Particle Filter Approach for Prognostics Using Exact Static Parameter Estimation and Consistent Prediction

Kai Hencken[1], Arthur Serres[1,2], and Giacomo Garegnani[1]

[1] *Corporate Research Center, ABB Switzerland Ltd, Baden-Dättwil, CH-5405*
kai.hencken@ch.abb.com
giacomo.garegnani@ch.abb.com

[2] *current mail address:*
arthurserres1510@icloud.com

Abstract

Particle filters are widely used in model-based prognostics. They estimate the future health state of an asset based on measurement data and an assumed degradation dynamics. Filters are in general applied to estimate only the states given a known dynamics of the process. In model-based prognostics, the dynamics is assumed to be known in an analytical form, but the parameters vary per device and need to be learned from the measurements as well. This is especially important for the calculation of the remaining useful life (RUL), as the prediction of the future evolution is needed.

There are commonly used approaches for this: Augmenting the state space with the parameter, together with assuming them to stay constant or adding an artificial diffusive evolution to them. The Liu–West filter improves on this by modifying the artificial evolution such that mean and standard deviation of the marginal parameter distribution are kept the same. Both approaches require to choose some tuning parameters, which might be difficult in practical applications. In addition, the model parameter is often assumed frozen for the prediction part, leading to an inconsistency. We propose how a modification of the parameter evolution in case of missing measurements can solve this in both cases.

More recently algorithms for combined state estimation and exact parameter estimation have been introduced, especially the Storvik filter, based on the usage of a sufficient statistic. We analyze how this can be applied to overcome difficulties with existing approaches, avoiding the need for tuning parameters. We also extend the Storvik filter in order to deal with time-steps with missing measurements. Two formally equivalent approaches are presented. These are applicable in all cases of missing measurements, coming either from irregular data acquisition, e.g. only during maintenance or inspection, or as part of the prediction step of the RUL calculation.

We study the different methods for two simple models in order to demonstrate potential issues with existing approaches and to explore the stability of the new one based on the Storvik filter. Finally we apply it to a practical application in the area of electrical distribution systems.

1. Introduction

The most common approach for predicting the end-of-life (EOL) of a device is to model its degradation. Let x_t be a degradation variable, describing the health of the device and evolving e.g. with time t. In the simplest case, we define the (soft) failure of the device as the condition that x_t reaches a predefined critical value x_{critical} (Goebel et al., 2017; Galar, Goebel, Sandborn, & Kumar, 2021). x_t can here be either a scalar or a vector, see e.g. (Peng, Ye, & Chen, 2018). We limit ourselves to the scalar case.

The evolution of x_t is in general described by a stochastic model. We restrict ourselves here to the discrete-time case, indexed by t, and assume that the state evolves from time instance t to $t+1$ as

$$x_{t+1} \sim p(x_{t+1} \mid x_t; \theta), \tag{1}$$

where p is a suitable probability model depending on a parameter vector θ. We assume the value of θ to be specific to each individual device rather than describing the behavior of a fleet and a prior distribution $p(\theta)$ to be known.

In most cases the degradation variable x_t is not directly measurable but needs to be inferred from an observable z_t. This might be a direct measurement of x_t corrupted by measurement error or a quantity that can be indirectly associated with it. Quite generally the relation between the degradation variable x_t and the observable z_t is described by

$$z_t \sim p(z_t \mid x_t; \theta), \tag{2}$$

where p is a probability model possibly also depending on the parameter vector θ.

The combination of Eqs. (1) and (2) forms a state-space model (SSM) with unknown parameter θ.

The aim of prognostics is to calculate the remaining useful life given the measurements $z_{0:t_0}$ up to the current time horizon t_0, using the predictive probability

$$p(x_t \mid z_{0:t_0}), \quad t > t_0. \tag{3}$$

Note that in Eq. (3) the parameter θ does not appear, as it has been marginalized over. The uncertainty associated with θ is therefore automatically considered. Due to the stochastic nature of the evolution and the uncertainty in the parameter, the RUL is itself a random variable. Its distribution is given by

$$\begin{aligned} \mathrm{RUL}(t \mid t_0) \sim p(t \mid \{x_{t+t_0} \geq x_{\text{critical}}\} \\ \cap \{x_{t'} < x_{\text{critical}} \; \forall t' < t + t_0\}). \end{aligned} \tag{4}$$

A review of the prognostics paradigm and its applications can be found, e.g., in (Si, Wang, Hu, & Zhou, 2011; Jouin, Gouriveau, Hissel, Péra, & Zerhouni, 2016; Goebel et al., 2017; Galar et al., 2021).

Determining the RUL requires a combined state and parameter estimation approach. In principle, the estimation can be obtained with any Bayesian method, e.g., a Markov Chain Monte Carlo (MCMC) approach, that determines the joint distribution $p(x_{0:t}, \theta \mid z_{0:t_0})$ of all past, present and future states and the parameter using all measurements until time t_0. In practice, this approach is not viable, since the evaluation has to be repeated each time a new measurement point is added. Hence, the method becomes computationally more demanding as time increases. A sequential approach is more appropriate, such as a sequential Monte Carlo method (SMC), see e.g. (Doucet, de Freitas, & Gordon, 2001; Chopin & Papaspiliopoulos, 2020). This requires to update the joint distribution $p(x_t, \theta \mid z_{0:t_0+1})$ at each increase of the horizon taking into account only the new measurement z_{t_0+1} and the already known joint distribution until t_0. The determination of the distribution of the state $x_{0:t_0}$, especially of only the current state x_{t_0}, is a well-studied problem for known parameter values, regularly solved with the help of particle filters. In contrast, the determination of the joint distribution of states and parameter is a more difficult one and often solved by applying some approximations.

