Inference for Non-Gaussian Dynamical Models with Time-varying Skew

Zixin Liu
Department of Engineering
University of Cambridge
Cambridge, UK

Zachary Tiller
Department of Engineering
University of Cambridge
Cambridge, UK

Simon Godsill
Department of Engineering
University of Cambridge
Cambridge, UK

***Abstract*—In this paper we introduce tracking models based on non-Gaussian continuous time stochastic processes with time-varying skewness. The idea behind this is that the skewness of the dynamical model may be able to model a propensity for an object to undergo manoeuvres of a particular type, for example velocities tending in a particular direction, but that these may change over time. This process is constructed based on a random series representation of conditionally Gaussian Lévy processes, which enables straightforward simulation of the models. We demonstrate the specific example of α-stable processes and find that such processes can capture abrupt changes owing to their heavy-tailed behaviour, and demonstrate the random changes in direction caused by the time-changing skewness of the distribution. We propose methods for joint tracking of both states and skewness for such processes, based on a marginalised particle filter, which are demonstrated to perform well even with limited numbers of particles.**

***Index Terms*—non-Gaussian stochastic process, sequential Monte Carlo, particle filter, Lévy process, stochastic differential equation**

I. Introduction

In order to model the movement of an object, it is common to use stochastic differential equations (SDEs) such as in [1], [2]. The noise term in the SDEs represent the randomness of the movement. It is well known that many real-world data exhibit heavy-tailed behavior which cannot be captured by Gaussian noise, such as in financial systems [3]–[5], communications [6]–[9], signal processing [10], image analysis [11], audio processing [12], [13], and in climatological sciences [14], [15]. Therefore, more general types of process should ideally be used to model the noise in such systems.

A natural choice of non-Gaussian model in continuous time is the Lévy process [5], [16]. An important property of Lévy processes is an infinitely divisible distribution that defines the process over time. However, in real life, the parameters of the system can also be randomly time-varying. Intuitively, moving objects do not necessarily follow the same process all the time as they undergo different manoeuvres. Therefore, we propose a more general time-varying process by setting the skewness of the process to be another stochastic process - a Brownian motion, in our case. Processes driven by such time-varying noise are also readily simulated within the framework we propose, and Monte Carlo likelihoods may be computed in order to run algorithms such as sequential Monte Carlo [17], [18].

Here an α-stable model is assumed for the underlying Lévy process as a a concrete example, where after some time T the process will reach a particular α-stable distribution, and we adopt a conditionally Gaussian representation of this process as in [19]. The skewness of the distribution is readily parameterised in terms of the mean and standard deviation of the conditional Gaussian form, and by allowing the mean to vary with time we propose to model time-varying skew. Simulation and inference for conditionally Gaussian Lévy-driven state-space models have been proposed in [19], while [1] extends the methods to spatial tracking scenarios, their work is limited to the symmetric (non-skewed) α-stable case. In our work here, we consider a time-varying skewness and state space models in multiple spatial dimensions. Inference for the time-varying mean parameter of the stochastic process is carried out within a marginalised particle filtering framework, using a novel extension of the approach for fixed mean and skewness in [19].

II. Modelling Framework

Consider now a stochastic process on a time axis $\tau = [0, T]$, which can be described by a linear stochastic differential equation (SDE) [20] taking the form

$$dX(t) = AX(t)dt + \mathbf{h}dW(t), X(t) \in \Re^P, W(t) \in \Re^1 \quad (1)$$

where $W(t)$ is the driving noise process, at time $t \in \tau$. Let $s \in \tau, s < t$; then, integrating (1) from time s to t gives

$$X(t) = e^{A(t-s)}X(s) + \int_s^t e^{A(t-u)}\mathbf{h}dW(u)\,.$$

Consider the stochastic integral in this equation, defining

$$\mathbf{I}(s,t) := \int_s^t e^{A(t-u)}\mathbf{h}dW(u)\,. \quad (2)$$

Now consider the form of $W(t)$, which will be modelled as a pure jump stochastic process [5]; we start from the series representation of $W(t)$ using a generalised shot noise representation, following [21]. Define $\mathcal{I}(A) = 1$ if event A is true and 0 otherwise. Let $\stackrel{d}{=}$ denote convergence in distribution. Then, according to [21],

$$W(t) \stackrel{d}{=} \sum_{i=1}^{\infty} J_i \mathcal{I}(V_i < t) - tc_i \quad (3)$$

where J_i are random jumps of the process occuring at random times $V_i \stackrel{iid}{\sim} \mathcal{U}(\tau)$ (Uniform distribution), and c_i are centering terms that are required for the convergence of the series in certain cases, such as the α-stable case with $\alpha \geqslant 1$. In this paper we deal only with cases where centering is not necessary and set all c_i to zero (more general cases with centering will be presented in future work). The jumps J_i are represented as a function $J_i = H(\Gamma_i, U_i)$ where $\{\Gamma_i\}$ are event times (epochs) of a unit-rate Poisson process, $H(\gamma, .)$ is a function that is non-increasing in γ, and U_i are iid random variates, here taken as standard normal $U_i \stackrel{iid}{\sim} \mathcal{N}(0,1)$. Two cases will be of interest to us here; firstly, define

$$H(\Gamma_i, U_i) = h(\Gamma_i)(\mu(V_i) + \sigma_W U_i),\ U_i \sim \mathcal{N}(0,1)\,. \quad (4)$$

We will denote this first case as the Normal-σ-Mean (NσM) mixture model, where $h()$ is a non-increasing function. This case will provide the basis for the examples presented in the paper. In particular the α-stable stochastic process is obtained with $h(\gamma) = \gamma^{-1/\alpha}$ ($\alpha \in (0,1)$), see [22]. $\alpha \in [1,2)$ is also possible but would require centering c_i, so we omit this case in the present context.

Secondly, the so-called Normal-Variance-Mean (NVM) model [5], [23] is obtained with $H(\Gamma_i, U_i) = h(\Gamma_i)\mu(V_i) + \sqrt{h(\Gamma_i)}\sigma_W U_i$. The NVM model is not taken further in this paper, but related developments based on this important class will be presented in future work.

Notice that in both cases the term $\mu(t)$ is allowed to be time-varying, in contrast with previous work in which $\mu(t) = \mu_W$, an unknown constant.

