

Filtering and Smoothing of Hidden Monotonic Trends and Application to Fouling Detection[★]

Ranjith Chiplunkar^{*}, Biao Huang^{**}

^{*} *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 1H9*

^{**} *(correspondence e-mail: bhuang@ualberta.ca).*

Abstract: In this paper, we present a filtering and smoothing scheme for process variables characterized by a hidden monotonic trend. The proposed method models the transition probability distribution of the hidden monotonic trend as a closed skew normal distribution and the observed data is assumed to have a Gaussian noise added onto this monotonic trend. The objective is to extract the monotonic trend given the noisy observations. The proposed method has advantages in process monitoring applications involving processes driven by a monotonic trend where vanilla Kalman filter may not be the apt option. The proposed method has been verified on an industrial dataset of a hot lime softener process to detect the fouling buildup.

Keywords: Monotonicity, Closed-skew normal distribution, State estimation, Fouling monitoring, Bayesian estimation

1. INTRODUCTION

Systems driven by a hidden monotonic trend (HMT) are a common occurrence in industrial processes. Quantities such as catalyst activity, equipment health factors (or equipment damage), certain types of fouling buildup, etc are monotonic. This monotonic trend may not be apparent in the datasets of such processes which makes monitoring such processes challenging. Hence one needs to extract the HMT from the observed data.

In regression approaches, such problems are formulated as optimization problems with monotonicity constraints. But for online applications, a linear dynamic model with a hidden monotonic model is more desirable (Gorinevsky (2004)). In these approaches, the state evolution is assumed to happen according to a monotonic random walk model and the output is assumed to be corrupted with Gaussian noise. Gorinevsky (2004) assumed an exponential distribution for the process noise as it has positive support and solved it as a quadratic programming problem. The same model was later solved in a moving horizon estimation framework in Samar et al. (2004) and Gorinevsky (2008). A Gamma process where the process noise follows a gamma distribution could be used for monotonic trending (van Noordwijk (2009)). Such hidden gamma process models are used to predict health factors in manufacturing and the filtering of such signals is implemented using a particle filter (Susto et al. (2018), Schirru et al. (2010)).

An important aspect of choosing a distribution for filtering and smoothing problems apart from the physical relevance is the ease of implementing the prediction and update steps. That is, if the distribution remains the same

or is tractable in the prediction and update (Bayesian inference) steps which are crucial to obtain recursive estimates. Gaussian distribution has this nice property of "closedness" while adding two Gaussian variables (prediction step) and while doing posterior inference (update step). But a Gaussian random walk model does not give monotonic signals. In this paper, we explore the usage of closed skew normal distribution (CSN) to model the HMT and propose a smoothing scheme for a CSN process.

González-Farías et al. (2004) proposed the CSN distribution which is a distribution with five parameters and as the name suggests is a skewed distribution. Similar to the Gaussian distribution, CSN has properties such as the linear transformation of a CSN is a CSN, and the posterior distribution in Bayesian inference is a CSN if the likelihood and prior are CSN. These properties allow the derivation of recursive filtering schemes for CSN processes. Naveau et al. (2005) derived a skewed filtering method when the observation noise is Gaussian and the states followed CSN. Karimi et al. (2010) derived Bayesian inversion equations for the case where likelihood and prior are CSN. Rezaie and Eidsvik (2014) discusses the implementation of CSN filtering for linear dynamic systems and propose ensemble KF with CSN distribution for nonlinear dynamic systems. Rezaie and Eidsvik (2016) propose a skewed unscented Kalman filter for nonlinear systems. Arellano-Valle et al. (2019) proposed a filtering and smoothing scheme for a linear model with Gaussian process noise and CSN noise for observations. The CSN filtering schemes have been applied to cases such as tracking velocity and position of a body entering the atmosphere (Rezaie and Eidsvik (2016)), saturation estimation in a petroleum reservoir using seismic data (Rezaie and Eidsvik (2014)), etc. Most of the applications are found in statistics literature and

[★] The authors would like to thank Natural Sciences and Engineering Research Council (NSERC), Canada for supporting this work.

practical implementations in process systems engineering are scarce.

In this paper, we present a method to extract the HMT from the observed noisy data. The HMT evolves as a CSN distribution and the output noise is assumed to be Gaussian. The key issue in practical implementations of filtering equations for a CSN process is the problem of blowing up of the dimensionality of the skewness term. We address this issue by approximating the high-dimensional skewness term as a low-dimension one at each time of the forward pass. For the backward smoothing pass, as far as we know, analytical expressions for the case where the hidden state evolves as a CSN process have not been derived. As will be shown later, such a derivation is not possible with low-dimension approximation of a CSN and we propose an importance sampling-based scheme for smoothing. Although approaches such as constrained state estimation based methods, or density truncation based constrained estimation (Simon and Simon (2010)) may be used to achieve our objective, we work with a CSN process. The CSN distribution is parametrized by five parameters giving a more flexible representation of the involved distributions. Also, the density truncation may not give good results for the monotonic trend process case if the observation model noise variance is high.

