frac{(d_K^{u,S}(r_1, \ldots, r_N) - (\frac{1}{2})^{K-1})}{\sqrt{(\frac{1}{2})^{K-1} \cdot \left(3 - (\frac{1}{2})^{K-2} \cdot K - 3 \cdot (\frac{1}{2})^{K-1}\right)}}.$$

In this thesis $K = 3$ and $K = 4$ is looked at, which gives the following formulas:

$$D_3^{u,S}(r_1, \ldots, r_N) = \sqrt{N-2} \frac{(d_3^{u,S}(r_1, \ldots, r_N) - 0.25)}{0.25\sqrt{5}} \to \mathcal{N}(0,1),$$

$$D_4^{u,S}(r_1, \ldots, r_N) = \sqrt{N-3} \frac{(d_4^{u,S}(r_1, \ldots, r_N) - \frac{1}{8})}{\sqrt{\frac{15}{8}}} \to \mathcal{N}(0,1).$$

These values in distribution are standard normally distributed for $N \to \infty$ (Kustosz, Müller, and Wendler 2016) and can be implemented via

```
sqrt(N-2)*(simple_3depth(residuals) - 0.25) / (0.25*sqrt(5))
```

and

```
sqrt(N-3) * (simple_4depth(residuals) - 1/8) / (sqrt(15)/8)
```

3.6.3 Full K-simplex depth

For bivariate residuals, simplex-based extensions are used. The full K-simplex depth

$$d_K^F(r_1, \ldots, r_N) = \frac{1}{\binom{N}{K+p}} \sum_{1 \le n_1 < \cdots < n_{K+p} \le N} \mathbf{1}\left\{0_p \in \bigcap_{k=1}^{K} \mathbb{S}(r_{n_k}, \ldots, r_{n_{k+p}})\right\},$$

with $\mathbb{S}(r_{n_k}, \ldots, r_{n_{k+p}})$ be the open simplex spanned by $r_{n_k}, \ldots, r_{n_{k+p}} \in \mathbb{R}^p$. The full 1-simplex depth is first introduced in Liu 1988 and Liu 1990. The extension to

the full K-simplex depth is made by Müller, Nagy, and Trippler 2024. In further proceedings, the simplex depth for $K = 1$ and $K = 2$ will be considered.

Quantile level	$K = 1$	$K = 2$
1%	0.2388	0.0784
1.667%	0.2401	0.0790
2%	0.2408	0.0792
5%	0.2437	0.0805
10%	0.2459	0.0813

Table 4: Simulated quantiles of the full K-simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 10\,000$, Sample size within a single simulation: $N = 200$).

The table 4 shows the empirical quantiles of this depth measure, based on $M = 10\,000$ simulations with a sample size of $N = 00$ for $K = 1$ and $N = 200$ for $K = 2$. The sample size was limited to $N = 200$ for $K = 21$ and $K = 2$.

The full K-simplex depth can be computed via

```
fullLSimplex(residuals)
```

from the package `SignDepth` for $K = 1$ and $K = 2$.

Standardised full K-simplex depth

The standardised version of the bivariate K-simplex depth is defined for $K = 1$ and $K = 2$ as follows (Müller, Nagy, and Trippler 2024):

$$D_1^F(r_1, \ldots, r_N) = N \cdot (d_1^F(r_1, \ldots, r_N) - \frac{1}{4}),$$

$$D_2^F(r_1, \ldots, r_N) = N \cdot (d_2^F(r_1, \ldots, r_N) - \frac{1}{12}).$$

Table 5 shows the estimated quantiles for $N = 200$ for $K = 1$ and $K = 2$. The distribution was approximated using $M = 10\,000$ independent simulation runs.

Quantile level	$K = 1$	$K = 2$
1%	-2.2382	-0.9836
1.667%	-1.9720	-0.8620
2%	-1.8358	-0.8169
5%	-1.2535	-0.5736
10%	-0.8250	-0.4014

Table 5: Simulated quantiles of the standardised full-simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 10\,000$, Sample size within a single simulation: $N = 200$).

3.6.4 Simplified K-Simplex Depth

The simplified version is computed using the function

`simplLSimplex(x)`

from the package `SignDepth` for $K = 1, 2$ and relies on local sign configurations instead of all combinations: let $S(x_1, \ldots, x_{p+1})$ denote the simplex spanned by these points. Then

$$d_K^S(r_1, \ldots, r_N) = \frac{1}{N - K - p + 1} \sum_{n=1}^{N-K-p+1} \mathbf{1}\left\{0_p \in \bigcap_{k=1}^{K} \mathbb{S}(r_{n+k-1}, \ldots, r_{n+k-1+p})\right\}.$$

This method is introduced in Müller, Nagy, and Trippler 2024. Table 6 summarises the empirical quantiles calculated from $M = 100000$ simulated depth values. Each simulation is based on a sample of size $N = 500$.

Quantile level	$K = 1$	$K = 2$
1%	0.2008	0.0503
1.667%	0.2048	0.0523
2%	0.2068	0.0523
5%	0.2149	0.0584
10%	0.2229	0.0644

Table 6: Simulated quantiles of the simplified simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 500$).

Standardised simplified K-simplex depth

Standardised version of the bivariate simplified simplex depth are defined as (Müller, Nagy, and Trippler 2024) with their following asymptotic distribution:

$$D_1^S(r_1, \ldots, r_N) = \sqrt{N-2} \cdot \frac{d_1^S(r_1, \ldots, r_N) - \frac{1}{4}}{\frac{1}{4} \cdot \sqrt{\frac{11}{3}}} \to \mathcal{N}(0, 1),$$

$$D_2^S(r_1, \ldots, r_N) = \sqrt{N-3} \cdot \frac{d_2^S(r_1, \ldots, r_N) - \frac{1}{12}}{\frac{1}{12} \cdot \sqrt{\frac{169}{10}}} \to \mathcal{N}(0, 1),$$

in distribution for $N \to \infty$. For these last two depths, the considered bivariate pairs are P_i, Q_i for $i = 1, 2, 3$.

3.6.5 Depth tests

For a general depth functional d_K^j, with $K \in \mathcal{K}$ (e.g. $K = 1, 2, 3, 4$) and $j \in \{S, F\}$, define the empirical quantile

$$q_{K,j}(\alpha)$$

as the α-quantile of the sample

$$\left\{ d_K^j(r_1^m, \ldots, r_N^m)\,,\ m = 1, \ldots, M \right\},$$

.

Then, the corresponding test rejects H_0 if

$$\sup_{\theta \in \Theta_0} d_K^j(r_1(\theta), \ldots, r_N(\theta)) < q_{K,j}(\alpha),$$

for $K \in \mathcal{K}$ and $j \in \{S, F\}$ (Müller, Nagy, and Trippler 2024).

For a generic depth function d, let

$$R_d(\alpha) = \mathbb{1}\left\{ \sup_{\theta \in \Theta_0} d(\theta) < q_d(\alpha) \right\}$$

denote the rejection indicator of the corresponding depth-based test at significance level α. Thus, $R_d(\alpha) = 1$ indicates the rejection of H_0, while $R_d(\alpha) = 0$ indicates non-rejection.

In particular, the rejection indicators corresponding to the depth functions considered in this paper are

$$R_{d_K^{u,F}}(\alpha), \quad R_{D_K^{u,F}}(\alpha), \quad R_{d_K^{u,S}}(\alpha), \quad R_{D_K^{u,S}}(\alpha),$$

and

$$R_{d_K^F}(\alpha), \quad R_{D_K^F}(\alpha), \quad R_{d_K^S}(\alpha), \quad R_{D_K^S}(\alpha).$$

3.7 Shapiro-Wilk test

Let $r_1, \ldots, r_n$ be ordered observations or like here residuals with ordered observations $r_{(1)} \leq \cdots \leq r_{(n)}$. The Shapiro–Wilk test is based on the test statistic

$$W = \frac{\left(\sum_{i=1}^n a_i r_{(i)}\right)^2}{\sum_{i=1}^n (r_i - \bar{r})^2},$$

where $\bar{r}$ denotes the sample mean (Shapiro and Wilk 1965). The coefficients a_i determine the weights of the ordered observations and are given by:

$$a' = (a1, \ldots, a_n) = \frac{m^\top G^{-1}}{\sqrt{m^\top G^{-1} G^{-1} m}},$$

where the vector $m = (m_1, \ldots, m_n)^\top$ denotes the vector of expected values of the order statistics, and the matrix G denotes the corresponding covariance matrix.

Small values of W provide evidence against the null hypothesis of a normal distribution. A significant p-value in the Shapiro–Wilk test indicates a deviation from the normal distribution. However, the test is extremely sensitive, particularly with large sample sizes, meaning that even very small and practically insignificant deviations can lead to the rejection of the null hypothesis (Field 2009, p. 148 & p. 793).

3.8 Implementation in R

The implementation of the methods and the simulation are programmed in the software program `R`, version 4.5.0 (R Core Team 2025). To import the admittance matrix, the `readxl` package is used (Wickham and Bryan 2025).

The packages `Matrix` (Bates and Maechler 2021) and `MASS` (Venables and Ripley 2002) are used as an extension for matrix calculations. For the simulations, the packages `numDeriv` (Gilbert and Varadhan 2019) and `mvtnorm` (Genz and Bretz 2009) are used. The package `pracma` (Borchers 2025) is used to calculate a matrix square root. Package `SignDepth` (Nagy 2026) is used for calculating depth values.

The package `ggplot2` (Wickham 2016) is used as well as the packages `corrplot` (Wei and Simko 2024), `patchwork` (Pedersen 2025), `dplyr` (Wickham, François, et al. 2026) and `tidyverse` (Wickham, Averick, et al. 2019) to create graphics.

4 Simulation study

In this chapter, a simulation study is used to investigate the behaviour of the residuals of the extended Kalman filter for different depth values. Therefore, various scenarios are considered. This chapter aims to determine the extent to which these scenarios affect the depth values, and how these values respond to a wide variety of situations. This should help assess the usefulness of depth values as performance criteria.

Section 4.1 presents the simulation framework for the classical model with load profiles. For this purpose, the process of generating data, the extended Kalman filter which is used, and the corresponding implementation are described. Based on this, section 4.2 analyses the residuals under correct model specifications. First, the model quality is assessed using NEES and NIS, and then the residuals are analysed using depth values.

Section 4.3 discusses the consequences of incorrect parameter assumptions and their impact on the resulting depth values. This analysis considers the influence of both the state transition parameter θ_v, and the initial reference state vector v_0. This is followed by an assessment of robustness in section 4.4. Both positive additive outliers and two-sided additive outliers are analysed in this way.

Furthermore, section 4.5 considers an alternative parameter setting for the state transition parameter $\theta_v = 0.9$. This additional analysis serves as a sensitivity test, investigating the behaviour of the depth values for a parameter setting closer to that of the real-world application.

4.1 Simulation model for the classical model with load profiles

This subsection examines the Kalman filter using simulation. This is used to evaluate the statistical effects of incorporating realized load profiles.

The aim is to compare two specifications of the extended Kalman filter: first a basic model that ignores deterministic load profile information in the state transition equation, and second an extended model that explicitly incorporates this information. The comparison focuses on the estimation accuracy, innovation behaviour, and residual distribution properties.

For this simulation study the initial condition is chosen as

$$v_0 = c_0 = c_0^1 = c_0^2,$$

with

$$e_n^0 = 0.9, \quad f_n^0 = 0.1, \quad \text{for } n = 1, \ldots, N.$$

The initial state covariance matrices for both filters are

$$C_0 = C_0^1 = C_0^2 = 0.1 \cdot \mathbb{I}_{2N \times 2N},$$

where $\mathbb{I}_{2N \times 2N}$ denotes the $2N \times 2N$ identity matrix.

Furthermore, process and measurement noise are assumed Gaussian and independent:

$$e_t^v \sim \mathcal{N}(0, \Sigma_v), \qquad e_t^z \sim \mathcal{N}(0, \Sigma_z).$$

4.1.1 Data generating process

To evaluate the filters under controlled conditions, data is simulated from a nonlinear state-space model. The voltage process in is simulated as follows

$$v_t = \theta_v v_{t-1} + (1 - \theta_v)v_0 + e_t^v,$$

where v_0 denotes a fixed reference state and θ_v controls the persistence of the process. Further, v_0 is set as the start vector $v_{0_n} = (e_n^0,\ f_n^0)$ for $n = 1, \ldots, N$.

Observations are generated through a nonlinear function

$$z_t = h(v_t) + e_t^z,$$

where z_t denotes the realized load profiles and $h(\cdot)$ represents the nonlinear link function.

4.1.2 Extended Kalman filter

State estimation is performed using an Extended Kalman filter. The prediction step is given by (Müller 2026)

$$a_t = \theta_v c_{t-1} + (1 - \theta_v)v_0,$$

$$R_t = \theta_v^2 C_{t-1} + \Sigma_v.$$

The nonlinear measurement prediction is

$$b_t = h(a_t),$$

$$H_t = h'(a_t).$$

The innovation covariance matrix is

$$S_t = H_t R_t H_t^\top + \Sigma_z,$$

and the Kalman gain is

$$K_t = R_t H_t^\top S_t^{-1}.$$

The update equations are

$$c_t = a_t + K_t(z_t - h(a_t)),$$

$$C_t = R_t - K_t H_t R_t.$$

The residuals used in the subsequent analysis are defined as

$$r_t = z_t - h(a_t),$$

or

$$r_t = S_t^{-\frac{1}{2}} \cdot (z_t - h(a_t)).$$

4.1.3 Simulation setup

The simulations were carried out for a network consisting of three nodes. For each node, both active power (P) and reactive power (Q) were considered, resulting in a six-dimensional state vector. The state transition parameter was set to $\theta_v = 0.1$.

The admittance matrix was taken from the real data. Since the network topology is independent of the observed measurements, the same admittance matrix can also be used for the simulated data. The admittance matrix obtained from the SimBench data set of Meinecke et al. 2020 is given by

$$A = \begin{bmatrix} 31.7 - 12.4\text{j} & 0 & 0 & -31.7 + 12.4\text{j} \\ 0 & 259.3 - 101.3\text{j} & 0 & -259.3 + 101.3\text{j} \\ 0 & 0 & 817.3 - 319.5\text{j} & -817.3 + 319.5\text{j} \\ -31.7 + 12.4\text{j} & -259.3 + 101.3\text{j} & -817.3 + 319.5\text{j} & 1108.3 - 433.2\text{j} \end{bmatrix}.$$

To investigate the robustness of the results with respect to the random initialisation, simulations were performed using two different random seeds (1 and 1234). Unless stated otherwise, the results shown in this thesis are based on the simulations using seed 1. The second seed is used to assess the robustness of the findings.

The results presented in this section were generated using a *constant-noise EKF* and a *state-dependent-noise EKF*. Both are Extended Kalman Filter (EKF) for the same nonlinear state-space model. The prediction and update equations are identical in both versions. The main differences concern the specification of the covariance matrices and the diagnostic quantities that are stored during the simulation. For each implementation, both the raw innovations (residuals) and their standardised

counterparts are analysed.

In the constant-noise EKF, both the process noise covariance matrix Σ_v and the observation noise covariance matrix Σ_z are constant throughout the simulation. The process covariance is defined as

$$\Sigma_v^{\text{const}} = \sigma_v^2 \mathbb{I}_{2N\times 2N},$$

with $\sigma_v^2 = 0.01$. The observation covariance is chosen as

$$\Sigma_z^{\text{const}} = 3\mathbb{I}_{2N\times 2N}.$$

Consequently, the assumed process and observation uncertainties remain constant over the entire simulation period.

In the state-dependent-noise EKF, both covariance matrices depend on the current scale of the system. The process covariance is defined as

$$\Sigma_v^{\text{state}} = \text{diag}(\sigma_v^2 \cdot v_0),$$

where $\sigma_v^2 = 10^{-5}$. The observation covariance is calculated as

$$\Sigma_z^{\text{state}} = \text{diag}(0.01 \cdot |h(v_0, A)|)\,,$$

where $|h(v_0, A)|$ denotes the vector obtained by taking the absolute value of each component of $h(v_0, A)$. Thus, larger state values and larger expected observations are associated with larger uncertainty. For the initial state used in this study,

$$h(v_0, A) = (-1.296,\ -4.162,\ -10.614,\ -34.034,\ -33.434,\ -107.290),$$

which yields

$$\Sigma_z^{\text{state}} = \begin{pmatrix} 0.01296 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04162 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.10614 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.34034 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33434 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.07290 \end{pmatrix}.$$

In the state-dependent-noise EKF, the updated covariance matrix is symmetrised after each filter step,

$$C_t = \frac{1}{2}\left(C_t + C_t^\top\right),$$

to reduce numerical asymmetries caused by floating-point arithmetic. The same procedure is applied to the innovation covariance matrix.

Both implementations store the innovation residuals. The considered filters analyse a set of ordinary innovation residuals, and a second set of standardised residuals obtained by scaling the innovations with their covariance matrix. Consequently, the results presented in this section are based on both residual definitions. All variables needed to estimate the Normalised Estimation Error Squared (NEES) and the Normalized Innovation Squared (NIS) are stored during the simulation.

4.2 Analysis of the residuals based on the correct model specification

This chapter analyses the results of the simulation study. First, the simulated (standardised) residuals generated using the correct model parameters are examined. The aim of this investigation is to verify whether the simulated data satisfy the distribution assumptions made in the model and whether the depth-based testing procedures lead to the expected results under the null model.

4.2.1 Graphical analysis of the residuals

First, the residuals from the two EKFs are analysed descriptively as an example. The non-standardised residuals from both filters, using seed 1, are used for this analysis. For each of the six components (P_1, Q_1 P_2, Q_2 P_3, Q_3), individual plots are examined. For each component, a Q-Q-plot, also known as a quantile-quantile plot, a density plot and a boxplot is analysed. These plots are shown collectively in the figures 29, 30 and 31 for the state-dependent-noise EKF and in the figures 32, 33 and 34 for the constant-noise EKF.

The Q-Q-plots are used to test the assumption of a normal distribution. Although the data were generated using normally distributed random variables, it is still useful to check this assumption. Q-Q-plots are created by plotting the empirical quantiles of the simulated residuals against the theoretical quantiles of a standard normal distribution. In addition, a reference line (here the red line) is plotted. If the data points lie close to the reference line, this supports a normal distributions

assumption. Should the plotted data points differ systematically from the reference line, this would argue against a normal distributions assumption (Wilk and Gnanadesikan 1968).

Density plots were used to analyse the distribution of the simulated residuals. These plots represent the distribution of the data as a smoothed curve and are well suited to identifying key characteristics such as skewness, multimodality or prominent outliers. The density plots were generated in R using the `density()` function. This is based on kernel density estimation, a non-parametric method for estimating a probability density. In this process, a small smooth curve (kernel) is fitted around each data point and then summed over all observations. This results in a continuous estimate of the distribution without the need for a fixed classification of classes, as is the case with a histogram. Formally, the estimated density function is given by:

$$\hat{f}(v) = \frac{1}{nh_{bw}} \sum_{i=1}^{n} K\left(\frac{v - v_i}{h_{bw}}\right),$$

where n denotes the number of observations, h_{bw} the bandwidth, and $K(\cdot)$ the kernel function. The bandwidth controls the degree of smoothing: small values result in a detailed but potentially noisy representation, whilst larger values smooth the curve more strongly (Silverman 1986).

In `R`, a Gaussian kernel is used by default. The bandwidth is automatically selected using the so-called 'nrd0' method, which is based on a rule by Silverman (Silverman 1986). This method uses a combination of the standard deviation and interquartile range of the data and scales these by the sample size to determine an appropriate smoothing level. This achieves a practical compromise between over-smoothing and under-smoothing.

Finally, for each component boxplots are analysed. These provide a compact overview of the simulated residuals. The box itself contains the middle 50% of the data, i.e. the range between the first and third quartiles, whilst the line in the middle represents the median. The whiskers extend up to 1.5 times the interquartile range in both directions, thus showing the range within which most values lie (Toit, Steyn, and Stumpf 2012, p. 29). Points outside this range are referred to as outliers and are marked in blue in the graphs.

Looking only at the middle section of the plotted data points from the state-dependent-noise EKF in the Q-Q plots, these are quite close to the reference line. However, a few data points in the outer regions differ significantly from the reference line in some cases. It is noticeable that larger deviations occur only in the Q values. Furthermore, these deviations do not occur evenly on both sides. For the components

Q_1 and Q_3, a deviation occurs on only one side. This indicates a skew in the distribution.

The density plots show the skewness suggested by the Q-Q plots. It can be seen that the distributions of the components Q_1 and Q_3 for the state-dependent-noise EKF clearly have a left-skewed distribution. Additionally, P_2 is slightly right-skewed and definitely has a heavy tail in this edge region.

The boxplots show that the boxes are mostly relatively small, which supports the observation of narrow peaks in the density plots. It can also be seen that all components except P_1 and Q_2 exhibit a few high outliers.

When looking at the Q-Q plots of the residuals of the constant-noise EKF the middle part is again close to the reference line. As before, there are some points in the outer areas that differ from this reference line. When comparing to the density plots it is clear to see that P_1, Q_1 and Q_3 have a slight left-skewed distribution. In comparison to before, the boxes of the boxplots are a bit wider. Outliers can be seen for the components Q_1 and Q_3.

In summary, it can be said that the distributions of the residuals of both filters at least resemble normal distributions; however, it is not possible to determine with certainty from the graphs whether the components follow a normal distribution.

An analysis of the NIS and NEES values obtained from the state-dependent-noise EKF values are shown in the figures 2 and 3. These show that the chosen model parameters are generally plausible.

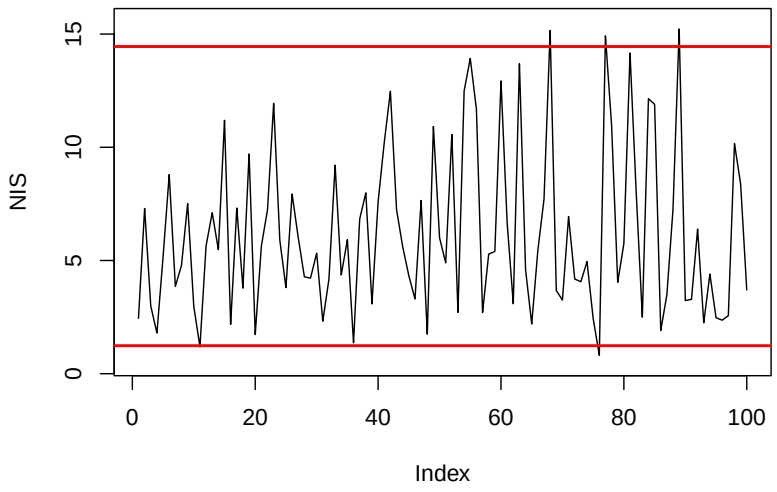

Figure 2: NIS values for the state-dependent-noise EKF of an illustrative simulation run.

For the NEES and NIS application $n_v = n_z = 6$ is considered. Therefore, the expected NEES and NIS equals 6, and the corresponding 95% confidence interval is approximately [1.24, 14.45].

The observed deviations from the theoretically expected limits are rare for both methods and do not suggest any systematic modelling errors. Therefore, it can be concluded that the model assumptions of the state-dependent-noise EKF provide an accurate enough description of the model, and that the filter is suitable application.

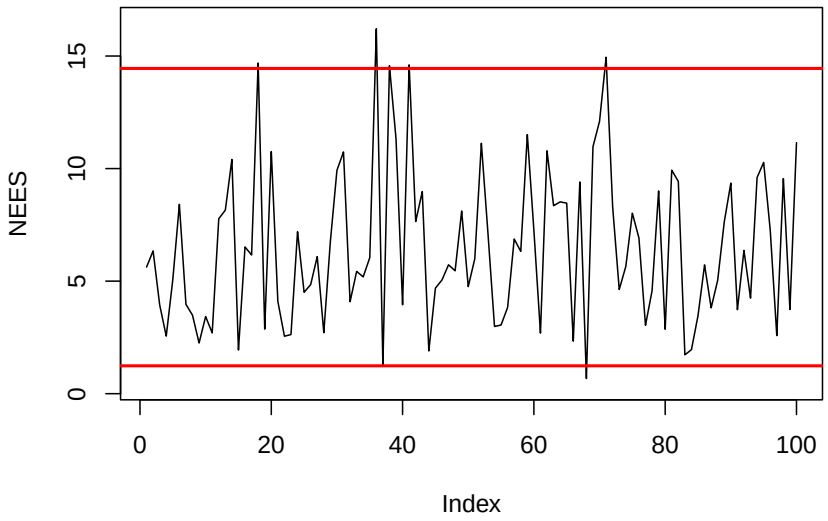

Figure 3: NEES values for the state-dependent-noise EKF of an illustrative simulation run.

The NIS values from the constant-noise EKF, shown in figure 4, remain within the corresponding confidence bounds, which indicates that the predicted measurement distribution is consistent with the observations. However, the NEES values (shown in figure 5) exceed the upper confidence limit, suggesting that the estimated state uncertainty has possibly been underestimated. While the constant-noise EKF can sufficiently describe the measurements, the internal state estimation is not fully consistent with the true model. This may be caused by a model mismatch or an inappropriate specification of the process noise covariance.

Although the NEES analysis of the constant-noise EKF indicates inconsistencies in the estimated state uncertainty, the NIS values remain within the expected confidence bounds. This suggests that the constant-noise EKF sufficiently represents the measurement predictions and resulting innovations.

Since the following analysis focuses on the residuals and their corresponding depth values rather than the accuracy of the estimated state itself, the model is still suitable for its intended purpose. The observed NEES deviations indicate limitations in state estimation and are taken into account when interpreting the results.

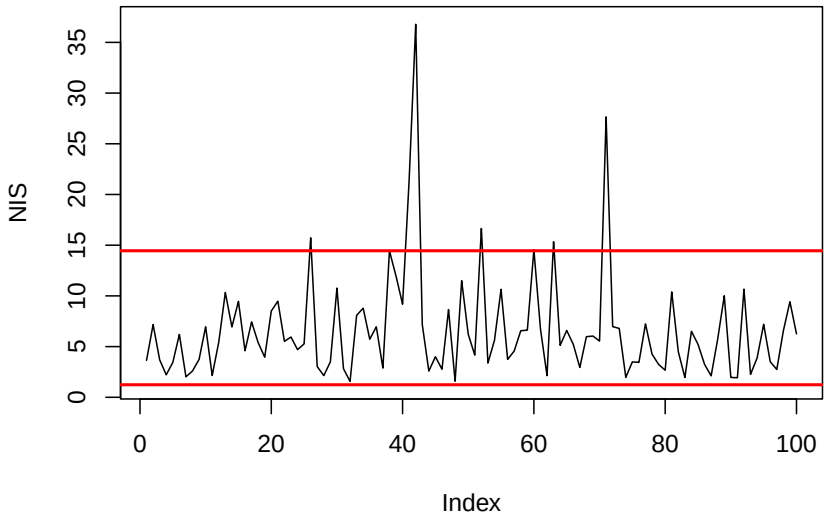

Figure 4: NIS values for the constant-noise EKF of an illustrative simulation run.

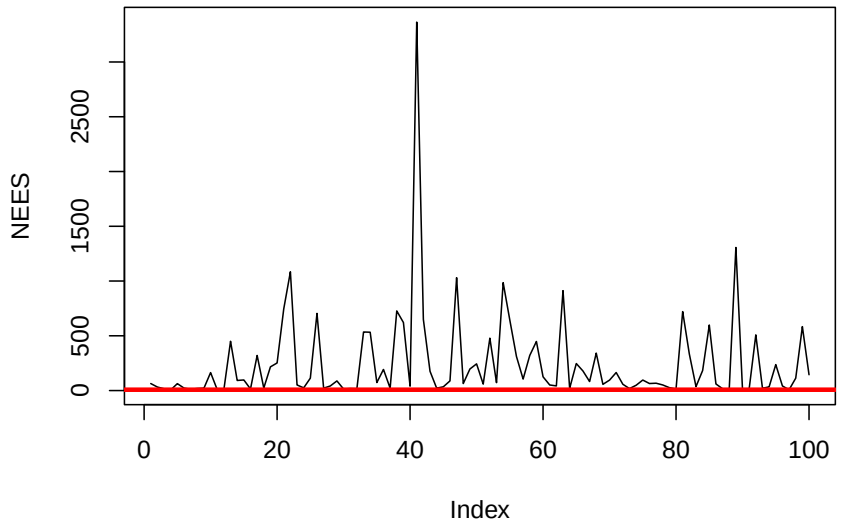

Figure 5: NEES values for the constant-noise EKF of an illustrative simulation run.

