Systematic Estimation of Noise Statistics for Nonlinear Kalman Filters[1]

Halvor Aarnes Krog*. Johannes Jäschke.*

**Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Sem Sælandsvei 4, Kjemiblokk 5, 7491 Trondheim, Norway (e-mail: halvor.a.krog@ntnu.no, johannes.jaschke@ntnu.no).*

Abstract: We propose two new systematic and easy-to-implement online tuning strategies for nonlinear Kalman filters with low computational cost. The tuning strategies assume the process and measurement noise are due to parametric uncertainty. We assume n_θ uncertain parameters which are translated into noise statistics by either i) generalized unscented transformation with $2n_\theta$ extra online model evaluations at every time step or ii) latin hypercube sampling, where the user sets the number of samples. Both approaches are distribution free, hence, the tuning strategies work for all kind of distributions. In the case study, it was found that the two proposed tuning strategies outperform the standard approach of fixed, diagonal noise matrices. In the case study, we further found that tuning based on the generalized unscented transformation seems to be more consistent than the method based on latin hypercube sampling for the same online computational cost. In addition, a Monte Carlo based tuning with modal noise adjustment is tested with promising performance. The modal noise adjustment is interesting as we can estimate the most likely point value of the noise (the mode of the noise distribution) and add this term to the state- and measurement equations at every time step.

Keywords: Estimation and Filtering, Generalized Unscented Transformation

1. INTRODUCTION

Many modern control algorithms require knowledge of the state variables $\boldsymbol{x}(t)$ of a process model. When it is not possible to measure the states directly, or the measurements are not of satisfactory quality, state estimation techniques are used. The Kalman Filter (KF) is arguably the most famous state estimator. By coupling the uncertainty in a process model with uncertainties of measurements, $\boldsymbol{y}(t)$, the KF obtains an estimate of the state, $\hat{\boldsymbol{x}}(t)$, and the corresponding covariance $\boldsymbol{P}$ of that estimate. This is the minimum variance estimate given that the measurement noise is Gaussian and the process model is linear. For non-linear models, the framework has been expanded to the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF), which both can appear in many forms.

The main issue with variants of Kalman filters is that they are difficult to tune to obtain satisfactory performance. This paper presents a systematic and easy-to-implement online tuning strategy of the process- and measurement noise covariance matrices $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$. The proposed approaches have a low additional on-line computational cost. The general idea is based on Valappil and Georgakis (2000). The major assumption in their work is that there is only parametric mismatch between the plant and the process model. However, their method also worked in presence of some structural plant-model mismatch (Valappil and Georgakis, 2000). In their case studies, the best performance of the filter was obtained by running Monte Carlo (MC) simulations on the parametric uncertainty distribution N_{MC} times. $\boldsymbol{Q}_k$ is then calculated at every time step as the sample covariance matrix of these N_{MC} runs. One drawback of their method is it may require N_{MC} to be very large to obtain a good estimate of the process noise covariance matrix. This is especially important if the probability distributions of the parameters have fat tails. A large N_{MC} imposes a heavy online computational burden, and if the sampling is random, there is no guarantee that the tails of the distribution are explored.

In this paper, we propose two new methods to address the issue of high computational cost while attaining similar or better information about the covariance matrix $\boldsymbol{Q}_k$. First, we propose to use latin hypercube sampling (LHS) in combination with MC simulations. Helton and Davis (2003) noted that LHS is the most broadly applicable approach to the propagation and analysis of uncertainty in complex systems, and often the only approach that is needed. LHS typically reduces the number of sample points required for MC simulations, and it ensures that the tails of the distribution are explored. A drawback of LHS is that it requires the parameter distributions to be independent. The second approach proposed in this paper handles dependent distributions in a natural manner. In particular, we propose to apply the generalized unscented transformation (GenUT), as presented by Ebeigbe et al. (2021), to estimate $\boldsymbol{Q}_k$. The GenUT works for all probability distributions, also correlated ones. If there are n_θ uncertain parameters, the extra online computational burden of obtaining $\boldsymbol{Q}_k$ by the GenUT is

[1] The authors acknowledge financial support from the Norwegian Research Council, SUBPRO, grant number: 237893. Code for the case study is available on: https://github.com/Process-Optimization-and-Control/Systematic-Estimation-of-Noise-Statistics-for-Nonlinear-Kalman-Filters

generally $2n_\theta + 2$ model evaluations. However, if the filter uses the mean values of the parameters the online cost is only $2n_\theta$. The method is computationally efficient for all distributions of the parameters, including distributions with heavy tails. These tuning strategies are applicable for both the UKF and EKF, but only the UKF is presented in this work.

This paper contains three contributions. To increase the accuracy and decrease the computational time when estimating the process noise, we propose to use either the GenUT or LHS. This is the focus in section 3.1 and 3.2. However, if the distribution of the process noise is not symmetrical, the most likely point is the mode and not the mean. The third contribution is therefore to use MC simulation to determine the mode of the process noise, and then use the mode instead of the mean in the Kalman filter. This is only possible for the proposed MC and LHS approaches. In some cases, this can improve the estimator performance. This method is described in section 3.3.