We apply the proposed method on an industrial hot lime softer(HLS) process to detect and monitor the fouling buildup. The rest of the paper is organized as follows. Section 2 discusses preliminaries regarding the CSN distribution and state estimation. The proposed model and the state estimation method is presented in section 3. Section 4 presents the case studies and the section 5 discusses the conclusions.

2. PRELIMINARIES

In this section, we revisit the CSN distribution and the overall procedure of state estimation.

2.1 Closed skew normal distribution

The CSN distribution can be described in terms of a Gaussian variable and an associated Gaussian variable with truncated observations (González-Farías et al. (2004), Iversen (2010)). Let us consider an n dimensional variable s and q dimensional v which jointly follow a Gaussian distribution.

$$\begin{bmatrix} s \\ v \end{bmatrix} \sim \mathcal{N}_{n+q} \left(\begin{bmatrix} \mu_s \\ \mu_v \end{bmatrix}, \begin{bmatrix} \Sigma_s & \Gamma_{sv} \\ \Gamma_{vs} & \Sigma_v \end{bmatrix} \right). \quad (1)$$

Let x be defined as $[s|v \geq 0]$. Hence we have

$$p(x) = p(s|v \geq 0) = \frac{p(v \geq 0|s) p(s)}{p(v \geq 0)} = (1 - \Phi_q(0; \mu_v, \Sigma_v))^{-1} (1 - \Phi_q(0; \mu_{v|s}, \Sigma_{v|s})) \mathcal{N}_n(x; \mu_s, \Sigma_s), \quad (2)$$

with

$$\mu_{v|s} = \mu_v + \Gamma_{vs} \Sigma_s^{-1} (s - \mu_s); \Sigma_{v|s} = \Sigma_v - \Gamma_{vs} \Sigma_s^{-1} \Gamma_{sv}. \quad (3)$$

Here, $\Phi_q(\cdot; \mu, \Sigma)$ represents a q dimensional cumulative distribution function (cdf) with mean μ and covariance

Σ . The distribution in (2) has a Gaussian term multiplied by a Gaussian cdf. The cdf term causes the skewness in the distribution. Equation (2) can be generalized to define the probability distribution function (pdf) of CSN as follows

$$p(x) = CSN_{n,q}(x; \mu, \Sigma, \Gamma, \nu, \Delta) = \Phi_q(0; \nu, \Delta + \Gamma \Sigma \Gamma')^{-1} \Phi_q(\Gamma(x - \mu); \nu, \Delta) \mathcal{N}_n(x; \mu, \Sigma).$$

Here μ and $\Sigma(n \times n)$ are the mean and covariance of the Gaussian part. It is to be noted that these are not the true mean and covariance of the CSN distribution. $\Gamma(q \times n)$ is called the skewness parameter. ν and $\Delta(q \times q)$ are the mean and covariance of the cdf part. It is to be noted that when $\Gamma = 0$, CSN reduces to a Gaussian distribution.

2.2 State estimation

A hidden Markov process is characterized by two probability distributions: the state transition probability distribution $p(x_t|x_{t-1})$ and the observation probability distribution $p(y_t|x_t)$. The estimation of the states x given the observations y is performed recursively in two passes. The forward called the filtering pass includes prediction and update steps, and the backward pass is called smoothing (Särkkä (2013)). The pdf calculations in these three steps are summarized as follows.

Prediction:

$$p(x_t|Y_{t-1}) = \int p(x_t|x_{t-1}) p(x_{t-1}|Y_{t-1}) dx_{t-1} \quad (4)$$

Update:

$$p(x_t|Y_t) = \frac{p(y_t|x_t) p(x_t|Y_{t-1})}{\int p(y_t|x_t) p(x_t|Y_{t-1}) dx_t} \quad (5)$$

Smoothing:

$$p(x_t|Y_T) = \int \frac{p(x_{t+1}|x_t) p(x_t|Y_t)}{\int p(x_{t+1}|x_t) p(x_t|Y_t) dx_t} p(x_{t+1}|Y_T) dx_{t+1} \quad (6)$$

Here $Y_t = [y_1, y_2, \dots, y_t]$ is a collection of measurements till time t and $T > t$ is a future time instant for t .

3. MONOTONIC HIDDEN TREND ESTIMATION

The proposed approach models the state transition probability of the HMT as a CSN.