The results shown are illustrative examples, generated from a single simulation run with a fixed random seed. They are intended for the qualitative analysis of the filter behaviour, allowing a comparison between the two approaches. This is not a statistically comprehensive evaluation; a larger number of independent simulation runs would be required for that. However, for the purpose of the following analysis, this is sufficient.

4.2.2 Analysis of the residuals using depth values

This section compares various depth values based on the simulated (standardised) residuals. An overview of the depth methods and the symbols used to denote them in the following results tables can be found in table 7. For each depth function, a hypothesis test is performed based on the corresponding quantile. The test decision is described by the indicator $R_d(\alpha)$. A value of $R_d(\alpha) = 1$ corresponds to a rejection of the null hypothesis, whilst $R_d(\alpha) = 0$ means that the null hypothesis is not rejected.

The critical values used for the depth tests were determined using simulations and thus form the empirical quantiles. That means, that these quantiles need to be seen critical. Quantiles simulated with a bigger sample size are listed in chapter 3. This shows that the simulated quantiles are changing a lot for different sample and simulation sizes. Therefore, some possible close test choices need to be seen critical. These are listed in detail in the tables in Appendix A.1 'Quantile tables' tables 24, 25, 26, 27, 28 and 29. For the non-standardised depth statistics, the quantiles correspond to the empirically determined critical values of the respective test statistic. For the standardised variants, the theoretical quantiles of the standard normal distribution are also used, as these statistics asymptotically converge to a standard normal distribution.

The depth tests were performed at a significance level of $\alpha = 0.05$ and at an adjusted significance level of $\alpha_{\text{Bonf}} = 0.0083$ for the univariate tests and $\alpha_{\text{Bonf}} = 0.0167$ for the bivariate tests.

Symbol	Method
$d_K^{u,F}$	full K-sign depth
$D_K^{u,F}$	standardised full K-sign depth
$d_K^{u,S}$	simplified K-sign depth
$D_K^{u,S}$	standardised simplified K-sign depth
d_K^F	K-simplex depth
D_K^F	standardised full K-simplex depth
d_K^S	simplified K-simplex depth
D_K^S	standardised simplified K-simplex depth
$R_d(\alpha)$	rejection indicator of the corresponding depth-based test at significance level α

Table 7: Overview of the considered depth measures and their notation.

First, the univariate depth values and the corresponding test decisions for the state-dependent-noise EKF are examined. These can be found in table 8 for the seed 1.

Since the data were simulated under the correct model, it is expected that the null hypothesis will not be rejected in most cases. For both significance levels, nearly all values of the full 3-sign depth and 4-sign depth are at their expected values. The full 3-sign depth is close to 0.25 for all components, which is the expected value of this depth method. The full 4-sign depths are also close to their expected value of 0.125 for all components.

In fact, almost all calculated depth values lie above the respective critical quantiles. Accordingly, the rejection indicators of the tests are zero for almost all tests. Only for the full 3-sign depth, full 4-sign depth and simplified 3-sign depth plus their standardised versions is a rejection observed for the component Q_3 at the $\alpha = 0.05$ level. However, after applying the Bonferroni correction, this rejection disappears. It is therefore likely to be a random false alarm, which can be explained by the multiple testing problem.

Given that this analysis is based on a single illustrative simulation, this discrepancy is not interpreted as an indication that the model is fundamentally unsuitable.

The table 9 with the results of the simplex-based methods does not show anything unusual either. Here all the depth values are within the expected range and the corresponding rejection indicators are all zero.

To investigate the influence of the chosen random seed on the depth values, another set of depth values from the state-dependent noise EKF is analysed for a different random seed in the tables 30 and 31. The results are very similar, with the

full and simplified K-sign depth and the simplex depth values all being very close to each other.

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2548	0.2538	0.2504	0.2556	0.2554	0.2337	0.2375	0.2267
$D_3^{u,F}$	0.4780	0.3791	0.0427	0.5622	0.5387	-1.6271	-1.2486	-2.3296
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1294	0.1273	0.1264	0.1300	0.1286	0.1135	0.1155	0.1078
$D_4^{u,F}$	0.4398	0.2321	0.1401	0.5041	0.3633	-1.1490	-0.9476	-1.7185
$R_{d_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2755	0.3265	0.2551	0.3673	0.2551	0.1531	0.1633	0.1224
$D_3^{u,S}$	0.4518	1.3553	0.0904	2.0781	0.0904	-1.7167	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,S}$	0.1546	0.2165	0.1134	0.2680	0.1134	0.0619	0.0515	0.0309
$D_4^{u,S}$	0.6030	1.8613	-0.2359	2.9100	-0.2359	-1.2846	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
Shapiro-Wilk	0.4404	0.0233	0.0649	0.8046	0.1738	0.1975		

Table 8: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF for Seed 1.

Note. $R_d(\alpha) = 1$ indicates rejection of the null hypothesis at significance level α, $R_T(\alpha) = 0$ indicates non-rejection. The significance levels are $\alpha = 0.05$ and the Bonferroni-adjusted level $\alpha_{\text{Bonf}} = \frac{0.05}{6} = 0.0083$. Critical values were obtained from the empirical quantile tables reported in Appendix A.1. For depth statistics converging in distribution to a standard normal random variable, the corresponding critical values are given by $\Phi^{-1}(0.05) = -1.6449$ and $\Phi^{-1}(0.0083) = -2.3940$.

Methods	**Nodes**			**Quantiles**	
	$(P_1,\ Q_1)$	$(P_2,\ Q_2)$	$(P_3,\ Q_3)$	α	α_{Bonf}
d_1^F	0.2543	0.2478	0.2433	0.2376	0.2303
D_1^F	0.4310	-0.2245	-0.6685	-1.2362	-1.9672
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0853	0.0828	0.0799	0.0775	0.0745
D_2^F	0.1938	-0.0572	-0.3409	-0.5867	-0.8810
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2245	0.3061	0.3163	0.1735	0.1531
D_1^S	-0.5275	1.1606	1.3716	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0515	0.0619	0.0928	0.0309	0.0206
D_2^S	-0.9138	-0.6175	0.2717	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 9: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF for Seed 1.

Note. $R_d(\alpha) = 1$ indicates rejection of the null hypothesis at significance level α, $R_T(\alpha) = 0$ indicates non-rejection. The significance levels are $\alpha = 0.05$ and the Bonferroni-adjusted level $\alpha_{\text{Bonf}} = \frac{0.05}{3} = 0.0167$. Critical values were obtained from the empirical quantile tables reported in Appendix A.1. For depth statistics converging in distribution to a standard normal random variable, the corresponding critical values are given by $\Phi^{-1}(0.05) = -1.6449$ and $\Phi^{-1}(0.0167) = -2.1280$.

This time, the null hypothesis for component Q_3 at the 3-sign depth and standardised 3-sign depth is rejected at the $\alpha = 0.05$ level. However, this changes as soon as the significance level is adjusted using the Bonferroni correction. Therefore, it can be concluded that the depth values are not dependent on the random seed.

The standardised residuals of the state-dependent noise EKF were then examined using depth values for seed 1, and the depth values remained comparable to those

of the non-standardised residuals (see tables 32 and 33). Interestingly, however, rejections no longer occur for the full 3-sign depth, full 4-sign depth and simplified 3-sign depth values and their standardised versions. This suggests that they are eliminated by the standardisation process. Otherwise, the rejection indicators remained unchanged and showed no significant deviations for this example simulation run.

Similar patterns emerge when comparing with the constant-noise EKF in the tables 34 and 35. Seed 1 was also used here to ensure the results are comparable. For the constant-noise EKF, the null hypothesis for the full 3-sign depth, full 4-sign depth and simplified 3-sign depth and their standardised versions is also rejected for the component Q_3 at the $\alpha = 0.05$ level. Unlike previously, the standardised residuals also reject these cases this time (for more details refer to the tables 36 and 37). The other depth values remain consistent across all scenarios.

The results show that the state-dependent-noise and constant-noise EKFs both satisfy the expected distribution properties of the simulated states. The Shapiro–Wilk statistics reveal lower values for certain nodes, particularly the standardised residuals. However, as the depth-based tests do not detect any significant differences, the results provide no evidence of a significant violation of the normality assumption.

4.3 Consequences of model parameter misspecification

The following section investigates the effect a misspecified parameter in the voltage process has on the simulated data, and to what extent this affects the depth tests. First, an examination is conducted to determine the effects of a mismatch between the state transition parameter θ_v, which is used to simulate the voltage process, and the θ parameter used for the state estimation. Secondly, an analysis with the same principle for the reference state vector v_0 is conducted.

4.3.1 Influence of the state transition parameter θ_v misspecification: state-dependent-noise EKF

A simulation study of 500 runs is carried out here to evaluate the extended Kalman filter. Each simulation considers 100 simulated data points. First, observations are generated and then processed by the filter.

The state process is simulated according to

$$v_t = \theta v_{t-1} + (1 - \theta)v_0 + e_t^v,$$

v_0 describes the reference state and $e_t^v \sim \mathcal{N}(0, \Sigma_t^v)$ represents the process noise term. The parameter θ defines the state transition dynamics of the system. For large values of the parameter θ the current state depends heavily on the previous state. If these parameter values are smaller, then the current state leans more towards the reference state vector v_0.

The observations are then calculated using the non-linear observation function

$$z_t = h(v_t) + e_t^z,$$

where $e_t^z \sim \mathcal{N}(0, \Sigma_t^z)$ represents the measurement error.

First, the case of a correctly specified model is being analysed. Furthermore, the study investigates how a discrepancy between the state transition parameter used in the simulation and the state transition parameter assumed by the filter affects the estimation performance.

Therefore, the system states are simulated using a parameter θ_{true},

$$v_t = \theta_{\text{true}} v_{t-1} + (1 - \theta_{\text{true}}) v_0 + e_t^v,$$

while the filter continues to use the parameter θ_{model} in its prediction step. The prediction model assumed by the filter is therefore given by

$$\hat{v}_{t|t-1} = \theta_{\text{model}} \hat{v}_{t-1|t-1} + (1 - \theta_{\text{model}}) v_0.$$

Here, $\hat{v}_{t|t-1}$ denotes the predicted state at time step t, based on the information available up to the previous time step $(t-1)$. This prediction is obtained by applying the state transition model before incorporating the new measurement at time t. The term $\hat{v}_{t-1|t-1}$ represents the estimated state at the previous time step after the measurement update has been performed.

In the case of a correctly specified model, the parameter used by the filter corresponds to the actual parameter of the simulated system, i.e.,$\theta_{\text{model}} = \theta_{\text{true}}$.

However, to investigate the influence of model misspecification, the filter can be intentionally provided with an incorrect parameter value such that $\theta_{\text{model}} \neq \theta_{\text{true}}$.

In this case, the filter prediction is based on a different state transition behaviour than the one underlying the simulated process. This results in a model mismatch, allowing the robustness of the filter with respect to incorrect assumptions about system dynamics to be analysed.

As the true parameter $\theta = 0.1$ was chosen. As false parameters,

$$\theta = 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9$$

were tested. To compare the results effectively, the rejection rates of the various depth tests are reviewed. In each simulation, data is simulated for three nodes, so for six components, and the combined rejection rate is then calculated. To do this, the number of depth tests rejections at the significance level $\alpha = 0.05$ is counted and then divided by $(500 \cdot 6)$ to get the reported rejection rates.

First, the behaviour of these rejection rates for the state-dependent-noise EKF is shown in figure 6. This figure shows the rejection rates dependent on an increasingly incorrect θ.

For $\theta_{\text{model}} = 0.1$, the filter model corresponds to the true system dynamics, and no model mismatch is present. In this case, the rejection rates are close to the significance level of $\alpha = 0.05$, indicating that the tests maintain the expected false rejection probability for a correctly specified model.

As the parameter θ_{model} increases, the difference between the assumed and the true state transition parameter becomes larger. This results in an increasing rejection probability for most of the investigated methods. The observed increase demonstrates that the tests are sensitive to the difference between the actual process dynamics and the dynamics assumed by the filter. The different standardised versions of the methods show nearly identical behaviour compared to their non-standardised counterparts. Therefore, the standardisation does not appear to significantly influence the rejection probability in this simulation setup.

The full simplex-based approaches and the simplified 1-simplex approach show the strongest increase in rejection rates. In particular, the simplified 1-simplex depth method reaches rejection rates above 95% for $\theta_{\text{model}} = 0.9$, indicating a high sensitivity to the model mismatch. The simplified 2-simplex depth displays in comparison to the other simplex-based method relatively low rejection rates.

The full sign-based depth approaches, including the full 3-sign depth and the full 4-sign depth, show a moderate but consistent increase in rejection probability. The full 4-sign depth method generally achieves slightly higher rejection rates than the full 3-sign depth. This indicates a somewhat higher sensitivity to differences in the assumed model dynamics. The lowest rejection rate of the depth methods is displayed by the simplified 4-sign depth.

In contrast, the Shapiro-Wilk test shows only a small increase in rejection rates over the investigated range of θ_{model}. The rejection probability remains close to the

significance level of $\alpha = 0.05$, even for larger differences between θ_{true} and θ_{model}. This indicates that the Shapiro-Wilk test is effective in controlling the false rejection rate for a correctly specified model but has limited power to detect the introduced model mismatch. Therefore, the test is less sensitive to changes in the assumed state transition dynamics compared to the other investigated approaches.

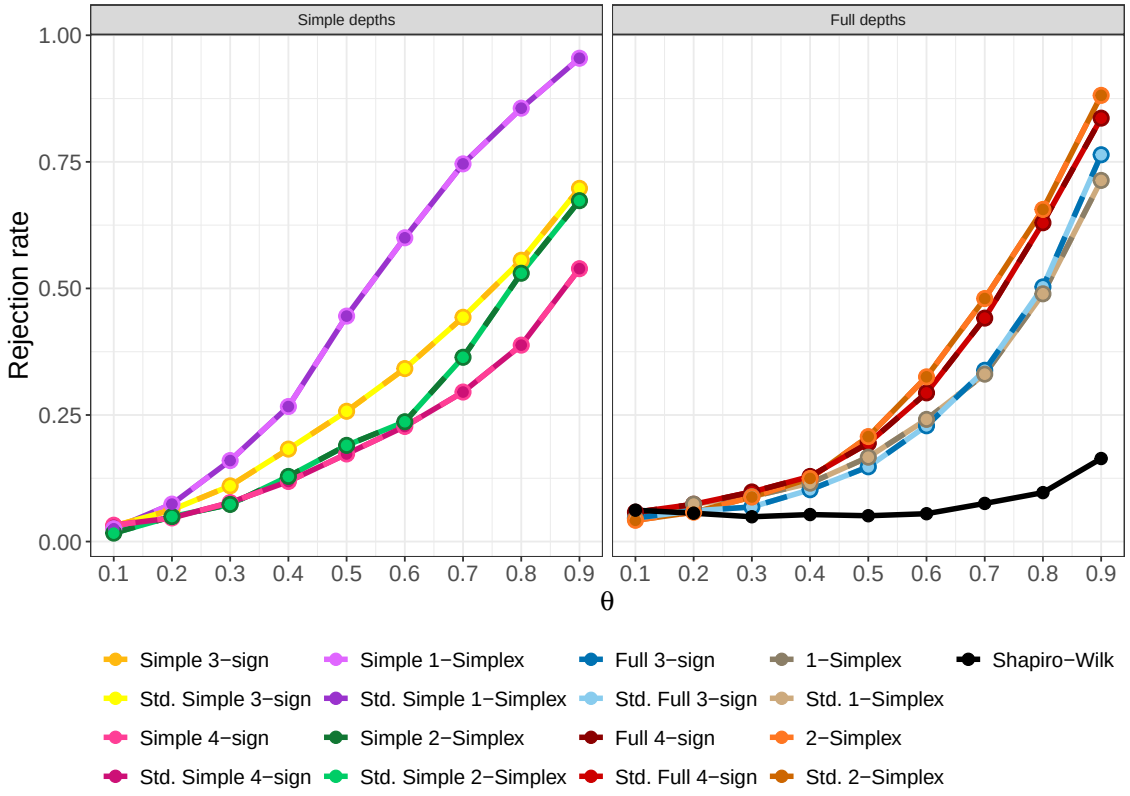

Figure 6: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the state-dependent-noise EKF and non standardised residuals (true parameter $\theta = 0.1$).

The figure 41 shows the rejection rates calculated using the standardised residuals from the state-dependent-noise EKF. The rejection rates obtained from the standardised and non-standardised residuals show nearly identical behaviour for all the relevant methods. This indicates that the standardisation of the residuals does not significantly affect the ability of the tests to detect the introduced model mismatch in this simulation setup. The same trends are observed for both set of residuals, with increasing rejection rates for increasing differences between the assumed and the true state transition parameter.

The similar performance suggests that the information contained in the residual structure is preserved after standardization. Therefore, the main factor influencing

the rejection probability is the differences in the assumed system dynamics rather than the scaling of the residuals.

As can be seen from the figures, the rejection rates do not all lie exactly at 0.05 when the model with the true state transition parameter is used. Therefore, a brief explanation of the Monte Carlo uncertainty to explain this behaviour follows.

Monte Carlo uncertainty

All empirical rejection probabilities reported in this thesis are based on a finite number of Monte Carlo replications. Consequently, the estimated rejection rates are themselves subject to sampling variability (Morris, White, and Crowther 2019, p. 2086).

Suppose that a hypothesis test is repeated independently M times under the same data-generating process. Let R denote the number of rejections and let

$$\hat{p} = \frac{R}{M}$$

be the empirical rejection probability. Since each simulation either results in a rejection or not, the number of rejections follows a binomial distribution,

$$R \sim \text{Bin}(M, p),$$

where p denotes the true rejection probability. Therefore,

$$\mathbb{E}(\hat{p}) = p$$

and

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{M}.$$

Hence, the Monte Carlo standard error of the estimated rejection probability is

$$\text{SE}(\hat{p}) = \sqrt{\frac{p(1-p)}{M}}.$$

In the simulation study presented in this thesis, $M = 500$ Monte Carlo replications are used. For a significance level of $\alpha = 0.05$, the Monte Carlo standard error is approximately

$$\sqrt{\frac{0.05 \cdot 0.95}{500}} \approx 0.0097.$$

Consequently, the empirical rejection rates are expected to fluctuate around the significance level due to Monte Carlo variability (Morris, White, and Crowther 2019, p. 2086). An approximate 95% confidence interval for the empirical rejection probability is therefore given by

$$0.05 \pm 1.96 \cdot 0.0097 = [0.031,\ 0.069].$$

Thus, rejection rates within this interval can reasonably be attributed to Monte Carlo variation, whereas substantially larger deviations indicate that the corresponding test is either liberal or conservative.

4.3.2 Influence of the state transition parameter θ_v misspecification: constant-noise EKF

The rejection rates for the constant-noise EKF are shown in Figure 7. For the correctly specified model $\theta_{\text{model}} = \theta_{\text{true}}$, the rejection rates of all methods remain close to the significance level of $\alpha = 0.05$. This indicates that the investigated approaches maintain an appropriate false rejection probability when the filter model is valid.

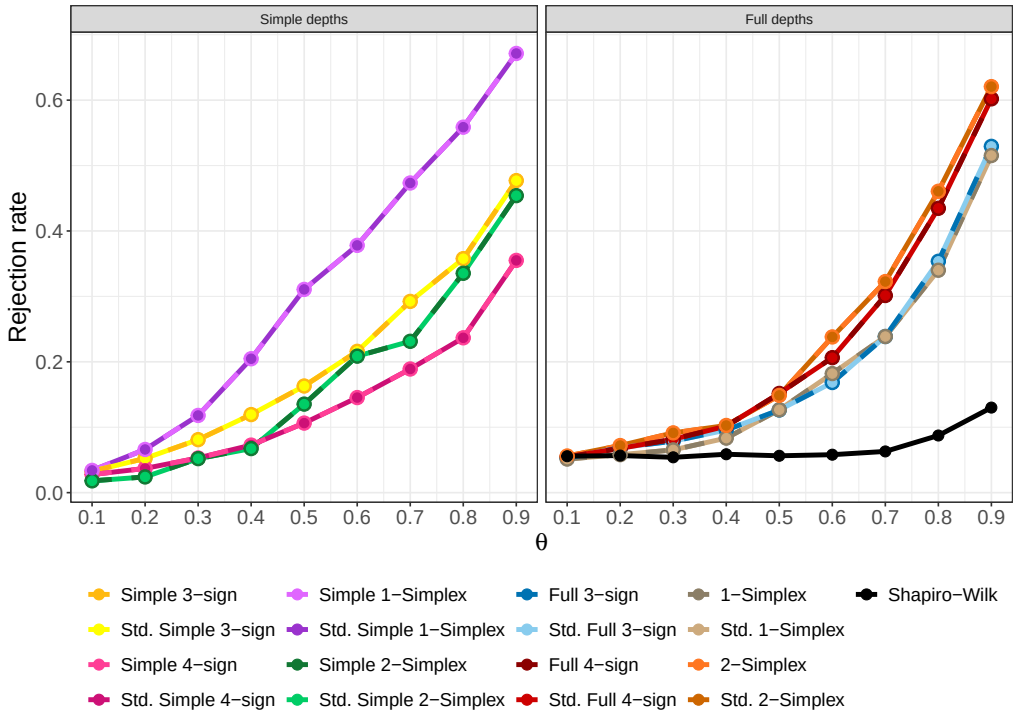

Figure 7: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the constant-noise EKF and non standardised residuals (true parameter $\theta = 0.1$).

With increasing values of θ_{model}, the rejection rates increase for most methods, showing that the tests are able to detect differences from the assumed system dynamics. However, the increase in rejection probability is less pronounced compared to the first simulation setup. The methods therefore show a lower sensitivity to the introduced model mismatch in this scenario.

The depth-based approaches show a gradual increase in rejection rates as the difference between θ_{model} and θ_{true} becomes larger. The full 4-sign depth method generally performs slightly better than the full 3-sign depth, while the inclusion of the simple version improves the sensitivity for some cases. This time the rejection rates for the full 3-sign depth and the full 1-simplex depth show a nearly identical behaviour. The simplex-based methods show a stronger increase compared to the sign-based depth approaches, although the achieved rejection rates remain below the observed ones in the first scenario.

As in the previous simulation, the simplified 1-simplex depth provides the highest rejection rates among the investigated methods. It shows a clear increase from values close to the significance level for the correctly specified model to approximately 67% for $\theta_{\text{model}} = 0.9$. Nevertheless, this increase is considerably smaller than the increase of the state-dependent-noise EKF, where this method reached a rejection rate above 95%. This indicates that the introduced model mismatch is more difficult to detect in this filter.

The Shapiro-Wilk test again remains close to the significance level over the investigated range. Although a slight increase in rejection probability is observed for large differences, its sensitivity remains considerably lower compared to the depth-based approaches.

When looking at the standardised residuals of the constant-noise EKF in this simulation study, a similar behaviour to the non standardised residuals can be observed (see figure 42).

Overall, the results show that rejection rates increase with the amount of discrepancy between the true and assumed state transition parameters. The investigated methods are therefore capable of detecting an incorrectly specified process model, with the simplex-based approaches providing the highest sensitivity among the considered methods. It can also be summarised that the same results were obtained for both the non-standardised and standardised residuals of both EKFs. Considering the Monte Carlo uncertainty, the rejection rates of all investigated methods remain close to the significance level of $\alpha = 0.05$ for the correctly specified model $\theta_{\text{model}} = \theta_{\text{true}}$. This indicates that the methods do not falsely reject the valid filter model.

4.3.3 Influence of the reference state vector v_0 misspecification: state-dependent-noise EKF

In this simulation study, the robustness of the EKF with respect to a misspecified reference state vector is investigated. The model remains unchanged. Only the reference state used by the filter is intentionally chosen incorrectly. The aim is to understand how much the filter and the depth values with their tests are influenced by this type of model misspecification.

The simulated data are therefore always generated from the correct model. In particular, the true reference state vector v_0 is fixed throughout all simulation scenarios. Consequently, the underlying stochastic process is identical in every experiment.

The misspecification is introduced only during the filtering step. Instead of using the true reference state v_0, the filter assumes a different value, denoted by $\tilde{v}_0$.

The prediction equation of the EKF therefore becomes

$$a_t = \theta_v c_{t-1} + (1 - \theta_v)\tilde{v}_0,$$

instead of

$$a_t = \theta_v c_{t-1} + (1 - \theta_v)v_0.$$

Thus, the difference between the predicted state and the observations is only caused by the incorrect reference state assumed by the filter.

Therefore, several levels of misspecification are considered. The true reference state vector is

$$v_0 = (0.9,\, 0.1,\, 0.9,\, 0.1,\, 0.9,\, 0.1)^\top.$$

The filter is then initialized with different assumed reference state vectors,

$$\tilde{v}_0 \in \left\{ \begin{array}{l} (0.9, 0.1, \ldots)^\top, \\ (0.8, 0.2, \ldots)^\top, \\ (0.7, 0.3, \ldots)^\top, \\ (0.6, 0.4, \ldots)^\top, \\ (0.5, 0.5, \ldots)^\top, \\ (0, 0, \ldots)^\top \end{array} \right\},$$

representing increasing degrees of misspecification. The first scenario corresponds to the correctly specified model and serves as a baseline. The remaining scenarios introduce progressively larger deviations from the true reference state. The additional

scenario of the components being zero is interesting, because here the prediction is only made based of the last estimation.

For each scenario, $M = 500$ independent simulation runs are performed. Each run consists of $T = 100$ observations. The rejection rates of all considered diagnostic tests are then estimated. The rejection rate is defined as the proportion of simulation runs in which the corresponding null hypothesis is rejected at the $\alpha = 0.05$ significance level.

This simulation isolates one specific source of model misspecification. All remaining model components are kept fixed. The state transition parameter is chosen as $\theta_v = 0.1$. Consequently, differences in the rejection rates can be attributed exclusively to the incorrect specification of the reference state used. This allows a direct assessment of the sensitivity of the proposed diagnostics to errors in the assumed reference state of the system.

An overview of the various values tested and their scenario names can be found in table 10.

Scenario	e_n	f_n
True	0.9	0.1
Mild	0.8	0.2
Medium	0.7	0.3
Severe	0.6	0.4
Extreme	0.5	0.5
Zero	0.0	0.0

Table 10: List of used reference state vectors for $v_0 = (e_1,\ f_1,\ e_2,\ f_2,\ e_3,\ f_3)$ with $n = 1, 2, 3$.

The simulation results for the state-dependent-noise EKF are shown for both the non standardised and standardised residuals. The results of these analysis are shown in the figure 8 for the non standardised residuals and in figure 44 for the standardised residuals.

For the correctly specified model, all rejection rates are close to the significance level. This is true for both the non standardised and standardised residuals. There are not exactly at $\alpha = 0.05$, but this is explained by the Monte Carlo uncertainty.

Both sets of residuals react for even the mild difference between the true and the wrong reference state vector. The rejection rates of the depth values in the mild scenario go up to at least 50%. The simplex-based methods even go up to a rejection rate of 100% while the sign depth methods react not as strongly. From the medium scenario onwards, all rejection rates of the depth methods are at a 100%.

This means, that all depth values detect even a relatively small difference between the true and assumed reference state reliably.

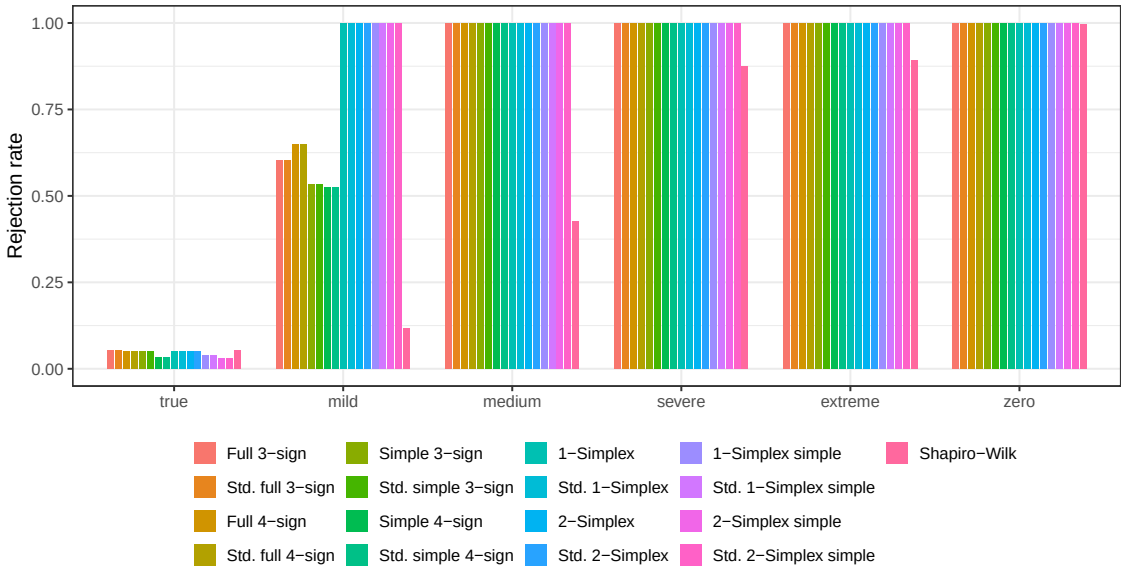

Figure 8: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the state-dependent-noise EKF and non standardised residuals.

