DOI: https://doi.org/10.15276/aait.06.2023.23
UDC 004.852:004.6

Systematic use of nonlinear data filtering methods in forecasting tasks

Aleksandr P. Gozhyj[1)]
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
Irina A. Kalinina[1)]
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193
Peter I. Bidyuk[2)]
ORCID: https://orcid.org/0000−0002−7421−3565; pbidyke@gmail.com. Scopus Author ID: 6602445011
[1)] Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str. Mykolaiv, 54000, Ukraine
[2)] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, Peremoga Ave. Kyiv, 03056, Ukraine

ABSTRACT

The article describes an approach to the systematic use of nonlinear data filtering methods in tasks of intelligent data analysis and machine learning. The concepts of filtering and non-linear filtering are considered. The analysis of modern methods of optimal and probabilistic nonlinear filtering of statistical data and the peculiarities of their application in solving the problems of estimating the states of dynamic systems is carried out. The application of the Kalman filter and its variants for solving nonlinear filtering problems is analysed. The classification of nonlinear filtering methods is given. In the basis of the classification are digital, optimal and probabilistic filters. Non-recursive and recursive digital filters are studied. The formulation of the problem of optimal filtering based on the Kalman filter is considered. The filtering equation for a free dynamic system based on the state space model of a discrete system is given. The extended Kalman filter and its modifications are considered. The Bayesian method of estimating the state of a nonlinear stochastic system is presented. The problem of linear and nonlinear probabilistic filtering is considered. Three filters are considered as examples of probabilistic filters: an unscented Kalman filter, a point mass filter, and a granular filter. The granular filtering algorithm and its modifications are considered in detail. The architecture of the information-analytical system for solving forecasting problems has been developed. The system consists of the following main components: user interface, information storage subsystem, data analysis and pre-processing subsystem, modelling subsystem, forecast construction and evaluation subsystem, visualization subsystem. As an example of forecasting based on the systematic use of non-linear filtering methods, the task of forecasting the prices of *Google* shares is considered. A comparison of the quality assessment results of basic models and forecast values without filtering and with different options for applying filters was carried out. To improve the quality of forecasting based on prepared data and based on nonlinear filtering methods, a method based on combined forecasts was used to solve the forecasting problem. The results of forecasting using the combined model are presented.

Keywords. Nonlinear filtering; optimal Kalman filter; extended Kalman filter; probabilistic filter; granular filtering algorithm; information analytic system; combined forecasts

For citation: Gozhyj A. P., Kalinina I. A., Bidyuk P. I. "Systematic use of nonlinear data filtering methods in forecasting tasks". ***Applied Aspects of Information Technology.*** **2023; Vol. 6 No. 4: 345–361. DOI: https://doi.org/10.15276/aait.06.2023.23**

INTRODUCTION

In today's data-driven world, huge amounts of information are processed every day. To analyze all this data, companies need to apply data mining information technologies and techniques to analyze and identify specific subsets of information to make informed decisions. Data mining is the process of discovering patterns in data by cleaning raw data, building models, and testing those models. With the help of methods of intelligent data analysis, large volumes of information are studied, and new information is generated, revealing regularities, correlations and anomalies of the collected data.

Data mining uses techniques at the intersection of statistics, database management, and machine learning. In other words, data mining includes different groups of techniques, from collection to visualization and extraction of information from data. One of these groups of intelligent data analysis methods is filtering.

Data filtering is the process of processing large amounts of data to identify specific subsets of information based on defined criteria. This process allows you to focus on specific data and exclude others that are not relevant to the process being investigated. Data filtering is an important component of the process of preprocessing a set of real data, which is used to solve problems of intelligent data analysis and machine learning

(classification, clustering, regression, etc.). The main goal of the filtering stage is to select a useful part of the data spectrum for further processing and modeling and to retain the noisy or simply unnecessary part of the data set for analysis. It helps you understand relevant data and then use that information to access possible outcomes. Data filtering is also necessary for data visualization. Data visualization tools allow you to quickly analyze data and make informed decisions based on the information obtained. When the data is filtered, it's easier to create charts and graphs that provide meaningful statistics. A wide variety of filtering methods are available in data mining. Filtering methods are based on machine learning methods, statistical models and deep learning algorithms.

Non-linear filtering methods are of greatest interest. The main task of linear and non-linear filtering is the formation/calculation of statistical or probabilistic inferences regarding the state of the system based on the available measurements. Within the Bayesian approach to data analysis, this is done by calculating or approximating the posterior distribution of the state vector, provided that all measurements and estimates of unmeasured components available at the time of calculation are used. Since the probability distribution function of measurements contains practically all available statistical information about the object under study, its evaluation is a fairly complete solution to the problem of assessing the condition and forecasting its development. The purpose of the paper is to study the systematic use of nonlinear filtering methods in machine learning problems.

ANALYSIS OF LITERARY DATA

The initial applications of filtering were used to solve technical problems. After the fundamental works on linear filtering by Kalman [1] and Kalman and Bucy [2], the theory of filtering was applied to solve the problems of determining satellite orbits and to solve various problems of navigation, as well as for the problems of controlling spacecraft [3]. Currently, the application of nonlinear filtering methods varies from engineering problems, machine learning problems [4], as well as various problems in economic sciences, finance, medicine, and natural sciences [5, 7], [8, 9], [10].

Modern methods of optimal and probabilistic nonlinear filtering of statistical (experimental) data and features of their application are used in solving the problem of assessing the states of dynamic systems, in particular, the tracking of moving objects [3]. Papers [6,7], [8,9], [10,11] consider the modeling and estimation of the trajectory of a moving object based on modern filtering methods, as well as the use of a filtering algorithm of the "granular filter" type, which is increasingly widely used in solving the problems of estimating and predicting the states of dynamic systems.

Using different filtering methods, analysts offer different approaches to data processing in order to better understand how to process and analyze a large volume of data and draw conclusions based on the results. Therefore, the application of filtering in the tasks of intelligent data analysis and machine learning requires the systematic use of various filtering methods.

For models of nonlinear systems, nonlinear filtering methods are most often used to increase the accuracy of estimation. The most widely used nonlinear filter in solving practical engineering problems is the extended Kalman filter (EKF) due to its simple algorithm and small amount of calculation [12]. The extended Kalman filter uses Taylor series expansion to approximate the model of a nonlinear system. When the nonlinearity is complex, the filtering accuracy will be reduced or even divergent due to the high-order truncation error [13]. Therefore, to increase the accuracy in this work, it is suggested to use the extended Kalman filter of the second order and the extended Kalman filter of the higher order consistently, but at the same time, the computational complexity of the filtering procedure is significantly increased.

In [14], the iterated extended Kalman filter (IEKF) is presented, which is obtained by dividing the one-stage EKF update into several time steps and gradually updating the states according to the nonlinear gradient of the measurement function.

In [14], numerical integration approximation methods were applied to nonlinear filtering. The Gaussian Hermite filter (GHF), the unscented Kalman filter (UKF) and the cubature Kalman filter (CKF) were successively proposed. Gaussian Hermite filter is a polynomial integral approximation filtering algorithm for nonlinear system models [15] that uses Gauss-Hermite polynomials to approximate the probability density in Gaussian filtering. Unscented Kalman filter takes

the UT criterion to select deterministic sigma sample points in the initial set of state distribution points, introduces the sample points into a nonlinear system, and obtains the mean and covariance of the posterior probability density function through the point set transformation [16, 17]. Unscented Kalman filter has less computation and better approximation performance than EKF. The CKF is based on the spherical-radial criterion and uses a group of cubature points with the same weight to calculate the mean and covariance of the state variables [18]. In works [19, 20], a comparative analysis of UKF and CKF is carried out for low- and high-dimensional models in nonlinear conditions. The simulation results show that the CKF has optimal numerical stability and filtering accuracy under high dimensional conditions. A granular (particle) filter (PF) is not limited by the linearization error or Gaussian noise assumption and approximates a probability density function corresponding to a non-linear function. However, it has a significant volume of calculations for a system that solves the problem of data processing in real time [21]. This phenomenon is often found in applied filtering systems with high accuracy [22]. Due to the non-linear distribution of the measurement domain, the existing non-linear filters have specific problems in terms of state estimation accuracy and filter consistency, and even filter divergence may occur in EKF [23]. The given examples demonstrate an increase in the accuracy of the system state assessment and the efficiency of filtering with the systematic use of nonlinear filtering methods.

FEATURES OF NONLINEAR DATA FILTERING METHODS

The article examines effectiveness of the systematic use of nonlinear filters during data preprocessing in machine learning tasks.

The article solves the following problems: research of modern methods of nonlinear filtering: digital, optimal and probabilistic; research on the systematic application of nonlinear filtering methods in machine learning problems; development of the architecture of an information-analytical system for solving forecasting problems; research of granular filtering algorithms; study of the effectiveness of the systematic use of nonlinear filtering methods in solving applied machine learning problems.