3.1 Model

Trend extraction models usually contain two equations: one to model the latent variable (8) and the other for the output (7). We proposed to add a third one in the form of (9) as shown below.

$$y_t = x_t + v_t \quad (7)$$

$$x_t = x_{t-1} + e_t \quad (8)$$

$$r_t = w_t \quad (9)$$

Here $v_t \sim \mathcal{N}(0, \sigma_v^2)$ is white noise and is uncorrelated with e_t and w_t . Noises e_t and w_t are Gaussian but are heavily correlated with a correlation coefficient ρ close to 1. We assume same variance of σ_e^2 for both e_t and w_t . With these assumptions, we have

$$\begin{bmatrix} x_t \\ r_t \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} x_{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & \rho \sigma_e^2 \\ \rho \sigma_e^2 & \sigma_e^2 \end{bmatrix} \right). \quad (10)$$

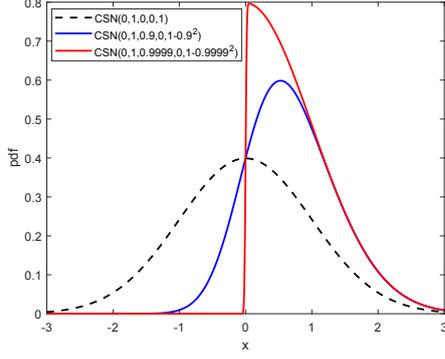


Fig. 1. The pdf of CSN at different values of ρ .

Here, we give the modeling and estimation procedure for a monotonically increasing signal. Extension to a monotonically decreasing one is straightforward. Since we want to have an increasing monotonic nature for the latent variable x_t , we want to calculate the probability $p(x_t|x_{t-1}, r_t \geq 0)$. If we condition the transition probability on $r_t \geq 0$, a high ρ means, $e_t \geq 0$ with a great chance, hence making $x_t \geq x_{t-1}$. We can write $p(x_t|x_{t-1}, r_t \geq 0)$ as

$$p(x_t|x_{t-1}, r_t \geq 0) = \frac{p(r_t \geq 0|x_{t-1}, x_t) p(x_t|x_{t-1})}{p(r_t \geq 0|x_{t-1})}. \quad (11)$$

We can see that

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t; x_{t-1}, \sigma_e^2), \quad (12)$$

$$p(r_t \geq 0|x_{t-1}) = p(r_t \geq 0) = 0.5 = \Phi(0; 0, \sigma_e^2). \quad (13)$$

The pdf $p(r_t \geq 0|x_{t-1}, x_t)$ can be calculated using (3) as

$$p(r_t \geq 0|x_{t-1}, x_t) = 1 - \Phi(0; \rho(x_t - x_{t-1}), \sigma_e^2(1 - \rho^2)) \\ = \Phi(\rho(x_t - x_{t-1}); 0, \sigma_e^2(1 - \rho^2)). \quad (14)$$

Substituting (12), (13) and (14) in (11), we get that

$$p(x_t|x_{t-1}, r_t \geq 0) = CSN_{1,1}(x_t; x_{t-1}, \sigma_e^2, \rho, 0, \sigma_e^2(1 - \rho^2)). \quad (15)$$

The transition pdf hence is a CSN with the skewness parameter as ρ . Since ρ is set to be close to 1, the resulting distribution is positively skewed ensuring that the mean of $x_t|x_{t-1}$ is greater than zero. Fig. 1 shows the pdf $p(x_t|x_{t-1}, r_t \geq 0)$ for $\rho = 0, 0.9$ and 0.9999 . For $\rho = 0$, we get the Gaussian distribution. As ρ increases and tends towards 1, the skewness increases and when it reaches 0.9999, it almost resembles a truncated Gaussian distribution with truncation below mean. This ensures that $x_t > x_{t-1}$ resulting in a monotonic trend. For the observation model, we have

$$p(y_t|x_t) = \mathcal{N}(y_t; x_t, \sigma_v^2). \quad (16)$$

Equations (16) and (15) together define the system.

3.2 Filtering

Once the pdfs are set up, the next task is to perform state estimation. The first step is the forward pass of filtering which contains the prediction and update steps.

Prediction step: To begin, let us assume that at time t we have

$$p(x_{t-1}|Y_{t-1}) = CSN_{1,1}(x_{t-1}; x_{t-1|t-1}, \Sigma_{t-1|t-1}, \\ \Gamma_{t-1|t-1}, \nu_{t-1|t-1}, \Delta_{t-1|t-1}) \quad (17)$$

with the dimension of Φ being 1. Since $p(x_t|x_{t-1}, r_t \geq 0)$ is $CSN_{1,1}(x_t; x_{t-1}, \sigma_e^2, \rho, 0, \sigma_e^2(1 - \rho^2))$, using (4), we have

$$p(x_t|Y_{t-1}, r_t \geq 0) = \int p(x_t|x_{t-1}, r_t \geq 0) p(x_{t-1}|Y_{t-1}) dx_{t-1} \\ = \int CSN_{1,1}(x_t; \dots) CSN_{1,1}(x_{t-1}; \dots) dx_{t-1}. \quad (18)$$

