ISSN 0005-1179, Automation and Remote Control, 2019, Vol. 80, No. 7, pp. 1230–1251. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Avtomatika i Telemekhanika, 2019, No. 7, pp. 61–88.

STOCHASTIC SYSTEMS

Conditionally Minimax Nonlinear Filter and Unscented Kalman Filter: Empirical Analysis and Comparison

A. V. Bosov[*,**,a] **and G. B. Miller**[*,b]

[*] *Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia*
[**] *Moscow Aviation Institute, Moscow, Russia*
e-mail: [a] *abosov@frccsc.ru,* [b] *gmiller@frccsc.ru*

Received November 15, 2018
Revised February 5, 2019
Accepted February 7, 2019

Abstract—We present the results of the analysis and comparison of the properties of two concepts in state filtering problems for nonlinear stochastic dynamic observation systems with discrete time: sigma-point Kalman filter based on a discrete approximation of continuous distributions and conditionally minimax nonlinear filter that implements the conditionally optimal filtering method based on simulation modeling. A brief discussion of the structure and properties of the estimates and justifications of the corresponding algorithms is accompanied by a significant number of model examples illustrating both positive applications and limitations of the efficiency for the estimation procedures. The simplicity and clarity of the considered examples (scalar autonomous regressions in the state equation and linear observations) allow us to objectively characterize the considered estimation methods. We propose a new modification of the nonlinear filter that combines the ideas of both considered approaches.

Keywords: nonlinear stochastic observation system, unscented transform, unscented Kalman filter, conditionally optimal filtering, conditionally minimax nonlinear filter, simulation modeling

DOI: 10.1134/S0005117919070026

1. INTRODUCTION

Stochastic filtering, state estimation, and identification of parameters for dynamic observation systems are problems important both for numerous applications and for the fundamental theory; they have attracted considerable research interest over many years. Nonlinear models and filtering algorithms occupy their own place in these studies. Due to the fact that nonlinear systems are extremely difficult to study by general rigorous methods, a certain mathematical freedom is allowed in this area, consisting in the possibility of informal justification of proposed solutions by considerations of reasonableness, engineering explanations, and approximate relations. Results obtained in this way are generally covered by the term "suboptimal filtering." In contrast with fundamental results in the field of optimal stochastic filtering, suboptimal filters have real prospects for practical application; for example, they allow to perform calculations for complex multidimensional models. And although estimates obtained with the help of these methods do not possess guaranteed properties characterizing their quality, there are examples that confirm their practical applicability and usefulness. A simple classic example of a suboptimal algorithm—the extended Kalman filter—is successfully used for a number of nonlinear problems; however, it exhibits divergence in other, less "successful" examples. A detailed review of suboptimal filtering algorithms goes beyond the

scope of this article; here we will focus on only one successful and developed area that has gained significant popularity: unscented Kalman filtering.

The concept of this modification of the Kalman filter (unscented Kalman Filter, UKF), first described in [1], has not undergone significant changes since then and has been successfully used to date, turning into a whole direction of research that finds applications in very different areas. The basic idea of UKF is to approximate continuous distributions of parameters in a nonlinear observation system by discrete ones using a specially chosen "basis," a set of sigma-points. This approximation allows to approximately calculate the moment characteristics of parameters that are used in recurrent equations of a Kalman structure filter. A good overview of the techniques, methods, and modifications related to unscented Kalman filtering algorithms can be found in [2], where significant effort has been made to systematize the results in this field. We also mention separately the work [3], which complements the original method with a solution to the high dimensionality problem (scaled unscented transformation), and also the works [4, 5] that explore the properties of unscented Kalman filter for systems of special type, approaches to filter robustification [6, 7], parameter optimization [8–10], improvements in computational efficiency [11], and generalization to continuous time [12]. The popularity of this research direction has been confirmed by a large number of practical publications devoted to using the UKF and adapting it to new models [13–18].

The heuristic approach, common for the suboptimal filtering direction, has alternatives that, on the one hand, provide a rigorous mathematical justification and, on the other hand, are suitable for widespread practical applications due to simplicity of implementation. Such an alternative is provided by the theory of Pugachev's conditionally optimal filtering and its development in relation to dynamical systems with discrete time, the theory of conditionally minimax nonlinear filtering (CMNF) [19–22]. At present, these results do not have as wide a scope as UKF. The purpose of this work is to provide a qualitative comparison of the concepts underlying these two practically important approaches and to illustrate the results of this analysis by a series of model experiments. A separate task is the development of the conditionally minimax approach based on ideas embodied in the UKF method.

2. BASIC MODEL

Consider a non-linear observation system defined by the following difference equations:

$$\begin{gathered} x_t = \varphi_t^{(1)}(x_{t-1}) + \varphi_t^{(2)}(x_{t-1})\, w_t, \quad t = 1, 2, \ldots, \quad x_0 = \eta, \\ y_t = \psi_t^{(1)}(x_t) + \psi_t^{(2)}(x_t) v_t. \end{gathered} \tag{1}$$

Here $x_t \in \mathbb{R}^p$ is a random process that defines the system state; $y_t \in \mathbb{R}^q$ are indirect observations; w_t and v_t are second-order discrete white noises; the random vector of initial conditions η also has a finite second moment; processes w_t and v_t together with the vector η are mutually independent. We denote their expectations and covariance matrices by $m_w(t), D_w(t), m_v(t), D_v(t), m_\eta, D_\eta$ respectively.

Consider the problem of estimating the state x_t from observations y_τ, $\tau = 1, \ldots, t$, using the mean squared difference as the evaluation criterion for the estimate $\hat{x}_t$:

$$\mathrm{E}\left\{||x_t - \hat{x}_t||^2\right\}.$$

Model (1) is a compromise description of the observation system, suitable for both unscented Kalman and conditionally minimax filters. Moreover, an additional assumption $\psi_t^{(2)}(x) = 1$ and/or requirement that all moments exist for the processes w_t and v_t often proves to be of practical importance. For the purposes of this work, it suffices to note that under constraints on the linear growth rate of functions at infinity and given the existence of second moments for all noises, the

processes defined by (1) are Hilbert processes. Accordingly, under such assumptions the filtering problem under consideration is well-posed and therefore we can speak of its solution, both the optimal solution in the form of the conditional expectation of x_t for all observations as well as various approximations of this solution.

3. METHODS AND APPROACHES

Both concepts for synthesizing nonlinear filtering algorithms that we study in this work are based on two simple static problems.

First, consider the problem whose solution illustrates the essence of the CMNF approach. Consider a random vector $z = \operatorname{col}(x, y)$, $x \in \mathbb{R}^p$, $y \in \mathbb{R}^q$ with known expectation $m_z = \mathrm{E}\{z\} = \operatorname{col}(m_x, m_y)$ and covariance matrix $D_z = \operatorname{cov}(z, z) = \begin{pmatrix} D_x & D_{xy} \\ D_{yx} & D_y \end{pmatrix}$. The distribution function $\mathcal{F}_z$ of vector z is unknown, and it is assumed that $\mathcal{F}_z \in \mathbf{\Phi}(m_z, D_z)$, i.e., it belongs to the class of all probability distributions with mean m_z and covariance matrix D_z. The problem is to find an estimate $\hat{x} = \theta(y)$ of the unobservable vector x given observations y, based on the mean squares criterion and a given uncertainty set $\mathbf{\Phi}$. We treat any Borel-measurable functions $\theta(y)$ as admissible estimators. We denote the estimation accuracy, provided by $\theta(y)$ if vector z has distribution $\mathcal{F}_z$, by $J(\theta, \mathcal{F}_z) = \mathrm{E}\{||x - \hat{x}||^2\}$.

The solution to this problem is given by a saddle point defined by inequalities $J(\theta^*, \mathcal{F}_z) \le J(\theta^*, \mathcal{F}_z^*) \le J(\theta, \mathcal{F}_z^*)$, where $\mathcal{F}_z^*$ is the Gaussian distribution with parameters m_z, D_z, $x^* = \theta^*(y) = D_{xy}D_y^+ y + (m_x - D_{xy}D_y^+ m_y)$ is the best linear estimate x given observations y with respect to the mean squares criterion [21, 22]. At the same time, $J(\theta^*, \mathcal{F}_z^*) = D_x - D_{xy}D_y^+ D_{yx}$. Above and below we denote by $\operatorname{cov}(x, y)$ the covariance matrix of random vectors x and y; by $\mathrm{E}\{x\}$, the expectation of a random vector x; by $+$, the matrix pseudo-inverse operation.

This result implies that when, in the estimation problem defined above, we construct the estimate x by observation y the *justified in the minimax sense* linear estimator is one whose parameters are determined by the moment characteristics of the joint distribution. This result underlies the CMNF method. Namely, the filtering estimate $\hat{x}_t$ of system state x_t is obtained by solving the following minimax problems with respect to the prediction $\tilde{x}_t$ and its correction:

$$\begin{gathered} \tilde{x}_t = \tilde{\theta}_t(\xi_t),\ \tilde{\theta}_t = \operatorname*{argmin}_{\tilde{\boldsymbol{\theta}}_t} \max_{\mathcal{F}_z} \mathrm{E}\left\{ ||x_t - \tilde{\boldsymbol{\theta}}_t(\xi_t)||^2 \right\},\ z = \operatorname{col}(x_t, \xi_t), \\ \hat{x}_t = \tilde{x}_t + \hat{\theta}_t(\zeta_t),\ \hat{\theta}_t = \operatorname*{argmin}_{\hat{\boldsymbol{\theta}}_t} \max_{\mathcal{F}_z} \mathrm{E}\left\{ ||x_t - \tilde{x}_t - \hat{\boldsymbol{\theta}}_t(\zeta_t)||^2 \right\},\ z = \operatorname{col}(x_t - \tilde{x}_t, \zeta_t). \end{gathered} \tag{2}$$

Here and below, $\xi_t = \xi_t(x)$ and $\zeta_t = \zeta_t(x, y)$ are some structural filter functions whose choice constitutes the practical part of the method.

Further, the actual description of the CMNF algorithm is as follows. Suppose that we have $\hat{x}_{t-1}$, the CMNF estimate of state x_{t-1} by observations y_τ, $\tau = 1, \ldots, t-1$. We search for a prediction $\tilde{x}_t$ in the form

$$\tilde{x}_t = F_t \xi_t + f_t, \quad \xi_t = \xi_t(\hat{x}_{t-1}), \quad \xi_t(x) = \varphi_t^{(1)}(x) + \varphi_t^{(2)}(x)\, m_w(t). \tag{3}$$

The prediction structure "by virtue of the system" is determined by the structural function $\xi_t(x)$ and coefficients F_t and f_t (the matrix and vector of the corresponding dimensions), determined by solving problem (2), namely $\tilde{\theta}_t(\xi_t) = F_t \xi_t + f_t$.

We search for the estimate $\hat{x}_t$ of the state x_t as

$$\begin{gathered} \hat{x}_t = \tilde{x}_t + H_t \zeta_t + h_t, \quad \zeta_t = \zeta_t(\tilde{x}_t,\ y_t), \\ \zeta_t(x, y) = y - \psi_t^{(1)}(x) - \psi_t^{(2)}(x)\, m_v(t). \end{gathered} \tag{4}$$

The structure of the correction in the form of a residual is determined by the structural function $\zeta_t(x,y)$ (residual) and coefficients H_t and h_t (matrix and vector of the corresponding dimensions) determined by solving problem (2), namely $\hat{\theta}_t(\zeta_t) = H_t\zeta_t + h_t$.

