СИСТЕМЫ АНАЛИЗА И ОБРАБОТКИ ДАННЫХ

ISSN 2782-2001

№ 2 (82) **2021**

УДК 681.51(06)
С 409

Учредитель

ФГБОУ ВО «Новосибирский государственный технический университет»

Редакционная коллегия

Главный редактор

Лемешко Борис Юрьевич, д-р техн. наук, проф., НГТУ, г. Новосибирск

Заместители главного редактора

Вострецов Алексей Геннадьевич, д-р техн. наук, проф., НГТУ, г. Новосибирск
Шокин Юрий Иванович, академик, д-р физ.-мат. наук, проф., Институт вычислительных технологий СО РАН, г. Новосибирск

Ответственный секретарь

Воевода Александр Александрович, д-р техн. наук, проф., НГТУ, г. Новосибирск

Члены редколлегии

Володин Андрей Игоревич, канд. физ.-мат. наук, доктор философии по статистике Университета Регины, проф., Канада
Воронин Владимир Викторович, д-р техн. наук, доцент, ТОГУ, г. Хабаровск
Воропай Николай Иванович, чл.-корр. РАН, д-р техн. наук, проф., Институт систем энергетики им. Мелентьева СО РАН, г. Иркутск
Воскобойников Юрий Евгеньевич, д-р физ.-мат. наук, проф., Сибстрин, г. Новосибирск
Гайдук Анатолий Романович, д-р техн. наук, проф., ЮФУ, г. Таганрог
Генералов Владимир Михайлович, д-р техн. наук, ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора, г. Новосибирск
Григорьев Юрий Дмитриевич, д-р техн. наук, проф., СПбГЭТУ «ЛЭТИ», г. Санкт-Петербург
Гриф Михаил Геннадьевич, д-р техн. наук, проф., НГТУ, г. Новосибирск
Дмитриев Александр Капитонович, д-р физ.-мат. наук, НГТУ, г. Новосибирск
Комаров Андрей Константинович, д-р физ.-мат. наук, проф., Институт автоматики и электрометрии СО РАН, г. Новосибирск
Крюков Андрей Васильевич, д-р техн. наук, проф., ИрГУПС, г. Иркутск
Лисицина Лилия Ивановна, д-р техн. наук, проф., НГТУ, г. Новосибирск
Наумова Елена Николаевна, д-р техн. наук, проф., Университет Тафтса, США
Носек Ярослав, д-р техн. наук, проф., Технический университет Либерец, Чехия
Ронжин Андрей Леонидович, д-р техн. наук, проф., СПИИРАН, г. Санкт-Петербург
Рубан Анатолий Иванович, д-р техн. наук, проф., СФУ, г. Красноярск
Сидоров Денис Николаевич, д-р физ.-мат. наук, проф., Институт систем энергетики им. Мелентьева СО РАН, г. Иркутск
Уранчимэг Тудевдагва, д-р техн. наук, проф., Хемницкий технический университет, Германия
Фомин Василий Михайлович, д-р физ.-мат. наук, проф., Институт теоретической и прикладной механики СО РАН, г. Новосибирск
Хабаров Валерий Иванович, д-р техн. наук, проф., СГУПС, г. Новосибирск
Ханова Анна Алексеевна, д-р техн. наук, проф., АГТУ, г. Астрахань
Хлебников Михаил Владимирович, д-р физ.-мат. наук, ИПУ им. В.А. Трапезникова РАН, г. Москва
Шандаров Станислав Михайлович, д-р физ.-мат. наук, проф., ТУСУР, г. Томск

Адрес редакции и издателя: 630073, г. Новосибирск, пр. К. Маркса, 20, НГТУ
Тел. (383) 346-56-88, тел., факс (383) 344-49-98
E-mail: nauchvest@corp.nstu.ru
Web Site: http://journals.nstu.ru/vestnik/

16+

Индекс: 82962 («РОСПЕЧАТЬ») **УДК 681.51(06)**

© Коллектив авторов, 2021
© Новосибирский государственный технический университет, 2021

СИСТЕМЫ АНАЛИЗА И ОБРАБОТКИ ДАННЫХ

ISSN 2782-2001

№ 2 (82) **2021**

СОДЕРЖАНИЕ

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

Абденова Г.А., Базикова К.М., Кенжегалым Ж.М. Параболическая модель в форме пространства состояний динамики накоплений 7

Булатов Ю.Н., Крюков А.В. Исследование кибербезопасности прогностических алгоритмов управления для установок распределенной генерации 19

Вагин Д.В. Структура и особенности программного комплекса для геофизических геометрических 3D-инверсий 35

Лемешко Б.Ю., Лемешко С.Б. Проблемы применения непараметрических критериев согласия в задачах обработки результатов измерений 47

Сивак М.А., Тимофеев В.С. Построение робастных нейронных сетей с различными функциями потерь 67

Усманов З.Д., Косимов А.А. Тестирование гамма-классификатора, настроенного на распознавание языков произведений на основе латинского алфавита 83

Якубовский Д.В., Крупенёв Д.С., Бояркин Д.А. Модель минимизации дефицита мощности электроэнергетических систем с учетом ограничений по контролируемым сечениям 95

ПРИБОРОСТРОЕНИЕ, МЕТРОЛОГИЯ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ ПРИБОРЫ И СИСТЕМЫ

Бабичев М.М. Реализация метода адаптивного квантования в измерительных генераторах с цифровым управлением 121

Боровикова Д.В., Гришин О.В., Ненько А.В., Юпашевский А.В., Казьмина А.С., Марков А.В., Мецлер К.А. Разработка аппаратно-программного комплекса анализа и коррекции речи 135

Правила для авторов 146

Корректор *Л.Н. Киншт*
Компьютерная верстка *С.И. Ткачева*

Новосибирский государственный технический университет
630073, г. Новосибирск, пр. К. Маркса, 20, НГТУ

Налоговая льгота – Общероссийский классификатор продукции
Издание соответствует коду 95 2000 ОК 005-93 (ОКП)

Журнал зарегистрирован Комитетом РФ по печати. Свидетельство о регистрации
средства массовой информации ПИ № ФС77–76010 от 03 июля 2019 г.
Периодичность издания – 4 номера в год.
Подписано в печать 11.06.2021. Выход в свет 18.06.2021.
Формат 70 × 108 1/16. Бумага офсетная. Тираж 300 экземпляров.
Уч.-изд. л. 12,95. Печатных листов 9,25. Издательский № 88. Заказ № 601
Цена договорная

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20

ANALYSIS AND DATA PROCESSING SYSTEMS

ISSN 2782-2001

№ 2 (82) **2021**

УДК 681.51(06)
С 409

Journal was established by

Novosibirsk State Technical University

Editorial board

Chief editor

Lemeshko B. Yu., D.Sc. (Eng.), Novosibirsk State Technical University, Novosibirsk, RF

Deputy chief editor

Vostretsov A. G., D.Sc. (Eng.), Novosibirsk State Technical University, Novosibirsk, RF
Shokin Yu. I., D.Sc. (Phys.& Math.), Institute of Computational Technologies, Novosibirsk, RF