The prediction step equations for the case where the prior and the transition pdf are CSN are given in Rezaie and Eidsvik (2014) and the readers are requested to refer to this paper for the detailed equations. Applying these equations for our model, we get

$$p(x_t|Y_{t-1}, r_t \geq 0) = \\ CSN_{1,2}(x_t; x_{t|t-1}, \Sigma_{t|t-1}, \Gamma_{t|t-1}, \nu_{t|t-1}, \Delta_{t|t-1}). \quad (19)$$

with

$$x_{t|t-1} = x_{t-1|t-1}; \quad \Sigma_{t|t-1} = \Sigma_{t-1|t-1} + \sigma_e^2; \\ \Gamma_{t|t-1} = \begin{bmatrix} \Gamma_{t-1|t-1} \Sigma_{t-1|t-1} \Sigma_{t|t-1}^{-1} \\ \rho \sigma_e^2 \Sigma_{t|t-1}^{-1} \end{bmatrix}; \quad \nu_{t|t-1} = \begin{bmatrix} \nu_{t-1|t-1} \\ 0 \end{bmatrix}; \\ \Delta_{t|t-1} = \begin{bmatrix} \Delta_{t-1|t-1}^{11} & \Delta_{t-1|t-1}^{12} \\ \Delta_{t-1|t-1}^{21} & \Delta_{t-1|t-1}^{22} \end{bmatrix}. \quad (20)$$

Here,

$$\Delta_{t|t-1}^{11} = \Delta_{t-1|t-1}^{11} + \Gamma_{t-1|t-1} \Sigma_{t-1|t-1} \Gamma_{t-1|t-1}^{-1} - \\ \Gamma_{t-1|t-1} \Sigma_{t-1|t-1} \Sigma_{t|t-1}^{-1} \Sigma_{t-1|t-1} \Gamma_{t-1|t-1}^{-1}, \\ \Delta_{t|t-1}^{22} = \sigma_e^2 - \rho^2 \sigma_e^4 \Sigma_{t|t-1}^{-1}, \\ \Delta_{t|t-1}^{12} = \Delta_{t-1|t-1}^{12} = -\Gamma_{t-1|t-1} \Sigma_{t-1|t-1} \Sigma_{t|t-1}^{-1} \rho \sigma_e^2. \quad (21)$$

We can see that the dimension of the skewness part gets increased by 1 in the prediction step. This will blow-up with t which will make the calculation of the moments of CSN difficult. Hence we need to approximate the two-dimensional skewness with a one-dimensional term. This will be shown after the update step.

Update step: Using (5) and ignoring terms without x_t , we have

$$p(x_t|Y_t, r_t \geq 0) \propto \mathcal{N}(y_t; x_t, \sigma_v^2) \mathcal{N}(x_t; x_{t|t-1}, \Sigma_{t|t-1}) \times \\ \Phi_2(\Gamma_{t|t-1}(x_t - x_{t|t-1}); \nu_{t|t-1}, \Delta_{t|t-1}). \quad (22)$$

Note that in the above equation we have a product of two Gaussians $\mathcal{N}(y_t|x_t)$ and $\mathcal{N}(x_t)$. Using the Bayesian rule expressions for Gaussians, we can write this as

$$\mathcal{N}(y_t; x_t, \sigma_v^2) \mathcal{N}(x_t; x_{t|t-1}, \Sigma_{t|t-1}) = \\ \mathcal{N}(y_t; x_{t|t-1}, \sigma_v^2 + \Sigma_{t|t-1}) \times \\ \mathcal{N}(x_{t|t-1} + K_t(y_t - x_{t|t-1}), (1 - K_t)\Sigma_{t|t-1}). \quad (23)$$

where $K_t = \Sigma_{t|t-1}/(\Sigma_{t|t-1} + \sigma_v^2)$. Using (23), we can write the posterior probability of our case as

$$p(x_t|Y_t, r_t \geq 0) \propto \mathcal{N}(x_t; x_{t|t}, \Sigma_{t|t}) \\ \times \Phi(\Gamma_{t|t-1}(x_t - x_{t|t-1}); \nu_{t|t-1}, \Delta_{t|t-1}), \quad (24)$$

with $x_{t|t-1} + K_t(y_t - x_{t|t-1}) = x_{t|t}$ and $(1 - K_t)\Sigma_{t|t-1} = \Sigma_{t|t}$. Also, we need to readjust the skewness term to make it a CSN. This can be done as

$$\Phi(\Gamma_{t|t-1}(x_t - x_{t|t-1} + x_{t|t} - x_{t|t}); \nu_{t|t-1}, \Delta_{t|t-1}) = \\ \Phi(\Gamma_{t|t-1}(x_t - x_{t|t}); \nu_{t|t-1} - \Gamma_{t|t-1}(x_{t|t} - x_{t|t-1}), \Delta_{t|t-1}). \quad (25)$$

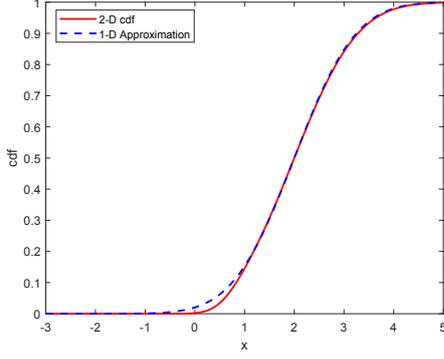


Fig. 2. Approximating a 2-D cdf with $\Gamma = [1 \ 2]'$, $\nu = [1 \ -1]'$ and $\Delta = [1 \ -0.5; -0.5 \ 1]$ as a 1-D cdf.

