

Non-linear Kalman filters comparison for generalised autoregressive conditional heteroscedastic clutter parameter estimation

ISSN 1751-9675

Received on 24th August 2018

Revised 1st February 2019

Accepted on 25th April 2019

doi: 10.1049/iet-spr.2018.5400

www.ietdl.org

Juan P. Pascual¹ ✉, Nicolás von-Ellenrieder², Javier Areta³, Carlos H. Muravchik⁴

¹LIAT, Dto. de Ing. en Telecomunicaciones, CNEA, Instituto Balseiro, CONICET, Av. Bustillo 9500, S.C. Bariloche, Argentina

²McGill University, Montreal, Canada

³Universidad Nacional de Río Negro (UNRN), CONICET, S.C. Bariloche, Argentina

⁴Instituto de Investigaciones en Electrónica, Control y Procesamiento de Señales (LEICI), UNLP and CICPBA, Av. 1 y 47, La Plata, Argentina

✉ E-mail: juanpablo.pascual@ib.edu.ar

Abstract: In this work, the authors analyse the estimation of the generalised autoregressive conditional heteroscedastic (GARCH) process conditional variance based on three non-linear filtering approaches: extended Kalman filter (EKF), unscented Kalman filter and cubature Kalman filter. The authors present a state model for a GARCH process and derive an EKF including second-order non-linear terms for simultaneous estimation of state and parameters. Using synthetic data, the authors evaluate the consistency and the correlation of the innovations for the three filters, by means of numerical simulations. The authors also study the performance of smoothed versions of the non-linear Kalman filters using real clutter data in comparison with a conventional quasi-maximum likelihood estimation method for the GARCH process coefficients. The authors show that with all methods the process coefficients estimates are of the same order and the resulting conditional variances are commensurable. However, the non-linear Kalman filters greatly reduce the computational load. These kind of filters could be used for the radar detector based on a GARCH clutter model that uses an adaptive threshold that demands the conditional variance at each decision instant.

1 Introduction

Clutter can greatly affect the performance of radar systems. Due to its random nature clutter is usually modelled as a stochastic process, whose distribution depends on the radar application. Given its mathematical tractability, radar detection based on Gaussian clutter models has been extensively investigated [1–3]. However, for sea radar and for high-resolution radar, clutter statistics deviates from the Gaussian model [4–7]. In such context, better suited distributions are log-normal, Weibull, K [8, 9], or the generalised compound probability density function (pdf) [10–12]. These distributions fit well heavy tailed clutter pdf. However, they are essentially time invariant, whereas often the radar environment may change abruptly [13], resulting in a degraded performance in real scenarios. Additionally, they are not easily mathematically tractable leading to suboptimal solutions to the detection problem [14].

Recently, we proposed a different approach to the adaptive detection problem considering a generalised autoregressive conditional heteroscedastic (GARCH) process to model the clutter [15]. A GARCH process is characterised by a time varying variance that depends on the history of the process and by a heavy tailed pdf [16], which is desirable for appropriately modelling clutter. The GARCH-based detector models the clutter considering the realisations in range (or fast time), i.e. it takes a realisation for each transmitted pulse, and the detection is performed for all the range cells in each pulse. One of the advantages of this detection scheme is that the threshold depends on the conditional variance, i.e. on the previous clutter samples for each cell under test. Thus, it is able to respond to fast changes in the clutter. However, for a real-time application, this detection algorithm requires a considerable amount of computational load, an optimisation problem has to be solved to find the estimates of the GARCH process coefficients before updating the conditional variance.

In order to estimate the unknown GARCH model parameters, the most commonly used estimation procedure has been the *quasi-maximum likelihood estimation* (QMLE) method [16, 17]. Since this method requires a large number of process samples to obtain

good parameter estimates in practical situations, we modified the procedure to use several short time series instead of a long one [15]. The disadvantage of the QMLE algorithms is that they are computationally intensive. An alternative estimation procedure is based on the autoregressive moving average model representation of the squared GARCH process. In [18], the use of the Yule-Walker estimator for the GARCH(1,1) model is proposed. Simulation results show that in some conditions this estimator performs similarly to the QMLE, but its convergence rate deteriorates as the number of model coefficients increase.

In [19], a discrete-time non-linear state-space formulation is used to characterise a GARCH (1,1) process and an extended Kalman filter (EKF) estimator is derived. This approach does not include the unknown model parameters in the state. Thus, the estimator works in two steps. In the first one, the QMLE method is used to find the unknown parameters. In the second step, the state estimate is calculated using the EKF with the estimated parameters in the first step. This mathematical representation was also extended to a GARCH model of arbitrary order [20]. In [21], we presented an adaptive linear minimum mean square error (LMMSE) estimator for the conditional variance of a GARCH process, analogous to the Kalman filter. However, it assumes that the coefficients are known and it also requires using the QMLE algorithm. This approach is similar to the estimation method proposed in [22], where a conditional variance estimator is derived from the minimum mean square error (MSE) criteria, assuming that the process parameters are known. However, they are estimated recursively with a maximum-likelihood procedure, numerically solved by means of a steepest descent method that updates the likelihood function gradient as new samples arrive, in order to keep a low computational load [23]. Recently, [24] proposed estimating GARCH time-varying parameters via the Kalman filter recursive equations, adapting the state-space representation of [20] to non-stationary GARCH models.

We consider here the discrete-time non-linear state-space formulation for GARCH models, in the scenario where the process coefficients are unknown and they are jointly estimated with the conditional variance, in a single step. This is akin to simultaneous

state and parameter estimation in linear models [25], cast as a non-linear estimation problem of an augmented state system. As far as we know, the non-linear model with augmented state including the conditional variance and the parameters is introduced here for GARCH models. With this formulation we set three Kalman filters which allow time updating the conditional variance at each instant with methods significantly less computationally intensive than the QMLE. We derive an EKF for the augmented state and, unlike the standard derivation, we include second-order non-linear terms to reduce the series expansion errors, which is another contribution of this work. We compare this EKF with the unscented Kalman filter (UKF) [26] and the cubature Kalman filter (CKF) [27].

The rest of the paper is organised as follows. In Section 2, we introduce the problem statement and the state model, we derive the EKF and we describe the sigma points and the cubature points used for the UKF and the CKF, respectively. We present numerical simulations results in Section 3 including a study of the filters consistency and of the correlation of the innovations. In Section 4, the three filters are tested on real sea clutter data to evaluate their performance in a real situation. These are also compared with the QMLE performance; because QMLE uses all the data inside the processing batch while the Kalman filters only use past data immediately before the filtering point, we combine each of the proposed Kalman filters with their smoothed versions to make a fair comparison [28]. Finally, we present conclusions in Section 5.

1.1 Notation

We adopt the following notation. Math italic is used for scalars x , uppercase bold for matrices \mathbf{X} and lowercase bold for vectors \mathbf{x} . For random variables x , y and z , the variable $x|y; z$ denotes the conditional random variable x given y and z . In addition, $\hat{\mathbf{x}}(r|k)$ denotes the estimate of the state $\mathbf{x}(r)$ given information available up to time k , the predicted value of the state if $k < r$ and the filtered value when $k = r$. The expectation and trace operators are denoted as $\mathbb{E}\{\cdot\}$ and $\text{tr}\{\cdot\}$ respectively, and \mathcal{CN} denotes a circularly-symmetric normal distribution.

