Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Multivariate Analysis 94 (2005) 382–400

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

A skewed Kalman filter

Philippe Naveau[a,*], Marc G. Genton[b], Xilin Shen[c]

[a] *Department of Applied Mathematics, Colorado University at Boulder, Boulder, CO 80309-0526, USA*
[b] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*
[c] *Department of Electrical Engineering, Colorado University at Boulder, Boulder, CO 80309-0526, USA*

Received 15 April 2002
Available online 6 August 2004

Abstract

The popularity of state-space models comes from their flexibilities and the large variety of applications they have been applied to. For multivariate cases, the assumption of normality is very prevalent in the research on Kalman filters. To increase the applicability of the Kalman filter to a wider range of distributions, we propose a new way to introduce skewness to state-space models without losing the computational advantages of the Kalman filter operations. The skewness comes from the extension of the multivariate normal distribution to the closed skew-normal distribution. To illustrate the applicability of such an extension, we present two specific state-space models for which the Kalman filtering operations are carefully described.

AMS 1991 subject classification: 62F99; 62H10

Keywords: State-space model; Closed skew-normal distribution

1. Introduction

The overwhelming assumption of normality in the Kalman filter literature can be understood for many reasons. A major one is that the multivariate distribution is completely characterized by its first two moments. In addition, the stability of the multivariate normal distribution under summation and conditioning offers tractability and simplicity. Therefore, the Kalman filter operations can be performed rapidly and efficiently whenever the

* Corresponding author.
E-mail addresses: naveau@colorado.edu (P. Naveau), genton@stat.ncsu.edu (M.G. Genton), xilin.shen@colorado.edu (Xilin Shen).

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.06.002

normality assumption holds. However, this assumption is not satisfied for a large number of applications. For example, some distributions used in a state-space model can be skewed. In this work, we propose a novel extension of the Kalman filter by working with a larger class of distributions than the normal distribution. This class is called *closed skew-normal distributions*. Besides introducing skewness to the normal distribution, it has the advantages of being closed under marginalization and conditioning. This class has been introduced by González-Farías et al. [9] and is an extension of the multivariate skew-normal distribution first proposed by Azzalini and his coworkers [1–**?**4]. These distributions are particular types of generalized skew-elliptical distributions recently introduced by Genton and Loperfido [8], i.e. they are defined as the product of a multivariate elliptical density with a skewing function.

This paper is organized as follows. In Section 2, we recall the definition of the closed skew-normal distribution and the basic framework of state-space and Kalman filtering. Section 3 presents the conditions under which the observation and state vectors of the state-space model follow closed skew-normal distributions. In Section 4, a sequential procedure based on the Kalman filter is proposed to estimate the parameters of such distributions. A simulated example illustrates the differences between the classical Kalman filter and our non-linear skewed Kalman filter. We discuss our strategy relative to other Kalman filters and conclude in Section 5.

2. Definitions and notations

2.1. The closed skew-normal distribution

The closed skew-normal distribution is a family of distributions including the normal one, but with extra parameters to regulate skewness. It allows for a continuous variation from normality to non-normality, which is useful in many situations, see e.g. Azzalini and Capitanio [4] who emphasized statistical applications for the skew-normal distribution.

An n-dimensional random vector X is said to have a multivariate closed skew-normal distribution [9,10], denoted by $CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$, if it has a density function of the form

$$\frac{1}{\Phi_m(0; \nu, \Delta + D\Sigma D^T)} \phi_n(x; \mu, \Sigma)\Phi_m(D(x-\mu); \nu, \Delta), \quad x \in \mathbb{R}^n, \tag{1}$$

where $\mu \in \mathbb{R}^n$, $\nu \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{n\times n}$ and $\Delta \in \mathbb{R}^{m\times m}$ are both covariance matrices, $D \in \mathbb{R}^{m\times n}$, $\phi_n(x; \mu, \Sigma)$ and $\Phi_n(x; \mu, \Sigma)$ are the n-dimensional normal pdf and cdf with mean μ and covariance matrix Σ. When $D = 0$, the density (1) reduces to the multivariate normal one, whereas it reduces to Azzalini and Capitanio's [4] density when $m = 1$ and $\nu = D\mu$. The matrix parameter D is referred to as a "shape parameter". The moment generating function $M_{n,m}(t)$ for a *CSN* distribution is given by

$$M_{n,m}(t) = \frac{\Phi_m(D\Sigma t; \nu, \Delta + D\Sigma D^T)}{\Phi_m(0; \nu, \Delta + D\Sigma D^T)} \exp\left\{\mu^T t + \frac{1}{2}(t^T \Sigma t)\right\} \tag{2}$$

for any $t \in \mathbb{R}^n$. This expression of the moment generating function is important to understand the closure properties of the *CSN* distribution for summation. It is straightforward to see that the sum of two *CSN* of dimension (n, m) is not another *CSN* of dimension (n, m). Despite this limitation, it is possible to show that the sum of two *CSN* of dimension (n, m) is a *CSN* of dimension $(n, 2m)$ [10]. Hence, the *CSN* is closed under summation whenever the dimension *m* is allowed to vary. Although important for specific applications (adding a relative small number of variables), this closure property is not appropriate when dealing with state space models. These models are based on sequential transformations from time $t - 1$ to time *t*. Implementing a sum at each time step rapidly increases the dimension *m* and the sizes of the matrix Δ and *D* quickly become unmanageable. For this reason, we will propose two new and different ways of introducing skewness without paying this dimensionality cost.

The three basic tools when implementing the Kalman filter are the closure under linear transformation, under summation and conditioning. In Section 3, we will present how the general skew-normal distribution behaves under such constraints.

2.2. The state-space model and the Kalman filter

The *state-space model* has been widely studied (e.g. [12,13,17,18,20]). This model has become a powerful tool for modeling and forecasting dynamical systems and it has been used in a wide range of disciplines such as biology, economics, engineerings and statistics [11,14]. The basic idea of the state-space model is that the *d*-dimensional vector of observation Y_t at time *t* is generated by two equations, the *observational* and the *system* equations. The first equation describes how the observations vary in function of the unobserved state vector X_t of length *h*

$$Y_t = F_t X_t + \varepsilon_t, \tag{3}$$

where ε_t represent an added noise and F_t is a $d \times h$ matrix of scalars. The essential difference between the state-space model and the conventional linear model is that the state vector X_t is not assumed to be constant but may change in time. The temporal dynamical structure is incorporated via the system equation

$$X_t = G_t X_{t-1} + \eta_t, \tag{4}$$

where η_t represents an added noise and G_t is an $h \times h$ matrix of scalars. There exists a long literature about the estimation of the parameters for such models. In particular, the Kalman filter provides an optimal way to estimate the model parameters if the assumption of gaussianity holds. Following the definition by Meinhold and Singpurwalla [15], the term "Kalman filter" used in this work refers to a recursive procedure for inference about the state vector. To simplify the exposition, we assume that the observation errors ε_t are independent of the state errors η_t and that the sampling is equally spaced, $t = 1, \ldots, n$. The results shown in this paper could be easily extended without such constraints. But, the loss of clarity in the notations would make this work more difficult to read without bringing any new important concepts.

3. Kalman filtering and closed skew-normal distributions

In order to obtain the closure under summation needed for the Kalman filtering, two options will be investigated in this work. The first one, that will be exposed in Section 3.1, is to determine under which conditions the observations and state vector follow closed skew-normal distributions. This question can be rewritten as: what kind of noise in Eqs. (3) and (4) should be added to a closed skew-normal distribution in order that the sum remains a closed skew-normal distribution? The second strategy is to extend the linear state-space model to a wider state-space model for which the stability under summation is better preserved. This approach will be described in Section 3.2.

In order to pursue our goals, we need the two following lemmas. The first one describes the stability of the closed skew-normal distribution under scalar transformation. For completeness, the proof of this lemma originally derived by González-Farías et al. [9] can be found in the appendix.

Lemma 1. *Let Y be a random vector with a closed skew-normal distribution $CSN_{n,m}(\mu, \Sigma, D, v, \Delta)$ and A an $r \times n$ matrix such that $A^T A$ is non-singular. If the random vector X is defined as the linear transform AY, then it also follows a closed skew-normal distribution,*

$$X = AY \sim CSN_{r,m}(A\mu, A\Sigma A^T, DA^{\leftarrow}, v, \Delta),$$

where $A^{\leftarrow}$ is the left inverse of A and $A^{\leftarrow} = A^{-1}$ when A is a $n \times n$ non-singular matrix.

The second lemma states that adding a Gaussian noise to a closed multivariate skew-normal vector of dimension (n, m) does not change the distribution class, i.e. the result is still a closed skew-normal vector of dimension (n, m) . In this paper, the proofs of our lemmas and propositions are presented in the appendix when needed.

Lemma 2. *Let X be a random vector with a closed skew-normal distribution $CSN_{n,m}(\psi, \Omega, D, v, \Delta)$ and Z be an n-dimensional Gaussian random vector with mean μ, covariance matrix Σ, and independent of X. Then, the sum $X + Z$ follows a closed skew-normal distribution*

$$CSN_{n,m}(\mu_{X+Z}, \Sigma_{X+Z}, D_{X+Z}, v_{X+Z}, \Delta_{X+Z}),$$

where $\mu_{X+Z} = \psi + \mu,\ \Sigma_{X+Z} = \Omega + \Sigma,\ D_{X+Z} = D\Omega(\Omega + \Sigma)^{-1},$

$$v_{X+Z} = v \quad \textit{and} \quad \Delta_{X+Z} = \Delta + (D - D_{X+Z})\Omega D^T.$$

3.1. Distribution of the state-space model variables

A direct application of Lemmas 2 and 1 allows us to derive the first proposition of this work.