Note that in the update step the dimension of the skewness term has remained same as the noise model is Gaussian. We now have

$$p(x_t|Y_t, r_t \geq 0) = CSN_{1,2}(x_t; x_{t|t}, \Sigma_{t|t}, \Gamma_{t|t}, \nu_{t|t}, \Delta_{t|t}). \quad (26)$$

where

$$\begin{aligned} x_{t|t} &= x_{t|t-1} + K_t(y_t - x_{t|t-1}); & \Sigma_{t|t} &= (1 - K_t)\Sigma_{t|t-1}; \\ \Gamma_{t|t} &= \Gamma_{t|t-1}; & \nu_{t|t} &= \nu_{t|t-1} - \Gamma_{t|t-1}(x_{t|t} - x_{t|t-1}); \\ \Delta_{t|t} &= \Delta_{t|t-1}. \end{aligned} \quad (27)$$

Issue of dimensionality: As we see in the prediction step, the dimensionality of $\Phi()$ in CSN increases by one at every time instant. To avoid the blowing up of dimensionality, we reduce it to 1 at the end of the update step. As pointed out in Rezaie and Eidsvik (2014), how to best make such approximations needs more research. However, for our case where we need to reduce the dimensionality from 2 to 1. We propose a simple procedure which is summarized in the following steps.

- (1) Say we have a two-dimensional cdf given for a one-dimensional x as follows.

$$\Phi\left(\begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} (x - \mu); \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}\right). \quad (28)$$

Generate the cdf as a function of x

- (2) Find the values of x when Φ is 0.8413, 0.5 and 0.1587.
- (3) Set the new parameters of the cdf as $\Gamma^\# = 1$, $\nu^\# = x_{(\Phi=0.5)} - \mu$ and $\Delta^\# = (x_{(\Phi=0.8413)} - x_{(\Phi=0.1587)})^2 \times 0.25$. Set $\Gamma_{t|t} = \Gamma^\#$, $\nu_{t|t} = \nu^\#$ and $\Delta_{t|t} = \Delta^\#$

The above steps reduce the dimensionality of cdf from 2 to 1. This approximation works well for a good range of Γ , ν and Δ . Once such approximation is depicted in Fig. 2.

Estimating the moments of CSN: At the end of the approximation step, we would have a CSN with a one-dimensional cdf term. We would then need to find the moments of the distribution to get the estimate of the states. The mean and covariance of the CSN pdf with one-dimensional cdf are given as

$$\begin{aligned} \hat{x}_{t|t} &= x_{t|t} + \Sigma_{t|t}\Gamma_{t|t}\eta; & \eta &= \frac{\mathcal{N}(0; \nu_{t|t}, \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma_{t|t})}{\Phi(0; \nu_{t|t}, \Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma_{t|t})} \\ \hat{P}_{t|t} &= \Sigma_{t|t} + \Sigma_{t|t}^2\Gamma_{t|t}^2\Lambda - \Sigma_{t|t}^2\Gamma_{t|t}^2\eta^2; & \Lambda &= \frac{\eta\nu_{t|t}}{\Delta_{t|t} + \Gamma_{t|t}\Sigma_{t|t}\Gamma_{t|t}} \end{aligned} \quad (29)$$

Here $\hat{x}_{t|t}$ and $\hat{P}_{t|t}$ are the mean and covariance of the approximated posterior CSN distribution.

3.3 Smoothing

In this step, we move backward estimating the pdf of the state given all the available observations in accordance with (6) with an added condition of $r_t \geq 0$. For compact notations, we leave out denoting this. As we see from this equation, there are two steps involved in deriving the smoothing equations. First we need to calculate the pdf $p(x_t|x_{t+1}, Y_T)$ by Bayesian inversion using $p(x_t|Y_t)$ as prior and $p(x_{t+1}|x_t)$ as the likelihood. Then we need to perform calculations similar to the prediction step in (4).