Solutions of both problems in (2) under the assumption that all necessary second moments are finite obviously exist, but may not be unique. Uniqueness is provided by an additional assumption on the minimum of the Euclidean norm of the solution and, therefore, by using the Moore–Penrose pseudo-inversion operation. Thus, the required coefficients are given in the form

$$\begin{gathered} F_t = \operatorname{cov}(x_t,\xi_t)\operatorname{cov}^{+}(\xi_t,\xi_t), \quad f_t = \mathrm{E}\{x_t\} - F_t\mathrm{E}\{\xi_t\}, \\ H_t = \operatorname{cov}(x_t-\tilde{x}_t,\zeta_t)\operatorname{cov}^{+}(\zeta_t,\zeta_t), \quad h_t = -H_t\mathrm{E}\{\zeta_t\}. \end{gathered} \tag{5}$$

Prediction $\tilde{x}_t$ and filtering estimate $\hat{x}_t$ are unbiased and provide the following estimation quality:

$$\begin{gathered} \tilde{K}_t = \operatorname{cov}(x_t-\tilde{x}_t, x_t-\tilde{x}_t) = \operatorname{cov}(x_t,x_t) - F_t\operatorname{cov}(\xi_t,x_t), \\ \hat{K}_t = \operatorname{cov}(x_t-\hat{x}_t, x_t-\hat{x}_t) = \tilde{K}_t - H_t\operatorname{cov}(\zeta_t, x_t-\tilde{x}_t), \end{gathered} \tag{6}$$

i.e., we ensure non-divergence and "meaningfulness" of the estimates, where "meaningfulness" is understood in the sense that both prediction and estimate have a guaranteed advantage as compared to the trivial estimate, since $F_t\operatorname{cov}(\xi_t,x_t) \geq 0$, $H_t\operatorname{cov}(\zeta_t, x_t-\tilde{x}_t) \geq 0$.

Relations (3)–(5) define a conditionally optimal Pugachev filter, and the CMNF paradigm supplements it with a minimax justification for the filter structure. In addition, an indispensable attribute of CMNF is a practical way to find the coefficients F_t, f_t, H_t and h_t by the Monte Carlo method, i.e., by computer simulation. The filter itself is obtained by replacing in (5) the expectations and covariances by their statistical estimates obtained as a result of computer simulation. For other questions, including existence conditions for the filters, we refer to the work [23], which additionally gives a more detailed overview of the method and directions of its further development. Note that the filter's structural functions, namely prediction $\xi_t(x)$ and correction $\zeta_t(x,y)$, can be defined in a flexible way: the presented structure of prediction "by virtue of the system" and correction in the form of a residual are, of course, only the simplest versions, and depending on the properties of a particular observation system $\xi_t(x)$ and $\zeta_t(x,y)$ can vary significantly.

Note also that, in principle, CMNF can be thought of as *non-linear Kalman filters* or, as we have mentioned in Section 1, as filters of Kalman structure. Let us clarify this by writing down the simplest of such filters: the extended Kalman filter (EKF). Here and below, we do not introduce additional notation but use already existing, for example, $\tilde{x}_t$ and $\hat{x}_t$ to denote the prediction and the filtering estimate, ξ_t for the prediction by virtue of the system," and ζ_t for the residual. The exact algorithms that we mean will be clear from context. For EKF, the relations are as follows:

$$\begin{gathered} \tilde{x}_t = \xi_t, \quad \xi_t = \xi_t(\hat{x}_{t-1}), \quad \xi_t(x) = \varphi_t^{(1)}(x) + \varphi_t^{(2)}(x)m_w(t), \\ \tilde{K}_t = f_t\hat{K}_{t-1}f_t^{\mathrm{T}} + \varphi_t^{(2)}D_w(t)\,\varphi_t^{(2)\mathrm{T}}, \quad f_t = \left.\frac{\partial\varphi_t^{(1)}(x)}{\partial x}\right|x=\tilde{x}_t, \\ \varphi_t^{(2)} = \varphi_t^{(2)}(\tilde{x}_t), \quad \hat{x}_t = \tilde{x}_t + H_t\zeta_t, \quad \zeta_t = y - \psi_t^{(1)}(x) - \psi_t^{(2)}(x)\,m_v(t), \\ H_t = \tilde{K}_t h_t^{\mathrm{T}}\left(h_t\tilde{K}_t h_t^{\mathrm{T}} + \psi_t^{(2)}D_v(t)\,\psi_t^{(2)\mathrm{T}}\right)^{+}, \\ h_t = \left.\frac{\partial\psi_t^{(1)}(x)}{\partial x}\right|x=\tilde{x}_t, \quad \psi_t^{(2)} = \psi_t^{(2)}(\tilde{x}_t), \\ \hat{K}_t = \tilde{K}_t - H_t h_t\tilde{K}_t. \end{gathered} \tag{7}$$

The rest of the notation that we reuse also makes sense. For instance, $f_t,\ h_t$ are some estimates of the moments, that is, more precisely, their approximations obtained by linearization, H_t has the same meaning as in (5), $\tilde{K}_t$ and $\hat{K}_t$ are the covariances of the prediction error and estimation error, only here, unlike (6), this approximation is due to elementary linearization of nonlinearities $\varphi_t^{(1)}(x)$ and $\psi_t^{(1)}(x)$, while CMNF uses exact values or their statistical estimates.

Now we turn to UKF. The Kalman filter structure is preserved in this method, but the main idea is fundamentally different. Namely, the conceptual basis of UKF is the unscented transform that consists of the following. Suppose, as above, that there exists a vector $z = \operatorname{col}(x, y) \in \mathbb{R}^{p+q}$, such that $y = \varphi(x)$. Regarding the distribution $\mathcal{F}_x$ of the vector x we assume that $\mathcal{F}_x \in \mathbf{\Phi}(m_x, D_x)$, i.e., its first two moments are known. Informally, we can say regarding $\mathcal{F}_x$ that this distribution is assumed to be continuous with a form "similar" to Gaussian. Moreover, in many publications on this topic, following the original assumptions [1], the discussion of CT filtering is illustrated by an example of calculation with a Gaussian vector x. Note that the reasoning here, as in the publications directly devoted to this subject, does not pretend to be formal and does not imply mathematical rigor of the assumptions and the results. As in any research on suboptimal estimate, it is the conceptual component that is important—the idea based on empirical considerations and supported by model examples. The transformation in question consists of representing $\mathcal{F}_x$ with its discrete counterpart. Namely, instead of the assumption that x can take arbitrary values, we assume that the range of values of x is constrained to a finite number of sigma-points $x^0, x^1, \ldots, x^{2p}$, each of which is associated with the corresponding probability $W^0, W^1, \ldots, W^{2p}$. Let us consider the relation between the dimension p of the vector x and the number of sigma-points $2p+1$. This is a fundamental point that emphasizes that the number of sigma-points, which should be few, are substantially bounded. Otherwise, choosing an unlimited number of x^i and selecting values of W^i, one can obviously achieve any given accuracy of approximation. The notation introduced above yields simple approximate relations for the moments of the vector y:

$$
\begin{gathered}
m_y \approx \hat{y} = \sum_{i=0}^{2p} W^i y^i, \quad D_y \approx \hat{D}_y = \sum_{i=0}^{2p} W^i (y^i - \hat{y})(y^i - \hat{y})^{\mathrm{T}}, \\
D_{xy} \approx \hat{D}_{xy} = \sum_{i=0}^{2p} W^i (x^i - m_x)(y^i - \hat{y})^{\mathrm{T}}, \quad y^i = \varphi\left(x^i\right).
\end{gathered}
\tag{8}
$$

The proximity of the moments m_y and $\hat{y}$, D_y and $\hat{D}_y$, D_{xy} and $\hat{D}_{xy}$ in (8) is exactly the goal of the unscented transform. This goal becomes more clear from the point of view of the next step: estimating the vector x from observations y. Indeed, the best linear estimate of x by y is $\theta^*(y) = D_{xy} D_y^+ y + (m_x - D_{xy} D_y^+ m_y)$, the unscented transform yields $\hat{x} = \hat{D}_{xy} \hat{D}_y^+ y + (m_x - \hat{D}_{xy} \hat{D}_y^+ \hat{y})$. In this case, it is intuitively clear that the more precisely $\hat{D}_{xy}, \hat{D}_y, \hat{y}$ approximate the moments D_{xy}, D_y, m_y, the more accurate the estimate $\hat{x}$ will be. On the other hand, this method is not burdened with any optimization settings regarding the best choice of sigma-points and the weights themselves, although it contains a fair number of reasonable, physically justified, and practically illustrated options for their choice (for more detail on this, see the survey [2]).

Note that the proposed interpretation of the unscented transform is needed only for illustration. Usually publications on UKF do not speak of a discrete distribution but rather only about sigma-points and "weights," which can even be negative. The choice of these weights can be quite diverse and non-universal even within the same model and can vary depending on the purpose of the transformation. One of the options used in the calculations below is as follows. We define the transformation parameters α, β, γ (recommendations on their choice and a physical interpretation are given, for example, in [24]) and define the "scale parameter" $\lambda = \alpha^2 (p + \gamma) - p$. Sigma-points

are constructed as follows:

$$x^0 = m_x, \quad x^i = m_x + \left(\sqrt{(p+\lambda)\,D_x}\right)_i, \quad i = 1, \ldots, p,$$
$$x^i = m_x - \left(\sqrt{(p+\lambda)\,D_x}\right)_i, \quad i = p+1, \ldots, 2p, \tag{9}$$

where $\left(\sqrt{(p+\lambda)\,D_x}\right)_i$ is the ith column of the matrix $\sqrt{(p+\lambda)\,D_x}$, and the square root of a matrix-valued argument denotes the result of the Cholesky decomposition. The weights are defined as

$$W^0 = \frac{\lambda}{p+\lambda} \quad \text{for computing} \quad \hat{y},$$
$$W^0 = \frac{\lambda}{p+\lambda} + 1 - \alpha^2 + \beta \quad \text{for computing} \quad \hat{D}_y \text{ and } \hat{D}_{xy}, \tag{10}$$
$$W^i = \frac{1}{2(p+\lambda)}, \quad i = 1, \ldots, 2p.$$

In the examples below, the above-mentioned recommendations from [24] were used to choose the basic values of parameters $\alpha^*, \beta^*, \gamma^*$. In addition, in every problem we additionally performed optimization where the basic values acted as initial conditions, and calculations were repeated many times for different values of the parameters, choosing the ones that resulted in a UKF with the greatest accuracy. It was possible to implement such an operation within the framework of this work only because of the simplicity of the chosen examples with scalar states and observations, and at the same time we had no limit on the simulation time. In practice, of course, such optimization of parameters will not always be possible.

We make two important remarks. First, by presenting the estimate of the unscented transform in the form $\hat{x} = m_x + \hat{D}_{xy}\hat{D}_y^+(y - \hat{y})$, we draw attention to the similarity, the functional identity of the difference $y - \hat{y}$ and residue ζ_t, as well as the coefficients $\hat{D}_{xy}\hat{D}_y^+$ and H_t in ratios (7) of the EKF. This similarity provides an empirical justification for the use of unscented transform in the relations of the non-linear Kalman filter.

The second point concerns the interpretation of the moments D_{xy}, D_y, m_y and the corresponding approximations $\hat{D}_{xy}$, $\hat{D}_y$, $\hat{y}$. This description of the unscented transform is based on a static estimation model, where these parameters are just first and second moments. If we presume that the state x_t of dynamical system (1) will play the role of the vector x, and the assumption on its distribution will change depending on the observations y_t that play the role of y, then it is appropriate to characterize $\mathcal{F}_x$ as a conditional distribution of x whose shape is unknown, but the mean value and covariance are the results of estimation made from previously obtained observations. In this sense, D_{xy}, D_y, m_y should be interpreted as conditional moments.

Now the exposition can be summarized by UKF relations for system (1). These relations are given according to [24].