2. BACKGROUND

2.1 Nonlinear Kalman Filtering

For thorough background theory of state estimation, refer to Simon (2006). A general discrete nonlinear system is given by

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \overline{\boldsymbol{\theta}}^{fx}, \boldsymbol{u}_{k-1}, t_k, \boldsymbol{w}_{k-1}) \quad (1)$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \overline{\boldsymbol{\theta}}^{hx}, t_k, \boldsymbol{v}_{k-1}) \quad (2)$$

where $\overline{\boldsymbol{\theta}}^{fx}, \overline{\boldsymbol{\theta}}^{hx}$ are the deterministic parameters for the state update and measurement equations, $\boldsymbol{u}_k$ are the manipulated variables, $\boldsymbol{w}_k \sim (\boldsymbol{0}, \boldsymbol{Q}_k)$ the zero-mean process noise with covariance $\boldsymbol{Q}_k$, and $\boldsymbol{v}_k \sim (\boldsymbol{0}, \boldsymbol{R}_k)$ the measurement noise. The subscript k denotes the discrete time step. The states are denoted by $\boldsymbol{x} \in \mathbb{R}^{n_x}$, the measurements by $\boldsymbol{y} \in \mathbb{R}^{n_y}$, and the continuous time by t. The filter is initialized with:

$$\hat{\boldsymbol{x}}_0^+ = \mathbb{E}[\boldsymbol{x}_0] \quad (3)$$

$$\boldsymbol{P}_0^+ = \mathbb{E}[(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0^+)(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0^+)^{\boldsymbol{T}}] \quad (4)$$

The $\hat{\boldsymbol{x}}$ indicates it is an estimate, while $\boldsymbol{x}$ is the true, uncertain state. The superscript "+" are the *a posteriori* estimate, while a "-" means the *a priori* estimate. *A priori* estimates have only used the state equations (1) for predicting the state at time k, while *a posteriori* estimates have also used equation (2) to process the measurements at time k. At every time step, the estimate of the state and covariance are propagated. If the propagation is done by linearizing equation (1) and (2), this leads to the EKF. If the unscented transformation (UT) is used to propagate the mean and covariance, this leads to the UKF, as proposed by Julier et al. (2000). There are many good sources which describes the EKF and UKF in detail, see e.g. Simon (2006) or Barfoot (2017).

The filter designer needs to provide the filter with reasonable values for $\hat{\boldsymbol{x}}_0^+, \boldsymbol{P}_0^+, \boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ to obtain satisfactory estimates. A good discussion about how to set these quantities are in Schneider and Georgakis (2013). Typically, the most challenging tuning parameter is the process noise matrix $\boldsymbol{Q}_k \in \mathbb{R}^{n_x \times n_x}$. If there is parametric uncertainty in the measurement equation, it can also be challenging to find a good value for $\boldsymbol{R}_k \in \mathbb{R}^{n_y \times n_y}$. Our paper proposes two methods for automatic tuning of these noise matrices. In the following discussion, we focus on obtaining $\boldsymbol{Q}_k$ for brevity, but the method is applicable analogous for obtaining information about $\boldsymbol{R}_k$.

2.2 Tuning of noise covariance matrices based on parametric uncertainty

An actual plant is described by (1). Our focus will be on the process noise, $\boldsymbol{w}_k$, which is the stochastic term. This noise can be due to unmodeled effects or model inaccuracies. It is assumed that the model inaccuracies dominate the process noise component, and that structural plant-model mismatch can be captured as uncertainty in the model parameters, see Valappil and Georgakis (2000). Model parameters are typically estimated from experiments, and it is a crucial step that the uncertainty of the parameter estimates is recorded. Let $\boldsymbol{\theta}^{fx} \sim \boldsymbol{F}_{fx}$, where $\boldsymbol{F}_{fx}$ can be any probability distribution, and similarly $\boldsymbol{\theta}^{hx} \sim \boldsymbol{F}_{hx}$. The actual plant in (1) can be reformulated as a stochastic process:

$$\boldsymbol{x}_k^{true} = \boldsymbol{f}\left(\boldsymbol{x}_{k-1}^{true}, \boldsymbol{\theta}^{fx}, \boldsymbol{u}_{k-1}, t_k\right) \quad (5)$$

The state estimator is using a fixed, deterministic value of the parameters, $\overline{\boldsymbol{\theta}}^{fx/hx}$, and a nominal value for the states as below:

$$\boldsymbol{x}_k^{nom} = \boldsymbol{f}\left(\boldsymbol{x}_{k-1}^{nom}, \overline{\boldsymbol{\theta}}^{fx}, \boldsymbol{u}_{k-1}, t_k\right) \quad (6)$$