In addition to the estimation problem for x_{t_0}, prognostics applications require the ability to make predictions. Indeed, in order to compute the RUL as in Eq. (4) one needs to calculate the future distribution for x_t for $t > t_0$, see Eq. (3). A related topic is the ability to evolve the sequential approach when measurements are sparse and obtained at irregular time intervals only. For instance, we expect measurements to be of this form, if they are obtained as part of a maintenance or inspection routine. If the time interval between measurements is long, it is beneficial to evolve the distribution of state and parameter and only keep their values at the current time. It allows to continue to evolve the distribution up to the next measurement without the need to restart from the last measurement point.

In this paper, we consider the two issues above and present:

- A critical review of the state of the art on particle based sequential methods for joint state and parameter estimation in SSMs;
- Proposals on how to extend the methods to deal with missing measurements, required especially for future predictions;
- Explore the use of the Storvik filter as an exact approach to the combined state and parameter estimation problem for prognostics applications.

The remainder of this paper is organized as follows. In Sec. 2 we briefly review the application of particle filters to SSMs. In Sec. 3 we present common approaches to perform joint state and parameter estimation, introducing also a possible way to handle missing measurements. In Sec. 4 we review the Storvik filter, an exact method for the combined state and parameter estimation, together with an extension of the method in case of missing measurements in Sec. 5. In Sec. 6 we analyze the applicability of the methods for two simple models, that are typical for prognostics applications, and in Sec. 7 we present results of the application of the Storvik filter on real data. We conclude the paper with an outcome and give potential future directions in Sec. 8.

2. The particle filter for SSMs

The particle filter is synonymous for the SMC approach to state estimation in nonlinear SSMs. For an introduction and review see, e.g., (Doucet, Godsill, & Andrieu, 2000; Doucet et al., 2001; Chopin & Papaspiliopoulos, 2020). Neglecting the parameter θ, particle filters approximate the probability distribution of the states by an empirical distribution based on a set of N particles $\{x_t^i\}_{i=1}^N$. The approximation incorporating weights for each particle is

$$p(x_t \mid z_{0:t}) \approx \widehat{p}(x_t \mid z_{0:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t),$$

where the weights $\{w_t^i\}_{i=1}^N$ are normalized to sum to one, and where $\delta_x()$ denotes the Dirac delta distribution centered in x. Particles and weights are propagated and updated according to Bayes' rule using the SSM defined by Eqs. (1) and (2).

The simplest implementation is the bootstrap particle filter, given for completeness in Algo. 1. More complex algorithms have been proposed in the literature in order to overcome the shortcomings of the bootstrap particle filter in practice. The main is the impoverishment of the particle set. This refers to the fact that without resampling most weights degenerate over time, or with resampling (as done in the bootstrap particle filter) only a few particles are retained after the resampling step.

Algorithm 1 Bootstrap particle filter

1: Initialize $\{x_0^i, w_0^i = 1/N\}_{i=1}^N$;
2: **for** $t = 1, \ldots, t_0$ **do**
3:     **for** $i = 1, \ldots, N$ **do** ▷ Propagate
4:         Sample $\widetilde{x}_t^i \sim p(x_t \mid x_{t-1}^i)$;
5:         Compute $\widetilde{w}_t^i = w_{t-1}^i p(z_t \mid \widetilde{x}_t^i)$;
6:     **end for**
7:     Normalize $\widehat{w}_t^i = \widetilde{w}_t^i / \sum_j \widetilde{w}_t^j$, $i = 1, \ldots, N$;
8:     **for** $i = 1, \ldots, N$ **do** ▷ Resample
9:         Sample k from the set $\{1, \ldots, N\}$
            with weights $\{\widehat{w}_t^j\}_{j=1}^N$;
10:         Set $x_t^i = \widetilde{x}_t^k, w_t^i = 1/N$;
11:     **end for**
12: **end for**

In this form, the particle filter assumes the model parameter θ to be known and fixed. Therefore, it cannot be directly applied to prognostics application, as the parameter values are typically specific to each individual device. Missing measurements can be treated on the other hand trivially: If the measurement z_t is unavailable, we marginalize Eq. (2) with respect to z_t and replace $p(z_t \mid \widetilde{x}_t^i)$ by $\int p(z_t' \mid \widetilde{x}_t^i) dz_t' = 1$. In practice, this corresponds to not updating the weights (line 5 of Algorithm 1) and only sampling $\widetilde{x}_t^i$ from the distribution $p(x_t \mid x_{t-1}^i)$. Resampling is not required in this case, but could still be done, even though it might lead to unnecessary impoverishment of the particles.

3. State and parameter estimation with particle filters

Estimating sequentially both state and parameter of a model is a difficult problem. Several methods have been proposed in the literature, see, e.g., (Doucet et al., 2000, 2001). Many of these do not treat the parameter estimation sequentially, and are therefore not further discussed here. In this section, we only describe three of the most common approaches in the prognostics literature, see, e.g., (Si et al., 2011; Jouin et al., 2016).

3.1. Parameter-augmented bootstrap particle filter

The straightforward approach to state and parameter estimation consists in augmenting the state-space x_t with the parameter θ, i.e., defining a new state space $X_t = (x_t, \theta_t)$. The underlying dynamics for x_t is unchanged and is defined by Eq. (3). The parameter θ is assumed to not evolve in time, i.e., it follows the trivial dynamics

$$\theta_t = \theta_{t-1},$$

and $\theta_0 = \theta$. Together with the prior distribution $p(\theta_0) = p(\theta)$ this is equivalent to the solution of the full problem. The augmented state of the resulting SSM can then be estimated using the bootstrap particle filter. The resulting method is given in Algo. 2.

With this method, the parameter θ_t does not evolve over time. Hence, the set of possible values for it is fixed throughout the algorithm, and is equal to the initial samples θ_0^i from the prior distribution $p(\theta)$. Due to resampling, only a few distinct values of θ survive after some time (in the worst case only one). Therefore, this algorithm leads in many cases to a strong overconfidence on the parameter uncertainty, and possibly to a wrong estimate of its value. Despite this shortcoming, this approach has been proposed in (An, Choi, & Kim, 2013), even if only in a tutorial setting.