The Shapiro–Wilk test also detects the wrong parameter but its increase in power is more gradual. Its rejection rate is approximately 14% for the mild scenario and increases to about 39% for the medium scenario. Only for the larger misspecifications does its power approach one. Consequently, the depth-based procedures identify the misspecification considerably earlier than the normality test.

The same behaviour can be observed for the standardised residuals of this EKF. This time the rejection rates of the sign depth methods for the mild scenario are even higher and lie at around 85%. This means that for standardised residuals the sign depth methods are even better to detect the wrong parameter even for small differences. The simplex depths and the Shapiro-Wilk test show the same rejection as for the non standardised residuals.

4.3.4 Influence of the reference state vector v_0 misspecification: constant-noise EKF

Now this simulation study for a misspecified reference state vector is analysed for the constant-noise EKF. Therefore, the rejection rates can be found in figure 9 for the non standardised residuals and in figure 45 for the standardised residuals. For

this EKF there is again no major difference in the rejection rates for the different residuals. Looking at the mild scenario, it is visible that the rejection rates for the sign depth of the non standardised residuals are lower that the rejection rates of the standardised residuals. Interestingly, the rejection rates of the simple sign depth methods of both residuals stay at around 90% for the medium and zero scenario.

The Shapiro-Wilk test only goes up to a rejection rate of 14% for the severe scenario. This means, that the test reacts far later to the wrong reference state parameter for the constant-noise EKF than it does for the state-dependent-noise EKF. For the zero scenario the Shapiro-Wilk test goes up to a rejection rate of about 65%.

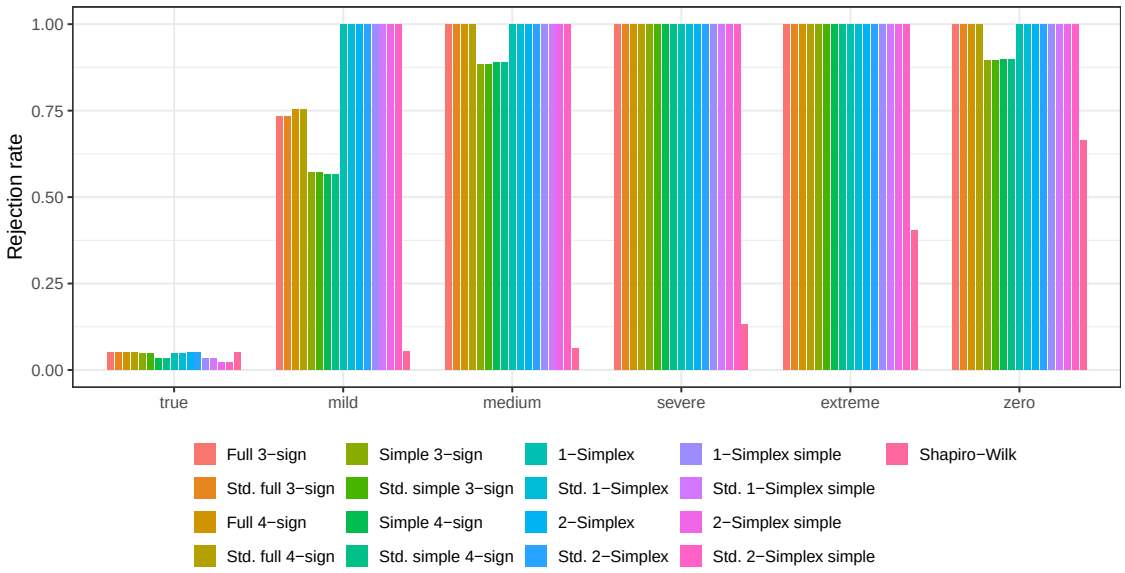

Figure 9: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the constant-noise EKF and non standardised residuals.

In summary, the depth value methods are really good at detecting an incorrectly specified reference state for both EKFs. The residuals, independent whether they are standardised or not, allow even mild differences from the true reference state to be identified with high probability. Across all considered scenarios, the depth-based procedures consistently achieve higher rejection rates than the Shapiro–Wilk test.

4.4 Robustness to outliers

In this section the robustness of the filter towards outliers is analysed. Therefore, first only positive additive outliers are considered. Afterwards two-sided additive

outliers are added to the simulated data and their influence is analysed.

4.4.1 Positive outliers

To evaluate the robustness of the Kalman filter, additive outliers (AO) are introduced into the measurement process. The state process is not modified and follows the state-space model described in the previous section. Only the observations are affected.

Without outliers, the observation model is given by

$$z_t = h(v_t) + e_t^z,$$

where $e_t^z \sim \mathcal{N}(0, \Sigma_z)$ denotes the measurement noise.

A constant offset of 1000 is added to the simulated observation vector at the selected time points. The same offset is added to every component of the observation vector. As a result, only the measurements are corrupted, while the underlying state remains unchanged. The observation model then becomes

$$z_t = \begin{cases} h(v_t) + e_t^z + 1000, & \text{if } t \in \mathcal{T}_{\text{AO}}^{(i)}, \\ h(v_t) + e_t^z, & \text{otherwise,} \end{cases}$$

where $\mathcal{T}_{\text{AO}}^{(i)}$ with $i = 1,\ 2,\ 3$ denotes the set of time points at which additive outliers are introduced into the observations. Three different outlier scenarios are considered. The outlier time points are randomly selected before the simulation and remain fixed during each simulation run. This allows the filter performance to be compared under different levels of measurement contamination. The first scenario contains two isolated outliers at the time points

$$\mathcal{T}_{\text{AO}}^{(1)} = \{10, 80\}.$$

The second scenario contains ten outliers at the time points

$$\mathcal{T}_{\text{AO}}^{(2)} = \{1,\ 14,\ 34,\ 39,\ 43,\ 51,\ 59,\ 68,\ 82,\ 87\},$$

which are randomly selected. The third scenario contains fifteen outliers with the time points

$$\mathcal{T}_{\text{AO}}^{(3)} = \{2,\ 6,\ 7,\ 15,\ 29,\ 37,\ 39,\ 42,\ 44,\ 47,\ 48,\ 53,\ 59,\ 73,\ 78\}.$$

These scenarios represent increasing levels of measurement contamination.

Since only the observations are affected, these disturbances are classified as additive outliers. The purpose of these scenarios is to investigate how isolated and repeated measurement errors influence the innovations and the resulting state estimates of the Kalman filter.

The figure 40 shows an example of an illustrative simulation for the first node run where ten additive outliers are added. It is clear to see that the influence of an outlier is not restricted to the measurement update at the corresponding time step. Due to the recursive nature of the Kalman filtering process, the corrupted measurement is incorporated into the state estimate and causes an inaccurate correction of the predicted state. As a result, the residuals remain significantly affected during the following time steps before returning to their nominal range. This indicates that the filter requires a certain recovery period to compensate for the erroneous measurement and restore an accurate state estimate.

The following figures show the rejection rates of the depth methods and the Shapiro-Wilk test for the different scenarios of outliers side by side. The rejection rates for the state-dependent-noise EKF are displayed in figure 10.

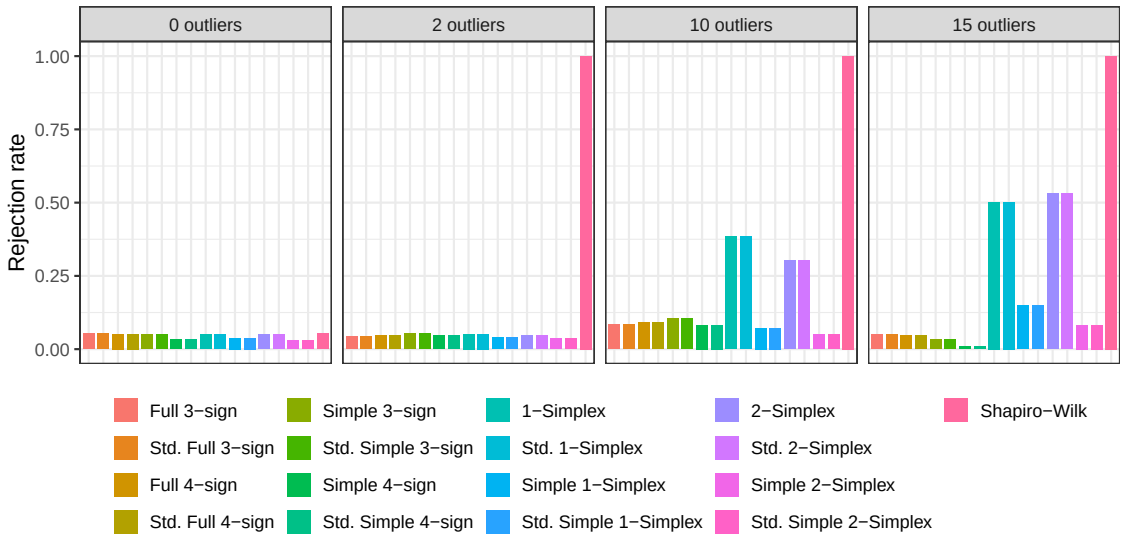

Figure 10: Rejection rates for the residuals of the state-dependent-noise EKF for positive additive outlier scenarios with 2, 10, and 15 outliers.

In this figure it is clear to see that already two outliers are enough to let the rejection rate of the Shapiro-Wilk test to go up to 100%. Meaning this test is especially sensitive to outliers. Since the sign-based depth values are purely based on the sign of the residuals, it is not surprising, that these tests do not change much over the scenarios. The simplex-based methods start to react with increasing their

rejection rates for ten and fifteen outliers. The rejection rate of the full and simple simplex depth go up to 50% for fifteen additive outliers. This behaviour is discussed in the following paragraph 'Influence of artificial outliers on the simplex depth'. The standardised simplex depth versions stay in comparison quite low.

The same results can be seen for the constant-noise EKF in figure 11. The sign-based depth do not show any big differences of the rejection rates. The rejection rates for the full sign depths go up to 10% for fifteen outliers. For this filter, the simplex methods also react to the outliers with a much higher rejection rate, but this time the maximum rejection rate is about 20%. This means, that the depth tests seem to be less influenced by the residuals with outliers when the constant-noise EKF is used.

Figure 11: Rejection rates for the residuals of the constant-noise EKF for positive additive outlier scenarios with 2, 10, and 15 outliers.

The standardised residuals of the EKFs and the rejection rates obtained from these residuals are shown in figure 47 and 48. The rejection rates of all considered methods behave the same. This means, that the standardisation of the residuals do not influence the rejection rate in this scenario as well.

Influence of artificial outliers on the simplex depth

Due to the dynamic nature of the filtering procedure, the addition of the outliers not only produced an extremely large residual at the contaminated observation but also caused a substantial negative residual at the next time point, typically of an approximately size of -90.

An example of the resulting simplices can be seen in the figures 35, 36, 37 and 38. These figures show an illustration of the simplices of the simplex depth for $K = 1$ on a subset of six observations generated with the state-dependent-noise EKF, including one positive outlier. The main panel shows the zoomed-in region used for the simplex depth, while the inset provides an overview of the complete subset and indicates the location of the zoomed region. Triangles containing the origin are highlighted in green, whereas triangles not containing the origin are outlined in red. In these graphics you can clearly see the influence of the outlier and the resulting smaller outlier.

This behaviour affects the simplex depth a lot. For example, the full 1-simplex depth is defined as the proportion of all triangles formed by triples of residuals that contain the origin. Under the null hypothesis, the residual vectors are approximately centered around the origin, implying that a relatively large proportion of these triangles contains the origin.

The introduction of artificial outliers changes the geometry of the residual cloud. Rather than observing a single isolated extreme point, the filtering procedure generates a characteristic pattern consisting of a very large positive residual followed by a pronounced negative residual. Resulting in residual vectors that are no longer approximately symmetrically distributed around the origin. Since the simplex depth depends on the geometric arrangement of all residual vectors, this distortion affects a large number of simplices.

Although for example only 10% of the residuals are directly contaminated, the effect on the simplex depth is considerably larger. For $N = 100$, the full 1-simplex depth is based on

$$\binom{100}{3} = 161,700$$

simplices. The number of simplices containing at least one contaminated residual equals

$$\binom{100}{3} - \binom{90}{3} = 161,700 - 117,480 = 44,220,$$

which corresponds to approximately 27.3% of all triangles. Hence, a relatively small fraction of contaminated observations influences a substantial proportion of the simplices contributing to the depth.

Since the rejection rule for the null hypotheses is given by

$$\sup_{\theta \in \Theta_0} d_K^j \left(r_1(\theta), \ldots, r_N(\theta) \right) < q_{K,j}(\alpha),$$

a reduction of the observed depth increases the probability of rejection. The simulated critical values $q_{K,j}(\alpha)$ are obtained under the uncontaminated null model and therefore correspond to approximately symmetric residual distributions. Artificial outliers together with the subsequent filter response alter the empirical residual distribution and may reduce the observed simplex depth relative to this reference distribution.

This indicates that the simplex depth is sensitive not only to the presence of extreme observations but also to the dynamic residual patterns generated by the filtering procedure after an outlier occurs.

4.4.2 Two-sided outliers

To investigate the influence of measurement outliers with varying directions, an additional outlier scenario is considered. The state process remains unchanged and follows the previously defined state-space model. Only the measurement process is modified at selected time points.

At predefined time points, an additional random shift is introduced into the observation vector. The modified observation model is given by

$$z_t = \begin{cases} h(v_t) + e_t^z + \eta_t, & \text{if } t \in \mathcal{T}_{\text{AO}}^{(i)} \text{ for } i = 1, 2, 3, \\ h(v_t) + e_t^z, & \text{otherwise,} \end{cases}$$

where η_t represents the outlier contribution. For each affected observation component, the shift is randomly selected from

$$\eta_t \in \{-1000, 1000\}.$$

Therefore, each component of the observation vector receives either a positive or a negative shift of scale 1000. The direction of the outlier is independent for each measurement component. An example of this can be seen in figure 39. This figure shows an example of ten added outliers to an illustrative run of the state-dependent-noise EKF. It is clear to see that the direction of the outliers are random and do not follow a specific pattern.

First, the rejection rate of the residuals obtained with the state-dependent-noise EKF are looked at in figure 12. This time the rejection rate of the depth-based methods only go up to maximal 10% for 10 and 15 outliers. It seems that the two-sided outliers do not effect the simplices as much as the one-sided outliers did. The Shapiro-Wilk tests rejection rate reacts the same as for one-sided outliers. These

results are not influenced by standardising the residuals of the state-dependent-noise EKF as shown in figure 49.

Figure 12: Rejection rates for the residuals of the state-dependent-noise EKF for two-sided additive outlier scenarios with 2, 10, and 15 outliers.

When looking at the residuals of the constant-noise EKF, not much difference can be seen to before. These rejection rates are shown in figure 13 and for the standardised residuals in figure 50. The depth-based rejection rate for this filter stay consistently around the significance level for both types of residuals. For the higher number of outliers the standardised simplex-based rates even get lower than for the scenario of no outliers. The Shapiro-Wilk test again performs as before and rejects the tests each time for even two outliers.

Overall, it can be noticed that one-sided residuals have a significant impact on the simplex-based method and the responding rejection rates. This changes, if two-sided additive outliers are introduced. The standardisation of residuals does not have any impact on the rejection rates of both filters. For all scenarios, where outliers are involved, the Shapiro-Wilk test react to this outliers and its rejection rate goes up to 100%.

Figure 13: Rejection rates for the residuals of the constant-noise EKF for two-sided additive outlier scenarios with 2, 10, and 15 outliers.

4.5 Analysis for an alternative parameter setting ($\theta_v = 0.9$)

In this section the studies from the previous sections are run again but this time the considered true state transition parameter is chosen as $\theta_{\text{true}} = 0.9$. For this parameter value, the model shows a strong temporal dependence, meaning that previous states have a much bigger influence on the current state than the reference state v_0.

4.5.1 Consequences of model parameter misspecification

The rejection rates for the scenario with a fixed true state transition parameter of $\theta_{\text{true}} = 0.9$ of the state-dependent-noise EKF are shown in figure 14. In this scenario, decreasing values of θ_{model} represent an increasing mismatch between the model used by the filter and the actual system dynamics. Since the previous analysis showed that the use of standardised residuals does not lead to a significant difference in the rejection rates, only the results based on the non-standardised residuals are considered for both EKF.

For the correctly specified model, where $\theta_{\text{model}} = \theta_{\text{true}} = 0.9$, all investigated methods show rejection rates close to the significance level of $\alpha = 0.05$ when applying the Monte Carlo uncertainty. This indicates that the tests maintain the expected rejection probability when the assumed model matches the true system dynamics.

However, in contrast to the previous simulation scenarios, the rejection rates

decrease as the mismatch between θ_{model} and θ_{true} increases. Most methods show rejection probabilities close to zero for smaller values of θ_{model}, indicating that the introduced model mismatch is not detected in this scenario. The decrease is particularly pronounced for the depth and simplex-based approaches, where the rejection rates drop below the nominal level even for small deviations from the true parameter.

The Shapiro-Wilk test shows a different behaviour. Its rejection rate remains approximately constant around the significance level of $\alpha = 0.05$ over the entire range of tested values. Therefore, similar to the previous experiments, it does not show a notable sensitivity to the introduced model mismatch, but it also maintains a stable rejection probability.

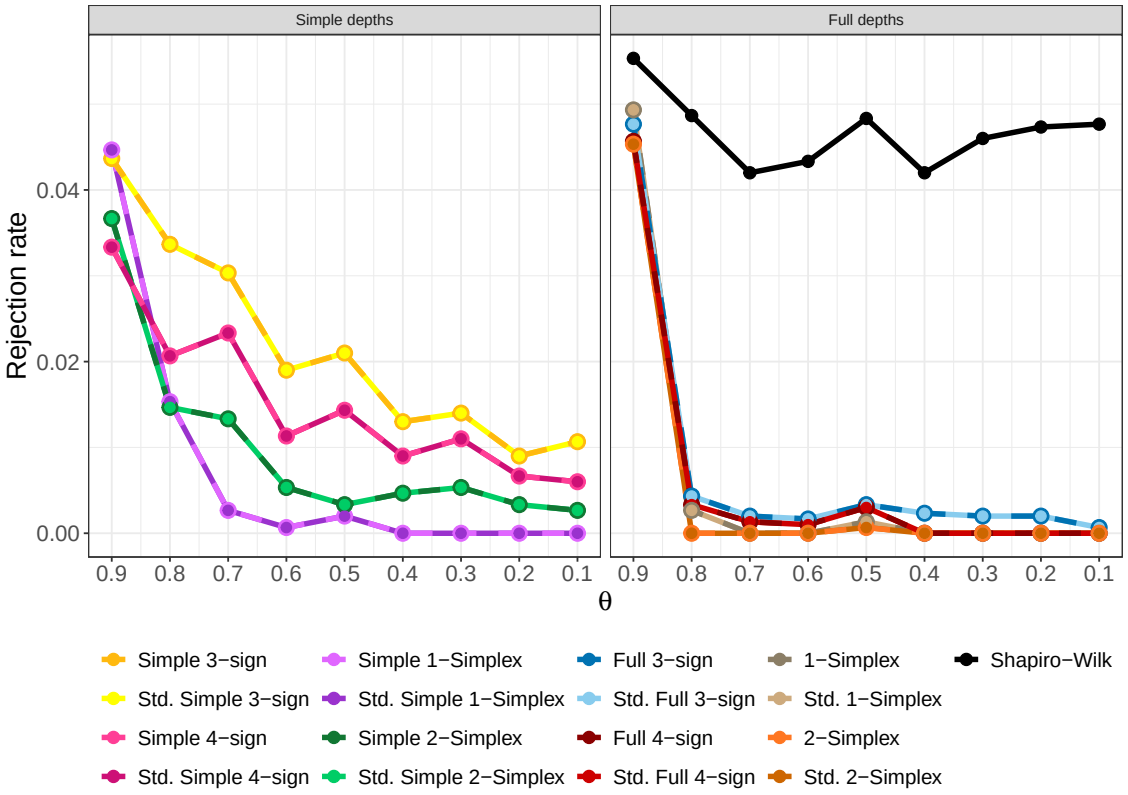

Figure 14: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the state-dependent-noise EKF and non standardised residuals (true parameter $\theta = 0.9$).

The results for the constant-noise EKF with $\theta_{\text{true}} = 0.9$ are shown in Figure 43.

For the correctly specified model, where $\theta_{\text{model}} = \theta_{\text{true}} = 0.9$, all investigated methods yield rejection rates close to the significance level of $\alpha = 0.05$. This indicates that the tests maintain an appropriate rejection probability when the assumed dynamics correspond to the true system dynamics.

Compared to the state-dependent-noise EKF, the rejection rates of the simple depth are generally lower for the constant-noise EKF. The overall behaviour, however, remains similar, suggesting that the direction of the model mismatch has a stronger influence on the detection performance than the specific noise model.

Overall, the results indicate that the investigated methods are sensitive to the direction of the model mismatch. While deviations from a smaller true state transition parameter were detected in the previous experiments, deviations from the true value $\theta_{\text{true}} = 0.9$ are not reflected by increasing rejection rates. This suggests that the detectability of a model mismatch depends on both the extent of the parameter deviation and the underlying system dynamics.

For this alternative parameter $\theta_v = 0.9$ the sensitiveness for a wrong reference state vector is also analysed. The figure 15 shows the rejection rates for the state-dependent-noise EKF and the figure 46 the ones for the constant-noise EKF. For both EKFs the simplex-based methods detect the wrong parameter setting even for the mild scenario. In these cases the rejection rate is at 100%.

Here, the sign-based methods also detect the wrong parameter, but the rejection rates only reach 100% for the extreme scenario when it comes to the state-dependent-noise EKF and for the severe scenario when it comes to the constant-noise EKF. This means, that if the state transition parameter is higher, the rejection rates of the sign-based depth methods are not as high for small differences in the reference state vectors.

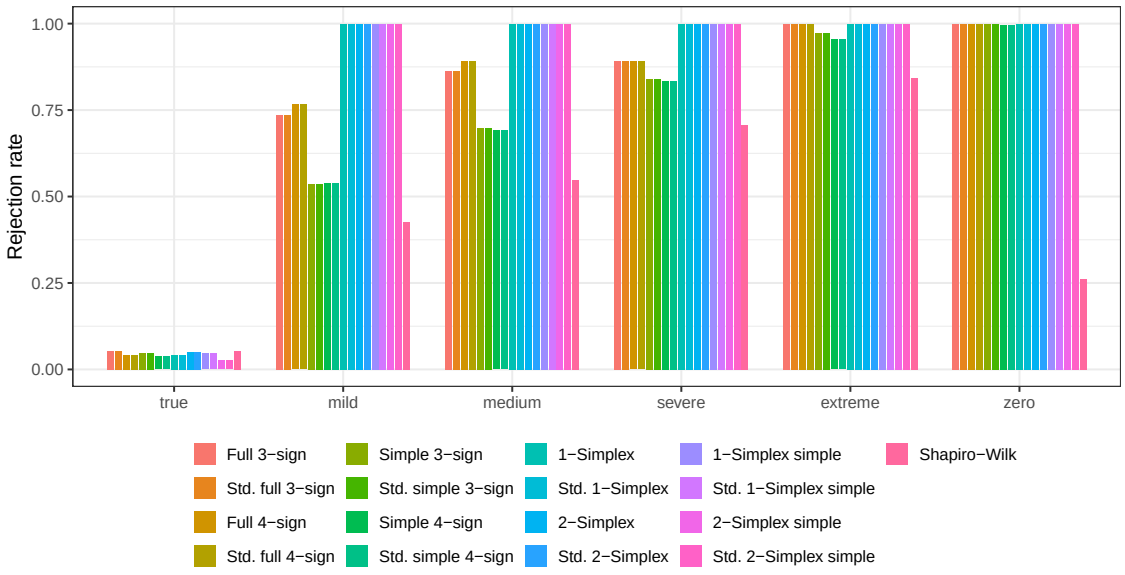

Figure 15: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the state-dependent-noise EKF and non standardised residuals with the parameter $\theta = 0.9$.

The Shapiro-Wilk test shows the same behaviour as before. Its rejection rates are much higher for the state-dependent-noise EKF than for the constant-noise EKF. However, there is one exception which is the zero scenario. Here the rejection rate of the Shapiro-Wilk test only lies at about 25%. The rejection rate for the constant-noise EKF for this scenario lies at about 70%.

Overall, the simplex-based methods are again more sensitive towards the wrong reference vector that the sign-based methods. However, all looked at methods detect the differences in the parameter quite good.

4.5.2 Robustness to outliers

In this section the influence of the outlier scenarios which have been defined before will be analysed for the different state transition parameter.

Positive outliers

First, the one-sided additive outliers are analysed. Therefore, the rejection rates obtained with the residuals of the state-dependent-noise EKF are shown in figure 16. This time the standardised residuals are not looked at because the results from before show that the standardisation does not has an influence on the rejection rate for this analysis.

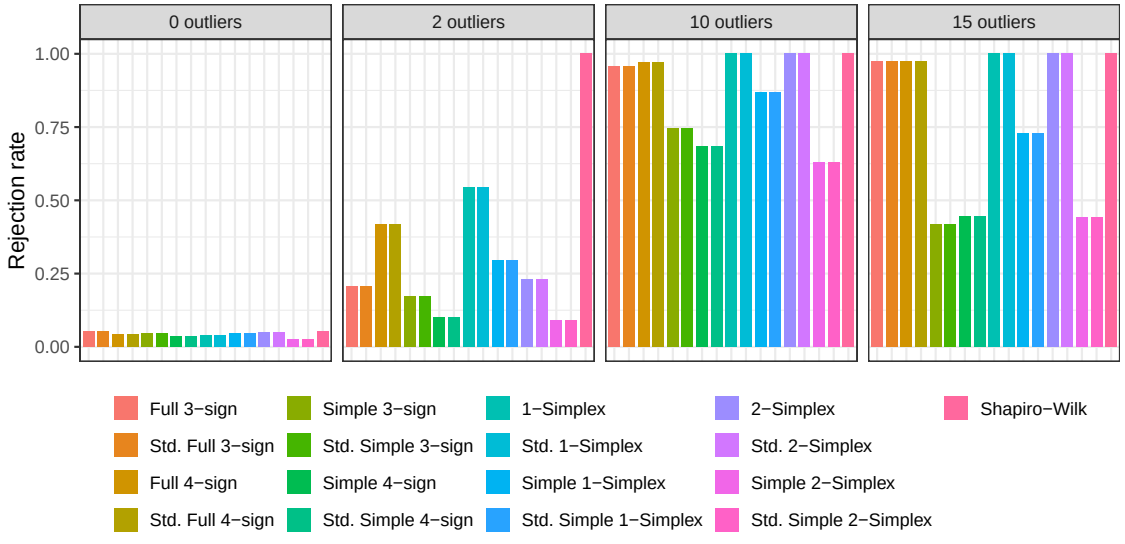

Figure 16: Rejection rates for the residuals of the state-dependent-noise EKF with $\theta = 0.9$ for positive additive outlier scenarios with 2, 10, and 15 outliers.

This time the rejection rates of all methods go up as soon as outliers are added. If only two positive outliers are added, the full 4-sign depth and the 1-simplex depth

with their standardised versions achieve the highest rejection rates with around 63%. The Shapiro-Wilk tests rejection rate is 100% as soon as outliers are introduce. For more outliers, the rejection rates of the full sign-methods and the full simplex-methods as well as their standardised versions lie at 100%. The simplified methods are less sensitive.

Similar to the results for the state-dependent-noise EKF, all methods have rejection rates for the constant-noise EKF close to the significance level when no outliers are added (see figure 17). With an increasing number of outliers, the rejection rates increase noticeable for most methods. The full sign-based methods, the full simplex-based methods and the Shapiro–Wilk tests show the highest rejection rates, whereas the simple methods remain less sensitive. Overall, the figure shows that the ability to detect non-normality improves as the number of outliers increases.

In comparison to the case where the state transition vector was chosen as $\theta_v = 0.1$, these rejection rates are much higher. This means that the outliers have a much bigger effect for this models. This can be explained by the model definition for v. When the state transition parameter $\theta_v = 0.9$ is close to one, the weight assigned to the reference state vector v_0 becomes very small. As a result, the current state v_t depends mainly on the previous state v_{t-1}. Therefore, an outlier in v_{t-1} has a strong influence on v_t, as most of its effect is carried over to the next time step. The impact of an outlier can spread through the process over several later observations because of this high level of persistence.