FEATURES OF BUILDING STRUCTURAL MODELS OF TIME SERIES

Data filtering is an important component of the process of preprocessing a set of real data, which is used in intelligent data analysis and in solving machine learning problems. The main purpose of using filtering methods is to select the necessary part of the data for further processing when solving applied problems. In modern data preprocessing procedures, the following types of filters are most common: digital, optimal, and probabilistic filters. The classification of filter types is presented in Fig. 1.

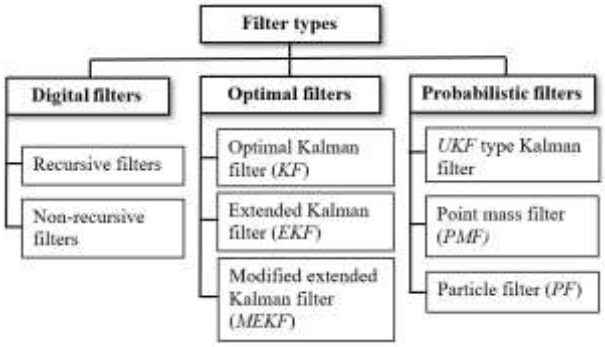

Fig. 1. **Classification of filters**
Source: **compiled by the authors**

Digital filtering. *Digital filters* (*DF*) are divided into two classes: non-recursive and recursive filters. Mathematically, the operation of a non-recursive filter can be represented, for example, by an autoregressive equation of the AR (*p*) type [21]:

$$y(k) = a_0 x(k) + a_1 x(k-1) + \ldots + a_p x(k-p),$$

where $y(k)$ is the observation value after filtering; $x(k), x(k-1), \ldots, x(k-p)$ is preliminary observation data; $a_0, a_1, \ldots, a_p$ are parameters (coefficients) that determine the amplitude-frequency response (AFR) of the filter. Next, the presented expression represents the convolution of the input signal with a certain set of filter coefficients.

Non-recursive filters have advantages compared to recursive ones: they are always stable and allow obtaining arbitrary frequency response. However, they require more resources and have long delays.

Mathematically, the operation of the recursive filter is described by the formula:

$$y(k) = a_0 x(k) + a_1 x(k-1) + \cdots + a_p x(k-p) -$$
$$-b_1 y(k-1) - b_2 y(k-2) - \cdots - b_q y(k-q).$$

Recursive filters have the following advantages: they are faster and simpler compared to non-recursive filters, and they also have analogy prototypes. Their disadvantages include the fact that their frequency response must be selected from real prototype filters, and they are not always stable.

Today, there are highly developed methods of optimization design of DF [24, 25], which allow designing effective filter structures with frequency characteristics of a predetermined shape. Many applied modelling systems have a toolkit for DF design, which consists of a set of appropriate functions, which greatly facilitates the calculation process.

Optimal filtering. The task of refining estimates of the process state under influence of random external disturbances and measurement noise (errors) is successfully solved using optimal filtering methods, in particular, the *Kalman filtering* (*KF*) algorithm. To date, there are several modifications of the filtering approach that provide the possibility of optimal data smoothing, calculation of short-term forecasts of states using optimal estimates, as well as estimation of immeasurable components of the state vector and some parameters of the mathematical model.

The main filtering equation for a free dynamic system (control actions are not taken into account) is based on the state space model of a discrete system and can be written as follows [1, 26]:

$$\hat{\mathbf{x}}(k) = \mathbf{A}\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{HA}\hat{\mathbf{x}}(k-1)],$$

where $\hat{\mathbf{x}}(k)$ is optimal estimate of the state vector $\mathbf{x}(k)$ at time k; $\mathbf{A}$ is matrix of state transitions (or matrix of system dynamics); $\mathbf{z}(k)$ is vector of initial system measurements; $\mathbf{H}$ is matrix of measurement coefficients; $\mathbf{K}(k)$ is optimal matrix of weighting coefficients, which is calculated as a result of the minimization of the functional:

$$J = \min_{\mathbf{K}} E\{[\hat{\mathbf{x}}(k) - \boldsymbol{x}(k)]^T[\hat{\boldsymbol{x}}(k) - \boldsymbol{x}(k)]\}.$$

The expression means minimization of the mathematical expectation of the squared errors of the state estimation. The value of $\mathbf{K}$ is determined by solving the corresponding Riccati equation.

The state estimation algorithm provides (automatically) the ability to estimate the state prediction one step ahead according to the equation:

$$\hat{\mathbf{x}}(k+1|k) = \mathbf{A}\hat{\boldsymbol{x}}(k|k).$$

This equation can be used to calculate multistep forecasts as follows:

$$\hat{\mathbf{x}}(k+s|k) = \mathbf{A}^s\hat{\boldsymbol{x}}(k|k),$$

where s is number of forecast steps.

Next, the Kalman filter performs the task of smoothing and forecasting taking into account statistical uncertainties such as covariance and mathematical expectation for two stochastic processes: external state perturbations and measurement errors. Therefore, the use of the filter expands the data processing system due to additional functionality aimed at combating statistical uncertainties. In addition, the adaptive version of the filter provides the possibility of real-time estimation of the statistical characteristics of two random processes. Indicators that cannot always be evaluated a priori lead to the need to build adaptive evaluation schemes.

The extended Kalman filter (*EKF*) is used to estimate the state of nonlinear non-Gaussian processes. This is a kind of linear Kalman filter, but it is applied to a linearized model of the system under study with Gaussian noise and the same moments of the first and second order. The extended KF approximates a non-linear function (the model of the system generating the data being processed) using a second-order Taylor expansion. However, the disadvantage of the approach is the replacement of the actual probability distribution of the data with a normal one, which leads to the use of the given model of system dynamics, and this may turn out to be unsuitable for further use [5, 10], [26].

Modified Extended Kalman Filter (*MEKF*). A more complex type of nonlinearity of the model, which is represented by the dependence of state variables $\mathbf{X}(t)$ with continuous time on possible discrete variables, $\mathbf{D}(k), k = 0,1,2, \dots$, which may have non-stationary probability distributions different from the distribution of continuous variables. Such situations require specific statements for all possible hypotheses related to possible values of discrete variables. The number of hypotheses can grow exponentially with the length of the discrete data sample, which can ultimately lead to high computational costs and generally unacceptable execution time of the filter implementation. To handle the following cases, another modification of the filter has been proposed, which involves the use of a random variable $\mathbf{H}(k)$, each value corresponding to one of the possible hypotheses. The distribution $\mathbf{H}(k)$, corresponds to the probability of the chosen hypothesis. In the MEKF implementation process, all combinations of values

$\mathbf{H}(k)$, and $\mathbf{D}(k+1)$ are taken into account, which makes it possible to analyse $\mathbf{K} \times |\mathbf{D}|$ component. Each new hypothesis is normalized on the incoming measurements, $\mathbf{Y}(k+1)$, and due to Bayesian conditioning, the mixture weights are adjusted, as well as the parameters of the multivariate Gaussians. The result of this procedure is an adequate model and, as a result, a higher quality of the final result - the quality of the state assessment, forecast and alternative decisions based on it.

The general form of the Bayesian state estimation method. The dynamics of a nonlinear stochastic system can be described by discrete state space equations as follows [26]:

$$\boldsymbol{x}(k) = \boldsymbol{f}[\boldsymbol{x}(k-1), \boldsymbol{w}(k-1)], \quad (1)$$

$$\boldsymbol{z}(k) = \boldsymbol{h}[\boldsymbol{x}(k), \boldsymbol{v}(k)], \quad (2)$$

where (1) is the equation of state; (2) is measurement equation; $\mathbf{x}(k)$ is vector of variable states with a non-Gaussian distribution $P_{x(k)}$; $z(k)$ is vector of real measurements (measurements can be complex numbers, but converted to real values); $\boldsymbol{w}(k)$ is vector of random external disturbances with a known probability distribution $P_{w(k)}$; $\boldsymbol{v}(k)$ is measurement noise vector (or measurement errors) with a known probability distribution $P_{v(k)}$; $\boldsymbol{f}$, $\boldsymbol{h}$ are non-linear deterministic functions; $k = 0,1,2,3, \ldots$ is discrete time.

Random disturbances in equations (1) and (2) are usually considered in an additive form, which facilitates estimation of model parameters, but makes it possible to build a model with a high degree of adequacy. If necessary, model (1), (2) can be extended by a vector of deterministic control influences $\mathbf{u}(k)$. The first measurement of $\boldsymbol{z}(1)$ makes it possible to estimate the state of $\mathbf{x}(1)$, and in the future, new measurements will lead to the estimation of future states.

We introduce the following notations for the sequence of state vectors:

$$\mathbf{x}(1:k) = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(k)\},$$

they will be used later.

In terms of conditional probability distributions, model (1), (2) can be written as follows [26]:

$$\boldsymbol{x}(k) \sim P[\boldsymbol{x}(k)|\boldsymbol{x}(k-1)],$$
$$\boldsymbol{z}(k) \sim P[\boldsymbol{z}(k)|\boldsymbol{z}(k-1)].$$

The problem of state estimation from the point of view of the Bayesian approach to data processing consists in the generation (estimation) of the posterior probability distribution $P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k)]$ based on the sequence of measurements $\boldsymbol{z}(1:k) = \{\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots, \boldsymbol{z}(k)\}$.

