

Методика Интеграции Формальных Методов Прогнозирования Временных Рядов в Data Assimilation

Ю.С. Тимошенкова, С.В. Поршневу, Н.Т. Сафиуллин

Аннотация—В статье описан разработанный авторами способ интеграции в метод *Data Assimilation (DA)* формальных методов прогнозирования временных рядов (ВР) (авторегрессии-проинтегрированного скользящего среднего, сингулярного спектрального анализа, метода группового учета аргумента, нейронных сетей с долговременной памятью), применяемых в тех случаях, когда неизвестна математическая модель динамической системы, породившей данный ВР (например, ВР, составленные из значений экономических показателей). Работоспособность предложенного способа интеграции проиллюстрирована на примере прогнозирования ВР «Air Passengers» методом *DA* на основе ансамблевого фильтра Калмана, в который интегрирован метод *ARIMA*. Получены оценки точности прогнозов значений выбранного ВР, вычисленные с помощью *ARIMA* и с помощью предложенного метода. Обсуждаются преимущества и недостатки предложенного способа интеграции метода *DA* и формальных методов прогнозирования ВР, а также направления его дальнейшего совершенствования.

Ключевые слова—Анализ временных рядов, прогнозирование временных рядов, *Data Assimilation*.

I. ВВЕДЕНИЕ

Задача прогнозирования одномерного временного ряда (ВР), под которым понимают значения некоторого процесса, порожденного изучаемой динамической системой (ДС), зафиксированные в упорядоченные в порядке возрастания моменты времени, является актуальной в различных областях человеческой деятельности, например, в экономике [1], метеорологии, климатологии [2], океанографии [3] и др. В общем случае данная задача состоит в предсказании на основе известных значений анализируемого ВР его значений в последующие моменты времени.

Множество задач прогнозирования ВР в зависимости от типа ВР можно разделить на следующие классы.

Статья получена 17 января 2022

Тимошенкова Юлия Сергеевна, Уральский Федеральный Университет, Институт радиоэлектроники и информационных технологий-РТФ, (e-mail: julia.timoshenkova@urfu.ru).

Поршневу Сергей Владимирович, Уральский Федеральный Университет, Институт радиоэлектроники и информационных технологий-РТФ, (e-mail: s.v.porshnev@urfu.ru).

Сафиуллин Николай Тахирович, Уральский Федеральный Университет, Институт радиоэлектроники и информационных технологий-РТФ, (e-mail: n.t.safiullin@urfu.ru).

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90006

1. Прогнозирование ВР, порожденных ДС, для описания эволюции которых существуют те или иные математические модели.
2. Прогнозирование ВР, порожденных ДС, не имеющих математических моделей вследствие их сложности.

Задачи первого класса возникают в автоматизированных системах управления различными техническими системами, например, ракетно-космической техники [4], задачи второго класса – в экономике, например, при прогнозировании курсов валют [5].

Соответственно, изучают два класса методов прогнозирования ВР:

- 1) методы прогнозирования ВР, опирающиеся на строгое математическое описание ДС, породившей данный ВР (неформальные методы);
- 2) методы прогнозирования ВР, опирающиеся только на сам ряд исходных наблюдений, без учета наличия или отсутствия информации об исходной ДС, породившей данный ВР (формальные методы);

К неформальным методам прогнозирования ВР относится, например, метод фильтра Калмана [6] (во всех его известных модификациях), к формальным методам прогнозирования ВР – метод авторегрессии и скользящего среднего (*Autoregressive Integrated Moving Average, ARIMA*) [7]; метод сингулярного спектрального анализа (*Singular spectrum analysis, SSA*), в русскоязычной научной литературе известный как метод «Гусеница» [8]; метод группового учета аргумента (МГУА), в англоязычной литературе – *Group Method of Data Handling (GMDH)* [9]; метод прогнозирования на основе использования рекуррентных нейронных сетей (*Recurrent Neural Network, RNN*) [7].

На практике решается, в том числе, задача прогнозирования таких ВР, у которых известны как спрогнозированные, так и соответствующие фактические (измеряемые) значения (например, краткосрочное прогнозирование погоды). Наличие точных и спрогнозированных значений ВР, синхронизированных во времени, обеспечивает возможность их сравнения и при необходимости последующей корректировки используемой модели ВР и/или уточнения спрогнозированных значений данного ВР в последующие моменты времени. Описанный подход реализован, например, в методе ассимиляции данных (*Data Assimilation, DA*) [4], относящийся к

классу неформальных методов прогнозирования ВР.

Напомним, что сегодня известны несколько модификаций *DA*, в том числе: 1) модификация, основанная на вычислении прогнозируемых значений ВР через построение функционала, обеспечивающего близость значений ВР, рассчитанных на основе математической модели динамической системы, породившей данный ВР, и наблюдаемых значений ВР; 2) модификация, основанная на использовании статической теории оценивания и фильтра Калмана, в которой ВР рассматривается как аддитивная смесь информативного сигнала и случайного процесса (шума) с известными статистическими свойствами.

Примеры успешного использования *DA* для прогнозирования ВР приведены в [10]–[13]. Однако, как показал анализ многочисленных работ, посвященных формальным методам прогнозирования ВР (см., например, [13] и др.), данные методы совместно с *DA* не использовались. В этой связи разработка способов интеграции формальных методов прогнозирования ВР и *DA* является актуальной задачей.

В статье описан разработанный авторами способ интеграции формальных методов прогнозирования ВР в метод *DA*, основанный на использовании статической теории оценивания и фильтра Калмана, приведены примеры использования для прогнозирования реальных ВР и получены оценки точности прогноза.

II. МАТЕРИАЛЫ И МЕТОДЫ РАБОТЫ

A. Классический *DA*

В методе *DA* полагается, что эволюция исследуемой ДС, характеризующейся в дискретном времени t_k вектором состояния \bar{x}_k , описывается следующей системой уравнений [2], [12]:

$$\begin{cases} \bar{x}_{k+1} = M(\bar{x}_k, t_k) + \bar{w}_k, \\ \bar{y}_k = H(\bar{x}_k, t_k) + \bar{v}_k, \end{cases} \quad (1)$$

где \bar{x}_k , $k=0,1,\dots$ – вектор состояния динамической системы в моменты времени t_k , $k=0,1,\dots$; M – оператор перехода (математическая модель, описывающая динамику системы); \bar{x}_{k+1} – спрогнозированное значение вектора состояния ДС в момент времени t_{k+1} , \bar{y}_k – вектор наблюдений, $H(\bar{x}_k, t_k)$ – оператор наблюдения, связывающий вектор состояния ДС с вектором наблюдений; \bar{w}_k , \bar{v}_k – ошибки модели и наблюдения, соответственно.