To exactly match the actual plant, an additive noise term $\widetilde{\boldsymbol{w}}_k \sim (\overline{\boldsymbol{w}}_k, \boldsymbol{Q}_k)$ is added to (6) which gives (7). Rearranging equation (7) and substituting (5) gives (8), which we will use to estimate $\widetilde{\boldsymbol{w}}_k$.

$$\boldsymbol{x}_k^{nom} + \widetilde{\boldsymbol{w}}_k = \boldsymbol{x}_k^{true} \quad (7)$$

$$\widetilde{\boldsymbol{w}}_k = \boldsymbol{f}\left(\boldsymbol{x}_{k-1}^{true}, \boldsymbol{\theta}^{fx}, \boldsymbol{u}_{k-1}, t_k\right) - \boldsymbol{x}_k^{nom} \quad (8)$$

By applying the information about the uncertainty of the parameters, it is possible to get accurate statistics of $\widetilde{\boldsymbol{w}}_k$ at every time step. This is in contrast to the "standard" way of using the filters, where equation (6) is used and a fixed hand-tuned diagonal matrix $\boldsymbol{Q}$ is added. Valappil and Georgakis (2000) proposed two methods of exploiting the information about the uncertainty of $\boldsymbol{\theta}^{fx}$. In the first method, the covariance matrix of the parameters, $\boldsymbol{P}_{\boldsymbol{\theta}}$, was propagated through the system by linearizing (8). This gives $\overline{\boldsymbol{w}}_k = \boldsymbol{0}$ and an estimate of $\boldsymbol{Q}_k$ accurate to the 1st order of the Taylor series, as in equation (9). This method has been successfully used in e.g. Tuveri et al. (2021) and Nagy and Braatz (2003).

$$\boldsymbol{J}_{p,nom} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{x}_{nom,k-1}, \overline{\boldsymbol{\theta}}^{fx}, \boldsymbol{u}_{k-1}, t_k}$$

$$\boldsymbol{Q}_k = \boldsymbol{J}_{p,nom} \boldsymbol{P}_{\boldsymbol{\theta}} \boldsymbol{J}_{p,nom}^T \quad (9)$$

In the second method, Valappil and Georgakis (2000) estimated $\widetilde{\boldsymbol{w}}_k$ by MC integration. Under the assumption that $\boldsymbol{x}_{k-1}^{true} \approx \hat{\boldsymbol{x}}_{k-1}^+$, we can take N_{MC} random samples from the distribution of $\boldsymbol{\theta}^{fx}$ and propagate them through (8). $\overline{\boldsymbol{w}}_k, \boldsymbol{Q}_k$ are then the sample mean and covariance of these N_{MC} runs.

Note that $\bar{\boldsymbol{w}}_k \neq \mathbf{0}$ in general, and that the mean of the process noise can be used in the state propagation step as below:

$$\hat{\boldsymbol{x}}_k^- = \mathbb{E}[\boldsymbol{f}(\boldsymbol{x}_{k-1}, \bar{\boldsymbol{\theta}}^{fx}, \boldsymbol{u}_{k-1}, t_k)] + \bar{\boldsymbol{w}}_{k-1} \tag{10}$$

In all the case studies in Valappil and Georgakis (2000), these two adaptive methods of setting $\bar{\boldsymbol{w}}_k, \boldsymbol{Q}_k$ showed superior performance compared to a fixed, diagonal hand-tuned $\boldsymbol{Q}$. As expected, the MC-based approach consistently performed better than the linearization approach, with the disadvantage of a significantly higher on-line computational cost.

2.3 Generalized unscented transformation

In the UKF, the mean and covariances were estimated by the UT. The UT approximates any nonlinear stochastic equation by creating a set of weights and sigma points. The sigma points are propagated through the nonlinear equation, and the weighted mean and covariance of the transformed points are estimated. The UT is accurate to the 3rd order of the Taylor series in approximating means and covariances, but it is only valid for symmetrical distributions. To approximate the mean and covariance for *any* distribution, one can instead use the GenUT as proposed by Ebeigbe et al. (2021).

The GenUT is accurate to the 3rd order of the Taylor series for independent distributions and 2nd order accuracy for correlated distributions. To generate the sigma points for the uncertain distribution $\boldsymbol{\theta}$, the mean, covariance, $CM3_i$ and $CM4_i$ for $i \in [1, n_\theta]$ must be available. Here, CMn is the central moment of order n. Refer to Ebeigbe et al. (2021) for a detailed discussion and details for how the weights and sigma-points $(W_{\boldsymbol{\theta}}^{(i)}, \boldsymbol{\chi}_{\boldsymbol{\theta}}^{(i)})$ for $i \in [0, 2n_\theta]$ are obtained. By assuming that $\boldsymbol{x}_{k-1}^{true} \approx \hat{\boldsymbol{x}}_{k-1}^+$, the mean and covariance of equation (8) is approximated by equation (11)-(13).