Equation (1) is the predicted conditional transition distribution

$$P[\boldsymbol{x}(k)|\boldsymbol{x}(k-1), \boldsymbol{z}(1:k-1)],$$

which is based on the states for previous moments and all available measurements, starting from the first and to $\boldsymbol{z}(1:k-1)$.

Equation (2) defines the likelihood function for a current measurement with known current state, $P[\boldsymbol{z}(k)|\boldsymbol{x}(1:k)]$.

The a priori probability of this state can be determined as follows:

$$P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k-1)],$$

and it can be calculated using the Bayes theorem according to the expression [26]:

$$P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k-1)] = \int P[\boldsymbol{x}(k)|\boldsymbol{x}(k-1), \boldsymbol{z}(1:k-1)] P[\boldsymbol{x}(k-1)|\boldsymbol{z}(1:k-1)] d\boldsymbol{x}(k-1). \quad (3)$$

The observation equation (2) defines the likelihood function for a current measurement with a known current state, $P[\boldsymbol{z}(k)|\boldsymbol{x}(1:k)]$.

On the other hand, the probability density function of the state of the previous time interval can be defined as follows:

$$P[\boldsymbol{x}(k-1)|\boldsymbol{z}(1:k-1)].$$

At the stage of correction, state estimates are calculated using distribution function of the following type:

$$P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k-1)] = cP[\boldsymbol{z}(k)|\boldsymbol{x}(1:k-1)] \rightleftarrows P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k-1)], \quad (4)$$

where c is normalizing constant.

The filtering problem consists in the recursive estimation of the first two moments of the state vector $\boldsymbol{x}(k)$ with known dimensions, $\boldsymbol{z}(1:k)$.

For some general type of distribution $P(\boldsymbol{x})$, the problem is to estimate the mathematical expectation for any (actual) function $\boldsymbol{x}(k)$, such as $\langle g(\boldsymbol{x}) \rangle_{p(x)}$, using equations (3) and (4) and calculating an integral of the type:

$$\langle g(\boldsymbol{x}) \rangle_{p(x)} = \int g(\boldsymbol{x}) P(\boldsymbol{x}) d\boldsymbol{x}. \quad (5)$$

But the integral cannot be taken in a closed form for the general type of multidimensional

distributions; its value should be approximated using known numerical procedures [26].

Equations (1) and (2) are often considered with additive random Gaussian components in the following form:

$$\boldsymbol{x}(k) = \boldsymbol{f}[\boldsymbol{x}(k-1)] + \boldsymbol{w}(k-1), \quad (6)$$

$$\boldsymbol{z}(k) = \boldsymbol{h}[\boldsymbol{x}(k)] + \boldsymbol{v}(k), \quad (7)$$

where $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are random vector Gaussian processes, which are represented in the simulation model by vector variables with zero mean and covariance matrices $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$, respectively.

The initial state $\boldsymbol{x}(0)$ is also modelled by random values, $\widehat{\boldsymbol{x}}_0$, which are independent of both noise processes and have a covariance matrix $\mathbf{P}_0^{\mathbf{xx}}$.

Assume that the nonlinear deterministic functions $\boldsymbol{f}, \boldsymbol{h}$ and the covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are stationary, that is, their parameters do not depend on time. Then

$$P[\boldsymbol{x}(k)|\boldsymbol{x}(k-1)|\boldsymbol{z}(1:k-1)] = N(\boldsymbol{x}(k); \boldsymbol{f}(\boldsymbol{x}(k-1)), \boldsymbol{Q}), \quad (8)$$

where $\boldsymbol{N}(\boldsymbol{t}; \boldsymbol{\tau}, \Sigma)$ **is** multidimensional Gaussian distribution, which is generally defined by the expression [26]:

$$PN(t, \tau, \Sigma) = \frac{1}{\sqrt{(2\pi)^k \|\Sigma\|}} exp\left\{-\frac{1}{2}[t - -\tau]^T (\Sigma)^{-1}[t-\tau]\right\}. \quad (9)$$

Now equation (3), which determines the a priori probability of states, can be written in the form:

$$P[\boldsymbol{x}(k)|\boldsymbol{z}(1:k-1)] = \int N[\boldsymbol{x}(k); \boldsymbol{f}(\boldsymbol{x}(k-1)), \boldsymbol{Q}] P[\boldsymbol{x}(k- -1)|\boldsymbol{z}(1:k-1)] d\boldsymbol{x}(k-1). \quad (10)$$

The expected value of $\boldsymbol{t}$ for the Gaussian distribution $N(\boldsymbol{t}; \mathbf{f}(\tau), \Sigma)$, can be represented by the expression [26]:

$$\langle \boldsymbol{t} \rangle = \int \boldsymbol{t}\, N(\boldsymbol{t}; \boldsymbol{f}(\boldsymbol{\tau}), \boldsymbol{\Sigma}) \boldsymbol{dt} = \boldsymbol{f}(\boldsymbol{\tau}). \quad (11)$$

It is known that the Kalman filter can be applied to any dynamic system represented in the form of a state space with additive Gaussian noises in both equations, regardless of the presence of nonlinearities. Although there may be a problem of convergence with nonlinearity.

This approach allows us to construct a Gaussian approximation to the posterior distribution $P(\boldsymbol{x}(k|k))$, with mean and covariance given by the expressions given below:

$$P\widehat{\boldsymbol{x}}(k|k) = \widehat{\boldsymbol{x}}(k|k-1) + \boldsymbol{K}(k)[\boldsymbol{z}(k) - \widehat{\boldsymbol{z}}(k|k-1)], \quad (12)$$

$$\boldsymbol{P}^{\boldsymbol{xx}}(k|k) = \boldsymbol{P}^{\boldsymbol{xx}}(k|k-1) - \boldsymbol{K}(k)\boldsymbol{P}^{\boldsymbol{zz}}\boldsymbol{K}^T(k), \quad (13)$$

where the optimal filter coefficient is calculated using the expression:

$$\boldsymbol{K}(k) = \boldsymbol{P}^{\boldsymbol{xz}}(k|k-1)[\boldsymbol{P}^{\boldsymbol{zz}}(k|k-1)]^{-1}. \quad (14)$$

The only approximation used in the above expressions is that of noise modelling using additive Gaussian sequences. Calculation of estimates of the state vector $\widehat{\boldsymbol{x}}(k|k)$, and covariance, $\boldsymbol{P}^{\boldsymbol{xx}}(k|k)$, is performed without approximation.

However, the practical implementation of the considered filter requires procedures for calculating integrals in equations that have the following form:

$$I = \int \boldsymbol{g}(\boldsymbol{x}) N(\boldsymbol{x}; \widehat{\boldsymbol{x}}, \boldsymbol{P}^{\boldsymbol{xx}}) d\boldsymbol{x}. \quad (15)$$

Here, $N(\boldsymbol{x}; \widehat{\boldsymbol{x}}, \boldsymbol{P}^{\boldsymbol{xx}})$ is a multidimensional Gaussian distribution with a vector of means $\widehat{\boldsymbol{x}}$ and a covariance matrix $\boldsymbol{P}^{\boldsymbol{xx}}$. There are three approximations for calculating the integral (15), considered in [27]. One of them can be chosen for a specific practical implementation.

Probabilistic filtering. The problem of linear and nonlinear probabilistic filtering is to compute a probabilistic inference about the state of the system using available measurements. In Bayesian data analysis, this is done by approximating the posterior distribution of the state vector using all available measurement information and estimates of the unmeasured components. Since the probability distribution function contains all the available information about the studied states of the system (processes), its assessment is a complete solution to the problem of assessing the state and forecasting its future development [26].

As examples of probabilistic filters, Figure 1 shows three filters: an unscented Kalman filter, a point mass filter, and a granular filter.

The implementation of the unscented Kalman filter (*U-filter* or *UKF*) is based on the principle that a set of discrete measurements can be used to estimate the mean of the data. The initial data distribution can be transformed into any other required to solve a particular problem statement by applying a nonlinear transformation to each dimension. The mean and covariance of the new distribution are the sought estimates needed by the filtering algorithm. Unlike the EKF, in which a nonlinear function (model) is approximated by a

linear one, the main advantage of this approach is the use of an available nonlinear function representing the data model. This means that there is no need to apply a linearization procedure based on the differentiation of a nonlinear function. This improves the quality of the estimates at the filter output, and also simplifies the filter implementation procedure due to the fact that the construction and implementation of the corresponding Jacobi matrix is excluded. Such a state estimation algorithm generates output data equivalent in quality to the optimal Kalman filter results for linear systems. But the advantage of this approach is that the filter is applied to nonlinear systems without applying the linearization procedure necessary to implement the EKF. It is analytically shown that the quality of filtering in this case exceeds the quality of the EKF and can be compared with the quality of the Gaussian filter of the second order [28, 29].