Executive Secretary

Voevoda A. A., D.Sc. (Eng.), Novosibirsk State Technical University, Novosibirsk, RF

The members of the editorial Board

Volodin A. I., D.Sc. (Phys. & Math.), Doctor of Philosophy in Statistics from the University of Regina, Professor of Statistics University of Regina, Canada
Voronin V. V., D.Sc. (Eng.), Pacific national University, Khabarovsk, RF
Voropay N. I., Corresponding Member, Russian Academy of Sciences, D.Sc. (Eng.), Melentiev Energy Systems Institute, Irkutsk, RF
Voskoboinikov Yu. E., D.Sc. (Phys. & Math.), Novosibirsk State University of Architecture and Civil Engineering, Novosibirsk, RF
Gaiduk A. R., D.Sc. (Eng.), Southern Federal University, Taganrog, RF
Generalov V. M., D.Sc. (Eng.), FBUN GNTS VB «Vektor» Rospotrebnadzora, Novosibirsk, RF
Grigor'ev Yu. D., D.Sc. (Eng.), Saint-Petersburg Electrotechnical University "LETI", Saint-Petersburg, RF
Grif M. G., D.Sc. (Eng.), Novosibirsk State Technical University, Novosibirsk, RF
Dmitriev A. K., D.Sc. (Phys. & Math.), Novosibirsk State Technical University, Novosibirsk, RF
Komarov A. K., D.Sc. (Phys. & Math.), Institute of Automation and Electrometry SB RAS, Novosibirsk, RF
Kryukov A. V., D.Sc. (Eng.), Irkutsk State Transport University, Irkutsk, RF
Lisitcina L. I., D.Sc. (Eng.), Novosibirsk State Technical University, Novosibirsk, RF
Naumova E. N., PhD. (Math.), Tufts University, USA
Nosek J., D.Sc. (Eng.), Technical University of Liberec, Czech Republic
Ronzhin A. L., D.Sc. (Eng.), St. Petersburg Institute for Informatics and Automation of the Russian Academy of sciences, Saint-Petersburg, RF
Ruban A. I., D.Sc. (Eng.), Siberian Federal University, Krasnoyarsk, RF
Sidorov D. N., D.Sc. (Phys. & Math.), Melentiev Energy Systems Institute, Irkutsk, RF
Uranchimeg T., D.Sc. (Eng.), Technische Universität Chemnitz, Germany
Fomin V. M., D.Sc. (Phys. & Math.), Khristianovich Institut of theoretical and applied mechanics SB RAS, Novosibirsk, RF
Khabarov V. I., D.Sc. (Eng.), Siberian Transport University, Novosibirsk, RF
Khanova A. A., D.Sc. (Eng.), Astrakhan State Technical University, Astrakhan, RF
Khlebnikov M. V., D.Sc. (Phys. & Math.), Institute of Control Sciences, Moscow, RF
Shandarov S. M., D.Sc. (Phys. & Math.), Tomsk State University of Control Systems and Radioelectronics, Tomsk, RF

Editorial Adress: 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation
Tel. (383) 346-56-88, Fax (383) 344-49-98
E-mail: nauchvest@corp.nstu.ru
Web Site: http://journals.nstu.ru/vestnik/

УДК 681.51(06)

© Authors, 2021
© Novosibirsk State Technical University, 2021

ANALYSIS AND DATA PROCESSING SYSTEMS

ISSN 2782-2001

№ 2 (82) **2021**

CONTENTS

INFORMATICS, COMPUTER ENGINEERING AND CONTROL

Abdenova G.A., Bazikova K.M., Kenzhegalym Zh.M. A parabolic model in the form of space states of the dynamics of savings 7

Bulatov Yu.N., Kryukov A.V. Study of cyber security of predictive control algorithms for distributed generation plants 19

Vagin D.V. The structure and features of the software for geophysical geometrical 3D inversions 35

Lemeshko B.Yu., Lemeshko S.B. Problems of nonparametric goodness-of-fit test application in tasks of measurement results processing 47

Sivak M.A., Timofeev V.S. Building robust neural networks using different loss functions 67

Usmanov Z.D., Kosimov A.A. Testing the γ-classifier adapted to recognize the languages of works based on the Latin alphabet 83

Iakubovskii D.V., Krupenev D.S., Boyarkin D.A. A minimization model of the power shortage of electric power systems with regard to the restrictions on controlled sections 95

INSTRUMENT MAKING, METROLOGY AND INFORMATION MEASUREMENT DEVICES AND SYSTEMS

Babichev M.M. Implementation of the adaptive quantization method in digitally controlled measuring generators 121

Borovikova D.V., Grishin O.V., Nenko A.V., Yupashevsky A.V., Kazmina A.S., Markov A.V., Metsler K.A. Development of a hardware and software complex for speech analysis and correction 135

Rules for authors 146

Editor *L.N. Kinsht*
Computer imposition *S.I. Tkacheva*

Novosibirsk State Technical University
20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation, NSTU

The journal is registered by Russian Federation Committee on the press.
The certificate on registration of mass media of PI No. ФС77-73010 of July 03, 2019.
Publication frequency – 4 numbers a year.
Signed in print June 11, 2021.
Date of publication June 18, 2021. Format 70 × 108 1/16
Offset Paper. Circulation is 300 copies. Educational-ed. liter. 12,95. Printed pages 9,25.
Publishing number 88. Order number 601

It is printed in printing house of Novosibirsk State Technical University
20 K. Marx Prospekt, Novosibirsk, 630073

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 7–18

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 7–18

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 517.95: 519.24 DOI: 10.17212/2782-2001-2021-2-7-18

Параболическая модель в форме пространства состояний динамики накоплений*

Г.А. АБДЕНОВА[a], К.М. БАЗИКОВА[b], Ж.М. КЕНЖЕҒАЛЫМ[c]

010008, Республика Казахстан, г. Нур-Султан, ул. Сатпаева, 2, Евразийский национальный университет имени Л.Н. Гумилева

[a] *gauhar.phd@gmail.com* [b] *kmbazikova@mail.ru* [c] *zhangul_flower@mail.ru*

Важное место в теории уравнений с частными производными и ее приложениях занимает уравнение теплопроводности, представитель класса так называемых параболических уравнений. Известно, что для проверки корректности математической модели, основанной на параболическом уравнении, очень важно существование ее решения, так как математическая модель не всегда адекватна конкретному явлению, и из существования решения реальной прикладной задачи не следует существование решения соответствующей математической задачи. Поэтому методы решения дифференциальных уравнений в частных производных – как аналитических, так и численных – всегда актуальны.

В настоящее время вычислительный эксперимент стал мощным средством теоретических исследований. Он проводится над математической моделью исследуемого объекта, но при этом по одним параметрам модели вычисляются другие параметры и делаются выводы о свойствах изучаемого объекта или явления. В работе рассматривается задача пассивной параметрической идентификации систем с распределенными параметрами для динамики аккумуляции ресурсов множества домохозяйств, а также преобразование ее в модель в форме пространства состояний с учетом белых шумов модели динамики исследуемого объекта и измерительной системы линейного типа. Используя метод конечных разностей, решение уравнений с частными производными параболического типа свели к решению системы линейных конечно-разностных и алгебраических уравнений. Для более точного оценивания поведения объекта было предложено использование алгоритма фильтрации по схеме Калмана.