$$\tilde{\boldsymbol{w}}_k^{(i)} = \boldsymbol{f}(\hat{\boldsymbol{x}}_{k-1}^+, \boldsymbol{\chi}_\theta^{(i)}, \boldsymbol{u}_{k-1}, t_k) - \boldsymbol{x}_k^{nom} \tag{11}$$

$$\bar{\boldsymbol{w}}_k = \sum_{i=0}^{2n_\theta} W_{\boldsymbol{\theta}}^{(i)} \tilde{\boldsymbol{w}}_k^{(i)} \tag{12}$$

$$\boldsymbol{Q}_k = \sum_{i=0}^{2n_\theta} W_{\boldsymbol{\theta}}^{(i)} (\tilde{\boldsymbol{w}}_k^{(i)} - \bar{\boldsymbol{w}}_k)(\dots)^T \tag{13}$$

where (...) means that it is a copy of the previous parenthesis.

2.4 Latin hypercube sampling

LHS is a well-known sampling procedure, which has the benefits that i) the samples cover the whole range of the uncertain parameter space and ii) it has an easy implementation. As the samples are guaranteed to be spread out over the whole parameter space, fewer samples are typically required for LHS compared to random sampling. This translates into a lower online computational cost when approximating the noise covariance matrices. The LHS can sample from any distribution, $\boldsymbol{\theta} \sim \boldsymbol{F}$, given that the parameters are independent. Correlations can be induced by combining LHS with the method in Iman and Conover (1982). For a detailed comparison about different sampling schemes, see Helton and Davis (2003).

3. PROPOSED METHODS FOR ESTIMATING NOISE STATISTICS IN NONLINEAR SYSTEMS

3.1 GenUT to estimate noise statistics

Assume the mean, covariance, 3rd and 4th central moments are available for our parameter distribution. These are denoted as $\boldsymbol{\mu_\theta}, \boldsymbol{P_\theta}, \boldsymbol{CM3}$ and $\boldsymbol{CM4}$, respectively. This information is readily available from the posterior distribution from a Bayesian parameter estimation, or it can be calculated from any specified probability distribution.

Before the filter is initialized, calculate offline $2n_\theta + 1$ sigma points $\boldsymbol{\chi_\theta} \in \mathbb{R}^{n_\theta \times (2n_\theta + 1)}$ and their weights $\boldsymbol{W_\theta} \in \mathbb{R}^{2n_\theta + 1}$. The i-th sigma point is denoted $\boldsymbol{\chi}_{\boldsymbol{\theta}}^{(i)} \in \mathbb{R}^{n_\theta}$, and it is equivalent to the $(i+1)$-th column in $\boldsymbol{\chi_\theta}$. Generating the sigma points is the most computational costly part of the GenUT algorithm as it involves calculating the matrix square root, which is typically the Cholesky factorization. It is emphasized that this step is performed offline and does not affect the online computational cost.

Initialize the filter. At every time step, propagate the $2n_\theta + 1$ sigma points and estimate the noise statistics by (11)-(13). Note that in this step, the state estimate $\hat{\boldsymbol{x}}_{k-1}^+$ is treated as a fixed quantity. The covariance matrix of the states, $\boldsymbol{P}_k$, is therefore not used in this step. A visualization of the process is shown in Figure 1. As $\boldsymbol{x}_k^{nom}$ is also calculated online, the total computational online cost is $2n_\theta + 2$ model evaluations. If we select $\bar{\boldsymbol{\theta}}^{fx} = \mathbb{E}[\boldsymbol{\theta}^{fx}] \stackrel{\text{def}}{=} \boldsymbol{\chi}_{\boldsymbol{\theta}}^{(0)}$, then $\tilde{\boldsymbol{w}}_k^{(0)} = \mathbf{0}$ and the additional cost is $2n_\theta$ model evaluations. Note that $\boldsymbol{x}_k^{nom}$ must still be calculated, but in the UKF that corresponds to propagating the zeroth sigma-point of the states, $\boldsymbol{\chi}_{k-1}^{(0)}$, through the state-update equation and it is therefore not an extra cost.

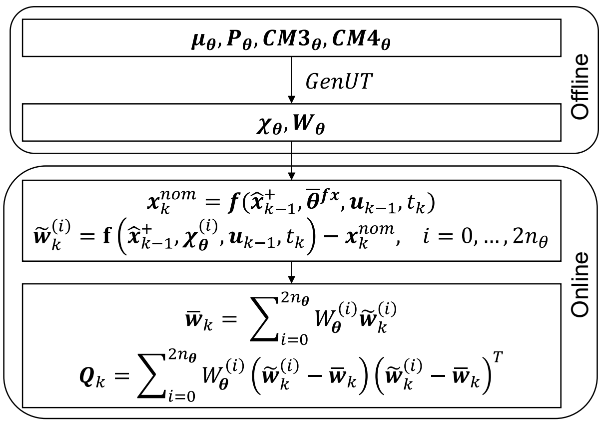

Figure 1: Tuning of noise statistics based on the GenUT. The sigma points and their weights are computed offline and propagated using the model equations in the online part.