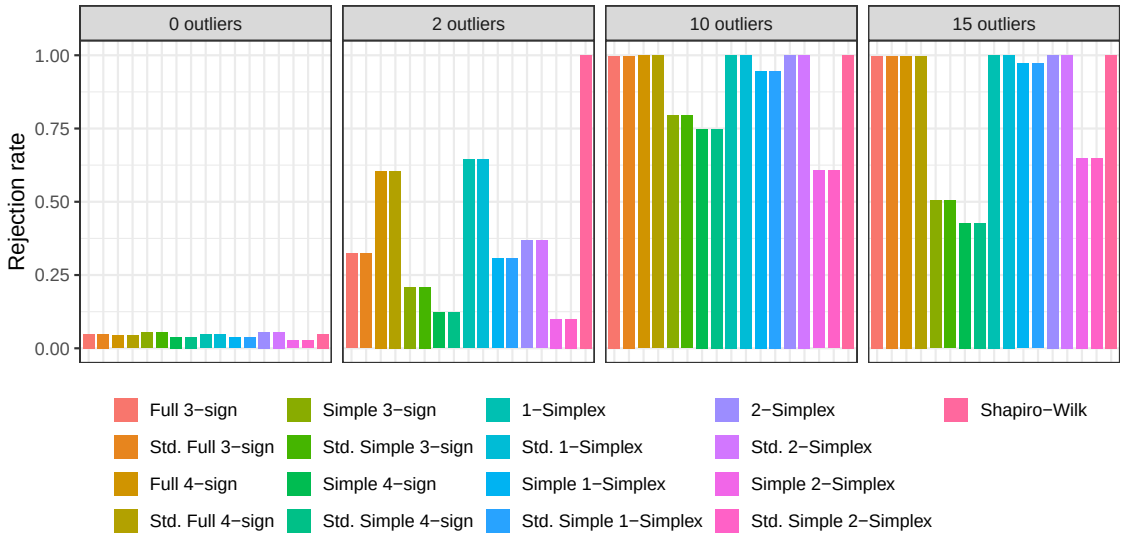

Figure 17: Rejection rates for the residuals of the constant-noise EKF with $\theta = 0.9$ for positive additive outlier scenarios with 2, 10, and 15 outliers.

Two-sided outliers

For the two-sided additive outliers added to the residuals of the state-dependent-noise EKF, all methods react to even two outliers with very high rejection rates (see figure 18). For the scenario with only two outliers, all rejection rates lie above 50%.

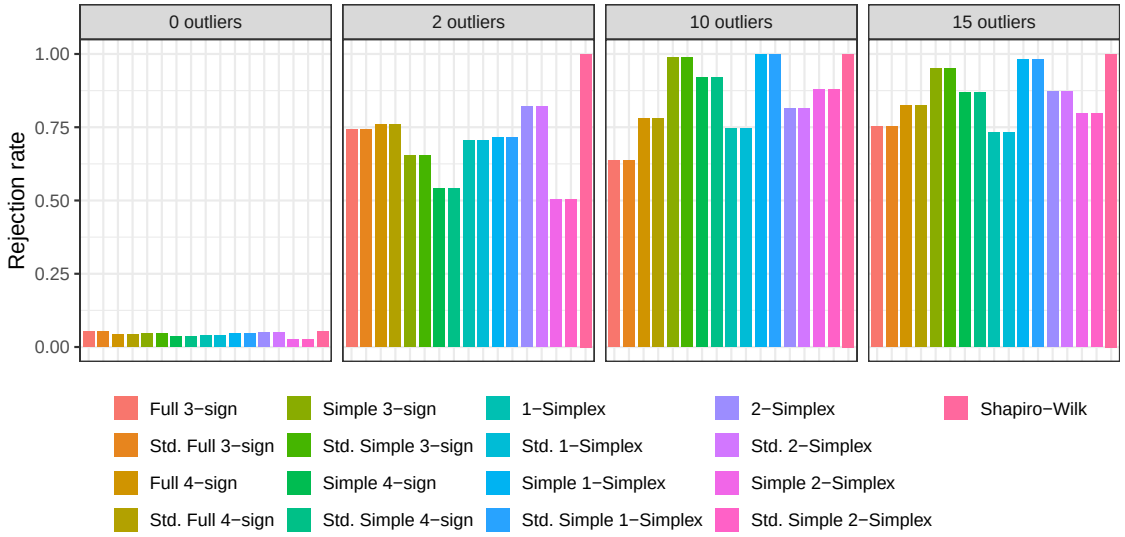

Figure 18: Rejection rates for the residuals of the state-dependent-noise EKF with $\theta = 0.9$ for two-sided additive outlier scenarios with 2, 10, and 15 outliers.

The figure shows that the rejection rates raise simultaneously. This means, that the outliers have an equal effect on the depth methods. For fifteen outliers, all rejection rates lie at least at 75%.

For the constant-noise EKF the rejection rates look a bit different, but the overall behaviour stays the same, as can be seen in figure 19. For this filter the rejection rates do not raise as much as for the state-dependent-noise EKF for two outliers. The maximum rejection rate in this scenario is 70%. For the scenarios with more that two outliers, the rejection rates behave like the ones from the state-dependent-noise filter.

Overall, it can be said, that outliers have much more effect on a model with a high state transition parameter. This is because the current state of the model then mainly depends on the previous state. Therefore the outliers have a lot of influence on the process.

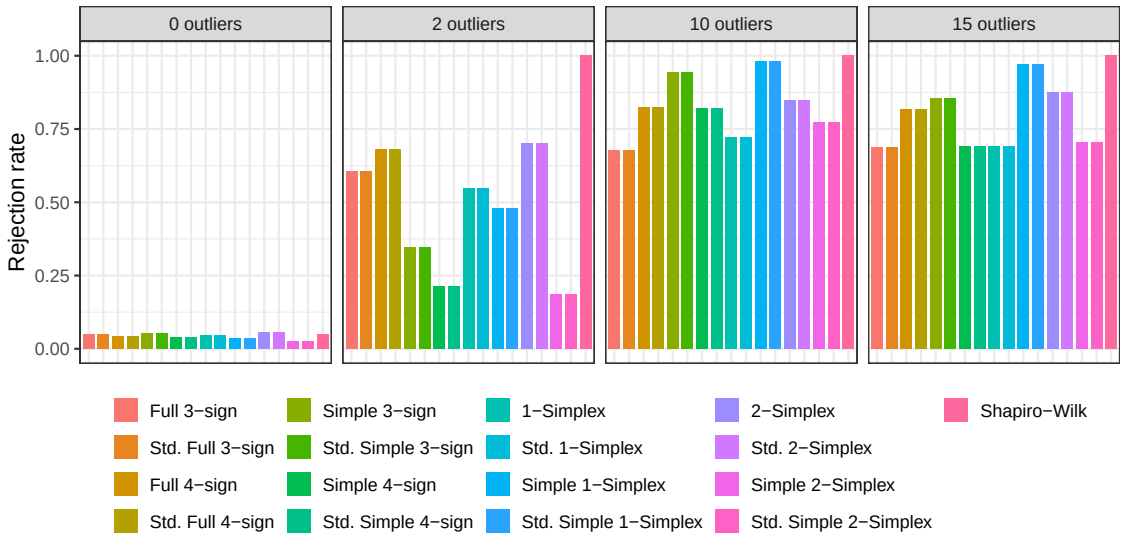

Figure 19: Rejection rates for the residuals of the constant-noise EKF with $\theta = 0.9$ for two-sided additive outlier scenarios with 2, 10, and 15 outliers.

5 Application to real data

This chapter shows how the suggested state estimation method can be applied to measurements from a real low-voltage distribution network. The idea is to try out the filtering approaches in real-life situations, not just on computer-generated data.

Therefore, this chapter is divided into seven sections. The first section explains the data preparation and provides a descriptive analysis of the considered nodes of the data. This is followed by a general description about the model set up of the considered models. The next four sections explain and analyse the considered models. A comparison of the models with each other can be found in the last section of this chapter.

5.1 Data preparation and descriptive analysis of the nodes

This section first describes the data preparation for the models and gives an descriptive analysis of the considered data. As described in chapter 2, the original network consists of fourteen nodes arranged in a radial structure. For clarity, the nodes have been renumbered using a consecutive numbering scheme, as shown in figure 20. The correspondence between the original and new node labels is provided in table 11. This renumbering was introduced solely to simplify the notation throughout the remainder of this thesis and does not affect the underlying electrical model.

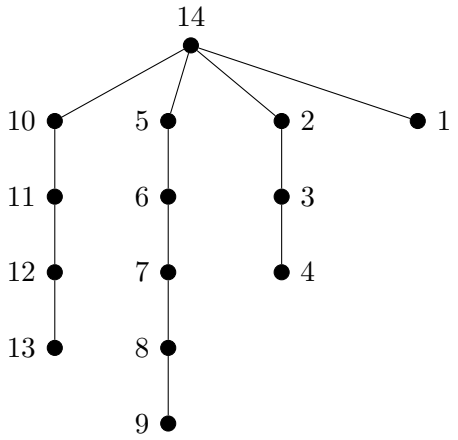

Figure 20: Graph representation of the electrical network with renumbered nodes.

Old ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
New ID	1	2	13	14	9	8	5	10	3	12	11	6	4	7

Table 11: Mapping of old node numbers (from original graph) to new node numbers (right to left, top-down).

Although measurements are available for the entire network, the following analysis only focused on three nodes. These correspond to nodes 1, 2 and 10 in the renumbered network and are directly connected to the slack node (see figure 21). Node 10 is referred to as Node 3 in the following. Restricting the study to these nodes reduces computational complexity while preserving the essential characteristics of the estimation problem. Active and reactive power measurements were available for each node, resulting in six observed variables: P_1, Q_1, P_2, Q_2, P_3, and Q_3. The application study is based on these variables.

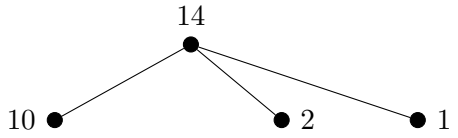

Figure 21: The electric network considered in this work.

Before the statistical methods that are described in chapter 3 are applied to the real data, the considered data itself is shortly looked at. Therefore boxplots for the considered components are shown in figure 22. This figure shows that the values of the reactive power components Q are much smaller that the values of the active power components P. The values of the node one, so P_1 and Q_1 have the highest values in comparison. Furthermore, the boxplots show that each component of the

considered data is right skewed. All components display outliers to the right of their distribution, but especially the component Q_3 got a lot of outliers in comparison to the other components.

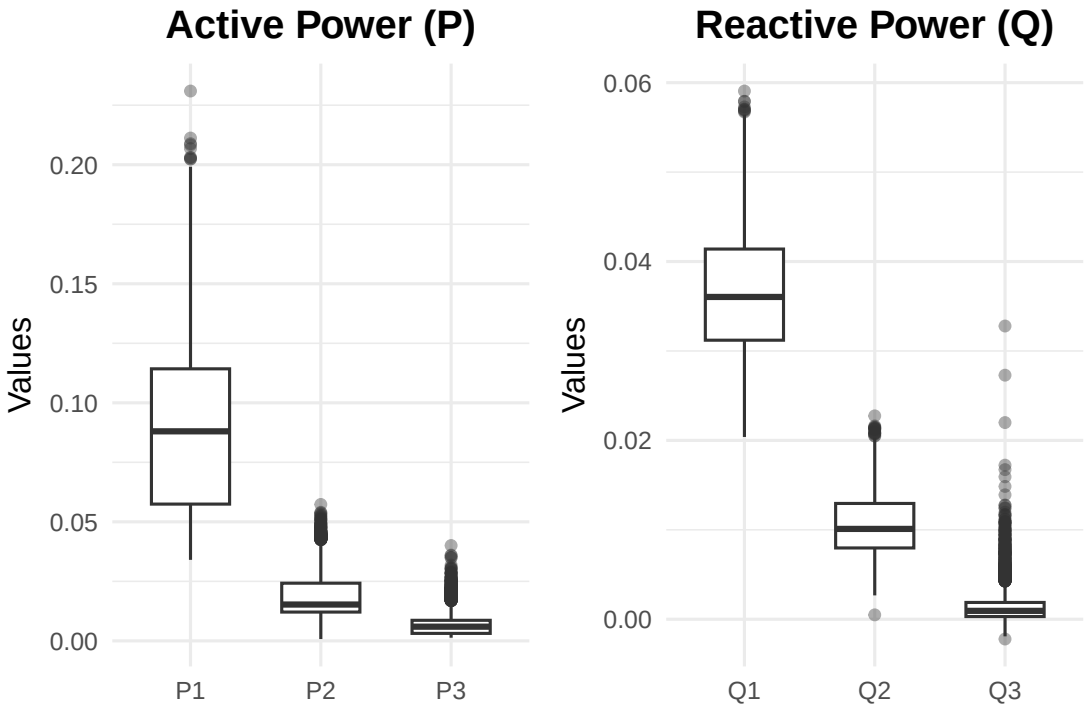

Figure 22: Boxplots of the consindered data in the following analysis (the first two weeks of the whole data set).

For the following application of the statistical methods the nodes should be independent of each other. To verify this assumption, a correlation matrix is displayed in figure 23. The correlation matrix shows that the strongest correlations occur between the active and reactive power measurements at the same node. In particular, the pair $(P_2,\ Q_2)$ exhibits a very high correlation of 0.89, while the corresponding correlations at nodes 1 and 3 are 0.72 and 0.49. Correlations between measurements from different nodes are generally weaker, indicating that the relationship between active and reactive power is strongest locally. Overall, the correlation structure suggests that the measurements contain both node-specific dependencies and weaker interactions across the network. To simplify the application and reduce the complexity of the following models, the components are assumed to be independent throughout the following analysis.

Figure 23: Correlation matrix of the components of the considered nodes.

5.2 General model setup

For the application study a period of continuous observations over the first fourteen days observed, with 336 hourly measurements taken for each variable is looked at. The original data are recorded every fifteen minutes. To reduce the number of observations, the data was aggregated to hourly values. This allows a longer observation period, covering more days, to be included in the analysis while maintaining a manageable dataset. The used data set for the following analysis contains both weekday and weekend consumption patterns. To model realistic data and ensure consistency with the simulation study presented in the previous chapter, each observation was given an independent Gaussian additive noise term. The standard deviation of this noise was set to 0.5% of the absolute measurement value. Consequently, it is assumed that the measurement covariance matrix Σ_z is known throughout the estimation procedure.

For some models the smoothed load profiles for the data are used. These are calculated with the definitions that are explained in section 3.2. These load profiles were calculated for the whole data set, but because only a subsection of the data set is later used, the smoothed profiles data was shortened accordingly. In figure 24 the observed data from the electrical grid and the smoothed load profiles are shown.

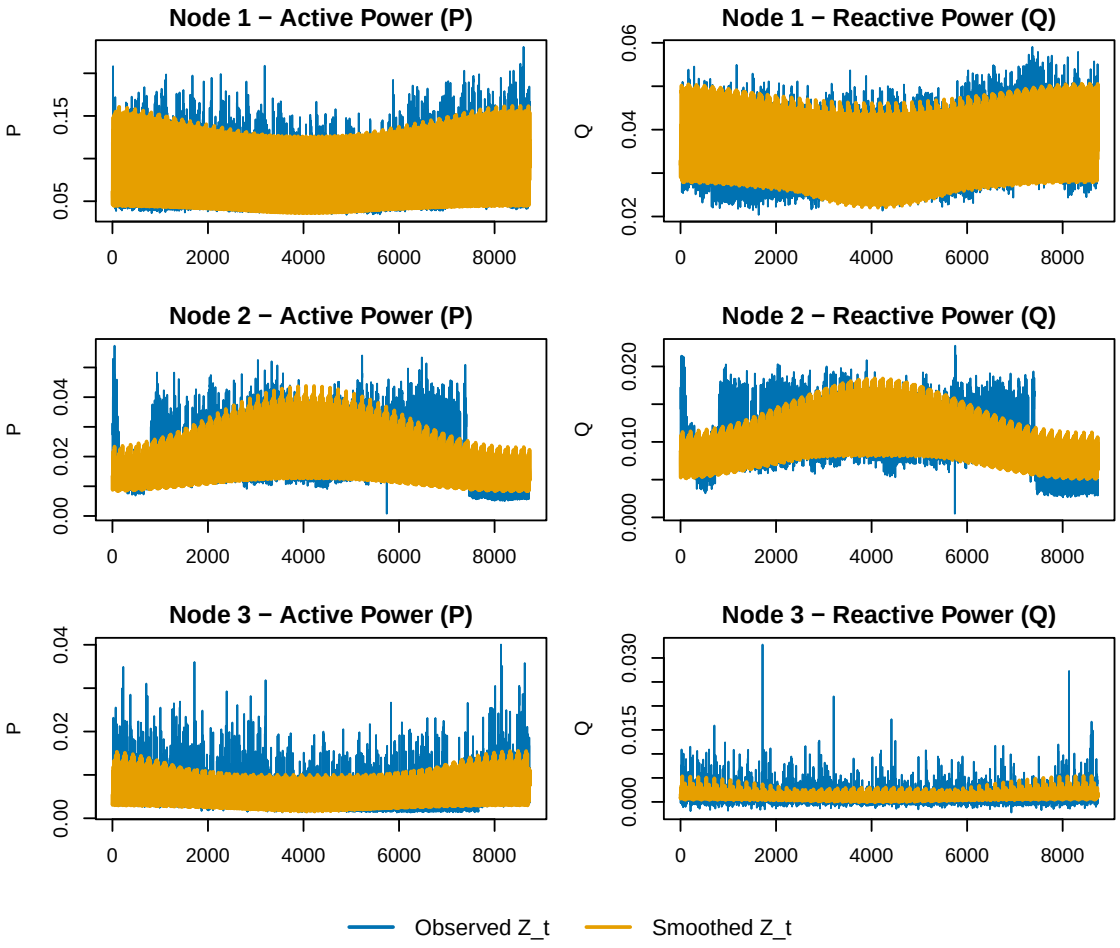

Figure 24: Observed data vs smoothed load profiles for the whole data set.

Parameter estimation and noise assumptions

For all considered models, the process and measurement noise assumptions are specified consistently. The process noise is assumed to follow a normal distribution,

$$E_t^v \sim \mathcal{N}(0, \Sigma_v),$$

where the covariance matrix is parameterised as

$$\Sigma_v = \sigma_v^2 \mathbb{I}_{2N \times 2N}.$$

Only one process variance parameter is estimated for all state components. This reduces the number of unknown parameters and avoids estimating individual noise levels for each voltage component.

The measurement noise is introduced to represent measurement uncertainty. The standard deviation is chosen proportional to the magnitude of the measured power

values:

$$\sigma_{z,i,t} = 0.005 \cdot |Z_{i,t}|.$$

Therefore, the measurement covariance matrix at time t is given by

$$\Sigma_z(t) = \text{diag}(\sigma^2_{z,1,t}, \ldots, \sigma^2_{z,2N,t}).$$

The unknown model parameters are estimated using maximum likelihood. For each parameter vector

$$\vartheta = (\theta_v, \log(\sigma_v^2)),$$

the Kalman filter is applied and the innovation likelihood is evaluated. The negative log-likelihood function is defined in section 3.3.

The logarithm of the process variance is optimized instead of the variance itself to guarantee positivity. The final parameter estimates are determined for every model individually.

The initial reference state vector v_0 is selected based on typical operating conditions of an electrical power system. In power flow formulations, active and reactive power represent the fundamental quantities describing the power exchange within the grid Frank and Rebennack 2016, p. 1175. Therefore, the chosen initial values should be within a physically reasonable operating range. The initial active power values are selected close to one per unit, representing a normalised operating level, while the reactive power values are initialized close to zero. This reflects an operating condition in which the power exchange is assumed to be primarily determined by active power, while reactive power is not the dominant component. These initial values provide a suitable starting point for the estimation procedure, while the actual states are subsequently adjusted based on the available measurements.

5.3 Model A: Random walk state-space model

The random walk model is used as a baseline model. It assumes that the state has no preferred operating point. Therefore, the future state only depends on the previous state and a random disturbance.

The state equation is

$$v_t = v_{t-1} + e_t^z.$$

This corresponds to

$$\theta_v = 1.$$

The process noise is defined as

$$e_t^z \sim \mathcal{N}(0,\ \sigma_v^2 \mathbb{I}_{6\times 6}).$$

The random walk model does not use information about typical load conditions. Therefore, it provides a simple reference for evaluating the benefit of additional load information.

For this model, θ_v is fixed and only the process variance is specified.

The chosen process variance is:

$$\hat{\sigma}_v^2 = 1.00 \cdot 10^{-7}.$$

The figure 25 shows the calculated NIS values for this model. Model A shows a lower-than-expected mean NIS value of 4.662, compared to the theoretical value of 6. This indicates that the filter is likely overestimating the uncertainty of its predictions. Almost 50% of the NIS values are below the lower confidence bound, suggesting that the estimated covariance is too large and the filter is operating too conservatively. The percentage of values above the upper bound is slightly higher than expected, indicating that occasional larger innovations still occur. All in all, the random walk model as a baseline model performs quite alright.

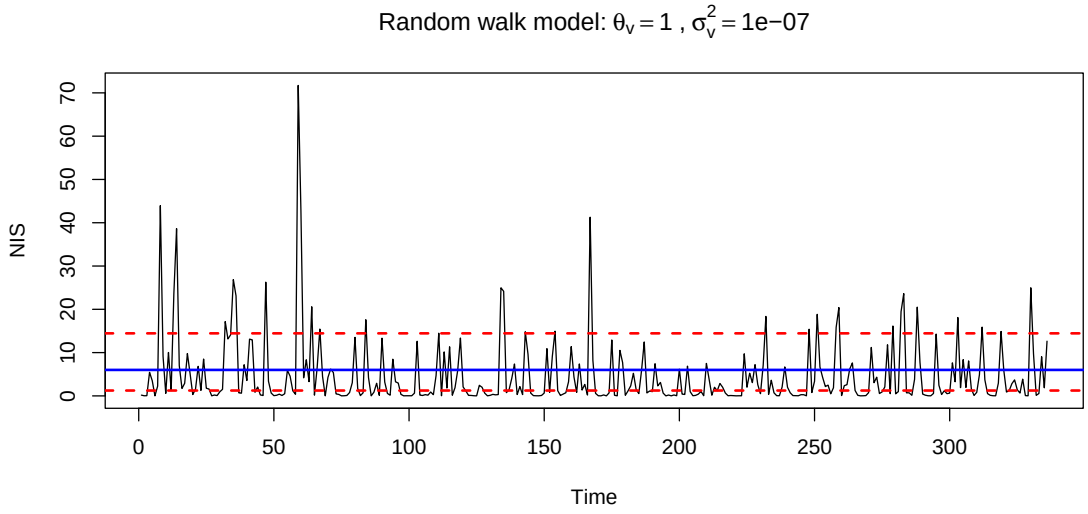

Figure 25: NIS values over time for the random walk model.

Table 38 summarises the results of the univariate depth-based tests for the random walk model of this model. The full K-sign depth does not reject the null hypothesis for most nodes, regardless of whether the standardised or non-standardised versions are considered. The one time where it is rejected for the $\alpha = 0.05$ level for the full 4-sign depth and its standardised version for P_1. However, after applying the Bonferroni correction, this rejection disappears.

In contrast, the simplified K-sign depth detects significant deviations at a few nodes. For K=3, the null hypothesis is rejected for the components P_2 and Q_2, both at the $\alpha = 0.05$ significance level and after the Bonferroni correction. For the simplified 4-sign depth, additional rejections are observed for Q_1 and Q_2 at the $\alpha = 0.05$ level, whereas only component P_2 remains significant after the Bonferroni adjustment. As shown in table 39, the bivariate K-simplex depth shows no rejections for any component pair. Overall, these results suggest that the random walk model provides a sufficient description of the data, with only a few local deviations detected by the simplified K-sign depth.

5.4 Model B: Mean-reverting state-space model with a constant reference state

The random walk model does not include any information about normal operating conditions. However, electrical loads usually fluctuate around typical values.

Therefore, model B introduces a mean-reverting behaviour. The state is attracted towards a constant reference state.

The state transition equation is

$$v_t = \theta_v v_{t-1} + (1 - \theta_v)v_0 + w_t.$$

The reference state is constant:

$$v_0 = \begin{bmatrix} 0.999 & 0.001 & 0.999 & 0.001 & 0.999 & 0.001 \end{bmatrix}^T.$$

The parameter θ_v determines the strength of the mean reversion. If θ_v is close to one, the model behaves similar to a random walk. Lower values force the state to return faster to the reference state.

The process noise is defined as

$$w_t \sim \mathcal{N}(0, \sigma_v^2 \mathbb{I}_{6\times 6}).$$

The parameters are estimated by maximum likelihood. The estimated parameters are:

$$\hat{\theta}_v = 0.988, \quad \hat{\sigma}_v^2 = 7.68 \cdot 10^{-8}.$$

This model achieves a mean NIS value of 6.021, which is very close to the expected value of 6. The NIS values are displayed in figure 26. However, the distribution of the NIS values does not fully match the expected chi-square distribution, as only 39.9% of the values are within the confidence band. A large fraction of values falls below the lower bound, while 14.3% exceed the upper bound. This suggests that although the average consistency is good, the uncertainty characterization is not optimal and the filter covariance may not accurately represent the actual estimation errors.

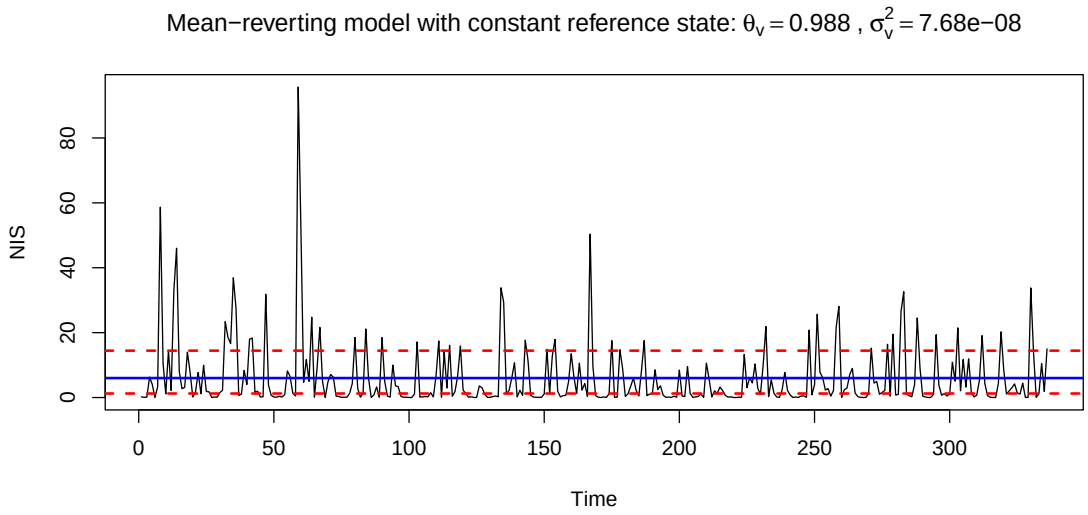

Figure 26: NIS values over time for the mean−reverting model with a constant reference state.

The tables 40 and 41 show the test results for this model. In contrast to the random walk model, most depth tests reject the null hypothesis. For the univariate analysis (refer to table 40), all depth measures detect significant differences at the components of node 2 and 3. At the $\alpha = 0.05$ significance level, component Q_1 is also rejected by most methods, whereas component P_1 is consistently not rejected. These conclusions remain unchanged after applying the Bonferroni correction, except for the full K-sign depth, where Q_1 is no longer significant. The bivariate results in table 41 show a similar pattern. Both versions of the full K-simplex depth reject the null hypothesis for all node pairs, while the simplified K-simplex depth rejects the

pairs $P_2,\ Q_2$ and $P_3,\ Q_3$ but not $P_1,\ Q_1$. In summary, the residuals show significant differences from the assumed reference distribution, as reflected by the large number of rejections across the different depth-based methods.

5.5 Model C: Mean-reverting state-space model with seasonal load profiles

Electrical loads usually show periodic behaviour. The expected load during nighttime differs from the expected load during daytime.

A constant reference state cannot represent these variations. Therefore, model C introduces a time-dependent seasonal reference profile.

The state transition model becomes

$$v_t = \theta_v v_{t-1} + (1 - \theta_v) v_{0,t} + w_t.$$

The time-dependent reference state $v_{0,t}$ is generated from the historical load data. Instead of using one constant reference value, the average load for each hour of the day is calculated and used as the corresponding reference state.