Substituting (4) and (3) into (2) with $c_i = 0$ we get a series representation of the form:

$$\mathbf{I}(s,t) \stackrel{d}{=} \sum_{i=1}^{\infty} h(\Gamma_i)\mathbf{f}_t(V_i)(\mu(V_i) + \sigma_W U_i)\mathcal{I}(V_i \in [s,t))$$

where $\mathbf{f}_t(u) = e^{A(t-u)}\mathbf{h}$.

Since it is impractical to consider an infinite number of series terms, we truncate the summation to consider only large jumps, in the manner of [19], [23], leading to the following truncated series representation,

$$\mathbf{I}^c(s,t) = \sum_{\Gamma_i < c} h(\Gamma_i)\mathbf{f}_t(V_i)(\mu(V_i) + \sigma_W U_i)\mathcal{I}(V_i \in [s,t)) \quad (5)$$

Thus far the development is much as in the standard Lévy-driven process detailed in [19], [23]. In these earlier papers, as stated above, $\mu(V_i) = \mu_W$ (a constant) and σ_W determine the skewness, $\beta \in [-1,1]$, and scale, γ, of the process, which are fixed (but unknown) over time; for example, in the α-stable case the relationships, according to [24], are

$$\beta = \frac{\mathbb{E}\left[|\mu_W + \sigma_W U_1|^\alpha \mathrm{sign}(\mu_W + \sigma_W U_1)\right]}{\mathbb{E}\left[|\mu_W + \sigma_W U_1|^\alpha\right]}$$

$$\gamma^\alpha = \frac{\mathbb{E}\left[|\mu_W + \sigma_W U_1|^\alpha\right]}{C_\alpha}$$

where C_α is a constant depending on α, and $\mathbb{E}$ denotes expectation.

Here though our contribution is to introduce a broader class of stochastic processes, in which the parameter $\mu(V_i)$, and hence the skewness of the process, may vary over time, for example as a Brownian motion[1]. The motivation for this is that the skewness of the process may indicate a propensity of an object to move preferentially in one direction rather than another (an example of 'intent' analysis), and that this propensity may be expected to vary with time in manoeuvring or evasive object bahaviours. In order to achieve this objective, define a new process $d\mu(t) = \sigma_\mu dB(t)$, where $B(t)$ is a standard Brownian motion. The resulting process $W(t)$ is no longer a Lévy process as the increments are not stationary, but it does have the advantage of allowing time-varying skewness in the model. In this current work we will ignore the error due to truncation of the infinite summation, which is expected to be small if c is set sufficiently large, a topic that will be explored in a future publication, (see [19], [23] for the full treatment of the residual error when $\mu(t) = \mu_W$ is constant).

With a direct truncation of the infinite series, the resulting state dynamical model is expressed as

$$X(t) = e^{A(t-s)}X(s) + \mathbf{I}^c(s,t) \quad (6)$$

and in the next section we summarise how to simulate the new process.

III. Simulation

First note that for a time interval $(s,t] \in \tau$ the Γ_is and V_is need not be generated for the whole interval τ. We can equivalently generate $\{\Gamma_i\}_{(s,t]}$ as a Poisson process of rate $(t-s)/T$, then for each such Γ_i, generate a uniform $V_i \sim \mathcal{U}(s,t]$. [2]

The process can then be simulated by generating sets of $\{\Gamma_i\}^c_{(s,t]}, \{V_i\}^c_{(s,t]}, \{\mu(V_i)\}^c_{(s,t]}, \{U_i\}^c_{(s,t]}$ at each time step, where the superscript c represents that all $\Gamma_i \leqslant c$ and the subscript $(s,t]$ denotes that $V_i \in (s,t]$ and hence that the jumps happen in the time interval $(s,t]$. Now, observing that $X(t)$ in (6) is Gaussian conditionally on the Γs, Vs and μs, we can write the conditional transition density of $X(t)$ as

$$X(t)|X(s), \{\Gamma_i\}^c_{(s,t]}, \{V_i\}^c_{(s,t]}, \{\mu(V_i)\}^c_{(s,t]} \sim \mathcal{N}\left(e^{A(t-s)}X(s) + \mathbf{m}_{(s,t]}, \mathbf{S}_{(s,t]}\right) \quad (7)$$

where

$$\mathbf{m}_{(s,t]} = \sum_{\Gamma_i < c} h(\Gamma_i)\mathbf{f}_t(V_i)\mu(V_i)\mathcal{I}(V_i \in [s,t)) \quad (8)$$

and

$$\mathbf{S}_{(s,t]} = \sigma_W^2 \sum_{\Gamma_i < c} h(\Gamma_i)^2 \mathbf{f}_t(V_i)\mathbf{f}_t^T(V_i)\mathcal{I}(V_i \in [s,t)) \quad (9)$$

[1] Essentially any linear Gaussian state space model for $\mu(t)$ would work here, but we study Brownian motion as a simple starting point.

[2] This method is exactly equivalent to generating the $\{\Gamma_i\}$ from a unit rate Poisson process with associated times $V_i \sim \mathcal{U}(\tau)$ and then *thinning* the points to select only those having $V_i \in (s,t]$ [25].

We can therefore just generate Γs, Vs and μs for each time interval required and sample $X(t)$ from its Gaussian conditional distribution. If simulation is required at a list of increasing times $0, t_1, t_2, ..., T$, then we simply apply the Gaussian sampling procedure successively over each sub-interval $(t_{j-1}, t_j]$ for $j = 1, ...,$ thus obtaining a 'skeleton' path of the process $X(0), X(t_1), ..., X(T)$. Note also that in practice the algorithm is easily extended to times beyond the initial axis τ by re-starting the generalised shot-noise model for $W(t)$ on a subsequent time axis $\tau + T$ and repeating as required.

The complete simulation algorithm is shown in Alg. 2.