Remark: If the measurement statistics $\bar{\boldsymbol{v}}_k, \boldsymbol{R}_k$ are to be estimated, the approach is the same. That is, use i) statistics from $\boldsymbol{\theta}^{hx}$ instead of $\boldsymbol{\theta}^{fx}$ to generate the sigma points, and ii) the a priori estimate $\hat{\boldsymbol{x}}_k^-$ and the measurement equation $\boldsymbol{h}(\cdot)$ in (11) instead of $\hat{\boldsymbol{x}}_{k-1}^+$ and $\boldsymbol{f}(\cdot)$.

3.2 LHS to estimate noise statistics

First, we specify the distribution of the parameters and generate N_{LHS} samples from the latin hypercube, which is a feature in most scientific computational packages. At every timestep, the N_{LHS} samples are propagated through (8). The noise statistics are now estimated as the sample mean and covariance of these N_{LHS} runs. This procedure is illustrated in Figure 2.

A key question is how large must N_{LHS} be to obtain a satisfactory approximation of the noise statistics. This is case dependent, and important factors are i) the distributions of the parameters, ii) the function every sample is propagated through and iii) available computational resources.

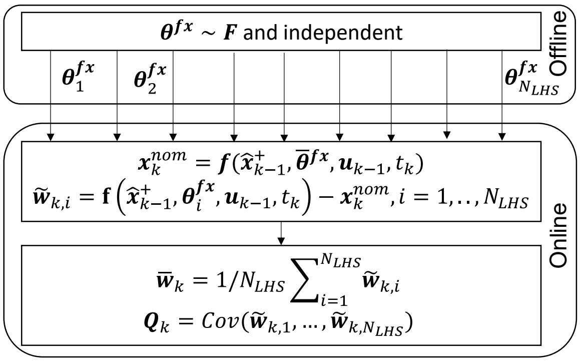

Figure 2: Tuning of noise matrix based on MC and LHS. Samples of parameters $\boldsymbol{\theta}^{fx}$ are generated offline using LHS, and these samples are used to estimate noise statistics in the online part.

3.3 Selecting the most likely noise term (the mode)

For skewed distributions, the mode and mean are not the same. Setting $\bar{\boldsymbol{w}}_k$ as the point value with highest probability density might therefore be a better choice than the mean. Hence, we are setting $\bar{\boldsymbol{w}}_k = \tilde{\boldsymbol{w}}_k^m = Mode(\tilde{\boldsymbol{w}}_k)$. See section 4.3 for how we find $\tilde{\boldsymbol{w}}_k^m$ in this paper. Note that we are implicitly breaking the assumption of $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ to be zero-mean in (1) and (2). However, the filter may still work even if all its assumptions are not fulfilled, see Simon (2006). Finding the mode is only possible for a sampling-based approach such as MC or LHS.

4. CASE STUDY: FALLING BODY

The problem investigated by Julier et al. (2000) and Simon (2006) is studied. A body falls into the atmosphere from a very high altitude and velocity. We want to estimate the altitude $x_1(t)$, velocity $x_2(t)$ and a constant ballistic coefficient $x_3(t)$. At our guidance, we have a radar measurement which is located at a height a and a horizontal distance M from the falling body. The dynamics of the system and the measurements are described by:

$$\dot{x}_1(t) = x_2(t) + w_1 \tag{14}$$

$$\dot{x}_2(t) = \frac{1}{2}\rho_0 e^{-\frac{x_1}{k}} x_2^2 x_3 - g + w_2 \tag{15}$$

$$\dot{x}_3(t) = w_3 \tag{16}$$

$$y(t_k) = \sqrt{M^2 + (x_1(t_k) - a)^2} + v_k \tag{17}$$

Here, ρ_0 is the air density at sea level, k is a constant relating the air density and altitude and g is the gravitational acceleration. Process noise is entering through $\boldsymbol{w}$ and the repeatability of the measurements are defined by $v \sim \mathcal{N}(0, \sigma_v^2)$ where $\sigma_v^2 = 1E4$ ft^2. A measurement is obtained every 0,5 second and a Runge-Kutta method of order 4 with adaptive step size is used to integrate the process model. The system and the state estimators are initialized with:

$$\boldsymbol{x}(0) = [3E5, -2E4, 1E-3]^T \tag{18}$$

$$\boldsymbol{P}_0^+ = diag([3E8, 4E6, 1E-6]) \tag{19}$$

$$\hat{\boldsymbol{x}}^+(0) = \boldsymbol{x}(0) + \sqrt{diag(\boldsymbol{P}_0^+)} \tag{20}$$

Following our assumption, $\boldsymbol{w} = \boldsymbol{0}$ as all process noise stems from a fixed parametric plant-model mismatch. Measurement noise is both due to parametric mismatch and white noise from v_k. The set of uncertain parameters are $\boldsymbol{\theta}^{fx} = [\rho_0, k]^T$ and $\boldsymbol{\theta}^{hx} = [M, a, v]^T$, while $g = 32{,}2\ ft/s^2$ is a deterministic parameter. We assume that parameter estimation of $\rho_0, k, M$and a has found them to be independent and following Gamma distributions given in Table 1. We select the mean of the parameter distribution for the UKF, while the system selects the most probable value, i.e. $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \boldsymbol{\theta}_{UKF} \neq \boldsymbol{\theta}_{true} = Mode(\boldsymbol{\theta})$.