Осуществлены расчеты с помощью математической системы Matlab на основе данных наблюдений за пять лет, взятых с сайта «Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан». Оценивание коэффициентов уравнений аккумуляции ресурсов домохозяйств в форме пространства состояний с использованием данной методики в достаточной степени универсально и может быть применено и в других областях науки и техники.

* *Статья получена 18 января 2021 г.*

Исследование выполнено в Евразийском национальном университете имени Л.Н. Гумилева Республики Казахстан.

Ключевые слова: дифференциальное уравнение с частными производными, стохастическая динамическая система, входные и выходные данные наблюдений, пассивная идентификация, аддитивный белый шум, линейное конечно-разностное уравнение, уравнение параболического типа, модель в пространстве состояний, фильтр Калмана

ВВЕДЕНИЕ

В настоящее время построение динамических моделей для трудно формализуемых областей, использующих дифференциальные уравнения с частными производными [1] и учитывающих случайные факторы и аддитивные шумы, вызывает особый интерес. Так, для моделирования аккумулирования ресурсов исследуемого объекта будем использовать стохастическую параболическую распределенную модель, описывающую динамику накоплений домохозяйств, в виде уравнений с частными производными, а модель измерительной системы – в виде линейных стохастических алгебраических моделей с аддитивными белыми шумами.

Анализ современных научных исследований показывает эффективность решения задач идентификации с использованием моделей пространства состояний (ПС) [2–4]:

$$U(t+1) = \Phi U(t) + B\nu(t), \quad U(1) = \overline{U}_1, \tag{1}$$

$$Y(t+1) = HU(t+1) + w(t+1), \quad t = 1,\ 2, \ldots, \tag{2}$$

где t – время; $U(t)$ – n-вектор состояния; Φ – переходная матрица состояния, размерность $n \times n$; $\nu(t)$ – r-вектор управления; B – переходная матрица управления, размерность $n \times r$; $U(1)$ – n-вектор начального состояния; $Y(t)$ – m-вектор выходных данных; H – матрица наблюдения, размерность $m \times n$; $w(t)$ – n-вектор белого шума измерительной системы, имеющий нулевое математическое ожидание и неизвестную ковариационную матрицу R.

1. ПОСТАНОВКА ЗАДАЧИ

Известно, что исследуемые объекты в каждый момент времени распределены неравномерно. В работах [5, 6] на основе применения принципа сплошных сред и введения функции, которая опишет распределение домохозяйств по аккумуляции ресурсов, получено параболическое уравнение с частными производными вида (3), которому удовлетворяет плотность исследуемых объектов на пространстве u:

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}((c+F)u) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(bu) + df, \tag{3}$$

с начальными условиями

$$x(t)|_{t=0} = \overline{x}_0, \tag{4}$$

где $x(t)|_{t=0}$ – гауссовская величина с математическим ожиданием $\bar{x}_0$ и дисперсией P_0.

В уравнении (3) $u(x,\ t)$ – искомые значения состояний эволюции аккумуляции ресурсов домохозяйств распределенного типа; $F(x,\ t)$ – заданная функция распределенного типа; $c,\ b,\ d$ – постоянные коэффициенты; $f(x,\ t)$ – число домохозяйств, которые попадут из других множеств домохозяйств на отрезок единичной длины пространства накоплений за единичный интервал времени в окрестности x и t. Уравнение (3) называется дифференциальным уравнением параболического типа, описывающим накопления множества домохозяйств. С учетом аддитивных шумов процесса аккумуляции ресурсов уравнение (3) можно записать в виде

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}((c+F)u) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(bu) + df + p_1 w, \tag{5}$$

где $p_1 = Q(x_k,\ t_s)$ – шумовой коэффициент в уравнении состояния; w – аддитивные белые шумы измерителей.

Предположим, что за исследуемыми объектами можно вести наблюдение, поэтому уравнение можно записать в виде

$$\begin{aligned} &y(x_k,\ t_s) = hu(x_k,\ t_s) + p_2\varepsilon(x_k,\ t_s), \\ &k = \overline{1,n},\quad s = \overline{1,m}, \end{aligned} \tag{6}$$

где $u(x,\ t)$ – пространственно-временная функция состояния, измеряемая в дискретных точках x_k и t_s: $\left\{u(x,\ t) \approx u(x_k,\ t_s) = u_{k,s},\quad k = \overline{1,n},\quad s = \overline{1,m}\right\}$; $y(x_k,\ t_s)$ – выходные наблюдения; h – весовой коэффициент; $\left\{\varepsilon(x_k,\ t_s) = \varepsilon_{k,s},\quad k = \overline{1,n},\quad s = \overline{1,m}\right\}$ – белый шум системы с неизвестной ковариацией $p_2 = R(x_k,\ t_s)$.

Учитывая указанные условия на основе дискретного выхода наблюдаемой системы $y(x_k,\ t_s)$ распределенного типа и начальное условие (4), можно сформулировать задачу оценивания параметров $c, b, d,\ p_1, p_2$.

2. АЛГОРИТМ ОЦЕНИВАНИЯ ПАРАМЕТРОВ В ПАРАБОЛИЧЕСКОМ УРАВНЕНИИ АККУМУЛЯЦИИ РЕСУРСОВ

Для реализации решения задачи пассивной идентификации используем дискретную форму выражения (5) в частных производных [7, 8]:

$$\frac{\partial u}{\partial t} = -a_1\frac{\partial}{\partial x}u + b_1\frac{\partial^2}{\partial x^2}u + d_1 f + p_1 w, \tag{7}$$

где $a_1 = c + F$, $b_1 = b/2$, $d_1 = d$, p_1 – коэффициенты; t – время, $t \geq 0$; x – координата, $0 \leq x \leq L$; $w(x,\ t)$ – белый шум с неизвестной ковариацией Q и нулевым математическим ожиданием; $u(x,\ t)$ – состояние эволюции аккумуляции ресурсов домохозяйств распределенного типа с граничным и начальным условиями:

$$\left.\frac{\partial u(x,\ t)}{\partial x}\right|_{x=0} = \tilde{u}_0, \quad \text{или} \quad u(x,\ t)|_{x=0} = \tilde{u}_0, \tag{8а}$$

$$u(x,\ t)|_{x=L} = \tilde{u}_L, \tag{8б}$$

$$u(x,\ t)|_{t=0} = u(x,\ 0) = u_0(x). \tag{9}$$

Для t и x интервалы квантования приравняем: $\Delta t = 1$ и $\Delta x = 1$, получим следующее выражение:

$$\frac{1}{\Delta t}(u_{k,s+1} - u_{k,s}) = \frac{b_1}{\Delta x^2}(u_{k,s+1} - 2u_{k,s} + u_{k,s-1}) -$$

$$-\frac{a_1}{\Delta x}(u_{k,s+1} - u_{k,s}) + d_1 f_{k,s} + p_1 w_{k,s},$$

где $k = \overline{1, n}$, $s = \overline{1, m}$.