Proposition 3. *Suppose that the initial state vector X_0 of the system composed by Eqs.* (3) *and* (4) *follows a closed skew-normal distribution, $CSN_{n,m}(\psi_0, \Omega_0, D_0, v_0, \Delta_0)$. If the noise ε_t,* respectively, *η_t, is an i.i.d. Gaussian vector with mean μ_ε and covariance*

Σ_ε, respectively, μ_η and Σ_η, then both the state vector X_t and the observation vector Y_t follow a closed skew-normal distribution, $X_t \sim CSN_{h,m}(\psi_t, \Omega_t, D_t, v_t, \Delta_t)$ and $Y_t \sim CSN_{d,m}(\mu_t, \Gamma_t, E_t, \gamma_t, \Theta_t)$. The parameters of these distributions satisfy the following relationships for $t = 1, 2, \ldots$,

$$\psi_t = G_t\psi_{t-1} + \mu_\eta, \quad \mu_t = F_t\psi_t + \mu_\varepsilon, \tag{5}$$

$$\Omega_t = G_t\Omega_{t-1}G_t^T + \Sigma_\eta, \quad \Gamma_t = F_t\Omega_t F_t^T + \Sigma_\varepsilon, \tag{6}$$

$$D_t = D_{t-1}\Omega_{t-1}G_t^T\Omega_t^{-1}, \quad E_t = D_t\Omega_t F_t^T\Gamma_t^{-1}, \tag{7}$$

$$v_t = v_{t-1}, \quad \gamma_t = v_t, \tag{8}$$

$$\Delta_t = \Delta_{t-1} + (D_{t-1} - D_t G_t)\Omega_t D_{t-1}^T, \tag{9}$$

and

$$\Theta_t = \Delta_t + (D_t - E_t F_t)\Omega_t D_t^T, \tag{10}$$

whenever $G_t^T G_t$ and $F_t^T F_t$ are non-singular matrices.

This proposition shows that the initial state X_0 and the Gaussian noises are the two key elements to obtain state and observation vectors with closed skew-normal distributions. Besides providing some fundamental relationships, this proposition is a good starting point to discuss some of the difficulties associated with skewness in the classical linear state-space model. In particular, the skewness in the observation vector would be better propagated in time if it was implemented, not exclusively with X_0, but at each time step. To develop a model with such a capability, we choose to extend the linear state-space model framework. In the next section, we will present the details of such an approach.

3.2. Extension of the linear state-space model

Our strategy to derive a model with a more flexible skewness is to directly incorporate a term for skewness into the observation equation

$$\begin{aligned} Y_t &= F_t X_t + \varepsilon_t \\ &= Q_t U_t + P_t S_t + \varepsilon_t, \quad \text{with } F_t = (Q_t, P_t) \quad \text{and} \quad X_t = (U_t^T, S_t^T)^T, \end{aligned} \tag{11}$$

where the random vector U_t of length k and the $d \times k$ matrix of scalar Q_t represent the linear part of the observation equation. In comparison, the random vector S_t of length l and the $d \times l$ matrix of scalar P_t correspond to the additional skewness. The most difficult task in this construction is to propose a simple dynamical structure of the skewness vector S_t and the "linear" vector U_t while keeping the independence between these two vectors (the last condition is not theoretically necessary but it is useful when interpreting the parameters). To reach this goal, we suppose that the bi-variate random vector $(U_t^T, V_t^T)^T$ is generated from a linear system

$$\begin{cases} U_t = K_t U_{t-1} + \eta_t^*, \\ V_t = -L_t V_{t-1} + \eta_t^+, \end{cases} \tag{12}$$

where the Gaussian noise $\eta_t^* \sim N_k(\mu_\eta^*, \Sigma_\eta^*)$ is independent of $\eta_t^+ \sim N_l(\mu_\eta^+, \Sigma_\eta^+)$ and where K_t, respectively, L_t represents a $k \times k$ matrix of scalars, respectively, a $l \times l$ matrix of scalars.

To continue our construction of the system, a few notations and a lemma are needed. The multivariate normal distribution of the vector $(U_t^T, V_t^T)^T$ is denoted by

$$\begin{pmatrix} U_t \\ V_t \end{pmatrix} \sim N_{k+l}\left(\begin{pmatrix} \psi_t^* \\ \psi_t^+ \end{pmatrix}, \begin{pmatrix} \Omega_t^* & 0 \\ 0 & \Omega_t^+ \end{pmatrix} \right). \tag{13}$$

The parameters of such vectors can be sequentially derived from an initial vector $(U_0^T, V_0^T)^T$ with a normal distribution.

Lemma 4. *Let $D_t^+ = \Omega_{t-1}^+ L_t^T (\Omega_t^+)^{-1}$, $\psi_t^+ = -L_t \psi_{t-1}^+ + \mu_\eta^+$, and $\Phi_t(\cdot) = \Phi_l(\cdot, \psi_t^+, \Omega_t^+)$. The skewness part S_t of the state vector $X_t = (U_t^T, S_t^T)^T$ is defined as*

$$S_t = \eta_t^+ - L_t W_{t-1}, \tag{14}$$

where the vector W_{t-1} is defined as follows:

If $D_t^+ \psi_t^+ \leqslant \psi_{t-1}^+$, then

$$W_{t-1} = \begin{cases} V_{t-1} & \text{if } V_{t-1} \leqslant D_t^+ \psi_t^+, \\ 2\psi_{t-1}^+ - V_{t-1} & \text{if } V_{t-1} \geqslant 2\psi_{t-1}^+ - D_t^+ \psi_t^+, \\ \Phi_{t-1}^{-1}\Bigg(\Phi_{t-1}(D_t^+\psi_t^+) & \\ \quad \times \frac{\Phi_{t-1}(V_{t-1}) - \Phi_{t-1}(D_t^+\psi_t^+)}{\Phi_{t-1}(2\psi_{t-1}^+ - D_t^+\psi_t^+) - \Phi_{t-1}(D_t^+\psi_t^+)} \Bigg) & \text{otherwise}, \end{cases} \tag{15}$$

$$W_{t-1} = \begin{cases} V_{t-1} & \text{if } V_{t-1} \leqslant D_t^+ \psi_t^+, \\ \Phi_{t-1}^{-1}\Big(\frac{\Phi_{t-1}(D_t^+\psi_t^+)}{1 - \Phi_{t-1}(D_t^+\psi_t^+)} (\Phi_{t-1}(V_{t-1}) & \\ \quad - \Phi_{t-1}(D_t^+\psi_t^+)) \Big) & \text{otherwise.} \end{cases} \tag{16}$$

With these definitions, the variable S_t follows a closed skew-normal distribution $S_t \sim CSN_{l,1}$ $(\psi_t^+, \Omega_t^+, D_t^+, \nu_t^+, \Delta_t^+)$, where we have $\nu_t^+ = \psi_{t-1}^+ - D_t^+\psi_t^+$, $\Delta_t^+ = \Omega_{t-1}^+ - D_t^+ \Omega_t^+ (D_t^+)^T$, and $\Omega_t^+ = L_t \Omega_{t-1}^+ L_t^T + \Sigma_\eta^+$.

Although Lemma 4 may look complex, it is easy to show that W_{t-1} has the same distribution than $[V_{t-1} \mid V_{t-1} \leqslant D_t^+ \psi_t^+]$ (see the proof of the lemma). It follows from (12) that the vector S_t defined from Eq. (14) has the same distribution than $[V_t \mid V_{t-1} \leqslant D_t^+ \psi_t^+]$. The former variable is usually used as a more classical definition of skew-normal vector [9]. In the context of time series analysis, the reason for defining W_t with (15) and (16) instead of using the simpler definition $W_{t-1} = [V_{t-1} \mid V_{t-1} \leqslant D_t^+ \psi_t^+]$ is that the latter one is not practical to generate simulated realizations of W_t. To illustrate this difficulty, suppose that v_{t-1} is one realization of V_{t-1} that does not satisfy $E_t = \{v_{t-1} \leqslant D_t^+ \psi_t^+\}$. In this situation, we have to re-simulate other realizations of V_{t-1} until E_t is true. A classical accept–reject algorithm can be time consuming if E_t occurs rarely (which is the case if a large amount of skewness is introduced). In comparison, defining W_{t-1} through (15) and (16) bypasses this computational obstacle. In this case, there is no need to simulate other realizations because E_t is always satisfied with the construction proposed in Lemma 4. The technique implemented to generate directly W_{t-1} is based on a folding construction that has been studied by Corcoran and Schneider [6]. The proof of Lemma 4 gives the details of such a folding. From a theoretical point of view, the reader only needs to keep in mind that S_t is stochastically equivalent to $[V_t \mid V_{t-1}D_t^+ \leqslant \psi_t^+]$.