Первое уравнение в (1) называется уравнением прогноза, второе уравнение – уравнением наблюдений. Соответственно, алгоритм *DA* разделяется на два этапа: этап прогнозирования и этап корректировки прогноза. Отметим, что корректировка вектора \bar{x}_k в методе *DA* может осуществляться не на каждом шаге прогноза, но в моменты времени t_{k+m} , t_{k+2m} , \dots , m – целое число, равное или большее единицы.

Изначально метод *DA*, обсуждаемый далее, был разработан для линейной ДС, в математической модели которой операторы M и H в (1) являются матрицами, поэтому

$$\begin{aligned} M(\bar{x}_k, t_k) &= M \cdot \bar{x}_k(t_k) = M \cdot \bar{x}_k = \bar{M}_k, \\ H(\bar{x}_k, t_k) &= H \cdot \bar{x}_k(t_k) = H \cdot \bar{x}_k = \bar{H}_k, \end{aligned}$$

а \bar{w}_k , \bar{v}_k – выборки из нормального белого шума с известными значениями ковариационных матриц ошибок Q_k , R_k .

На этапе прогнозирования на основе известного в момент времени t_k вектора состояния ДС \bar{x}_k^f и вектора наблюдений \bar{y}_k вычисляют скорректированное значение вектора состояния ДС по формуле:

$$\bar{x}_k^a = \bar{x}_k^f + K_k^*(\bar{e}_k), \quad (2)$$

где

$\bar{e}_k = \bar{y}_k - H(\bar{x}_k^f)$ – отклонение реального вектора наблюдений \bar{y}_k от спрогнозированного вектора наблюдений, полученного на основе вектора состояния ДС \bar{x}_k^f , в линейном случае: $\bar{e}_k = \bar{y}_k - H \cdot \bar{x}_k^f = \bar{y}_k - \bar{H}_k$;

K_k^* – передаточный коэффициент фильтра Калмана, вычисляемый по формуле:

$$\bar{K}_k^* = P_k^f \bar{H}_k^T (\bar{H}_k P_k^f \bar{H}_k^T + R_k)^{-1}, \quad (3)$$

Здесь P_k^f – ковариационная матрица ошибок прогноза, вычисляемая по формуле:

$$P_k^f = \bar{M}_k (I - \bar{K}_{k-1}^* \bar{H}_{k-1}) P_{k-1}^f \bar{M}_k^T + Q_k, \quad (4)$$

здесь I – единичная матрица.

Также на данном этапе вычисляют ковариационную матрицу ошибок прогноза P_k^a :

$$P_k^a = (I - \bar{K}_k^* \bar{H}_k) P_k^f. \quad (5)$$

На этапе корректировки прогноза с учетом вектора наблюдений \bar{y}_k вычисляют уточненное значение вектора состояния ДС \bar{x}_k^a по формуле:

$$\bar{x}_k^a = \bar{x}_k^f + \bar{K}_k^* (\bar{y}_k - H(\bar{x}_k^f)) = \bar{x}_k^f + \bar{K}_k^* (\bar{y}_k - \bar{H}_k), \quad (6)$$

а также новую ковариационную матрицу ошибок прогноза P_{k+1}^f :

$$P_{k+1}^f = \bar{M}_{k+1} P_k^a \bar{M}_{k+1}^T + Q_k, \quad (7)$$

используемую далее для вычисления \bar{x}_{k+1}^f по формуле:

$$\bar{x}_{k+1}^f = M(\bar{x}_k^a). \quad (8)$$

Затем повторяется описанная выше последовательность действий.

Блок-схема алгоритма *DA*, описанного, следуя [2], [14], представлена на рисунке 1.

Поясним смысл оператора модели и оператора наблюдения в (1) на примере линейного гармонического осциллятора (например, точечного груза массой m , прикрепленного к одному из концов пружины с жесткостью k , второй конец которой закреплен неподвижно), совершающего свободные колебания, который описывается дифференциальным уравнением (ДУ) второго порядка:

$$\ddot{r} + \omega^2 r = 0, \quad (9)$$

где r – смещение точечного груза относительно равновесного положения, $\omega = \sqrt{k/m}$ – циклическая частота колебаний гармонического осциллятора, с начальными условиями $r(0) = r_0$, $\dot{r}(0) = v_0$;

и на примере системы дифференциальных уравнений первого порядка (СДУ), предложенной Э.Н. Лоренцом для описания процесса конвекции в плоском двумерном слое морской воды [15], [16]:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(r - z) - y, \\ \dot{z} = xy - bz \end{cases} \quad (10)$$

где $\sigma = \nu/k$ – число Прандтля, $r = Ra/Ra_c$ – нормированное число Рэлея, $b = 4/(1+a^2)$ – геометрический фактор, x – интенсивность конвекции, y – разность температур между восходящим и нисходящим потоками; z – отклонение вертикального температурного профиля от линейного. СДУ (10) также возникает в следующих математических моделях: конвекции в замкнутой петле; вращения водяного колеса; одностороннего лазера; диссипативного гармонического осциллятора с инерционной нелинейностью. Напомним, что СДУ (10) замечательна тем, что при $\sigma = 10$, $r = 28$, $b = 8/3$ ее решение, получившее название аттрактор Лоренца, оказывается хаотическим.

Для записи в явном виде оператора эволюции M гармонического осциллятора представим ДУ второго порядка (10) в виде системы ДУ первого порядка:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} z_2 \\ -\omega^2 z_1 \end{bmatrix}, \quad (11)$$

где $z_1 = r$, $z_2 = \dot{z}_1$, $\dot{z}_2 = \ddot{z}_1 = \ddot{r} = -\omega^2 r = -\omega^2 z_1$. Из (11) видно, что в пространстве состояний линейный гармонический осциллятор описывается вектором $[z_1 \ z_2]^T$.