Значения функций $u(x,\ t)$ объединим по координатам k, s узлов сетки при интервалах $\Delta t = 1$ и $\Delta x = 1$.

Получим систему уравнений при $k = 1,\ 2, \ldots,\ n$:

$$\begin{cases} u_{1,s+1} = c_1 u_{0,s} - c_2 u_{1,s} + c_3 u_{2,s} + u_{1,s} + d_1 f_{1,s} + p_1 w_{1,s}, \\ u_{2,s+1} = c_1 u_{1,s} - c_2 u_{2,s} + c_3 u_{3,s} + u_{2,s} + d_1 f_{2,s} + p_1 w_{2,s}, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ u_{n,s+1} = c_1 u_{n,s} - c_2 u_{n,s} + c_3 u_{n+1,s} + u_{1,s} + d_1 f_{n,s} + p_1 w_{n,s}. \end{cases} \tag{10}$$

Сделаем следующие обозначения:

$$U_{s+1} \overset{\Delta}{=} \begin{pmatrix} u_{1,s+1} \\ u_{2,s+1} \\ \vdots \\ u_{n,s+1} \end{pmatrix}_{n\times 1}; \quad U_s \overset{\Delta}{=} \begin{pmatrix} u_{1,s} \\ u_{2,s} \\ \vdots \\ u_{n,s} \end{pmatrix}_{n\times 1}; \quad F_s \overset{\Delta}{=} \begin{pmatrix} f_{1,s} \\ f_{2,s} \\ \vdots \\ f_{n,s} \end{pmatrix}_{n\times 1};$$

$$W_s \overset{\Delta}{=} \begin{pmatrix} w_{1,s} \\ w_{2,s} \\ \vdots \\ w_{n,s} \end{pmatrix}_{n\times 1} ; \quad E_2 \overset{\Delta}{=} \begin{pmatrix} c_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{n\times 1} ; \quad B_2 \overset{\Delta}{=} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ c_3 \end{pmatrix}_{n\times 1} . \tag{11}$$

Граничные условия примут вид

$$\theta_s = u_{0,s} = 0 ; \quad \eta_s = u_{n+1,s} . \tag{12}$$

Уравнение (7) в векторно-матричной форме примет вид

$$M \overset{\Delta}{=} \begin{pmatrix} -c_2 & c_3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ c_1 & -c_2 & c_3 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & c_1 & -c_2 & c_3 \\ 0 & 0 & 0 & 0 & \cdots & 0 & c_1 & -c_2 \end{pmatrix}_{n\times n} ; \tag{13}$$

$$\overline{G}^{\Delta} = d_1 E; \quad \Omega^{\Delta} = p_1 E, \tag{14}$$

где E – единичная матрица размерностью $n \times n$. Используя (11)–(14), выражение (10) представим в векторно-матричной форме:

$$U_{s+1} = MU_s + \overline{G}U_s + NF + \Omega W_s + \left(E_2\theta_s + B_2\eta_s\right) . \tag{15}$$

В [9, 10] приведены обозначения, которые позволяют представить соотношение (15) в виде модели в ПС.

В итоге уравнение (10) запишется в виде

$$U_{s+1} = MU_s + GH_s + NF + \Omega W_s . \tag{16}$$

Соотношение (16) по виду соответствует выражению (1), где время может иметь значения: $s = \{1,\ 2, \ldots,\ m\}$, а начальные условия можно записать следующим образом:

$$U_0 = \begin{pmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{n,1} \end{pmatrix} = \begin{pmatrix} u_{1,0} \\ u_{2,0} \\ \vdots \\ u_{n,0} \end{pmatrix}_{n\times 1} . \tag{17}$$

Далее, при условии h = const, введем обозначения для уравнения (6):

$$Y_s = \begin{pmatrix} y_{1,s} \\ y_{2,s} \\ \vdots \\ y_{n,s} \end{pmatrix}_{n\times 1}; \quad U_s = \begin{pmatrix} u_{1,s} \\ u_{2,s} \\ \vdots \\ u_{n,s} \end{pmatrix}_{n\times 1}; \quad \boldsymbol{\varepsilon}_s = \begin{pmatrix} \varepsilon_{1,s} \\ \varepsilon_{2,s} \\ \vdots \\ \varepsilon_{n,s} \end{pmatrix}_{n\times 1},$$

$$Y_0 = y_0 / h. \tag{18}$$

В итоге уравнение (6) запишем в векторном виде

$$Y_s = hU_s + p_2\varepsilon_s, \qquad s = 0,\ 1,\ 2, \ldots,\ m. \tag{19}$$

Полученные соотношения (16), (17), (19) – модель стохастической системы с распределенными параметрами, представленные как модель ПС (1)–(2).

При значениях n и s сгруппируем в (16) дискретные данные наблюдений $u_{k,j}$ за четыре года и выведем четыре уравнения с неизвестными $c_1,\ -c_2,\ c_3,\ d_1$:

$$\begin{cases} u_{n-4,s+1} = c_1u_{n-5,s} - c_2u_{n-4,s} + c_3u_{n-3,s} + \\ \qquad + u_{n-4,s} + d_1f_{n-4,s} + p_1w_{n-4,s}; \\ u_{n-3,s+1} = c_1u_{n-4,s} - c_2u_{n-3,s} + c_3u_{n-2,s} + \\ \qquad + u_{n-3,s} + d_1f_{n-3,s} + p_1w_{n-3,s}; \\ u_{n-2,s+1} = c_1u_{n-3,s} - c_2u_{n-2,s} + c_3u_{n-1,s} + \\ \qquad + u_{n-2,s} + d_1f_{n-2,s} + p_1w_{n-2,s}; \\ u_{n-1,s+1} = c_1u_{n-2,s} - c_2u_{n-1,s} + c_3u_{n,s} + \\ \qquad + u_{n-1,s} + d_1f_{n-1,s} + p_1w_{n-1,s}. \end{cases} \tag{20}$$

В уравнениях (16) и (19) p_1, p_2 есть коэффициенты при шумовых неконтролируемых переменных W и ε, которые дают количественную характеристику дисперсии аддитивных шумов динамики и измерительной системы, являются белыми гауссовскими последовательностями с нулевым средним и дисперсиями, равными единице. В работах [11–13] предлагается использование алгоритма расчета всех дисперсионных характеристик модели динамики

и измерительной системы с учетом совокупности выходных данных $\left\{y_{k,s},\ k=\overline{1,n},\ s=\overline{1,m}\right\}$.

Можно найти средние оценки дисперсий для $p_1 \approx \widehat{Q} = Q$ и $p_2 \approx \widehat{R} = R$. Данный алгоритм можно использовать для каждой строки $k=\overline{1,n}$. Оценки дисперсий используются при параметрической идентификации состояния объекта с применением калмановской фильтрации [2, 11, 13, 14].