Table 1: Parameter values used by the system, $\boldsymbol{\theta}_{true}$, and the UKF, $\boldsymbol{\theta}_{UKF}$. The distribution for $\boldsymbol{\theta}$ is $\boldsymbol{\theta} \sim Offset + \boldsymbol{\Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

$\boldsymbol{\theta}$	Offset	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\theta}_{true}$	$\boldsymbol{\theta}_{UKF}$
ρ_0(lb-s^2/ft^4)	1,8	4	2E1	1,95	2
k (ft)	1,8E4	4	2E-3	1,95E4	2E4
M (ft)	9,0E4	4	4E-4	9,75E4	1E5
a (ft)	9,0E4	4	4E-4	9,75E4	1E5

The performance of the state estimator is now only dependent on the noise statistics $(\tilde{\boldsymbol{w}}_k, \tilde{v}_k)$ since $\hat{\boldsymbol{x}}_0^+, \boldsymbol{P}_0^+$ are given by equation (19)-(20). The three proposed methods will be compared to two benchmark methods (UKF with MC (Valappil and Georgakis (2000), and fixed $\boldsymbol{Q}$). See section 4.1-4.4 for details on the five methods for this case study. The performance indicators for the five estimators are the additional online computational cost and the accuracy of the method. Table 2 compares the computational cost for the methods in terms of function evaluations. The accuracy of the methods is assessed by equation (21), which is the same cost function as in Valappil and Georgakis (2000).

$$J_i = \frac{\sqrt{\sum_{k=0}^{t_{end}} \left(\hat{x}_{i,k}^{+} - x_{i,k}\right)^2}}{\sqrt{\sum_{k=0}^{t_{end}} \left(x_{i,k}^{OL} - x_{i,k}\right)^2}} \tag{21}$$

Here, the state estimate error is compared against the open loop (OL) error for every state i. The OL simulation has the same parameters and is initialized with the same estimate as the filters, but it does not take measurements into account. Hence, if the cost function is less than 1, it means that the filter is improving its estimates by using the measurements.

4.1 Tuning based on the GenUT

The sigma points and weights $((\boldsymbol{\chi}_{\boldsymbol{\theta}}^{fx}, \boldsymbol{W}_{\theta}^{fx}), (\boldsymbol{\chi}_{\theta}^{hx}, \boldsymbol{W}_{\theta}^{hx}))$ for the GenUT were calculated before the filter was initialized. Since $\boldsymbol{\theta}^{fx} \in \mathbb{R}^2$, then $\boldsymbol{W}^{fx} \in \mathbb{R}^5$ and $\boldsymbol{\chi}_{\boldsymbol{\theta}}^{fx} \in \mathbb{R}^{2\times 5}$. In accordance with section 3.1, the additional computational cost of estimating $\boldsymbol{Q}_k$ by the GenUT is therefore 4 evaluations of $f(\cdot)$ at every sampling time. For estimating $\boldsymbol{R}_k$ we have that $\boldsymbol{\theta}^{hx} \in \mathbb{R}^3$, so 6 evaluations of $h(\cdot)$ was done online at each time step. Note that we set $(\overline{\boldsymbol{w}}_k, \bar{v}_k) = (\mathbf{0}, 0)$. The reason is that the mean of $\widetilde{\boldsymbol{w}}_k, \tilde{v}_k$ is further away from the mode than the point $\mathbf{0}$.

4.2 Tuning based on LHS

The noise matrices $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ were estimated by LHS. N_{LHS} were set to 10 in order to explore the parameter space while keeping a reasonable computational cost. We set $(\overline{\boldsymbol{w}}_k, \bar{v}_k) = (\mathbf{0}, 0)$ for the same reason as in section 4.1.

4.3 Tuning based on modal noise adjustment

A MC filter with adaptive noise matrices and $(\overline{\boldsymbol{w}}_k, \bar{v}_k) = (\widetilde{\boldsymbol{w}}_k^{\mathrm{m}}, \tilde{v}_k^{\mathrm{m}})$ was tested, in accordance with section 3.3. A brute force approach was used to find the mode. We ran a MC simulation with $N_{MC} = 500$ random samples at every time step, and a histogram of $\widetilde{\boldsymbol{w}}_k$ with 20 bins were built. The mode was found as the bin in the histogram with the most points. The filter is denoted MCm.