Используем в (11) конечно-разностную аппроксимацию производных первого порядка точности:



Рис. 1 – Классический алгоритм метода *DA*: m – длина прогноза, k – номер отсчета прогноза, len – длина вектора наблюдений

$$\begin{bmatrix} \dot{z}_1(t_k) \\ \dot{z}_2(t_k) \end{bmatrix} = \begin{bmatrix} \frac{[z_1]_{k+1} - [z_1]_k}{\tau} \\ \frac{[z_2]_{k+1} - [z_2]_k}{\tau} \end{bmatrix},$$

где τ – шаг интегрирования СДУ(11). Тогда можно записать модель линейного гармонического осциллятора в дискретной форме:

$$\begin{bmatrix} [z_1]_{k+1} \\ [z_2]_{k+1} \end{bmatrix} = \begin{bmatrix} [z_1]_k + [z_2]_k \tau \\ -\omega^2 \tau [z_1]_k + [z_2]_k \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ -\omega^2 \tau & 1 \end{bmatrix} \begin{bmatrix} [z_1]_k \\ [z_2]_k \end{bmatrix}. \quad (12)$$

Из (12) видно, что искомым оператор эволюции линейного гармонического осциллятора есть матрица

$$M = \begin{bmatrix} 1 & \tau \\ -\omega^2 \tau & 1 \end{bmatrix}.$$

В тех случаях, когда проводятся измерения только координаты или скорости линейного гармонического осциллятора, вектор наблюдений \bar{y}_k во втором уравнении системы (1) вырождается в один из скаляров:

$$y_k = [z_1]_k,$$

$$y_k = [z_2]_k,$$

соответственно. Следовательно, в рассматриваемых случаях оператор наблюдений H представляет собой одну из вырожденных матриц:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$H = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

соответственно.

В тех случаях, когда проводятся одновременные измерения и координаты и скорости гармонического

осциллятора, вектор наблюдений будет $\bar{y}_k = \begin{bmatrix} [z_1]_k \\ [z_2]_k \end{bmatrix}$.

Соответственно, оператор наблюдений H представляет собой единичную матрицу.

В тех случаях, когда измеряется ускорение гармонического осциллятора $\ddot{r}(t) = \dot{z}_2(t)$, вектор \bar{y}_k во втором уравнении системы (1) вырождается в скаляр $y_k = H([z_1]_k, [z_2]_k, [\dot{z}_2]_k)$. Принимая во внимание связь между $[z_1]_k$, $[z_2]_k$, $[\dot{z}_2]_k$, понятно, что в рассматриваемом случае оператор наблюдений H является оператором вычисления второй производной от $z_1(t)$, для вычисления которого можно использовать следующую конечно разностную аппроксимацию:

$$H([z_2]_k) = \frac{[z_2]_k - [z_2]_{k-1}}{\tau}.$$

Для нахождения в явном виде оператора эволюции M для СДУ (10), используем конечно разностную аппроксимацию первых производных

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \frac{x_{t+1} - x_t}{\tau} \\ \frac{y_{t+1} - y_t}{\tau} \\ \frac{z_{t+1} - z_t}{\tau} \end{bmatrix},$$

где τ – шаг интегрирования СДУ (10), запишем модель Лоренца в дискретной форме:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 1 - \sigma\tau & \sigma\tau & 0 \\ r\tau & 1 - \tau & -x_t\tau \\ y_t\tau & 0 & 1 - b\tau \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}. \quad (13)$$

Из (13) видно, что при использовании классического метода DA оператор наблюдения в (1) представляет собой матрицу:

$$M = \begin{bmatrix} 1 - \sigma\tau & \sigma\tau & 0 \\ r\tau & 1 - \tau & -x_t \\ y_t\tau & 0 & 1 - b\tau \end{bmatrix}.$$

Оператор наблюдений H при предположении о том, что одновременно наблюдаются значения всех трех переменных x, y, z , – оказывается единичной матрицей.

Отметим, что сегодня разработано множество различных модификаций фильтра Калмана, используемых в DA , в том числе: EKF (*Extended Kalman filter*), $EnKF$ (*Ensemble Kalman Filter*), UKF (*Unscented Kalman Filter*) и др. [2], из которых далее в работе мы используем $EnKF$ (ансамблевый фильтр Калмана). Его

выбор обусловлен тем, что, с одной стороны, он имеет относительно невысокую в сравнении с другими модификациями вычислительную сложность, а с другой стороны, обеспечивает приемлемую точность прогнозирования [12]. Блок-схема DA , в котором используется $EnKF$ -фильтр, представлена на рисунке 2.

Из рисунка 2 видно, что в данной модификации метода DA в каждый момент времени t_k генерируется ансамбль наблюдений:

$$\bar{y}_{k,i}^{(ensemble)} = \bar{y}_k + \bar{u}_i, \quad i = 0, N^{(ensemble)}, \quad (14)$$

где \bar{u}_i – шум с известным законом распределения. Далее, используя случайные значения реализации $\bar{y}_{k,i}^{(ensemble)}$, в соответствие с (6), вычисляется ансамбль скорректированных значений $\bar{x}_{k,i}^{(ensemble)}$, а в качестве скорректированного значения прогноза – усредненное по ансамблю $\bar{x}_{k,i}^{(ensemble)}$ значение $\bar{x}_k^a = \text{mean}(\bar{x}_{k,i}^{(ensemble)})$. При этом вместо корреляционной матрицы ошибок наблюдения используется ее несмещенная оценка, вычисляемая по формуле:

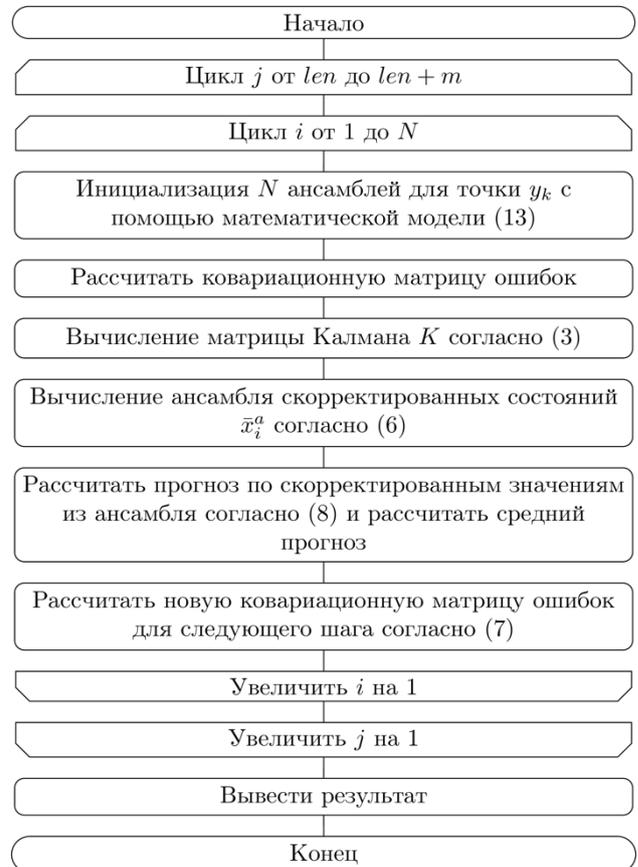


Рис. 2 – Схема алгоритма метода DA с использованием $EnKF$ -фильтра: m – длина прогноза, k – номер отсчета прогноза, len – длина вектора наблюдений, N – размер ансамбля фильтра