The Bayesian inversion can be performed similar to the update step. To keep the dimensionality to one, we take $p(x_t|Y_t) = \mathcal{N}(x_t; \hat{x}_{t|t}, \hat{P}_{t|t})$ which is a Gaussian with the actual mean and variance of CSN. The Bayesian inversion step leads to the following results.

$$p(x_t|x_{t+1}, Y_t) \propto p(x_{t+1}|x_t)p(x_t|Y_t). \quad (30)$$

Plugging the respective pdfs and proceeding similar to (22)-(25), we get

$$p(x_t|x_{t+1}, Y_t) = CSN(x_t, \mu_t^*, \Sigma_t^*, \Gamma_t^*, \nu_t^*, \Delta_t^*). \quad (31)$$

$$\mu_t^* = \hat{x}_{t|t} + C_t(x_{t+1} - \hat{x}_{t|t}); \quad \Sigma_t^* = (1 - C_t)\Sigma_{t|t};$$

$$\Gamma_t^* = -\rho; \quad \nu_t^* = -\rho(x_{t+1} - \mu_t^*); \quad \Delta_t^* = \sigma_e^2(1 - \rho^2). \quad (32)$$

Here, $C_t = \Sigma_{t|t}/(\Sigma_{t|t} + \sigma_e^2)$. We can see that ν_t^* is a function of x_{t+1} . As a result of this, the smoothing equation would look as follows.

$$p(x_t|Y_T) = \int \frac{\mathcal{N}() \Phi(\dots \nu_t^*(x_{t+1}))}{\Phi(\dots \nu_t^*(x_{t+1}))} p(x_{t+1}|Y_T) dx_{t+1}. \quad (33)$$

Because of ν_t^* being a function of x_{t+1} , we have a cdf term in the denominator which is a function of x_{t+1} . This term makes the integration difficult. Hence, recursive smoothing equations cannot be derived easily. We hence resort to an importance sampling-based method for the smoothing step.

Importance sampling: Let $f(x)$ be a function whose expected value is to be found according to a distribution $p(x)$. With a sampling approach, we can estimate it as

$$\mathbb{E}[f(x)] = \int f(x) p(x) dx = \frac{1}{N} \sum_{i=1}^N f(x^{(i)}); \quad x^{(i)} \sim p(x). \quad (34)$$

However, if we cannot sample from $p(x)$, we can generate samples from a sampling distribution $q(x)$ (from which sampling can be done easily) and then proceed as follows.

$$\mathbb{E}[f(x)] = \int f(x) \frac{p(x)}{q(x)} q(x) dx = \int w(x) f(x) q(x) d(x) \quad (35)$$

Now, we can consider $f(x)w(x)$ as the function and $q(x)$ as the distribution. We have

$$\mathbb{E}[f(x)] = \frac{1}{N} \sum_{i=1}^N w(x^{(i)}) f(x^{(i)}); \quad x^{(i)} \sim q(x). \quad (36)$$

Smoothing via importance sampling: At $t = T - 1$ we have

$$p(x_{T-1}|Y_T) = \int CSN(x_{T-1}|x_T, Y_T) \times CSN(x_T|Y_T) dx_T. \quad (37)$$

Since, sampling from CSN is difficult (and later we would have a mixture CSN for $p(x_t|Y_T)$), we sample from $q(x_T|Y_T) = \mathcal{N}(x_T; \hat{x}_{T|T}, \hat{P}_{T|T})$, where $\hat{x}_{T|T}$ and $\hat{P}_{T|T}$ are the means of the posterior CSN at the end of update step calculated according to (29). We now have

$$p(x_{T-1}|Y_T) = \frac{1}{N} \sum_{i=1}^N w_T^{(i)} CSN(x_{T-1}|x_T^{(i)}), \quad (38)$$

where

$$w_T^{(i)} = \frac{CSN(x_T^{(i)}|Y_T)}{\mathcal{N}(x_T^{(i)}; \hat{x}_{T|T}, \hat{P}_{T|T})}; \quad x_T^{(i)} \sim q(x_T|Y_T). \quad (39)$$

Equation (38) is a mixture of CSN distributions and the moments can be estimated as weighted moments of each CSN. Hence we have

$$\begin{aligned} \hat{x}_{T-1|T} &= \frac{1}{N} \sum_{i=1}^N w^{(i)} \hat{x}_{T-1|T}^{(i)}, \\ \hat{P}_{T-1|T} &= \frac{1}{N} \sum_{i=1}^N w^{(i)} (\hat{P}_{T-1|T}^{(i)} + (\hat{x}_{T-1|T}^{(i)} - \hat{x}_{T-1|T})^2). \end{aligned} \quad (40)$$

For the next time step, we would have

$$p(x_{T-2}|Y_T) = \int CSN(x_{T-2}|x_{T-1}, Y_T) p(x_{T-1}|Y_T) dx_{T-1}. \quad (41)$$

Since $p(x_{T-1}|Y_T)$ is a mixture CSN, we will sample from $\mathcal{N}(x_{T-1}; \hat{x}_{T-1|T}, \hat{P}_{T-1|T})$. The resulting weights would be

$$w_{T-1}^{(i)} = \frac{\frac{1}{N} \sum_{j=1}^N w_T^{(j)} CSN(x_{T-1}^{(i)}|x_T^{(j)})}{\mathcal{N}(x_{T-1}^{(i)}; \hat{x}_{T-1|T}, \hat{P}_{T-1|T})}; \quad x_{T-1}^{(i)} \sim q(x_{T-1}|Y_T) \quad (42)$$

This means $p(x_{T-2}|Y_T)$ is a mixture of CSN pdf and we continue the process till we reach $p(x_1|Y_T)$. We can observe that $p(x_t|x_{t+1}, Y_t)$ is a CSN with negative skewness. Since $p(x_t|Y_T)$ is a weighted sum of CSN distributions all with a negative skewness, if ρ is selected to be close to 1, we will have $\hat{x}_{t|T} < \hat{x}_{t+1|T}$.