4.4 Tuning based on benchmark methods

For comparison, we use UKFs where i) the noise matrices were found by MC with $N_{MC} = 500$ random samples as in Valappil and Georgakis (2000) and ii) a manually tuned UKF with fixed noise matrices. For the MC filter, we set $(\overline{\boldsymbol{w}}_k, \bar{v}_k) = (\mathbf{0}, 0)$, see section 4.1 for the reasoning behind this. For the fixed filter we set $\boldsymbol{Q}_{fixed} = diag([1E3, 1E3, 1E-8])$ and $R_{fixed} = 1E6 \gg \mathbb{E}[v^2]$. Note that normal approach is to set $R_{fixed} = \mathbb{E}[v^2]$. In this case study the filter will diverge with that tuning as the additional parametric uncertainty is not considered, meaning that the filter trusts the measurements too much.

Table 2: Computational cost in the case study for each method, given by the additional model evaluations at each time step compared to the "fixed" noise filter.

Model	GenUT	LHS	MC	MCm
$\tilde{w}_k(\cdot) = f(\cdot_k) - x_k^{nom}$	4	10	500	500
$\tilde{v}(\cdot) = h(\cdot_k) - y_k^{nom}$	6	10	500	500

4.5 Simulation results

The system was simulated for 30 seconds. The trajectories of the true state, state estimates and measurements sequence are shown in Figure 3.

The simulation shows that initially, the velocity x_2 is very high due to the low drag effect at high altitude x_1. After about 10 seconds the altitude has decreased which leads to an increase in the air density. The drag effect becomes more important, and x_2 converges to the terminal velocity. If we look at the measurements $y(t)$, we see that after about 10 seconds the body is at the same altitude as the radar measurement. The body is still falling and $y(t)$ reports that the distance to the body increases.

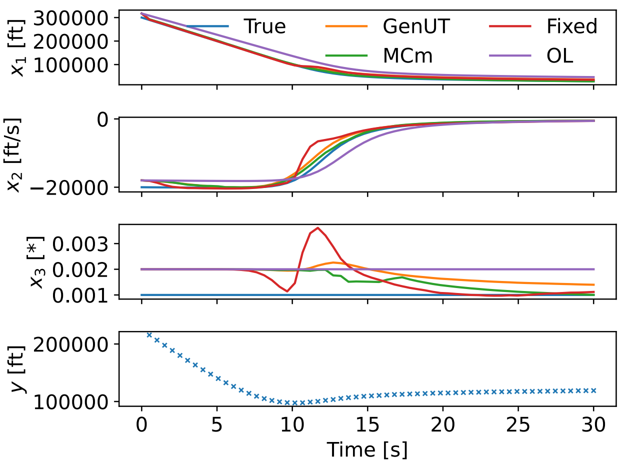

Figure 3: State trajectory, state estimates obtained by different UKF tuning strategies and measurement. Open loop (OL) simulation from $\hat{\boldsymbol{x}}^+(0)$ also included, while LHS and MC are resembling GenUT and are therefore not shown. *units of x_3 is $\frac{ft^3}{lb-s^2}$.

The simulation was repeated 100 times to observe if the results were consistent. The distribution of the cost function (21) is shown as a violin plot in Figure 4.

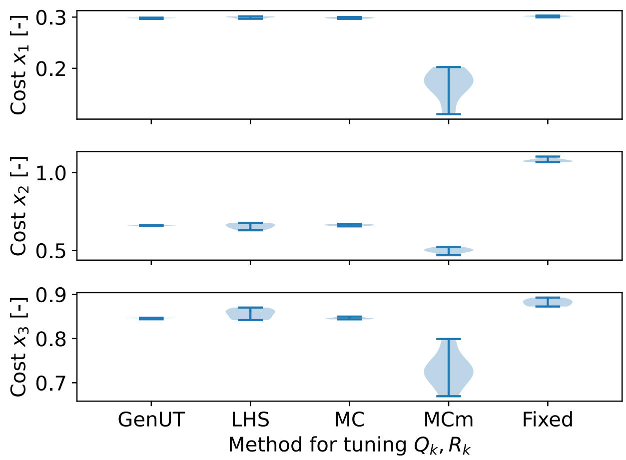

Figure 4: Distribution of the cost function after 100 runs for each state and tuning method. $N_{LHS} = 10$ and $N_{MC} = N_{MCm} = 500$.

To structure the discussion of the contributions of this paper, we first discuss the performance of the MCm tuning. Afterwards, we will discuss GenUT, LHS, MC and the Fixed filters.

4.6 Performance improvement of using the mode, MCm

As Figure 4 shows, a significant improvement of the cost function (21) was obtained by setting $(\bar{\boldsymbol{w}}_k, \bar{v}_k) = (Mode(\tilde{\boldsymbol{w}}_k), Mode(\tilde{v}_k))$. This is not a general result, but modal adjustment might improve estimator performance in certain cases. The standard deviation of the cost function for MCm is high due to the inaccurate method used to find the mode. Future work will explore this tuning method further.