В дальнейшем для расчетов оценок коэффициентов c_1, $-c_2$, c_3, d_1 можно использовать уравнения фильтра Калмана. Если рассматривать систему (21) как систему четырех уравнений с четырьмя неизвестными, то можно рассчитать четверку оценок c_1, $-c_2$, c_3, d_1 [15] на основе следующих формул:

$$\widehat{a}_1 = \left(\frac{u_{n-1,s+1}-u_{n-1,s}}{u_{n,s}} - \frac{u_{n-2,s+1}-u_{n-2,s}}{u_{n-1,s}}\right)\Bigg/\left(\frac{u_{n-2,s}-2u_{n-1,s}+u_{n,s}}{u_{n,s}} - \right.$$

$$\left. - \frac{u_{n-3,s}-2u_{n-2,s}+u_{n-1,s}}{u_{n-1,s}}\right), \tag{21}$$

$$\widehat{b}_1 = \frac{u_{n-2,s+1}-u_{n-2,s}}{u_{n-1,s}} - a_1\left(\frac{u_{n-3,s}-2u_{n-2,s}+u_{n-1,s}}{u_{n-1,s}}\right). \tag{22}$$

3. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ АЛГОРИТМА ИДЕНТИФИКАЦИИ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ

Для примера возьмем данные аккумуляции ресурсов домохозяйств за четыре года (2014–2017 гг.).

Пример.

а) Чтобы осуществить расчеты для проведения исследований на основе вычисления коэффициентов уравнений накоплений домохозяйств и в последующем для получения наиболее достоверных оценок поведения исследуемого объекта, по схеме фильтра Калмана промоделируем значения при следующих значениях индексов узлов сетки модели: $n = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$, $s = \{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8\}$. Данный пример рассчитывался в математической системе Matlab [16].

б) Задача численного оценивания коэффициентов распределенного уравнения параболического типа аккумуляции ресурсов множества домохозяйств при заданных данных наблюдений и краевых, начальных условиях. Получили значения коэффициентов уравнений: $\widehat{c}_1 = -0.0006477$; $\widehat{c}_2 = 1.000606$; $\widehat{c}_3 = -0.018023$; $\widehat{a}_1 = 74.6055$, использовав выражения (21), (22). Смоделированы значения наблюдений расходов R_0 и R_1 исследуемого объекта за 2014–2017 гг.

в) Более точно была решена задача оценивания состояния динамики объекта с применением калмановской фильтрации [11, 13]. Для расчета оце-

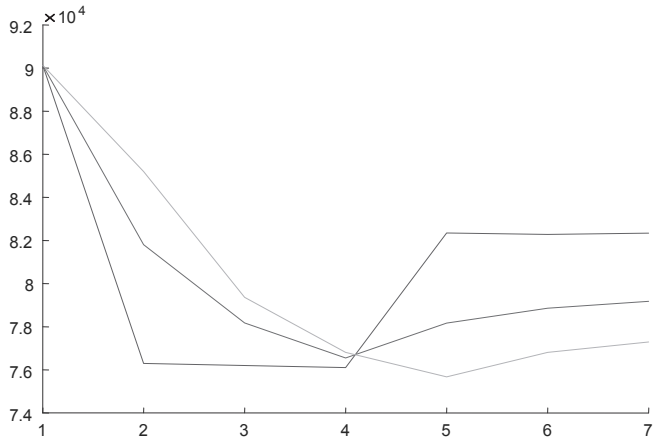
нок предсказаний и оценок фильтрации использованы данные наблюдений: $n = 24$, $y = [76200\ 76097\ 82343\ 82285\ 82339\ 79187\ 79111\ 79013\ 82485\ 82393\ 82307\ 83346\ 83286\ 83161\ 90188\ 90049\ 89992\ 86508\ 86425\ 86299\ 91191\ 91025\ 90960\ 94894]$; дисперсия динамики объекта $Q = 1282$; дисперсия измерительной системы $R = 33177$; расчетные коэффициенты дифференциального уравнения: $a = 0.7012$; $b = 21997$.

Результаты расчета и графики поведений наблюдений, оценок предсказания и фильтрации представлены на рисунке.

$xp = [90132\ 85198\ 79357\ 76812\ 75670\ 76806\ 77289]$; (Пр) – % оценки предсказания.

$xf = [90132\ 81803\ 78173\ 76544\ 78164\ 78853\ 79176]$; (Ф) – % оценки фильтрации.

$yS = [90132\ 76291\ 76200\ 76097\ 82343\ 82285\ 82339]$; (Н) – % наблюдения.



Графики наблюдений, оценок предсказания и фильтрации

Plots of Observation and Prediction and Filtering Estimates

г) Обратная задача оценивания коэффициентов уравнения аккумуляции ресурсов домохозяйств решена с применением фильтрационных оценок поведения объекта по схеме Калмана. Расчетные оценки $\hat{c}_1$, $\hat{c}_2$, $\hat{c}_3$, $\hat{d}_1$ были использованы для наиболее точного моделирования выходных данных системы.

ЗАКЛЮЧЕНИЕ

Рассмотренный алгоритм оценивания коэффициентов параболического уравнения аккумуляции ресурсов множества домохозяйств с дальнейшим применением фильтра Калмана более достоверен по сравнению с остальными численными методами. Для нахождения более точных значений коэффициентов уравнения (7) можно воспользоваться современными методами планирования эксперимента, использовать другие приближенные формулы для внутренних узлов сетки и уменьшить шаг сетки между соседними точками.

СПИСОК ЛИТЕРАТУРЫ

1. *Самарский А.А., Гулин А.В.* Численные методы математической физики. – М.: Научный мир, 2003. – 316 с.

2. *Горский В.Г., Адлер Ю.П., Талалай А.М.* Планирование промышленных экспериментов (модели динамики). – М.: Металлургия, 1978. – 112 с.

3. *Анисимов А.С.* Идентификация объектов управления: учебное пособие. – Новосибирск: НЭТИ, 1985. – 80 с.

4. *Mehra R.R.* Optimal input signals for parameter estimation in dynamic systems – Servey and new results // IEEE Transactions on Automatic Control. – 1974. – Vol. 19, iss. 6. – P. 753–768. – DOI: 10.1109/TAC.1974.1100701.

5. *Ерофеенко В.Т., Козловская И.С.* Уравнения с частными производными и математические модели в экономике: курс лекций. – Минск: Изд-во БГУ, 2004. – 245 с.

6. *Чернавский Д.С., Попков Ю.С., Рахимов А.Х.* Математические модели типологии семейных накоплений // Экономика и математические методы. – 1994. – Т. 30, вып. 2. – С. 98–106.

7. *Араманович И.Г., Левин В.И.* Уравнения математической физики. – М.: Наука, 1969. – 288 с.

8. *Абденова Г.А.* Структурно-параметрическая идентификация систем с распределенными параметрами с использованием модели типа «вход–состояние–выход» // Научный вестник НГТУ. – 2006. – № 1 (38). – С. 9–16.

9. *Абденов А.Ж., Абденова Г.А.* Методика пассивной идентификации коэффициентов уравнения теплопроводности с учетом ошибок оценок состояния объекта и измерительной системы // Автометрия. – 2016. – Т. 52, № 2. – С. 43–51. – DOI: 10.15372/AUT20160205.