$$\mathbf{R}_k = \begin{pmatrix} \overline{u_0 u_0} & \overline{u_0 u_1} & \cdots & \overline{u_0 u_M} \\ \overline{u_1 u_0} & \overline{u_1 u_1} & \cdots & \overline{u_1 u_M} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{u_M u_0} & \overline{u_M u_1} & \cdots & \overline{u_M u_M} \end{pmatrix}.$$

Здесь чертой сверху обозначена процедура усреднения по ансамблю реализаций \bar{u}_i , а вместо произведений $\mathbf{P}_k^f \bar{\mathbf{H}}_k^T$, $\bar{\mathbf{H}}_k \mathbf{P}_k^f \bar{\mathbf{H}}_k^T$ – используются их статистические оценки:

$$\mathbf{P}_k^f \bar{\mathbf{H}}_k^T = \frac{1}{I+1} \sum_i (\bar{x}_i^f - \text{mean}(\bar{x}_{k,i}^f)) (H(\bar{x}_{k,i}^f) - \text{mean}(\bar{y}_{k,i}^f))^T,$$

$$\begin{aligned} \bar{\mathbf{H}}_k \mathbf{P}_k^f \bar{\mathbf{H}}_k^T &= \\ &= \frac{1}{I+1} \sum_i (H(\bar{x}_{k,i}^f) - \text{mean}(\bar{x}_{k,i}^f)) \cdot \\ &\cdot (\bar{x}_i^f - \text{mean}(\bar{x}_{k,i}^f)) (H(\bar{x}_{k,i}^f) - \text{mean}(\bar{y}_{k,i}^f))^T, \end{aligned}$$

и вместо ковариационной матрицы ошибок прогноза – ее оценка, рассчитываемая по формуле:

$$\mathbf{P}_a^f = \frac{1}{I-1} \sum_i (x_{i-m}^f - \bar{y}_k) (x_i^f - \bar{y}_k)^T, m = \overline{0, I}. \quad (15)$$

Результаты сравнительного анализа качества прогнозирования классическим методом *DA* и методом *DA*, использующим *EnKF*-фильтр, представлены в [17].

В. Интеграция формальных методов прогнозирования ВР в метод DA

При решении задачи прогнозирования ВР, порожденных ДС, для которых на данный момент не существует соответствующих математических моделей, с помощью формальных методов прогнозирования, полагают, что

$$x_{k+1} = F(x_k, x_{k-1}, \dots, x_{k-p}, \alpha_1, \dots, \alpha_{p+1}), \quad (16)$$

где x_{k+1} – прогнозируемое значение; F – функция, определяемая выбранным формальным методом прогнозирования ВР; $x_k, x_{k-1}, \dots, x_{k-p}$ – предыдущие значения ВР, $\alpha_1, \dots, \alpha_{p+1}$ – параметры модели.

Сравнивая (16) и уравнение прогноза в (1) видим, что с точностью до аддитивного члена \bar{w}_k (ошибки прогноза), данные уравнения имеют одинаковую структуру, поэтому функция $F(x_k, x_{k-1}, \dots, x_{k-p}, \alpha_1, \dots, \alpha_m)$, по сути, аналогична оператору $M(\bar{x}_k, t_k)$ в (1). Следовательно, использование (16) эквивалентно переходу от описания ДС в пространстве состояний к ее описанию в p -мерном пространстве координатном пространстве. Используя отмеченную аналогию, можно сразу записать систему уравнений, объединяющую метод *DA* и формальные методы прогнозирования ВР в виде:

$$\begin{cases} x_{k+1} = F(\bar{x}_k, \bar{\alpha}_k, t_k) + w_k, \\ y_k = x_k + v_k, \end{cases} \quad (17)$$

где $\bar{x}_k^T = (x_k, x_{k-1}, \dots, x_{k-p})^T$, $\bar{\alpha}_k^T = (\alpha_k, \alpha_{k-1}, \dots, \alpha_{k-(p-1)})$,

F – функция, определяемая выбранным формальным методом прогнозирования ВР; y_k – наблюдаемое значение вектора состояния ДС, здесь, оператор наблюдения, связывающий x_k с наблюдаемым значением y_k , $H \equiv 1$, w_k , v_k – ошибки модели и наблюдения, соответственно. Вектор параметров математической модели (16) при использовании формальной модели в методе *DA* идентифицируется на основе использования $k, k-1, \dots, k-(p+1)$ значений ВР.

Блок-схема алгоритма интеграции формальных методов прогнозирования ВР в метод *DA* (17), основанном на использовании *EnKF*-фильтра Калмана, представлен на рисунке 3.

Из рисунка 3 видно, что данный алгоритм (далее алгоритм интеграции) реализуется выполнением следующей последовательности действий:

1. Выбор формального метода прогнозирования.
2. Задать параметры метода *DA* и формального метода прогноза.