When applying a *point mass filter* (*PMF*), a network of points is superimposed on the state space, which is used to recursively estimate the posterior distribution of states. This filtering procedure is suitable for handling any nonlinear and non-Gaussian processes and can represent almost any posterior probability distribution with high accuracy. The main disadvantage of PMF is the high dimensionality of the distribution network in the case of a high order of the state space. This filtering procedure is used in "non-standard" cases of multidimensional distributions that require high-quality data processing results.

One of the types of probabilistic Bayesian filters is called *particle filter* (*PF*) (granular filtering). The task to be solved by the filter is the construction (approximation) of the posterior probability density for the unknown states taking into account the necessary measurements, i.e. the estimate of $P[x(k)|x(1:k)]$. There are alternative particle filtering algorithms based on pseudorandom sequences generated by Monte Carlo methods to estimate desired multivariate distributions [30].

An example of the implementation of a recursive Bayesian filter. The implementation method of the recursive Bayesian filter using Monte Carlo pseudo-random sequence generation is performed using the Sequential Importance Sampling (SIS) algorithm. It is the basic algorithm of particle filtering (*granular filtering*). The idea of filtering is to represent the desired posterior probability density as a sequence of random values with appropriate weights, which is used to compute the filtered estimates. With a significant increase in the number of elements of the sequence, the characteristic of the result of the Monte Carlo program becomes equivalent to the functional description for the posterior density, and the SIS filter approaches the optimal Bayesian estimate.

Let $\{x^i(1:k), w^i(k)\}_{i=1}^{N_S}$ be a random measure of the posterior density, $P[x(1:k),\ z(1:k)]$, where $\{x^i(1:k), i = 0,1, \dots, N_S\}$ is a set of k steps of the evolution trajectory for reference points (particles) with individual normalized weight coefficients, $\{w^i(k), i = 0,1, \dots, N_S\}$, $\sum_{i=1}^{N_S} w^i(k) = 1$. Where N_S is the number of particles that will be used to estimate the state.

The posterior density at time k can be represented as follows:

$$P[x(1:k)|z(1:k)] \approx \sum_{1}^{N_S} w^i(k)\delta\left(x(1:k) - x^i(1:k)\right), \quad (16)$$

where $\delta(x)$ is Dirac's δ – function, i.e. $P\left[x^i(1:k)|z(1:k)\right] \approx w^i(k)$.

Weighting factors are generated according to the principle *of importance sampling*. The procedure can be characterized as follows. Suppose it is desired to generate a probability distribution, $P(x) \propto \pi(x)$ (the symbol "$\propto$" stands for proportion), where $\{x(\cdot)\}$ is a desired random process that cannot be generated from its true distribution, but for which the approximation $\pi(x)$ can be calculated.

Let $x^i \sim P(x)$, $i = 0, \dots, N_S$ be realizations of random variables that can be easily generated from the $Q(\cdot)$, distribution, called *the proposal density* or *the importance density*.

Then the weighted approximation of the desired distribution $P(\cdot)$ looks like this:

$$P(x) \approx \sum_{i=1}^{N_S} w^i(k)\delta\left(x - x^i\right),$$

where

$$w^i \propto \frac{\pi\left(x^i\right)}{Q(x^i)}, \quad (17)$$

is the normalized weighting factor for the i-th particle.

After calculating the ratio, the coefficients are normalized to satisfy the condition: $\sum_{i=1}^{N_S} w^i(k) = 1$.

If the realizations of the processes $\{x^i(1:k)\}$ were generated from the distribution with the density of the offer, $Q[x(1:k)|z(1:k)]$, then the weighting coefficients in equation (16) will be calculated as follows :

$$w^i \propto \frac{P[x(1:k)|z(1:k)]}{Q[x(1:k)|z(1:k)]}. \quad (18)$$

In the case of successive computations, at each iteration of the generation procedure, a weighted sample is generated that approximates the posterior density $P[x(1:k-1)|z(1:k-1)]$, and then a new sample can be generated to approximate the density $P[x(1:k)|z(1:k)]$.

If we use Bayes' theorem, we can write:

$$P\big(x(k)\big|z(1:k)\big) = \\ = \frac{P(z(k)|x(k))P(x(k)|z(1:k-1))}{P(z(k)|z(1:k-1))}.$$

If the condition is fulfilled

$$Q\big(x(k)\big|x(1:k-1), z(1:k)\big) = Q(x(k)|x(k-1), z(k)),$$

that is, the density supply depends only on $x(k-1)$ and $z(k)$.

The following expression can be used for repeated (recursive) evaluation of weighting factors [30]:

$$w^i(k) \propto w^i(k-1)\frac{P(z(k)|x^i(k))P(x^i(k)|x^i(k-1))}{P(x^i(k)|x^i(k-1),z(k))}. \quad (19)$$

The filtered posterior distribution can be approximated as follows:

$$P(x(k)|z(1:k)) \approx \\ \approx \sum_{i=1}^{N_S} w^i(k)\delta(x(k) - x^i(k)). \quad (20)$$

It should be emphasized that the weighting coefficients, $w^i(k)$, must be normalized in such a way that $\sum_{i=1}^{N_S} w^i(k) = 1$.

The selection of the proposed density is one of the most important points in the particle filter design procedure. Possible methods of its selection, as well as their advantages and disadvantages, are considered in [30].

Often, the prior distribution of the data is used as the supply density:

$$Q\left(x(k)\middle|x^i(k-1), z(k)\right) \\ = P\left(x(k)\middle|x^i(k-1)\right). \quad (21)$$

In this case, the expression is simplified to the form:

$$w^i(k) \propto w^i(k-1)\,P(z(k)\,|\,x^i(k)). \quad (22)$$

But such a choice of supply density cannot be used to solve all problems.

Basic algorithm. The sequential importance sampling algorithm is proposed as the basic one for the granular filter.

The elements $x^i(1:1)$, in the weighted sample at the first stage $\left\{x^i(1:1), \frac{1}{N_S}\right\}_{i=1}^{N_S}$ are generated from the initial distribution $P(x(1))$. Since this distribution is relevant, no adjustment of values is required, and all weighting factors must have the same values, i.e.: $w^i(1) = 1/N_S$. If we have a verified sample at step $(k-1)$, then the procedure for generating a weighted sample at step k can be represented by the pseudocode from Table 1.

For this and all subsequent filtering algorithms, the posterior distribution is approximated using (20), and the estimate of the conditional mathematical expectation of the state, $x(k)$, is determined as follows:

$$\hat{x}(k) = \sum_{i=1}^{N_S} w^i(k)x^i(k). \quad (23)$$

Resampling of particle samples. The implementation of the SIS filter often leads to the problem of degeneracy of the weight coefficients, when after a certain number of iterations all coefficients, except one, take on small weights.

Since the dispersion of the weighting coefficients increases over time, it is impossible to avoid the phenomenon [26, 30]. This degeneracy is caused by the fact that a significant part of the calculations is spent on updating particles that practically do not affect the approximated distribution, $P\big(x(k)\big|z(1:k)\big)$.

An approach to reducing the degeneracy effect through the use of particle resembling is proposed. The main idea of the algorithm is to remove particles with a small weight and focus on particles with a large weight.

Table 1. **Algorithms of granular filtering**

No.	Algorithm name	Algorithm pseudocode
1	Algorithm of Sequential Sampling by Importance (SIS). *Base*	Algorithm 1: SIS Particle Filter $\left[\{x^i(k), w^i(k)\}_{i=1}^{Ns}\right]$= SIS $\left[\{x^i(k-1), w^i(k-1)\}_{i=1}^{Ns}, z(k)\right]$ FOR – generate $x^i(k) \sim q\left(x(k)\middle\|x^i(k-1), z(k)\right)$. – assign the particle $x^i(k)$ weight $w^i(k)$ according to (28) END FOR
2	Particle resampling algorithm	Algorithm 2: Resampling Algorithm $\left[\{x^{j^*}(k), w^j(k), i^j\}_{j=1}^{Ns}\right]$= RESAMPLE $\left[\{x^i(k), w^i(k), i^j\}_{i=1}^{Ns}\right]$ Initialize distribution function (DF): $c(1) = 0$ FOR $i = \overline{2, N_S}$ – Construct DF: $c(i) = c(i-1) + w^i(k)$ END FOR Start DF from beginning: $i = 1$ Generate initial point: $u(1) \sim U[0, N_S^{-1}]$. FOR $j = \overline{1, N_S}$ – Move along DF: $u(j) = u(1) + N_S^{-1}(j-1)$ – WHILE $u(j) > c(i)$ – $i = i + 1$ – END WHILE – Assign new value: $x^{j^*}(k) = x^i(k)$ – Assign weight: $w^j(k) = N_S^{-1}$ – Assign basic index: $i^j = i$ END FOR
3	Sequential Importance Sampling with Resampling Filter (SISR)	Algorithm 3: SIR Particle Filter $\left[\{x^i(k), w^i(k)\}_{i=1}^{Ns}\right]$= SIR $\left[\{x^i(k), w^i(k)\}_{i=1}^{Ns}, z(k)\right]$ FOR $i = \overline{1, N_S}$ – Generate $x^i(k) \sim p(x(k)\|x^i(k-1))$ – Compute $w^i(k) = p(z(k)\|x^i(k))$ END FOR Compute total weight: $t = \sum_{i=1}^{Ns} w^i(k)$ FOR $i = \overline{1, N_S}$ – Normalize i-th weight: $w^i(k) = t^{-1} w^i(k)$ END FOR Perform resampling using algorithm 2 (Resampling Algorithm): – $\left[\{x^i(k), w^i(k), -\}_{i=1}^{Ns}\right]$= RESAMPLE $\left[\{x^i(k), w^i(k)\}_{i=1}^{Ns}\right]$