Algorithm 1: Sample Latent Variables

Input : Time interval $(s, t]$
Output: Ordered latent variables $\{\{\Gamma, V\}_j\}^c_{(s,t]}$
begin
  Initialize $\Gamma = 0$, $\{\Gamma_i, V_i\}^c_{(s,t]} = \varnothing$
  while $\Gamma < c$ **do**
    $\Gamma += r$, $r \sim \exp(T/(t-s))$
    $V \sim \mathcal{U}(s, t]$
    add (Γ, V) to $\{\Gamma_i, V_i\}^c_{(s,t]}$
  Sort the set of pairs in the order of increasing Vs, to obtain the ordered set $\{\Gamma_j, V_j\}^c_{(s,t]}$

Algorithm 2: Simulation of sample paths

Input: An increasing set of observation times, $\{t_n\}_{n=1}^N$
Output: Sample path $\{X(t_n)\}_{n=1}^N$
begin
  Initialise $X(0)$, $t_0 = 0$, $\mu(0)$
  for $n = 1 \dots N$ **do**
    Sample $\{\Gamma_j, V_j\}^c_{(t_{n-1},t_n]}$ by Alg. 1
    for $(\Gamma_j, V_j) \in \{\Gamma_j, V_j\}^c_{(t_{n-1},t_n]}$ **do**
      Sample $\mu(V_j)$ from $\mathcal{N}(\mu(V_{j-1}), (V_j - V_{j-1})\sigma_\mu^2)$ defining $V_0 = t_{n-1}$
    Extract largest time $V_{max} = \max_j \{V_j\}^c_{(t_{n-1},t_n]}$
    Sample $\mu(t_n)$ from $\mathcal{N}(\mu(V_{max}), (t_n - V_{max})\sigma_\mu^2)$
    Compute $\mathbf{m}_{(s,t]}$ and $\mathbf{S}_{(s,t]}$ by (8) and (9)
    Sample $X(t_n)|X(t_{n-1})$ from the normal distribution (7)

IV. State Inference and Tracking

We now describe how to carry out inference in the new model, including estimation of the kinematic state $X(t)$ and the time-varying mean process $\mu(t)$.

A. State space model

In order to pose the inference problem, we first rewrite the state transition model using an extended state vector that includes both $X(t)$ and $\mu(t)$, following the fixed-μ case in [19],

$$\boldsymbol{\alpha}_t = \begin{bmatrix} X(t) \\ \mu(t) \end{bmatrix}, \ \boldsymbol{\alpha}_t = \mathbf{A}\boldsymbol{\alpha}_s + \mathbf{B}\mathbf{e}_{s,t}, \mathbf{e}_{s,t} \sim \mathcal{N}(0, \mathbf{C}_{s,t}) \quad (10)$$

Now, suppose that measurements $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n, ...$ are made at discrete times $t_1, t_2, ..., t_n, ...$ through a linear Gaussian observation model:

$$\mathbf{y}_n = \mathbf{H}\mathbf{X}_{t_n} + \mathbf{v}_{t_n}, \mathbf{v}_{t_n} \overset{iid}{\sim} \mathcal{N}(0, \kappa_V^2\sigma_W^2\mathbf{I}), \mathbf{y}_n \in \Re^D \quad (11)$$

where σ_W^2 is once again the variance term in Eq. (4) and κ_V is a variance inflation term, as in [19].

Key to the new development is the specification of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}_{s,t}$ for the proposed time-varying model (10). Theorem 1 below gives the required result for the new time-varying model.

Theorem 1. *For the extended state transition model in (10) where the transition model for $X(t)$ is specified as (7), the terms $\mathbf{A}, \mathbf{B}, \mathbf{C}_{s,t}$ take the following form,*

$$\mathbf{A} = \begin{bmatrix} e^{A(t-s)} & \sum_{i=1}^I h(\Gamma_i)\mathbf{f}_t(V_i) \\ \mathbf{0}_{1\times P} & 1 \end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix} \tilde{\mathbf{m}}_{(s,t]} & \mathbf{I}_P \\ [\mathbf{0}_{1\times I} \quad 1] & \mathbf{0}_{1\times P} \end{bmatrix}$$
$$\mathbf{C}_{s,t} = \begin{bmatrix} \mathbf{C}_\mu & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{(s,t]} \end{bmatrix}$$

with the additional required terms defined in (13) and (15). Here $\mathbf{0}_{R\times C}$ represents an all zero matrix with R rows and C columns, $\mathbf{1}_{R\times C}$ is defined similarly as an $R \times C$ matrix of all unity elements, and $\mathbf{I}_R$ is the identity matrix with size $R \times R$.

Proof. First observe that (8) is a linear function of the unknown variables $\mu(V_i)$. If there are I jumps within the interval $(s, t]$ (I is a random Poisson variable on any time interval $(s, t]$), then we may collect together all of the terms corresponding to jumps at times $\{V_i \in (s, t]\} = \{V_1, ..., V_I\}$ as a vector ordered by increasing times V_i, plus the end point $\mu(t)$,

$$\boldsymbol{\mu}_{V_{(s,t]}} := \left[\mu(V_1), \mu(V_2), \ldots, \mu(V_I), \mu(t)\right]^T$$

then (8) can be written as

$$\mathbf{m}_{(s,t]} = \tilde{\mathbf{m}}_{(s,t]}\boldsymbol{\mu}_{V_{(s,t]}}, \quad (12)$$

where

$$\tilde{\mathbf{m}}_{(s,t]} = [h(\Gamma_1)\mathbf{f}_t(V_1), h(\Gamma_2)\mathbf{f}_t(V_2), ..., h(\Gamma_I)\mathbf{f}_t(V_I), 0] \quad (13)$$

so the state update from $(X(s), \boldsymbol{\mu}_{V_{(s,t]}})$ to $(X(t), \mu(t))$, can be expressed as follows,

$$\begin{bmatrix} X(t) \\ \mu(t) \end{bmatrix} = \mathbf{A}_2 \begin{bmatrix} X(s) \\ \boldsymbol{\mu}_{V_{(s,t]}} \end{bmatrix} + \mathbf{B}_2\mathbf{e}_t, \ \mathbf{e}_t \sim \mathcal{N}(0, \mathbf{S}_{(s,t]}) \quad (14)$$

where $\mathbf{S}_{(s,t]}$ is given by (9) and

$$\mathbf{A}_2 = \begin{bmatrix} e^{A(t-s)} & \tilde{\mathbf{m}}_{(s,t]} \\ \mathbf{0}_{1\times P} & [\mathbf{0}_{1\times I} \quad 1] \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} \mathbf{I}_P \\ \mathbf{0} \end{bmatrix}.$$