Suppose that we have $\hat{x}_{t-1}$, an estimate of the UKF for the state x_{t-1} according to observations y_τ, $\tau = 1, \ldots, t-1$, and $\hat{K}_{t-1}$ is an estimate of the conditional covariance matrix of the error $x_{t-1} - \hat{x}_{t-1}$. We choose sigma-points $\hat{x}_{t-1}^i$ according to (9) and weights W^i according to (10), using $D_x = \hat{K}_{t-1}$, and recalculate the sigma-points $\tilde{x}_t^i = \varphi_t^1\left(\hat{x}_{t-1}^i\right) + \varphi_t^2\left(\hat{x}_{t-1}^i\right) m_w(t)$, $i = 0, \ldots, 2p$. We are looking for a prediction $\tilde{x}_t$ in the following form:

$$\tilde{x}_t = \sum_{i=0}^{2p} W^i \tilde{x}_t^i,$$
$$\tilde{K}_t = \sum_{i=0}^{2p} W^i \left(\tilde{x}_t^i - \tilde{x}_t\right)\left(\tilde{x}_t^i - \tilde{x}_t\right)^{\mathrm{T}} + \varphi_t^2\left(\hat{x}_{t-1}^i\right) D_w(t) \left(\varphi_t^2\left(\hat{x}_{t-1}^i\right)\right)^{\mathrm{T}}, \tag{11}$$

where $\tilde{K}_t$ is the UKF estimate of the conditional covariance matrix of the prediction error $x_t - \tilde{x}_t$. Next, the following sigma-points are recalculated:

$$y_t^i = \psi_t^1\left(\tilde{x}_t^i\right) + \psi_t^2\left(\tilde{x}_t^i\right) m_v(t).$$

We look for an estimate $\hat{x}_t$ of the state x_t in the form

$$\hat{y}_t = \sum_{i=0}^{2p} W^i y_t^i, \quad \hat{K}_t^y = \sum_{i=0}^{2p} W^i\left(y_t^i - \hat{y}_t\right)\left(y_t^i - \hat{y}_t\right)^{\mathrm{T}} + \psi_t^2\left(\tilde{x}_t^i\right) D_v(t)\left(\psi_t^2\left(\tilde{x}_t^i\right)\right)^{\mathrm{T}},$$

$$\hat{K}_t^{xy} = \sum_{i=0}^{2p} W^i\left(\tilde{x}_t^i - \tilde{x}_t\right)\left(y_t^i - \hat{y}_t\right)^{\mathrm{T}}, \tag{12}$$

$$H_t = \hat{K}_t^{xy}(\hat{K}_t^y)^+, \quad \hat{x}_t = \tilde{x}_t + H_t\zeta_t, \quad \zeta_t = y_t - \hat{y}_t, \quad \hat{K}_t = \tilde{K}_t - H_t\hat{K}_t^y H_t^{\mathrm{T}}.$$

Here we note the values $\hat{K}_t^y$ and $\hat{K}_t^{xy}$, which are UKF estimates of the conditional covariances of the residual ζ_t and the mixed prediction covariance of the prediction $\tilde{x}_t$ and the residual ζ_t, respectively. Note that the above statements regarding conditional moments are based on a specific condition, the assumption that the state x_{t-1} has the distribution $\mathcal{F}_{x_{t-1}} \in \mathbf{\Phi}\left(\hat{x}_{t-1}, \hat{K}_{t-1}\right)$, i.e., the values $\hat{x}_{t-1}$ and $\hat{K}_{t-1}$ computed at the previous filtering step are interpreted as conditional expectation and conditional covariance, or at least their approximations. This is the reason why a remark was made earlier regarding the interpretation of approximations $\hat{D}_{xy}$, $\hat{D}_y$, $\hat{y}$ obtained as a result of the unscented transform. In our opinion, this is an extremely important observation regarding UKF methodology, which explains the practical efficiency of filtering algorithms and high accuracy of the estimates in specific experiments. The term "conditional" here is in no way connected with the "truly" conditional distribution of x_t with respect to observations y_τ, $\tau = 1, \ldots, t$, or some other set of random variables. But the meaning of this sentence is intuitively clear, and its form is illustrated by the strict relations of the linear Kalman filter and the normal correlation theorem [25].

Finally, we note that notation used in (11) and (12) was specifically chosen to be similar to the notation used in relations (3)–(6) of CMNF and (7) of EKF in order to emphasize the similarity and even identity in their interpretation.

4. MODEL CALCULATIONS

4.1. Static Regression and Polar Coordinates

Let us consider a series of model examples that allow us to illustrate the presence of both "good" and "unsuccessful" conditions for using UKF and analyze the behavior of CMNF under the same conditions. Note that, in general, the concept of a UKF estimate seems to be quite viable and useful, it has visible advantages and even prospects that have not yet been realized. But at the same time there are limitations whose systematic study has so far escaped proper attention from the researchers in this field. One of the goals of this work is to reveal the limitations of UKF, at least with examples of an academic nature, and make sure that such models have the possibility of practical filtering (for this, we use the CMNF).

Let us consider a simple static example whose model has been used in many publications to illustrate the results of the unscented transform (the full argument is well illustrated in [24]). Consider a Gaussian vector $\mathrm{col}(x, y)$, where the values x, y are interpreted as Cartesian coordinates of a point on the plane. We will assume these coordinates to be independent, unknown, and observable indirectly. By ρ and ϕ we denote the observations of the corresponding polar coordinates

of x and y together with Gaussian additive noise:

$$\rho = \sqrt{x^2 + y^2} + v_\rho, \quad \phi = \arctan(x/y) + v_\phi, \tag{13}$$

where v_ρ, v_ϕ are independent centered values with standard deviations $\sigma_{v_\rho} = 30$ m, $\sigma_{v_\phi} = 5$ deg. The problem is to estimate x and y given ρ and ϕ. We use the following algorithms for estimation.

1. *Best linear estimate*

$$\theta^*(\rho,\phi) = D_{x,y,\rho,\phi} D^+_{\rho,\phi} \mathrm{col}(\rho,\phi) + \left(\mathrm{col}(m_x, m_y) - D_{x,y,\rho,\phi} D^+_{\rho,\phi} \mathrm{col}(m_\rho, m_\phi)\right),$$

$$D_{x,y,\rho,\phi} = \begin{pmatrix} \mathrm{cov}(x,\rho) & \mathrm{cov}(x,\phi) \\ \mathrm{cov}(y,\rho) & \mathrm{cov}(y,\phi) \end{pmatrix}, \quad D_{\rho,\phi} = \begin{pmatrix} \mathrm{cov}(\rho,\rho) & \mathrm{cov}(\rho,\phi) \\ \mathrm{cov}(\phi,\rho) & \mathrm{cov}(\phi,\phi) \end{pmatrix}.$$

All moment characteristics appearing in these relations are calculated with the Monte Carlo method for a sample size of 10 000.

2. *UKF estimate.* It is calculated with exactly the same formulas, the only difference being that the moments m_ρ, m_ϕ, $D_{x,y,\rho,\phi}$, $D_{\rho,\phi}$ are computed by the UKF formulas (12) (prediction (11) is not needed here because of the static nature of the problem), the sigma-points needed for this, by formulas (9), the weights, by (10). At the same time, parameters α, β, γ are subject to additional optimization, namely, we choose the best values of these parameters from the point of view of estimation quality in the considered problem.

Remark. Here and below, by the optimization of parameters α, β, and γ in a UKF we mean the following computational procedure. According to the recommendations from [24], we choose the following basic parameter values: $\alpha^* = 0.5$, $\beta^* = 2.0$, $\gamma^* = 3 - p$. Then we form a uniform grid of possible values of α, β, γ around these base values, and then choose out of the grid the best in the sense of minimizing the second moment of the error for the UKF estimate at the last filtering step (the only step in the static task of recalculating polar coordinates). In the vicinity of this combination of parameters, we then perform Nelder–Mead optimization. As a result, we consider about 10 000–20 000 combinations of possible values for α, β, γ. We note that in some cases, variation of the unscented transform parameters relative to the basic α^*, β^*, γ^* does not lead to visible changes in estimation quality. In such cases, namely when the difference between the minimum and maximum values of the second moment of the estimation error at the last filtering step does not exceed 0.1 %, the base values are used for UKF estimation.

3. *CMNF estimation.* Let us first note that the best linear estimate in this problem is also a CMNF estimate if we assume that the structure function (in this static example, there is only one structure function) is $\zeta = \mathrm{col}(\rho, \phi)$. But as a "truly" conditional-minimax estimate we will use $\zeta = \mathrm{col}\,(\zeta_x, \zeta_y)$, $\zeta_x = \rho\cos\phi$, $\zeta_y = \rho\sin\phi$, i.e., relations of the transition from polar to Cartesian coordinates, "inverse" in the sense of the observation model (13). This is a typical example of choosing the CNMF evaluation structure using the physical meaning of the model.

In the model (13), only conditions for the location of the initial point remain undefined, i.e., parameters $\mathrm{col}\,(m_x, m_y)$ and $\begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}$. The second moments will not add any fundamental variety in the experiment, so we assume $D_x = D_y = 30^2$ everywhere. But the values of m_x and m_y will reflect the understanding of the physical property of "nearness" and "farness." As "near" we let $m_x = 30$ and $m_y = 40$; as "far," $m_x = 300$ and $m_y = 400$. The results are presented in the table, where $\mathrm{D}[x - \hat{x}]$ and $\mathrm{D}[y - \hat{y}]$ denote error variances of the estimates for coordinates x and y by the corresponding algorithms. For the UKF, in addition, we show in parentheses the estimate of the corresponding variance, obtained by averaging the values of $\hat{K}_t$ from (12). Similar "theoretical" values for linear estimation and CMNF are not given, since they are close to the "statistical" ones

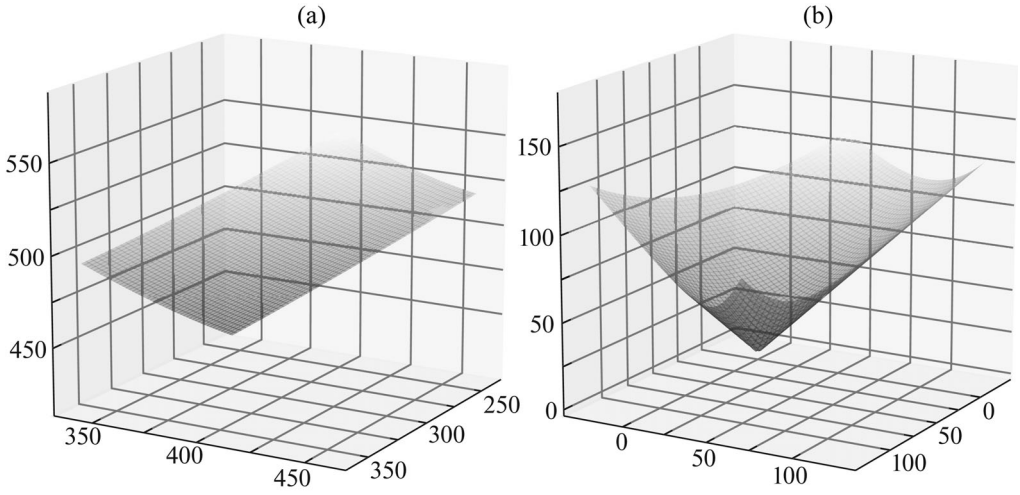

Fig. 1. Fragments of the surface $\rho = \sqrt{x^2 + y^2}$: (a) the "far" model; (b) the "near" model.

given in the table and become closer as we carry out more experiments within the framework of the Monte Carlo method.

For calculations in the "near" model, we used the following parameters of the unscented transform: $\alpha = 1.06$, $\beta = 1.35$, $\gamma = 0.91$. These parameters significantly improve the accuracy of estimation compared to the baseline, for which $\mathrm{D}[x - \hat{x}] = 711$, $\mathrm{D}[y - \hat{y}] = 477$. Optimization in the "far" model did not have a significant effect on estimation quality, therefore, for the unscented transform we used the basic values of α^*, β^* and γ^*.

The first thing that attracts attention in the results above is the fact that properties of the estimates in the "far" model are identical, a small loss of CMNF is well within statistical error. In this case, accuracy estimates for the UKF are also quite close, which suggests that the estimates of conditional moments used in (12) can also be approximated well with sigma-points. The second interesting observation is that the UKF loses by a factor of more than two in the "near" model, together with a significant error in approximating the variance of the estimation error. At the same time, the linear estimate's error is a little less but also too large. CMNF copes with the task much better and, which is especially interesting, much more successful than in the "far" model. The reason for such results, in our opinion, is quite simple. The "far" model shows "close to linear" behavior, while the "near" model is essentially nonlinear. This is confirmed by Fig. 1, which shows fragments of the surface $\rho = \sqrt{x^2 + y^2}$ for $x \in [m_x - 2\sqrt{D_x}, m_x + 2\sqrt{D_x}]$ and $y \in [m_y - 2\sqrt{D_y}, m_y + 2\sqrt{D_y}]$; you can see that in the "far" model (Fig. 1a) this fragment is visually indistinguishable from the plane, and in the "near" model (Fig. 1b) it is not a plane at all, and this nonlinearity is very informative, as CNMF convincingly demonstrates.