This results in a repeating 24-hour profile:

$$v_{0,t} = v_{0,t+24}.$$

For example, the first reference state of the profile is

$$v_{0,1} = \begin{bmatrix} 0.0555 & 0.0308 & 0.0143 & 0.0095 & 0.0059 & 0.0011 \end{bmatrix}^T.$$

The initial state of the Kalman filter is defined separately:

$$v_{0,init} = \begin{bmatrix} 0.999 & 0.001 & 0.999 & 0.001 & 0.999 & 0.001 \end{bmatrix}^T.$$

This value only initializes the filter and does not represent the expected load behaviour. The difference between $v_{0,init}$ and $v_{0,t}$ is caused by the independent initialization of the state estimation. Due to the high value of θ_v, the influence of the reference state in one time step is small. Therefore, the state converges gradually towards the seasonal profile instead of immediately adopting its values.

The process noise allows deviations from this typical behaviour:

$$w_t \sim \mathcal{N}(0, \sigma_v^2 \mathbb{I}_{6\times 6}).$$

Compared to model A, the reference state is no longer constant. The model can follow systematic daily variations while still applying mean reversion.

The estimated parameters are:

$$\hat{\theta}_v = 0.999, \quad \hat{\sigma}_v^2 = 5.41 \cdot 10^{-7}.$$

The here introduced model provides the most consistent NIS behaviour (see figure 27). The mean NIS value of 5.983 is very close to the theoretical expectation of 6, and 98.2% of the values remain within the confidence bounds. This indicates that the filter predictions are highly consistent with the measurements and that the estimated uncertainty is well matched to the actual errors. The very small percentages below and above the bounds suggest a stable filter behaviour with few unexpected innovations. However, the unusually high percentage of NIS values within the confidence bounds may suggests that the innovation sequence may have a lower variance than expected from the theoretical chi-square distribution.

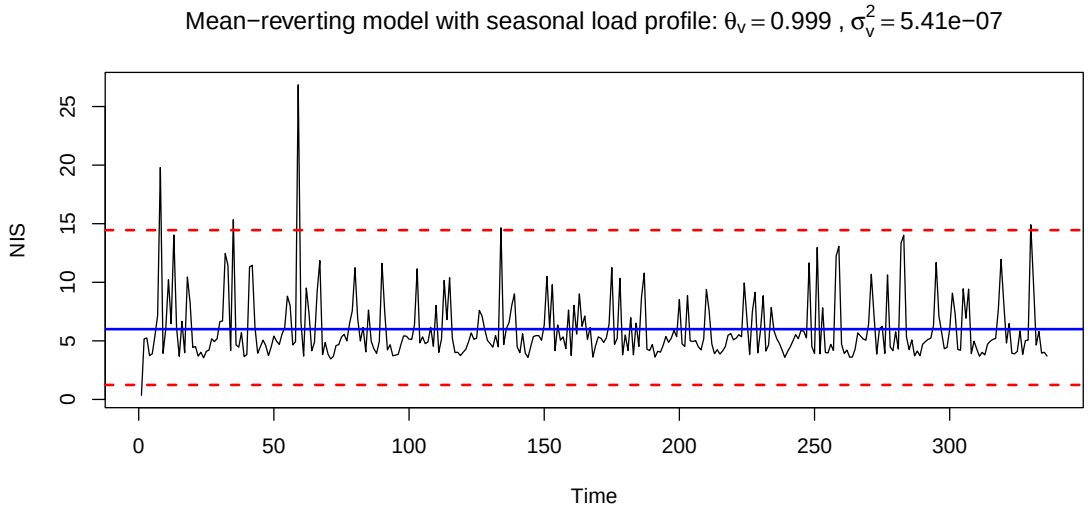

Figure 27: NIS values over time for the mean−reverting model with seasonal load profiles.

The depth test results for the mean-reverting state-space model with seasonal load profiles are presented in the tables 42 and 43. In this case, all depth-based tests indicate significant differences of the residuals from the assumed reference distribution. The univariate results in table 42 show that both the full and simplified K-sign depths reject the null hypothesis for all nodes. This result is consistent across different values of K and remains unchanged after applying the Bonferroni correction. The bivariate analysis in table 43 shows the same pattern, with all

node pairs being rejected by both simplex depth methods. Overall, the results suggest that the residuals contain systematic differences that are not explained by the assumed model structure.

5.6 Model D: Mean-reverting state-space model with a time-varying seasonal reference state

Model C assumes that the seasonal profile is directly available in the state-space. However, the measurements are related to the states through a nonlinear measurement function. Therefore, model D reconstructs the seasonal reference state from the measurement space.

The seasonal reference state is obtained from

$$h(v_{0,t}) = \hat{z}_t,$$

where $\hat{z}_t$ is the smoothed seasonal load profile.

Because the measurement function is nonlinear, Newton's method is used:

$$v_{0,t}^{(k+1)} = v_{0,t}^{(k)} - H(v_{0,t}^{(k)})^{-1}(h(v_{0,t}^{(k)}) - \hat{z}_t).$$

For example, the first reconstructed reference state is

$$v_{0,1} = \begin{bmatrix} 1.0020 & -0.0002 & 1.0001 & -0.00001 & 1.0000 & 0.0000 \end{bmatrix}^T.$$

The reconstructed states are close to the nominal operating point. The active power components are close to one, while the reactive power components are close to zero. This corresponds to a typical operating condition of the electrical system (Frank and Rebennack 2016, p.1175).

Compared to model C, the reference state is not calculated from an average load profile. Instead, it is obtained by solving the inverse measurement problem and therefore directly matches the smoothed measurements.

The reconstructed state is then used as the time-dependent reference state in the mean-reverting model:

$$v_t = \theta_v v_{t-1} + (1 - \theta_v) v_{0,t} + w_t.$$

The advantage of this approach is that the seasonal information is consistent with the nonlinear measurement model. The reference state changes over time and represents the expected hidden system state.

The parameters are estimated by maximum likelihood. The estimated parameters are:

$$\hat{\theta}_v = 0.450, \quad \hat{\sigma}_v^2 = 4.83 \cdot 10^{-8}.$$

The NIS values for this model are shown in figure 28. This model has a mean NIS value of 6.013, which is close to the expected value. Nevertheless, only about half of the values are within the confidence band, while a large number of samples fall below the lower threshold and some exceed the upper threshold. This indicates that the mean consistency is acceptable, but the uncertainty model is not fully accurate. The filter appears to be somewhat conservative most of the time, but it still produces occasional larger-than-expected innovations.

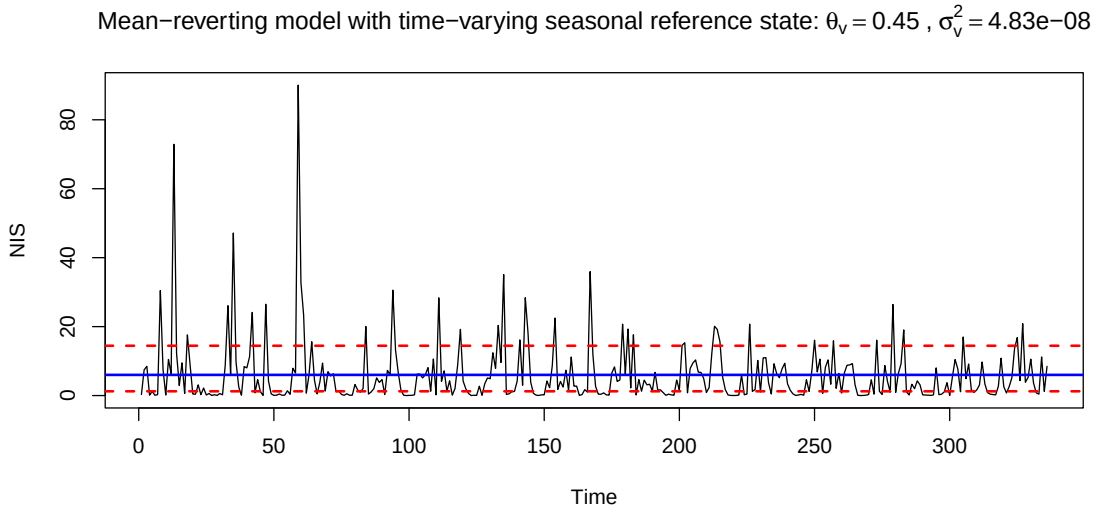

Figure 28: NIS values over time for the mean−reverting model with time−varying seasonal reference state.

The depth test results for this methods are presented in tables 44 and 45. The univariate tests in table 44 show that the null hypotheses is rejected for all depth methods for the components of the first two nodes. For the components of the third node, the null hypotheses is never rejected after the Bonferroni correction is applied to the significance level. The same pattern is shown for the bivariate test (see table 45). Overall, the results indicate that the time-varying seasonal reference state captures part of the systematic structure in the data, although some differences in the residuals remain.

5.7 Comparison of the models

This section compares the different state-space models from the previous sections with each other. The goal is not to identify a perfect model, but to determine which model provides a suitable description of the observed data.

The main differences between the model structures are shown in table 12. In the following, model A is used as the baseline model. It is a random walk model and does not use a reference state. The models B, C and D are mean reverting state-space models. Model B uses a constant reference state, while models C and D use seasonal reference information. The difference between model C and model D is that the reference state of model D is reconstructed in the state-space using the measurement model.

Model	State transition	Reference state
A	$v_t = v_{t-1} + w_t$	None
B	$v_t = \theta_v v_{t-1} + (1-\theta_v) v_0 + w_t$	Constant v_0
C	$v_t = \theta_v v_{t-1} + (1-\theta_v) v_{0,t} + w_t$	Seasonal profile $v_{0,t}^{\text{seasonal}}$
D	$v_t = \theta_v v_{t-1} + (1-\theta_v) v_{0,t} + w_t$	Reconstructed $v_{0,t} = h^{-1}(\hat{z}_t)$

Table 12: Overview of the investigated state-space models.

The estimated parameters and the NIS results are given in table 13. Models B and C have values of $\hat{\theta}_v$ close to one, which means that these models put a higher weight on the last prediction and less weight on the reference state vector. The estimated value of model D is with $\hat{\theta}_v = 0.450$ way lower. This shows that the time-varying reference state has a stronger effect on the state evolution.

Model	$\hat{\theta}_v$	$\hat{\sigma}_v^2$	Mean NIS	Within band	Below	Above
A	1	$1.00 \cdot 10^{-7}$	4.662	41.1%	49.7%	9.2%
B	0.988	$7.68 \cdot 10^{-8}$	6.021	39.9%	45.8%	14.3%
C	0.999	$5.41 \cdot 10^{-7}$	5.983	98.2%	0.3%	1.5%
D	0.450	$4.83 \cdot 10^{-8}$	6.013	50.9%	37.5%	11.6%

Table 13: Estimated model parameters and NIS evaluation.

The mean NIS values of all models are close to the expected value of 6. However, not the expected 95% of the NIS values lie within the confidence band. Model C shows the best NIS behaviour, with 98.2% of the values inside the confidence band. This shows that the seasonal reference profile gives a good estimate of the uncertainty. Models A and B have more values below the confidence band, which indicates that the estimated uncertainty is slightly too large. Model D has a similar mean NIS value, but only 50.9% of the values are inside the confidence band.

The empirical process noise variances, which are shown in table 14, are calculated via

$$\hat{w}_t = v_t - \theta_v v_{t-1} - (1 - \theta_v) v_{0,t}.$$

Models A, B, and C have very similar values. This means that the additional reference information mainly changes the expected state behaviour and not the total amount of process variation. Model D has a slightly lower process noise variance. This indicates that the time-varying reference state explains more of the changes in the system.

Model	Mean process noise variance: mean($Var(\hat{w}_t)$)
A	7.8150×10^{-8}
B	7.7130×10^{-8}
C	7.8140×10^{-8}
D	4.3337×10^{-8}

Table 14: Empirical process noise variance obtained from the reconstructed process noise $\hat{w}_t$.

The depth-based tests offer another approach to evaluate the model performance. Model A shows only a few rejected tests. Models B and C show more rejected tests, even though model C has the best NIS results. Model D reduces the number of rejected tests compared to model C, especially for the last node.

The differences between the NIS results and the depth-based tests show that both methods evaluate different aspects of the residuals. The NIS analyses whether the estimated uncertainty matches the measurement errors. The depth-based tests tests whether the residuals follow the assumed reference distribution. Therefore, rejected depth-based tests do not necessarily mean that the model is not suitable. It just shows that the residuals maybe have some additional structures that are not described by the models and therefore do not fit for the structure that depth-values test.

Overall, the results show that the models with seasonal information describe the load behaviour better than the random walk baseline. Model C provides the best NIS results, while model D reduces some of the differences found by the depth-based tests. Therefore, the most suitable model depends on the evaluation method and the desired application.

6 Summary

This thesis analyses the use of K-sign depth as a criteria for comparing state-space models of electrical power grids. These models provide a flexible framework for modelling the dynamic behaviour of complex systems, but selecting the most suitable model can be challenging. Conventional model selection criteria often depend on distributional assumptions and can be affected by outliers or model misspecifications, which motivates the search for more robust alternatives.

The focus of this thesis is on evaluating whether K-sign depth can offer a robust and reliable method for evaluating the performance of different state-space models. It investigates whether depth values can tell the difference between well-specified and misspecified models, and whether they remain informative in the presence of outliers. Additionally, it compares different methods of K-sign depth to determine the most suitable depth measure for model evaluation.

To answer these questions, the analysis is carried out in two stages. First, a simulation study is performed in which different state-space models are analysed under controlled conditions. This allows the effects of correct and incorrect parameter assumptions, different model specifications and outliers on the model residuals and resulting depth values to be examined systematically. Secondly, the methods are applied to data from a realistic electrical power network. Several state-space models with different assumptions regarding state dynamics and seasonal structure are fitted and compared based on the depth values of their residuals. This combination of simulation and real-world application allows the theoretical properties and practical usefulness of K-sign depth for model comparison to be evaluated.

The simulation study shows that the investigated depth-based methods perform as expected when the model assumptions are correctly specified. A state-dependent-noise and constant-noise extended Kalman filter is considered. When a misspecification of the state transition parameter occurs, the depth-based values react to this misspecification. The rejection rates increase with more pronounced differences between the assumed and the true model parameters. This indicates that the proposed methods are able to detect deviations from the underlying state-space model. The simplex-based depth measures show the highest sensitivity to parameter misspecification. At the same time, all investigated depth measures are able to identify incorrect reference state vectors with high probability, even for relatively small deviations. This only goes for the case when the parameter used in the state process is smaller than the parameter used in the prediction model. The classical Shapiro-Wilk test is unable to detect the misspecified parameter. Is a misspecification for

the reference vector looked at, all considered methods detect this. Even though, the Shapiro-Wilk test reacts slower than the depth-based methods.

The robustness analysis shows that the different depth measures react differently to additive outliers. The Shapiro-Wilk test is strongly affected by even a small number of outliers. In contrast, the sign-based depth methods remain more stable because they only rely on the signs of the residuals. The simplex-based methods are more sensitive to additive outliers when these are entirely positive, which can be explained by the influence of extreme observations on the geometric structure of the residuals. Therefore, the different depth measures provide complementary information: sign-based depth offers a more robust assessment, whereas simplex-based depth is more sensitive to changes in the residual distribution.

The application to electrical power grid data shows that the considered models which incorporate seasonal information as load profiles describe the behaviour of the real data sufficiently. A mean-reverting state-space model with seasonal load profiles achieves the best Normalized Innovation Squared values, which were used to compare the considered models. However, the corresponding curve of these values suggest that maybe the innovation sequence of the data has a lower variance than expected. Another factor that should be mentioned is, that the reference state vector which was constructed using the load profiles differs quite a lot from the assumed reference vector which was used by the other considered models. The depth test results for the mean-reverting state-space model with seasonal load profiles indicate significant differences of the residuals from the assumed reference distribution.

Additionally, a mean-reverting state-space model with time-varying seasonal reference state was considered. This model constructs the reference state for every step using the measurement model. When evaluating this model via the Normalized Innovation Squared, the mean of the values looks good, but only half of the considered values lie within the confidence band. The large number of values below this confidence band indicate that the uncertainty model is not fully accurate. However, the depth test results indicate that the time-varying seasonal reference state model captures part of the systematic structure in the data, although some differences in the residuals remain.

These real data applications show that state-space models work well for electrical power grid data. Models which include additional information as load profiles capture the underlying behaviour of the data even better. Both the considered depth-based methods and the Normalized Innovation Squared proved to be a useful evaluation methods.

Even though the depth-based methods seem to be quite robust and reliable

when it comes to the simulation studies, they reach their limitations when they are applied to residuals of state-space models of real data applications. Here they seem to work better if the model is able to describe the underlying behaviour to a high quality. They are however still able to provide more distinguished methods for model comparison than the classical Shapiro-Wilk test. Therefore, the depth-based tests are a promising approach as a comparison criteria.

A critical point that should be discussed is, that the empirical quantiles of the depth values depend on the number of observations. That means that if a large data set is considered, the computation of these quantiles for the depth tests can be computationally demanding and is therefore unpractical. Especially for the simplex-based depth these quantiles may be not computable due to the high amount of considered simplices.

In conclusion, this thesis shows promising results with the depth-based methods for state-space models used on simulated data and real electrical power grid data. Furthermore, an analysis of state-space models used in an application shows which approach to the modelling of electrical power grid data seems most promising.

Moving forward, it would be interesting to analyse further state-space models, especially models that include additional information about the load profile behaviour of electrical power systems. The aim would be to identify models that describe the underlying structure of the data well and at the same time achieve good results according to both the Normalized Innovation Squared and the depth-based methods.

Furthermore, the combination of different evaluation criteria could be analysed in more detail. The results of this thesis show that the Normalized Innovation Squared and the depth-based methods provide different information about the models performance. While the Normalized Innovation Squared evaluates the consistency of the assumed uncertainty model, the depth-based methods provide additional information about differences in the residual distribution. Therefore, using both approaches together could provide a more detailed evaluation of state-space models.

Additionally, the influence of other types of model misspecifications and different state-space model structures could be analysed. This would provide further insight into the strengths and limitations of depth-based methods and help to evaluate their potential as a criteria for state-space model comparison.

List of Symbols

Data, Dimensions, and Indices

Symbol	Brief description
d	Calendar-day index.
$d(t)$	Calendar day at time t.
h	Hour index.
$h(t)$	Hour at time t.
i	General observation, component, or hypothesis-test index.
j	Index identifying simplified or full depth method.
k	Summation or iteration index.
m	Simulation index for the quantile calculation or dimension of the observation vector in the likelihood expression.
n	Node or observation index.
n_v	Dimension of the state vector used in the NEES distribution.
n_z	Dimension of the measurement vector used in the NIS distribution.
N	Number of nodes (excluding slack) in the electrical network, number of residuals or observation size.
p	Dimension of the residual vectors in the simplex-depth definition.
t	Time index.
T	Number of observed time points.
w	Day-of-the-week index.
$w(t)$	Day of the week at time t.

Depth Measures and Depth-Based Tests

Symbol	Brief description
d	Generic depth function.
$d(\theta)$	Depth evaluated for the residuals generated under parameter value θ.
d_K^j	General K-depth function, where j denotes the simplified or full version.
$d_K^{u,F}(r_1,\ldots,r_N)$	Full K-sign depth of the residuals $r_1,\ldots,r_N$.
$d_K^{u,S}(r_1,\ldots,r_N)$	Simplified K-sign depth of the residuals $r_1,\ldots,r_N$.
$d_K^{F}(r_1,\ldots,r_N)$	Full K-simplex depth of the residuals $r_1,\ldots,r_N$.
$d_K^{S}(r_1,\ldots,r_N)$	Simplified K-simplex depth of the residuals $r_1,\ldots,r_N$.
$D_K^{u,F}(r_1,\ldots,r_N)$	Standardised full K-sign depth of the residuals $r_1,\ldots,r_N$.
$D_K^{u,S}(r_1,\ldots,r_N)$	Standardised simplified K-sign depth of the residuals $r_1,\ldots,r_N$.
$D_K^{F}(r_1,\ldots,r_N)$	Standardised full K-simplex depth of the residuals $r_1,\ldots,r_N$.
$D_K^{S}(r_1,\ldots,r_N)$	Standardised simplified K-simplex depth of the residuals $r_1,\ldots,r_N$.
F	Superscript denotes the full version of a depth measure.
K	Parameter of the K-sign or K=simplex depth.
$\mathcal{K}$	Set of considered depth orders, such as $\{1,2,3,4\}$.
0_p	Origin of the p-dimensional Euclidean space.
$q_d(\alpha)$	Critical quantile for a generic depth-based test.
$q_{K,j}(\alpha)$	Empirical α-quantile of the depth measure with order K and version j.
$R_d(\alpha)$	Rejection indicator for the depth-based test associated with depth function d.
S	Superscript denotes the simplified version of a depth measure.
$\mathbb{S}(r_1,\ldots,r_{p+1})$	Open simplex spanned by the residual vectors $r_1,\ldots,r_{p+1}$.
u	Superscript identifying a univariate depth measure.
Θ_0	Parameter space specified by the null hypothesis of a depth-based test.

Electrical Network and Power Quantities

Symbol	Brief description
A	Admittance matrix of the electrical network.
a_{nk}	Admittance coefficient between nodes n and k.
a_{nk}^*	Complex conjugate of the admittance coefficient a_{nk}.
$\overline{A}$	Real-valued block representation of the admittance matrix.
$\overline{A}_{nk}$	Real-valued block associated with admittance coefficient a_{nk}.
$\overline{A}_{n,\text{slack}}$	Admittance block connecting node n to the slack node.
b_{nk}	Imaginary part of the admittance coefficient a_{nk}.
e_n	Real part of the complex voltage at node n.
e_n^0	Initial real voltage part at node n.
f_n	Imaginary part of the complex voltage at node n.
f_n^0	Initial imaginary voltage part at node n.
g_{nk}	Real part of the admittance coefficient a_{nk}.
j	Imaginary unit.
P	Active electrical power.
$P_n(t)$	Active power at node n and time t.
$P_{n,d,h}$	Active power at node n, calendar day d, and hour h.
Q	Reactive electrical power.
$Q_n(t)$	Reactive power at node n and time t.
$Q_{n,d,h}$	Reactive power at node n, calendar day d, and hour h.
S_n	Complex power injection at node n.
U_n	Matrix representing of the complex voltage v_n.
$\overline{U}_n$	Alternative matrix representing of the complex voltage v_n.
v_n	Complex voltage $e_n + jf_n$ at node n.

Functions, Operators, and Distributions

Symbol	Brief description
$\lvert \cdot \rvert$	Componentwise absolute-value operator when applied to a vector.
$\binom{N}{K}$	Number of possible subsets of size K selected from N elements.
Cov	Covariance function.
$\det(\cdot)$	Matrix determinant.
$\operatorname{diag}(\cdot)$	Operator constructing a diagonal matrix from its arguments.
$\operatorname{Diag}(\cdot)$	Operator constructing a block-diagonal matrix.
$\mathbb{E}$	Expected value.
$F^{-1}(\alpha, n)$	Inverse cumulative distribution function of the chi-square distribution with n degrees of freedom.
$h(v)$	Nonlinear link function computing voltage states to power injections.
$H(v)$	Jacobian matrix of the nonlinear link function h evaluated at state vector v.
$\mathbb{I}_{q\times q}$	Identity matrix of dimension $q \times q$.
$\mathbb{1}(\cdot)$	Indicator function.
$\operatorname{Im}(\cdot)$	Imaginary-part operator.
$K(\cdot)$	Kernel function used in kernel density estimation.
$\log(\cdot)$	Natural logarithm.
$\mathcal{N}(\mu, \Sigma)$	Normal distribution with mean μ and covariance matrix Σ.
$\mathcal{N}_q(\mu, \Sigma)$	q-dimensional multivariate normal distribution.
$\lVert \cdot \rVert$	Euclidean norm.
$\mathbb{P}(\cdot)$	Probability operator.
$\operatorname{Re}(\cdot)$	Real-part operator.
$\mathbb{R}$	Real numbers.
sup	Supremum.
$(\cdot)^{-1}$	Matrix inverse or inverse-function operator.
$(\cdot)^{\top}$	Vector or matrix transpose.
$\chi^2(n)$	Chi-square distribution with n degrees of freedom.

Symbol	Brief description
$\Phi^{-1}(\alpha)$	α-quantile of the standard normal distribution.

Kalman Filter and State-Space Models

Symbol	Brief description
a_t	Predicted state at time t.
b_t	Predicted observation at time t.
c_0	Initial filtered state.
c_t	Updated state after observing z_t.
C_0	Initial state covariance matrix.
C_t	Updated state covariance matrix at time t.
$H_t(a_t)$	Jacobian matrix of h_t evaluated at the predicted state a_t.
K_t	Kalman gain matrix at time t.
L_t	Lower-triangular Cholesky factor of the innovation covariance S_t.
R_t	Predicted state covariance matrix.
S_t	Predicted innovation covariance matrix at time t.
V_t	State process vector containing the real and imaginary voltage components at time t.
v_t	Realisation of the voltage-state vector at time t.
v_0	Initial or reference state vector.
$v_{0,t}$	Time-varying reference state vector.
$v_{0,\text{init}}$	State vector used only to initialise the Kalman filter.
w_t	Vector of explanatory variables included in the general state-transition equation or empirical process noise variance in the real-data state-space models.
Z_t	Observation process vector containing active and reactive power values at time t.
z_t	Observed vector at time t.
$z_{1:t-1}$	Collection of observations from time 1 through time $t-1$.

Symbol	Brief description
$\Theta_{v,t}$	State transition matrix describing the autoregressive effect of past states.
$\Theta_{w,t}$	Coefficient matrix describing the effect of explanatory variables.
θ_v	State transition parameter.

Load Profiles and Regression

Symbol	Brief description
d_0	Calendar day representing the beginning of spring.
$e^P_{n,d,h}$	Regression-error term for active power at node n, day d, and hour h.
$e^Q_{n,d,h}$	Regression-error term for reactive power at node n, day d, and hour h.
$\mathbf{x}$	Sinusoidal regression-basis function.
$\mathbf{x}(d)$	Regression-design vector evaluated for calendar day d.
$\mathbf{X}_w$	Design matrix containing the regression-basis vectors for weekday w.
$\mathbf{Y}_{n,w,h}$	Response vector used to estimate the load profile at node n, weekday w, and hour h.
$\beta^P_{n,w,h}$	Regression-coefficient vector for the active-power load profile.
$\beta^Q_{n,w,h}$	Regression-coefficient vector for the reactive-power load profile.
$\hat{\beta}_{n,w,h}$	Least-squares estimate of a load-profile regression-coefficient vector.
$\hat{P}_{n,d,h}$	Smoothed active-power value at node n, day d, and hour h.
$\hat{Q}_{n,d,h}$	Smoothed reactive-power value at node n, day d, and hour h.
$\hat{Z}_t$	Smoothed load-profile vector at time t.
$\hat{z}_t$	Smoothed seasonal measurement vector used to reconstruct a reference state.
$v^{\text{seasonal}}_{0,t}$	Seasonal reference-state profile used in the state-space model.

Noise, Covariance, and Model Parameters

Symbol	Brief description
e_t^v	Gaussian process noise term.
e_t^z	Gaussian measurement noise term.
E_t^v	Noise term used in the real-data model formulation.
Σ_t^v	Process noise covariance matrix at time t.
Σ_t^z	Measurement noise covariance matrix at time t.
Σ_v^{const}	Constant process noise covariance matrix.
Σ_v^{state}	State-dependent process noise covariance matrix.
$\Sigma_z(t)$	Time-dependent measurement noise covariance matrix.
Σ_z^{const}	Constant measurement noise covariance matrix.
Σ_z^{state}	State-dependent measurement noise covariance matrix.
σ_v^2	Process-noise variance.
$\hat{\sigma}_v^2$	Estimated process-noise variance.
σ_z^2	Measurement-noise variance.
$\sigma_{z,i,t}$	Measurement-noise standard deviation for component i at time t.

Residual Diagnostics and Statistical Tests

Symbol	Brief description
a_i	Shapiro–Wilk coefficient of the ith ordered observation.
a'	Row vector containing the Shapiro–Wilk coefficients.
b_1	Lower confidence bound for a NEES or NIS consistency test.
b_2	Upper confidence bound for a NEES or NIS consistency test.
E_i	Event that a type I error occurs in hypothesis test i.
G	Covariance matrix of order statistics in the Shapiro–Wilk test.
H_0	Null hypothesis of a statistical test.
m	Vector of expected normal-order statistics in the Shapiro–Wilk test.