From Lemma 4, we deduce that the state vector has also a closed skew-normal distribution

$$X_t = \begin{pmatrix} U_t \\ S_t \end{pmatrix} \sim CSN_{k+l,k+1}(\psi_t, \Omega_t, D_t, v_t, \Delta_t) \quad \text{with } \psi_t = \begin{pmatrix} \psi_t^* \\ \psi_t^+ \end{pmatrix} \tag{17}$$

and

$$\Omega_t = \begin{pmatrix} \Omega_t^* & 0 \\ 0 & \Omega_t^+ \end{pmatrix}, \quad D_t = \begin{pmatrix} 0 & 0 \\ 0 & D_t^+ \end{pmatrix}, \quad v_t = \begin{pmatrix} 0 \\ v_t^+ \end{pmatrix}, \quad \Delta_t = \begin{pmatrix} I & 0 \\ 0 & \Delta_t^+ \end{pmatrix}.$$

Hence, the variable S_t through the matrix L_t introduces at each time step a different skewness(if needed) in the state vector whose temporal structure is defined by V_t in (12). The price for this gain in skewness flexibility is that this state vector (because of (15) and (16)) does not have anymore a linear structure like the one defined by (4).

To illustrate the distribution of the skewness vector S_t, two histograms of S_t are plotted at two different instants $t = 0$ (no skewness, see left panel) and $t = 40$ (large skewness, see right panel) (Fig. 1). These simulated data were generated by setting $F_t = P_t = (-1)^t/2$, $\mu_\varepsilon = 0$, $\mu_\eta^+ = 2$, $Q_t = K_t = \Sigma_\eta^* = \mu_\eta^* = 0$, and $\Sigma_\varepsilon = \Sigma_\eta^+ = 1$. The other parameters were set according to Fig. 2 that describes the temporal evolution of L_t, ψ_t, Ω_t, D_t, v_t and Δ_t for this simulation. A more detailed discussion of this example will be presented in Section 5 (Discussions and conclusions).

The next proposition summarizes our findings and can be seen as a more general result than Proposition 3 (if $P_t = 0$ or $L_t = 0$ then the classical state-space model is obtained).

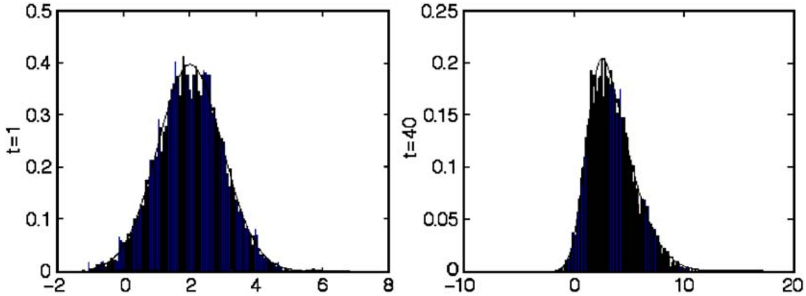

Fig. 1. Density of S_t with histograms from simulated values. The left panel corresponds to the initial time, $t = 0$, (no skewness) and the right panel to the time $t = 40$ for the parameters described in Fig. 2.

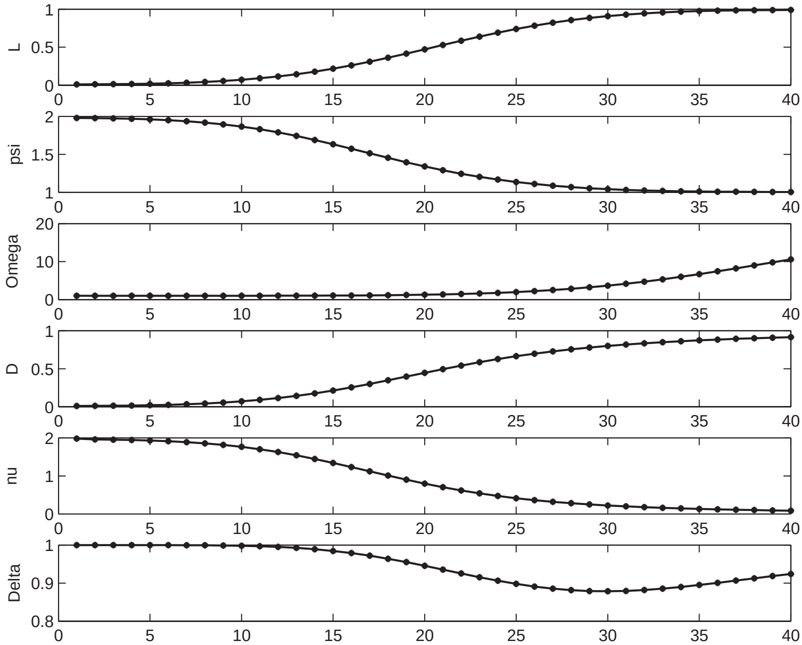

Fig. 2. Temporal evolution of the parameters used to simulate S_t in Fig. 1. We set $\mu_\varepsilon = \mu_\eta^+ = 2$, $Q_t = K_t = \Sigma_t^* = \mu_\eta^* = 0$, and $\Sigma_\varepsilon = \Sigma_\eta^* = 1$.

Proposition 5. *Suppose that the initial vector* $(U_0^T, V_0^T)^T$ *of the linear system defined by* (12) *follows the normal distribution defined by*

$$\begin{pmatrix} U_0 \\ V_0 \end{pmatrix} \sim N_{k+l}\left(\begin{pmatrix} \psi_0^* \\ \psi_0^+ \end{pmatrix}, \begin{pmatrix} \Omega_0^* & 0 \\ 0 & \Omega_0^+ \end{pmatrix} \right). \tag{18}$$

Then both the state vector $X_t = (U_t^T, S_t^T)^T$ *and the observation vector* Y_t *of the non-linear state-space model defined by Eqs.* (11), (12), *and* (14) *follow closed skew-normal distributions,* $X_t \sim CSN_{h,m}(\psi_t, \Omega_t, D_t, v_t, \Delta_t)$ *and* $Y_t \sim CSN_{d,m}(\mu_t, \Gamma_t, E_t, \gamma_t, \Theta_t)$ *for* $t \geqslant 1$. *The parameters of these distributions satisfy the following relationships*:

$$\psi_t^* = K_t \psi_{t-1}^* + \mu_\eta^*, \quad \psi_t^+ = -L_t \psi_{t-1}^+ + \mu_\eta^+, \quad \mu_t = F_t \psi_t + \mu_\varepsilon,$$

$$\Omega_t^* = K_t \Omega_{t-1}^* K_t^T + \Sigma_\eta^*, \quad \Omega_t^+ = L_t \Omega_{t-1}^+ L_t^T + \Sigma_\eta^+, \quad \Gamma_t = F_t \Omega_t F_t^T + \Sigma_\varepsilon,$$

$$D_t^+ = \Omega_{t-1}^+ L_t^T (\Omega_t^+)^{-1}, \quad E_t = D_t \Omega_t F_t^T \Gamma_t^{-1}, \quad v_t^+ = \psi_{t-1}^+ - D_t^+ \psi_t^+$$

$$\gamma_t = v_t = (0^T, (v_t^+)^T)^T, \quad \Delta_t^+ = \Omega_{t-1}^+ - D_t^+ \Omega_t^+ (D_t^+)^T$$

and

$$\Theta_t = \Delta_t + (D_t - E_t F_t) \Omega_t D_t^T.$$

Although similar to Eqs. (5)–(10), the relationships presented in the above proposition show important differences. The main one is between the most important skewness parameter D_t in (7) and D_t^+. For the former, if $D_{t-1} = 0$ then $D_t = 0$ for all t. In comparison, if $D_{t-1}^+ = 0$ then D_t^+ can be very different from 0. This means that the skewness can be easily changed in time for the latter model. Another advantage of working with a state vector defined by (17) is that this model gives the power to clearly identify the skewness sources, and therefore the parameter interpretation is much easier.

4. Sequential estimation procedure: Kalman filtering

Following the work of Meinhold and Singpurwalla [15], we use a Bayesian formulation to derive the different steps of the Kalman filtering for the two models presented in the previous section, i.e. the skewed linear state-space model in Section 3.1 and the extended state-space models in Section 3.2. The key notion is that given the data $\mathbf{Y}_t = (Y_1, \ldots, Y_t)$, inference about the state vector values can be carried out through a direct application of Bayes' theorem. In the Kalman literature, the conditional distribution of $(X_{t-1} \mid \mathbf{Y}_{t-1})$ is usually assumed to follow a Gaussian distribution at time $t-1$. In our case, this assumption at time $t-1$ is expressed in function of the closed skew-normal distribution

$$(X_{t-1} \mid \mathbf{Y}_{t-1}) \sim CSN_{n,m}(\hat{\psi}_{t-1}, \hat{\Omega}_{t-1}, \hat{D}_{t-1}, \hat{v}_{t-1}, \hat{\Delta}_{t-1}), \tag{19}$$

where $\hat{.}$ represents the location, scale, shape, and skewness parameters of $(X_{t-1} \mid \mathbf{Y}_{t-1})$. Then, we look forward in time t, but in two stages: prior to observing Y_t, and after observing Y_t. To implement these two steps, we need to determine the conditional distribution of a closed skew-normal distribution. The following lemma which can be found in González-Farías et al. [9] gives such a result.