Source: **compiled by the authors**

At this stage, a new set of random values is generated $\{x^{i^*}(k)\}_{i=1}^{N_S}$, an approximate discrete distribution is used $P\left(x(k)\middle|z(1:k)\right)$, which is calculated using (20), thus

$$P\{x^{i^*}(k) = x^j(k)\} = w^j(k).$$

The numbers generated in this way create a sequence of independent identically distributed random numbers from the distribution (20) with weighting coefficients; $w^i(k) = 1/N_S$. The numbers generated in this way create a sequence of independent identically distributed random numbers from the distribution (20) with weighting coefficients; $w^i(k) = 1/N_S$.

The pseudocode of the proposed resampling procedure is given in Table 1. The procedure is computationally simple point of view, and also provides for saving the indices of each element of the new sample, due to the use of the index from the previous sample for further use.

The procedure has a number of disadvantages: it reduces the possibilities for parallel calculations;

particles with a large weight can be used repeatedly. This is called impoverishment of the sample, and then all the particles can converge into one particle in a few iterations.

Sequential importance sampling with resampling filter (*SISR*). A Monte Carlo procedure is proposed that can be used to solve the problems of recursive Bayesian filtering. The procedure has practically no restrictions on its application.

The functions $\boldsymbol{f}(\cdot,\cdot)$ and $\boldsymbol{h}(\cdot,\cdot)$ in (1) and (2) must be known; it should also be possible to generate pseudo-random sequences of the noise distribution, $P\big(v(k-1)\big)$, and the prior distribution, $P(x(k)|x(k-1))$, as well as determine the value of the density distribution $P(z(k)|x(k))$, at certain points with an accuracy of at least up to a common constant.

The SISR algorithm was obtained from the SIS algorithm with the appropriate selection of the following elements:

- the supply density,

$Q\left(x(k)\middle|x^i(k-1),z(k)\right)$, can be replaced by the prior distribution, $P\left(x(k)\middle|x^i(k-1)\right)$;

- a resampling step is performed at each time point.

This choice of supply density proves the necessity of selecting implementations from $P\left(x(k)\middle|x^i(k-1)\right)$.

The implementation of $x^i(k)\sim P(x(k)|x^i(k-1))$ can be done by first generating noise and then computing $x^i(k)=f\left(x^i(k-1),v^i(k-1)\right)$.

For this particular choice of supply density, the weight update expression takes the form (22). Given that resampling is carried out at each moment of time, we have $w^i(k-1)=1/N_S\ \forall i$, and then

$$w^i(k)\propto P\left(z(k)\middle|x^i(k)\right). \qquad (24)$$

The weights specified in (24) are normalized before the resampling phase. The pseudocode of the algorithm is given in Table 1.

DEVELOPMENT OF THE ARCHITECTURE OF THE INFORMATION ANALYTIC SYSTEM

To solve forecasting problems, the architecture of the information-analytical system is proposed (Fig. 2). The system consists of the following main components: user interface, information storage subsystem, data analysis and preprocessing subsystem, modeling subsystem, forecast construction and evaluation subsystem, visualization subsystem.

The information storage subsystem contains the necessary computational procedures, sets of models and forecast quality criteria, statistical data, and relevant expert assessments. The data and knowledge required for their further processing are collected and stored in a database and knowledge base (DKB).

The subsystem of data analysis and preliminary processing consists of the following components: a unit of analysis and evaluation of probabilistic statistical information and a unit of data filtering. In turn, the analysis block provides the following data pre-processing procedures: identification and filling in of gaps in data, detection of anomalous values and their processing, identification of nonlinearities, non-stationarity of data and their types, normalization of data. Because different types of filters produce different effects on the data, they are best applied to combined statistical data filtering procedures capable of producing the desired smoothing effects. The information-analytic forecasting system uses a block of combined filtering based on digital, optimal and probabilistic Bayesian filters.

The simulation subsystem consists of a procedure for dividing the data set prepared for simulation into two samples (training and test) and a simulation block. The modelling block involves the development of basic alternative forecast models and their quality assessment based on a set of criteria.

The subsystem for building and evaluating forecasts consists of a block for building forecasts based on basic models, a block for combining forecasts, a block for ensemble learning, and a block for assessing the quality of forecasts based on a set of quality metrics. The functional capabilities of the system are easily modified and expanded thanks to the block design of the system. Blocks for combining forecast values and ensemble learning are provided in the forecasting subsystem for the opportunity to improve the quality of forecast values of basic models.

The visualization subsystem is designed to visualize the performance of each subsystem and make quick decisions at each step of data processing.

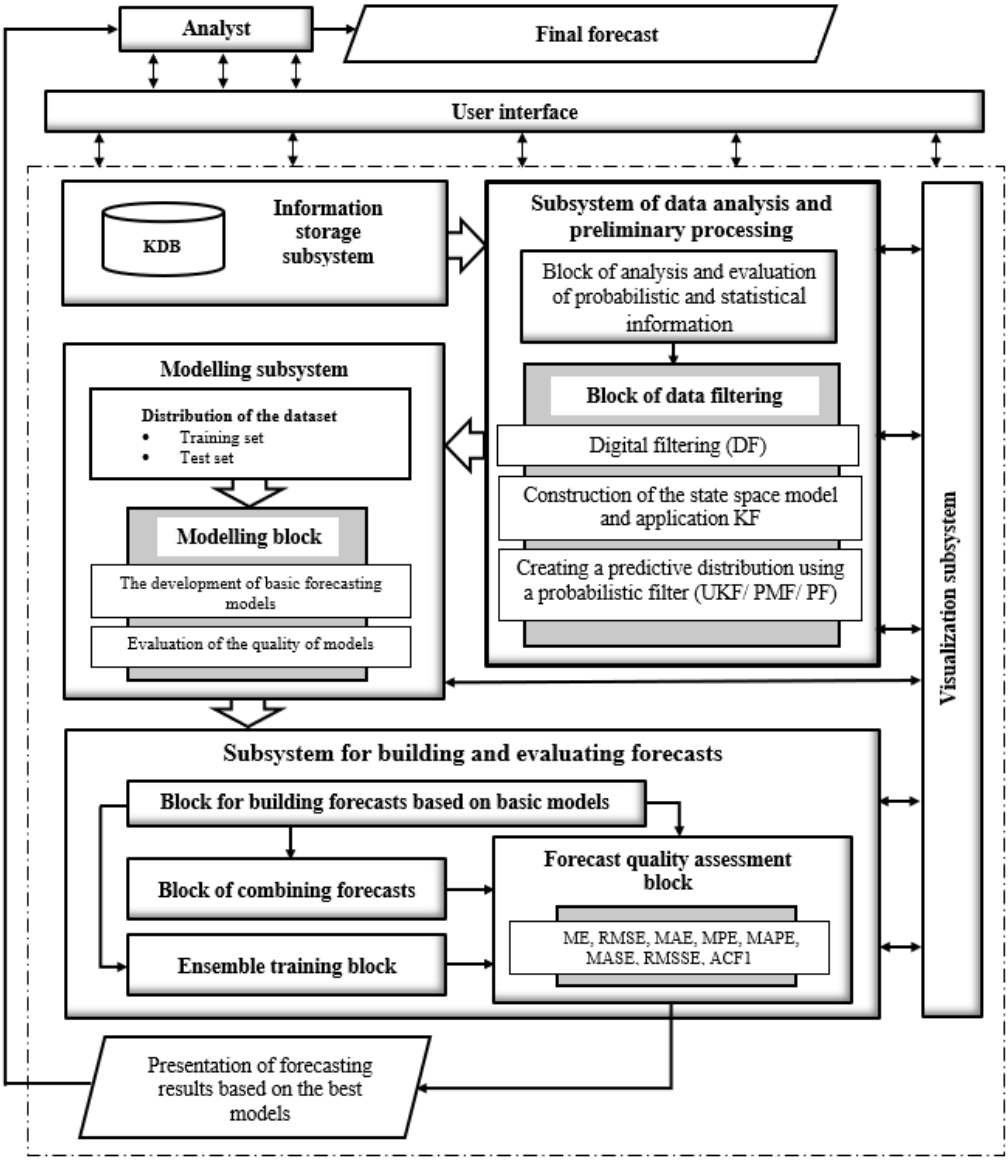

Fig. 2. **The architecture of the information-analytic forecasting system**

Source: **compiled by the authors**

Example. As an example of forecasting based on the systematic use of non-linear filtering methods, the task of forecasting the prices of *Google* shares is considered. A data set is loaded into the information storage system, which contains information about the value of the company in the period from January 1, 2016 to May 26, 2019. These data were collected from the site *https://finance.yahoo.com/*.