2 State estimation for a GARCH process

2.1 Problem formulation

For simplicity, consider a complex GARCH (1,1) process $c(r)$ described by the set of equations [16]

$$c(r) = \sigma(r)z(r), \quad (1)$$

$$\sigma^2(r) = k + \alpha_1 \sigma^2(r-1) + \beta_1 |c(r-1)|^2, \quad (2)$$

where $z(r) \sim \mathcal{CN}(0, 1)$ is a white Gaussian process, $\sigma^2(r)$ is the conditional variance of the GARCH process and k , α_1 and β_1 are constant process coefficients.

The idea is to develop an algorithm to update a conditional variance estimate at each instant r , treating $\sigma^2(r)$ as a state and (2) as the plant equation of a discrete-time dynamic system. However, the process coefficients are unknown, leading to a linear system with unknown parameters and we tackle the problem augmenting the state to include the parameters, turning the problem non-linear [29].

Thus, the augmented state, consisting of the basic state $\sigma^2(r)$ and the parameters k , α_1 and β_1 , is

$$\mathbf{x}(r) \triangleq [x_1(r)x_2(r)x_3(r)x_4(r)]^T = [\sigma^2(r)k\alpha_1\beta_1]^T. \quad (3)$$

We can then write the system as

$$\mathbf{x}(r+1) = \mathbf{f}(\mathbf{x}(r)), \quad (4)$$

where $\mathbf{f}(\mathbf{x}(r))$ is a vector function whose first element corresponds to the non-linear plant equation which from (2) is $f_1(\mathbf{x}(r)) = x_2(r) + x_3(r)x_1(r) + x_4(r)|c(r)|^2$, and assuming the

parameters are time invariant, their dynamic equations are $f_i(\mathbf{x}(r)) = x_i(r)$, $i = 2, 3, 4$.

The observations $y(r)$ consist of the squared magnitude of the process samples, i.e. $|c(r)|^2 = \sigma^2(r)|z(r)|^2$ thus,

$$y(r) = |z(r)|^2 x_1(r) = [v(r) + m_2]x_1(r) = h(x_1(r), v(r)), \quad (5)$$

where $v(r) = |z(r)|^2 - m_2$ with $m_2 = \mathbb{E}\{|z(r)|^2\}$. Note that $|z(r)|^2$ plays the role of a multiplicative noise.

2.2 Approximation of the non-linear problem

It is not feasible to obtain the optimal filter that minimises the MSE to estimate the states $\mathbf{x}(r)$, i.e. the conditional mean $\mathbb{E}\{\mathbf{x}(r)|Y^r\}$, where

$$Y^r \triangleq \{y(i), i \leq r\}, \quad (6)$$

denotes the information set available at time r . Then, as in the EKF, an estimator will be presented using series expansion of the non-linear dynamic equation.

Assuming that at time r an estimate of the state, $\hat{\mathbf{x}}(r|r)$, exists, i.e. an approximation of the conditional mean given by

$$\hat{\mathbf{x}}(r|r) \simeq \mathbb{E}\{\mathbf{x}(r)|Y^r\}. \quad (7)$$

Then to obtain the EKF, the non-linear function in (4) is expanded in a Taylor series around the latest estimate $\hat{\mathbf{x}}(r|r)$ including terms up to second order [29]

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\simeq \mathbf{f}(\hat{\mathbf{x}}(r|r)) + \mathbf{F}(r)[\mathbf{x} - \hat{\mathbf{x}}(r|r)] \\ &+ \frac{1}{2} \sum_{i=1}^4 \mathbf{e}_i [\mathbf{x} - \hat{\mathbf{x}}(r|r)]^T \nabla^2 f_i(r) [\mathbf{x} - \hat{\mathbf{x}}(r|r)], \end{aligned} \quad (8)$$

where \mathbf{e}_i is a vector whose i th-element is one and the rest are zero,

$$\mathbf{F}(r) \triangleq \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{x}}(r|r)}, \quad (9)$$

with

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \nabla f_1^T \\ \nabla f_2^T \\ \nabla f_3^T \\ \nabla f_4^T \end{bmatrix} = \begin{bmatrix} x_3 & 1 & x_1 & |c(r)|^2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

where ∇f_i is the gradient of f_i and $\nabla^2 f_i(r)$ is the Hessian of f_i given by

$$\nabla^2 f_i(r) \triangleq \left. \frac{\partial^2 f_i}{\partial \mathbf{x} \partial \mathbf{x}^T} \right|_{\mathbf{x} = \hat{\mathbf{x}}(r|r)}. \quad (11)$$

Taking into account (4), it is easily obtained that

$$\nabla^2 f_i(r) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (12)$$

and $\nabla^2 f_i(r) = 0$, $i = 2, 3, 4$. Then (8) reduces to

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\simeq \mathbf{f}(\hat{\mathbf{x}}(r|r)) + \mathbf{F}(r)[\mathbf{x} - \hat{\mathbf{x}}(r|r)] \\ &+ \frac{1}{2} \mathbf{e}_1 [\mathbf{x} - \hat{\mathbf{x}}(r|r)]^T \nabla^2 f_1(r) [\mathbf{x} - \hat{\mathbf{x}}(r|r)], \end{aligned} \quad (13)$$

where the argument in the Hessian was omitted because it is constant.

The observation (5) is linearised to first-order around $\hat{x}_i(r|r)$ and the mean value of the noise $\mathbb{E}\{v(r)\} = 0$. Dropping the time index, leads to

$$h(x, v) \simeq m_2 x + \hat{x}_i(r|r)v. \quad (14)$$

From (13) and (14), the dynamic and observation equations can be written as

$$\begin{aligned} x(r+1) &= f(\hat{x}(r|r)) + F(r)\Delta x(r|r) \\ &+ \frac{1}{2}e_i \Delta x^T(r|r) \nabla^2 f_i \Delta x(r|r) + w(r) \end{aligned} \quad (15)$$

$$y(r) = h^T x(r) + \hat{x}_i(r|r)v(r), \quad (16)$$

where $h^T = [1000]$, $\Delta x(r|r) = x(r) - \hat{x}(r|r)$, $w(r)$ and $v(r)$ are the process noise and the observation noise, respectively. The noise sequence $w(r)$ is assumed to be zero-mean white and Gaussian, with covariance $\mathbb{E}\{w(r)w^T(r)\} = Q(r)$. Given that $z(r)$ is Gaussian, the sequence $v(r)$ defined in (5) has a Chi-squared distribution with 2 degrees of freedom with its mean shifted to 0. The noise sequences are assumed to be mutually independent.

The first component of the process noise, $w_1(r)$, is included to compensate errors introduced by the linearisation. Additionally, the last three components of the process noise, associated with the coefficients to be estimated, are artificial noise sequences preventing the filter-calculated variances of the parameter estimates from converging to zero and providing adaptivity. The EKF is not an optimal estimation procedure, and in general it will not yield consistent estimates of the parameters, i.e. the estimates will not always converge to the true values. To maintain adaptability, it is desirable that the covariance does not converge to zero, otherwise it may lead to estimation errors much larger than the filter-calculated covariance [30].

2.3 EKF algorithm

We follow the approach of [30] for discrete-time linear systems, to describe the EKF for (15) and (16). Notice two distinctive points: (i) the quadratic dynamic equation differs from the standard EKF linearised to order 1; (ii) the measurement noise is Chi-squared distributed and appears multiplied by a state dependent gain.