Lemma 6. *Suppose that Y is a closed skew-normal random vector $Y \sim CSN_{n,m}(\mu, \Sigma, D, v, \Delta)$ and it is partitioned into two components, Y_1 and Y_2, of dimensions h and $n-h$, respectively, and with a corresponding partition for μ, Σ, D, and v. Then the conditional distribution of Y_2 given $Y_1 = y_1$ is:*

$$CSN_{n-h,m}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, D_2, v - D_1 y_1, \Delta). \tag{20}$$

Note that the converse is also true, i.e. if (20) is the conditional distribution of Y_2 given $Y_1 = y_1$ and $Y_1 \sim CSN_{h,m}(\mu_1, \Sigma_{11}, D_1, v_1, \Delta)$, then the joint distribution of Y_1 and Y_2 is $CSN_{n,m}(\mu, \Sigma, D, v, \Delta)$.

4.1. Skewed linear state-space model

In this section, we assume that the model presented in Proposition 3 holds. In particular, the noises, ε_t and η_t, added at each time step are supposed to be normally distributed. The next proposition summarizes the different Kalman filtering steps necessary to sequentially update the state of this particular model.

Proposition 7. *Suppose that the initial state vector X_0 of the system composed by Eqs.* (3) *and* (4) *follows a closed skew-normal distribution, $CSN_{n,m}(\psi_0, \Omega_0, D_0, v_0, \Delta_0))$ and that the noise ε_t, respectively, η_t, is an i.i.d. Gaussian vector with mean μ_ε and covariance Σ_ε, respectively, μ_η and Σ_η. Then, the parameters of the posterior distribution of X_t defined by* (19) *are computed through the next cycle by the following sequential procedure*:

$$\hat{\psi}_t = G_t\hat{\psi}_{t-1} + \mu_\eta + \tilde{\Omega}_t F_t^T(\Sigma_\varepsilon + F_t\tilde{\Omega}_t F_t^T)^{-1}[Y_t - F_t[G_t\hat{\psi}_{t-1} + \mu_\eta] - \mu_\varepsilon]$$

with

$$\tilde{\Omega}_t = G_t\hat{\Omega}_{t-1}G_t^T + \Sigma_\eta, \quad \hat{\Omega}_t = \tilde{\Omega}_t - \tilde{\Omega}_t F_t^T(\Sigma_\varepsilon + F_t\tilde{\Omega}_t F_t^T)^{-1}F_t\tilde{\Omega}_t,$$

$$\hat{D}_t = \hat{D}_{t-1}\hat{\Omega}_{t-1}G_t^T\tilde{\Omega}_t^{-1}, \quad \hat{v}_t = \hat{v}_{t-1} \quad \textit{and}$$
$$\hat{\Delta}_t = \hat{\Delta}_{t-1} + (\hat{D}_{t-1} - \hat{D}_t G_t)\hat{\Omega}_{t-1}\hat{D}_{t-1}.$$

This series of equations constitutes the Kalman filtering steps for the skewed linear state-space model.

This proposition shows that adding skewness does not change fundamentally the classical Kalman filtering operations for the skewed linear state-space model. The only difference with the classical Gaussian Kalman filter is the equalities dealing with the new parameters $\hat{D}_t$, $\hat{v}_t$ and $\hat{\Delta}_t$. They characterize the added skewness and they have the advantage to be easy to implement. Note that the estimators of v are time invariant. This corroborates the result found in Proposition 3.

4.2. Extended state-space model

Proposition 8. *Suppose that the initial vector $(U_0^T, V_0^T)^T$ of the linear system defined by* (12) *follows the normal distribution defined by* (18). *Then, the posterior distribution of $(X_t \mid \mathbf{Y}_t)$ defined from* (11), (12), *and* (14) *follows a $CSN(X_t|\mathbf{Y}_t) \sim CSN_{k+l,k+l}(\hat{\psi}_t, \hat{\Omega}_t, \hat{D}_t, \hat{v}_t, \hat{\Delta}_t)$ with*

$$\hat{\psi}_t = \begin{pmatrix} \hat{\psi}_t^* \\ \hat{\psi}_t^+ \end{pmatrix}, \quad \hat{\Omega}_t = \begin{pmatrix} \hat{\Omega}_t^* & 0 \\ 0 & \hat{\Omega}_t^+ \end{pmatrix}, \quad \hat{D}_t = \begin{pmatrix} 0 & 0 \\ 0 & \hat{D}_t^+ \end{pmatrix},$$

$$\hat{v}_t = \begin{pmatrix} 0 \\ \hat{v}_t^+ \end{pmatrix} \quad \textit{and} \quad \hat{\Delta}_t = \begin{pmatrix} I & 0 \\ 0 & \hat{\Sigma}_v^+ \end{pmatrix}.$$

The parameters of the posterior distributions are computed through the next cycle by the following sequential procedure:

$$\begin{pmatrix} \hat{\psi}_t^* \\ \hat{\psi}_t^+ \end{pmatrix} = \begin{pmatrix} K_t\hat{\psi}_{t-1}^* + \mu_\eta^* + \tilde{\Omega}_t^* Q_t^T\Sigma_t^{-1}e_t \\ -L_t\hat{\psi}_{t-1}^+ + \mu_\eta^+ + C_t P_t^T\Sigma_t^{-1}e_t \end{pmatrix},$$

where $e_t = Y_t - Q_t[K_t\hat{\psi}_{t-1}^* + \mu_\eta^*] - P_t[-L_t\hat{\psi}_{t-1}^+ + \mu_\eta^+ + \tau_t^{(1)}] - \mu_\varepsilon$, C_t is the conditional covariance $C_t = cov(V_t, S_t \mid \mathbf{Y}_{t-1})$, $\tilde{\Omega}_t^+ = L_t\hat{\Omega}_{t-1}^+L_t^T + \Sigma_\eta^+$, $\tilde{\Omega}_t^* = K_t\hat{\Omega}_{t-1}^*K_t^T + \Sigma_\eta^*$, and

$\Sigma_t = Q_t \tilde{\Omega}_t^* Q_t^T + P_t(\tilde{\Omega}_t^+ + \tau_t^{(2)}))P_t^T + \Sigma_\varepsilon$, *with*

$$\tau_t^{(i)} = \left. \frac{\partial^i \log(\Phi_m(\tilde{D}_t^+ \tilde{\Omega}_t^+ \theta; \tilde{\nu}_t^+, \tilde{\Delta}_t^+ + \tilde{D}_t^+ \tilde{\Omega}_t^+ (\tilde{D}_t^+)^T))}{\partial \theta^i} \right|_{\theta=0} \quad for\ i = 1, 2 \tag{21}$$

with $\tilde{D}_t^+ = \hat{\Omega}_{t-1}^+ L_t^T (\tilde{\Omega}_t^+)^{-1}$, $\tilde{\nu}_t^+ = \hat{\psi}_{t-1}^+ - \tilde{D}_t^+ \tilde{\psi}_t^+$, *and* $\tilde{\Delta}_t^+ = \hat{\Omega}_{t-1}^+ - \tilde{D}_t^+ \tilde{\Omega}_t^+ (\tilde{D}_t^+)^T$. *The covariance matrices are equal to*

$$\begin{pmatrix} \hat{\Omega}_t^* \\ \hat{\Omega}_t^+ \end{pmatrix} = \begin{pmatrix} \tilde{\Omega}_t^* - \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} Q_t \tilde{\Omega}_t^* \\ \tilde{\Omega}_t^+ - C_t P_t^T \Sigma_t^{-1} P_t C_t \end{pmatrix},$$

$\hat{D}_t^+ = \overline{\Omega}_{t-1}^+ \overline{L}_t^T (\hat{\Omega}_t^+)^{-1}$, $\hat{\nu}_t^+ = \hat{\psi}_{t-1}^+ - \hat{D}_t^+ \hat{\psi}_t^+$, *and* $\hat{\Delta}_t^+ = \overline{\Omega}_{t-1}^+ - \hat{D}_t^+ \hat{\Omega}_t^+ (\hat{D}_t^+)^T$, *where* $\overline{\Omega}_{t-1}^+ = \hat{\Omega}_{t-1}^+ - C_t P_t^T \Sigma_t^{-1} P_t C_t$, $\overline{L}_t = L_t + \Sigma_\nu^+ P_t^T \Sigma_t^{-1} P_t \tilde{C}_t (\overline{\Omega}_{t-1}^+)^{-1}$, *and* $C_t = \Sigma_\eta^+ - L_t \tilde{C}_t$.

This series of equations constitutes the Kalman filtering steps for the skewed extended state-space model.

Although the notations are a little more complex than in the previous proposition, the Kalman filtering steps for the skewed extended state-space model do not present any computational difficulties. As previously mentioned, the advantage of this model over the linear one is that the temporal structure offers more flexibility.