Algorithm 2 Parameter-augmented bootstrap particle filter

Initialize $\{x_0^i, \theta_0^i, w_0^i = 1/N\}_{i=1}^N$;
for $t = 1, \ldots, t_0$ **do** ▷ Propagate
    for $i = 1, \ldots, N$ **do**
        Set $\widetilde{\theta}_t^i = \theta_{t-1}^i$
        Sample $\widetilde{x}_t^i \sim p(x_t \mid x_{t-1}^i; \widetilde{\theta}_t^i)$;
        Compute $\widetilde{w}_t^i = w_{t-1}^i p(z_t \mid \widetilde{x}_t^i; \widetilde{\theta}_t^i)$;
    end for
    Normalize $\widehat{w}_t^i = \widetilde{w}_t^i / \sum_j \widetilde{w}_t^j$, $i = 1, \ldots, N$;
    for $i = 1, \ldots, N$ **do** ▷ Resample
        Sample k from the set $\{1, \ldots, N\}$
            with weights $\{\widehat{w}_t^j\}_{j=1}^N$;
        Set $x_t^i = \widetilde{x}_t^k$, $\theta_t^i = \widetilde{\theta}_t^k$, $w_t^i = 1/N$;
    end for
end for

3.2. Diffusive bootstrap particle filter

The main limitation of the parameter-augmented bootstrap particle filter is the impossibility to create new parameter values θ_t^i. This can be overcome by increasing their variability over time and in particular by exploring values close to the particles that survive the resampling. Since we only have information regarding the likelihood function or posterior distribution of values of state and parameter represented by some particles, some approximation is needed.

The most popular approach to create variability in the parameter consists in adding a stochastic dynamic term to its time evolution. In almost all practical cases, this dynamics takes the form of a Brownian motion, i.e.,

$$\theta_t \sim \mathcal{N}(\theta_{t-1}, \Sigma_\theta), \tag{5}$$

where Σ_θ is a suitable covariance matrix.

Whereas the motivation for the stochastic evolution of the parameter θ is purely to improve the particle filter method, it is often proposed, that it attempts to capture the mismatch between the model and the real underlying process, even if the parameter is not – in principle – changing in time. Despite this mismatch being potentially a valid point, using a stochastic dynamics for this in prognostics is difficult to justify. Indeed, the variation of the degradation variable is often rather limited and deviations will tend to be rather systematic than random. Another case made is that it allows to capture change-points of the parameter, where the time evolution changes abruptly, e.g., due to a transition to a faulty state. Such transitions are often handled better using dedicated approaches. In addition, the diffusive nature of Eq. (5) leads to past measurements being considered progressively less by the filter, leading to a larger parameter and prediction uncertainty.

On a more practical side, introducing the covariance matrix Σ_θ adds hyperparameters to the algorithm that are often difficult to tune. Unfortunately, the performance of the algorithm relies strongly on a good choice of them. If the covariance has too small elements the parameter is essentially static and the method has the same issues as in Sec. 3.1. Conversely, if Σ_θ has too large elements, the dynamics introduces overdispersion to the parameter. This second case is particularly concerning in case of missing measurements and especially in the prediction phase. Indeed, without measurements, which are the driving force constraining the parameter, the diffusive dynamics leads to a strong and purely artificial increase in uncertainty. The calculation of the RUL is most susceptible to this, as a prediction over a long time horizon is made. A hybrid approach is often employed to overcome this difficulty: The parameter is evolved using the stochastic model for the estimation phase, but is then frozen for the prediction phase. This inconsistency is listed as one of the open questions in (Jouin et al., 2016).

We propose here the introduction of an improved parameter evolution by using a time dependent covariance matrix Σ_θ in order to mitigate this issue. The time dependence is defined in the following way: the parameter is only updated when measurements are done, otherwise it remains unchanged. This corresponds formally to setting $\Sigma_\theta = 0$ for time steps without measurements. With this practical approach, we do not incur an artificial but unneeded overdispersion and still retain the better exploration of the parameter space with respect to the method of Sec. 3.1. We also remove the inconsistency in calculating the evolution in the parameter estimation and prediction phase. Despite these improvements, the diffusive bootstrap filter still strongly relies on choosing appropriately the covariance matrix to avoid either the impoverishment of the particles, or the loss of information carried by past measurements.

The diffusive bootstrap particle filter algorithm including the improvement for missing measurements is shown in Algo. 3.

Algorithm 3 Diffusive augmented bootstrap particle filter including treating of missing measurements

Initialize $\{x_0^i, \theta_0^i, w_0^i = 1/N\}_{i=1}^N$;
for $t = 1, \ldots, t_0$ **do** ▷ Propagate
  for $i = 1, \ldots, N$ **do**
    if z_t available **then**
      Sample $\widetilde{\theta}_t^i \sim p(\theta_t | \theta_{t-1}^i, \Sigma_\theta)$;
    else
      Set $\widetilde{\theta}_t^i = \theta_{t-1}^i$;
    end if
    Sample $\widetilde{x}_t^i \sim p(x_t \mid x_{t-1}^i; \widetilde{\theta}_t^i)$;
    Compute $\widetilde{w}_t^i = w_{t-1}^i p(z_t \mid \widetilde{x}_t^i; \widetilde{\theta}_t^i)$;
  end for
  Normalize $\widehat{w}_t^i = \widetilde{w}_t^i / \sum_j \widetilde{w}_t^j$, $i = 1, \ldots, N$;
  for $i = 1, \ldots, N$ **do** ▷ Resample
    Sample k from the set $\{1, \ldots, N\}$ with weights $\{\widehat{w}_t^j\}_{j=1}^N$;
    Set $x_t^i = \widetilde{x}_t^k$, $\theta_t^i = \widetilde{\theta}_t^k$, $w_t^i = 1/N$;
  end for
end for

3.3. Liu–West filter

A popular approach trying to avoid the artificial overdispersion of the parameter due to the stochastic evolution of the parameter has been proposed in (Liu & West, 2001), referred to here as "Liu–West filter". Their approach has been widely used due to two advantages: it is independent of the specific model and it is easy to implement. Examples of its use in prognostics application are e.g.(Hu, Baraldi, Di Maio, & Zio, 2015; Peng et al., 2018).