Moreover, since $\mu(t)$ is Brownian motion we know that:

$$p(\boldsymbol{\mu}_{V_{(s,t]}}|\mu(s)) = \mathcal{N}(\mu(s)\mathbf{1}_{(I+1)\times 1}, \mathbf{C}_\mu)$$

where

$$\mathbf{C}_\mu = \begin{bmatrix} (V_1-s)\sigma_\mu^2 & & (V_1-s)\sigma_\mu^2 & (V_1-s)\sigma_\mu^2 \\ & \ddots & & \\ (V_1-s)\sigma_\mu^2 & & (V_I-s)\sigma_\mu^2 & (V_I-s)\sigma_\mu^2 \\ (V_1-s)\sigma_\mu^2 & & (V_I-s)\sigma_\mu^2 & (t-s)\sigma_\mu^2 \end{bmatrix}. \tag{15}$$

Hence a further state update proceeding from $(X(s), \mu(s))$ to $(X(s), \boldsymbol{\mu}_{V_{(s,t]}})$ can be written as

$$\begin{bmatrix} X(s) \\ \boldsymbol{\mu}_{V_{(s,t]}} \end{bmatrix} = \mathbf{A}_1 \begin{bmatrix} X(s) \\ \mu(s) \end{bmatrix} + \mathbf{B_1}\mathbf{e}_{\delta t}, \mathbf{e_{\delta t}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mu) \tag{16}$$

where

$$\mathbf{A_1} = \begin{bmatrix} \mathbf{I}_P & 0 \\ \mathbf{0}_{(I+1)\times P} & \mathbf{1}_{(I+1)\times 1} \end{bmatrix}, \quad \mathbf{B_1} = \begin{bmatrix} \mathbf{0}_{P\times(I+1)} \\ \mathbf{I}_{I+1} \end{bmatrix}.$$

The two steps are now combined by substituting (16) into (14) to give

$$\begin{aligned} \begin{bmatrix} X(t) \\ \mu(t) \end{bmatrix} &= \mathbf{A}_2\mathbf{A}_1 \begin{bmatrix} X(s) \\ \mu(s) \end{bmatrix} + \mathbf{A}_2\mathbf{B}_1\mathbf{e}_{\delta t} + \mathbf{B}_2\mathbf{e}_t \\ &:= \mathbf{A} \begin{bmatrix} X(s) \\ \mu(s) \end{bmatrix} + \mathbf{B}\mathbf{e}_{s,t}, \ \mathbf{e}_{s,t} \sim \mathcal{N}(0, \mathbf{C}_{s,t}) \end{aligned}$$

and now the model is fully specified in the required form (10), since $\boldsymbol{\alpha}_s = \left[X(s), \, \mu(s)\right]^T$. After simplification, we get the desired result. $\square$

B. Monte Carlo filtering

For simplicity of notation, denote $\boldsymbol{\alpha}_{t_n}$ by $\boldsymbol{\alpha}_n$, $\{\Gamma_i\}^c_{(t_{n-1},t_n]}$ by $\{\Gamma_i\}_n$ and $\{V_i\}^c_{(t_{n-1},t_n]}$ by $\{V_i\}_n$, etc.. The ultimate inference goal ('filtering') is to estimate the extended state posterior distribution,

$$p(\boldsymbol{\alpha}_n|\mathbf{y}_{1:n}).$$

The sequential update however operates in the marginal space of the latent variables Γ and V as follows, for time step $n-1$ to n:

$$p(\{\Gamma_i, V_i\}_{1:n}|y_{1:n}) = \frac{p(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \{\Gamma_i, V_i\}_{1:n})p(\{\Gamma_i, V_i\}_n)}{p(\mathbf{y}_n|\mathbf{y}_{1:n-1})} p(\{\Gamma_i, V_i\}_{1:n-1}|\mathbf{y}_{1:n-1}) \tag{17}$$

Following a similar approach to [19], we first define a rescaled covariance matrix for the dynamical noise in the extended state space model (10) as

$$\tilde{\mathbf{C}}_{s,t} = \mathbf{C}_{s,t}/\sigma_W^2$$

and note that the observation noise variance is assumed to also be proportional to σ_W^2, see (11). In addition we assume $\sigma_\mu^2 = \kappa_\mu^2\sigma_W^2$, so that $\tilde{\mathbf{C}}_{s,t}$ does not depend on σ_W. The reason for making this structural assumption in the model is to attain tractability of the posterior distribution, which will then allow σ_W^2 to be marginalised out and estimated by Monte Carlo filtering. This assumption could be omitted at the expense of a more expensive Monte Carlo filter with one additional parameter.

The Kalman filter is utilised in order to compute the incremental weight terms in the marginal space, $p(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \{\Gamma_i, V_i\}_{1:n}, \sigma_W^2)$, and also the conditional posterior, $p(\boldsymbol{\alpha}_n|\{\Gamma_i, V_i\}_{1:n}, \mathbf{y}_{1:n}, \sigma_W^2)$. First, consider the posterior density. From time t_n to t_{n+1}, given

$$p(\boldsymbol{\alpha}_n|\{\Gamma_i, V_i\}_{1:n}, \mathbf{y}_{1:n}, \sigma_W^2) = \mathcal{N}(\boldsymbol{\mu}_{n|n}, \sigma_W^2\tilde{\mathbf{P}}_{n|n})$$

and the state space model specified in (10) and (11). In the prediction step, perform the standard Kalman prediction, expressed explicitly in terms of the rescaled covariance matrices,

$$p(\boldsymbol{\alpha}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n}, \sigma_W^2) = \mathcal{N}(\boldsymbol{\mu}_{n+1|n}, \sigma_W^2\tilde{\mathbf{P}}_{n+1|n})$$

where

$$\begin{aligned} \boldsymbol{\mu}_{n+1|n} &= \mathbf{A}\boldsymbol{\mu}_{n|n} \\ \tilde{\mathbf{P}}_{n+1|n} &= \mathbf{A}\tilde{\mathbf{P}}_{n|n}\mathbf{A}^T + \mathbf{B}\tilde{\mathbf{C}}_{n,n+1}\mathbf{B}^T \end{aligned} \tag{18}$$

then carry out the Kalman update step to obtain the conditional posterior and the incremental weight terms:

$$\begin{aligned} &p(\mathbf{y}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n}, \sigma_W^2) \\ &\quad = \mathcal{N}(\hat{\mathbf{y}}_{n+1|n}, \sigma_W^2\tilde{\mathbf{P}}^y_{n+1|n}) \end{aligned} \tag{19}$$