For a filtered state estimate or state prediction at time r , $\hat{x}(r|k)$, the estimation error is defined as

$$\Delta x(r|k) \triangleq x(r) - \hat{x}(r|k) \quad k \leq r, \quad (17)$$

and the MSE matrix of $x(r)$ given the data Y^k is

$$P(r|k) \triangleq \mathbb{E}\{\Delta x(r|k)\Delta x^T(r|k) | Y^k\}. \quad (18)$$

Thus, the goal is to get a recursion that allows to update the filtered state estimate at $r+1$ and its MSE estimate matrix. The set of EKF equations can be obtained from the static estimation equations for the LMMSE estimator [30]. The recursion obtained in this way is optimal if the problem is linear, and if process and measurement noises and the initial conditions are Gaussian and independent among themselves. In our problem, it leads to a suboptimal solution, i.e. $\hat{x}(r+1|r+1) \simeq \mathbb{E}\{x(r+1) | Y^{r+1}\}$ due to the linearisation of the non-linear model and observation equation, and its particular noise distribution.

Table 1 summarises the algorithm steps of the proposed EKF, once $\hat{x}(0|0)$ and $P(0|0)$ have been initialised, for $r \geq 0$.

2.4 UKF and CKF algorithms

We follow the approach in [29] to implement the UKF algorithm, based on the system (4) and (5). To propagate the state estimate

and error covariance matrix in the time update stage, the $2N+1$ sigma points $c_i(r+1/r+1)$ are chosen as [26]

$$\begin{aligned} \chi_0(r+1|r+1) &= \hat{x}(r+1|r+1) \\ \chi_i(r+1|r+1) &= \hat{x}(r+1|r+1) + \sqrt{N+\kappa}[\Sigma(r+1|r+1)]_i \\ \chi_{i+N}(r+1|r+1) &= \hat{x}(r+1|r+1) - \sqrt{N+\kappa}[\Sigma(r+1|r+1)]_i, \end{aligned} \quad (19)$$

for $i = 1, \dots, N$, where $\kappa \in \mathbb{R}$, N is the state dimension, $[\Sigma(r+1|r+1)]_i$ is the i th row of $\Sigma(r+1|r+1)$, with $P(r+1|r+1) = \Sigma^T(r+1|r+1)\Sigma(r+1|r+1)$. The weights ω_i associated with the sigma points are

$$\begin{aligned} \omega_0 &= \kappa/(N+\kappa) \\ \omega_i &= 0.5/(N+\kappa) \quad i = 1, \dots, 2N. \end{aligned} \quad (20)$$

To propagate the observation prediction and its covariance in the measurement update stage, the $2N+1$ sigma points $\chi_i(r+1/r)$ are computed with a set of equations analogous to (19), using $\hat{x}(r+1|r)$ and $\Sigma(r+1|r)$ instead of $\hat{x}(r+1|r+1)$ and $\Sigma(r+1|r+1)$ respectively, where

$$P(r+1|r) = \Sigma^T(r+1|r)\Sigma(r+1|r).$$

The CKF also works in two steps: the time update and the measurement update. We use the CKF algorithm detailed in Appendix A of [31], where the cubature-point set $\{\xi_i, \omega_i\}$ used to numerically compute the integrals involved is given by

$$\begin{aligned} \xi_i &= \sqrt{N}[1]_i \quad i = 0, \dots, 2N-1 \\ \omega_i &= 0.5/N \quad i = 0, \dots, 2N-1, \end{aligned} \quad (21)$$

where $[1]_i$ is a vector of $N \times 1$ whose element $i \bmod(N) + 1$ is 1 for $i = 0, \dots, N-1$ and -1 for $i = N, \dots, 2N-1$ and the rest are zero.

3 Numerical simulations

A series of Monte-Carlo simulations were conducted in order to verify the correct operation of the filters and compare their performances. To generate the synthetic data, we set the values of the coefficients to $k = 0.5$, $\alpha_1 = 0.2$ and $\beta_1 = 0.3$. With these coefficients the GARCH process satisfies the stationarity condition [16], $\alpha_1 + \beta_1 < 1$, and the unconditional variance of the process results $k/(1 - \alpha_1 - \beta_1) = 1$. In all cases, we fix the initial value of the conditional variance state equal to the square of the first data sample, we fix the second state as a random fraction of the square of the first data sample, and we choose the third and fourth states randomly but within the corresponding restrictions for the respective coefficients, which are that $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$ [16]. To choose the initial MSE state estimation matrix,

Table 1 Algorithm steps of the EKF ($\gamma = m_4 - m_2^2$)

Step	Compute
1. system matrix	$F(r-1)$ through (9)
2. state prediction	$\hat{x}(r+1 r) = f(\hat{x}(r r)) + \frac{1}{2}e_i \text{tr}\{\nabla^2 f_i P(r r)\}$
3. state prediction MSE	$P(r+1 r) = F(r)P(r r)F^T(r) + Q(r) + \frac{1}{2}e_i e_i^T \text{tr}\{\nabla^2 f_i P(r r)\nabla^2 f_i P(r r)\}$
4. innovations	$\Delta y(r+1 r) = y(r+1) - h^T \hat{x}(r+1 r)$
5. obs. prediction MSE	$s(r+1 r) = h^T P(r+1 r)h + \gamma \hat{x}_i^2(r r)$
6. gain	$K(r+1) = P(r+1 r)h s^{-1}(r+1 r)$
7. state estimation	$\hat{x}(r+1 r+1) = \hat{x}(r+1 r) + K(r+1)\Delta y(r+1 r)$
8. state estimation MSE	$P(r+1 r+1) = [I - K(r+1)h^T]P(r+1 r)$

we assume that the initial state is a normally distributed random variable with known mean, such that $\Delta \mathbf{x}(0|0)^T \mathbf{P}(0|0)^{-1} \Delta \mathbf{x}(0|0)$ is Chi-square distributed with 4 degrees of freedom. Then we fix $\mathbf{P}(0|0)$ in order to satisfy the test $\Delta \mathbf{x}(0|0)^T \mathbf{P}(0|0)^{-1} \Delta \mathbf{x}(0|0) < a$ where a is the upper limit of the 95% confidence region from the Chi-square distribution [30]. The simulations are performed for $R = 1 \times 10^5$ instants or range cells. Since $z(r) \sim \mathcal{CN}(0, 1)$, then $\gamma = m_4 - m_2^2 = 1$. We adjust the covariance of the noise by means of simulations to make the filter consistent, with an acceptable rate of convergence [30, 32]. The covariance matrix is assumed diagonal. Initially, we fixed the standard deviations of the process noise as a few percent of the guessed values of the parameters (especial 10% of the square of the first data sample for the first two components and, based on the constraint $\alpha_1 + \beta_1 < 1$, 0.1 for the last two components). We simulate the system and we monitor the errors that are defined in the filters consistency analysis of the next sections. Finally, we adjust the noise variances until the filter yields estimation errors commensurate with the calculated MSE state estimation matrix.