Table

Comparison for model (13)		"Near"	"Far"
1. Linear estimate	$\mathrm{D}[x - \hat{x}]$	585	562
	$\mathrm{D}[y - \hat{y}]$	456	514
2. UKF estimate	$\mathrm{D}[x - \hat{x}]$	614 (444)	562 (553)
	$\mathrm{D}[y - \hat{y}]$	467 (395)	514 (508)
3. CMNF estimate	$\mathrm{D}[x - \hat{x}]$	276	564
	$\mathrm{D}[y - \hat{y}]$	323	516

Note that this behavior of the UKF estimate is not unexpected, many researchers have noted it, stating the efficiency condition for the unscented transform as the "conservatism" of the covariance of estimation error, which in this example is provided by the linear behavior of the "far" model and violated by the non-linear behavior in the "near" model.

4.2. Rational Function Regression and the Cubic Sensor

Below we will consider examples of autonomous scalar dynamical systems. In such examples, it is easy to analyze the clearly expressed properties of the model and their influence, ultimately, on the accuracy of the estimates. The first model in this series is:

$$x_t = \frac{x_{t-1}}{1+x_{t-1}^2} + w_t, \quad t = 1, \dots, T, \quad T = 50, \quad x_0 = \eta,$$
$$y_t = x_t + x_t^3 + v_t. \tag{14}$$

Noises w_t and v_t are assumed to be standard Gaussian, the initial condition η is also Gaussian with $m_\eta = 0.1$ and $D_\eta = 1$. The observations in (14) are well known under the name "cubic sensor," and they are very convenient (informative) for estimation since for large values of x_t they significantly increase the signal-to-noise ratio, while for small values of x_t they are close to linear. As for the dynamics, the regression function $x/(1+x^2)$, albeit non-linear (Fig. 2a), demonstrates rather inert behavior: it has only two extremes, smooth derivatives, and asymptotically tends to zero at infinity.

Note also that the phase process x_t has one useful property. It is obviously ergodic (it is easy to show this following basic results on nonlinear regression, as described, for example, in [26]), and convergence to the limit distribution is almost instantaneous: already at the second step, under the initial conditions very different from the limit, we obtain moments close to the limit distribution: $\mathrm{E}[x_T] = 0$, $\mathrm{D}[x_T] = 1.16$ (on Fig. 2b the limit distribution is illustrated by the histogram x_T). Thus, system (14) actually has no transition process, and therefore it seems perfect for both UKF and CMNF.

Here and everywhere below, the UKF is implemented according to relations (11), (12), supplemented by the optimization of parameters α, β, γ mentioned in the first example. The structure of CMNF is basic: $\xi_t(x) = x/(1+x^2)$, i.e., the prediction is constructed "by virtue of the system,"

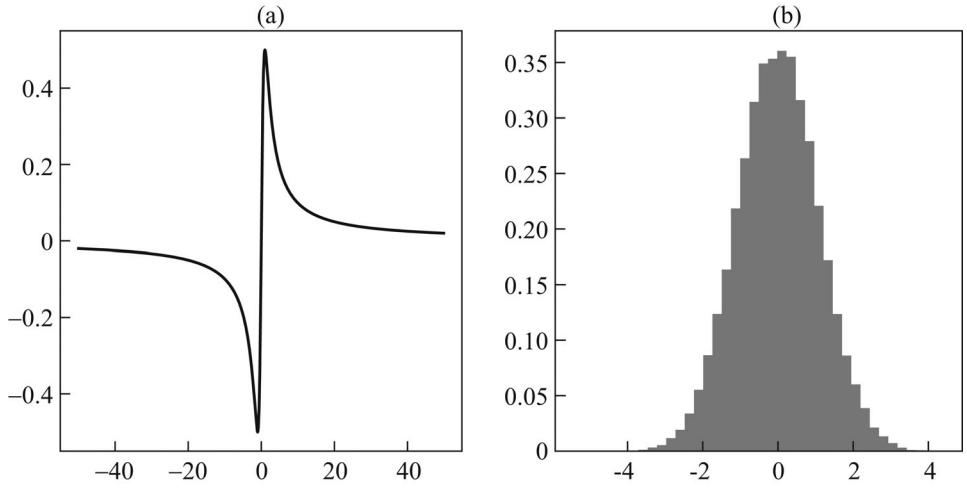

Fig. 2. Characterization of model (14): (a) regression function; (b) histogram for x_T, $\mathrm{E}[x_T] = 0$, $\mathrm{D}[x_T] = 1.16$.

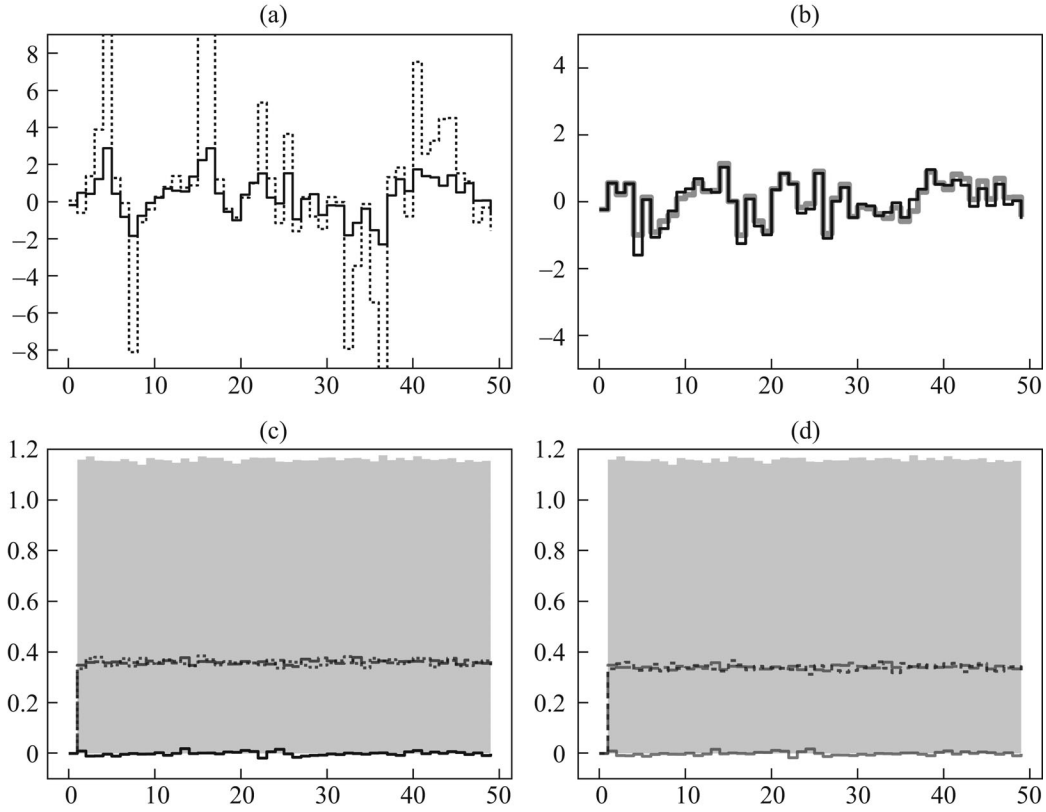

Fig. 3. Sample trajectories and quality indicators of estimates in the model (14): (a) sample trajectories x_t (solid line), y_t (dotted line); (b) filtering error trajectories $\hat{x}_t - x_t$ for the CMNF (black line), UKF (gray line); (c) and (d) accuracy indicators of CMNF and UKF on the background of the variance of the process $\mathrm{D}[x_t]$ (upper boundary of the gray area): $\mathrm{E}[x_t - \hat{x}_t]$ (solid lines), $\mathrm{D}[x_t - \hat{x}_t]$ (dashed lines), $\hat{K}_t$ (dotted lines).

$\zeta_t(x, y) = y - x - x^3$ is the correction in the form of residual. The CMNF parameters were calculated using the Monte Carlo method for a bundle of 10^5 trajectories (it would be appropriate to call it the training set), and the same bundle was used to optimize the parameters of the UKF. The quality of filtering estimates and moment characteristics of x_t were estimated on another bundle of 10^6 trajectories. The above conditions of calculations were used in all the following examples.

Figure 3 shows sample trajectories (states, observations—Fig. 3a), estimation errors for the UKF and CMNF—Fig. 3b, quality indicators for the estimates—mean and variances of the estimates and process errors are shown for CMNF on Fig. 3c, for the UKF—on Fig. 3d. Sample moments of the limit distributions of estimation errors are $\mathrm{E}[x_T - \hat{x}_T] = 0$, $\mathrm{D}[x_T - \hat{x}_T] = 0.35$ for CMNF and $\mathrm{E}[x_T - \hat{x}_T] = 0$, $\mathrm{D}[x_T - \hat{x}_T] = 0.34$ for UKF. In addition, Fig. 3c shows a priori CMNF accuracy, i.e., the value $\hat{K}_t$ from (6), calculated by the Monte Carlo method, $\hat{K}_T = 0.35$; Fig. 3d shows the "theoretical" accuracy, i.e., the average value $\hat{K}_t$ from (12), $\hat{K}_T = 0.11$. The best found values of the unscented transform parameters in this model are $\alpha = 0.20$, $\beta = 1.99$, $\gamma = 1.23$. For the basic values of the parameters $\alpha^* = 0.5$, $\beta^* = 2.0$, $\gamma^* = 2.0$, the UKF estimate turned out to be worse than trivial, providing the accuracy of $\mathrm{D}[x_T - \hat{x}_T] = 1.57$ with $\mathrm{D}[x_T] = 1.16$.

As we can see, both estimates exhibit similar qualities, both are unbiased and meaningful in the sense that the variance of the error estimates are less, and significantly so, than the variance of the process itself and the noise variance in the observations. Note also that the UKF works slightly better than the CMNF, but the latter has a simpler structure (the prediction is "by virtue of the system," with a correction in the form of residual), and we have made no attempts to improve it although they are easy to imagine. For example, an increase in the dimension of the basic

correction $\zeta_t(x, y)$ will improve CMNF estimation quality due to "additional" observations y_t^2, y_t^3. Also note that the "theoretical" accuracy of the UKF is far from reality, which does not prevent it from solving the problem successfully.