$$\begin{aligned} &p(\boldsymbol{\alpha}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1}, \sigma_W^2) \\ &\quad = \mathcal{N}(\boldsymbol{\mu}_{n+1|n+1}, \sigma_W^2\tilde{\mathbf{P}}_{n+1|n+1}) \end{aligned} \tag{20}$$

where

$$\begin{aligned} \hat{\mathbf{y}}_{n+1|n} &= \mathbf{H}\boldsymbol{\mu}_{n+1|n} \\ \tilde{\mathbf{P}}^y_{n+1|n} &= \mathbf{H}\tilde{\mathbf{P}}_{n+1|n}\mathbf{H}^T + \kappa_V^2\mathbf{I} \\ \mathbf{K} &= \tilde{\mathbf{P}}_{n+1|n}\mathbf{H}^T\{\tilde{\mathbf{P}}^y_{n+1|n}\}^{-1} \\ \boldsymbol{\mu}_{n+1|n+1} &= \boldsymbol{\mu}_{n+1|n} + \mathbf{K}(\mathbf{y}_{n+1} - \hat{\mathbf{y}}_{n+1|n}) \\ \tilde{\mathbf{P}}_{n+1|n+1} &= (\mathbf{I} - \mathbf{K}\mathbf{H})\tilde{\mathbf{P}}_{n+1|n} \end{aligned} \tag{21}$$

From (19) it is then possible to further marginalise σ_W^2 from the likelihood. This is as a result of the special rescaled form of the covariance matrices. For this to apply in closed form, set the conjugate prior [26], an Inverse Gamma (IG) distribution, $\sigma_W^2 \sim \mathcal{IG}(\alpha_W, \beta_W)$. Then, if at time t_{n+1}, $p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n}, \mathbf{y}_{1:n}) = \mathcal{IG}(\alpha_n, \beta_n)$, we can calculate $p(\mathbf{y}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n})$ by

$$\begin{aligned} &p(\mathbf{y}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n}) \\ &= \int p(y_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n}, \sigma_W^2)\times \\ &\qquad p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n}, \mathbf{y}_{1:n})d(\sigma_W^2) \\ &= \int \frac{1}{\sqrt{2\pi|\tilde{\mathbf{P}}^y_{n+1|n}|}} \left(\frac{1}{\sigma_W^2}\right)^{\alpha_n+1+D/2} \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \exp\{-\frac{\beta_{n+1}}{\sigma_W^2}\}d(\sigma_W^2) \\ &= \frac{1}{\sqrt{2\pi|\tilde{\mathbf{P}}^y_{n+1|n}|}} \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \frac{\Gamma(\alpha_n + D/2)}{\beta_{n+1}^{\alpha_n+1/2}} \end{aligned} \tag{22}$$

recall that $\mathbf{y}_n$ is of dimension D. Then, we can use this result to calculate $p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1})$

$$p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1}) = \mathcal{IG}(\alpha_{n+1}, \beta_{n+1})$$

where

$$\alpha_{n+1} = \alpha_n + D/2 \tag{23}$$

$$\beta_{n+1} = \beta_n + (\mathbf{y}_{n+1} - \hat{\mathbf{y}}_{n+1|n})^T (\tilde{\mathbf{P}}^y_{n+1|n})^{-1} (\mathbf{y}_{n+1} - \hat{\mathbf{y}}_{n+1|n})/2 \tag{24}$$

We can therefore use (23) and (24) to update recursively $p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1})$.

Marginalised Monte Carlo filtering can then be carried out in a fairly standard way, following the steps in [18]. Briefly, suppose the Monte Carlo filter has successfully sampled up to time n, i.e. we have M weighted random samples $\{\Gamma_i, V_i\}_{1:n}^{(m)}$ as follows:

$$p(\{\Gamma_i, V_i\}_{1:n}|\mathbf{y}_{1:n}) \approx \sum_{m=1}^{M} w_n^{(m)} \delta_{\{\Gamma_i, V_i\}_{1:n}^{(m)})}(\{\Gamma_i, V_i\}_{1:n})$$

We then draw a new set of random jumps for each sample according to Alg. 1,

$$\{\Gamma_i, V_i\}_{n+1}^{(m)} \overset{iid}{\sim} p(\{\Gamma_i, V_i\}_{n+1}),\ m = 1, ..., M$$

The update formula (17) then shows that the importance weights should be:

$$w_{n+1}^{(m)} \propto w_n^{(m)} p(\mathbf{y}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}^{(m)}, \mathbf{y}_{1:n}), \tag{25}$$

which have been computed for each sample using the Kalman filter in (22). Weights are normalised in the usual way such that $\sum_m w_{n+1}^{(m)} = 1$. Resampling of the Monte Carlo population may then be carried out in the standard way, as required.

The final so-called Rao-Blackwellised posterior distribution over all unknowns is then given by a mixture of Normal-IG distributions as follows:

$$p(\boldsymbol{\alpha}_{n+1}, \sigma_W^2|\mathbf{y}_{1:n+1}) \approx \sum_{m=1}^{M} w_{n+1}^{(m)} \mathcal{N}(\boldsymbol{\alpha}_{n+1}; \boldsymbol{\mu}_{n+1|n+1}^{(m)}, \sigma_W^2 \tilde{\mathbf{P}}_{n+1|n+1}^{(m)}) \times \mathcal{IG}(\sigma_W^2; \alpha_{n+1}^{(m)}, \beta_{n+1}^{(m)})$$

If required it is then straightforward to obtain marginal posteriors for σ_W^2 and $\boldsymbol{\alpha}_{n+1}$, as follows,

$$p(\sigma_W^2|\mathbf{y}_{1:n+1}) \approx \sum_{m=1}^{M} w_{n+1}^{(m)} \mathcal{IG}(\sigma_W^2; \alpha_{n+1}^{(m)}, \beta_{n+1}^{(m)}) \tag{26}$$