The Liu–West filter, similarly to the diffusive particle filter of Sec. 3.2, evolves the parameter in time with a stochastic process. Unlike it, the process is tuned such that the mean and covariance of the marginal parameter distribution stays invariant during the parameter update process. The overdispersion of the parameter estimation is therefore kept under control, avoiding the main drawback of the diffusive bootstrap filter.

At each time t, the parameter value for the i^{th} particle is sampled from the modified stochastic process

$$\theta_t^i \sim N(m_t^i, \Sigma_t),$$

with suitable values for m_t^i and Σ_t. To calculate these, the (weighted) mean m_t and covariance $\Sigma_{m,t}$ of the marginal distribution over all θ_{t-1}^i are determined. These are then used to get

$$m_t^i = a\theta_{t-1}^i + (1-a)m_t \tag{6}$$

and

$$\Sigma_t = (1-a^2)\Sigma_{m,t}, \tag{7}$$

In Eqs. (6) and (7), the scalar tuning parameter $a \in [0,1]$ is used in both m_t^i and Σ_t such that the marginal distribution of the newly sampled $\{\theta_t^i\}_{i=1}^N$ have the same mean m_t and covariance Σ_t as before. In this way the overdispersion from the sampling of θ_t^i is kept at a minimum.

Two limiting cases can be seen: If the coefficient $a \to 1$ the parameter values of the particles do not move over a time step, the Liu–West filter approaches the static particle filter. Conversely if $a \to 0$ all particles have parameters drawn from a common normal distribution, independent of the individual parameter values of the particles at the previous time step. The parameter a is often chosen very close to $a \to 1$ (e.g. 0.995) or even adapted over time to cope with the improved knowledge of the parameter. The method was inspired by the analogy of the marginal parameter distribution in the diffusive update step with a kernel density estimation or a Gaussian mixture model centered around the m_t^i.

A possible implementation of the Liu–West filter is given in Algo. 4; the main difference is the modified calculation of the update of the parameter value of the particles. We give here only the simplest implementation, whereas in (Liu & West, 2001) some additional importance sampling steps are used in addition.

Algorithm 4 The Liu–West particle filter

Initialize $\{x_0^i, \theta_0^i, w_0^i = 1/N\}_{i=1}^N$;
for $t = 1, \ldots, T$ **do** ▷ Propagate
  for $i = 1, \ldots, N$ **do**
    Determine mean m_t and variance Σ_t of the marginal of the θ_{t-1}^i;
    Determine $m_t^i = a\theta_{t-1}^i + (1-a)m_t$;
    Sample $\widetilde{\theta}_t^i \sim \mathcal{N}(m_t^i, (1-a^2)\Sigma_t)$;
    Sample $\widetilde{x}_t^i \sim p(x_t | x_{t-1}^i, \widetilde{\theta}_t^i)$;
    Evaluate the corresponding weights $w_t^i \propto w_{t-1}^i p(z_t | x_t^i, \widetilde{\theta}_t^i)$;
  end for
  for $i = 1, \ldots, N$ **do** ▷ Resample
    Sample k from the set $\{1, \ldots, N\}$ with weights $\{\widehat{w}_t^j\}_{j=1}^N$;
    Set $x_t^i = \widetilde{x}_t^k$, $\theta_t^i = \widetilde{\theta}_t^k$, $w_t^i = 1/N$;
  end for
end for

The Liu–West filter often works in practice, but requires tuning of the hyperparameter a, which might be difficult to set to a reasonable value in a real application. A wrong value of a can lead, as before to an incorrect prediction.

Dealing with missing measurements can be done in two possible ways: The first one replaces $p(z_t \mid \cdots)$ by one and evolves the parameter θ_t^i in the same way as for time steps with measurements. The second one follows the proposal above and keeps the parameter fixed for that time step. This corresponds formally to choosing $a = 1$ for them.

4. Storvik filter

In (Storvik, 2002; Johannes & Polson, 2006; Erol, Li, Ramsundar, & Russell, 2013) the authors propose a class of particle filter approaches that are exact with respect to the parameter distribution. Even though they slightly differ in their specific implementation, they are based on the same basic concept and we refer to them together as "Storvik filter".

The main problem with parameter estimation in SSMs, and therefore also in SMC, is the increasing number of measurement data and hidden states, which makes evolving the parameter distribution progressively harder over time. The Storvik filter assumes the existence of a finite-dimensional sufficient statistic $s(x,z)$ for the distribution of the parameter given the states $x_{0:t}$ and the measurements $z_{0:t}$. Denoting $s_t = s(x_{0:t}, z_{0:t})$, sufficiency means that

$$p(\theta \mid x_{0:t}, z_{0:t}) = p(\theta \mid s_t).$$

The value of s_t carries all the relevant information contained in the history of $x_{0:t}$ and $z_{0:t}$. In addition, the Storvik filter requires a recursive rule

$$s_t = S(s_{t-1}, x_t, z_t).$$

to update the sufficient statistic with each new state and measurement.

The existence of a sufficient statistic with a finite and fixed dimension independent of t is not guaranteed. The Fisher–Pitman–Koopman–Darmois theorem states that such a finite sufficient statistic s_t exists if and only if the distribution of θ belongs to the exponential family, see e.g. (Barankin & Maitra, 1963). This is often the case and thus guarantees a wide applicability of the Storvik filter. Especially many models use normal distributed process noise terms together with a linear dependency of the parameter, which can be addressed with this approach as discussed in (Erol et al., 2013). Extensions to distributions that are not members of the exponential family can be found in (Johannes & Polson, 2006), where the authors consider mixtures of exponential families for this case. Finally, in (Joyce & Marjoram, 2008) the authors discuss the determination of approximately sufficient statistics from data, if exact sufficient statistics are not available.