5. Discussions and conclusions

To illustrate the difference between the classical Gaussian Kalman filter and our non-linear skewed Kalman filter, both filters were used to estimate the temporal evolution of the state vector X_t from simulated observations Y_t. These observations were generated by setting $F_t = P_t = (-1)^t/2$, $\mu_\varepsilon = 0$, $\mu_\eta^+ = 2$, $Q_t = K_t = \Sigma_\eta^* = \mu_\eta^* = 0$, and $\Sigma_\varepsilon = \Sigma_\eta^+ = 1$. This is the same setting as in Fig. 2 that shows the evolution of all parameters used to simulate our observations. In Fig. 3, the solid line represents the observed path for X_t and the circles correspond to the estimated $\hat{X}_t$ from classical Gaussian KF (white circles) and non-linear skewed KF (black circles). For small t ($t < 15$), the skewness introduced by L_t (top panel) is still weak and the difference between both estimators is small. But for larger t and therefore greater skewness, the classical KF cannot capture the slow temporal increase in X_t values. In comparison, the non-linear skewed KF follows more closely this tendency. For a numerical point of view, the mean-square point error $\sum_{t=1}^{40}(X_t - \hat{X}_t)^2/40$ was computed for both filters, yielding 0.84 for our skewed Kalman filter and 1.58 for the classical one. This clearly indicates that the classical Kalman filter lost some efficiency when skewness was introduced.

Obviously, there have been many other attempts to deal with non-Gaussian state space models in the past. To name a few, Smith and Miller [19], Bradley et al. [5] and Meinhold and Singpurwalla [16] have proposed alternative approaches to the classical Kalman filter. Meinhold and Singpurwalla [16] assumed a multivariate distribution with Student-t marginals. Bradley et al. [5] proposed a methodology based on normal scale mixtures.

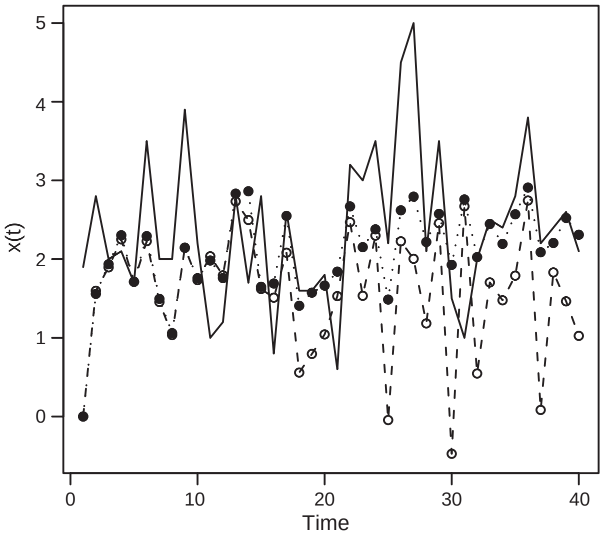

Fig. 3. Estimation of the temporal evolution of the state space variable X_t by using the classical Gaussian Kalman filter (white circles) and by implementing the non-linear skewed Kalman filter (black circles). The solid line represents the simulated values of X_t and the circles the estimated values of X_t. The skewness introduced through the time evolution of L_t is shown in the top panel of Fig. 2.

Smith and Miller [19] worked with exponential variables conditionally on unobserved variables. A common characteristic between these three studies is that they were all based on a Bayesian framework. In comparison, our approach does not make use of prior and posterior distributions. But this is by no means essential to the implementation of our strategy. Propositions 3 and 5 show that our models propagate the closed skew-normal distribution in time. Consequently, one could use a Bayesian approach if wanted. Lemma 6 will be then the cornerstone for deriving conditional densities. The limitations of our approach are elsewhere. Because of the very nature of skew-normal distributions, it is not possible to model heavy tail behaviors and/or to represent multi-modal distributions. For the latter point, we believe that the two approaches (skew normals and mixture of normals) are in fact complementary when modeling data and they could be combined. One is more adapted when dealing with skewness and the other is better representing multi-modality. For highly complex observations (multi-modal and skewed), more research has to be done to implement a method based on a mixture of closed skew-normal distributions and to compare it with other mixture approaches. Concerning heavy tail distributions, current research is undertaken to introduce skewness with the same strategy used in (1), i.e. a density multiplied by another distribution function. Finally, we would like to stress that the skewness is clearly identifiable in our parametrization. The interpretation of parameters in mixture models is sometimes not as clear.

In this work, we showed that extending the normal distribution to the closed skew-normal distribution for state-space models did neither reduce the flexibility nor the traceability of the operations associated with Kalman filtering. To the contrary, the introduction of a few skewness parameters provides a simple source of asymmetry needed for many applications. Further research is currently conducted to illustrate the capabilities of such extended state-space models for real case studies.

Acknowledgements

Part of this research was done in Boulder USA and supported by the NCAR Geophysical Statistics Project (GSP) whose generous hospitality is greatly acknowledged. Another part of this project was sponsored by the NSF-GMS Grant ATM-0327936. We thank an Associate Editor and two anonymous referees for a number of suggestions which led to a better presentation of the material. P. Naveau thank Uli Schneider from the GSP team for very interesting discussions about the folding coupler.

AppendixA.

Proof of Lemma 1. Following the work of González-Farías et al. [9] and using Eq. (2), we can write that the moment generating function of $X = AY$ is equal to

$$\begin{aligned} M_X(t) &= M_Y(A^T t) \\ &= \frac{\Phi_m(D(\Sigma A^T t); \nu, \Delta + D\Sigma D^T)}{\Phi_m(0; \nu, \Delta + D\Sigma D^T)} \exp\left\{\mu^T A^T t + \frac{1}{2}(t^T A\Sigma A^T t)\right\} \\ &= \frac{\Phi_m((DA^{\leftarrow})(A\Sigma A^T t); \nu, \Delta + (DA^{\leftarrow})A\Sigma A^T (DA^{\leftarrow})^T)}{\Phi_m(0; \nu, \Delta + (DA^{\leftarrow})A\Sigma A^T (DA^{\leftarrow})^T)} \\ &\quad \times \exp\left\{\mu^T A^T t + \frac{1}{2}(t^T A\Sigma A^T t)\right\}, \end{aligned}$$

where $A^{\leftarrow}$ is the left inverse of A. $\square$

Proof of Lemma 2. It is well known that the moment generating function of the Gaussian random vector Z is equal to $M_Z(t) = \exp(\mu^T t + t^T \Sigma t/2)$, whereas the moment generating of X is given by (2). Since the moment generating function of the sum of independent vectors is simply the product of each moment generating function, we have

$$M_{X+Z}(t) = \frac{\Phi_m(D(\Omega t); \nu, \Delta + D\Omega D^T)}{\Phi_m(0; \nu, \Delta + D\Omega D^T)} \exp\left\{(\mu + \psi)^T t + \frac{1}{2}(t^T(\Sigma + \Omega)t)\right\}.$$

Clearly, we have to set $\mu_{X+Z} = \mu + \psi$ and $\Sigma_{X+Z} = \Sigma + \Omega$. The difficulty is to show that $\Phi_m(D(\Omega t); \nu, \Delta + D\Omega D^T)$ can be rewritten as

$$\Phi_m(D_{X+Z}(\Sigma_{X+Z} t); \nu_{X+Z}, \Delta_{X+Z} + D_{X+Z}\Sigma_{X+Z} D_{X+Z}^T),$$

for the appropriate Σ_{X+Z}, D_{X+Z} and ν_{X+Z}. A little algebra allows us to verify that Σ_{X+Z}, D_{X+Z} and ν_{X+Z} stated in Lemma 2 satisfies the required specification. $\square$

Proof of Lemma 4. Introduce the constant $c = D_t^+ \psi_t^+$ and assume that $c \leqslant \psi_{t-1}^+$ (the case $c > \psi_{t-1}^+$ can be treated with the same argument). First, we will show that W_{t-1} has the same distribution that $[V_{t-1} \mid V_{t-1} \leqslant c]$. Introduce $V_{t-1}^* = 2\psi_{t-1}^+ - V_{t-1}$ and $\tilde{V}_{t-1} = \Phi_{t-1}^{-1}(h_{t-1}(V_{t-1}))$ where $h_{t-1}(x) = a\Phi_{t-1}(x) + b$ with

$$a = \frac{\Phi_{t-1}(c)}{\Phi_{t-1}(2\psi_{t-1}^+ - c) - \Phi_{t-1}(c)} \quad \text{and} \quad b = -\Phi_{t-1}(c)a.$$

Note that $\{x : c < x < 2\psi_{t-1}^+ - c\} = \{x : -\infty < \Phi_{t-1}^{-1}(h_{t-1}(x)) < c\}$. With these new notations, the vector W_{t-1} defined by (15) can be rewritten as

$$W_{t-1} = \begin{cases} V_{t-1} & \text{if } V_{t-1} \leqslant c, \\ V_{t-1}^* & \text{if } V_{t-1}^* \leqslant c, \\ \tilde{V}_{t-1} & \text{if } \tilde{V}_{t-1} \leqslant c. \end{cases}$$

It follows that the distribution of W_{t-1} is equal to

$$\begin{aligned} P(W_{t-1} \leqslant x) = {} & P(V_{t-1} \leqslant x \mid V_{t-1} \leqslant c) P(V_{t-1} \leqslant c) \\ & + P(V_{t-1}^* \leqslant x \mid V_{t-1}^* \leqslant c) P(V_{t-1} \geqslant 2\psi_{t-1}^+ - c) \\ & + P(\tilde{V}_{t-1} \leqslant x \mid \tilde{V}_{t-1} \leqslant c) P(c < V_{t-1} < 2\psi_{t-1}^+ - c). \end{aligned}$$