Fig. 1a shows the true conditional variance of the GARCH process and the values estimated by the three filters for the same run. In Fig. 1b, we present the estimation errors, in order to appreciate the differences between the true and estimated value of the conditional variance. Note that once the transient finished, the error remains $< 10\%$ most of the time. Similarly, Fig. 2 shows the true and the estimated value of the process coefficients for the same run and in Table 2 the mean and the estimation root-mean-square error (RMSE) of the coefficient estimates are presented for 5000 and 20,000 samples. All the coefficient estimates present the same behaviour. A transient at the beginning and then a stationary behaviour, where they vary slightly around a mean value. The k and α_i estimates of the UKF and of the CKF have a higher bias than the EKF. On the other hand, for α_i estimates the EKF shows a longer transient than the UKF and CKF.

The mean of the conditional variance of a GARCH process is equal to the unconditional variance of the process, i.e. $\mathbb{E}\{[c(r)]^2\} = \mathbb{E}\{\sigma^2(r)\}$, because $\sigma^2(r)$ and $z(r)$ are independent and $\mathbb{E}\{[z(r)]^2\} = 1$. Thus, a measure of the unconditional variance can be also used to verify the filter evolution. Fig. 3 shows the estimated values of the unconditional variance obtained in the simulation by means of the time average, computed as

$$\langle \theta(i) \rangle_{i,r} = \frac{1}{r} \sum_{i=1}^r \theta(i), \quad (22)$$

where $\theta(i) = y(i) = [c(i)]^2$ when directly estimating from the data, and $\theta(i) = \hat{x}_i(i|i)$ when estimated from the state of each filter. Note that once the transient disappears all the estimates converge approximately to the same value and within a 1% of error with respect to the true value.

These results show that all filters exhibit a similar response. In the next two subsections, we present an analysis of the consistency of the proposed filters and of the statistical properties of the innovations.

3.1 Filter consistency

An estimator of a deterministic parameter is said to be consistent if the estimate converges to the true value of the parameter in some stochastic sense [33]. This implies that as the information about the parameter increases in the estimation procedure, the uncertainty about its true value reduces asymptotically to zero. This sort of convergence cannot be guaranteed in non-linear filtering procedures. In the present problem, state estimation will be considered consistent (i) when the estimates are unbiased and (ii) they have a MSE matrix approaching that calculated by the filter, i.e. if the state estimation errors satisfy the two moment conditions [30]

$$\mathbb{E}\{\Delta \mathbf{x}(r|r)\} = 0, \quad (23)$$

$$\mathbb{E}\{\Delta \mathbf{x}(r|r) \Delta \mathbf{x}^T(r|r)\} = \mathbf{P}(r|r), \quad (24)$$

despite the inherent approximations. We will check consistency by means of Monte Carlo simulations.

To verify the first consistency condition (23), we evaluated the time average $\langle e_i(r) \rangle_{i,M}$ of the state estimation errors, $e_i(r) = \theta_i(r) - \hat{\theta}_i(r)$, from $M = 50$ independent estimates realisations, where $\theta_i(r)$ and $\hat{\theta}_i(r)$ denote the true and the estimated parameter values respectively for the i -th run. In Fig. 4, we show the average of the state estimation errors, which gives an idea of the estimation accuracy. A small bias in the estimation of the coefficients k and α_i can be seen in these results for the three filters. However, in the case of the CKF the bias is slightly higher than the bias of the EKF and of the UKF. This is also noticeable in Figs. 2a and b.

To test the second condition (24), we computed the RMS of each component of the estimation errors. The RMS errors were obtained from the same $M = 50$ runs above and they were computed as $\sqrt{\langle e_i^2(r) \rangle_{i,M}}$. Fig. 5 presents the RMS estimation error for each state, and the RMS error predicted by each filter, $\sqrt{[\mathbf{P}(r|r)]_{ii}}$, $i = 1, 2, 3, 4$, for the single run used to get the estimated values shown in Figs. 1a and 2. Note that the RMSE of the three filters are of the same order of the corresponding $\sqrt{[\mathbf{P}(r|r)]_{ii}}$. This is the expected behaviour in order to satisfy the second consistency condition (24). However, for the CKF the RMSE of the coefficients slightly depart from their corresponding $\sqrt{[\mathbf{P}(r|r)]_{ii}}$. In view of the results of Fig. 5, the mean of the state estimation errors do not depart from zero more than a standard deviation. Thus, we can argue from a practical point of view that the filters satisfy the consistency conditions.

3.2 Innovation statistics

Another criterion often used to check the proper behaviour of the filter is that the innovations are expected to be uncorrelated.

The previous test assumes that M independent runs have been made. However, assuming ergodicity of the innovation sequence, correlation is estimated from a single run, using range (or time) averages. Then the range-average autocorrelation

$$\begin{aligned} \bar{\rho}_{rr}(l, r) &= \frac{\sum_{i=r}^{r+K-1} \Delta y(i|i-1) \Delta y(i+l|i+l-1)}{\sqrt{\sum_{i=r}^{r+K-1} \Delta y^2(i|i-1) \sum_{i=r}^{r+K-1} \Delta y^2(i+l|i+l-1)}}, \end{aligned} \quad (25)$$

can be used as the statistic in a real-time implementation to verify the innovation property. Fig. 6 shows $\bar{\rho}_{rr}(l, r)$ for the run used to get the estimated values showed in Figs. 1a and 2, with a range window of $K = 100$ samples. Similar results hold for larger values of l . These results present a low variability and are close to 0 for the three filters, as expected.

4 Performance analysis

In this section, the filters are tested on real sea clutter data to evaluate their performance in a realistic scenario. The data were collected by the McMaster University IPIX radar, at the Osborne Head Gunnery Range, Dartmouth, Nova Scotia, Canada [4], and are available on-line. Specifically, the used data were recorded on 10 November 1993 at 00:34:24 a.m. which corresponds to the dataset *stare6*. The IPIX radar has full polarimetric information, shown results correspond to vertical polarisation only. The dataset corresponds to inhomogeneous sea clutter without a target. The height of the sea waves was of ~ 0.9 m. The fast time or range dimension consists of $R = 68$ samples, the sampling interval is 15 m and the radar range resolution is 30 m. The number of transmitted pulses, i.e. the number of samples in the slow time dimension, is $T = 8192$, with a pulse repetition frequency of 500 Hz.

The GARCH model for the clutter proposed in [15] considers the realisations in range. Hence, it is necessary to rearrange the data because the number of range cells in each pulse is not enough to extinguish the transient and obtain a reasonable estimation error. To solve this problem, we concatenated different pulses in a single vector. Since the coherence time of the environment is greater than the pulse repetition interval, the pulse samples for a given range cell are correlated. To avoid this, we do not take consecutive pulses. Fig. 7 shows the estimated autocorrelation function for the data in the slow time dimension. It can be seen after 20 ms, or equivalently after 10 pulses, the autocorrelation function is approximately zero; in other words, samples from two pulses

separated a time interval >20 ms are not correlated. Hence, to form the data vector we take 1 in 10 pulses. Note that such rearrangement can be done in real time and it does not become a limitation in practical scenarios.

The initial conditions are chosen in the same way as described for the numerical simulations. The covariance of the noise is chosen to have an intermediate value as a trade-off between convergence rate and estimates variability.