Then, to obtain a marginalised estimate of the states, we first calculate that for each particle

$$\begin{aligned} & p(\boldsymbol{\alpha}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1}) \\ &= \int_0^\infty p(\boldsymbol{\alpha}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}, \mathbf{y}_{1:n+1}, \sigma_W^2) \\ &\qquad p(\sigma_W^2|\{\Gamma_i, V_i\}_{1:n+1}^{(m)}, \mathbf{y}_{1:n+1}) d(\sigma_W^2) \\ &= \int_0^\infty \mathcal{IG}(\sigma_W^2; \alpha_{n+1}, \beta_{n+1}) \mathcal{N}(\boldsymbol{\alpha}_{n+1}; \boldsymbol{\mu}_{n+1}, \sigma_W^2 \tilde{\mathbf{P}}_{n+1}) d(\sigma_W^2) \\ &= t_{2\alpha_{n+1}}(\boldsymbol{\mu}_{n+1}, \frac{\beta_{n+1}}{\alpha_{n+1}} \tilde{\mathbf{P}}_{n+1}) \end{aligned}$$

where $t_v(\boldsymbol{\mu}, \mathbf{P})$ represents the Student distribution with degrees of freedom v, mean $\boldsymbol{\mu}$ and scale $\mathbf{P}$. Thus, the marginalised estimate of states is given by

$$p(\boldsymbol{\alpha}_{n+1}|\mathbf{y}_{1:n+1}) \approx \sum_{m=1}^{M} w_{n+1}^{(m)} p(\boldsymbol{\alpha}_{n+1}|\{\Gamma_i, V_i\}_{1:n+1}^{(m)}, \mathbf{y}_{1:n+1}) \tag{27}$$

Note that α_{n+1} is not to be confused with the state vector $\boldsymbol{\alpha}_{n+1}$.

The algorithm for a single iteration is shown in Alg. 3.

Algorithm 3: A single time iteration of the Marginalised PF

Input : $\{\boldsymbol{\mu}_n^{(m)}, \tilde{\mathbf{P}}_n^{(m)}, w_n^{(m)}\}_{m=1:M}$
Output: $\{\boldsymbol{\mu}_{n+1}^{(m)}, \tilde{\mathbf{P}}_{n+1}^{(m)}, w_{n+1}^{(m)}\}_{m=1:M}$
begin
  if *Resample* **then**
    Resample Particles by their weights
  for $m = 1 \ldots M$ **do**
    Sample $\{\{\Gamma, V\}_j\}^c_{(t_n, t_{n+1}]}$ by Alg. 1;
    (10);
    Carry out prediction step as in (18) using (10);
    Do the update step as in (21), update weights by (25);
    Store $\boldsymbol{\mu}_{n+1}^{(m)}, \tilde{\mathbf{P}}_{n+1}^{(m)}, w_{n+1}^{(m)}$;

V. Experiments

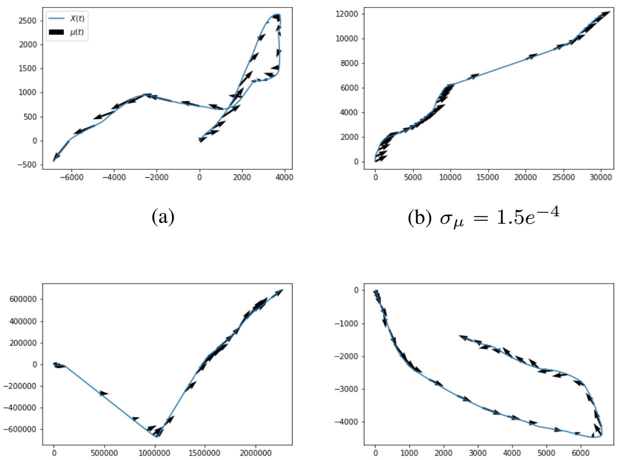
Initial evaluations are carried out for the α-stable processes. Other forms of NVM and NσM models will be reported in future publications. Recall that for α-stable process, $h(\gamma) = \gamma^{-1/\alpha}$ so we can rewrite (5) as

$$\mathbf{I}^c(s,t) \overset{d}{=} \sum_{\Gamma_i < c} \left[\Gamma_i^{-1/\alpha} \mathbf{f}_t(V_i)(\mu(V_i) + \sigma_W U_i)\right]$$

and (8) (9) also change accordingly.

As for the SDE, we choose a 2-D Langevin model, so we have in (1),

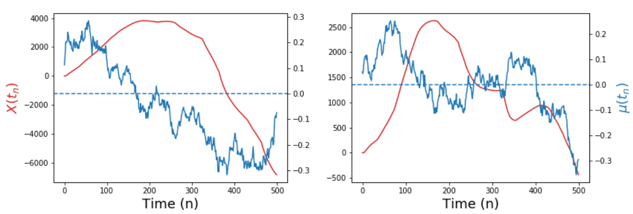
$$A = \begin{bmatrix} 0 & 1 \\ 0 & \lambda \end{bmatrix}, \mathbf{h} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$



(a) (b) $\sigma_\mu = 1.5e^{-4}$

(c) $\alpha = 0.5$ (d) $\mu_0 = [0.05, -0.1]$

Fig. 1: Simulated trajectories $X(t)$ (blue) plotted with vector $\boldsymbol{\mu}(t)$ (black arrows) for Langevin model driven by α-stable Lévy noise, parameters: (a) $\alpha = 0.9$, $\lambda = 0.05$, $\mu_0 = [0.1, 0.05]$, $\sigma_\mu = 0.015$, $N = 500$, $c = 10$; (b) $\sigma_\mu = 1.5e^{-4}$, otherwise as (a); (c) $\alpha = 0.5$, otherwise (a); (d) $\mu_0 = [0.05, -0.1]$, others as (a).



(a) Dimension 1 (b) Dimension 2

Fig. 2: Simulated trajectories $X(t)$ (red) plotted with $\mu(t)$ (blue), dashed line corresponds to zero μ, parameters used are $\alpha = 0.9$, $\lambda = 0.05$, $\mu_0 = [0.1, 0.05]$, $\sigma_\mu = 0.015$, $c = 10$.

Examples of trajectories (blue curve) and vector $\boldsymbol{\mu}(t)$ (black arrows) are shown in Fig. 1. The trajectories are aimed at emulating real data where we observe both sharp changes in direction caused by the heavy-tailed noise distribution and overall changes in direction over time, modelled by the time-varying skewness, included as a proxy for the time-varying directional intent of the object. We have used $\alpha = 0.9$, $\lambda = 0.05$, $\mu(0) = [0.1, 0.05]$, $\sigma_\mu = 0.015$ for both dimensions, $c = 10$ in Fig. 1a. The remaining panels use different parameters: Fig. 1b uses a small σ_μ which causes $\mu(t) \approx \mu(0)$ so the direction of the movement hardly changes. Because of the positive starting value for skewness, the trajectory moves towards the upper right all the time. The effect of $\mu(0)$ can be seen by comparing Fig. 1b with Fig. 1d. In Fig. 1d, $\mu(0)$ for horizontal movement is positive and negative for the transverse movement, resulting in the trend towards the lower right at the start. Smooth and small changes of direction can be observed in these 3 trajectories, while Fig. 1c shows the opposite. Because as α decreases, the α-stable distribution becomes heavier-tailed, the velocity is more likely to jump to extreme values. Hence the trajectory is less smooth with variations in a larger scale. Comparing to Fig. 1b where the adjustments of velocity are in a small scale, here the changes of velocity are more abrupt.