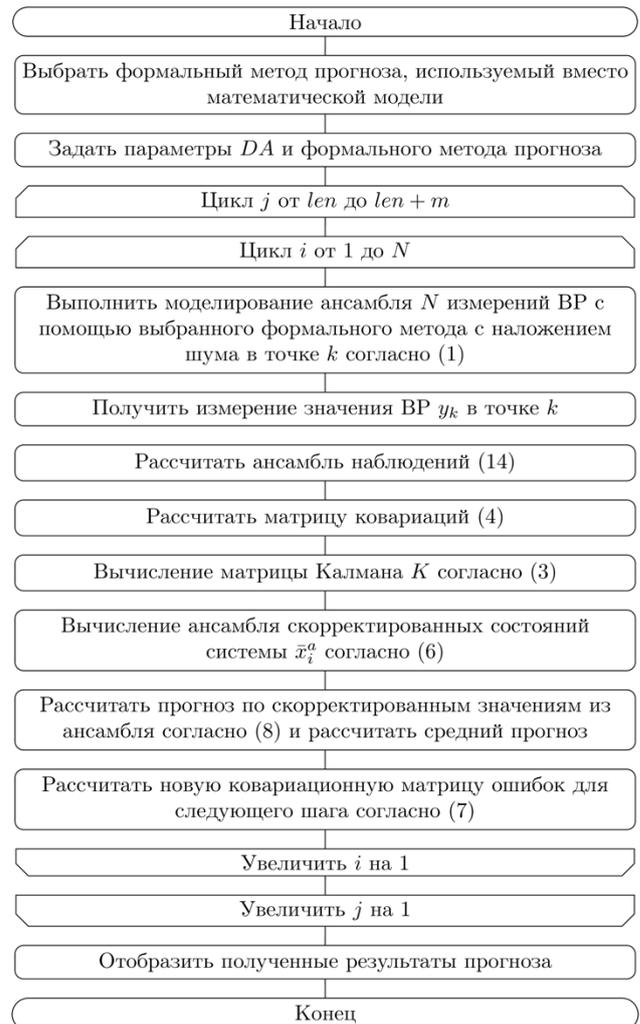


Рис. 3 – Алгоритм интеграции формальных методов прогнозирования ВР в метод *DA*

3. Получить ансамбль прогнозов в точке с помощью формального метода (в рамках работы — это прогноз с использованием AP) и добавления шума.

4. Получить данные наблюдений системы в прогнозируемый момент времени.

5. Рассчитать матрицы ковариации ошибок прогноза.

6. Выполнить коррекцию прогноза путем применения матриц ковариации и передаточного коэффициента фильтра Калмана K для каждой точки ансамбля.

7. Обновить каждый член ансамбля прогноза и матрицы ковариаций в соответствии с полученным значением K .

8. Выполнить пункты 3–7 для каждой прогнозируемой точки.

9. Оценить точность полученного прогноза.

В данном решении параметры используемых моделей задаются один раз в начале и не меняются со временем выполнения прогнозирования.

С. Описание прогнозируемого ВР

Тестирование разработанной методики проводилось на наборе данных «Air Passengers» [18], одного из популярных ВР, используемых для тестирования различных методов прогнозирования ВР [19]. Данный ВР составлен из ежемесячных объемов международных пассажироперевозок авиалиний США в период с января 1949 по декабрь 1960 гг., измеряемых в тысячах человек. Число членов ВР (длина ВР) – $N = 144$. Визуализация ВР в виде зависимости «объема ежемесячных пассажироперевозок от времени» представлена на рисунке 4.

Анализ ВР, представленного на рисунке 4, позволяет высказать предположение о том, что он представляет собой линейную комбинацию отсчетов монотонно возрастающей (тренд) и периодической функции времени. Можно показать [20], что наилучшей аппроксимацией обсуждаемого тренда является линейная функция (см. Рисунок 5).

Разность между обсуждаемым ВР и линейным трендом представлена на рисунке 6.

Из рисунка 6 видно, что в данном ВР присутствует периодическая (с периодом 12 месяцев) и сезонная компоненты, ответственная за увеличение амплитуды

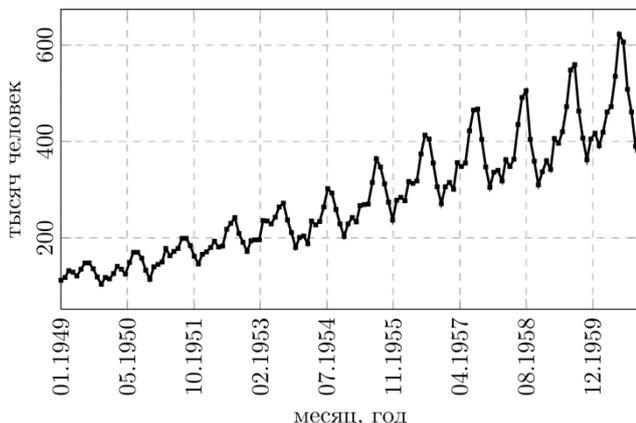


Рис. 6 – Временной ряд «Air Passengers»

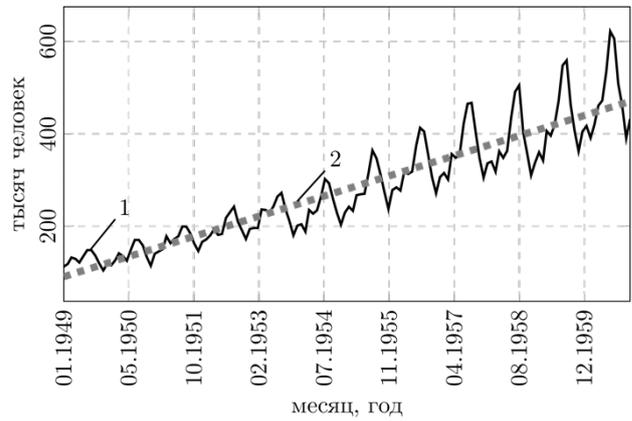


Рис. 4 – Визуализация тренда и ВР «Air Passengers»: 1 – отсчеты ВР, 2 – тренд

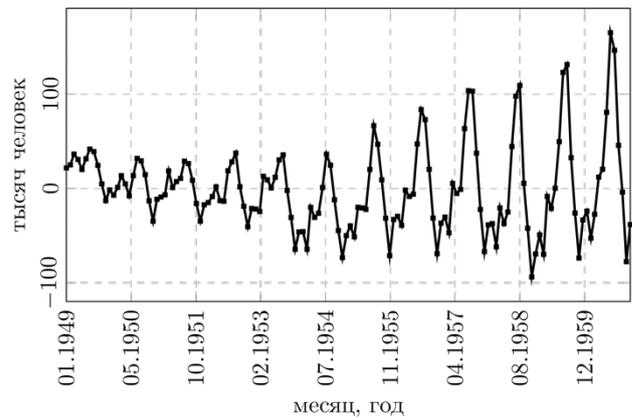


Рис. 5 – Визуализация ВР «Air Passengers» с удаленным линейным трендом

периодической составляющей в период с июня по сентябрь, которые в [20] выделены на основе количественного анализа ВР «Air Passengers». Здесь наличие сезонной компоненты обусловлена тем, что в США большинство людей используют отпуска в летние месяцы. Также отметить, что анализируемый ВР ряд не имеет явных аномалий и выбросов, в связи с чем не требуется его предобработка с целью удаления выбросов.