We compare the estimates obtained with the three versions of Kalman filters to the results obtained by a conventional GARCH

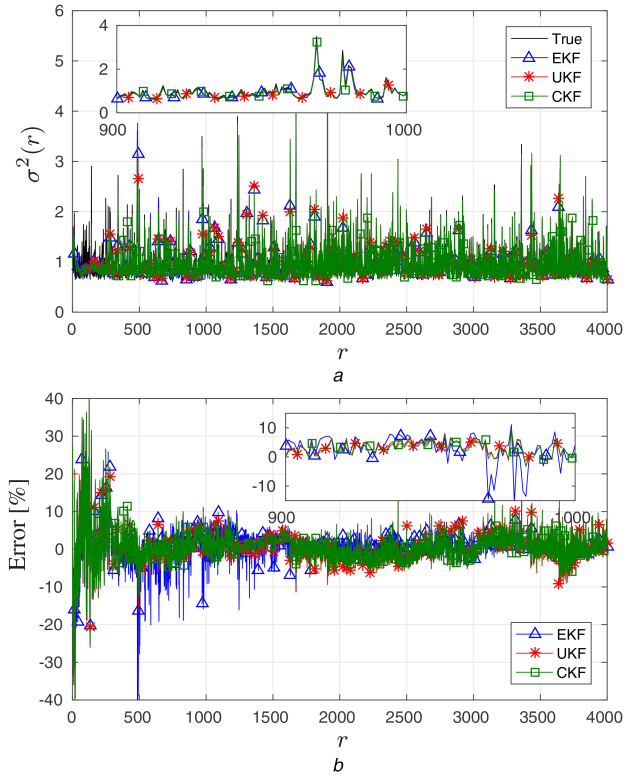


Fig. 1 Conditional variance
(a) Estimated values, (b) Estimation error

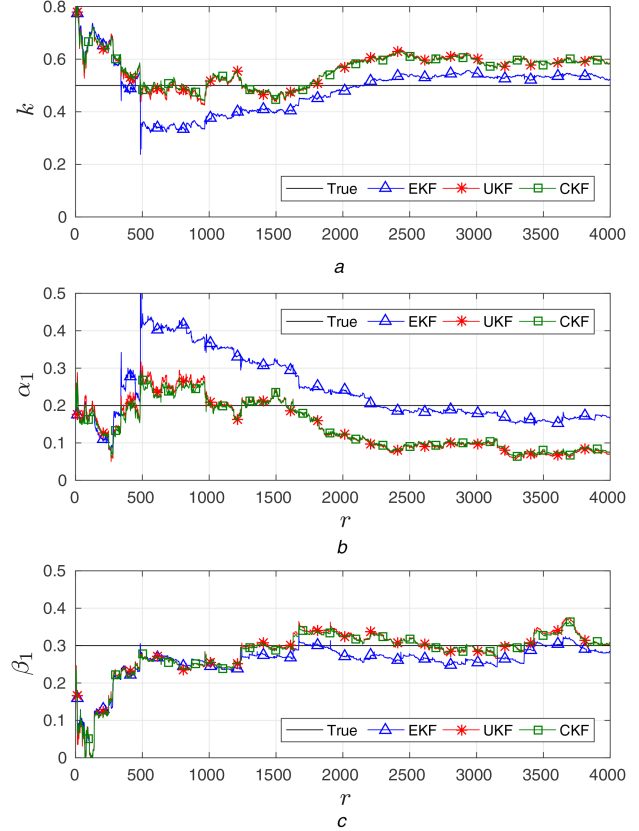


Fig. 2 Estimated values of the process coefficients
(a) Coefficient k , (b) Coefficient α_1 , (c) Coefficient β_1

Table 2 Mean and RMSE of the coefficient estimates

Filter	Coefficient	5000 samples		20,000 samples	
		Mean	RMSE	Mean	RMSE
EKF	k	0.50	0.08	0.52	0.05
	α_1	0.22	0.09	0.17	0.07
	β_1	0.27	0.06	0.29	0.04
UKF	k	0.56	0.09	0.54	0.06
	α_1	0.13	0.09	0.12	0.09
	β_1	0.29	0.06	0.32	0.06
CKF	k	0.57	0.08	0.54	0.06
	α_1	0.13	0.09	0.13	0.08
	β_1	0.29	0.06	0.32	0.06

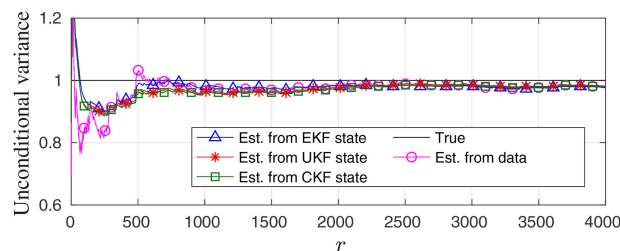


Fig. 3 Unconditional variance of the GARCH process

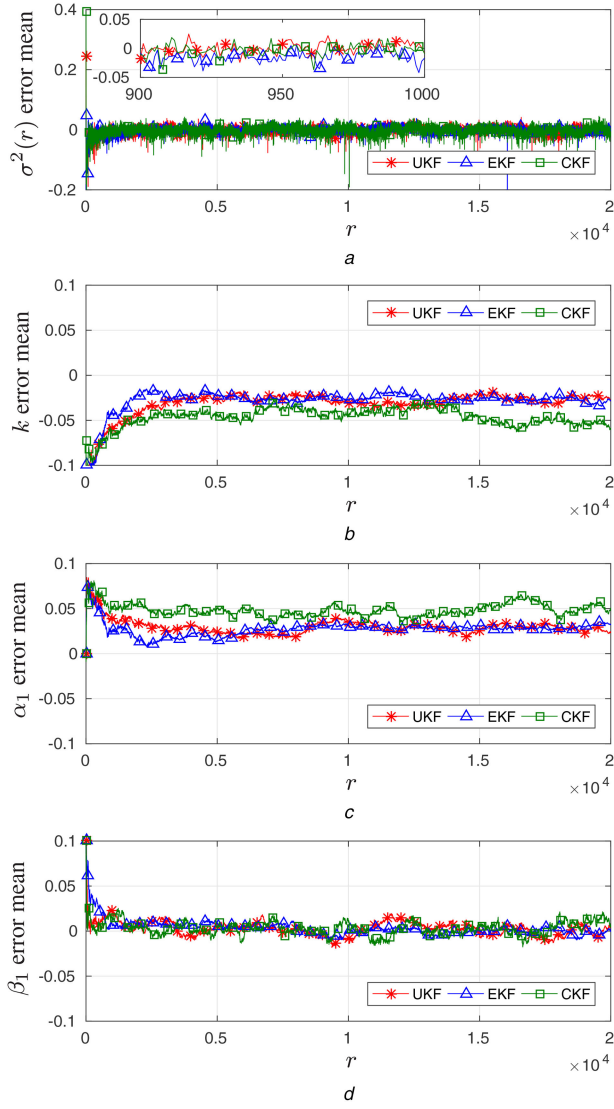


Fig. 4 Mean of the state estimation error from 50 Monte Carlo runs
(a) Conditional variance $\sigma^2(r)$, (b) Coefficient k , (c) Coefficient α_1 , (d) Coefficient β_1

coefficient estimation method, the QMLE. A detailed description of this method can be found in [15].

The coefficients estimation by the QMLE method is performed using first a variable window which was increased to include up to 103 realisations, in all cases separated by 10 pulses, as explained above. Each pulse has $R = 68$ samples, giving a total of 7004 samples, enough to achieve an acceptable estimation error [15].