4.3. Inverse Proportional Regression

We now turn to models that represent less comfortable conditions for nonlinear estimation. In this case, as before, states are described by scalar autonomous regressions with Gaussian perturbations that have the ergodic property. In this subsection, we consider two models that use the function of inverse proportionality, while observations are straight linear:

$$x_t = \frac{1}{\sqrt[3]{x_{t-1}}} + 100w_t, \quad t = 1, \ldots, T, \quad T = 50, \quad x_0 = \eta,$$
$$y_t = x_t + 100v_t, \tag{15}$$

$$x_t = \min\left(10^5, \frac{1}{x_{t-1}^2}\right) + 100w_t, \quad t = 1, \ldots, T, \quad T = 50, \quad x_0 = \eta,$$
$$y_t = x_t + 100v_t. \tag{16}$$

The noises w_t and v_t are assumed to be standard Gaussian, the initial condition η is also Gaussian with $m_\eta = 3$ and $D_\eta = 1$. For correctness, relations (15), (16) must be supplemented with the equation $x_t = 10^5$, if $x_{t-1} = 0$. In both models, the regression function of one class $1/x^\delta$,

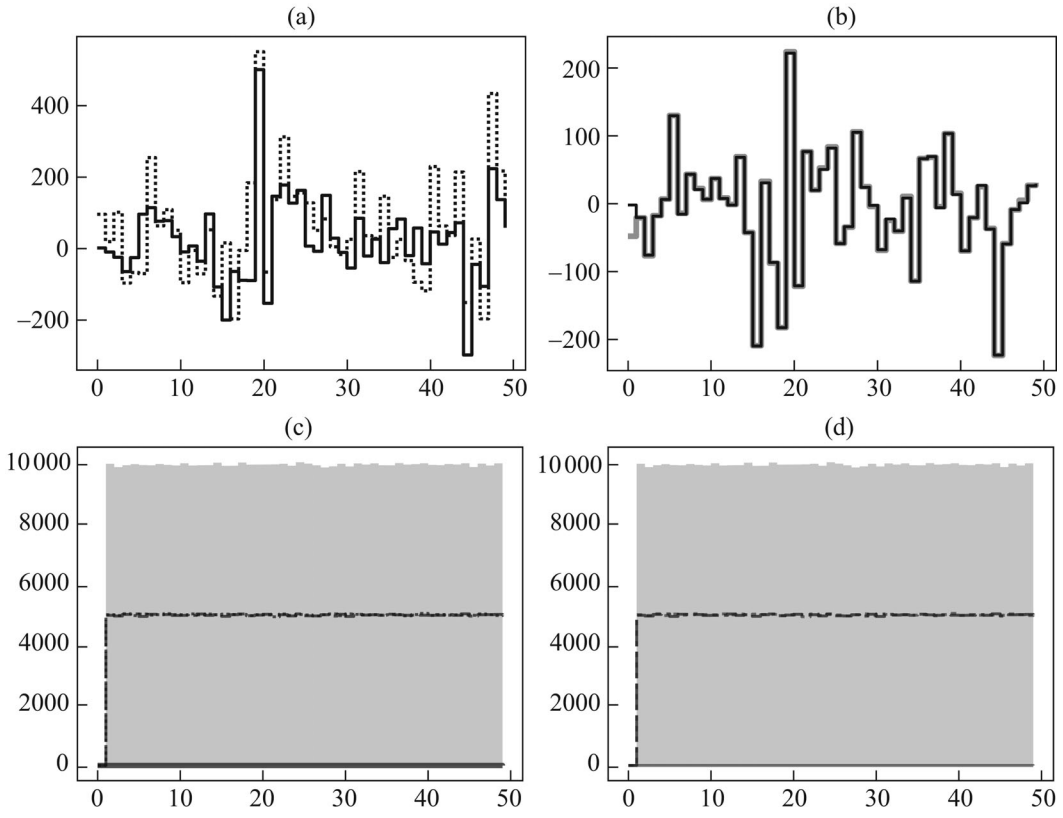

Fig. 4. Sample trajectories and quality indicators of estimates in the model (15): (a) sample trajectories x_t (solid line), y_t (dotted line); (b) filtering error trajectories $\hat{x}_t - x_t$ for the CMNF (black line) and UKF (gray line); (c) and (d) accuracy indicators of CMNF and UKF on the background of the process variance $\mathrm{D}[x_t]$ (upper boundary of the gray area): $\mathrm{E}[x_t - \hat{x}_t]$ (solid lines), $\mathrm{D}[x_t - \hat{x}_t]$ (dotted lines), $\hat{K}_t$ (dotted lines).

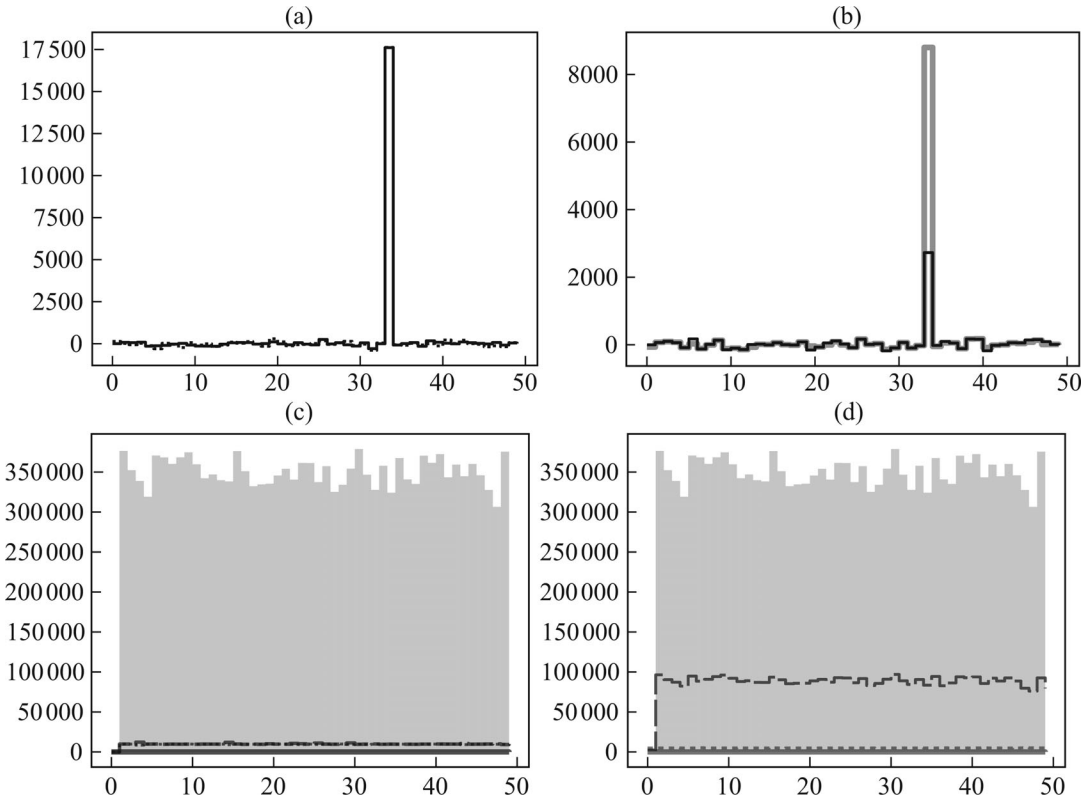

Fig. 5. Sample trajectories and quality indicators of estimates in the model (16): (a) sample trajectories x_t (solid line), y_t (dotted line); (b) filtering error trajectories $\hat{x}_t - x_t$ for the CMNF (black line) and UKF (gray line); (c) and (d) accuracy indicators of CMNF and UKF on the background of the process variance $\mathrm{D}[x_t]$ (upper boundary of the gray area): $\mathrm{E}[x_t - \hat{x}_t]$ (solid lines), $\mathrm{D}[x_t - \hat{x}_t]$ (dotted lines), $\hat{K}_t$ (dotted lines).

$\delta = \frac{1}{3}$ in (15), $\delta = 2$ in (16). The difference, as is easy to see, is that the degree of δ in (15) ensures the existence of the second moments of x_t, and in (16), on the contrary, even the first moment may not exist. That is why, in order to make the filtering problem correct, an upper bound 10^5 was added to (16), which ensures the existence of all moments. The limit distributions for processes x_t in (15) and (16) are characterized by typical unimodal probability densities, the limit sample mean values $\mathrm{E}[x_T]$ are equal to 0.18 and 6.06, variances $\mathrm{D}[x_T]$ are equal to 10 314.41 and 361 201.04 respectively.

Both processes x_t have a similar behavior: close to zero expectation and "many" samples near zero, large variance provided by the intermittently appearing "large" values, when the realizations are $x_{t-1} \approx 0$. Note that both models do not fall into the scope of robust estimation, there are no "missing" or "incorrect" observations, outliers, etc. Robustification in the traditional sense will not give a good result in this model since in this case we do need to evaluate rare bursts, as they represent the actual state of the process being evaluated and cannot be ignored. The results of the calculations are illustrated by the following figures. For (15), examples of characteristic trajectories (states, observations—on Fig. 4a), error estimates for UKF and CMNF on Fig. 4b), quality indicators for the estimates—means and variances of the estimates and process errors are shown for CMNF on Fig. 4c, for the UKF, on Fig. 4d. Sample moments of the limit distributions of estimation errors are $\mathrm{E}[x_T - \hat{x}_T] = -1.87$, $\mathrm{D}[x_T - \hat{x}_T] = 5107.98$ for CMNF and $\mathrm{E}[x_T - \hat{x}_T] = -1.62$, $\mathrm{D}[x_T - \hat{x}_T] = 5109.28$ for UKF. In addition, Fig. 4c shows the value $\hat{K}_t$ from (6), $\hat{K}_T = 5000.11$; Fig. 4d shows the averaged value $\hat{K}_t$ from (12), $\hat{K}_T = 5000.04$. For the model (16), similar results are shown in Fig. 5. Sample moments of the limit distributions of estima-

tion errors here are $\mathrm{E}[x_T - \hat{x}_T] = -0.78$, $\mathrm{D}[x_T - \hat{x}_T] = 9667.38$ for CMNF and $\mathrm{E}[x_T - \hat{x}_T] = 2.51$, $\mathrm{D}[x_T - \hat{x}_T] = 92\,521.33$ for UKF. In addition, Fig. 5c shows the value $\hat{K}_t$ from (6), $\hat{K}_T = 9804.96$; Fig. 5d shows the average value $\hat{K}_t$ from (12), $\hat{K}_T = 5007.19$. In both models (15), (16), variation of the basic values of the unscented transform parameters $\alpha^*, \beta^*, \gamma^*$ does not lead to an improvement in the estimation quality.

The obtained results are in line with the expectations: under comfortable conditions for the model (15), both estimates behave equally well, and only the CMNF works in the "inconvenient" model (16), its accuracy, albeit not much, is higher than the accuracy of direct observations. The reason for this is not a poor implementation of the UKF, but the inapplicability in this model of the very concept of sigma-points, since such nonlinearity as in (16) apparently cannot be characterized by a small finite set of transformation samples (sigma-points). But the worst thing, in our opinion, for UKF in this model is the following. In this simple model example, it is possible to predict in advance its bad behavior simply by visually analyzing the nature of the nonlinearity in the state equation. In practice, this kind of behavior is unlikely to be predicted a priori, since it is hardly possible to count on successful visual analysis of equations that are even slightly more complex than the presented scalar autonomous regression.

4.4. Logistic Model

Next, we consider a model based on a discrete counterpart of the well-known logistic model (population size model), the logistic mapping or Feigenbaum mapping $x_t = \delta x_{t-1}(1 - x_{t-1})$ [27]. Adding a Gaussian perturbation to this model, we obtain the dynamics of the observation system. The model thus turns out to be unstable and, as before, it has to be artificially constrained. Moreover, upon reaching the boundary we propose to return the process to its original state so that the explosive phenomena would be repeated time after time. Thus, in this section we will consider the following observation system:

$$\begin{gathered} x_t = \begin{cases} x_{t-1}\,(1 - x_{t-1})\,, & \text{if} \quad x_{t-1}\,(1 - x_{t-1}) > -10^3 \\ 0, \text{ otherwise} \end{cases} \qquad + w_t,\ t = 1, \ldots, 50,\ x_0 = \eta, \\ y_t = x_t + 100 v_t. \end{gathered} \tag{17}$$

Here, the dynamics is based on the logistic mapping with $\delta = 1$, so if the lower bound (-10^3), did not appear in (17), then there would be a "explosion" to the negative infinity: $x_t \to -\infty$. But at the same time, even if we remove from (17) the limit (-10^3), formally the state x_t has all finite moments at any given time, so the solution to the filtering problem exists for every t. Another thing is that the moments and absolute values of x_t grow exponentially, and therefore in computer calculations any trajectory in the model without a boundary will exceed the computer maximum in a short time.

The results are illustrated on Fig. 6. Sample moments of the marginal distributions of estimation errors are $\mathrm{E}[x_T - \hat{x}_T] = -0.83$, $\mathrm{D}[x_T - \hat{x}_T] = 6208.12$ for CMNF and $\mathrm{E}[x_T - \hat{x}_T] = -17.07$, $\mathrm{D}[x_T - \hat{x}_T] = 13\,527.82$ for UKF, whereas the sample variance of the process is $\mathrm{D}[x_T] = 16\,341.25$. In addition, Fig. 6c shows the value $\hat{K}_t$ from (6), $\hat{K}_T = 6300.53$; Fig. 6d shows the averaged value $\hat{K}_t$ from (12), $\hat{K}_T = 3068.76$. In this model, optimization of the unscented transform parameters also gives no advantage, therefore the basic α^*, β^*, γ^* are used.