Because the mean and the variance of V_{t-1}^* are equal to those of V_{t-1}, we have $P(V_{t-1}^* \leqslant x \mid V_{t-1}^* \leqslant c) = P(V_{t-1} \leqslant x \mid V_{t-1} \leqslant c)$. The variable $\Phi_{t-1}(V_{t-1})$ follows an uniform distribution on the interval $[0, 1]$. Consequently, the variable $[\Phi_{t-1}(V_{t-1}) \mid c < V_{t-1} < 2\psi_{t-1}^+ - c]$ is also uniformly distributed but on the interval $[\Phi_{t-1}(c), \Phi_{t-1}(2\psi_{t-1}^+ - c)]$. It follows that

$$\begin{aligned} P(\tilde{V}_{t-1} \leqslant x \mid \tilde{V}_{t-1} \leqslant c) &= P(h_{t-1}(V_{t-1}) \leqslant \Phi_{t-1}(x) \mid c < V_{t-1} < 2\psi_{t-1}^+ - c) \\ &= P\left(\Phi_{t-1}(V_{t-1}) \leqslant \frac{\Phi_{t-1}(x) - b}{a} \,\middle|\, c < V_{t-1} < 2\psi_{t-1}^+ - c\right) \\ &= \frac{\Phi_{t-1}(x)}{\Phi_{t-1}(c)} = P(V_{t-1} \leqslant x \mid V_{t-1} \leqslant c). \end{aligned}$$

Hence, we have $P(W_{t-1} \leqslant x) = P(V_{t-1} \leqslant x \mid V_{t-1} \leqslant c)$ and then W_{t-1} has the same distribution that $[V_{t-1} \mid V_{t-1} \leqslant c]$.

Because of (14) and (12), S_t has the same distribution as $[V_t \mid V_{t-1} \leqslant c]$. The second part of this proof is to show that $[V_t \mid V_{t-1} \leqslant c]$ follows a closed skew-normal distribution. The argument is classical and it is shown for completeness.

We deduce from Eqs. (12) and (13) that

$$\begin{pmatrix} V_t \\ V_{t-1} \end{pmatrix} \sim N_{2l}\left(\begin{pmatrix} \psi_t^+ \\ \psi_{t-1}^+ \end{pmatrix}, \begin{pmatrix} \Omega_t^+ & -L_t \Omega_{t-1}^+ \\ -\Omega_{t-1}^+ L_t^T & \Omega_{t-1}^+ \end{pmatrix} \right).$$

The conditional distribution of $[V_{t-1} \mid V_t = y]$ also follows a normal distribution with mean $m_t(y) = \psi_{t-1}^+ - \Omega_{t-1}^+ L_t^T (\Omega_t^+)^{-1}(y - \psi_t^+)$ and variance

$$\Delta_t^+ = \Omega_{t-1}^+ - \Omega_{t-1}^+ L_t^T (\Omega_t^+)^{-1} L_t \Omega_{t-1}^+.$$

Define $D_t^+ = \Omega_{t-1}^+ L_t^T (\Omega_t^+)^{-1}$, then $m_t(y) = \psi_{t-1}^+ - D_t^+(y - \psi_t^+)$, and $\Delta_t^+ = \Omega_{t-1}^+ - D_t^+ \Omega_t^+ (D_t^+)^T$. These equalities allow us to write the density of the skewness vector S_t

$$
\begin{aligned}
f_{S_t}(y \mid V_{t-1} \leqslant D_t^+ \psi_t^+) &= \frac{f_{V_t}(y)\ P(V_{t-1} \leqslant D_t^+ \psi_t^+ \mid V_t = y)}{P(V_{t-1} \leqslant D_t^+ \psi_t^+)} \\
&= \frac{\phi_l(y; \psi_t^+, \Omega_t^+)\ \Phi_l(D_t^+ \psi_t^+; m_t(y), \Delta_t^+)}{\Phi_l(D_t^+ \psi_t^+; \psi_{t-1}^+, \Omega_{t-1}^+)} \\
&= \frac{\phi_l(y; \psi_t^+, \Omega_t^+) \Phi_l(D_t^+ (y - \psi_t^+); \psi_{t-1}^+ - D_t^+ \psi_t^+, \Delta_t^+)}{\Phi_l(0; \psi_{t-1}^+ - D_t^+ \psi_t^+, \Delta_t^+ + D_t^+ \Omega_t^+ (D_t^+)^T)}.
\end{aligned}
$$

Comparing the rhs of the last equality with the definition of the closed skew-normal distribution (1) gives the required result. □

Proof of Lemma 6. The proof of this lemma is the same as for the multivariate Gaussian distribution. It is based on the moment generating function defined by (2). □

Proof of Proposition 7. Because of (19) and (4), we have

$$
\begin{aligned}
(X_t \mid \mathbf{Y}_{t-1}) &= (G_t X_{t-1} + \eta_t \mid \mathbf{Y}_{t-1}) \\
&= G_t (X_{t-1} \mid \mathbf{Y}_{t-1}) + \eta_t.
\end{aligned}
$$

Since the noise η_t is assumed to follow a normal distribution, we can apply Lemmas 1 and 2 to deduce the state of knowledge X_t prior to observing Y_t

$$
(X_t \mid \mathbf{Y}_{t-1}) \sim CSN(G_t \hat{\psi}_{t-1} + \mu_\eta, \tilde{\Omega}_t, \hat{D}_t, \hat{v}_t, \hat{\Delta}_t) \tag{A.1}
$$

with $\tilde{\Omega}_t = G_t \hat{\Omega}_{t-1} G_t^T + \Sigma_\eta$, $\hat{D}_t = \hat{D}_{t-1} \hat{\Omega}_{t-1} G_t^T \tilde{\Omega}_t^{-1}$, $\hat{v}_t = \hat{v}_{t-1}$, and $\hat{\Delta}_t = \hat{\Delta}_{t-1} + (\hat{D}_{t-1} - \hat{D}_t G_t) \hat{\Omega}_{t-1} \hat{D}_{t-1}$. On observing Y_t, our objective is to compute the posterior of X_t, i.e. $P(X_t \mid \mathbf{Y}_t)$. To reach this goal, we introduce

$$
e_t = Y_t - F_t[G_t \hat{\psi}_{t-1} + \mu_\eta] - \mu_\varepsilon,
$$

the error in predicting Y_t from point $t-1$. From this definition and the observation Eq. (3), it follows that

$$
\begin{aligned}
(e_t \mid X_t, \mathbf{Y}_{t-1}) &= (F_t[X_t - G_t \hat{\psi}_{t-1} - \mu_\eta] + \varepsilon_t - \mu_\varepsilon \mid X_t, \mathbf{Y}_{t-1}) \\
&= (F_t[X_t - G_t \hat{\psi}_{t-1} - \mu_\eta] \mid X_t, \mathbf{Y}_{t-1}) + \varepsilon_t - \mu_\varepsilon.
\end{aligned}
$$

This last equality in distribution and the normality of ε_t imply that

$$
(e_t \mid X_t, \mathbf{Y}_{t-1}) \sim N_d(F_t[X_t - G_t \hat{\psi}_{t-1} - \mu_\eta], \Sigma_\varepsilon). \tag{A.2}
$$

To link the posterior of X_t with the error e_t, we notice that

$$
P(X_t \mid \mathbf{Y}_t) = P(X_t \mid Y_t, \mathbf{Y}_{t-1}) = P(X_t \mid e_t, \mathbf{Y}_{t-1}).
$$

Applying the converse of Lemma 6 to the distributions of $(e_t \mid X_t, \mathbf{Y}_{t-1})$ and $(X_t \mid \mathbf{Y}_{t-1})$, respectively, described by (A.1) and (A.2), allows us to derive that the vector $(X_t^T, e_t^T \mid \mathbf{Y}_{t-1})^T$ follows

$$CSN\left(\begin{pmatrix} G_t\hat{\psi}_{t-1}+\mu_\eta \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tilde{\Omega}_t & \tilde{\Omega}_t F_t^T \\ F_t\tilde{\Omega}_t & \Sigma_\varepsilon + F_t\tilde{\Omega}_t F_t^T \end{pmatrix}, \begin{pmatrix} \hat{D}_t \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \hat{v}_t \\ 0 \end{pmatrix}, \hat{\Delta}_t\right).$$

From Lemma 6, we can then deduce the desired posterior distribution

$$(X_t \mid \mathbf{Y}_t) = (X_t \mid e_t, \mathbf{Y}_{t-1}) \sim CSN(\hat{\psi}_t, \hat{\Omega}_t, \hat{D}_t, \hat{v}_t, \hat{\Delta}_t)$$

with parameters stated in Proposition 7. □

Proof of Proposition 8. For the non-linear state-space model defined by (11), (12) and (14), we assume that we have up to time $t-1$ (this is true for $t=0$)

$$\left(\begin{matrix} U_{t-1} \\ \\ V_{t-1} \end{matrix}\middle| \mathbf{Y}_{t-1}\right) \sim N_{k+l}\left(\begin{pmatrix} \hat{\psi}^*_{t-1} \\ \hat{\psi}^+_{t-1} \end{pmatrix}, \quad \begin{pmatrix} \hat{\Omega}^*_{t-1} & \hat{\Omega}^{*+}_{t-1} \\ \hat{\Omega}^{*+}_{t-1} & \hat{\Omega}^+_{t-1}, \end{pmatrix}\right), \tag{A.3}$$

where $\hat{.}$ represents the posterior mean and covariance. From Eq. (12), we deduce that

$$\left(\begin{matrix} U_t \\ V_t \\ V_{t-1} \end{matrix}\middle| \mathbf{Y}_{t-1}\right) = \left(\begin{matrix} K_t U_{t-1} \\ -L_t V_{t-1} \\ V_{t-1} \end{matrix}\middle| \mathbf{Y}_{t-1}\right) + \begin{pmatrix} \eta^*_t \\ \eta^+_t \\ 0 \end{pmatrix}.$$