4. CASE STUDIES

We demonstrate the effectiveness of the proposed method on two case studies: a simulation case study and an industrial one related to a hot lime softener process (HLS).

4.1 Simulation case study

This case study is to demonstrate that the proposed method is able to extract monotonically increasing function. In this case study, we generated 250 samples of a sigmoid function as the HMT x . The sigmoid function is as follows

$$x_t = 10 / (1 + \exp(-0.08 * (t - 125))). \quad (43)$$

The observed output y_t is generated by adding Gaussian noise to x_t . We implement the proposed method on this data to extract the monotonic x_t and compare it with Kalman filter (KF) results. Fig. 3 depicts the extracted

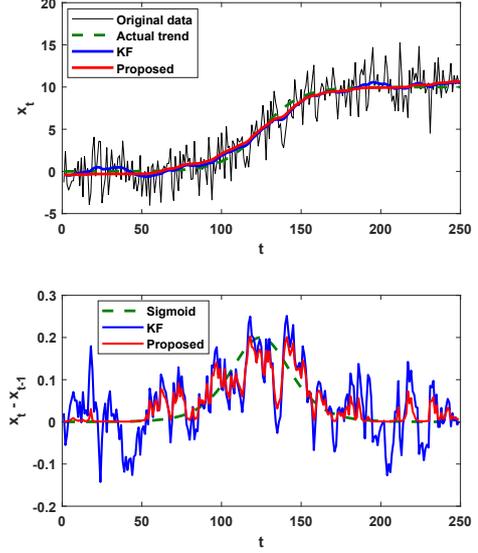


Fig. 3. Comparing the extracted signals (top figure) and their rate of change (bottom figure) against the original sigmoid function.

signals and their rate of change using KF and the proposed method. It is clear that the proposed method is able to extract the monotonic signal well and the KF one has a few unnecessary oscillations. For processes with monotonic drifts, the rate of change of the monotonic drift also would be of interest. In such cases, KF might give misleading results with negative rates, while the monotonic signals will not. Also, for a given RMSE with the original sigmoid, monotonic signals are smoother than the ones obtained from the KF enabling easier and more accurate monitoring of the process.

4.2 Fouling prediction in hot lime softener process

The data in this case study is obtained from an industrial HLS unit. HLS is used to soften the produced water before sending it to a steam generator in a steam-assisted gravity drainage process. Produced water is mixed in a tank with certain chemicals and is sent to filters to remove suspended solids. Fouling deposition occurs in the pipeline between the HLS tank and the filters. The approach is based on the Darcy-Weisbach equation for flow through a pipe:

$$(F/\sqrt{\Delta P})^{0.8} \propto A. \quad (44)$$

where F is the flowrate, ΔP is the pressure drop across the pipe and A is the cross-section area available for the flow. During fouling, the cross-section area A decreases gradually, and hence the quantity $(F/\sqrt{\Delta P})^{0.8}$ which will be referred to as the flow-coefficient, also decreases. Therefore, we may monitor the flow coefficient for fouling buildup. However, monitoring the flow coefficient as such is not reliable and apparent as it is noisy. Hence we need to filter the noises to isolate the trend. As mentioned in Alsadaie and Mujtaba (2017), fouling rates can be classified as linear, falling rate, asymptotic, and sawtooth nature. The first three are monotonic while the sawtooth behavior is monotonic initially but starts oscillating after some time because of the flushing away of the deposited material. Flushing away usually happens when there is sufficient deposition of fouling and hence can be an indicator of

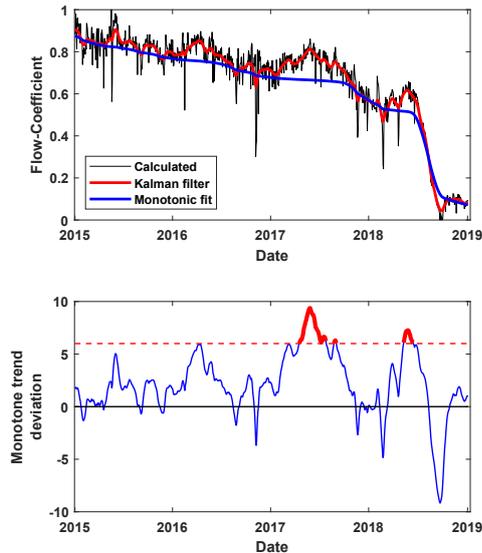


Fig. 4. The original and the filtered flow coefficient (top figure). The error between KF and the monotonic trend is monitored to detect fouling (bottom figure).

sufficient fouling buildup. The proposed method can be applied to all these categories to detect fouling.