This example is a vivid illustration of the potential difficulties in practical applications of UKF. It is a simple polynomial regression that ensures the formal existence of all moments and informative linear observations, but the phase process has an "explosive" character that cannot be tracked by the sigma-point method. The UKF turns out to be significantly biased, and the accuracy of the estimate is close to trivial and does not stand to comparison with the accuracy of CMNF. The

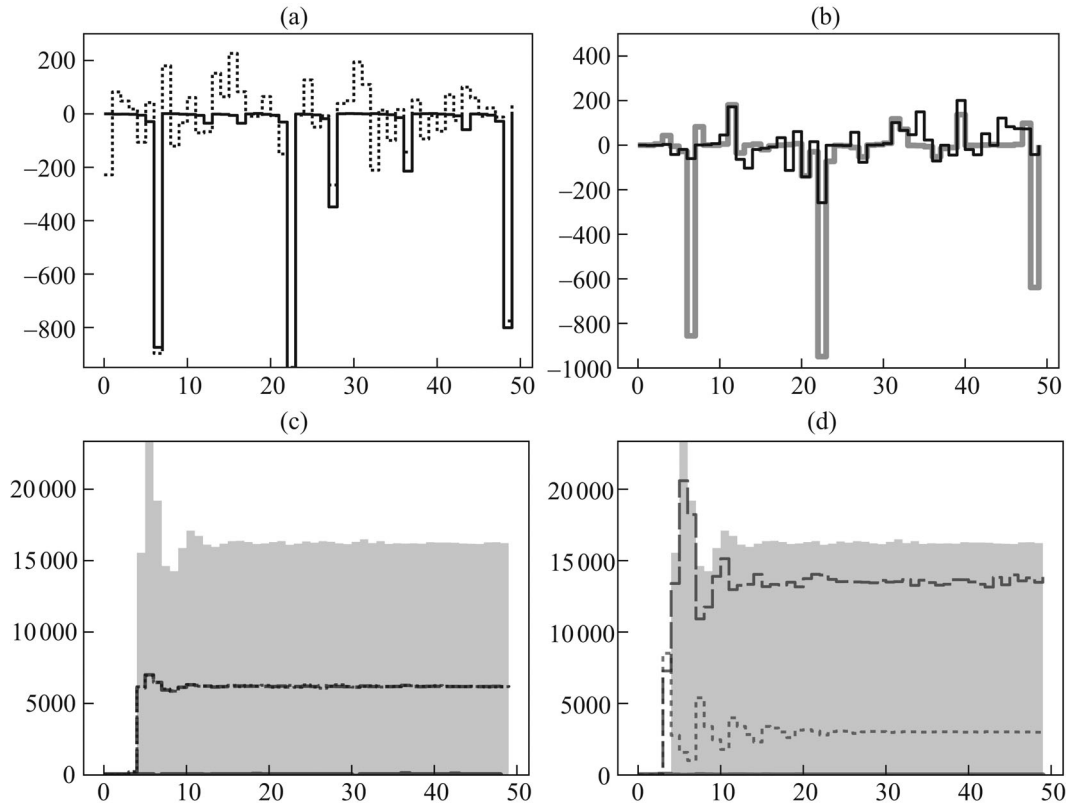

Fig. 6. Sample trajectories and quality indicators of estimates in the model (17): (a) sample trajectories x_t (solid line), y_t (dotted line); (b) filtering error trajectories $\hat{x}_t - x_t$ for the CMNF (black line) and the UKF (gray line); (c) and (d) accuracy indicators of CMNF and UKF on the background of the process variance $\mathrm{D}[x_t]$ (upper boundary of the gray area): $\mathrm{E}[x_t - \hat{x}_t]$ (solid lines), $\mathrm{D}[x_t - \hat{x}_t]$ (dashed lines), $\hat{K}_t$ (dotted lines).

estimation error for CMNF is also large, but estimation by the conditional minimax method makes sense, since the error variance is definitely less than the noise variance in the observations.

4.5. Model of Switching Observation Channels

This model is based on the idea of additional external influence on the characteristics of the observer, which had first attracted attention about 50 years ago (see, for example, [28]) and was later generalized in models of randomly structured systems and hidden Markov models. Here we consider one of its simplest variations.

Suppose that there is a white noise δ_t that does not depend on the state x_t, which is defined by a collection of discrete random variables taking values δ^i, $i = 1, \ldots, n$, with a known distribution $P\{\delta_t = \delta^i\} = p^i$. For convenience, we assume that δ^i are n-dimensional unit vectors, i.e., $\delta^i = (0, \ldots, 1, \ldots 0)^{\mathrm{T}}$, where 1 is at position i. Let d, $\sigma \in \mathbb{R}^n$. The expression $d^{\mathrm{T}}\delta_t$ gives the ith coordinate of the vector $d = (d^1, \ldots, d^n)^{\mathrm{T}}$. We assume that the state x_t at time t can be observed by one of the n linear observation channels $y_t = d^i x_t + \sigma^i v_t$, and the choice is determined by the value δ_t. With the help of the notation above, such observations can be written as $y_t = d^{\mathrm{T}}\delta_t x_t + \sigma^{\mathrm{T}}\delta_t v_t$. The state x_t is defined by a simple linear regression, thus obtaining the following observation system:

$$\begin{gathered} x_t = ax_{t-1} + b + cw_t, \quad t = 1, \ldots, T, \quad T = 50, \quad x_0 = \eta, \\ y_t = d^{\mathrm{T}}\delta_t x_t + \sigma^{\mathrm{T}}\delta_t v_t, \end{gathered} \tag{18}$$

where η, w_t, v_t are standard Gaussian.

For the calculations, we let $n = 3,\ p^1 = p^3 = 0.25,\ p^2 = 0.5,\ d^{\mathrm{T}} = (4;\ 1;\ 0.5),\ \sigma^{\mathrm{T}} = (1;\ 1;\ 3),$ $a = 0.8,\ b = 0.2,\ c = 6.0$. It is easy to see that the limit distribution of x_t is Gaussian with unit mean and variance 100.

In this example, neither relations (11), (12) for UKF nor relations (1)–(4) for CMNF cannot be used due to the formal discrepancy between models (18) and (1). This inconsistency can be fixed in the following way. In addition to the noise w_t, we introduce one more independent Gaussian white noise $\{w_t^\delta\}$ and non-intersecting intervals Δ^i: $P\{w_t^\delta \in \Delta^i\} = p^i,\ i = 1, 2, 3$. Next, we supplement the state x_t with the second coordinate

$$x_t^\delta = \left(I\left(w_t^\delta \in \Delta^1\right), I\left(w_t^\delta \in \Delta^2\right), I\left(w_t^\delta \in \Delta^3\right)\right)(1, 2, 3)^{\mathrm{T}},$$

i.e., the currently available observation channel number δ_t. In the equation of observations, instead of $d^{\mathrm{T}}\delta_t$ we use the function $\psi_t^{(1)}\left(x_t, x_t^\delta\right) = d^i x_t$, if $x_t^\delta = i,\ i = 1, 2, 3$, and similarly instead of $\sigma^{\mathrm{T}}\delta_t$ we use the function $\psi_t^{(2)}(x_t, x_t^\delta) = \sigma^i x_t$ if $x_t^\delta = i$. This notation formally reduces the model (18) to the form (1), and while the state becomes two-dimensional, the observations remain scalar. Thus, in the implementation of the UKF five sigma-points will be used according to (9). Note that the model in question does not yield any additional considerations in the construction of the structure of the UKF. The situation is different for CMNF. The specific structure of the observation system has to be reflected in the filter structure, in its correction term $\zeta_t(x, y)$. Indeed, having constructed the standard prediction $\tilde{x}_t$ "by virtue of the system" and determined its accuracy $\tilde{K}_t$ according to (3) and (6), the correction step can be viewed as an auxiliary estimation task for x_t by single observation $y_t = d^{\mathrm{T}}\delta_t x_t + \sigma^{\mathrm{T}}\delta_t v_t$. Under the additional assumption that the distribution of x_t is Gaussian with parameters $\tilde{x}_t$ and $\tilde{K}_t$, its solution is easy to write in the form of a weighted sum of linear estimates x_t for every possible observation $y_t = d^i x_t + \sigma^i v_t$, where the weights are the likelihood ratios. This approach yields the following basic correction:

$$\zeta_t(x, y) = \frac{\sum\limits_{i=1}^{3} p^i N\left(y;\ d^i\tilde{x}_t,\ {d^i}^2\tilde{K}_t + {\sigma^i}^2\right)\left(d^i\tilde{K}_t\left({d^i}^2\tilde{K}_t + {\sigma^i}^2\right)^{-1}(y - d^i x)\right)}{\sum\limits_{i=1}^{3} p^i N\left(y;\ d^i\tilde{x}_t,\ {d^i}^2\tilde{K}_t + {\sigma^i}^2\right)},$$

where $N\left(y; m, \sigma^2\right)$ denotes a Gaussian density with parameters m, σ^2, calculated at point y.

Commenting on this, we can say that the expressions

$$\frac{p^i N\left(y_t; d^i\tilde{x}_t, {d^i}^2\tilde{K}_t + {\sigma^i}^2\right)}{\sum\limits_{i=1}^{3} p^i N\left(y_t; d^i\tilde{x}_t, {d^i}^2\tilde{K}_t + {\sigma^i}^2\right)}$$

represent an estimate of the posterior probability of the event $\{\delta_t = \delta^i\}$, the values $d^i\tilde{K}_t\left({d^i}^2\tilde{K}_t + {\sigma^i}^2\right)^{-1}(y - d^i\tilde{x}_t)$, in fact, are the corrective components of the linear Kalman filter, provided that at the time t the ith observation channel was available.

In this model, an additional attempt was made to optimize the values of the unscented transform parameters separately in the prediction and correction steps, i.e., additional recalculation of the points y_t^i in (12) according to (9). The best found values are: $\alpha = 2.46,\ \beta = 3.82,\ \gamma = 2.59$ for the prediction step and $\alpha = 0.52,\ \beta = 4.15,\ \gamma = 1.24$ for the correction step. In this case, it is impossible to achieve a serious improvement in the accuracy of UKF: for the basic values of the parameters $\alpha^*, \beta^*, \gamma^*$ the estimation accuracy is $\mathrm{D}[x_T - \hat{x}_T] = 71.81$.