The Storvik filter is given in Algo. 5. It shares a number of similarities with the already discussed filters in that the parameter is evolved as well in each time step. The particles contain in addition to x_t^i and θ_t^i also s_t^i, which are used to sample new values of $\theta_t^i \sim p(\theta_t^i \mid s_t^i)$.

5. Storvik Filter with missing measurements

Incorporating missing measurements into the Storvik filter is not straightforward, as was in the other cases, due to the re-

Algorithm 5 The Storvik particle filter

```
1:  Initialize $\{x_0^i\}_{i=1}^N$;
2:  Compute $\{s_0^i = s(x_0^i, z_0)\}_{i=1}^N$;
3:  for $t = 1, \dots, T$ do                                   ▷ Propagate
4:      for $i = 1, \dots, N$ do
5:          Sample $\theta_t^i \sim p(\theta_t \mid s_{t-1}^i)$;
6:          Sample $\widetilde{x}_t^i \sim p(x_t \mid x_{t-1}, \theta_t^i)$;
7:          Evaluate $\widetilde{w}_t^i = w_{t-1}^i p(z_t \mid \widetilde{x}_t^i, \theta_t^i)$;
8:          Compute $\widetilde{s}_t^i = S(s_{t-1}^i, \widetilde{x}_t^i, z_t)$;
9:      end for
10:     Normalize $\widehat{w}_t^i = \widetilde{w}_t^i / \sum_j \widetilde{w}_t^j$, $i = 1, \dots, N$;
11:     for $i = 1, \dots, N$ do                                ▷ Resample
12:         Sample $k$ from the set $\{1, \dots, N\}$
                with weights $\{\widehat{w}_t^j\}_{j=1}^N$;
13:         Set $x_t^i = \widetilde{x}_t^k$, $s_t^i = \widetilde{s}_t^k$, $w_t^i = 1/N$;
14:     end for
15: end for
```

quired update of the sufficient statistic in each step. We have identified two possible approaches:

1. Resampling the parameter at each step, despite missing measurements. We denote this choice by "U" as in "Updating".
2. Freezing the parameter to the value at the last observed time. We denote this choice by "F" as in "Frozen".

The two approaches can be shown to result from splitting the joint posterior distribution for the evolution of the state from $t+1$ to $t+k$ without measurements into

$$p(x_{t+1:t+k}, \theta) = p(\theta | s_{t+k}) p(x_{t+1:t+k}),$$

showing that the sufficient statistic needs to be updated using all $x_{t+1:t+k}$ to get the correct distribution of θ at the end. The distribution $p(x_{t+1:t+k})$ on the other hand can be decomposed in two different ways, either as

$$\begin{aligned} p(x_{t+1:t+k}) \quad = \quad & \int p(x_{t+k}|\theta) p(x_{t+k-1}|\theta) \dots \\ & p(x_{t+1}|\theta)\ p(\theta|x_{1:t}, z_{1:t}) d\theta \end{aligned}$$

which corresponds to sampling one θ at the last time step with a measurement and sampling all new values of x with it, which is the "F" approach. Alternatively one can write

$$\begin{aligned} p(x_{t+1:t+k}) = & \int p(x_{t+k}|x_{t+k-1}, \theta) p(\theta|s_{t+k-1}) d\theta \times \\ & p(x_{t+1:t+k-1}) \end{aligned}$$

which corresponds to updating the sufficient statistic after each step and sampling a new θ from it. This is the "U" approach. This shows that both approaches are identical in principle, but could still be more or less efficient in applications.

A possible implementation of both approaches is given in Algos. 6 and 7 for the "U" and "F" approach replacing lines 3 to 9 in Algo. 5, respectively. Please note the similarity of the two approaches to the one previously discussed.

Algorithm 6 Storvik particle filter "U" for predictions

```
for $i = 1, \dots, N$ do
    Sample $\theta_t^i \sim p(\theta_t \mid s_{t-1}^i)$;
    Sample $x_t^i \sim p(x_t \mid x_{t-1}, \theta_t^i)$
    Set $w_t^i = w_{t-1}^i$
    Compute $s_t^i = S(s_{t-1}^i, x_t^i)$
end for
```

Algorithm 7 Storvik particle filter "F" for predictions

```
Set $\theta_t^i = \theta_{t-1}^i$ for $i = 1, \dots, N$
for $i = 1, \dots, N$ do
    Sample $x_t^i \sim p(x_t \mid x_{t-1}, \theta_t^i)$
    Set $w_t^i = w_{t-1}^i$
    Compute $s_t^i = S(s_{t-1}^i, x_t^i)$
end for
```

6. Simulation Study

We compare the performances of the four algorithms (parameter-augmented bootstrap particle filter, diffusive bootstrap particle filter, Liu–West filter, and Storvik filter) when applied to two simple models mimicking typical degradation dynamics. In addition we use a MCMC implementation in order to get the exact posterior distribution for all cases, using the JAGS probabilistic programming language (Plummer, 2003).

6.1. The linear model

The simplest model is the one of a Brownian motion with drift for x_t

$$x_{t+1} \sim \mathcal{N}(x_t + \alpha, \sigma_x) \tag{8}$$

together with a normal distributed measurement error

$$z_t \sim \mathcal{N}(x_t, \sigma_z) \tag{9}$$

This model is also referred to in the literature as the Whitmore model (Whitmore, 1995). For this study, we assume that only α is unknown and that σ_x and σ_z are known, so that only α and x_t need to be determined. This model falls into the class of having normal distributed process noise and being linear in the parameter. Therefore the sufficient statistic is known to be the mean and standard deviation of the distribution of the parameter α.