В [21] обосновано, что данный временной ряд может быть описан одной из известных формальных авторегрессионных моделей, из которых, без потери общности полученных результатов, в нашем исследовании будут использоваться AR-модели вида:

$$x_{k+1} = c + \sum_{i=1}^p \alpha_i x_{k-i} + \varepsilon_k, \quad (18)$$

где c – константа, ε_k – белый шум, $\alpha_1, \alpha_2, \dots, \alpha_p$ – параметры модели (коэффициенты авторегрессии). Напомним, что параметры AR-модели (18) находятся как решения уравнения Юла-Уокера [1], которое при определенных условиях сводится к методу наименьших квадратов.

III. МЕТОДИКА ИССЛЕДОВАНИЯ И АНАЛИЗ РЕЗУЛЬТАТОВ ПРОГНОЗИРОВАНИЯ ВР «AIR PASSENGERS»

Исследование было проведено в соответствие с методикой, реализующейся выполнением следующей

последовательности действий:

1. Разбиение прогнозируемого ВР на известные и прогнозируемые наблюдения, число которых составили $N_{test} = 84$ и $N_f = 60$ отсчетов, соответственно.
2. Выбор порядка AR -модели p .
3. Идентификация параметров AR -модели (18) [1].
4. Прогнозирование 60 значений ВР (в период с января 1956 по декабрь 1960 гг.) с помощью формальной AR -модели (18) и предложенной интегрированной методики.
5. Оценка точности прогноза на основе сравнения спрогнозированных значений ВР и соответствующих реальных известных наблюдений с помощью критерия:

$$RMSE_p = \frac{1}{N_f} \left[\sum_{i=1}^{N_f} (Y_{i+N_{test}} - \hat{Y}_{i+N_{test}})^2 \right]^{1/2}, \quad (19)$$

где N_f – число прогнозируемых наблюдений, \hat{Y} – спрогнозированные значения ВР, Y – известные отсчеты ВР в прогнозируемых точках.

Были вычислены оценки значений $RMSE$ для AR -моделей, порядок которых изменялся от 1 до 10. Для получения статистически значимых результатов размер ансамбля фильтра в $EnKF$ выбирался равным 50. Далее с помощью AR -модели (18) и разработанного алгоритма интеграции были вычислены в соответствии с (17) зависимости дисперсии разностей между точными и спрогнозированными $RMSE_{1,2} = f(p)$, представленные на рисунке 7.

Из рисунка 7 видно, что функция $RMSE_1 = f(p)$ при увеличении порядка AR -модели от 1 до 10 монотонно убывает от 94,4 до 43,2, а $RMSE_2 = f(p)$ оказывается возрастающей от 3,4 до 27,1. Диапазоны изменения зависимостей $RMSE_1 = f(p)$ для других порядков AR -моделей представлены в таблице 2.

Для пояснения данного результата рассмотрим зависимости спрогнозированных значений от времени для случаев $p = 1, 10$, представленные на рисунках 8, 9.

Из рисунков 8, 9 видно, что при увеличении порядка AR -модели улучшается качество прогнозирования тренда, который оказывает определяющее влияние на

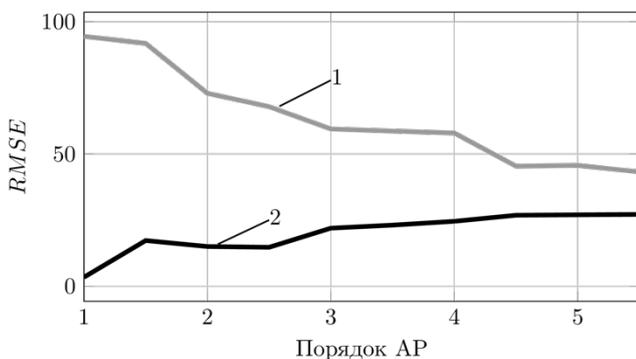


Рис. 7 – Зависимость $RMSE$ от порядка AR -модели, используемой для прогнозирования: 1 – AR -модель, 2 – алгоритм интеграции

Таблица 1. Диапазоны изменения параметра $RMSE$

Порядок AR	RMSE			
	Алгоритм интеграции		AR-модель	
	min	max	min	max
1	2,75	3,381	92,35	94,40
2	15,32	17,24	89,43	91,74
3	13,33	14,98	69,41	72,87
4	13,00	14,69	65,15	67,83
5	18,14	21,92	57,01	59,40
6	22,94	23,07	57,27	58,60
7	20,74	24,51	55,91	57,86
8	24,51	26,80	42,61	45,36
9	24,97	26,96	43,012	45,63
10	26,51	27,11	40,66	43,17

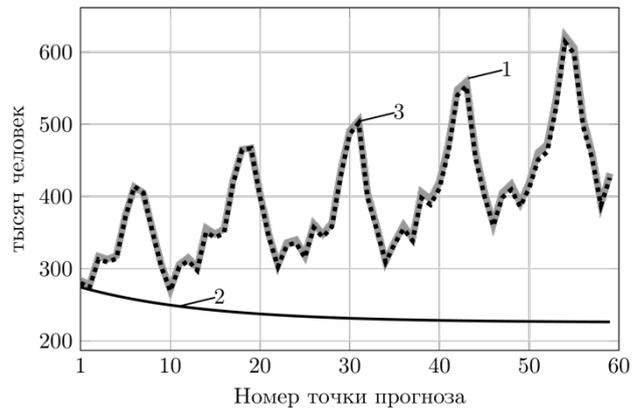


Рис. 8 – Результат прогнозирования ВР «Air Passengers» на основе AR -модели 1-го порядка: 1 – истинные значения ВР (серая линия), 2 – AR -модель, 3 – алгоритм интеграции (пунктирная линия)