10. Оценивание коэффициентов уравнения теплопроводности с учетом шумов измерительной системы / А.Ж. Абденов, Ж.К. Нурбекова, В.Б. Уткин, Г.А. Абденова // Вестник ЕНУ им. Л.Н. Гумилева. – 2015. – № 4 (107). – С. 5–13.

11. *Mehra R.R.* Identification and adaptive Kalman filtering // Mechanics. – 1971. – N 3. – P. 34–52.

12. *Абденова Г.А., Воевода А.А.* Оценивание параметров и характеристик шумов нестационарных процессов в стохастических системах, описываемых в пространстве состояний // Сборник научных трудов НГТУ. – 2010. – № 3 (61). – С. 11–18.

13. *Абденова Г.А.* Прогнозирование значений уровня временного ряда на основе уравнений фильтра Калмана // Ползуновский вестник. – 2010. – № 2. – С. 4–6.

14. *Эйкхофф П.* Основы идентификации систем управления. – М.: Мир, 1975. – 680 с.

15. *Abdenov A., Abdenova G., Kulbaev D.* Estimation of equation coefficients of free and forced string vibrations in continuous medium with consideration of dynamic noise and measurement // Materials Today: Proceedings. – 2019. – Vol. 16. – P. 336–342.

16. *Дьяконов В.П.* MATLAB. Полный самоучитель. – М.: ДМК Пресс, 2012. – 768 с.

Абденова Гаухар Амирзаевна, доцент кафедры «Математическое и компьютерное моделирование» механико-математического факультета Евразийского национального университета им. Л.Н. Гумилева, кандидат технических наук, PhD. Область научных интересов: математическое моделирование в экономике, параметрическая идентификация систем. Автор более 40 научных работ. E-mail: gauhar.phd@gmail.com

Базикова Карлыгаш Манаповна, докторант специальности «Математическое и компьютерное моделирование» Евразийского национального университета им. Л.Н. Гумилева. Направление научных интересов – математическое и компьютерное моделирование. Имеет 2 печатные работы. E-mail: kmbazikova@mail.ru

Кенжегалым Жангүл Медетқызы, магистрант специальности «Математическое и компьютерное моделирование» Евразийского национального университета им. Л.Н. Гу-

милева. Направление научных интересов – математическое и компьютерное моделирование. Научных работ и учебных пособий пока не имеет. E-mail: zhangul_flower@mail.ru

Abdenova Gaukhar A., associate professor at the Department of Mathematical and Computer Modeling, Faculty of Mechanics and Mathematics, L.N. Gumilyov Eurasian National University, PhD (Eng.). Her research interests are focused on mathematical modeling in economics and parametric identification of systems. She has more than 40 scientific publications. E-mail: gauhar.phd@gmail.com

Bazikova Karlygash M., a postdoctoral student specializing in "Mathematical and Computer Modeling", L.N. Gumilyov Eurasian National University. Her research interests are focused on mathematical and computer modeling. She has 2 scientific publications. E-mail: kmbazikova@mail.ru

Kenzhegalym Zhangul M., a master degree student specializing in "Mathematical and Computer Modeling", L.N. Gumilyov Eurasian National University. His research interests include mathematical and computer modeling. He has not published any papers yet. E-mail: zhangul_flower@mail.ru

DOI: 10.17212/2782-2001-2021-2-7-18

A parabolic model in the form of space states of the dynamics of savings[*]

G.A. ABDENOVA[a], *K.M. BAZIKOVA*[b], *Zh.M. KENZHEGALYM*[c]

L.N. Gumilyov Eurasian National University, 2 Satpayev Street, Nur-Sultan, 010008, Republic of Kazakhstan

[a] *gauhar.phd@gmail.com* [b] *kmbazikova@mail.ru* [c] *zhangul_flower@mail.ru*

Abstract

An important place in the theory of partial differential equations and its applications is occupied by the heat equation, a representative of the class of the so-called parabolic equations. It is known that to check the correctness of a mathematical model based on a parabolic equation, the existence of its solution is very important since a mathematical model is not always adequate to a specific phenomenon and the existence of a solution to a corresponding mathematical problem does not follow from the existence of a solution to a real applied problem. Therefore, methods for solving partial differential equations, both analytical and numerical, are always relevant.

Nowadays, a computational experiment has become a powerful tool for theoretical research. It is carried out over a mathematical model of the object under study, but at the same time, other parameters are calculated using one of the parameters of the model and conclusions are drawn about the properties of the object or phenomenon under study. The problem of passive parametric identification of systems with distributed parameters for resource accumulation dynamics of many households using a stochastic distributed model in the form of a state space with regard to the white noise of the dynamics model of the object under study and the white noise of the model of a linear-type measuring system is considered in the paper. The use of the finite difference method allowed us to reduce the solution of partial differential equations of a parabolic type to the solution of a system of linear finite difference and algebraic equations represented by models in the form of a state space. It was also proposed to use a filtering algorithm based on the Kalman scheme for reliable estimation of the object behavior.

Calculations were carried out using the Matlab mathematical system based on statistical data for five years, taken from the site “Agency for Strategic Planning and Reforms of the Republic of Kazakhstan Bureau of National Statistics”. Estimation of the coefficients of the equa-

[*] *Received 18 January 2021.*

tions for the household resource accumulation in the form of a state space using this technique is sufficiently universal and can be applied in other fields of science and technology.

Keywords: partial differential equation, stochastic dynamical system, input and output observational data, passive identification, additive white noise, linear finite-difference equation, parabolic equation, model in the form of a state space, Kalman filter

REFERENCES

1. Samarskii A.A., Gulin A.V. *Chislennye metody matematicheskoi fiziki* [Numerical methods of mathematical physics]. Moscow, Nauchnyi mir Publ., 2003. 316 p.

2. Gorskii V.G., Adler Yu.P., Talalai A.M. *Planirovanie promyshlennykh eksperimentov (modeli dinamiki)* [Planning of industrial experiments. Models of dynamics]. Moscow, Metallurgiya Publ., 1978. 112 p.

3. Anisimov A.S. *Identifikatsiya ob"ektov upravleniya* [Identifying control objects]. Novosibirsk, NETI Publ., 1985. 80 p.

4. Mehra R.R. Optimal input signals for parameter estimation in dynamic systems – Servey and new results. *IEEE Transactions on Automatic Control*, 1974, vol. 19, iss. 6, pp. 753–768. DOI: 10.1109/TAC.1974.1100701.

5. Erofeenko V.T., Kozlovskaya I.S. *Uravneniya s chastnymi proizvodnymi i matematicheskie modeli v ekonomike* [Equations with private derivatives and mathematical models in economics]. Minsk, BSU Publ., 2004. 245 p.

6. Chernavskii D.S., Popkov Yu.S., Rakhimov A.Kh. Matematicheskie modeli tipologii semeinykh nakoplenii [Mathematical models of family savings typology]. *Ekonomika i matematicheskie metody = Economics and Mathematical Methods* , 1994, vol. 30, iss. 2, pp. 98–106.