By comparing the blue curves with the arrows, the effect of time-varying skewness on the 'intent' of the object can be seen. The movement of $X(t)$ loosely aligns with the direction of $\mu(t)$.

More detailed plots of $\mu(t)$ and $X(t)$ of Fig. 1a for each dimension are shown in Fig. 2, where the red and blue curves are $X(t)$ and $\mu(t)$ respectively. For the y-axis movement, looking at Fig. 2b, when $n < 150$, $\mu(t)$ is positive most of the time, which causes an increasing trend of X. Then μ suddenly drops below zero and stays negative, which corresponds to the small drop in $X(t)$. Then, as μ fluctuates around zero after $n \approx 300$, the curve also fluctuates. So, it is clear that the sign of μ corresponds to a tendency for increase/decrease in $X(t)$. When combined with $\beta(t)$ values for other coordinate dimensions, this will lead to preferential motions in particular spatial directions, which change over time. This is because individual jumps will on average be biased in the direction of the spatial $\mu(t)$ values.

We do note however that the relationship between intent and the direction of $\mu(t)$ may become a little more complex than indicated here, since in cases with $\alpha \in (1, 2)$ (not currently included in the models) a compensator term will also affect the 'intent' of the object. This case ($\alpha \in (1, 2)$) and the relationship with 'intent' will be explored more fully in future work, for example by the inclusion of additional unknown drift terms in the underlying Lévy process.

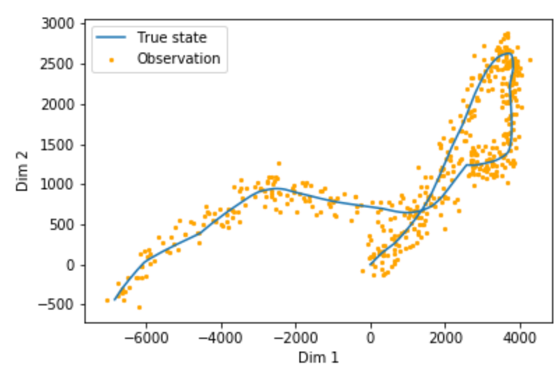


Fig. 3: Trajectory (blue) and noisy observations (orange) for Langevin model, $\kappa_V = 2800$, $\alpha = 0.9$, $\lambda = 0.05$, $\mu_0 = [0.1, 0.05]$, $\sigma_\mu = 0.015$, $c = 10$.

Finally, we illustrate marginalised particle filtering for the dynamic states and μ from the noisy data, as shown in Fig. 3, where $\sigma_W = 0.05$, standard deviation of the noise $\sigma_V = \kappa_V \times \sigma_W, \kappa_V = 2800$. In the particle filter, 300 particles are used. Filtered estimation of $\mu(t)$ and $X(t)$ with 3-standard deviation regions are shown in Fig. 4. The estimated values are the mean of the mixture distributions in (26) and (27). The standard deviation is obtained by calculating variance of the mixture distributions. We can see that the estimated $\mu(t)$ and $X(t)$ in each dimension can track the true latent value quite accurately, with suitable confidence intervals. Fig. 5 shows the estimated $X(t)$ with vector $\boldsymbol{\mu}(t)$ overlaid. From this figure, we can again see that the estimated μ is aligned well with the typical direction of motion of X. The estimated posterior distribution $p(\sigma_W^2|\mathbf{y}_{1:N})$ is shown in Fig. 6, where the true value falls in the region with high probability mass.

VI. Conclusion

We have presented a framework for modelling and inference of tracking models with time-varying skew. Skewness is modelled as a Brownian motion process driving a shot-noise representation of the stochastic process. Applications are presented in the case of α-stable processes, demonstrating how such a model can represent the change of an object's directional intent over time. On test data the methods perform well even with a small number of particles and noisy observations. Such methods can also easily extend to other processes that can be represented by generalised shot noise processes, such as the NVM processes, where we can simply change the form of $H(\Gamma, U)$ and the rest follows as here. Further work is addressing theoretical issues of how to compensate the time-varying case in the α-stable model with $1 \leqslant \alpha < 2$ and how to approximate the residual due to truncation, both of which were possible in the fixed β case. The current simulations can only be viewed as a proof of principle on limited data. Future work will present intent analysis for real object motions, as well as incorporation of other key elements such as clutter measurements and data associations. Finally, other applications will be explored such as animal behaviour tracking and financial data prediction.

References

[1] R. Gan, B. I. Ahmad, and S. J. Godsill, "Lévy state-space models for tracking and intent prediction of highly maneuverable objects," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 4, 2021.

[2] Q. Li, R. Gan, J. Liang, and S. J. Godsill, "An Adaptive and Scalable Multi-Object Tracker Based on the Non-Homogeneous Poisson Process," *IEEE Transactions on Signal Processing*, vol. 71, pp. 105–120, 2023.

[3] B. Mandelbrot, "New methods in statistical economics," *Journal of Political Economy*, vol. 71, no. 5, pp. pp. 421–440, 1963.

[4] E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, vol. 38, no. 1, pp. pp. 34–105, 1965.

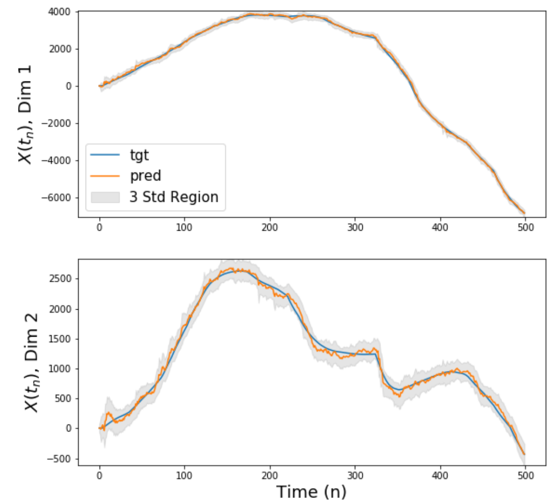
[5] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, 2003.