After loading in the analysis subsystem and pre-preparation of the data, the analysis of the structure and types of the data was first performed, and the missing values were processed. The data is characterized by irregular registration of observations, which leads to a large number of missing values and masking of possible seasonal fluctuations. This makes the forecasting task quite difficult. Kalman smoothing was used to restore gaps in the time series [31, 32]. Using a set of statistical tests (ADF, KPSS, PP), the original series was checked for stationarity. The result of the verification was a conclusion about the non-stationarity of the process, which is reflected by the set of observed time series values. No stationarity of the process is confirmed by the nature of the values of the sample autocorrelation functions ACF and PACF. Visual analysis of the data made it possible to decide on the choice of modeling method. First of all, one should take into account the dominant role of the trend present in the data, which represents non-linear and non-stationary behavior. There are also templates that reflect the seasonal behavior of

the data to be reflected in the models. However, the degree of their influence is much smaller. To implement the filtering procedure, various types of nonlinear filters from the filtering block were used.

ARIMA statistical models, the method of fitting generalized additive models (GAM), Bayesian structural time series models (BSTS) [33] and forward propagation artificial neural networks (NNAR) are used in the modeling block as basic forecasting models. These methods were chosen because of their ability to recognize complex patterns in time series.

Table 2, Table 3 and Table 4 show a comparison of the results of the quality assessment of basic models and forecast values without application and with various options for applying filters.

Table 2. **The quality of models and forecasts without the use of a filtering unit**

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
ARIMA $(0,1,0)(2,0,0)_7$	0.99	25487.25	2.18	67.93	62.57	5.19	0.047
GAM (annual and weekly seasonal components)	0.99	26655.77	2.21	87.02	83.88	6.13	0.052
BSTS (the component of the linear local trend + the component of the autoregressive process)	0.99	25391.39	2.13	42.81	40.56	4.27	0.033
NNAR (n=10, Sigmoid, maxit=5000)	0.99	25088.74	2.11	37.29	32.72	3.99	0.026

Source: **compiled by the authors**

Table 3. **Quality of models and forecasts using digital filtering**

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
ARIMA $(0,1,0)(2,0,0)_7$	0.99	23355.54	2.10	65.01	60.73	4.54	0.045
GAM (annual and weekly seasonal components)	0.99	24132.15	2.08	85.29	79.34	5.06	0.048
BSTS (the component of the linear local trend + the component of the autoregressive process)	0.99	23861.65	2.07	38.15	36.11	3.79	0.030
NNAR (n=10, Sigmoid, maxit=5000)	0.99	21887.54	2.05	33.35	29.52	3.04	0.021

Source: **compiled by the authors**

Table 4. **Quality of models and forecasts with systematic application of nonlinear filtering methods**

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
ARIMA (0,1,0)(2,0,0)	0.99	24453.1	2.12	62.80	57.65	4.09	0.037
GAM (annual and weekly seasonal components)	0.99	24335.12	2.13	83.45	76.12	4.88	0.035
BSTS (the component of the linear local trend + the component of the autoregressive process)	0.99	25061.08	2.10	34.07	30.75	3.27	0.029
NNAR (n=10, Sigmoid, maxit=5000)	0.99	23881.14	2.07	29.24	23.13	2.71	0.019

Source: **compiled by the authors**

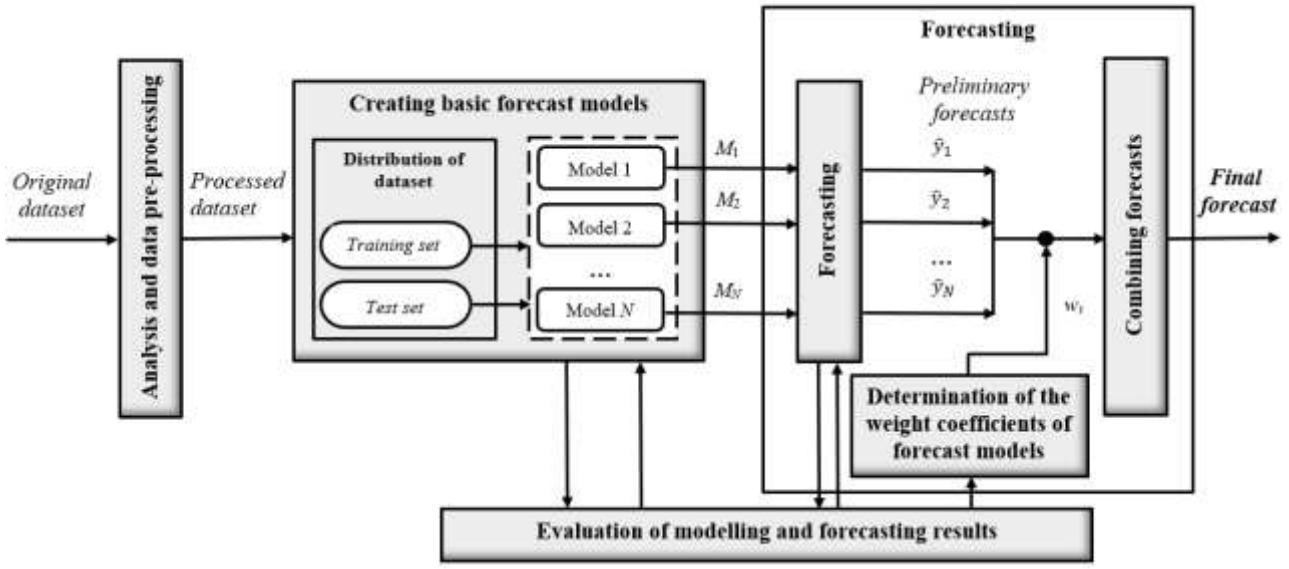

Fig. 3. **Scheme of the method of improving the quality of forecast values**

Source: **compiled by the authors**

The digital filter is used in the form of exponential smoothing based on the Holt and Holt-Winters methods, due to the presence of trend and seasonality in the data. The filter prepares the data to build the state space model that is needed to apply the optimal filters. An extended Kalman filter of the first order is implemented as an optimal filter. The evaluation of the process, which is made by EKF, is used to build an acceptable model of dispersion dynamics. The probabilistic filter generates a predictive variance distribution that is needed to estimate the predictive values. A granular filter with a basic sequential importance sampling (SIS) algorithm was used as a probabilistic filter. The combination of filters in the filtering unit was selected experimentally. Filter parameters were determined experimentally. From the above results, it can be concluded that the systematic use of nonlinear filtering methods significantly improves the quality indicators of basic models.

To improve the quality of forecasting on the basis of prepared data and on the basis of nonlinear filtering methods, the method [34] was applied to solve the forecasting problem, the structural diagram of which is presented in Fig. 3. The first stage of the method is the process of analysis and preprocessing of the data set. At this stage, the following procedures are implemented: detection and processing of gaps in the data set, detection of anomalies, checking for non-linearity, non-stationarity and their consideration, filtering and smoothing of data, etc. After this stage, the primary data set is completely prepared for the modeling process. At the second stage, the data set is divided into two parts: training and test. The next step of the modeling stage is the construction of basic predictive models. The base models are built on the basis of selected methods.

They are checked for adequacy using quality metrics, the values of which are transferred to the model evaluation results block. Preliminary forecasts are formed from the basic models at the forecasting stage. Assessments of the quality of models are the basis for the formation of weighting factors when combining forecasts. The final stage of the methodology is the stage of combining, at which the method of combining is determined and its effectiveness is determined. If an improvement in forecast accuracy is not found, it is necessary to return to the stage of forming basic models, or to change their number and type of combination. Such a structural scheme fully corresponds to the process of building combined forecasts for time series based on simple averaging of forecasts, weighted combination of forecasts and regression [33]. To increase the accuracy of the combined forecast, the forecasting procedure is performed on the models with close variance values. The GAM and ARIMA models have variance values that are significantly different from the variance of the other two models. Therefore, these models were not considered in the next iteration of combining forecasts. Table 5 shows a comparison of the forecast scores for the Google

time series for the BSTS model, NNAR, and the combined model.