Symbol	Brief description
$P(k)$	State-estimation error covariance matrix used in the NEES statistic.
$P(k \mid k-1)$	Predicted state covariance at time k before incorporating the current measurement.
r_t	Innovation or prediction residual at time t.
r_i	Individual residual used in a depth or normality calculation.
$r_{(i)}$	ith ordered residual in the Shapiro–Wilk statistic.
$\bar{r}$	Sample mean of the residuals.
$R(k)$	Measurement noise covariance matrix in the NIS formulation.
$S(k)$	Innovation covariance matrix used in the NIS statistic.
W	Shapiro–Wilk test statistic.
$z(k)$	Measurement vector at time k in the NIS formulation.
$\hat{z}(k)$	Predicted measurement vector at time k.
α	Significance level of a statistical test.
α_{Bonf}	Bonferroni-adjusted significance level.
$\epsilon_v(k)$	Normalized Estimation Error Squared at time k.
$\epsilon_\nu(k)$	Normalized Innovation Squared at time k.
$\hat{v}(k)$	Estimated state vector at time k.
$\tilde{v}(k)$	State-estimation error $v(k) - \hat{v}(k)$.
$\nu(k)$	Innovation or measurement residual at time k.

Simulation, Estimation, and Outliers

Symbol	Brief description
$\hat{f}(v)$	Estimated Kernel density function for value v.
h_{bw}	Bandwidth used in kernel density estimation.
M	Number of Monte Carlo replications or simulated depth values.
p	True rejection probability in a Monte Carlo experiment.
$\hat{p}$	Empirical Monte Carlo rejection probability.
R	Number of rejections observed in a Monte Carlo experiment.

Symbol	**Brief description**
$\mathrm{SE}(\hat{p})$	Monte Carlo standard error of the estimated rejection probability.
$\mathcal{T}_{\mathrm{AO}}^{(i)}$	Set of time points contaminated by additive outliers in scenario i.
$\hat{v}_{t\|t-1}$	State predicted for time t using information available through time $t-1$.
$\hat{v}_{t-1\|t-1}$	Updated state estimate at time $t-1$.
$\tilde{v}_0$	Misspecified reference state vector assumed by the filter.
$\hat{w}_t$	Reconstructed process noise vector used for empirical variance calculations.
y_t	Solution of the triangular system $L_t y_t = r_t$ in the likelihood calculation.
η_t	Additive outlier contribution applied to the observation vector at time t.
θ	Generic model parameter or parameter value considered in a depth test.
θ_{model}	State-transition parameter assumed by the Kalman filter.
θ_{true}	State-transition parameter used to generate the simulated process.
ϑ	Parameter vector used for maximum-likelihood estimation.
$\ell(\theta)$	Log-likelihood evaluated at parameter value θ.

Bibliography

Bar-Shalom, Yaakov, Xiao-Rong Li, and Thiagalingam Kirubarajan (2001). *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Ltd. ISBN: 9780471221272. DOI: https://doi.org/10.1002/0471221279.fmatter_indsub. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471221279.fmatter_indsub.

Bates, Douglas and Martin Maechler (2021). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.3-3. URL: https://CRAN.R-project.org/package=Matrix.

Borchers, Hans W. (2025). *pracma: Practical Numerical Math Functions*. R package version 2.4.6. DOI: 10.32614/CRAN.package.pracma. URL: https://CRAN.R-project.org/package=pracma.

Cavus, Emre (2025). "Dynamics of Modern Power Systems: Challenges and Opportunities". In: *Electronics* 14.6, p. 1159. DOI: 10.3390/electronics14061159. URL: https://www.mdpi.com/2079-9292/14/6/1159.

Durbin, James and Siem Jan Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd ed. Oxford Statistical Science Series. Oxford: Oxford University Press. ISBN: 9780199641178.

Field, Andy (2009). *Discovering Statistics Using SPSS*. 3rd ed. Los Angeles [i.e. Thousand Oaks, Calif.]: SAGE Publications. ISBN: 978-1-84787-906-6.

Frank, Stephen and Steffen Rebennack (2016). "An introduction to optimal power flow: Theory, formulation, and examples". In: *IIE Transactions* 48.12, pp. 1172–1197. DOI: 10.1080/0740817X.2016.1189626. URL: https://doi.org/10.1080/0740817X.2016.1189626.

Genz, Alan and Frank Bretz (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag. ISBN: 978-3-642-01688-2.

Gilbert, Paul and Ravi Varadhan (2019). *numDeriv: Accurate Numerical Derivatives*. R package version 2016.8-1.1. DOI: 10.32614/CRAN.package.numDeriv. URL: https://CRAN.R-project.org/package=numDeriv.

Goeman, Jelle J. and Aldo Solari (2014). "Multiple hypothesis testing in genomics". In: *Statistics in Medicine* 33.11, pp. 1946–1978. DOI: https://doi.org/10.1002/sim.6082. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6082.

Horn, Roger A. and Charles R. Johnson (2012). *Matrix Analysis*. 2nd ed. New York: Cambridge University Press. ISBN: 9780521548236.

Kustosz, Christoph P., Christine H. Müller, and Martin Wendler (2016). "Simplified simplicial depth for regression and autoregressive growth processes". In: *Journal of Statistical Planning and Inference* 173, pp. 125–146.

Liu, Regina Y. (1988). "On a notion of simplicial depth". In: *Proceedings of the National Academy of Sciences of the United States of America* 85, pp. 1732–1734.

Liu, Regina Y. (1990). "On a notion of data depth based on random simplices". In: *The Annals of Statistics* 18, pp. 405–414.

Malcherczyk, Dominik, Kerstin Leckey, and Christine H. Müller (2021). "K-sign depth: From asymptotics to efficient implementation". In: *Journal of Statistical Planning and Inference* 215, pp. 344–355. DOI: 10.1016/j.jspi.2021.02.003.

Meinecke, Steffen, Džanan Sarajlić, Simon Ruben Drauz, Annika Klettke, Lars-Peter Lauven, Christian Rehtanz, Albert Moser, and Martin Braun (2020). "SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis". In: *Energies* 13.12, p. 3290. DOI: 10.3390/en13123290.

Morris, Tim P., Ian R. White, and Michael J. Crowther (2019). "Using simulation studies to evaluate statistical methods". In: *Statistics in Medicine* 38.11, pp. 2074–2102. DOI: 10.1002/sim.8086.

Müller, Christine H. (Mar. 5, 2026). "Optimal Designs for the Kalman Filter Applied to Electrical Power Distribution Grids". Unpublished manuscript.

Müller, Christine H., Stanislav Nagy, and Samuel Trippler (2024). "Multivariate sign depth and related distribution-free tests for model fit". Manuscript Number: AISM-D-24-00108, Original Paper, submitted for publication.

Nagy, Stanislav (2026). *SignDepth: Sign depths for one-dimensional and bivariate data*. R package version 0.1.7, commit 25b89d1ad510362a00df4eb37d6f2a87a112fac3. URL: https://github.com/NagyStanislav/SignDepth.

Pedersen, Thomas Lin (2025). *patchwork: The Composer of Plots*. R package version 1.3.2. DOI: 10.32614/CRAN.package.patchwork. URL: https://CRAN.R-project.org/package=patchwork.

Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli (2009). *Dynamic Models with R*. Springer-Verlag New York.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Shapiro, Samuel S. and Martin B. Wilk (Dec. 1965). "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3-4, pp. 591–611. DOI: 10.

1093/biomet/52.3-4.591. URL: https://doi.org/10.1093/biomet/52.3-4.591.

Silverman, Bernard W. (1986). "Density estimation for statistics and data analysis". In: *Biometrical Journal* 30.7, pp. 876–877. DOI: https://doi.org/10.1002/bimj.4710300745. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710300745.

Toit, Stephanus H. C., A. G. W. Steyn, and Rolf H. Stumpf (2012). *Graphical Exploratory Data Analysis*. 1st ed. Springer Texts in Statistics. New York, NY: Springer. ISBN: 978-1-4612-4950-4. DOI: 10.1007/978-1-4612-4950-4.

Venables, William N. and Brian D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer. URL: https://www.stats.ox.ac.uk/pub/MASS4/.

Wang, Zhaoyuan, Siqi Bu, Jiaxin Wen, and Can Huang (2025). "A comprehensive review on uncertainty modeling methods in modern power systems". In: *International Journal of Electrical Power & Energy Systems* 166, p. 110534. DOI: 10.1016/j.ijepes.2025.110534. URL: https://www.sciencedirect.com/science/article/pii/S0142061525000857.

Wei, Taiyun and Viliam Simko (2024). *R package 'corrplot': Visualization of a Correlation Matrix*. (Version 0.95). URL: https://github.com/taiyun/corrplot.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: 10.21105/joss.01686.

Wickham, Hadley and Jennifer Bryan (2025). *readxl: Read Excel Files*. R package version 1.4.5. DOI: 10.32614/CRAN.package.readxl. URL: https://CRAN.R-project.org/package=readxl.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan (2026). *dplyr: A Grammar of Data Manipulation*. R package version 1.2.0. DOI: 10.32614/CRAN.package.dplyr. URL: https://CRAN.R-project.org/package=dplyr.

Wilk, Martin B. and Ramanathan Gnanadesikan (Mar. 1968). "Probability plotting methods for the analysis for the analysis of data". In: *Biometrika* 55.1, pp. 1–17. ISSN: 0006-3444. DOI: 10.1093/biomet/55.1.1. URL: https://doi.org/10.1093/biomet/55.1.1.

A Additional tables

A.1 Quantile tables

Quantile level	$K = 3$	$K = 4$
0.8333%	0.2267	0.1078
1%	0.2276	0.1085
2%	0.2320	0.1115
5%	0.2375	0.1155
10%	0.2419	0.1186

Table 24: Simulated quantiles of the K-sign depth for $K = 3$ and $K = 4$ (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 100$).

Quantile level	$K = 3$	$K = 4$
0.8333%	−2.3296	−1.7185
1%	−2.2369	−1.6467
2%	−1.8040	−1.3489
5%	−1.2486	−0.9476
10%	−0.8145	−0.6400

Table 25: Simulated quantiles of the standardised K-sign depth for $K = 3$ and $K = 4$ (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 100$).

Quantile level	$K = 3$	$K = 4$
0.8333%	0.1224	0.0309
1%	0.1327	0.0309
2%	0.1429	0.0412
5%	0.1633	0.0515
10%	0.1837	0.0619

Table 26: Simulated quantiles of the simplified 3- and 4-sign depth (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 100$).

Quantile level	$K = 1$	$K = 2$
1%	0.2270	0.0733
1.667%	0.2303	0.0745
2%	0.2312	0.0750
5%	0.2376	0.0775
10%	0.2419	0.0793

Table 27: Simulated quantiles of the full K-simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 10\,000$, Sample size within a single simulation: $N = 100$).

Quantile level	$K = 1$	$K = 2$
1%	-2.3049	-1.0080
1.667%	-1.9672	-0.8810
2%	-1.8782	-0.8326
5%	-1.2362	-0.5867
10%	-0.8108	-0.4007

Table 28: Simulated quantiles of the standardised full-simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 10\,000$, Sample size within a single simulation: $N = 100$).

Quantile level	$K = 1$	$K = 2$
1%	0.1429	0.0206
1.667%	0.1531	0.0206
2%	0.1531	0.0206
5%	0.1735	0.0309
10%	0.1837	0.0412

Table 29: Simulated quantiles of the simplified simplex depth for $K = 1$ and $K = 2$ (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 100$).

A.2 Result tables for the simulations

State-dependent-noise EKF

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2527	0.2561	0.2503	0.2431	0.2527	0.2543	0.2375	0.2267
$D_3^{u,F}$	0.2665	0.6141	0.0291	-0.6908	0.2727	0.4348	-1.2486	-2.3296
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1288	0.1304	0.1237	0.1263	0.1269	0.1288	0.1155	0.1078
$D_4^{u,F}$	0.3788	0.5410	-0.1325	0.1277	0.1884	0.3850	-0.9476	-1.7185
$R_{d_4^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2653	0.3163	0.1837	0.2755	0.3367	0.3673	0.1633	0.1224
$D_3^{u,S}$	0.2711	1.1746	-1.1746	0.4518	1.5360	2.0781	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,S}$	0.1340	0.1753	0.0825	0.1134	0.2474	0.1959	0.0515	0.0309
$D_4^{u,S}$	0.1835	1.0224	-0.8651	-0.2359	2.4905	1.4419	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
Shapiro-Wilk	0.9787	0.6223	0.8867	0.5140	0.6729	0.1180		

Table 30: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF for Seed 1234.

Methods		**Nodes**		**Quantiles**	
	$(P_1,\ Q_1)$	$(P_2,\ Q_2)$	$(P_3,\ Q_3)$	α	α_{Bonf}
d_1^F	0.2532	0.2511	0.2548	0.2376	0.2303
D_1^F	0.3173	0.1119	0.4842	-1.2362	-1.9672
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0851	0.0796	0.0854	0.0775	0.0745
D_2^F	0.1759	-0.3708	0.2052	-0.5867	-0.8810
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2143	0.2755	0.2449	0.1735	0.1531
D_1^S	-0.7385	0.5275	-0.1055	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0412	0.1340	0.0309	0.0309	0.0206
D_2^S	-1.2102	1.4572	-1.5066	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 31: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF for Seed 1234.

State-dependent-noise EKF with standardised residuals

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2534	0.2520	0.2514	0.2550	0.2559	0.2379	0.2375	0.2267
$D_3^{u,F}$	0.3383	0.1973	0.1404	0.4978	0.5894	-1.2066	-1.2486	-2.3296
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1289	0.1258	0.1270	0.1294	0.1286	0.1160	0.1155	0.1078
$D_4^{u,F}$	0.3868	0.0763	0.2028	0.4413	0.3596	-0.9004	-0.9476	-1.7185
$R_{d_4^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2347	0.2959	0.2653	0.3571	0.2449	0.1735	0.1633	0.1224
$D_3^{u,S}$	-0.2711	0.8132	0.2711	1.8974	-0.0904	-1.3553	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,S}$	0.1134	0.1753	0.1134	0.2680	0.0825	0.0722	0.0515	0.0309
$D_4^{u,S}$	-0.2359	1.0224	-0.2359	2.9100	-0.8651	-1.0749	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
Shapiro-Wilk	0.4609	0.0188	0.0872	0.8097	0.3360	0.1093		

Table 32: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF with standardised residuals for Seed 1.

Methods		**Nodes**		**Quantiles**	
	$(P_1,\ Q_1)$	$(P_2,\ Q_2)$	$(P_3,\ Q_3)$	α	α_{Bonf}
d_1^F	0.2543	0.2478	0.2433	0.2376	0.2303
D_1^F	0.4310	-0.2245	-0.6685	-1.2362	-1.9672
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0853	0.0828	0.0799	0.0775	0.0745
D_2^F	0.1938	-0.0572	-0.3409	-0.5867	-0.8810
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2245	0.3061	0.3163	0.1735	0.1531
D_1^S	-0.5275	1.1606	1.3716	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0515	0.0619	0.0928	0.0309	0.0206
D_2^S	-0.9138	-0.6175	0.2717	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 33: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the state-dependent-noise EKF with standardised residuals for Seed 1.

Constant-noise EKF

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2552	0.2536	0.2378	0.2562	0.2525	0.2352	0.2375	0.2267
$D_3^{u,F}$	0.5238	0.3643	-1.2189	0.6240	0.2517	-1.4799	-1.2486	-2.3296
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1286	0.1289	0.1174	0.1309	0.1284	0.1128	0.1155	0.1078
$D_4^{u,F}$	0.3560	0.3883	-0.7608	0.5870	0.3359	-1.2215	-0.9476	-1.7185
$R_{d_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2653	0.2857	0.3367	0.3367	0.2755	0.1531	0.1633	0.1224
$D_3^{u,S}$	0.2711	0.6325	1.5360	1.5360	0.4518	-1.7167	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,S}$	0.1134	0.1753	0.2062	0.2371	0.1443	0.0515	0.0515	0.0309
$D_4^{u,S}$	-0.2359	1.0224	1.6516	2.2808	0.3932	-1.4943	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
Shapiro-Wilk	0.2898	0.0211	0.4345	0.6597	0.8283	0.4321		

Table 34: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the constant-noise EKF for Seed 1.

Methods		**Nodes**		**Quantiles**	
	$(P_1,\ Q_1)$	$(P_2,\ Q_2)$	$(P_3,\ Q_3)$	α	α_{Bonf}
d_1^F	0.2570	0.2393	0.2456	0.2376	0.2303
D_1^F	0.6970	-1.0742	-0.4447	-1.2362	-1.9672
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0872	0.0795	0.0811	0.0775	0.0745
D_2^F	0.3821	-0.3805	-0.2218	-0.5867	-0.8810
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2449	0.3469	0.2857	0.1735	0.1531
D_1^S	-0.1055	2.0046	0.7385	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0722	0.1443	0.1031	0.0309	0.0206
D_2^S	-0.3211	1.7536	0.5681	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 35: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the constant-noise EKF for Seed 1.

Constant-noise EKF with standardised residuals

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2552	0.2536	0.2378	0.2562	0.2554	0.2362	0.2375	0.2267
$D_3^{u,F}$	0.5238	0.3643	-1.2189	0.6240	0.5387	-1.3772	-1.2486	-2.3296
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1286	0.1289	0.1174	0.1309	0.1309	0.1142	0.1155	0.1078
$D_4^{u,F}$	0.3560	0.3883	-0.7608	0.5870	0.5889	-1.0829	-0.9476	-1.7185
$R_{d_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{D_4^{u,F}}(0.05)$	0	0	0	0	0	1		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2653 0	.2857	0.3367	0.3367	0.2959	0.1531	0.1633	0.1224
$D_3^{u,S}$	0.2711	0.6325	1.5360	1.5360	0.8132	-1.7167	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{D_3^{u,S}}(0.05)$	0	0	0	0	0	1		
$R_{d_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,S}$	0.1134	0.1753	0.2062	0.2371	0.1340	0.0619	0.0515	0.0309
$D_4^{u,S}$	-0.2359	1.0224	1.6516	2.2808	0.1835	-1.2846	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.05)$	0	0	0	0	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	0	0	0	0		
Shapiro-Wilk	0.2899	0.0212	0.4298	0.6287	0.7248	0.5218		

Table 36: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the constant-noise EKF with standardised residuals for Seed 1.

Methods		**Nodes**		**Quantiles**	
	$(P_1,\ Q_1)$	$(P_2,\ Q_2)$	$(P_3,\ Q_3)$	α	α_{Bonf}
d_1^F	0.2570	0.2393	0.2456	0.2376	0.2303
D_1^F	0.6970	-1.0742	-0.4447	-1.2362	-1.9672
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0872	0.0795	0.0811	0.0775	0.0745
D_2^F	0.3821	-0.3805	-0.2218	-0.5867	-0.8810
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2449	0.3469	0.2857	0.1735	0.1531
D_1^S	-0.1055	2.0046	0.7385	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0722	0.1443	0.1031	0.0309	0.0206
D_2^S	-0.3211	1.7536	0.5681	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 37: Depth statistics, critical values, and rejection decisions of an illustrative simulation from the constant-noise EKF with standardised residuals for Seed 1.

A.3 Result tables for the data application

Model A: Random walk state-space model

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2473	0.2519	0.2502	0.2506	0.2506	0.2507	0.2464	0.2431
$D_3^{u,F}$	-0.8933	0.6403	0.0838	0.1871	0.1953	0.2285	-1.2218	-2.3279
$R_{d_3^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.05)$	0	0	0	0	0	0		
$R_{d_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_3^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_4^{u,F}$	0.1216	0.1268	0.1256	0.1253	0.1260	0.1262	0.1222	0.1199
$D_4^{u,F}$	-1.1569	0.5905	0.2054	0.0948	0.3240	0.4109	-0.9383	-1.7199
$R_{d_4^{u,F}}(0.05)$	1	0	0	0	0	0		
$R_{D_4^{u,F}}(0.05)$	1	0	0	0	0	0		
$R_{d_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$R_{D_4^{u,F}}(0.0083)$	0	0	0	0	0	0		
$d_3^{u,S}$	0.2006	0.2006	0.1707	0.1557	0.2246	0.3952	0.2006	0.1796
$D_3^{u,S}$	-1.6150	-1.6150	-2.5939	-3.0833	-0.8320	4.7473	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	0	1	1	0	0		
$R_{D_3^{u,S}}(0.05)$	0	0	1	1	0	0		
$R_{d_3^{u,S}}(0.0083)$	0	0	1	1	0	0		
$R_{D_3^{u,S}}(0.0083)$	0	0	1	1	0	0		
$d_4^{u,S}$	0.1081	0.0811	0.0601	0.0691	0.1141	0.2372	0.0841	0.0661
$D_4^{u,S}$	-0.6367	-1.6555	-2.4478	-2.1082	-0.4103	4.2306	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	1	1	1	0	0		
$R_{D_4^{u,S}}(0.05)$	0	1	1	1	0	0		
$R_{d_4^{u,S}}(0.0083)$	0	0	1	0	0	0		
$R_{D_4^{u,S}}(0.0083)$	0	0	1	0	0	0		
Shapiro-Wilk	0	0	0	0	0	0		

Table 38: Depth statistics, critical values, and rejection decisions for the random walk state-space model (Quantile simulations: $M = 100\,000$, Sample size within a single simulation: $N = 336$).

Methods		**Nodes**		**Quantiles**	
	(P_1, Q_1)	(P_2, Q_2)	(P_3, Q_3)	α	α_{Bonf}
d_1^F	0.2475	0.2473	0.2509	0.2463	0.2443
D_1^F	-0.8260	-0.9102	0.2859	-1.2269	-1.9009
$R_{d_1^F}(0.05)$	0	0	0		
$R_{D_1^F}(0.05)$	0	0	0		
$R_{d_1^F}(0.0167)$	0	0	0		
$R_{D_1^F}(0.0167)$	0	0	0		
d_2^F	0.0832	0.0817	0.0839	0.0816	0.0808
D_2^F	-0.0440	-0.5418	0.1977	-0.5872	-0.8650
$R_{d_2^F}(0.05)$	0	0	0		
$R_{D_2^F}(0.05)$	0	0	0		
$R_{d_2^F}(0.0167)$	0	0	0		
$R_{D_2^F}(0.0167)$	0	0	0		
d_1^S	0.2275	0.2605	0.3323	0.2066	0.1946
D_1^S	-0.8573	0.4001	3.1433	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	0	0		
$R_{D_1^S}(0.05)$	0	0	0		
$R_{d_1^S}(0.0167)$	0	0	0		
$R_{D_1^S}(0.0167)$	0	0	0		
d_2^S	0.0571	0.0661	0.1141	0.0541	0.0450
D_2^S	-1.3997	-0.9198	1.6396	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	0	0		
$R_{D_2^S}(0.05)$	0	0	0		
$R_{d_2^S}(0.0167)$	0	0	0		
$R_{D_2^S}(0.0167)$	0	0	0		

Table 39: Depth statistics, critical values, and rejection decisions for the random walk state-space model(Quantile simulations: $M = 100\,000$ - except for the full K-simplex depth and the standardisation $M = 10\,000$, Sample size within a single simulation: $N = 336$).

Model B: Mean-reverting state-space model with constant reference state

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.2514	0.2447	0.1247	0.0101	0.0781	0.0000	0.2464	0.2431
$D_3^{u,F}$	0.4836	-1.7855	-42.1175	-80.6166	-57.7445	-84.0000	-1.2218	-2.3279
$R_{d_3^{u,F}}(0.05)$	0	1	1	1	1	1		
$R_{D_3^{u,F}}(0.05)$	0	1	1	1	1	1		
$R_{d_3^{u,F}}(0.0083)$	0	0	1	1	1	1		
$R_{D_3^{u,F}}(0.0083)$	0	0	1	1	1	1		
$d_4^{u,F}$	0.1259	0.1204	0.0369	0.0002	0.0109	0.0000	0.1222	0.1199
$D_4^{u,F}$	0.3073	-1.5595	-29.6123	-41.9224	-38.3450	-42.0000	-0.9383	-1.7199
$R_{d_4^{u,F}}(0.05)$	0	1	1	1	1	1		
$R_{D_4^{u,F}}(0.05)$	0	1	1	1	1	1		
$R_{d_4^{u,F}}(0.0083)$	0	0	1	1	1	1		
$R_{D_4^{u,F}}(0.0083)$	0	0	1	1	1	1		
$d_3^{u,S}$	0.2126	0.1647	0.0898	0.0150	0.0838	0.0000	0.2006	0.1796
$D_3^{u,S}$	-1.2235	-2.7896	-5.2367	-7.6837	-5.4324	-8.1731	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	0	1	1	1	1	1		
$R_{D_3^{u,S}}(0.05)$	0	1	1	1	1	1		
$R_{d_3^{u,S}}(0.0083)$	0	1	1	1	1	1		
$R_{D_3^{u,S}}(0.0083)$	0	1	1	1	1	1		
$d_4^{u,S}$	0.1111	0.0571	0.0210	0.0000	0.0180	0.0000	0.0841	0.0661
$D_4^{u,S}$	-0.5235	-2.5610	-3.9193	-4.7117	-4.0325	-4.7117	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	0	1	1	1	1	1		
$R_{D_4^{u,S}}(0.05)$	0	1	1	1	1	1		
$R_{d_4^{u,S}}(0.0083)$	0	1	1	1	1	1		
$R_{D_4^{u,S}}(0.0083)$	0	1	1	1	1	1		
Shapiro-Wilk	0	0	0	0	0	0		

Table 40: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with constant reference state (Quantile simulations: $M = 100\,000$, Sample size within a single simulation: $N = 336$).

Methods		**Nodes**		**Quantiles**	
	(P_1, Q_1)	(P_2, Q_2)	(P_3, Q_3)	α	α_{Bonf}
d_1^F	0.2402	0.0002	0.0000	0.2463	0.2443
D_1^F	-3.2995	-83.9290	-84.0000	-1.2269	-1.9009
$R_{d_1^F}(0.05)$	1	1	1		
$R_{D_1^F}(0.05)$	1	1	1		
$R_{d_1^F}(0.0167)$	1	1	1		
$R_{D_1^F}(0.0167)$	1	1	1		
d_2^F	0.0805	0.0001	0.0000	0.0816	0.0808
D_2^F	-0.9410	-27.9636	-28.0000	-0.5872	-0.8650
$R_{d_2^F}(0.05)$	1	1	1		
$R_{D_2^F}(0.05)$	1	1	1		
$R_{d_2^F}(0.0167)$	1	1	1		
$R_{D_2^F}(0.0167)$	1	1	1		
d_1^S	0.2186	0.0000	0.0000	0.2066	0.1946
D_1^S	-1.2002	-9.5442	-9.5442	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	0	1	1		
$R_{D_1^S}(0.05)$	0	1	1		
$R_{d_1^S}(0.0167)$	0	1	1		
$R_{D_1^S}(0.0167)$	0	1	1		
d_2^S	0.0691	0.0000	0.0000	0.0541	0.0450
D_2^S	-0.7598	-4.4389	-4.4389	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	1	1		
$R_{D_2^S}(0.05)$	0	1	1		
$R_{d_2^S}(0.0167)$	0	1	1		
$R_{D_2^S}(0.0167)$	0	1	1		

Table 41: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with constant reference state (Quantile simulations: $M = 100\,000$ - except for the full K-simplex depth and the standardisation $M = 10\,000$, Sample size within a single simulation: $N = 336$).

Model C: Mean-reverting state-space model with seasonal load profiles

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.0743	0.0065	0.0000	0.0000	0.0000	0.0000	0.2464	0.2431
$D_3^{u,F}$	-59.0201	-81.8018	-84.0000	-84.0000	-84.0000	-84.0000	-1.2218	-2.3279
$R_{d_3^{u,F}}(0.05)$	1	1	1	1	1	1		
$R_{D_3^{u,F}}(0.05)$	1	1	1	1	1	1		
$R_{d_3^{u,F}}(0.0083)$	1	1	1	1	1	1		
$R_{D_3^{u,F}}(0.0083)$	1	1	1	1	1	1		
$d_4^{u,F}$	0.0127	0.0001	0.0000	0.0000	0.0000	0.0000	0.1222	0.1199
$D_4^{u,F}$	-37.7373	-41.9563	-42.0000	-42.0000	-42.0000	-42.0000	-0.9383	-1.7199
$R_{d_4^{u,F}}(0.05)$	1	1	1	1	1	1		
$R_{D_4^{u,F}}(0.05)$	1	1	1	1	1	1		
$R_{d_4^{u,F}}(0.0083)$	1	1	1	1	1	1		
$R_{D_4^{u,F}}(0.0083)$	1	1	1	1	1	1		
$d_3^{u,S}$	0.0838	0.0060	0.0000	0.0000	0.0000	0.0000	0.2006	0.1796
$D_3^{u,S}$	-5.4324	-7.9774	-8.1731	-8.1731	-8.1731	-8.1731	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	1	1	1	1	1	1		
$R_{D_3^{u,S}}(0.05)$	1	1	1	1	1	1		
$R_{d_3^{u,S}}(0.0083)$	1	1	1	1	1	1		
$R_{D_3^{u,S}}(0.0083)$	1	1	1	1	1	1		
$d_4^{u,S}$	0.0150	0.0000	0.0000	0.0000	0.0000	0.0000	0.0841	0.0661
$D_4^{u,S}$	-4.1457	-4.7117	-4.7117	-4.7117	-4.7117	-4.7117	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	1	1	1	1	1	1		
$R_{D_4^{u,S}}(0.05)$	1	1	1	1	1	1		
$R_{d_4^{u,S}}(0.0083)$	1	1	1	1	1	1		
$R_{D_4^{u,S}}(0.0083)$	1	1	1	1	1	1		
Shapiro-Wilk	0	0	0	0	0	0		

Table 42: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with seasonal load profiles (Quantile simulations: $M = 100\,000$, Sample size within a single simulation: $N = 336$).