[6] N. Azzaoui and L. Clavier, "Statistical channel model based on α-stable random processes and application to the 60 GHz ultra wide band channel," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1457–1467, May 2010.

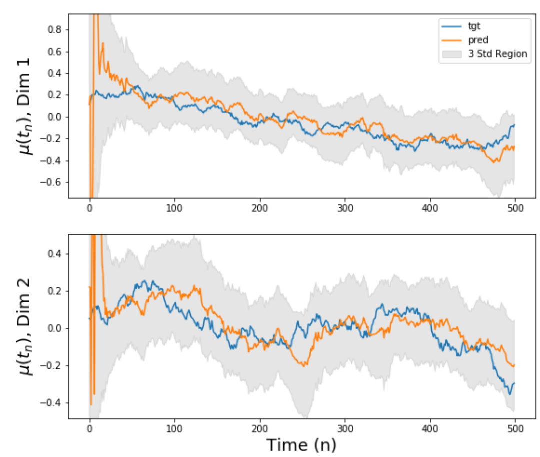
[7] J. Fahs and I. Abou-Faycal, "On the capacity of additive white alpha-stable noise channels," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 294–298.

[8] M. de Freitas, M. Egan, L. Clavier, A. Goupil, G. Peters, and N. Azzaoui, "Capacity Bounds for Additive Symmetric α-Stable Noise Channels," *IEEE Transactions on Information Theory*, vol. 63, pp. 5115–5123, 2017.

[9] D. J. Warren and J. B. Thomas, "Asymptotically robust detection and estimation for very heavy-tailed noise," *IEEE Transactions on Information Theory*, vol. 37, no. 3, pp. 475–481, May 1991.



(a) $X(t)$



(b) $\mu(t)$

Fig. 4: Filtered estimates for both states and μ from the particle filter for a Langevin model driven by the α-stable Lévy noise, using 300 particles, $\alpha = 0.9$, $\lambda = 0.05$, $\mu(0) = [0.1, 0.05]$, $\sigma_\mu = 0.015$, $c = 10$. Blue curves show the mean and grey region is the three-standard deviation region, calculated from the mixture distribution in (27).

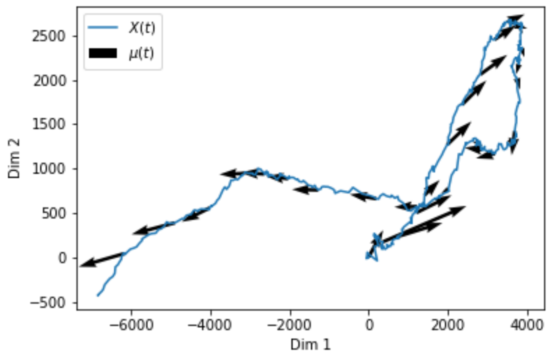


Fig. 5: Filtered Estimates $X(t)$ (blue) and vector $\boldsymbol{\mu}(t)$ (black arrows) from the particle filter, i.e. mean of the mixture Student distributions in (27).

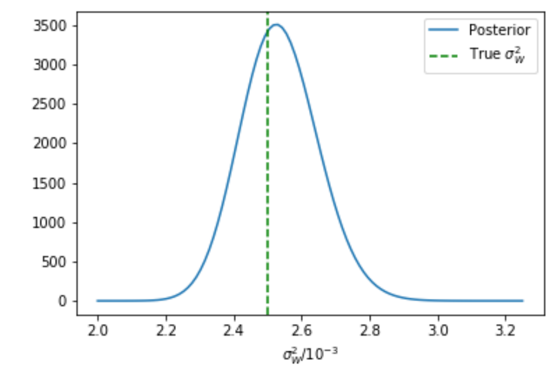


Fig. 6: Posterior of σ_W^2 in (26) from the particle filter.

[10] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*, ser. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, 1995.

[11] A. Achim, E. E. Kuruoǧlu and J. Zerubia, "SAR image filtering based on the heavy-tailed Rayleigh model," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2686–2693, 2006.

[12] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 352–372, 1998.

[13] M. J. Lombardi and S. J. Godsill, "On-line Bayesian estimation of signals in symmetric α-stable noise," *Signal Processing, IEEE Transactions on*, vol. 54, no. 2, pp. 775–779, 2006.

[14] R. W. Katz and B. G. Brown, "Extreme events in a changing climate: Variability is more important than averages," *Climatic Change*, vol. 21, no. 3, pp. 289–302, 1992.

[15] R. W. Katz, M. B. Parlange, and P. Naveau, "Statistics of extremes in hydrology," *Advances in Water Resources*, vol. 25, no. 8–12, pp. 1287 – 1304, 2002.

[16] J. Bertoin, *Lévy Processes*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1996. [Online]. Available: https://books.google.co.uk/books?id=A2CMMQEACAAJ

[17] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[18] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[19] S. Godsill, M. Riabiz, and I. Kontoyiannis, "The Lévy state space model," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 487–494.

[20] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications (Universitext)*, 6th ed. Springer, Jan. 2014.

[21] J. Rosiński, "Series representations of Lévy processes from the perspective of point processes," in *Lévy Processes*, O. Barndorff-Nielsen, S. Resnick, and T. Mikosch, Eds. Birkhauser Boston, 2001, pp. 401–415.

[22] G. Samorodnitsky, M. S. Taqqu, and R. Linde, "Stable non-gaussian random processes: stochastic models with infinite variance," *Bulletin of the London Mathematical Society*, vol. 28, no. 134, pp. 554–555, 1996.

[23] M. T. Costa, I. Kontoyiannis, and S. Godsill, "Generalised shot noise representations of stochastic systems driven by non-gaussian lévy processes," *J. Applied Probabity (In press)*, 2024.

[24] G. Samorodnitsky and M. L. Taqqu, *Non-Gaussian Stable Processes*. Chapman and Hall, 1994.

[25] J. F. C. Kingman, *Poisson processes*, ser. Oxford Studies in Probability. New York: The Clarendon Press Oxford University Press, 1993, vol. 3, oxford Science Publications.

[26] J. M. Bernardo and A. F. Smith, *Bayesian theory*. John Wiley & Sons, 2009, vol. 405.