4.7 Performance improvement of structured noise estimation

From Figure 4 we observe that the tuning methods have similar performance for x_1. This is reasonable as x_1 is the only state in the measurement equation, and $R_{fixed} \ll R_{other}$. For the remaining states, the proposed adaptive tuning methods outperform the manual tuning. Even with careful manual tuning of the noise matrices, the cost of x_2 is greater than one. The MC filter with many samples and high online computational cost is the reference method from Valappil and Georgakis (2000) which GenUT and LHS is compared to. For all the states, we observe that GenUT, LHS and MC has the same average performance, but there is less variation in the cost function by the GenUT-method. One way to decrease the variance in the cost function for LHS and MC is to increase the number of samples. Figure 5 shows how the variance of the cost decreases for the MC-based tuning when the sample number increases. To calculate the sample variance, every simulation is repeated 100 times. E.g. there has been 100 runs to calculate the sample variance when $N_{LHS} = N_{MC} = 1000$.

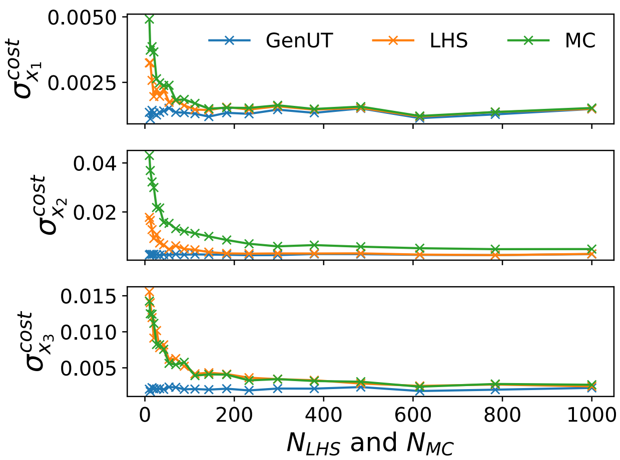

Figure 5: Variation of the cost function decreases for the LHS and MC-based tuning when the sample number increases. They converge to the standard deviation of the GenUT-based tuning.

As Figure 5 shows, the standard deviation in the cost function for LHS and MC approaches the GenUT-based tuning, given that the number of sample points are sufficiently high. Also note that the LHS-based tuning always has a lower standard deviation in the cost function than the MC-based tuning for the same computational cost. This is as expected from the theory.

Note that Gamma distributions were used for the parameters in this case study. This is a heavy-tailed distribution which are difficult to sample from. If the parametric distributions are e.g. Gaussian or Uniform, it is expected that σ^{LHS}_{cost} would be more similar to σ^{GenUT}_{cost} for a lower sampling number.

5. CONCLUSION

By assuming only parametric uncertainty in the process and measurement model, we proposed a GenUT-based and a LHS-based tuning method to estimate the noise statistics. The methods are i) easy to implement, ii) computationally fast and accurate and iii) eliminates the need of cumbersome hand-tuning of the state estimators. The drawback is the unavoidable increase in online computational cost compared to a fixed, but sub-optimal, noise matrix. The GenUT-based tuning easily handles correlated parameter distributions, while the LHS-based tuning only works for independent distributions in its original form.

A method for estimating the modal value for the noise, and utilizing this information in the state estimator, was also proposed. This method showed promising performance and will be subject for further work.

BARFOOT, T. D. 2017. *State Estimation for Robotics*, Cambridge University Press.

EBEIGBE, D., BERRY, T., NORTON, M. M., WHALEN, A. J., SIMON, D., SAUER, T. & SCHIFF, S. J. 2021. A Generalized Unscented Transformation for Probability Distributions. *ArXiv*.

HELTON, J. C. & DAVIS, F. J. 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety,* 81**,** 23-69.

IMAN, R. L. & CONOVER, W. J. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation,* 11**,** 311-334.

JULIER, S., UHLMANN, J. & DURRANT-WHYTE, H. F. 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE transactions on automatic control,* 45**,** 477-482.

NAGY, Z. K. & BRAATZ, R. D. 2003. Robust nonlinear model predictive control of batch processes. *AIChE Journal,* 49**,** 1776-1786.

SCHNEIDER, R. & GEORGAKIS, C. 2013. How To NOT Make the Extended Kalman Filter Fail. *Ind. Eng. Chem. Res,* 52**,** 3354-3362.

SIMON, D. 2006. Optimal state estimation : Kalman, H [infinity] and nonlinear approaches. Hoboken, N.J.: Wiley-Interscience.

TUVERI, A., PÉREZ-GARCÍA, F., LIRA-PARADA, P. A., IMSLAND, L. & BAR, N. 2021. Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring. *Journal of process control,* 106**,** 195-207.

VALAPPIL, J. & GEORGAKIS, C. 2000. Systematic Estimation of State Noise Statistics for Extended Kalman Filters. *AIChE journal,* 46**,** 292-308.