The QMLE uses all the data inside the processing batch while the Kalman filters only use data up to the filtering point. Thus, the resulting estimates of the QMLE will present a lower variability, especially when filtering did not advance for long enough time. To make a fair comparison, we combine each of the proposed Kalman filters with their smoothed versions so that they also use all the available data. Once the filtering has been performed and $\hat{x}(r|r)$, $\hat{x}(r+1|r)$, $P(r|r)$ and $P(r+1|r)$ for $r = 0, 1, \dots, N-1$ have been stored, the smoothing requires a backward iteration that consists of the steps presented in Table 3 for $r = N-1, \dots, 1, 0$ [30].

4.1 Results

We ran all the estimation methods in the described conditions. Fig. 8a shows the estimates of the conditional variance obtained by means of (2) with the QMLE estimated coefficients, and by the EKF, the UKF and the CKF after smoothing. In Fig. 8b, we present the time average and the RMS value of the difference between the conditional variance approximations obtained with each Kalman filter and the conditional variance computed with the QMLE. Fig. 9 shows the coefficients estimated through QMLE and the

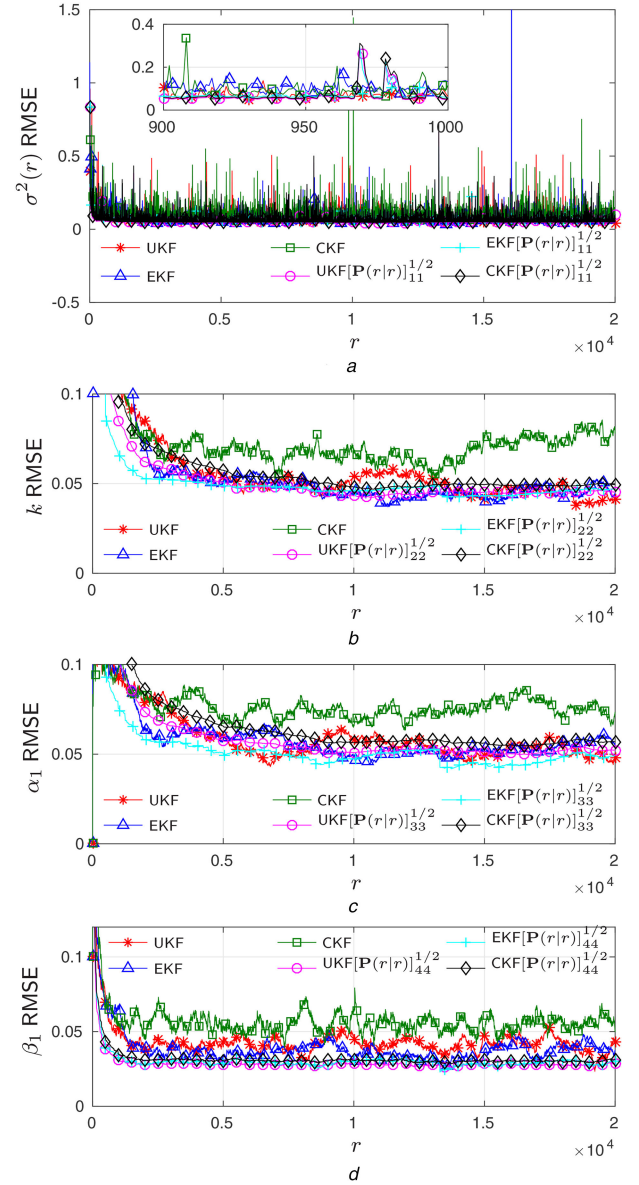


Fig. 5 RMSE of the state estimation from 50 Monte Carlo runs
(a) Conditional variance $\sigma^2(r)$, (b) Coefficient k , (c) Coefficient α_1 , (d) Coefficient β_1

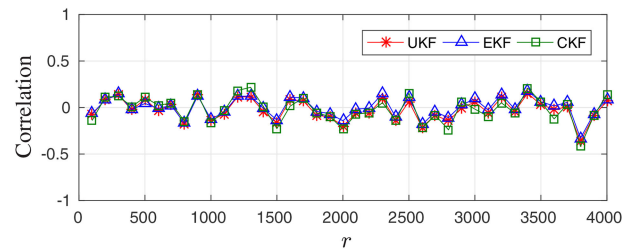


Fig. 6 Innovation autocorrelation for $l = 1$

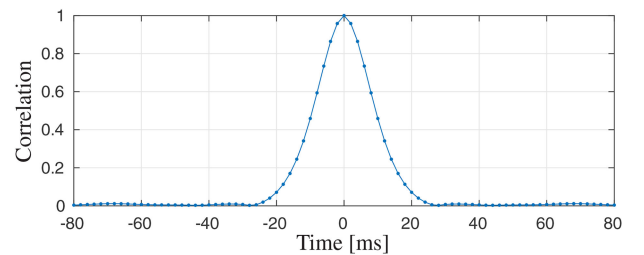


Fig. 7 Autocorrelation function estimate of the IPIX radar stare6 data set

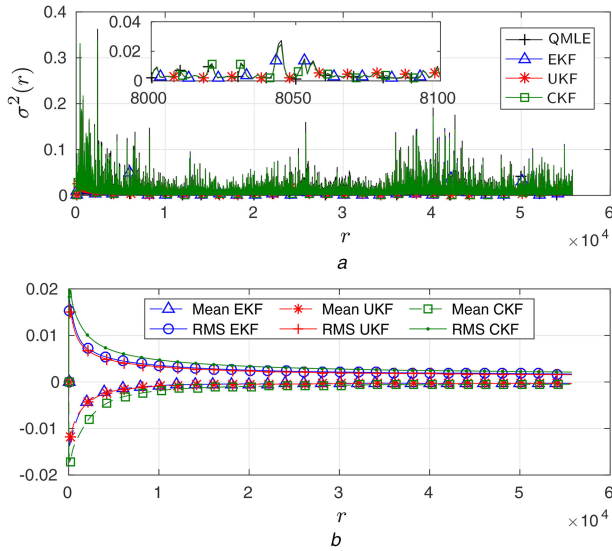


Fig. 8 Conditional variance
(a) Estimated values, (b) Estimated difference

Kalman filters. Despite the slight differences, they are all of the same order. In Fig. 10, we present the unconditional variance estimated as the sample average from the square magnitude of the data; from the conditional variance obtained by means of (2) with the coefficients estimated by QMLE; and from the conditional variance estimated by the smoothed versions of the EKF, the UKF and the CKF. While the transient is longer in the CKF case, and shorter in the data and the QMLE cases, all the unconditional variance estimates are of the same order and tend approximately to the same value after the transient. Finally, Fig. 11 shows the real-time innovation correlation. It is usually ~ 0.15 instead of 0 as expected. An explanation for this behaviour is that the data are correlated between adjacent range cells, since the sampling interval is 15 m and the radar resolution is 30 m.

5 Discussion and conclusion

We presented three algorithms to sequentially estimate the conditional variance of a GARCH process. The estimation procedures are based on non-linear Kalman filters, the EKF, the UKF and the CKF. The GARCH process presents a better characterisation of the sea clutter statistics than most of the time invariant distributions [15]. The detector based on the GARCH clutter model is an adaptive detection algorithm with a threshold scaled by the process conditional variance. The proposed filters aim at updating the conditional variance at each decision instant.