Methods	**Nodes**			**Quantiles**	
	(P_1, Q_1)	(P_2, Q_2)	(P_3, Q_3)	α	α_{Bonf}
d_1^F	0.0054	0.0000	0.0000	0.2463	0.2443
D_1^F	-82.1790	-84.0000	-84.0000	-1.2269	-1.9009
$R_{d_1^F}(0.05)$	1	1	1		
$R_{D_1^F}(0.05)$	1	1	1		
$R_{d_1^F}(0.0167)$	1	1	1		
$R_{D_1^F}(0.0167)$	1	1	1		
d_2^F	0.0009	0.0000	0.0000	0.0816	0.0808
D_2^F	-27.6980	-28.0000	-28.0000	-0.5872	-0.8650
$R_{d_2^F}(0.05)$	1	1	1		
$R_{D_2^F}(0.05)$	1	1	1		
$R_{d_2^F}(0.0167)$	1	1	1		
$R_{D_2^F}(0.0167)$	1	1	1		
d_1^S	0.0060	0.0000	0.0000	0.2066	0.1946
D_1^S	-9.3156	-9.5442	-9.5442	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	1	1	1		
$R_{D_1^S}(0.05)$	1	1	1		
$R_{d_1^S}(0.0167)$	1	1	1		
$R_{D_1^S}(0.0167)$	1	1	1		
d_2^S	0.0030	0.0000	0.0000	0.0541	0.0450
D_2^S	-4.2790	-4.4389	-4.4389	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	1	1	1		
$R_{D_2^S}(0.05)$	1	1	1		
$R_{d_2^S}(0.0167)$	1	1	1		
$R_{D_2^S}(0.0167)$	1	1	1		

Table 43: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with seasonal load profiles (Quantile simulations: $M = 100\,000$ - except for the full K-simplex depth and the standardisation $M = 10\,000$, Sample size within a single simulation: $N = 336$).

Model D: Mean-reverting state-space model with time-varying seasonal reference state

Methods	**Nodes**						**Quantiles**	
	P_1	Q_1	P_2	Q_2	P_3	Q_3	α	α_{Bonf}
$d_3^{u,F}$	0.1959	0.2233	0.1535	0.1343	0.2450	0.2500	0.2464	0.2431
$D_3^{u,F}$	-18.1622	-8.9696	-32.4245	-38.8770	-1.6666	0.0016	-1.2218	-2.3279
$R_{d_3^{u,F}}(0.05)$	1	1	1	1	1	0		
$R_{D_3^{u,F}}(0.05)$	1	1	1	1	1	0		
$R_{d_3^{u,F}}(0.0083)$	1	1	1	1	0	0		
$R_{D_3^{u,F}}(0.0083)$	1	1	1	1	0	0		
$d_4^{u,F}$	0.0868	0.1055	0.0782	0.0436	0.1207	0.1236	0.1222	0.1199
$D_4^{u,F}$	-12.8409	-6.5354	-15.7317	-27.3347	-1.4301	-0.4717	-0.9383	-1.7199
$R_{d_4^{u,F}}(0.05)$	1	1	1	1	1	0		
$R_{D_4^{u,F}}(0.05)$	1	1	1	1	1	0		
$R_{d_4^{u,F}}(0.0083)$	1	1	1	1	0	0		
$R_{D_4^{u,F}}(0.0083)$	1	1	1	1	0	0		
$d_3^{u,S}$	0.1168	0.1826	0.0719	0.1138	0.1976	0.3353	0.2006	0.1796
$D_3^{u,S}$	-4.3557	-2.2023	-5.8240	-4.4536	-1.7129	2.7896	-1.6449	-2.3940
$R_{d_3^{u,S}}(0.05)$	1	1	1	1	1	0		
$R_{D_3^{u,S}}(0.05)$	1	1	1	1	1	0		
$R_{d_3^{u,S}}(0.0083)$	1	0	1	1	0	0		
$R_{D_3^{u,S}}(0.0083)$	1	0	1	1	0	0		
$d_4^{u,S}$	0.0390	0.0571	0.0180	0.0360	0.1021	0.1952	0.0841	0.0661
$D_4^{u,S}$	-3.2402	-2.5610	-4.0325	-3.3534	-0.8631	2.6459	-1.6449	-2.3940
$R_{d_4^{u,S}}(0.05)$	1	1	1	1	0	0		
$R_{D_4^{u,S}}(0.05)$	1	1	1	1	0	0		
$R_{d_4^{u,S}}(0.0083)$	1	1	1	1	0	0		
$R_{D_4^{u,S}}(0.0083)$	1	1	1	1	0	0		
Shapiro-Wilk	0	0	0	0	0	0		

Table 44: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with time-varying seasonal reference state (Simulations: $M = 100\,000$, Sample size within a single simulation: $N = 336$).

Methods		**Nodes**		**Quantiles**	
	(P_1, Q_1)	(P_2, Q_2)	(P_3, Q_3)	α	α_{Bonf}
d_1^F	0.1969	0.1110	0.2471	0.2463	0.2443
D_1^F	-17.8556	-46.6977	-0.9710	-1.2269	-1.9009
$R_{d_1^F}(0.05)$	1	1	0		
$R_{D_1^F}(0.05)$	1	1	0		
$R_{d_1^F}(0.0167)$	1	1	0		
$R_{D_1^F}(0.0167)$	1	1	0		
d_2^F	0.0571	0.0198	0.0821	0.0816	0.0808
D_2^F	-8.8082	-21.3516	-0.4107	-0.5872	-0.8650
$R_{d_2^F}(0.05)$	1	1	0		
$R_{D_2^F}(0.05)$	1	1	0		
$R_{d_2^F}(0.0167)$	1	1	0		
$R_{D_2^F}(0.0167)$	1	1	0		
d_1^S	0.1826	0.0449	0.2994	0.2066	0.1946
D_1^S	-2.5718	-7.8296	1.8860	-1.6449	-2.1280
$R_{d_1^S}(0.05)$	1	1	0		
$R_{D_1^S}(0.05)$	1	1	0		
$R_{d_1^S}(0.0167)$	1	1	0		
$R_{D_1^S}(0.0167)$	1	1	0		
d_2^S	0.0691	0.0120	0.1231	0.0541	0.0450
D_2^S	-0.7598	-3.7991	2.1195	-1.6449	-2.1280
$R_{d_2^S}(0.05)$	0	1	0		
$R_{D_2^S}(0.05)$	0	1	0		
$R_{d_2^S}(0.0167)$	0	1	0		
$R_{D_2^S}(0.0167)$	0	1	0		

Table 45: Depth statistics, critical values, and rejection decisions from the mean-reverting state-space model with time-varying seasonal reference state (Quantile simulations: $M = 100\,000$ - except for the full K-simplex depth and the standardisation $M = 10\,000$, Sample size within a single simulation: $N = 336$).

B Additional graphics

B.1 Descriptive graphics of the simulated data

State-dependent-noise EKF

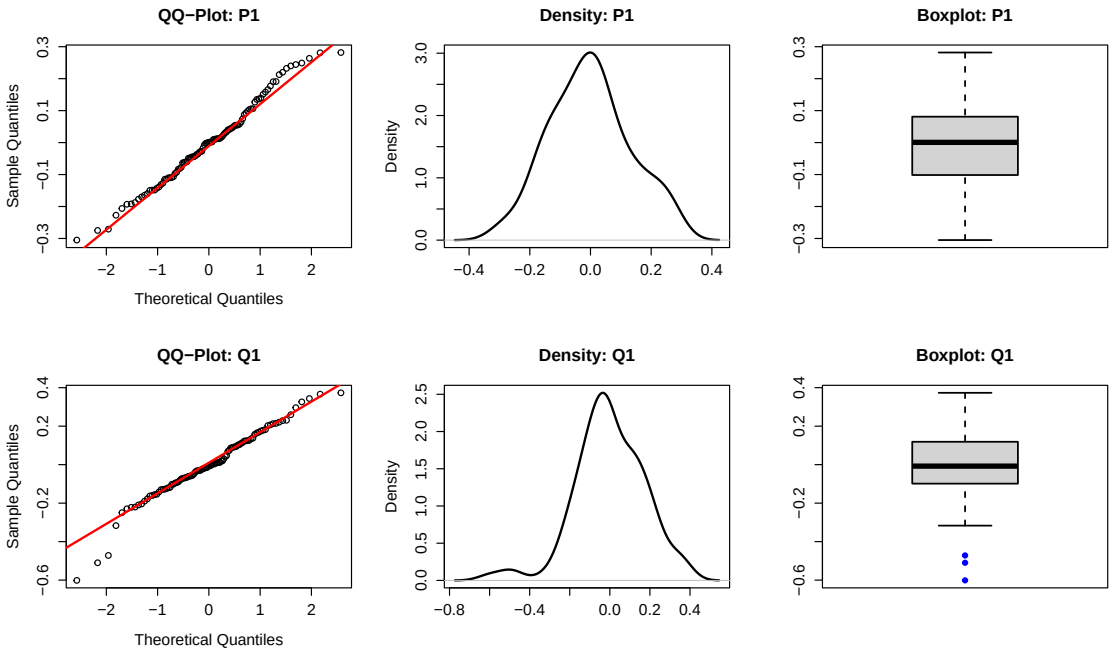

Figure 29: Descriptive graphics from data of an illustrative simulation run from the state-dependent-noise EKF for the first node with the seed = 1.

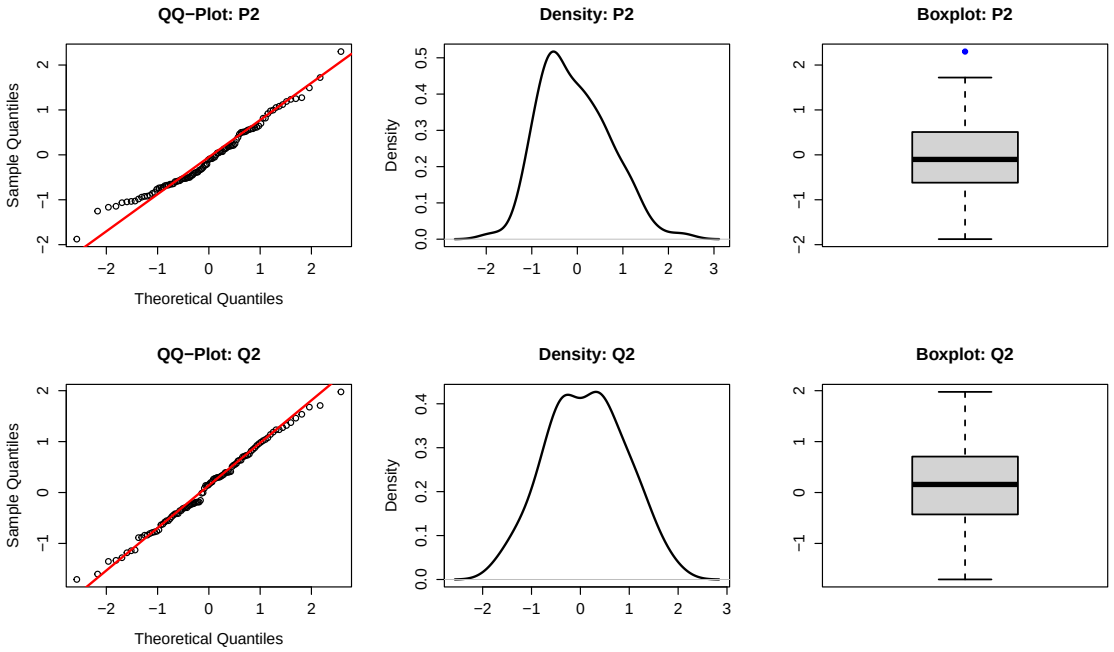

Figure 30: Descriptive graphics from data of an illustrative simulation run from the state-dependent-noise EKF for the second node with the seed = 1.

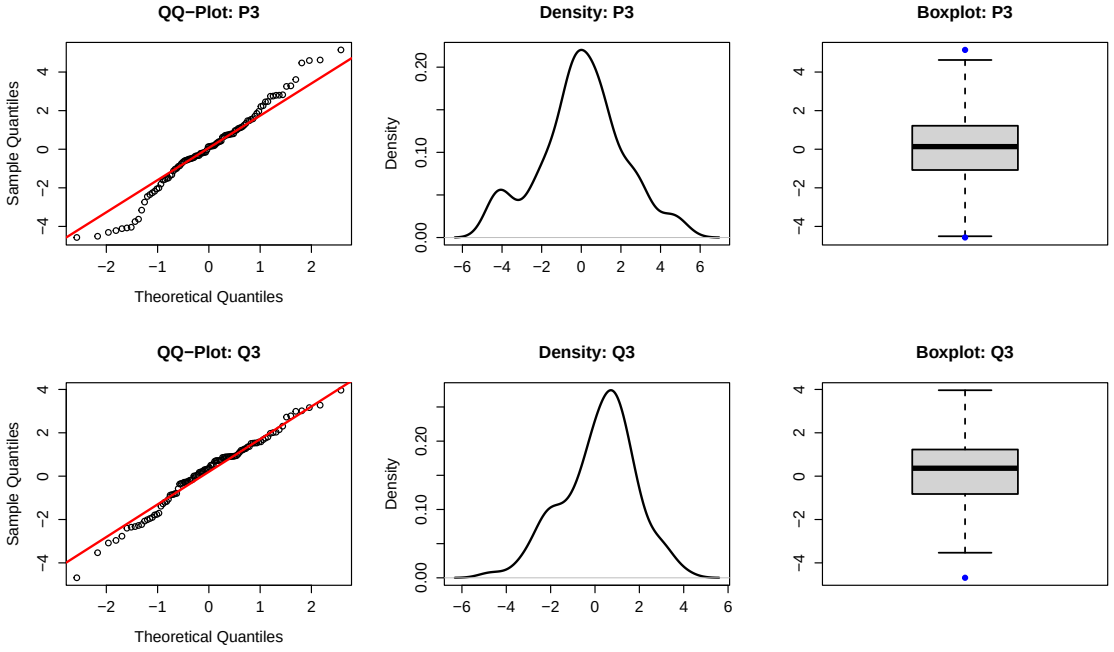

Figure 31: Descriptive graphics from data of an illustrative simulation run from the state-dependent-noise EKF for the third node with the seed = 1.

Constant-noise EKF

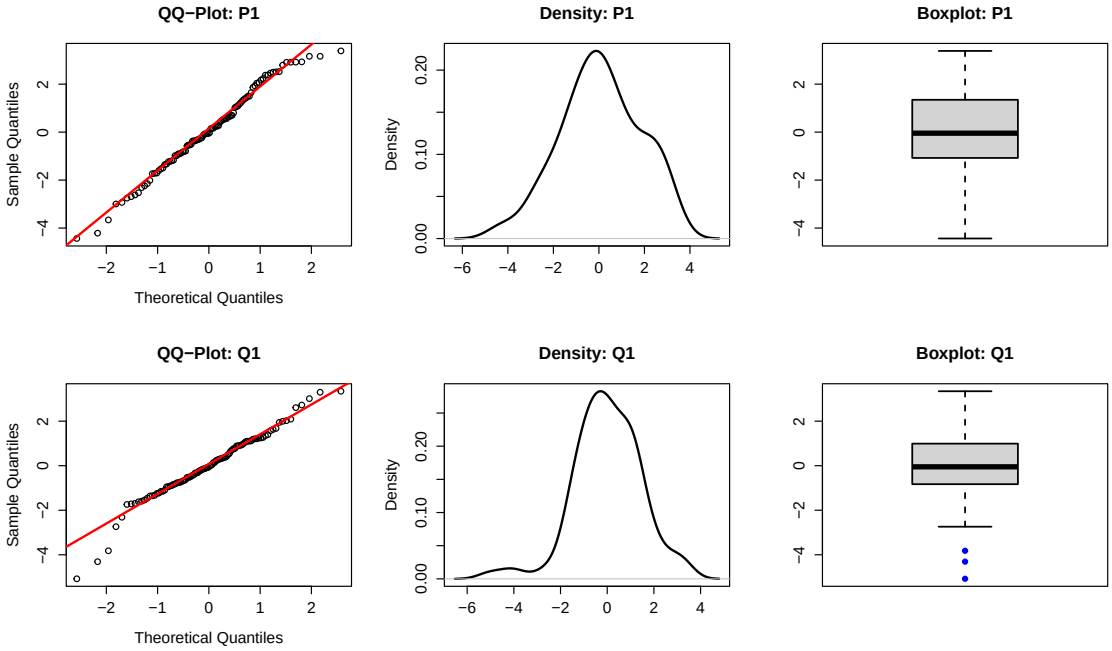

Figure 32: Descriptive graphics from data of an illustrative simulation run from the constant-noise EKF for the first node with the seed = 1.

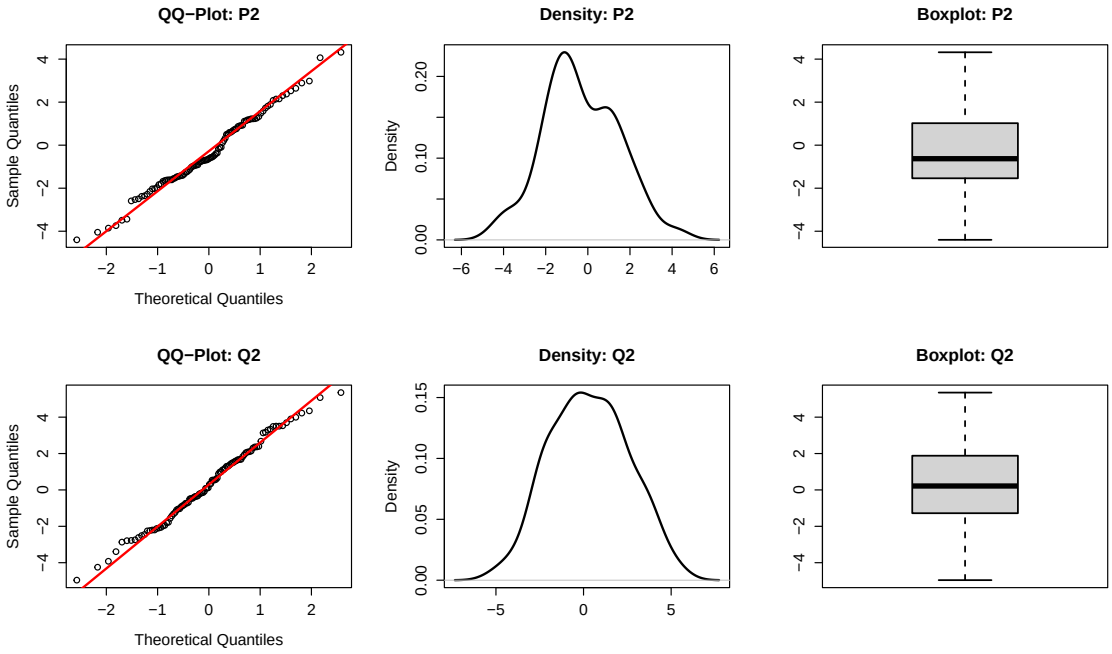

Figure 33: Descriptive graphics from data of an illustrative simulation run from the constant-noise EKF for the second node with the seed = 1.

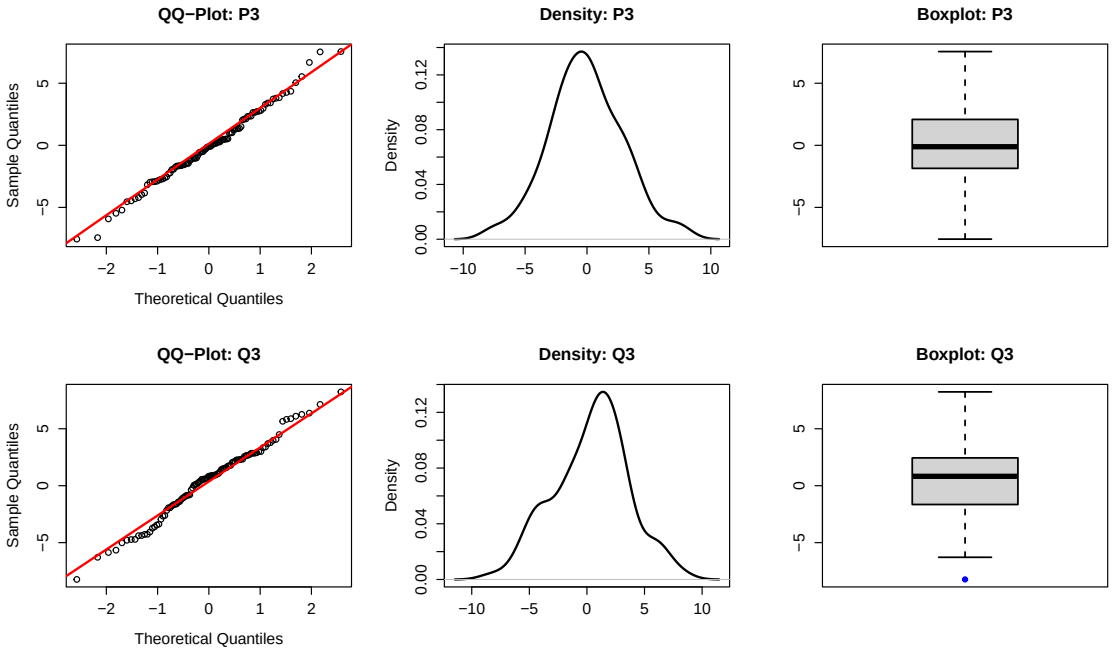

Figure 34: Descriptive graphics from data of an illustrative simulation run from the constant-noise EKF for the third node with the seed = 1.

B.2 Simplices

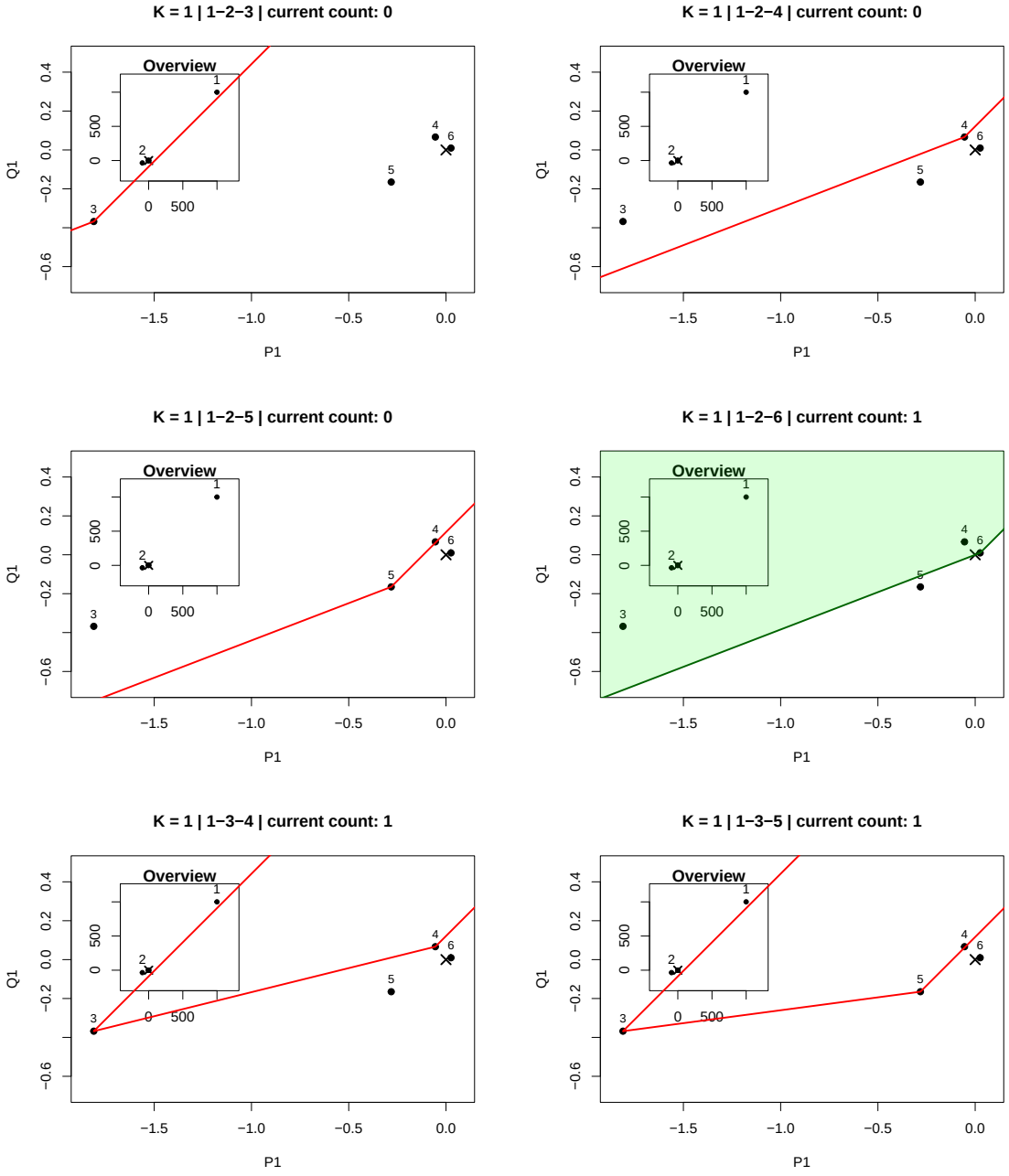

Figure 35: Illustration of the simplex depth for $K = 1$ on a subset of six observations generated with the state-dependent-noise EKF, including one positive outlier (part 1). The main panel shows the zoomed-in region used for the simplex depth, while the inset provides an overview of the complete subset and indicates the location of the zoomed region. Triangles containing the origin are highlighted in green, whereas triangles not containing the origin are outlined in red.