7. Aramanovich I.G., Levin V.I. *Uravneniya matematicheskoi fiziki* [Equations of mathematical physics]. Moscow, Nauka Publ., 1969. 288 p.

8. Abdenova G.A. Strukturno-parametricheskaya identifikatsiya sistem s raspredelennymi s ispol'zovaniem modeli tipa "vkhod-sostoyanie-vykhod" [Structural and parametric identification of the systems with distributed parameters of using the model of the type "entry-state-exit"]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Science bulletin of the Novosibirsk state technical university*, 2006, no. 1 (38), pp. 9–160.

9. Abdenov A.Zh., Abdenova G.A. Metodika passivnoi identifikatsii koeffitsientov uravneniya teploprovodnosti s uchetom oshibok otsenok sostoyaniya ob"ekta i izmeritel'noi sistemy [Passive identification of heat equation coefficients with account for errors in estimating the state of the object and measuring system]. *Avtometriya = Optoelectronics, Instrumentation and Data Processing*, 2016, vol. 52, no. 2, pp. 43–51. DOI: 10.15372/AUT20160205. (In Russian).

10. Abdenov A.Zh., Nurbekova Zh.K., Utkin V.B., Abdenova G.A. Otsenivanie koeffitsientov uravneniya teploprovodnosti s uchetom shumov izmeritel'noi sistemy [Coefficients evaluation of heat equation with account of noise measuring system]. *Vestnik ENU im. L.N. Gumileva = Herald of the L.N. Gumilyov Eurasian National University*, 2015, no. 4 (107), pt. 1, pp. 5–13.

11. Mehra R.R. Identification and adaptive Kalman filtering. *Mechanics*, 1971, no. 3, pp. 34–52.

12. Abdenova G.A., Voevoda A.A. Otsenivanie parametrov i kharakteristik shumov nestatsionarnykh protsessov v stokhasticheskikh sistemakh, opisyvaemykh v prostranstve sostoyanii [The parameters estimation and noise characteristics of time-varying processes in the stochastic systems in the state space]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Transaction of scientific papers of the Novosibirsk state technical university*, 2010, no. 3 (61), pp. 11–18.

13. Abdenova G.A. Prognozirovanie znachenii urovnya vremennogo ryada na osnove uravnenii fil'tra Kalmana [Prediction of time series level values based on Kalman filter equations]. *Polzunovskii vestnik = Polzunov Bulletin*, 2010, no. 2, pp. 4–6.

14. Eykhoff P. *System identification*. London, New York, Wiley-Interscience, 1974 (Russ. ed.: Eikkhoff P. *Osnovy identifikatsii sistem upravleniya*. Moscow, Mir Publ., 1975. 680 p.).

15. Abdenov A., Abdenova G., Kulbaev D. Estimation of equation coefficients of free and forced string vibrations in continuous medium with consideration of dynamic noise and measurement. *Materials Today: Proceedings*, 2019, vol. 16, pp. 336–342.

16. D'yakonov V.P. *MATLAB. Polnyi samouchitel'* [MATLAB. Complete tutorial]. Moscow, DMK Press Publ., 2012. 768 p.

Для цитирования:

Абденова Г.А., Базикова К.М., Кенжеғалым Ж.М. Параболическая модель в форме пространства состояний динамики накоплений // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 7–18. – DOI: 10.17212/2782-2001-2021-2-7-18.

For citation:

Abdenova G.A., Bazikova K.M., Kenzhegalym Zh.M. Parabolicheskaya model' v forme prostranstva sostoyanii dinamiki nakoplenii [A parabolic model in the form of space states of the dynamics of savings]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 7–18. DOI: 10.17212/2782-2001-2021-2-7-18.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 19–34

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 19–34

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 681.5 DOI: 10.17212/2782-2001-2021-2-19-34

Исследование кибербезопасности прогностических алгоритмов управления для установок распределенной генерации*

Ю.Н. БУЛАТОВ[1,a], А.В. КРЮКОВ[2,3,b]

[1] *665709, РФ, г. Братск, ул. Макаренко, 40, Братский государственный университет*
[2] *664074, РФ, г. Иркутск, ул. Чернышевского, 15, Иркутский государственный университет путей сообщения*
[3] *664074, РФ, г. Иркутск, ул. Лермонтова, 83, Иркутский национальный исследовательский технический университет*

[a] *bulatovyura@yandex.ru* [b] *and_kryukov@mail.ru*

В электроэнергетике активно развивается направление, связанное с применением установок распределенной генерации (РГ), расположенных вблизи энергопринимающих устройств потребителей. Внедрение установок РГ связано с решением множества технических задач, одной из которых является оптимизация настроек автоматических регуляторов возбуждения (АРВ) и скорости (АРС) синхронных генераторов во всех возможных режимах работы. Это требует использования сложных моделей систем электроснабжения, установок распределенной генерации и их регуляторов, а также трудоемких расчетов, учитывающих большое количество взаимосвязанных параметров. Однако существует и другой подход, основанный на применении прогностических регуляторов, для настройки которых при линейных прогнозных моделях нужен только один параметр.

В статье приведено описание метода построения и настройки предлагаемого прогностического АРС синхронного генератора, а также компьютерных моделей установок распределенной генерации, используемых для проведения исследований. Цель исследований состояла в определении кибербезопасности систем электроснабжения, оснащенных установками распределенной генерации с прогностическими регуляторами скорости, которые могут быть реализованы на базе микропроцессорной техники. Исследования проводились в системе MATLAB с применением пакетов имитационного моделирования Simulink и SymPowerSystems на компьютерных моделях установок РГ с одним турбогенератором, работающим на выделенную нагрузку, а также группой гидрогенераторов, связанных с электроэнергетической системой большой мощности. Результаты моделирования показали эффективность предлагаемых прогностических алгоритмов управления, а также то, что их кибербезопасность можно повысить

* *Статья получена 26 января 2021 г.*
Работа выполнена при финансовой поддержке по гранту государственного задания Министерства науки и высшего образования России (проект № 0667-2020-0039).

путем введения аппаратных ограничений на диапазон изменения постоянной времени прогнозирующего звена.

Ключевые слова: установки распределенной генерации, синхронные генераторы, прогностические алгоритмы управления, кибербезопасность, автоматический регулятор скорости, линейное прогнозирующее звено, система электроснабжения, компьютерное моделирование

ВВЕДЕНИЕ

В настоящее время активно развивается направление, связанное с использованием в электроэнергетике установок распределенной генерации (РГ) [1–5], расположенных вблизи энергопринимающих устройств потребителей. Применение таких установок позволяет уменьшить нагрузки питающей сети, снизить потери, повысить надежность электроснабжения потребителей и качество электроэнергии [2, 6]. Однако внедрение установок РГ требует решения множества задач, одной из которых является оптимизация настроек автоматических регуляторов возбуждения (АРВ) и скорости (АРС) синхронных генераторов [7–10] во всех возможных режимах с целью обеспечения их надежной и устойчивой работы в системе электроснабжения (СЭС), имеющей связь с электроэнергетической системой (ЭЭС) большой мощности. Это требует использования сложных моделей ЭЭС и СЭС, установок РГ и их автоматических регуляторов, а также трудоемких расчетов, учитывающих большое количество взаимосвязанных параметров.