Based on a state model for a GARCH process, we derived the EKF algorithm using a linearisation up to second-order terms. By means of Monte-Carlo simulations we numerically evaluated the performance of the estimation methods. We also analysed the consistency of the filters and the statistics of the innovations using synthetic data. The EKF and UKF filters present a better performance than the CKF from the consistency viewpoint. However, all three have acceptable results, with only a small bias in the estimation of two of the process coefficients. One possible reason for this behaviour is the observability of the state model, following the approach presented in [34]. The approach suggests that a primary cause of inconsistency occurs when the dimension of the observable space of the linearised model is larger than that from the non-linear system. A possible next step is to study the consistency of the proposed GARCH parameter estimators based on the Kalman filters from the perspective of observability. However, in [15] we have shown that the GARCH detector is robust with respect to changes in the parameters.

Regarding the results with real clutter data, the QMLE method seems to have smaller variability in the coefficients estimates than the non-linear Kalman filters. However, this is due to the Kalman filters use only past data immediately before the filtering point while QMLE uses the whole data batch. In order to compare both

Table 3 Smoother Kalman filter algorithm steps

Step	Compute
1. smoother gain	$C(r) = P(r r)F^T(r)P^{-1}(r+1 r)$
2. smoothed state	$\hat{x}(r N) = \hat{x}(r r) + C(r)[\hat{x}(r+1 N) - \hat{x}(r+1 r)]$
3. covariance of the smoothed state	$P(r N) = P(r r) + C(r)[P(r+1 N) - P(r+1 r)]C^T(r)$

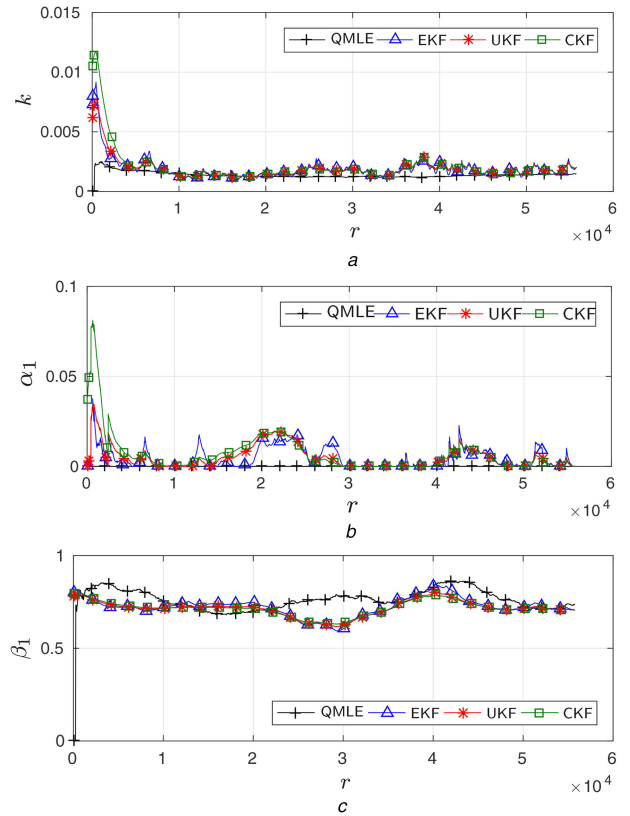


Fig. 9 Estimated values of the process coefficients from data
(a) k , (b) α_1 , (c) β_1

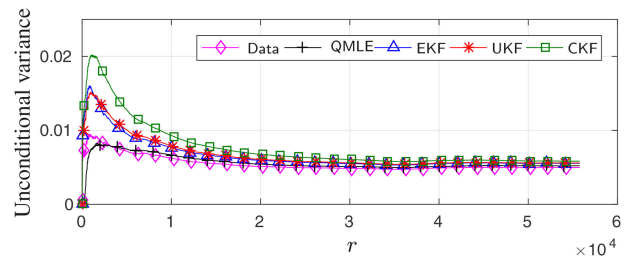


Fig. 10 Estimated values of the data unconditional variance

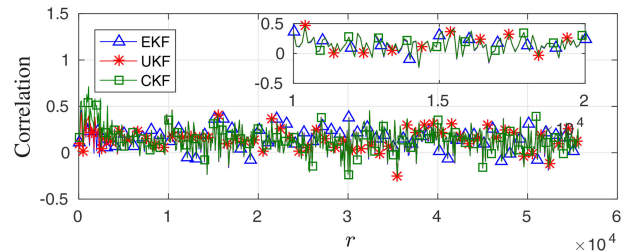


Fig. 11 Real time autocorrelation of the innovation for $l = 1$

approaches, smoothed versions for all three Kalman filters were included obtaining similar performance to the QMLE.

It is worth pointing out that the proposed filters have a limitation. The covariance matrix of the process noise is adjusted in order to get errors commensurate with the calculated MSE state estimation matrix, with an acceptable rate of convergence.

However, one of the sources of uncertainty while working with Kalman filters, even with linear models, is due to the unknown a priori statistics of the noise. This suggests modifying the filters using an adaptive algorithm to estimate the noise covariance matrix. A possible approach for the EKF is the extended expectation-maximisation proposed in [35], which avoids the computational efforts involved in computing the gradients of the objective function in the maximum likelihood framework. On the other hand, adaptive algorithms for the UKF and the CKF proposed in this work might be developed based on the robust Masreliez–Martin UKF [36].

The great advantage of the Kalman filters is the computational load reduction when compared to QMLE. It is not straightforward to make a detailed analysis of the computational burden, especially in the case of the QMLE due to its iterative nature. However, all estimation methods were run in the same desktop computer exclusively dedicated to this task. In this condition, the EKF and the smoother EKF evaluated the 55,759 state estimates in ~4 and 7 s, respectively, and the smoothed versions of the UKF and of the CKF spend ~15 s to evaluate the 55,759 state estimates. Meanwhile, the QMLE spent 450 s to obtain 816 coefficients estimates, and this does not include the computation of the conditional variance. Clearly, the proposed method presents a significant advantage in the computational load. To further reduce the computational complexity of the EKF, the decoupled EKF could be used [37], which has shown good performance in multilayer neural network training.

Finally, it might be of interest to investigate the use of the Kalman filters to GARCH clutter parameter estimation in conjunction with the radar detector derived in [15] and to compare with the detection performance when the standard algorithm for GARCH parameter estimation is used.