The prior distributions are assumed to be given as

$$p(\alpha) = \mathcal{N}(\alpha_0, \sigma_{\alpha_0}) \tag{10}$$

and

$$p(x_0) = \mathcal{N}(z_0, \sigma_{x_0}) \tag{11}$$

centered around the first measurement z_0.

To simulate the data set we have used $x_0 = 1000$, $\alpha = -8.65$, $\sigma_x = 1$, and $\sigma_z = 5$. The prior distribution parameters are $\alpha_0 = \alpha = -8.65$, $\sigma_{\alpha_0} = \sqrt{5}$ and $\sigma_{x_0} = 10$.

6.2. The stretched exponential or Weibull model

For many devices, e.g. for batteries or capacitors, the degradation process accelerates over time. The Weibull function, also known as stretched exponential, is often employed to capture this. We use a model of the form

$$x_{t+1} \sim \mathcal{N}\left(x_t - 3\alpha\left(-\ln\left(\frac{x_t}{X}\right)\right)^{1-1/3} x_t, \sigma_x\right),$$

which is chosen to follow approximately a Weibull function with shape parameter 3, but in the form of a time-independent stochastic process. The measurement model reads as before

$$z_t \sim \mathcal{N}(x_t, \sigma_z),$$

We assume the same prior distribution for α and x_0 as in Eqs. (10) and (11).

As in the linear model, the process noise is modeled as normal distributed and the dependency on the parameter is linear, even if the evolution of x_t is not. This makes the sufficient statistic to be as before the mean and standard deviation of the distribution for α.

For the simulation we have used initial condition $x_0 = 995$, maximum value $X = 1000$, $\sigma_x = 1$, $\sigma_z = 5$, and $\alpha = 1/80$. For the prior we use $\alpha_0 = \alpha = 1/80$ and $\sigma_{\alpha_0} = 6 \cdot 10^{-3}$ and where $\sigma_{x_0} = 1$ to avoid sampling impossible values $x_0 \geq X$. Parameters were chosen, such that the two models are comparable in terms of the degradation path.

For all tests we simulate 100 time steps with the measurements thinned, such that only every 5th time step was recorded to verify how efficiently the methods can treat missing measurement. We also stopped the estimation phase at either time horizon $t_0 = 30$ or 75 and continued with the prediction part only. The data set used, as well as the mean degradation curve for the two models are shown in Fig. 1.

6.3. Results

Figure 2 demonstrates the issues that can affect the static and the diffusive particle filters and the Liu–West filter. Measurements are available until $t_0 = 75$, after which only predictions were done. Note that we selected hyperparameters in order to exaggerate the issues. A more fine-tuned approach would lead to better agreement with the MCMC results.

With respect to the MCMC based reference result, given in Figure 2(d) we observe:

- Figures 2(a) and (b) demonstrate the underestimation and overestimation of the parameter uncertainty when using the static parameter-augmented particle filter and the diffusive particle filter, respectively.
- Figure 2(c) demonstrates that also the Liu–West filter can yield overdispersed results, even if overall less severe than for the diffusive particle filter case.

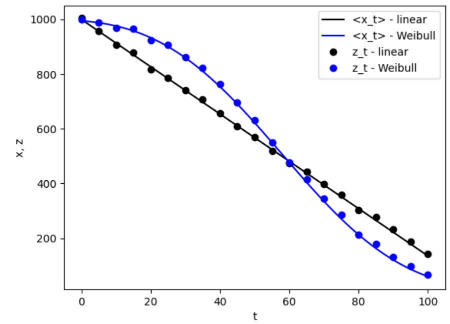
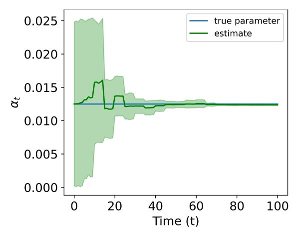
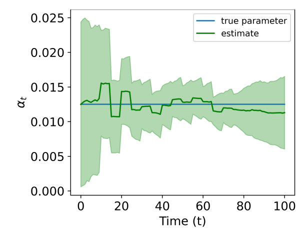


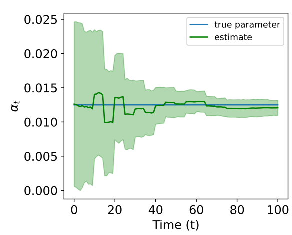
Figure 1. The mean value for x_t and the thinned measurement data z_t for the two models is shown.



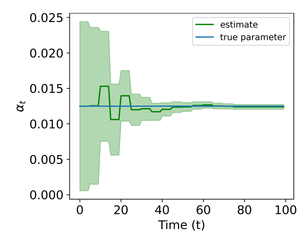
(a) Parameter augmented filter



(b) Diffusive filter



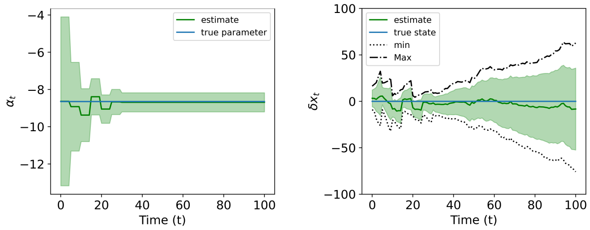
(c) Liu–West filter



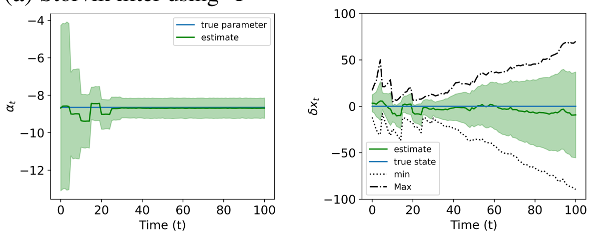
(d) MCMC

Figure 2. Parameter estimation for the Weibull model for the three different particle filter: (a) static bootstrap, (b) diffusive, and (c) Liu–West filter. (d) gives the reference results using the MCMC approach. Mean values and uncertainty in terms of two standard deviations are given.