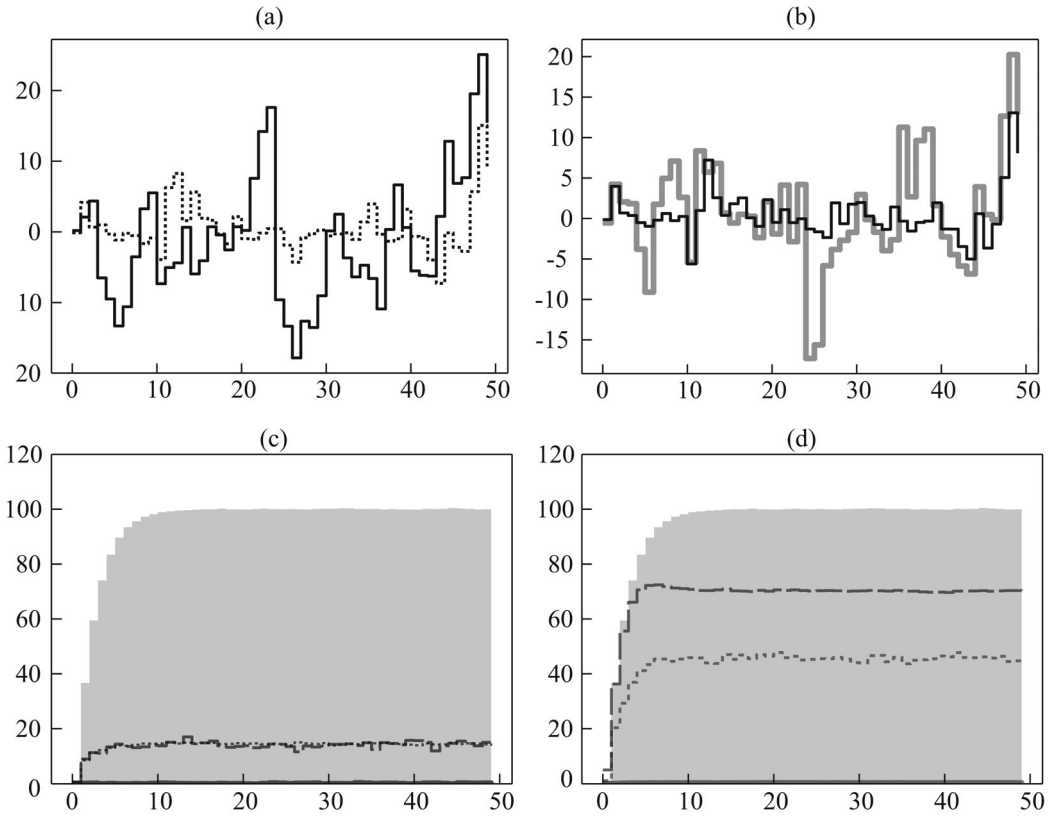

Fig. 7. Sample trajectories and quality indicators of estimates in the model (18): (a) sample trajectories for x_t (solid line) and y_t (dotted line); (b) filtering error trajectories $\hat{x}_t - x_t$ for CMNF (black line) and the UKF (gray line); (c) and (d) accuracy indicators of CMNF and UKF on the background of the process variance $\mathrm{D}[x_t]$ (the upper boundary of the gray area): $\mathrm{E}[x_t - \hat{x}_t]$ (solid lines), $\mathrm{D}[x_t - \hat{x}_t]$ (dashed lines), $\hat{K}_t$ (dotted lines).

The calculation results are illustrated on Fig. 7. Sample moments of the limit distributions of estimation errors are $\mathrm{E}[x_T - \hat{x}_T] = 0.18$, $\mathrm{D}[x_T - \hat{x}_T] = 14.30$ for CMNF and $\mathrm{E}[x_T - \hat{x}_T] = 0.29$, $\mathrm{D}[x_T - \hat{x}_T] = 70.18$ for UKF. Additionally, Fig. 7c shows the value $\hat{K}_t$ from (6), $\hat{K}_T = 14.41$; Fig. 7d shows the average value $\hat{K}_t$ from (12), $\hat{K}_T = 45.27$.

The results demonstrate the expected significant loss of UKF with respect to CMNF. It is clear that the high quality of CMNF estimates is ensured by choosing the correction $\zeta_t(x, y)$ that reflects the specifics of the model. An illustration of the flexibility of the CMNF method with respect to the filter structure is one of the objectives of this example and differentiates it from all previous ones. The UKF, on the contrary, is deliberately put into undesirable setting here by formally reducing the model (18) to the form (1). In this example, it is easy to see and implement a much better UKF, choosing sigma-points that take into account the discrete nature of the distribution δ_t, namely for each of the three sigma-points selected separately for the first coordinate x_t, define three two-dimensional sigma vectors (x_t^i, δ^j), $j = 1, 2, 3$, thus forming the final set of nine sigma-points for state (x_t, δ_t). This option was not implemented for the following reasons. First, it violates an essential part of the UKF concept, the assumption that there is a small number of sigma-points compared to the dimension of the problem (in the two-dimensional case, five sigma-points are standardly recommended). Second, such an approach for UKF will not be universal. It is applicable in this particular situation with three switching channels, since 9 sigma-points are not a very large number, but as the number of channels grows, it will become unacceptable. Finally,

the optimization procedure for UKF parameters α, β, γ from relations (9), (10) becomes too computationally intensive.

5. MODIFIED CMNF

The results of the experiments performed with UKF and their comparison with CMNF generally confirms the effectiveness of both concepts. However, the application of UKF in any practical problem is always a challenge with an unpredictable result, and getting good results always requires specific parameters (sigma-points) tuning. The conditional minimax filter always works, the structure is chosen naturally and easily, but the estimate may turn out to be too conservative, which is typical for minimax problems. A combination of the two concepts seems promising. One possible idea in this direction is to modify the CMNF by using estimates of conditional moment characteristics, covariances calculated with the state estimates at each filtering step in the filter structure, similar to how this is done in UKF. But in contrast we propose to use not the sigma-point transformation, but the Monte Carlo method, i.e., simulation modeling.

To describe the modified CMNF (MCMNF) algorithm, let us return to relations (5), (6), where CMNF coefficients are expressed in terms of expectations and covariances of the state x_t, basic prediction ξ_t, prediction error $x_t - \tilde{x}_t$, and basic correction ζ_t. In CMNF implementations, operations $\mathrm{E}\{\cdot\}$ and $\operatorname{cov}(\cdot,\cdot)$ were replaced with the corresponding statistical estimates calculated a priori by the simulated set of sample trajectories. We denote the corresponding averaging operators by $\overline{\mathrm{E}}\{\cdot\}$ and $\overline{\operatorname{cov}}(\cdot,\cdot)$. For example, in (5) instead of $\mathrm{E}\{x_t\}$ one uses $\overline{\mathrm{E}}\{x_t\} = \frac{1}{N}\sum_{i=1}^{N} x_t^i$, in (6) instead of $\operatorname{cov}(\zeta_t, x_t - \tilde{x}_t)$ we substitute $\overline{\operatorname{cov}}(\zeta_t, x_t - \tilde{x}_t) = \frac{1}{N}\sum_{i=1}^{N} \zeta_t^i \left(x_t^i - \tilde{x}_t^i\right)^{\mathrm{T}}$, where N is the number of simulated trajectories, and x_t^i, $\tilde{x}_t^i$, ζ_t^i, $i = 1, \ldots, N$ are realizations. In the modified filter, we propose to exclude prior modeling and estimate filter parameters during the calculation process for each trajectory. Accordingly, the MCMNF algorithm is as follows.

Suppose that at time moment t we have $\hat{x}_{t-1}$—the MCMNF estimate of the state x_{t-1} according to observations y_τ, $\tau = 1, \ldots, t-1$, $\hat{K}_{t-1}$—an estimate of the conditional covariance of the error. On step t, we simulate the following samples: $\{x_{t-1}^i\}_{i=1}^N$—from the Gaussian distribution with mean $\hat{x}_{t-1}$ and covariance $\hat{K}_{t-1}$, $\{w_t^i, v_t^i\}_{i=1}^N$—discrete white noise according to model (1), and sets $\{x_t^i, y_t^i\}_{i=1}^N$ are calculated using formulas (1). The MCMNF prediction $\tilde{x}_t$ is sought in the form $\tilde{x}_t = \overline{\mathrm{E}}\{x_t\}$.

Note that the basic prediction function ξ_t is not needed here, coefficient $F_t = 0$, since the estimate samples $\hat{x}_{t-1}$ are not simulated and the calculation is performed in the same way as in the first step of CMNF. Prediction accuracy is defined as $\tilde{K}_t = \overline{\operatorname{cov}}(x_t - \tilde{x}_t, x_t - \tilde{x}_t)$. Next, the set $\{\zeta_t^i\}_{i=1}^N$ is constructed and the coefficients $H_t = \overline{\operatorname{cov}}(x_t - \tilde{x}_t, \zeta_t)\,\overline{\operatorname{cov}}^{+}(\zeta_t, \zeta_t)$, $h_t = -H_t\overline{\mathrm{E}}\{\zeta_t\}$, $\zeta_t^i = \zeta_t(\tilde{x}_t,\ y_t^i)$ and the estimate accuracy $\hat{K}_t = \tilde{K}_t - H_t\overline{\operatorname{cov}}(\zeta_t, x_t - \tilde{x}_t)$ are computed. The estimate $\hat{x}_t$ of the state x_t, as in CMNF, is sought in the form $\hat{x}_t = \tilde{x}_t + H_t\zeta_t + h_t$, $\zeta_t = \zeta_t(\tilde{x}_t,\ y_t)$.

Thus, the idea of MCMNF consists in approximating the conditional distribution of x_{t-1} with respect to y_τ, $\tau = 1, \ldots, t-1$, with Gaussian distribution with moments given by the filtering results, and in calculating filter parameters through estimates of expectations and covariances obtained by simulation. We note that the assumption of Gaussian approximation of the conditional distribution can be considered in the same minimax context as in CMNF. It is also important to note that the size of the simulated samples in CMNF and MCMNF should be fundamentally different, since the calculation of the CMNF coefficients is performed a priori and thus is not limited by anything, while the calculation of MCMNF coefficients is performed during the filtering process, i.e., it assumes real-time feasibility. For this reason, in experiments for CMNF we used samples of size 10^5 and in some examples 10^6 simulated trajectories, while in the calculation of MCMNF N is assumed to be 10^3.

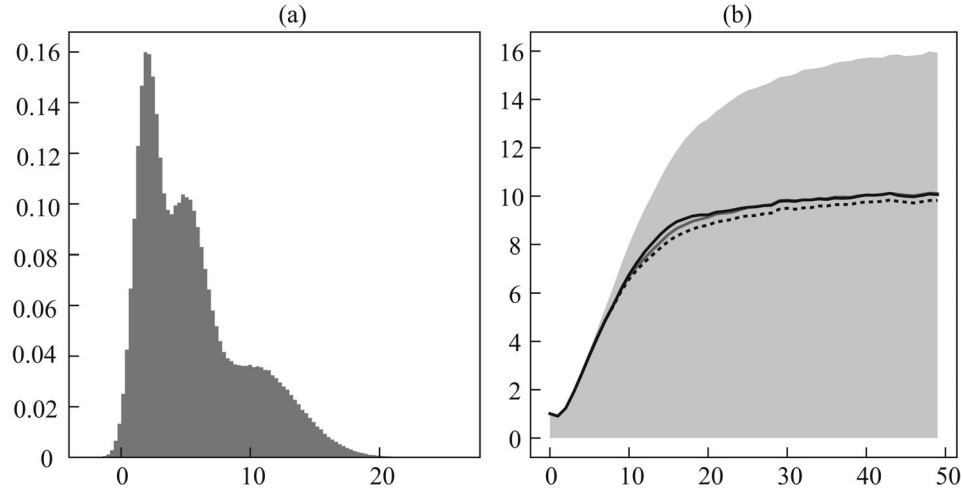

Fig. 8. Model (19): (a) histogram for x_T, $\mathrm{E}[x_t] = 5.56$, $\mathrm{D}[x_T] = 15.93$; (b) error variance in the UKF estimate (gray line), CMNF (black solid line), and MCMNF (black dotted line) against the background of the process variance $\mathrm{D}[x_t]$ (upper boundary of the gray area).

To test MCMNF and demonstrate the efficiency of combining the concepts of CMNF and UKF, consider a regression model with thresholds. This is the simplest model out of all our experiments that demonstrates identical behavior of CMNF and UKF, on the one hand, and the advantage of MCMNF, on the other. The model is based on the following idea. Suppose that for a scalar x_t n we define intervals $(x^{i-1}, x^i]$, $i = 1, \ldots, n$: $-\infty = x^0 < x^1 < \ldots < x^{n-1} < x^n = +\infty$. By $e(x_t)$ we denote the n-dimensional unit vector: $e(x_t) = (0, \ldots, 1, \ldots 0)^{\mathrm{T}}$, where the unit is at position i if $x_t \in (x^{i-1}, x^i]$. Let $a, b, c \in \mathbb{R}^n$. The expression $a^{\mathrm{T}} e(x_t)$ defines the ith coordinate of the vector a, and its position is determined by x_t falling into the ith segment. Using this simple technique, it is easy to write, for example, n linear regressions: the ith regressions will have the form $a_i x + b_i + c_i w_t$. The final model is $x_t = a^{\mathrm{T}} e\,(x_{t-1})\, x_{t-1} + b^{\mathrm{T}} e(x_{t-1}) + c^{\mathrm{T}} e\,(x_{t-1})\, w_t$, i.e., it is a combination of n first order autoregressions. Supplementing the above ratios with simple linear observations, we obtain the following observation system:

$$\begin{gathered} x_t = a^{\mathrm{T}} e\,(x_{t-1})\, x_{t-1} + b^{\mathrm{T}} e(x_{t-1}) + c^{\mathrm{T}} e(x_{t-1}) w_t, \quad t = 1, 2, \ldots, \quad x_0 = \eta, \\ y_t = x_t + 10 v_t, \end{gathered} \tag{19}$$

where η, w_t, v_t are standard Gaussian.