Table 5. **Comparison of estimates of forecasting results for time series *Google***

Model	RMSE	MAE	MAPE	Theil
BSTS	34.07	30.75	3.28	0.027
NNAR	29.24	23.13	2.95	0.019
Combination	27.24	21.13	2.71	0.017

Source: **compiled by the authors**

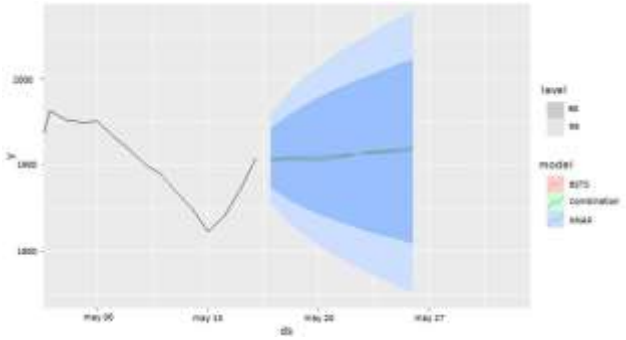

Fig. 4. **Results of time series forecasting using a combined model**

Source: **compiled by the authors**

From the analysis of the table, it follows that the combined forecast model exhibits the best quality indicators compared to the base models. A graphical representation of the prediction results using the combined model is shown in Fig. 4. The 80 % and 95 % prediction intervals for each component and their combination are shown. The figure shows only the forecast part.

CONCLUSIONS

The study investigated the systematic use of nonlinear data filtering methods in the problems of intelligent data analysis and machine learning. The analysis of modern methods of digital, optimal and probabilistic nonlinear filtering of statistical data and the peculiarities of their application in solving the problems of evaluating the states of dynamic systems is carried out. The application of the Kalman filter and its modifications for solving nonlinear filtering problems is analyzed. The classification of nonlinear filtering methods is presented. The basis of the classification is digital, optimal and probabilistic filters. Non-recursive and recursive digital filters are studied. The formulation of the problem of optimal filtering based on the Kalman filter is considered. The filtering equation for a free dynamic system based on the state space model of a discrete system is given. The extended Kalman filter and its modifications are considered. The Bayesian method of estimating the state of a nonlinear stochastic system is presented. The problem of linear and nonlinear probabilistic filtering is considered. Three filters are considered as examples of probabilistic filters: an unscented Kalman filter, a point mass filter, and a granular filter. The granular filtering algorithm and its modifications are considered in detail. The architecture of the information-analytical system for solving forecasting problems has been developed. As an example of forecasting non-stationary process based on the systematic use of non-linear filtering methods, the task of forecasting the prices of Google shares is considered. A comparison of the quality assessment results of basic models and forecast values without filtering and with different options for applying filters was carried out. To improve the quality of forecasting based on prepared data and based on nonlinear filtering methods, a method based on combined forecasts was used to solve the forecasting problem. The systematic use of non-linear filtering methods increases the efficiency of data preparation when solving the problems of intelligent data analysis and machine learning.

REFERENCES

1. Kalman, R. E. “A new approach to linear filtering and prediction problems”. ASME. *Journal of Basic Engineering*. 1960; 82 (1): 35–45. DOI: http://dx.doi.org/10.1115/1.3662552.

2. Kalman, R. E., & Bucy, R. S. “New results in linear filtering and prediction theory”. ASME. *Journal of Basic Engineering*. 1961: 83(1): 95–108. DOI: http://dx.doi.org/10.1115/1.3658902.

3. Jazwinski, A. H. “Stochastic processes and filtering theory”. *Academic Press.* 1970, https://www.scopus.com/authid/detail.uri?authorId=7801321826.

4. Bishop, C. M. “Pattern recognition and machine learning”. *Information Science and Statistics*. *Springer-Verlag.* 2006. 4. Berlin: Heidelberg. https://www.scopus.com/authid/detail.uri?authorId=7202946665.

5. Anderson, B. D. O. & Moore, J. B. “Optimal Filtering”. *Englewood Cliffs: Prentice-Hall*. 1979, https://www.scopus.com/authid/detail.uri?authorId=7404263205.

6. Bidyuk, P. I. et al. “Time series analysis” (in Ukrainian). *Polytechnika, NTUU KPI*. Kyiv: Ukraine. 2013, https://www.scopus.com/authid/detail.uri?authorId=6602445011.

7. Kay, S. M. "Fundamentals of statistical signal processing: Estimation theory". *Upper Saddle River: Prentice Hall PTR*. 1993, https://www.scopus.com/authid/detail.uri?authorId=7202568468.

8. Chui, C. K. & Chen, G. "Kalman filtering with real-time applications". *Springer-Verlag*. 2009, https://www.scopus.com/authid/detail.uri?authorId=24297847200. DOI: http://dx.doi.org/10.1007/978-3-540-87849-0.

9. Haykin, S. "Adaptive filtering theory". *4th ed. Upper Saddle River: Prentice-Hall*. 2002, https://www.scopus.com/authid/detail.uri?authorId=7101764513.

10. Gibbs, B. P. "Advanced Kalman Filtering, Least-Squares and modeling". *Hoboken: John Wiley & Sons.* 2011, https://www.scopus.com/authid/detail.uri?authorId=7103103137.
DOI: http://dx.doi.org/10.1002/9780470890042.

11. Pantyeyev, R. L., Timoshchuk, O. L., Huskova, V. H. & Bidyuk, P. I. "Data filtering techniques in decision support systems". *KPI Science News*. 2021; 1: 16–31, https://www.scopus.com/authid/detail.uri?authorId=57188696085.
DOI: http://dx.doi.org/10.20535/kpisn.2021.1.231205.

12. Valipour, M. & Ricardez-Sandoval, L.A. "Abridged gaussian sum extended Kalman filter for nonlinear state estimation under non-Gaussian process uncertainties". *Computers & Chemical Engineering.* 2021; 155 (10–12): 107534, https://www.scopus.com/authid/detail.uri?authorId=57221165001.
DOI: http://dx.doi.org/10.1016/j.compchemeng.2021.107534.

13. Wang, Q., Sun, X. & Wen, C. "Design method for a higher order extended Kalman Filter Based on Maximum Correlation Entropy and a Taylor Network system". *Sensors.* 2021; 21 (17): 5864, https://www.scopus.com/authid/detail.uri?authorId=57211301492.
DOI: http://dx.doi.org/10.3390/s21175864.

14. Liu, M., Gao, S., Li, W. & Xie, B. "Modified Iterative extended Kalman filter based on state augmentation". *Appl. Electron. Tech.* 2020; 46: 79–81, https://www.scopus.com/authid/detail.uri?authorId=56517197100.

15. Ito, K. & Xiong, K. "Gaussians filters for nonlinear filtering problems". *IEEE Trans. Autom. Control.* 2000; 45: 910–927, https://www.scopus.com/authid/detail.uri?authorId=7410401247.
DOI: http://dx.doi.org/10.1109/9.855552.

16. Gao, B., Hu, G. & Gao, S. "Multi-sensor optimal data fusion based on the adaptive fading unscented Kalman filter". *Sensors*. 2018; 18 (2): 488, https://www.scopus.com/authid/detail.uri?authorId=56047606300. DOI: http://dx.doi.org/10.3390/s18020488.

17. Arasaratnam, I. & Haykin, S. "Cubature Kalman Filter". *IEEE Transactions on Automatic Control*. 2009; 54 (6): 1254–1269, https://www.scopus.com/authid/detail.uri?authorId=7101764513.
DOI: http://dx.doi.org/10.1109/TAC.2009.2019800.

18. Yang, F., Luo, Y. & Zheng, L. "Double-Layer cubature Kalman filter for nonlinear estimation". *Sensors.* 2019; 19(5): 986, https://www.scopus.com/authid/detail.uri?authorId=55511198800.
DOI: http://dx.doi.org/10.3390/s19050986.

19. Garcia, R. V., Pardal, P. & Kuga, H. K. "Nonlinear filtering for sequential spacecraft attitude estimation with real data: Cubature Kalman filter, unscented Kalman filter and extended Kalman filter". *Advances in Space Research.* 2019; 63 (2): 1038–1050, https://www.scopus.com/authid/detail.uri?authorId=35723676800. DOI: https://doi.org/10.1016/j.asr.2018.10.003.

20. Zhang, A., Bao, S. D. & Gao, F. "A novel strong tracking cubature Kalman filter and its application in maneuvering target tracking". *Chinese Journal of Aeronautics.* 2019; 32 (11): 2489–2502, https://www.scopus.com/authid/detail.uri?authorId=8855736400.
DOI: http://dx.doi.org/10.1016/j.cja.2019.07.025.

21. Fan, Z. E., Weng, S. Q. & Jiang, J. "Particle filter object tracking algorithm based on sparse representation and nonlinear resampling". *Journal of Beijing Institute of Technology*. 2018; 27 (1): 51–57, https://www.scopus.com/authid/detail.uri?authorId=36985917200.
DOI: http://dx.doi.org/10.15918/j.jbit1004-0579.201827.0107.

22. Davis, B. & Blair, W. D. "Adaptive gaussian mixture modeling for tracking of long-range targets". In *Proceedings of the IEEE Aerospace Conference, Big Sky*, MT. USA. 2016. p. 1–9, https://www.scopus.com/authid/detail.uri?authorId=16686127400.