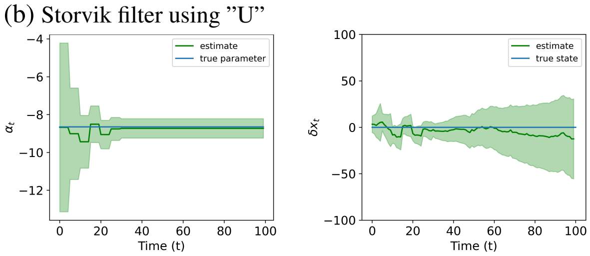
In Fig. 3 we give the results for the two implementations (”F” and ”U”) of the Storvik filter for missing measurements for the linear model, in Fig. 4 for the Weibull model. In order to focus on the performance of the two algorithm in the prediction phase, we set the time horizon to $t_0 = 30$. We observe that both implementations of the Storvik filter for missing measurements lead to results that are consistent with the MCMC result, despite (or due to) the absence of tuning hyperparameters. We also observe that the two implementations are practically indistinguishable, with the exception of a slightly more unstable behavior of the U implementation, visible in the initial time period of the Weibull model in



(a) Storvik filter using "F"

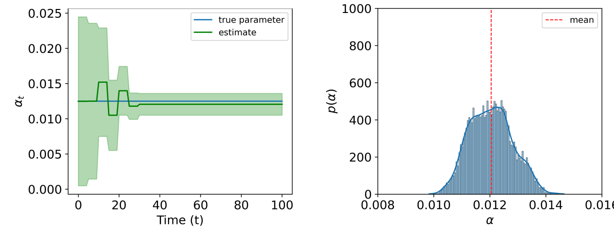


(b) Storvik filter using "U"

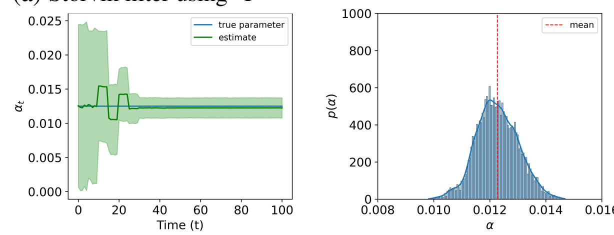


(c) MCMC

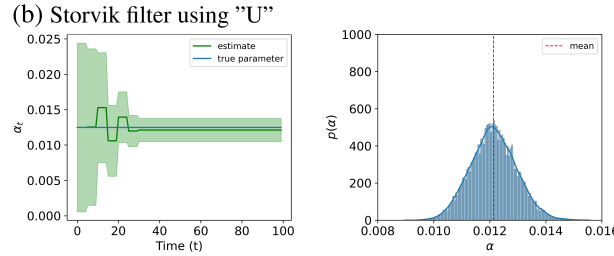
Figure 3. Parameter and state estimation of the linear model with the Storvik filter: Results using the two different treatments in case of missing data are compared using (a) the Frozen and (b) the Update approach. Results are compared with the reference result using the MCMC approach. On the left the evolution of the parameter estimation, on the right the difference between predicted and the true state is given. In both cases the mean and the uncertainty in terms of two standard deviation is shown.



(a) Storvik filter using "F"



(b) Storvik filter using "U"



(c) MCMC

Figure 4. Parameter estimation of the Weibull model with the Storvik filter: Results using the two different treatments in case of missing data are compared using (a) the Frozen and (b) the Update approach. Results are compared with the reference result using the MCMC approach. On the left the evolution of the parameter estimation is given showing the mean and the uncertainty in terms of two standard deviation, whereas on the right we give the full distribution for the final time step.

Fig. 4(b).

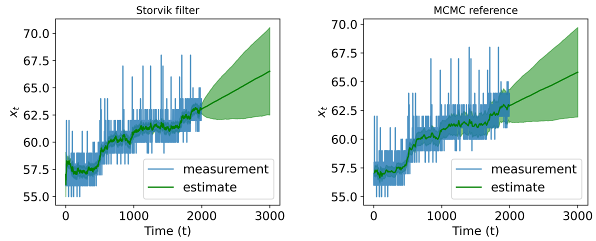
7. Application to real data with breaker opening times

In this section, we test the application of the Storvik filter against real data from an application with circuit breakers. In this case model misspecification is present and could undermine the applicability of the approach.

Circuit breakers are protection devices to interrupt short circuit currents occurring in an electric network. They are operated by mechanical mechanisms whose malfunction is one of dominant failure modes for them. The time required to open or close the contacts is the commonly monitored property. For instance, a reduction in a spring force or an increase in friction leads to an increase of this time. Hence, tracking it as a function of the number of operations enables to predict the end-of-life of these devices.

The evolution of the time x_t of the mechanical opening/closing operation is in general stochastic. Please note that t in this case typically refers to the number of operations performed instead of the time in operation. We describe it by the linear model as given in Eq. (8). The main issue with the data is that the measurement error is not following a normal distribution as assumed in Eq. (9). In fact, because of the signal processing performed during acquisition, the data is strongly quantized, as can be seen in Fig. 5(a). This was already discussed and analysed in (Hencken, 2021), which concluded that assuming normal distributed error gives reasonable results in a full analysis.

As circuit breaker often perform a larger number of operations, the use of a sequential approach will be an advantage in practice. We therefore explore here whether a particle based approach based on the Storvik filter is suitable. The model features three unknown parameters: the drift α and the two standard deviation for the process σ_x and the measurement σ_z, which leads to a more complex sufficient statistic. Following the usual normal-inverse-gamma model, the sufficient statistic consists of six variables, which are mean and standard deviation of the normal distribution of the drift α and



(a) Storvik filter (b) MCMC

Figure 5. Results of the application of the Storvik filter to the estimation of the degradation process of opening times x_t of a circuit breaker as a function of the number of oerations t. (a) shows the original data together with the state estimations and the prediction of the future evolution for the Storvik filter using the "U" method, (b) the same but for the MCMC reference approach.

the location and scale parameter of the inverse Gamma distribution for σ_x and similar for σ_z. For more details, we refer to (Storvik, 2002), where the sufficient statistic, as well as their update rules, are given.