6 Acknowledgments

This work was supported by the Universidad Nacional de Cuyo (UNCuyo) C020 3853/16, the ANPCyT PICT PICT2014-1232, the UNLP 11-I-209 and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

7 References

- [1] Kelly, E.J.: 'An adaptive detection algorithm', *IEEE Trans. Aerosp. Electron. Syst.*, 1986, **AES-22**, (1), pp. 115–127
- [2] Bose, S., Steinhardt, A.: 'A maximal invariant framework for adaptive detection with structured and unstructured covariance matrices', *IEEE Trans. Signal Process.*, 1995, **43**, (9), pp. 2164–2175
- [3] Sheikhi, A., Nayeibi, M.M., Aref, M.R.: 'Adaptive detection algorithm for radar signals in autoregressive interference', *Proc. Inst. Electr. Eng.-Radar Sonar Navig.*, 1998, **145**, (5), pp. 309–314
- [4] Haykin, S., Krasnor, C., Nohara, T., *et al.*: 'A coherent dual-polarized radar for studying the ocean environment', *IEEE Trans. Geosci. Remote Sens.*, 1991, **29**, (1), pp. 189–191
- [5] Farina, A., Gini, F., Greco, M.V., *et al.*: 'High resolution sea clutter data: statistical analysis of recorded live data', *Proc. Inst. Electr. Eng.-Radar Sonar Navig.*, 1997, **144**, (3), pp. 121–130
- [6] Gini, F., Greco, M.V., Diani, M., *et al.*: 'Performance analysis of two adaptive radar detectors against non-Gaussian real sea clutter data', *IEEE Trans. Aerosp. Electron. Syst.*, 2000, **36**, (4), pp. 1429–1439
- [7] Xu, S., Shui, P., Yan, X.: 'Non-coherent detection of radar target in heavy-tailed sea clutter using bi-window non-linear shrinkage map', *IET Signal Process.*, 2016, **10**, (9), pp. 1031–1039
- [8] Shnidman, D.A.: 'Generalized radar clutter model', *IEEE Trans. Aerosp. Electron. Syst.*, 1999, **35**, (3), pp. 857–865
- [9] Sangston, K., Gerlach, K.: 'Coherent detection of radar targets in a non-Gaussian background', *IEEE Trans. Aerosp. Electron. Syst.*, 1994, **30**, (2), pp. 330–340
- [10] Sangston, K.J., Gini, F., Greco, M.S.: 'Coherent radar target detection in heavy-tailed compound-Gaussian clutter', *IEEE Trans. Aerosp. Electron. Syst.*, 2012, **48**, (1), pp. 64–77
- [11] Shang, X., Song, H.: 'Radar detection based on compound-Gaussian model with inverse gamma texture', *IET Radar Sonar Navig.*, 2011, **5**, (3), pp. 315–321
- [12] Cui, G., Kong, L., Yang, X., *et al.*: 'Distributed target detection with polarimetric MIMO radar in compound-Gaussian clutter', *Digit. Signal Process.*, 2012, **22**, (3), pp. 430–438
- [13] Li, Y., Moran, W., Sira, S.P., *et al.*: 'Monte-Carlo based estimation methods for rapidly-varying sea clutter'. Proc. 13th IEEE Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009, DSP/SPE 2009, Marco Island, FL, 2009, pp. 256–261
- [14] Gini, F., Greco, M.V.: 'Suboptimum approach to adaptive coherent radar detection in compound-Gaussian clutter', *IEEE Trans. Aerosp. Electron. Syst.*, 1999, **35**, (3), pp. 1095–1104
- [15] Pascual, J.P., von Ellenrieder, N., Hurtado, M., *et al.*: 'Radar detection algorithm for GARCH clutter model', *Digit. Signal Process.*, 2013, **23**, (4), pp. 1255–1264
- [16] Bollerslev, T.: 'Generalized autoregressive conditional heteroscedasticity', *J. Econ.*, 1986, **31**, (3), pp. 307–327
- [17] Berkes, I., Horváth, L., Kokoszka, P.: 'GARCH processes: structure and estimation', *Bernoulli*, 2003, **9**, (2), pp. 201–227
- [18] Kristensen, D., Linton, O.: 'A closed-form estimator for the GARCH (1,1) model', *Econ. Theory*, 2006, **22**, (2), pp. 323–337
- [19] Ossandón, S., Bahamonde, N.: 'On the nonlinear estimation of GARCH models using an extended Kalman filter'. Proc. World Congress on Engineering, London, U.K., 2011, vol. 1
- [20] Ossandón, S., Bahamonde, N.: 'A new nonlinear formulation for GARCH models', *C.R. Math.*, 2013, **351**, (5), pp. 235–239
- [21] Pascual, J.P., von Ellenrieder, N., Muravchik, C.H.: 'Conditional variance LMMSE estimator for a GARCH process clutter model'. Proc. 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2014), A Coruña, Spain, 2014, pp. 309–312
- [22] Mousazadeh, S., Cohen, I.: 'AR-GARCH in presence of noise: parameter estimation and its application to voice activity detection', *IEEE Audio Speech Lang. Process.*, 2011, **19**, (4), pp. 916–926
- [23] Mousazadeh, S., Cohen, I.: 'Simultaneous parameter estimation and state smoothing of complex GARCH process in the presence of additive noise', *Signal Process.*, 2010, **90**, (11), pp. 2947–2953
- [24] Ferreira, G., Navarrete, J.P., Rodríguez Cortés, F.J., *et al.*: 'Estimation and prediction of time-varying GARCH models through a state-space representation: a computational approach', *J. Stat. Comput. Simul.*, 2017, **87**, (12), pp. 2430–2449
- [25] Ljung, L., Söderström, T.: 'Theory and practice of recursive identification' (The MIT Press, Cambridge, MA, 1983)
- [26] Julier, S., Uhlmann, J., Durrant-Whyte, H.: 'A new method for the nonlinear transformation of means and covariances in filters and estimators', *IEEE Trans. Autom. Control*, 2000, **45**, (3), pp. 477–482
- [27] Roy, A., Mitra, D.: 'Multi-target trackers using cubature kalman filter for Doppler radar tracking in clutter', *IET Signal Process.*, 2016, **10**, (8), pp. 888–901
- [28] Long, T., Zheng, L., Chen, X., *et al.*: 'Improved probabilistic multi-hypothesis tracker for multiple target tracking with switching attribute states', *IEEE Trans. Signal Process.*, 2011, **59**, (12), pp. 5721–5733
- [29] Simon, D.: 'Optimal state estimation - Kalman, H_∞, and nonlinear approaches' (John Wiley & Sons, New York, 2006)
- [30] Bar-Shalom, Y., Li, X., Kirubarajan, T.: 'Estimation with applications to tracking and navigation' (John Wiley & Sons, New York, 2001)
- [31] Arasaratnam, I., Haykin, S.: 'Cubature Kalman filters', *IEEE Trans. Autom. Control*, 2009, **54**, (6), pp. 1254–1269
- [32] Dhaouadi, R., Mohan, N., Norum, L.: 'Design and implementation of an extended Kalman filter for the state estimation of a permanent magnet synchronous motor', *IEEE Trans. Power Electron.*, 1991, **6**, (3), pp. 491–497
- [33] Kay, S.M.: 'Fundamentals of statistical signal processing, estimation theory' (Prentice Hall, Upper Saddle River, NJ, 1993)
- [34] Huang, G.P., Mourikis, A.I., Roumeliotis, S.I.: 'Analysis and improvement of the consistency of extended kalman filter based SLAM'. Proc. IEEE Int. Conf. Robotics and Automation, Pasadena, CA, USA, 2008, pp. 473–479
- [35] Bavdekar, V.A., Deshpande, A.P., Patwardhan, S.C.: 'Identification of process and measurement noise covariance for state and parameter estimation using extended kalman filter', *J. Process. Control*, 2011, **21**, pp. 585–601
- [36] Li, W., Sun, S., Jia, Y., *et al.*: 'Robust unscented Kalman filter with adaptation of process and measurement noise covariances', *Digit. Signal Process.*, 2016, **48**, pp. 93–103
- [37] Puskorius, G.V., Feldkamp, L.A.: 'Parameter-based kalman filter training: theory and implementation', in Haykin, S., (ed.): 'Kalman filtering and neural networks' (John Wiley & Sons, New York, 2001), pp. 23–67