In the HLS process, we observe the saw-tooth nature of fouling. If we fit a monotonic trend to a saw-tooth curve, it becomes easier to detect the saw-tooth nature, hence the beginning of fouling. Based on this rationale, we propose the following steps for a fouling monitoring scheme.

- (1) Calculate the flow coefficient. Smoothen this using the KF to get the red curve in the top figure of Fig. 4.
- (2) Implement the proposed method on the calculated flow coefficient to get the monotonic trend (blue curve in the top figure of Fig. 4).
- (3) Calculate the error between the red and the blue curves and monitor it.

In the initial phases of fouling, there is steady deposition. Hence the error between the monotonic curve and the red curve will be small. As the sawtooth nature starts, the monotonic trend will not be able to track the periodic increasing trends. Hence the error increases. Based on trial and error, one can set a threshold to the error and detect when the saw-tooth behavior starts which is an indicator of sufficient fouling buildup. We have used the Kalman filtered signal to calculate the error with the monotonic trend because using the calculated coefficient directly results in very noisy errors making deductions difficult. From Fig. 4 we can see that the proposed scheme has successfully issued a warning of fouling at about mid of 2017 before a rapid deterioration of the flow coefficient. Despite the usage of importance sampling, given that there is only one dimension of x , the method can be applied for both online and offline trend visualization.

5. CONCLUSION

The proposed method for extracting the HMT models the hidden signal as a CSN distribution. Similar to Gaussian distribution, filtering schemes have been developed for CSN in literature and we have proposed a smoothing

scheme based on importance sampling. The efficacy of the proposed methods has been demonstrated in two case studies encouraging further exploration of CSN hidden Markov models for analyzing industrial datasets.

REFERENCES

- Alsadaie, S.M. and Mujtaba, I.M. (2017). Dynamic modelling of heat exchanger fouling in multistage flash (msf) desalination. *Desalination*, 409, 47–65.
- Arellano-Valle, R.B., Contreras-Reyes, J.E., Quintero, F.O.L., and Valdebenito, A. (2019). A skew-normal dynamic linear model and bayesian forecasting. *Computational Statistics*, 34(3), 1055–1085.
- González-Farías, G., Dominguez-Molina, A., and Gupta, A.K. (2004). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, 126(2), 521–534.
- Gorinevsky, D. (2004). Monotonic regression filters for trending deterioration faults. In *proceedings of the 2004 American Control Conference*, volume 6, 5394–5399. IEEE.
- Gorinevsky, D. (2008). Efficient filtering using monotonic walk model. In *2008 American Control Conference*, 2816–2821. IEEE.
- Iversen, D.H. (2010). *Closed-skew distributions: Simulation, inversion and parameter estimation*. Master’s thesis, Institutt for matematiske fag.
- Karimi, O., Omre, H., and Mohammadzadeh, M. (2010). Bayesian closed-skew gaussian inversion of seismic avo data for elastic material properties. *Geophysics*, 75(1), R1–R11.
- Naveau, P., Genton, M.G., and Shen, X. (2005). A skewed kalman filter. *Journal of multivariate Analysis*, 94(2), 382–400.
- Rezaie, J. and Eidsvik, J. (2014). Kalman filter variants in the closed skew normal setting. *Computational Statistics & Data Analysis*, 75, 1–14.
- Rezaie, J. and Eidsvik, J. (2016). A skewed unscented kalman filter. *International Journal of Control*, 89(12), 2572–2583.
- Samar, S., Gorinevsky, D., and Boyd, S. (2004). Moving horizon filter for monotonic trends. In *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, volume 3, 3115–3120. IEEE.
- Särkkä, S. (2013). *Bayesian filtering and smoothing*, volume 3. Cambridge University Press.
- Schirru, A., Pampuri, S., and De Nicolao, G. (2010). Particle filtering of hidden gamma processes for robust predictive maintenance in semiconductor manufacturing. In *2010 IEEE International Conference on Automation Science and Engineering*, 51–56. IEEE.
- Simon, D. and Simon, D.L. (2010). Constrained kalman filtering via density function truncation for turbofan engine health estimation. *International Journal of Systems Science*, 41(2), 159–171.
- Susto, G.A., Schirru, A., Pampuri, S., Beghi, A., and De Nicolao, G. (2018). A hidden-gamma model-based filtering and prediction approach for monotonic health factors in manufacturing. *Control Engineering Practice*, 74, 84–94.
- van Noortwijk, J.M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1), 2–21.