Hence, the variable $(U_t, V_t, V_{t-1} \mid \mathbf{Y}_{t-1})^T$ is Gaussian with mean and variance equal to

$$N\left(\begin{pmatrix} K_t\hat{\psi}^*_{t-1}+\mu^*_\eta \\ -L_t\hat{\psi}^+_{t-1}+\mu^+_\eta \\ \hat{\psi}^+_{t-1} \end{pmatrix}, \begin{pmatrix} \tilde{\Omega}^*_t & -K_t\hat{\Omega}^{*+}_{t-1}L_t^T & K_t\hat{\Omega}^{*+}_{t-1} \\ -L_t\hat{\Omega}^{*+}_{t-1}K_t^T & \tilde{\Omega}^+_t & -L_t\hat{\Omega}^+_{t-1}, \\ \hat{\Omega}^{*+}_{t-1}K_t^T & -\hat{\Omega}^+_{t-1}L_t^T & \hat{\Omega}^+_{t-1} \end{pmatrix}\right) \tag{A.4}$$

with $\tilde{\Omega}^+_t = L_t\hat{\Omega}^+_{t-1}L_t^T + \Sigma^+_\eta$ and $\tilde{\Omega}^*_t = K_t\hat{\Omega}^*_{t-1}K_t^T + \Sigma^*_\eta$.

On observing Y_t, our objective is to first compute the posterior of (U_t, V_t, V_{t-1}), i.e. $P(U_t, V_t, V_{t-1} \mid \mathbf{Y}_t)$ and then to obtain $P(X_t \mid \mathbf{Y}_t)$. For the former, we introduce $e_t = Y_t - Q_t[K_t\hat{\psi}^*_{t-1} + \mu^*_\eta] - P_t[E(S_t \mid \mathbf{Y}_{t-1})] - \mu_\varepsilon$, where $E(S_t \mid \mathbf{Y}_{t-1})$ is the conditional expectation of S_t given $\mathbf{Y}_{t-1}$. To compute this quantity, we follow the same procedure used in Lemma 4 to deduce that the variable $(S_t \mid \mathbf{Y}_{t-1})$ follows approximately a closed skew-normal distribution function $CSN_{l,1}(\tilde{\psi}^+_t, \tilde{\Omega}^+_t, \tilde{D}^+_t, \tilde{v}^+_t, \tilde{\Delta}^+_t)$ with $\tilde{\psi}^+_t = -L_t\hat{\psi}^+_{t-1} + \mu^+_\eta$, $\tilde{D}^+_t = \hat{\Omega}^+_{t-1}L_t^T(\tilde{\Omega}^+_t)^{-1}$, $\tilde{v}^+_t = \hat{\psi}^+_{t-1} - \tilde{D}^+_t\tilde{\psi}^+_t$, and $\tilde{\Delta}^+_t = \hat{\Omega}^+_{t-1} - \tilde{D}^+_t\tilde{\Omega}^+_t(\tilde{D}^+_t)^T$. To compute the mean and the variance of $(S_t \mid \mathbf{Y}_{t-1})$, we use the moment generating function $M(\theta)$ from (2) (Genton et al. [7] also computed these moments in the special case $v=0$)

$$M(\theta) = \frac{\Phi_m(\tilde{D}^+_t\tilde{\Omega}^+_t\theta; \tilde{v}^+_t, \tilde{\Delta}^+_t + \tilde{D}^+_t\tilde{\Omega}^+_t(\tilde{D}^+_t)^T)}{\Phi_m(0; \tilde{v}^+_t, \tilde{\Delta}^+_t + \tilde{D}^+_t\tilde{\Omega}^+_t(\tilde{D}^+_t)^T)} \exp\{(\tilde{\psi}^+_t)^T\theta + \frac{1}{2}(\theta^t\tilde{\Omega}^+_t\theta)\}.$$

The cumulant function ($K(\theta) = \log M(\theta)$) becomes

$$K(\theta) = c_m + \log(\Phi_m(\tilde{D}_t^+ \tilde{\Omega}_t^+ \theta; \tilde{v}_t^+, \tilde{\Delta}_t^+ + \tilde{D}_t^+ \tilde{\Omega}_t^+ (\tilde{D}_t^+)^T)) + (\tilde{\psi}_t^+)^T \theta + \frac{1}{2}(\theta^T \tilde{\Omega}_t^+ \theta).$$

Taking the first and second derivatives of the cumulant function provides $E(S_t \mid \mathbf{Y}_{t-1}) = -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ + \tau_t^{(1)}$ and $V(S_t \mid \mathbf{Y}_{t-1}) = \tilde{\Omega}_t^+ + \tau_t^{(2)}$ where $\tau^{(i)}$ is defined by (21).

From the observation Eq. (11), $(e_t \mid U_t, V_t, V_{t-1}, \mathbf{Y}_{t-1})$ is equal to

$$(Q_t[U_t - K_t \hat{\psi}_{t-1}^* - \mu_\eta^*] + P_t[S_t - E(S_t \mid \mathbf{Y}_{t-1})] + \varepsilon_t - \mu_\varepsilon \mid U_t, V_t, V_{t-1}, \mathbf{Y}_{t-1}).$$

This last equality, the normality of ε_t and the fact that the variable S_t is entirely defined from (η_t^+, V_{t-1}) (see Lemma 4) imply that the variable $(e_t \mid U_t, V_t, V_{t-1}, \mathbf{Y}_{t-1})$ follows

$$N(Q_t[U_t - K_t \hat{\psi}_{t-1}^* - \mu_\eta^*] + P_t[S_t - E(S_t \mid \mathbf{Y}_{t-1})], \Sigma_\varepsilon). \tag{A.5}$$

Applying the classical properties of the multivariate normal distribution to the variables $(e_t \mid U_t, V_t, V_{t-1}, \mathbf{Y}_{t-1})$ and $(U_t, V_t, V_{t-1} \mid \mathbf{Y}_{t-1})$, respectively, described by (A.4) and (A.5), allows us to derive that the vector $(U_t, V_t, V_{t-1}, e_t \mid \mathbf{Y}_{t-1})^T$ follows

$$N\left(\begin{pmatrix} K_t \hat{\psi}_{t-1}^* + \mu_\eta^* \\ -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ \\ \hat{\psi}_{t-1}^+ \\ 0 \end{pmatrix}, \begin{pmatrix} \tilde{\Omega}_t^* & -K_t \hat{\Omega}_{t-1}^{*+} L_t^T & K_t \hat{\Omega}_{t-1}^{*+} & \tilde{\Omega}_t^* Q_t^T \\ -L_t \hat{\Omega}_{t-1}^{*+} K_t^T & \tilde{\Omega}_t^+ & -L_t \hat{\Omega}_{t-1}^+ & C_t P_t^T \\ \hat{\Omega}_{t-1}^{*+} K_t^T & -\hat{\Omega}_{t-1}^+ L_t^T & \hat{\Omega}_{t-1}^+ & \tilde{C}_t P_t^T \\ Q_t \tilde{\Omega}_t^* & P_t C_t & P_t \tilde{C}_t & \Sigma_t \end{pmatrix} \right)$$

with $C_t = \operatorname{cov}(V_t, S_t \mid \mathbf{Y}_{t-1})$, $\tilde{C}_t = \operatorname{cov}(V_{t-1}, S_t \mid \mathbf{Y}_{t-1})$ and $\Sigma_t = Q_t \tilde{\Omega}_t^* Q_t^T + P_t V(S_t \mid \mathbf{Y}_{t-1}) P_t^T + \Sigma_\varepsilon$. Since we have $(U_t, V_t, V_{t-1} \mid \mathbf{Y}_t)^T = (U_t, V_t, V_{t-1} \mid e_t, \mathbf{Y}_{t-1})^T$, the distribution of this vector is a multivariate normal with mean

$$\begin{pmatrix} K_t \hat{\psi}_{t-1}^* + \mu_\eta^* \\ -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ \\ \hat{\psi}_{t-1}^+ \end{pmatrix} + H_t \Sigma_t^{-1} e_t, \quad \text{where } H_t = \begin{pmatrix} \tilde{\Omega}_t^* Q_t^T \\ C_t P_t^T \\ \tilde{C}_t P_t^T \end{pmatrix}$$

and its covariance matrix is equal to

$$\begin{pmatrix} \tilde{\Omega}_t^* & -K_t \hat{\Omega}_{t-1}^{*+} L_t^T & K_t \hat{\Omega}_{t-1}^{*+} \\ -L_t \hat{\Omega}_{t-1}^{*+} K_t^T & \tilde{\Omega}_t^+ & -L_t \hat{\Omega}_{t-1}^+ \\ \hat{\Omega}_{t-1}^{*+} K_t^T & -\hat{\Omega}_{t-1}^+ L_t^T & \hat{\Omega}_{t-1}^+ \end{pmatrix} - H_t \Sigma_t^{-1} H_t^T.$$