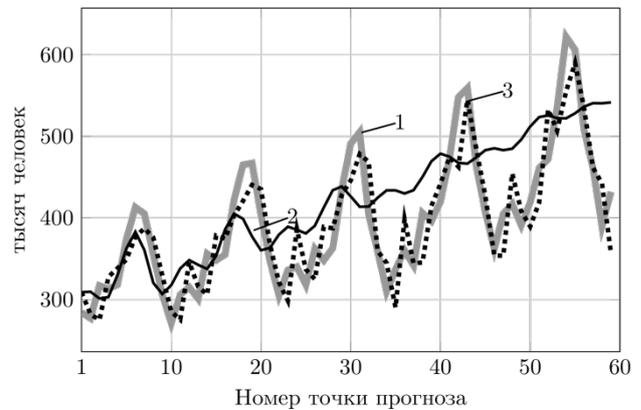


Рис. 9 – Результат прогнозирования ВР «Air Passengers» на основе AR -модели 10-го порядка: 1 – истинные значения ВР, 2 – AR -модель, 3 – алгоритм интеграции

значения анализируемого ВР. Отмеченное свойство AR -модели объясняет монотонное убывание функции $RMSE_1 = f(p)$. Отметим, что AR -модель 10-го порядка пытается отследить наличие периодической и сезонной составляющих анализируемого ВР, однако, точность ее прогнозирования оказывается недостаточно высокой. Здесь можно ожидать, что увеличить точность прогнозирования анализируемого ВР удастся при использовании метода $SARIMA$, разработанного для прогнозирования ВР с сезонной составляющей (*Seasonal ARIMA*) [22].) Таким образом, для повышения точности прогнозирования ВР «Air Passengers» с помощью (17) целесообразно использовать AR -модели высокого порядка. Однако увеличение порядка AR -модели

приводит к увеличению размерности системы уравнений Юла-Уокера и, соответственно, к пропорциональному увеличению числа необходимых известных пред-прогнозных точек, что не всегда возможно на практике. При использовании предложенного алгоритма интеграции, наоборот, наименьшая погрешность прогнозирования оказывается в случае использования *AR*-модели первого порядка. Данный результат, с нашей точки зрения, обусловлен тем, что для модели первого порядка уравнение прогноза и уравнение коррекции оказываются линейны относительно входящих в них переменных.

IV. ЗАКЛЮЧЕНИЕ

Предложен алгоритм интеграции формальных методов прогнозирования ВР в метод *DA*, в котором используется *EnKF*-фильтр Калмана, работоспособность которого подтверждена результатами прогнозирования ВР «*Air Passengers*». Проведен сравнительный анализ точности прогнозов, полученных с помощью формальной модели ВР, и алгоритма интеграции, результаты которого позволяют сделать следующие выводы:

1. На примере одномерного ВР продемонстрирована возможность достоверного прогнозирования его значений с помощью метода *DA*, в котором вместо математической модели, описывающей закон изменения во времени состояния ДС, породившей изучаемый ВР, используется его формальная модель ВР.

2. Точность прогнозирования ВР «*Air Passengers*», оцениваемая дисперсией разностей между точными и спрогнозированными значениями изучаемого ВР, с помощью оригинальной методики, основанной на интеграции формальных методов прогнозирования ВР и метода *DA*, оказалась выше аналогичной величины в случае прогнозирования на основе *AR*-модели ВР любого порядка.

Подтверждение работоспособности предложенной методики прогнозирования состояния ДС в пространстве состояний больших размерностей является предметом последующих публикаций.

БИБЛИОГРАФИЯ

- [1] G. E. Box, G. M. Jenkins, G. C. Reinsel, и G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [2] L. Nerger и др., «SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications EU FP7 SPACE-2011-1 project 283580».
- [3] M. Kanamitsu, «Description of the NMC global data assimilation and forecast system», *Weather and forecasting*, т. 4, вып. 3, сс. 335–342, 1989.
- [4] А. В. Костров и А. М. Мардук, «Об одной математической задаче теории движения ракет», *Журнал вычислительной математики и математической физики*, т. 34, вып. 10, сс. 1427–1443, 1994.
- [5] L. Kilian и M. P. Taylor, «Why is it so difficult to beat the random walk forecast of exchange rates?», *Journal of International Economics*, т. 60, вып. 1, сс. 85–107, 2003.
- [6] G. Welch и G. Bishop, «An introduction to the Kalman filter», 1995.
- [7] G. P. Zhang, «Time series forecasting using a hybrid ARIMA and neural network model», *Neurocomputing*, т. 50, сс. 159–175, 2003.
- [8] N. Golyandina, V. Nekrutkin, и А. А. Zhigljavsky, *Analysis of time series structure: SSA and related techniques*. CRC press, 2001.
- [9] А. Г. Ивахненко, *Долгосрочное прогнозирование и управление сложными системами*. Техніка, 1975.
- [10] Y. Timoshenkova, S. Porshnev и N. Safiullin, «On the possibility of correction of the forecasting of the Lorenz attractor dynamic characteristics using experimental data and data assimilation», в *Journal of Physics: Conference Series*, 2018, т. 1053, с. 012004.

- [11] Y. Timoshenkova, S. Porshnev, N. Safiullin и O. Ponomareva, «On the Possibility of the Forecast Correction Lorenz System Use Particle Filter», в *2018 International Conference on Applied Mathematics and Computational Science, ICAMCS. NET 2018*, 2018, сс. 27–30.
- [12] P. L. Houtekamer и H. L. Mitchell, «Data assimilation using an ensemble Kalman filter technique», *Monthly Weather Review*, т. 126, вып. 3, сс. 796–811, 1998.
- [13] M. Ghil и P. Malanotte-Rizzoli, «Data assimilation in meteorology and oceanography», *Advances in geophysics*, т. 33, сс. 141–266, 1991.
- [14] B. K. W. Lahoz и R. Menard, *Data assimilation*. Springer, 2010.
- [15] E. N. Lorenz, «Deterministic nonperiodic flow», *Journal of atmospheric sciences*, т. 20, вып. 2, сс. 130–141, 1963.
- [16] М. Табор, *Хаос и интегрируемость в нелинейной динамике*. 2001.
- [17] M. Roth, G. Hendeby, C. Fritsche, и F. Gustafsson, «The Ensemble Kalman filter: a signal processing perspective», *EURASIP Journal on Advances in Signal Processing*, т. 2017, вып. 1, сс. 1–16, 2017.
- [18] «Air Passengers». <https://kaggle.com/abhisheknamidi/air-passengers> (просмотрено дек. 01, 2021).
- [19] M. Bocquet, «Introduction to the principles and methods of data assimilation in geosciences», *Notes de cours, École des Ponts ParisTech*, 2014.
- [20] *Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова, А.А.Жигляевского*. 1997. Просмотрено: дек. 20, 2021. [Онлайн]. Доступно на: https://rusneb.ru/catalog/001980_000024_RU_FESSL_MAIN_1331818796268482813/
- [21] А. Н. Тырсин, «Построение моделей авторегрессии временных рядов при наличии помех», *Математическое моделирование*, т. 17, вып. 5, сс. 10–16, 2005.
- [22] A. E. Permanasari, I. Hidayah, и I. A. Bustoni, «SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence», в *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2013, сс. 203–207.