DOI: http://dx.doi.org/10.1109/AERO.2016.7500839.

23. Davis, B. & Blair, W.D. “Gaussian mixture approach to long-range radar tracking with high range resolution”. In: *Proceedings of the IEEE Aerospace Conference, Big Sky*. MT. USA. 2015. p. 1–9, https://www.scopus.com/authid/detail.uri?authorId=16686127400.
DOI: http://dx.doi.org/10.1109/AERO.2015.7119222.

24. Koopman, S. J. “State space time series analysis”. *Department of Econometrics VU University Amsterdam, Tinbergen Institute*. 2011, https://www.scopus.com/authid/detail.uri?authorId=6701473976. – Available from: https://personal.vu.nl/s.j.koopman/documents/2011TSEweek1.pdf. – [Accessed: 14 November 2022].

25. Smith, J. O. “Introduction to digital Filters with audio applications”. *Center for Computer Research in Music and Acoustics (CCRMA)*. Stanford University. 2007, https://www.scopus.com/authid/detail.uri?authorId=7410167523.

26. Trofymchuk, O., Bidyuk, P., Kalinina, I. & Gozhyj, A. “Financial risk estimation in conditions of stochastic uncertainties”. *Lecture Notes in Computational Intelligence and Decision Making. Lecture Notes on Data Engineering and Communications Technologies*. *Springer, Cham*. 2022; 77: 3–24, https://www.scopus.com/authid/detail.uri?authorId=56110310300. DOI: http://dx.doi.org/10.1007/978-3-030-82014-5_1.

27. Haug, A. J. “A tutorial on bayesian estimation and tracking techniques applicable to nonlinear and non-gaussian processes”. *MITRE Corporation Technical Report (McLean, Virginia), MTR 05W0000004*. 2005, https://www.scopus.com/authid/detail.uri?authorId=7102013560.

28. Julier, S. J. & Uhlmann, J. K. “Unscented filtering and nonlinear estimation”. *Proc. of the IEEE*. 2004; 92 (3): 401–422, https://www.scopus.com/authid/detail.uri?authorId=7003972937. DOI: http://dx.doi.org/10.1109/JPROC.2003.823141.

29. Luo, X. & Moroz, L. M. “Ensemble Kalman filters with the unscented transform”. *Computer Science, Engineering*. *Physica D Nonlinear Phenomena*. 2009; 238 (5): 549–562, https://www.scopus.com/authid/detail.uri?authorId=8367899300.
DOI: http://dx.doi.org/10.1016/j.physd.2008.12.003.

30. Gustafsson, F. “Particle filter theory and practice with positioning applications”. *IEEE. Aerospace and Electronic Systems Magazine*. 2010; 25 (7): 53–82, https://www.scopus.com/authid/detail.uri?authorId=35546404000.
DOI: http://dx.doi.org/10.1109/MAES.2010.5546308.

31. Kim, Y. & Bang H. “Introduction to Kalman filter and its applications”. *Open Access Peer-Reviewed Chapter*. 2018, https://www.scopus.com/authid/detail.uri?authorId=55699623600. DOI: http://dx.doi.org/10.5772/intechopen.80600.

32. Pei, Y., Biswas, S., Fussell D. S. & Pingali K. “An elementary introduction to Kalman filtering”. *Communications of the ACM*. 2017; 62 (11), https://www.scopus.com/authid/detail.uri?authorId=7003277701. DOI: http://dx.doi.org/10.1145/3363294.

33. Kalinina, I. A. & Gozhyj A. P. “Modeling and forecasting of nonlinear nonstationary processes based on the Bayesian structural time series”. *Applied Aspects of Information Technology*. 2022; 5 (3): 240–255. DOI: http://dx.doi.org/10.15276/aait.05.2022.17.

34. Kalinina, I., Bidyuk, P., Gozhyj, A. & Malchenko P. “Combining forecasts based on time series models in machine learning tasks”. *CEUR Workshop Proceedings*. 2023; 3426: 25–35, https://www.scopus.com/authid/detail.uri?authorId=57198354193. – Available from: CEUR-WS.org/Vol-3426/paper2.pdf. – [Accessed: 14 June 2023].

Conflicts of Interest: the authors declare no conflict of interest

Received 25.09.2023
Received after revision 30.11.2023
Accepted 14.12.2023

DOI: https://doi.org/10.15276/aait.06.2023.23
УДК 004.852:004.6

Системне використання нелінійних методів фільтрації даних в задачах прогнозування

Гожий Олександр Петрович [1)]
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
Калініна Ірина Олександрівна [1)]
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193
Бідюк Петро Іванович [2)]
ORCID: https://orcid.org/0000−0002−7421−3565; pbidyke@gmail.com. Scopus Author ID: 6602445011
[1)] ЧНУ імені Петра Могили, вул. 68 Десантників, 10. Миколаїв, 54000, Україна
[2)] Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Перемоги, 37. Київ, 03056, Україна

АНОТАЦІЯ

У статті описано підхід до системного використання методів нелінійної фільтрації даних в задачах інтелектуального аналізу даних та машинного навчання. Розглянуто поняття фільтрації та нелінійної фільтрації. Проведено аналіз сучасних методів оптимальної та ймовірнісної нелінійних фільтрацій статистичних даних й особливості їх застосування в розв'язанні задач оцінювання станів динамічних систем. Проаналізовано застосування фільтра Калмана та його різновидів для вирішення задач нелінійної фільтрації. Наведено класифікацію методів нелінійної фільтрації. Основу класифікації складають цифрові, оптимальні та ймовірнісні фільтри. Досліджено нерекурсивні та рекурсивні цифрові фільтри. Розглянуто постановка задачі оптимальної фільтрації на основі фільтра Калмана. Приведено рівняння фільтрації для вільної динамічної системи, засноване на моделі простору станів дискретної системи. Розглянуто розширений фільтр Калмана і його модифікації. Представлено байєсівський метод оцінки стану нелінійної стохастичної системи**.** Розглянуто проблема лінійної та нелінійної ймовірнісної фільтрації. В якості прикладів ймовірнісних фільтрів розглянуто три фільтра: фільтр Калмана без запаху, фільтр точкової маси та гранулярний фільтр. Детально розглянуто алгоритм гранулярної фільтрації та його модифікації. Розроблено архітектуру інформаційно аналітичної системи для вирішення задач прогнозування. Система складається з наступних основних компонентів: інтерфейс користувача, підсистема зберігання інформації, підсистема аналізу та попередньої обробки даних, підсистема моделювання, підсистема побудови та оцінки прогнозів, підсистема візуалізації. В якості прикладу прогнозування на основі системного використання методів нелінійної фільтрації розглянуто завдання прогнозування цін акцій компанії *Google*. Проведено порівняння результатів оцінювання якості базових моделей та прогнозних значень без фільтрації та з різними варіантами застосування фільтрів. Для покращення якості прогнозування на підготовлених даних та на основі методів нелінійної фільтрації для вирішення задачі прогнозування застосовано метод на основі комбінованих прогнозів. Представлено результати прогнозування з використанням комбінованої моделі.

Ключові слова: нелінійна фільтрація; оптимальний фільтр Калмана; розширений фільтр Калмана; ймовірнісний фільтр; алгоритми гранулярної фільтрації; інформаційно-аналітична система; комбіновані прогнози

ABOUT THE AUTHORS

Aleksandr P. Gozhyj - Doctor of Engineering Sciences, Professor, Professor Department of Intelligent Information Systems, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str, Mykolaiv, 54000, Ukraine
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
***Research field*:** artificial intelligence methods; machine learning; probabilistic-statistical methods; intelligent web technologies

Гожий Олександр Петрович - доктор технічних наук, професор, професор кафедри Інтелектуальних інформаційних систем ЧНУ імені Петра Могили, вул. 68 Десантників, 10. Миколаїв, 54000, Україна

Irina A. Kalinina - PhD, Associate Professor, Associate Professor Department of Intelligent Information Systems, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str, Mykolaiv, 54000, Ukraine
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193
Research field: probabilistic-statistical methods; machine learning; forecasting methods; coloured Petri nets

Калініна Ірина Олександрівна - кандидат технічних наук, доцент, доцент кафедри Інтелектуальних інформаційних систем ЧНУ імені Петра Могили, вул. 68 Десантників, 10. Миколаїв, 54000, Україна

Peter I. Bidyuk - Doctor of Engineering Sciences, Professor, Professor Department of Mathematical Methods of System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, Peremoga Ave. Kyiv, 03056, Ukraine
ORCID: https://orcid.org/0000−0002−7421−3565; pbidyke@gmail.com. Scopus Author ID: 6602445011
***Research field*:** probabilistic-statistical methods; analysis of nonlinear systems; forecasting methods; methods of time series analysis

Бідюк Петро Іванович - доктор технічних наук, професор, професор кафедри Математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Перемоги, 37. Київ, 03056, Україна