With the same kind of argument, the vector $(U_t, V_t \mid \mathbf{Y}_t)^T$ follows:

$$N\left(\left[\begin{pmatrix} K_t \hat{\psi}_{t-1}^* + \mu_\eta^* \\ -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ \end{pmatrix} + J_t \Sigma_t^{-1} e_t \right], \left[\begin{pmatrix} \tilde{\Omega}_t^* & -K_t \hat{\Omega}_{t-1}^{*+} L_t^T \\ -L_t \hat{\Omega}_{t-1}^{*+} K_t^T & \tilde{\Omega}_t^+ \end{pmatrix} - J_t \Sigma_t^{-1} J_t^T \right] \right)$$

with

$$J_t = \begin{pmatrix} \tilde{\Omega}_t^* Q_t^T \\ C_t P_t^T \end{pmatrix}, \quad \text{and} \quad J_t \Sigma_t^{-1} J_t^T = \begin{pmatrix} \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} Q_t \tilde{\Omega}_t^* & \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} P_t C_t \\ C_t P_t^T \Sigma_t^{-1} Q_t \tilde{\Omega}_t^* & C_t P_t^T \Sigma_t^{-1} P_t C_t \end{pmatrix}.$$

This distribution is used to implement the first update of the Kalman filter, i.e. the parameters of (A.3) are now set for a new cycle

$$\begin{pmatrix} \hat{\psi}_t^* \\ \hat{\psi}_t^+ \end{pmatrix} = \begin{pmatrix} K_t \hat{\psi}_{t-1}^* + \mu_\eta^* + \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} e_t \\ -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ + C_t P_t^T \Sigma_t^{-1} e_t \end{pmatrix}$$

and the covariance matrix $\begin{pmatrix} \hat{\Omega}_t^* & \hat{\Omega}_t^{*+} \\ \hat{\Omega}_t^{*+} & \hat{\Omega}_t^+ \end{pmatrix}$ is equal to

$$\begin{pmatrix} \tilde{\Omega}_t^* - \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} Q_t \tilde{\Omega}_t^* & -K_t \hat{\Omega}_{t-1}^{*+} L_t^T - \tilde{\Omega}_t^* Q_t^T \Sigma_t^{-1} P_t C_t \\ -L_t \hat{\Omega}_{t-1}^{*+} K_t^T - C_t P_t^T \Sigma_t^{-1} Q_t \tilde{\Omega}_t^* & \tilde{\Omega}_t^+ - C_t P_t^T \Sigma_t^{-1} P_t C_t \end{pmatrix}.$$

To get the final part, i.e. $P(X_t \mid \mathbf{Y}_t)$, we use the fact that the vector $(V_t, V_{t-1} \mid \mathbf{Y}_t)^T$ follows

$$N\left(\begin{pmatrix} -L_t \hat{\psi}_{t-1}^+ + \mu_\eta^+ + C_t P_t^T \Sigma_t^{-1} e_t \\ \hat{\psi}_{t-1}^+ + \tilde{C}_t P_t^T \Sigma_t^{-1} e_t \end{pmatrix}, \left[\begin{pmatrix} \tilde{\Omega}_t^+ & -L_t \hat{\Omega}_{t-1}^+ \\ -\hat{\Omega}_{t-1}^+ L_t^T & \hat{\Omega}_{t-1}^+ \end{pmatrix} - I_t \Sigma_t^{-1} I_t^T \right] \right)$$

with

$$I_t = \begin{pmatrix} C_t P_t^T \\ \tilde{C}_t P_t^T \end{pmatrix} \quad \text{and} \quad I_t \Sigma_t^{-1} I_t^T = \begin{pmatrix} C_t P_t^T \Sigma_t^{-1} P_t C_t & C_t P_t^T \Sigma_t^{-1} P_t \tilde{C}_t \\ \tilde{C}_t P_t^T \Sigma_t^{-1} P_t C_t & \tilde{C}_t P_t^T \Sigma_t^{-1} P_t \tilde{C}_t \end{pmatrix}.$$

Define $\overline{\Omega}_{t-1} = \hat{\Omega}_{t-1}^+ - C_t P_t^T \Sigma_t^{-1} P_t C_t$, and $\overline{L}_t = L_t + \Sigma_v^+ P_t^T \Sigma_t^{-1} P_t \tilde{C}_t (\overline{\Omega}_{t-1}^+)^{-1}$. The covariance matrix of the vector $(V_t, V_{t-1} \mid \mathbf{Y}_t)$ is then equal to

$$\begin{pmatrix} \hat{\Omega}_t^+ & -\overline{L}_t \overline{\Omega}_{t-1} \\ -\overline{\Omega}_{t-1} \overline{L}_t^T & \overline{\Omega}_{t-1} \end{pmatrix}.$$

It follows from (14) that $(S_t \mid \mathbf{Y}_t) \sim CSN_{l,l}(\hat{\psi}_t^+, \hat{\Omega}_t^+, \hat{D}_t^+, \hat{v}_t^+, \hat{\Delta}_t^+)$, with $\hat{D}_t^+ = \overline{\Omega}_{t-1}^+ \overline{L}_t^T (\hat{\Omega}_t^+)^{-1}$, $\hat{v}_t^+ = \hat{\psi}_{t-1}^+ - \hat{D}_t^+ \hat{\psi}_t^+$, and $\hat{\Delta}_t^+ = \overline{\Omega}_{t-1}^+ - \hat{D}_t^+ \hat{\Omega}_t^+ (\hat{D}_t^+)^T$. We deduce that the state vector has also a closed skew-normal distribution

$$(X_t \mid \mathbf{Y}_t) = \left(\begin{matrix} U_t \\ S_t \end{matrix} \middle| \mathbf{Y}_t \right) \sim CSN_{k+l,k+l}(\hat{\psi}_t, \hat{\Omega}_t, \hat{D}_t, \hat{v}_t, \hat{\Delta}_t), \quad \text{with } \hat{\psi}_t = \begin{pmatrix} \hat{\psi}_t^* \\ \hat{\psi}_t^+ \end{pmatrix},$$

$$\hat{\Omega}_t = \begin{pmatrix} \hat{\Omega}_t^* & 0 \\ 0 & \tilde{\Omega}_t \end{pmatrix}, \quad \hat{D}_t = \begin{pmatrix} 0 & 0 \\ 0 & \hat{D}_t^+ \end{pmatrix}, \quad \hat{v}_t = \begin{pmatrix} 0 \\ \hat{v}_t^+ \end{pmatrix}, \quad \text{and}$$
$$\hat{\Delta}_t = \begin{pmatrix} I & 0 \\ 0 & \hat{\Delta}_t^+ \end{pmatrix}. \qquad \square$$

References

[1] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.
[2] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 46 (1986) 199–208.
[3] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[4] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. Roy. Statist. Soc. B 61 (1999) 579–602.
[5] P.C. Bradley, N.G. Polson, D.S. Stoffer, A Monte Carlo approach to nonnormal and nonlinear state-space modeling, J. Amer. Statist. Assoc. 87 (1992) 493–500.
[6] J.N. Corcoran, U. Schneider, Coupling methods for perfect simulation, Probab. Eng. Inform. Sci. 17 (2003) 277–303.
[7] M.G. Genton, L. He, X. Liu, Moments of skew-normal random vectors and their quadratic forms, Statist. Probab. Lett. 51 (2001) 319–325.
[8] M.G. Genton, N. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Ann. Inst. Statist. Math, to appear.
[9] G. González-Farías, J.A. Domínguez-Molina, A.K. Gupta, The closed skew-normal distribution, in: M.G. Genton (Ed.), Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality, CRC/Chapman & Hall, London, 2004, pp. 25–42.
[10] G. González-Farías, J.A. Domínguez-Molina, A.K. Gupta, Additive properties of skew normal random vectors, J. Statist. Plann. Inference, to appear.
[11] W. Guo, Y. Wang, M. Brown, A signal extraction approach to modeling hormones time series with pulses and a changing baseline, J. Amer. Stat. Assoc. 94 (1999) 746–756.
[12] P.J. Harrison, C.F. Stevens, A Bayesian approach to short-term forecasting, Oper. Res. Quart. 22 (1971) 341–362.
[13] P.J. Harrison, C.F. Stevens, Bayesian forecasting, J. Roy. Statist. Soc. Ser. B 38 (1976) 205–247.
[14] G. Kitagawa, W. Gersch, A smoothness priors state-space modeling of times series with trend and seasonality, J. Amer. Statist. Assoc. 79 (1984) 378–389.
[15] R.J. Meinhold, N.D. Singpurwalla, Understanding the Kalman filter, Amer. Statist. 37 (1983) 123–127.
[16] R.J. Meinhold, N.D. Singpurwalla, Robustification of Kalman filter models, J. Amer. Statist. Assoc. 84 (1989) 479–486.
[17] N. Shephard, Partially non-Gaussian state-space models, Biometrika 81 (1994) 115–131.
[18] R.H. Shumway, D.S. Stoffer, Dynamic linear models with switching, J. Amer. Statist. Assoc. 86 (1991) 763–769.
[19] R.L. Smith, J.E. Miller, A non-Gaussian state space model and application to prediction of records, J. Roy. Statist. Soc. B 48 (1986) 79–88.
[20] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, Springer, New York, 1997.