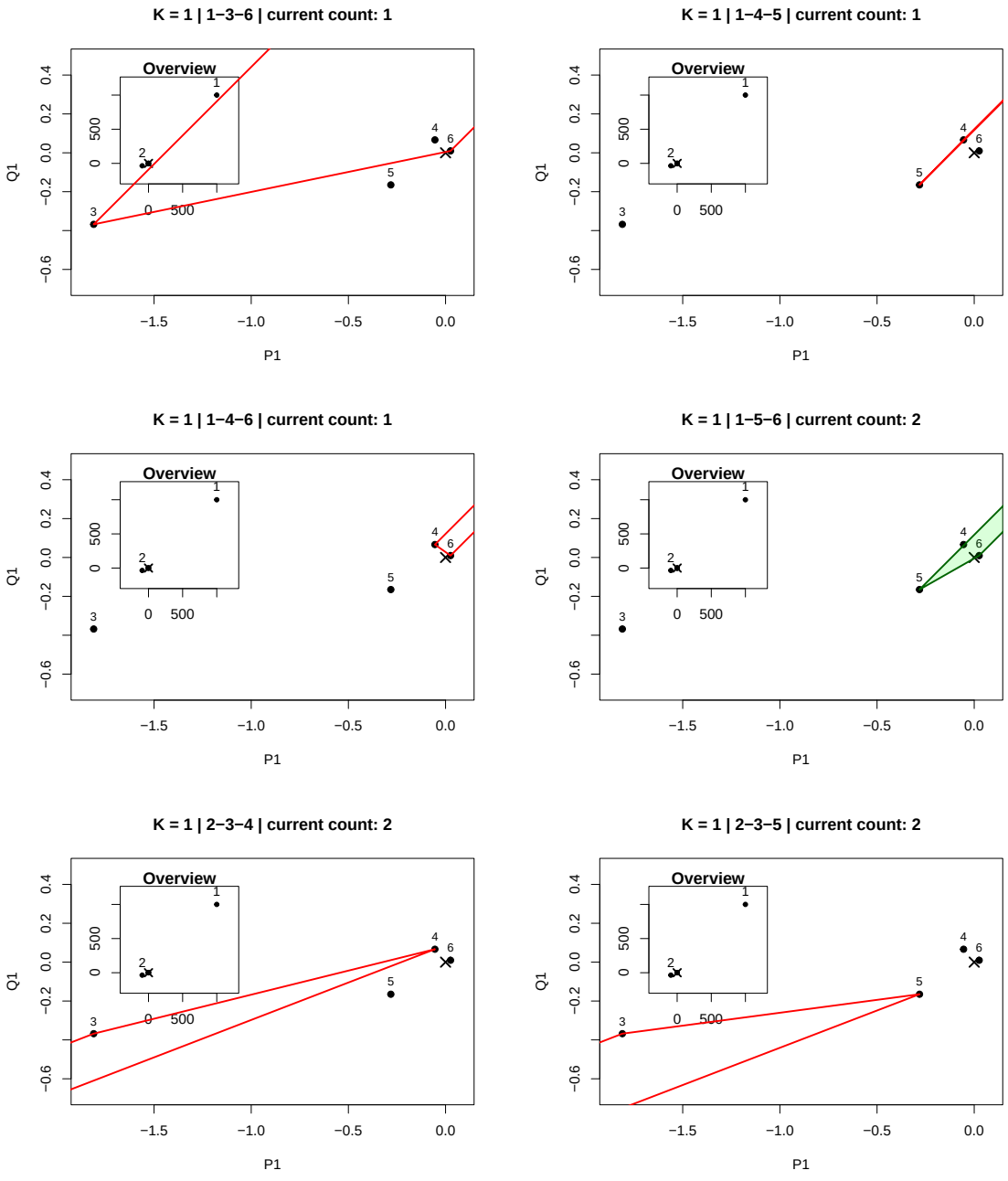

Figure 36: Illustration of the simplex depth for $K = 1$ on a subset of six observations generated with the state-dependent-noise EKF, including one positive outlier (part 2). The main panel shows the zoomed-in region used for the simplex depth, while the inset provides an overview of the complete subset and indicates the location of the zoomed region. Triangles containing the origin are highlighted in green, whereas triangles not containing the origin are outlined in red.

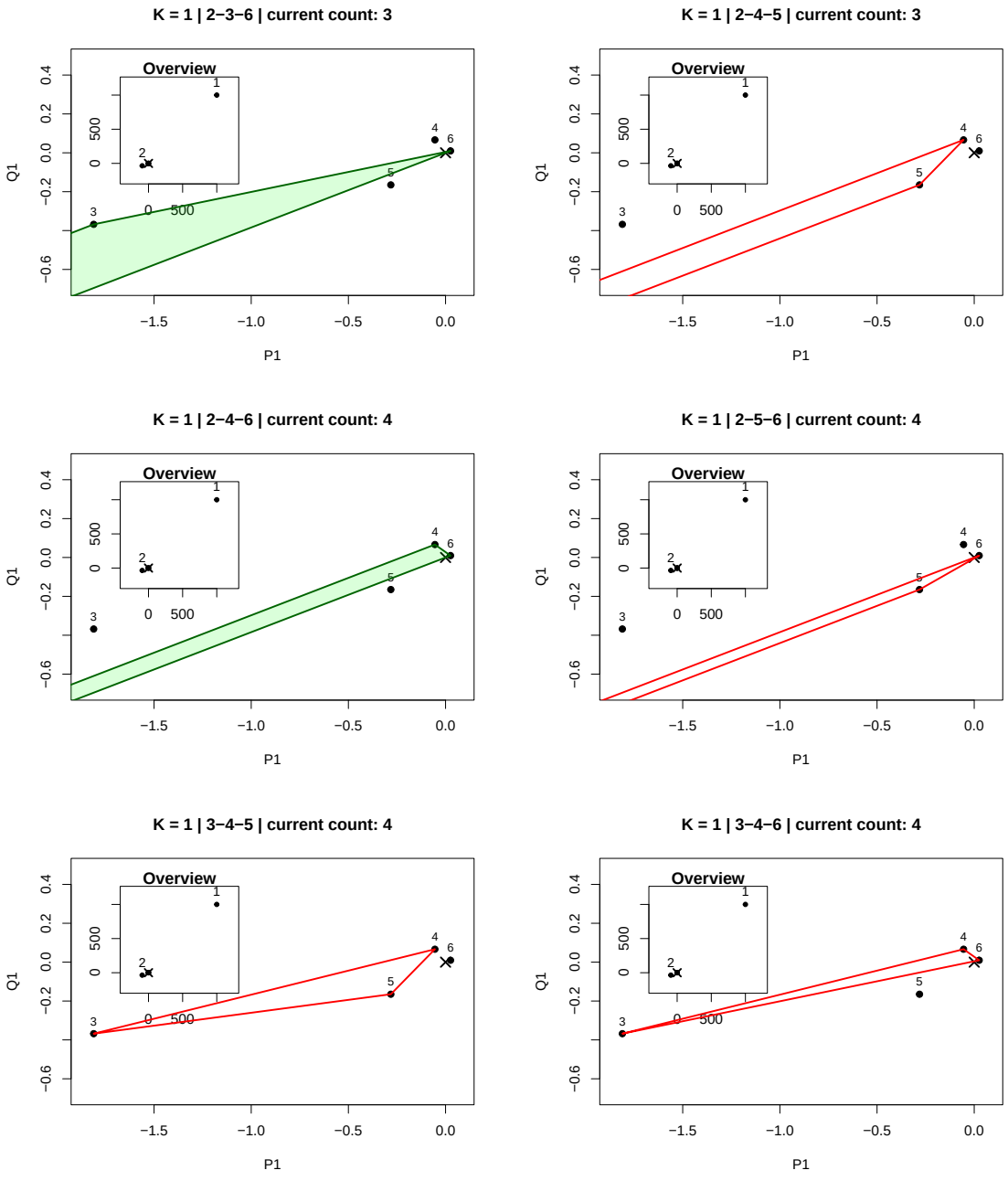

Figure 37: Illustration of the simplex depth for $K = 1$ on a subset of six observations generated with the state-dependent-noise EKF, including one positive outlier (part 3). The main panel shows the zoomed-in region used for the simplex depth, while the inset provides an overview of the complete subset and indicates the location of the zoomed region. Triangles containing the origin are highlighted in green, whereas triangles not containing the origin are outlined in red.

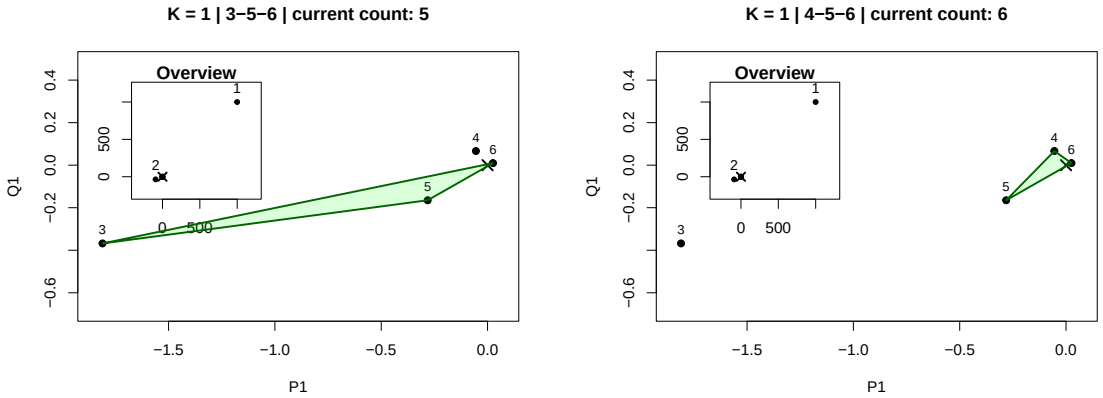

Figure 38: Illustration of the simplex depth for $K = 1$ on a subset of six observations generated with the state-dependent-noise EKF, including one positive outlier (part 4). The main panel shows the zoomed-in region used for the simplex depth, while the inset provides an overview of the complete subset and indicates the location of the zoomed region. Triangles containing the origin are highlighted in green, whereas triangles not containing the origin are outlined in red.

B.3 Outliers

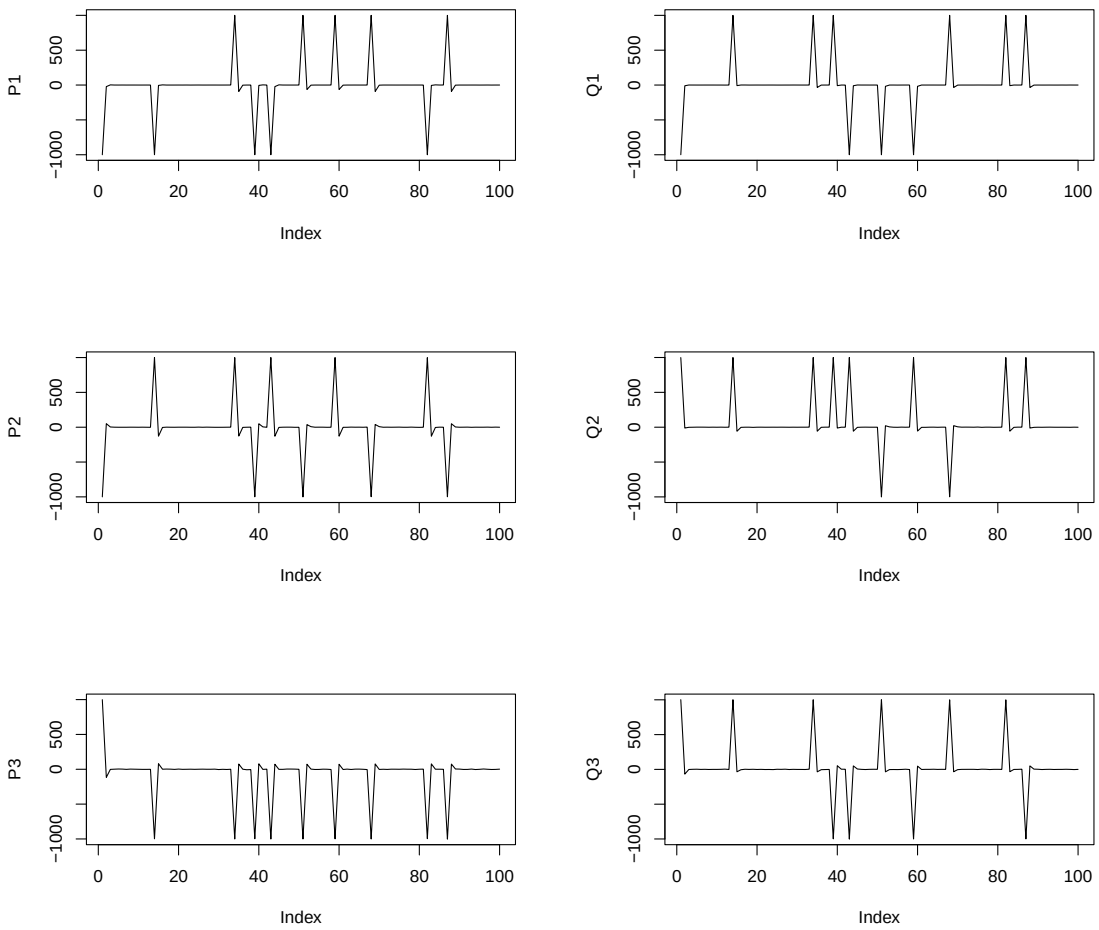

Figure 39: Example 10 Outliers two sided, sdn EKF.

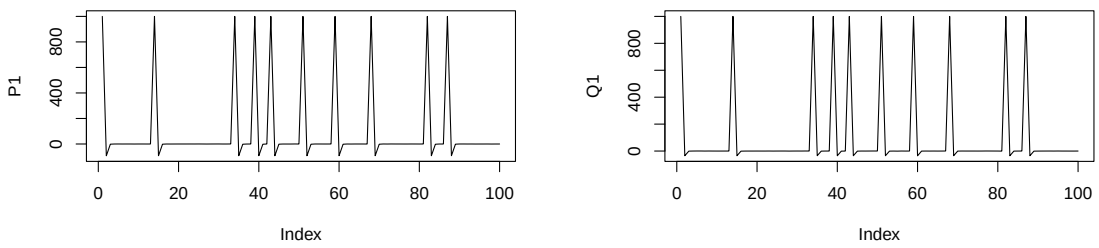

Figure 40: Example 10 Outliers one sided, sdn EKF.

B.4 Rejection rates of the simulation study

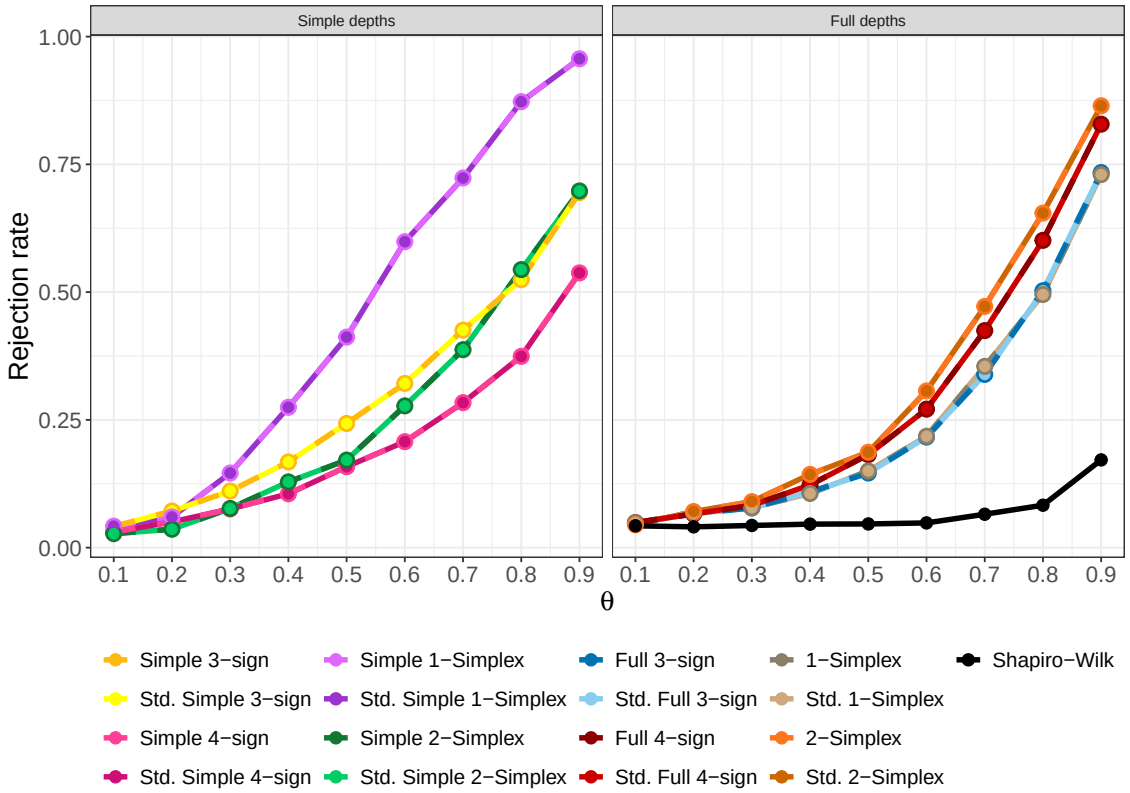

Figure 41: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the state-dependent-noise EKF and standardised residuals (true parameter $\theta = 0.1$).

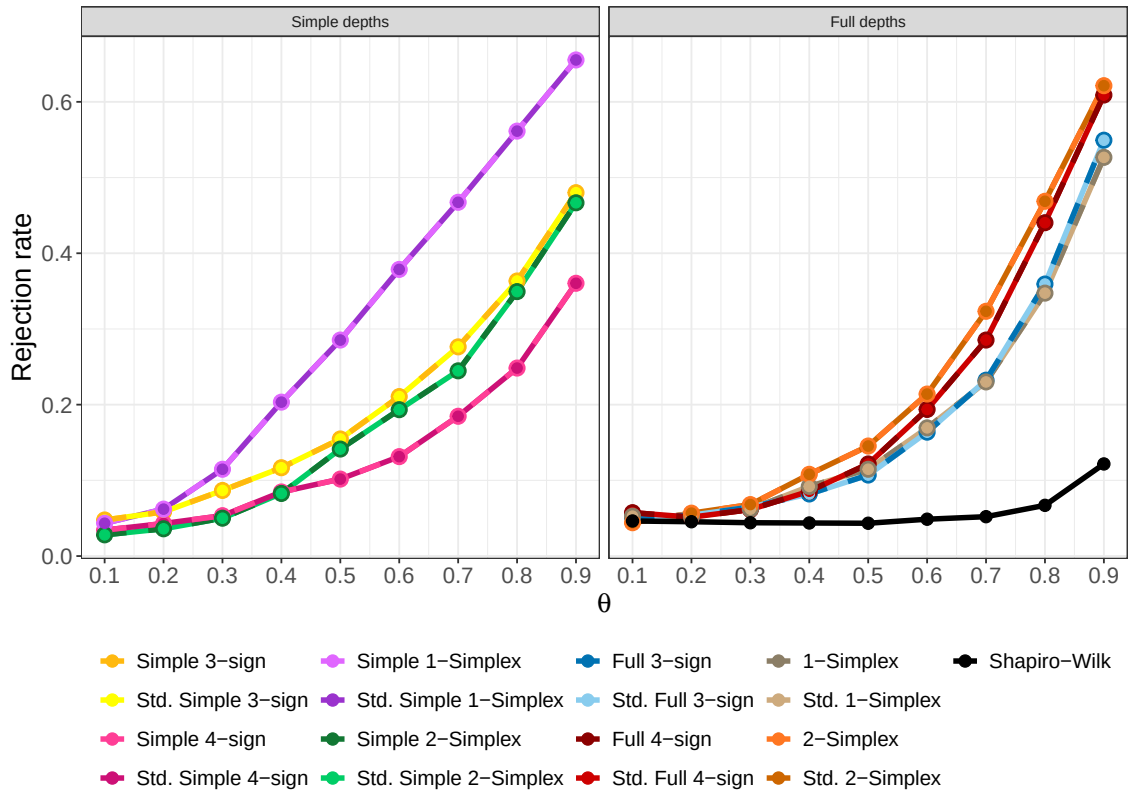

Figure 42: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the constant-noise EKF and standardised residuals (true parameter $\theta = 0.1$).

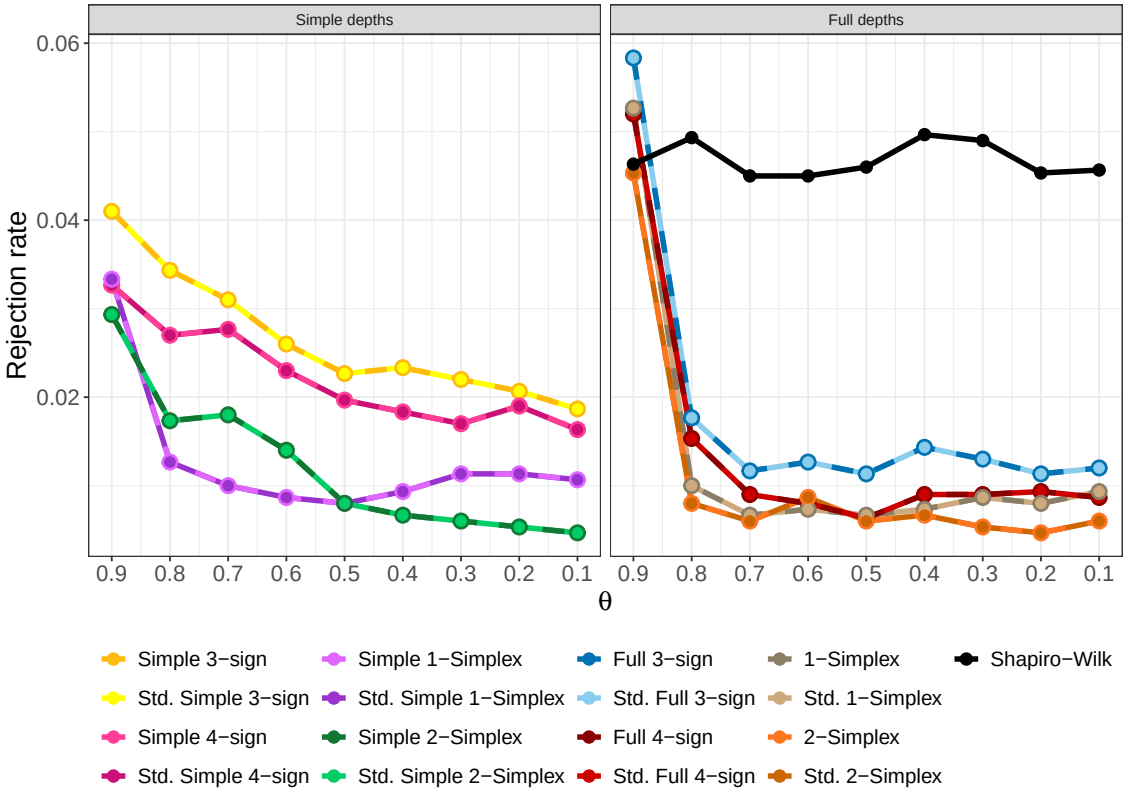

Figure 43: Rejection rates of the depth value methods and the Shapiro-Wilk test for different wrong parameter θ used in the voltage process in the simulation with the constant-noise EKF and non standardised residuals (true parameter $\theta = 0.9$).

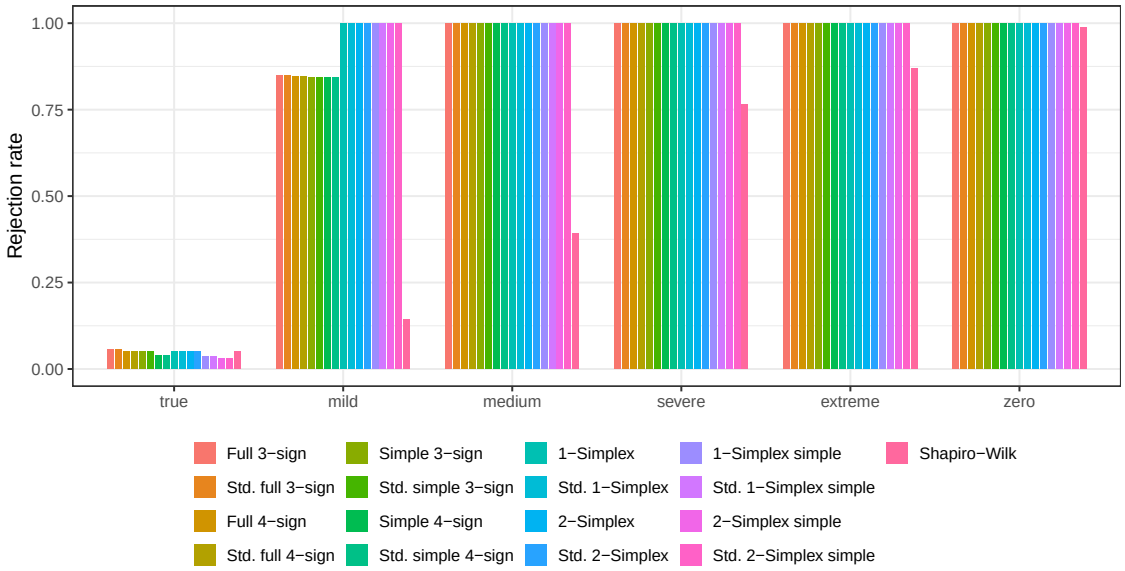

Figure 44: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the state-dependent-noise EKF and standardised residuals.

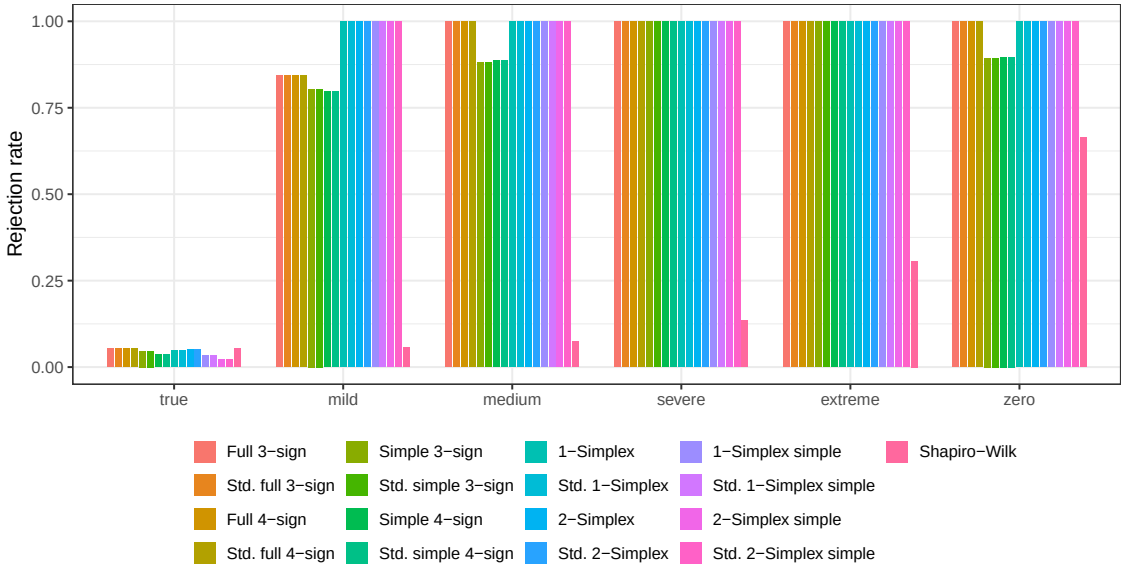

Figure 45: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the constant-noise EKF and standardised residuals.

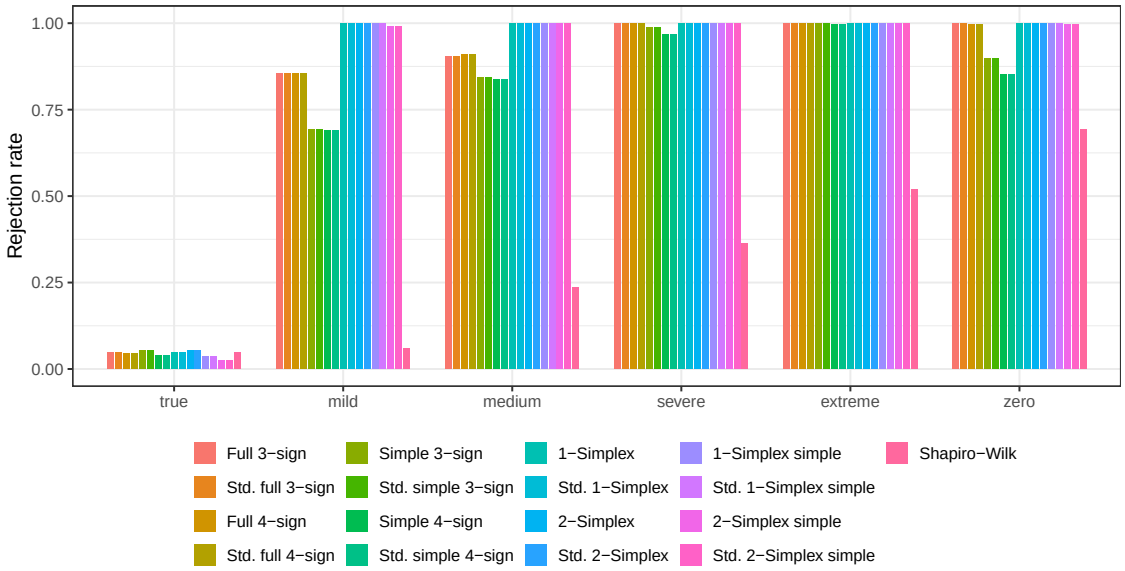

Figure 46: Rejection rate of the depth value methods and the Shapiro-Wilk test for different wrong parameters v_0 used in the prediction of the filter with the constant-noise EKF and non standardised residuals with the parameter $\theta = 0.9$.

Figure 47: Rejection rates for the standardised residuals of the state-dependent-noise EKF for positive additive outlier scenarios with 2, 10, and 15 outliers.

Figure 48: Rejection rates for the standardised residuals of the constant-noise EKF for positive additive outlier scenarios with 2, 10, and 15 outliers.

Figure 49: Rejection rates for the standardised residuals of the state-dependent-noise EKF for two-sided additive outlier scenarios with 2, 10, and 15 outliers.

Figure 50: Rejection rates for the standardised residuals of the constant-noise EKF for two-sided additive outlier scenarios with 2, 10, and 15 outliers.