As all measurements are available, we focus on a comparison of the basic Storvik filter without missing measurements with the exact result as given by the MCMC approach. The results of the state estimation and the expected future evolution of the two models are shown in Fig. 5. The Storvik filter is able to estimate the states quite similar to the ones found in the reference approach and in addition is able to capture the future evolution. Some slight deviations are visible, especially of the MCMC results showing a slightly larger uncertainty at the end of the measurements. But this demonstrates in a first step the possible application of the Storvik filter in real applications.

8. Conclusions and Outlook

Model-based prognostics requires joint state and parameter estimation. A sequential approach is most suitable to avoid increase in computational complexity over time. Several approaches involving particle filters and their potential issues have been discussed. We have focused also on the need of a robust treatment of time steps with missing measurements either due to irregular data acquisition or for the predictions needed for the RUL calculation. We have explored the use of the Storvik filter for prognostics application as an exact parameter estimation approach. We have shown that it can be naturally extended to incorporate missing measurements in two ways, which are similar to the ones discussed for the other particle filter approaches. Its main limitation is that it is restricted to problems allowing for the existence of a sufficient statistic. Simulations using two simple models showed the robustness and reliability of the Storvik filter, whereas we demonstrated as well, that other approaches can lead to erroneous results. We have also applied it to one real world examples, in order to test its applicability in a case, where the assumed model is not valid.

Prognostics using particle filters is an active area of research and development of real applications. The promising results with the Storvik filter should be further explored and its applicability to more complex problems, including higher dimensional state space and parameter vectors, but also to models beyond the restricted class studied here, should be explored. Finding suitable sufficient statistics in these more general models, even outside the exponential family, is another line of research to be undertaken.

References

An, D., Choi, J.-H., & Kim, N. H. (2013, July). Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. *Reliability Engineering & System Safety*, *115*, 161–169. doi: 10.1016/j.ress.2013.02.019

Barankin, E. W., & Maitra, A. P. (1963). Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics. *Sankhyā: The Indian Journal of Statistics, Series A*, 217–244.

Chopin, N., & Papaspiliopoulos, O. (2020). *An Introduction to Sequential Monte Carlo*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-47845-2

Doucet, A., de Freitas, N., & Gordon, N. (2001). An introduction to Sequential Monte Carlo methods. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential monte carlo methods in practice* (pp. 3–14). New York, NY: Springer New York.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On Sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*, 197 – 208.

Erol, Y., Li, L., Ramsundar, B., & Russell, S. J. (2013). *The extended parameter filter.*

Galar, D., Goebel, K., Sandborn, P., & Kumar, U. (2021). *Prognostics and remaining useful life (rul) estimation: Predicting with confidence*. CRC Press.

Goebel, K., Daigle, M., Saxena, A., Sankararaman, S., Roychoudhury, I., & Celaya, J. (2017). *Prognostics: The science of making predictions*. Createspace Independent Pub.

Hencken, K. (2021). Parameter estimation of a Wiener process of mechanical degradation through censored measurement of timings. *Proceedings of the 31st European Safety and Reliability Conference*, 317-324.

Hu, Y., Baraldi, P., Di Maio, F., & Zio, E. (2015). A particle filtering and kernel smoothing-based approach for new design component prognostics. *Reliability Engineering & System Safety*, *134*, 19-31. doi: https://doi.org/10.1016/j.ress.2014.10.003

Johannes, M. S., & Polson, N. G. (2006). Exact particle filtering and parameter learning..

Jouin, M., Gouriveau, R., Hissel, D., Péra, M.-C., & Zerhouni, N. (2016). Particle filter-based prognostics: Review, discussion and perspectives. *Mechanical Systems and Signal Processing*, *72*, 2–31.

Joyce, P., & Marjoram, P. (2008). Approximately sufficient statistics and bayesian computation. *Statistical applications in genetics and molecular biology*, *7*(1).

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential monte carlo methods in practice* (pp. 197–223). New York, NY: Springer New York.

Peng, W., Ye, Z.-S., & Chen, N. (2018). Joint online rul prediction for multivariate deteriorating systems. *IEEE Transactions on Industrial Informatics*, *15*(5), 2870–2878.

Plummer, M. (2003, 04). JAGS: a program for analysis of bayesian graphical models using gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011, August). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research*, *213*(1), 1–14. doi: 10.1016/j.ejor.2010.11.018

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, *50*(2), 281-289. doi: 10.1109/78.978383

Whitmore, G. A. (1995). Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime data analysis*, *1*(3), 307–319.

BIOGRAPHIES

Kai Hencken is corporate research fellow at ABB corporate research. He obtained a PhD in Theoretical Physics from the University of Basel in 1994. He was a post-doc at the University of Washington from 1995 to 1997 and at the University of Basel from 1997 to 2005, where he received his habilitation in 2000 and is a lecturer since. In 2005 he joined the theoretical Physics group at ABB corporate research. His research interests include the combination of physical modeling with statistical methods to solve problems related to industrial devices, as well as developing diagnostics and prognostics approaches.

Arthur Serres studied mathematics in EPFL, Imperial College London, and ETHZ. He is currently enrolled in the final year of a Master program at ETH, where he focuses on graph neural networks for transaction monitoring. He worked as a research intern in ABB corporate research and he is now a quantitative researcher. His interests include theoretical machine learning and probability theory.

Giacomo Garegnani obtained a PhD in Mathematics from EPFL in 2021, with a thesis on inverse problems involving partial and stochastic differential equations, and on probabilistic numerical methods. He is now a scientist at ABB corporate research. His research interests include uncertainty quantification of numerical solvers, identifiability of differential models, and methods of statistical inference for condition monitoring.