Technique for Formal Methods of Time Series Forecasting Integration in Data Assimilation

Y.S. Timoshenkova, S.V. Porshnev, N.T. Safiullin

Annotation—The article describes the method developed by the authors for integration of formal methods of time series (TS) forecasting (autoregressive integrated moving average (ARIMA), singular spectrum analysis, group method of data handling, artificial recurrent neural network with long short-term memory) into the Data Assimilation (DA), in cases where mathematical model of the dynamic system generating the TS is not known (for example TS consisting of economic indicators). The performance of the proposed integration method is illustrated on the example of forecasting TS named "Air Passengers" using the DA method based on the ensemble Kalman filter with integrated ARIMA method. Estimations of the accuracy of "Air Passengers" TS forecasts is calculated using ARIMA and the proposed integration method. Article discusses the advantages and disadvantages of the proposed method for integrating formal TS forecasting methods into the DA and also directions for its further improvement.

Key Words—Time series analysis, time series forecast, Data Assimilation.

REFERENCES

- [1] G. E. Box, G. M. Jenkins, G. C. Reinsel, i G. M. Ljung, Time series analysis: forecasting and control. John Wiley & Sons, 2015.
- [2] L. Neger i dr., «SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications EU FP7 SPACE-2011-1 project 283580».
- [3] M. Kanamitsu, «Description of the NMC global data assimilation and forecast system», Weather and forecasting, t. 4, vyp. 3, ss. 335–342, 1989.
- [4] A. V. Kostrov i A. M. Mardjuk, «Ob odnoj matematicheskoj zadache teorii dvizhenija raket», Zhurnal vychislitel'noj matematiki i matematicheskoj fiziki, t. 34, vyp. 10, ss. 1427–1443, 1994.
- [5] L. Kilian i M. P. Taylor, «Why is it so difficult to beat the random walk forecast of exchange rates?», Journal of International Economics, t. 60, vyp. 1, ss. 85–107, 2003.
- [6] G. Welch i G. Bishop, «An introduction to the Kalman filter», 1995.
- [7] G. P. Zhang, «Time series forecasting using a hybrid ARIMA and neural network model», Neurocomputing, t. 50, ss. 159–175, 2003.
- [8] N. Golyandina, V. Nekrutkin, i A. A. Zhigljavsky, Analysis of time series structure: SSA and related techniques. CRC press, 2001.
- [9] A. G. Ivahnenko, Dolgosrochnoe prognozirovanie i upravlenie slozhnymi sistemami. Tehnika, 1975.
- [10] Y. Timoshenkova, S. Porshnev i N. Safiullin, «On the possibility of correction of the forecasting of the Lorenz attractor dynamic characteristics using experimental data and data assimilation», v Journal of Physics: Conference Series, 2018, t. 1053, s. 012004.
- [11] Y. Timoshenkova, S. Porshnev, N. Safiullin i O. Ponomareva, «On the Possibility of the Forecast Correction Lorenz Systeme Use Particle Filter», v 2018 International Conference on Applied Mathematics and Computational Science, ICAMCS. NET 2018, 2018, ss. 27–30.
- [12] P. L. Houtekamer i H. L. Mitchell, «Data assimilation using an ensemble Kalman filter technique», Monthly Weather Review, t. 126, vyp. 3, ss. 796–811, 1998.
- [13] M. Ghil i P. Malanotte-Rizzoli, «Data assimilation in meteorology and oceanography», Advances in geophysics, t. 33, ss. 141–266, 1991.
- [14] B. K. W. Lahoz i R. Menard, Data assimilation. Springer, 2010.
- [15] E. N. Lorenz, «Deterministic nonperiodic flow», Journal of atmospheric sciences, t. 20, vyp. 2, ss. 130–141, 1963.
- [16] M. Tabor, Haos i integriruemost' v nelinejnoj dinamike. 2001.
- [17] M. Roth, G. Hendeby, C. Fritsche, i F. Gustafsson, «The Ensemble Kalman filter: a signal processing perspective», EURASIP Journal on Advances in Signal Processing, t. 2017, vyp. 1, ss. 1–16, 2017.
- [18] «Air Passengers». <https://kaggle.com/abhishekmamidi/air-passengers> (prosmotreno dek. 01, 2021).
- [19] M. Bocquet, «Introduction to the principles and methods of data assimilation in geosciences», Notes de cours, École des Ponts ParisTech, 2014.
- [20] Glavnye komponenty vremennyh rjadov: metod «Gusenica» / Pod red. D.L.Danilova, A.A.Zhigljavskogo. 1997. Prosmotreno: dek. 20, 2021. [Onlajn]. Dostupno na: https://rusneb.ru/catalog/001980_000024_RU_FESSL_MAIN_1331818796268482813/
- [21] A. N. Tyrsin, «Postroenie modelej avtoregressii vremennyh rjadov pri nalichii pomeh», Matematicheskoe modelirovanie, t. 17, vyp. 5, ss. 10–16, 2005.
- [22] A. E. Permanasari, I. Hidayah, i I. A. Bustoni, «SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence», v 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), 2013, ss. 203–207.

Yulia Timoshenkova, Ural Federal University, Yekaterinburg, Russia, <https://orcid.org/0000-0002-4137-0484> (e-mail: julia.timoshenkova@urfu.ru).

Sergey Porshnev, Ural Federal University, Yekaterinburg, Russia, <https://orcid.org/0000-0001-6884-9033> (e-mail: s.v.porshnev@urfu.ru).

Nikolai Safiullin, Ural Federal University, Yekaterinburg, Russia, <https://orcid.org/0000-0002-0004-5634> (e-mail: n.t.safiullin@urfu.ru).

The reported study was funded by RFBR according to the research project No 20-37-90006