For the calculations, we choose $n = 3$, intervals $-\infty < 3 < 7 < +\infty$, $a^{\mathrm{T}} = (0.3;\ 0.4;\ 0.7)$, $b^{\mathrm{T}} = (1.4;\ 3.0; 3.0)$, $c^{\mathrm{T}} = (0.9;\ 1.5;\ 2.5)$. As in all models above, here the process x_t is ergodic, its limit distribution is illustrated on Fig. 8a, and the sample moments are $\mathrm{E}[x_T] = 5.56$, $\mathrm{D}[x_T] = 15.93$.

With this distribution, it is easy to see three "peaks" corresponding to three stable autoregressions $x_t = a_i x_{t-1} + b_i + c_i w_t$. The limit distribution of each of their regressions, considered separately from the others, will be Gaussian with mean $m_i = b_i/(1 - a_i)$ and variance $\sigma_i^2 = c_i^2/(1 - a_i^2)$. Forming together with given thresholds the process x_t according to (19), the selected parameters provide the limit density shown on Fig. 8a, which corresponds to the following probabilities: $p^1 = P(x_t \le 3.0) \approx 0.33$, $p^2 = P(3.0 < x_t \le 7.0) \approx 0.37$, $p^3 = P(x_t > 7.0) \approx 0.3$.

In addition to the multimodal limit distribution, model (19) also makes it possible to demonstrate the possibility of the flexible construction of CMNF structure. And here, unlike the model with switching observation channels (18), the structure of the filter is determined by the prediction. As in most of the examples shown above, the correction in the form of a residual is natural for linear observations. We define the predicting structural function $\xi_t\,(x)$ by analogy with the correction for

model (18) based on likelihood ratios, using the notation $N\left(x; m, \sigma^2\right)$ for Gaussian density:

$$\xi_t\left(x\right) = \frac{\sum\limits_{i=1}^{3} p^i N\left(x; m_i, \sigma_i^2\right)(a_i x + b_i)}{\sum\limits_{i=1}^{3} p^i N\left(x; m_i, \sigma_i^2\right)}.$$

Commenting on such a basic forecast, we can say that the expression

$$\sum_{i=1}^{3} p^i N(x; m_i, \sigma_i^2)$$

should be considered as an elementary estimate of the limit density shown on Fig. 8a, and the ratio

$$p^i N(x; m_i, \sigma_i^2) \Big/ \sum_{i=1}^{3} p^i N(x; m_i, \sigma_i^2)$$

is an estimate of the conditional probability x_{t-1} falling into the ith interval.

Computational results are illustrated on Fig. 8b. All three filters considered above demonstrate good estimation quality, MCMNF with the limit sample variance of the estimation error $\mathrm{D}[x_T - \hat{x}_T] = 9.82$ has a slight advantage about 2–3 %, UKF with $\mathrm{D}[x_T - \hat{x}_T] = 10.11$ slightly losing to CMNF with $\mathrm{D}[x_T - \hat{x}_T] = 10.06$. At the same time, recall that for the UKF we used preliminary parameter optimization, CMNF also requires a large amount of prior computations, whereas for the MCMNF there is no need for prior computations, which means that there are no restrictions on the estimation horizon. The best found values of unscented transform parameters in this model are: $\alpha = 0.71$, $\beta = 1.45$, $\gamma = 2.56$ for the prediction step and $\alpha = 0.41$, $\beta = 1.80$, $\gamma = 1.84$ for the correction step; they allow to improve the accuracy of estimation of the UKF $\mathrm{D}[x_T - \hat{x}_T] = 11.77$ for the basic values of $\alpha^*, \beta^*, \gamma^*$.

In conclusion, we note that choosing a simple model to illustrate MCMNF results proved to be rather difficult. The demonstrated result is apparently ensured by a characteristic limit distribution (Fig. 8a). Greater effect can be achieved if we abandon scalar equations.

6. CONCLUSION

In general, the experiments we have conducted with all the considered filters allows us to characterize the computational complexity and resource demands of the algorithms at a qualitative level. The need for prior calculations and storage of CMNF coefficients makes this filter the most resource-intensive. At the same time, the volume of stored parameters is determined by the specified filtering horizon, and the considerable costs of prior modeling are compensated by the extreme simplicity of calculating posterior estimates: it is characterized by the same computational complexity that the classical Kalman filter has for the linear problem. The "optimized" UKF in the scalar case assumes the same small amount of computation, but this volume increases with the dimension of the problem, i.e., as the number $2p + 1$ of sigma-points $x_t^i \in \mathbb{R}^p$ grows, it becomes more complex since we need to compute the matrix square root at each estimation step. In addition, we should note that the absence of prior calculations for the UKF does not even guarantee the "successful" choice of its parameters, and optimization performed in the considered experiments requires the same resources as modeling CMNF parameters in scalar examples. The computational complexity of such optimization grows exponentially with dimension, so the possibility of optimal selection of parameters α, β, γ for the UKF in practically significant tasks seems doubtful. The absence of prior calculations and the cost of storing parameters is a characteristic feature and an advantage

of MCMNF. It is achieved by transferring the modeling steps to posterior calculations. At the same time, the volume of modeling is less than the prior calculations of the CMNF coefficients (in examples, we used 10^3 versus 10^5 and 10^6), but still significant, and it seems reasonable only if there are appropriate computational capabilities, which can be provided by modern computer systems.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 19-07-00187-a.

REFERENCES

1. Julier, S.J., Uhlmann, J.K., and Durrant-Whyte, H.F., A New Approach for Filtering Nonlinear Systems, *Proc. IEEE Am. Control Conf. (ACC'95)*, 1995, pp. 1628–1632.
2. Menegaz, H.M.T., Ishihara, J.Y., Borges, G.A., and Vargas, A.N., A Systematization of the Unscented Kalman Filter Theory, *IEEE Trans. Autom. Control*, 2015, vol. 60, no. 10, pp. 2583–2598.
3. Julier, S.J., The Scaled Unscented Transformation, *Proc. IEEE Am. Control Conf. (ACC'02)*, 2002, pp. 4555–4559.
4. Xu, L., Ma, K., and Fan, H., Unscented Kalman Filtering for Nonlinear State Estimation with Correlated Noises and Missing Measurements, *Int. J. Control Autom. Syst.*, 2018, vol. 16, no. 3, pp. 1011–1020.
5. Li, L. and Xia, Y., Stochastic Stability of the Unscented Kalman Filter with Intermittent Observations, *Automatica*, 2012, vol. 48, no. 5, pp. 978–981.
6. Lee, D., Vukovich, G., and Lee, R., Robust Unscented Kalman Filter for Nanosat Attitude Estimation, *Int. J. Control Autom. Syst.*, 2017, vol. 15, no. 53, pp. 2161–2173.
7. Zhao, Y., Gao, S.-S., Zhang, J., and Sun, Q.-N., Robust Predictive Augmented Unscented Kalman Filter, *Int. J. Control Autom. Syst.*, 2014, vol. 12, no. 5, pp. 996–1004.
8. Scardua, L.A. and da Cruz, J.J., Complete Offline Tuning of the Unscented Kalman Filter, *Automatica*, 2017, vol. 80, pp. 54–61.
9. Straka, O., Dunik, J., and Simandl, M., Unscented Kalman Filter with Advanced Adaptation of Scaling Parameter, *Automatica*, 2014, vol. 50, no. 10, pp. 2657–2664.
10. Dunik, J., Simandl, M., and Straka, O., Unscented Kalman Filter: Aspects and Adaptive Setting of Scaling Parameter, *IEEE Trans. Autom. Control*, 2012, vol. 57, no. 9, pp. 2411–2416.
11. Biswas, S.K., Qiao, L., and Dempster, A.G., A Novel a Priori State Computation Strategy for the Unscented Kalman Filter to Improve Computational Efficiency, *IEEE Trans. Autom. Control*, 2017, vol. 62, no. 4, pp. 1852–1864.
12. Sarkka, S., On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems, *Trans. Autom. Control*, 2007, vol. 52, no. 9, pp. 1631–1641.
13. Li, X., Liu, A., Yu, C., and Su, F., Widely Linear Quaternion Unscented Kalman Filter for Quaternion-Valued Feedforward Neural Network, *IEEE Signal Process. Lett.*, 2017, vol. 24, no. 9, pp. 1418–1422.
14. Bhotto, M.Z.A. and Bajic, I.V., Constant Modulus Blind Adaptive Beamforming Based on Unscented Kalman Filtering, *IEEE Signal Process. Lett.*, 2015, vol. 22, no. 4, pp. 474–478.
15. Li, L. and Xia, Y., Unscented Kalman Filter over Unreliable Communication Networks with Markovian Packet Dropouts, *Trans. Autom. Control*, 2013, vol. 58, no. 12, pp. 3224–3230.
16. Wu, P., Li, X. and Bo, Y., Iterated Square Root Unscented Kalman Filter for Maneuvering Target Tracking Using TDOA Measurements, *Int. J. Control Autom. Syst.*, 2013, vol. 11, no. 4, pp. 761–767.

17. Jochmann, G., Kerner, S., Tasse, S., and Urbann, O., Efficient Multi-Hypotheses Unscented Kalman Filtering for Robust Localization, *Lect. Notes Comput. Sci.* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*), vol. 7416 LNCS, 2012, pp. 222–233.

18. Leven, W.F. and Lanterman, A.D., Unscented Kalman Filters for Multiple Target Tracking with Symmetric Measurement Equations, *Trans. Autom. Control*, 2009, vol. 54, no. 2, pp. 370–375.

19. Pugachev, V.S., Recurrent Estimation of Variables and Parameters in Stochastic Systems Defined by Difference Equations, *Dokl. Math.*, 1978, vol. 243, no. 5, pp. 1131–1133.

20. Pugachev, V.S., Estimation of Variables and Parameters in Discrete-Time Nonlinear Systems, *Autom. Remote Control*, 1979, vol. 40, no. 4, pp. 39–50.

21. Pankov, A.R., Recurrent Conditionally Minimax Filtering of Processes In Nonlinear Difference Stochastic Systems, *J. Comput. Syst. Sci. Int.*, 1993, vol. 31, no. 4, pp. 54–60.

22. Pankov, A.R. and Bosov A.V., Conditionally Minimax Algorithm for Nonlinear System State Estimation, *IEEE Trans. Autom. Control*, 1994, vol. 39, no. 8, pp. 1617–1620.

23. Borisov, A.V., Bosov, A.V., Kibzun, A.I., Miller, G.B., and Semenikhin, K.V., The Conditionally Minimax Nonlinear Filtering Method and Modern Approaches to State Estimation in Nonlinear Stochastic Systems, *Autom. Remote Control*, 2018, vol. 79, no. 1, pp. 1–11.

24. Wan, E.A. and Van der Merwe, R., The Unscented Kalman Filter, in *Kalman Filtering and Neural Networks*, Haykin, S., Ed., New York: Wiley, 2001, pp. 221–280.

25. Shiryaev, A.N., *Veroyatnost'* (Probability), Moscow: Nauka, 1989.

26. Bhattacharya, R.N. and Lee, C., Ergodicity of Nonlinear First Order Autoregressive Models, *J. Theor. Probab.*, 1995, vol. 8, no. 1, pp. 207–219.

27. May, R.M., Simple Mathematical Models with Very Complicated Dynamics, *Nature*, 1976, vol. 261, pp. 459–467.

28. Nahi, N., Optimal Recursive Estimation with Uncertain Observation, *IEEE Trans. Inform. Theory*, 1969, vol. 15, no. 4, pp. 457–462.

This paper was recommended for publication by M.M. Khrustalev, a member of the Editorial Board