Существует другой подход, основанный на применении прогностических алгоритмов [11, 12], предполагающих работу регулятора по рассчитанному прогнозу погрешности $\varepsilon\ (t + \Delta t)$. Прогнозирование с помощью линейной модели может быть реализовано по двум значениям регулируемой величины: текущему и предыдущему. Проведенные исследования на компьютерных моделях СЭС с установками РГ [13, 14] показали, что применение линейного прогнозирующего звена в АРВ и АРС позволяет улучшить показатели качества управления напряжением и частотой при различных возмущениях, а также позволяет придать регулятору скорости адаптивные свойства [15].

Построение прогностических регуляторов установок РГ возможно с использованием цифровых (микропроцессорных) АРВ и АРС. Для этого необходимо изучить влияние произвольного изменения постоянной времени прогнозирующего звена на работу систем управления (СУ) распределенными генераторами СЭС. Такое изменение может быть проведено в результате кибератаки на СУ энергообъектов.

Ниже приводится описание математической модели автопрогностического (самонастраивающегося) АРС синхронного генератора, метода его настройки и компьютерных моделей установок РГ, с использованием которых были проведены исследования. Цель исследований состояла в определении устойчивости линейных прогностических алгоритмов управления к несанкционированным изменениям времени прогноза. Исследования проводились в системе MATLAB на компьютерных моделях установок РГ с одним генератором и группой генераторов.

1. МЕТОД ПОСТРОЕНИЯ И НАСТРОЙКИ ПРОГНОСТИЧЕСКОГО РЕГУЛЯТОРА СКОРОСТИ СИНХРОННОГО ГЕНЕРАТОРА

Частота переменного тока является важным параметром, характеризующим качество электрической энергии, и к ее стабильности предъявляются жесткие требования. Выполнить эти требования можно на основе использования алгоритмов, позволяющих с минимальными затратами усовершенствовать классический АРС. Например, применение алгоритмов прогноза позволяет повысить эффективность работы систем автоматического управления при сохранении настроек типового регулятора [14, 16]. Прогностический регулятор может состоять из двух последовательно соединенных элементов: пропорциально-интегрально-дифференциального (ПИД) регулятора и прогнозирующего звена. Наиболее просто реализуется линейное прогнозирующее звено (ЛПЗ) со следующей передаточной функцией [14]:

$$W_p(s) = T_p s + 1, \tag{1}$$

где T_p – постоянная времени ЛПЗ, или время прогноза; s – оператор Лапласа.

При использовании прогностических регуляторов возникает задача определения времени прогноза для различных режимов работы объекта регулирования. В работе [14] показано, что величина T_p должна быть близкой к десятой доле периода собственных колебаний оптимально настроенной замкнутой системы автоматического регулирования. При этом исследования, проведенные в [14] для двухконтурной системы регулирования, показывают, что наибольший эффект от прогнозирования удается достичь при увеличении времени прогноза до 100 с.

Для настройки прогностического регулятора скорости предлагается использовать частоту f_p собственных колебаний ротора генератора [17, 18]:

$$f_p = \frac{\sqrt{P_s \dfrac{\omega_0}{T_{je}}}}{2\pi}; \qquad P_s = \frac{E_q U_g}{X_d}\cos\delta, \tag{2}$$

где P_s – синхронизирующая мощность генератора, о. е.; U_g – номинальное напряжение, о. е.; δ – угол между напряжением и ЭДС E_q, эл. град.; X_d – индуктивное сопротивление по продольной оси, о. е.; ω_0 – номинальная угловая частота, рад/с; T_{je} – эквивалентная постоянная механической инерции агрегата, с.

Постоянная времени ЛПЗ определяется так: $T_p = 1/f_p$.

Проведенные исследования [13, 15] показали, что наибольший эффект удается получить при использовании усилителя, включенного последовательно с ЛПЗ.

Частота f_p и обратно пропорциональная ей постоянная времени T_p зависят от угла нагрузки δ. Поэтому целесообразно корректировать T_p в режиме реального времени при вариации режима работы синхронного генератора. Струк-

турная схема предлагаемой системы регулирования скорости с автоматически настраиваемым ЛПЗ показана на рис. 1. Следует отметить, что функция $f_p(\delta)$ зависит от параметров конкретного агрегата и сети, к которой он подключен.

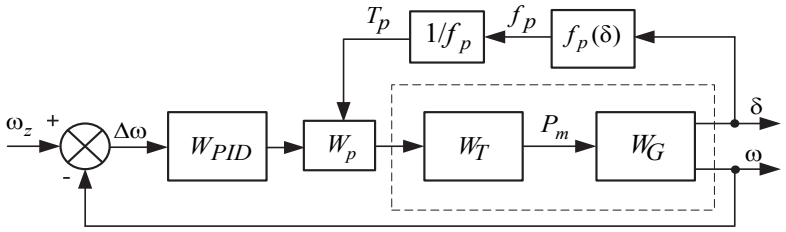


Рис. 1. Система регулирования частоты, использующая автопрогностический АРС; отдельным блокам отвечают следующие передаточные функции:

W_G – генератор; W_T – турбина; W_{PID} – АРС (ПИД-регулятор); W_p – ЛПЗ

Fig. 1. A frequency control system using an auto-predictive automatic speed regulator (ASR); the following transfer functions correspond to individual units:

W_G is the generator; W_T is the turbine; W_{PID} is ASR (a PID controller); W_p is the linear predictor

Предлагаемая структура системы регулирования частоты позволяет модернизировать классический АРС и автоматически в режиме реального времени вычислять постоянную T_p прогнозирующего звена. Для ее правильного определения необходимо задать параметры агрегата установки РГ в соответствии с выражением (2). Представленная структура системы регулирования частоты может использоваться для различных установок РГ: турбогенераторных, гидрогенераторных, газотурбинных, дизельных и других.

2. ОПИСАНИЕ КОМПЬЮТЕРНЫХ МОДЕЛЕЙ УСТАНОВОК РГ, ИСПОЛЬЗУЮЩИХ АВТОПРОГНОСТИЧЕСКИЙ АРС

Компьютерные модели установок РГ с автопрогностическим АРС создавались в системе MATLAB с использованием пакетов Simulink и SymPowerSystems для автономно работающей турбогенераторной установки (ТГУ) и для малой гидроэлектростанции (ГЭС), состоящей из трех агрегатов.

Для турбогенератора, приводимого во вращение паровой турбиной, использовался блок библиотеки SymPowerSystems системы MATLAB – Synchronous Machine. Схема модели ТГУ, показанная на рис. 2, состоит из следующих основных блоков: синхронная машина (блок Synchronous Machine); возбудитель (блок Exciter); АРВ и АРС (блоки MARE и Automatic regulator of rotor speed); паровая турбина (блок Steam turbine).

На рис. 3 представлена схема модели паровой турбины с промежуточным отбором пара.

Для системы возбуждения использовалась передаточная функция [18] $\frac{1}{0{,}025s+1}$.

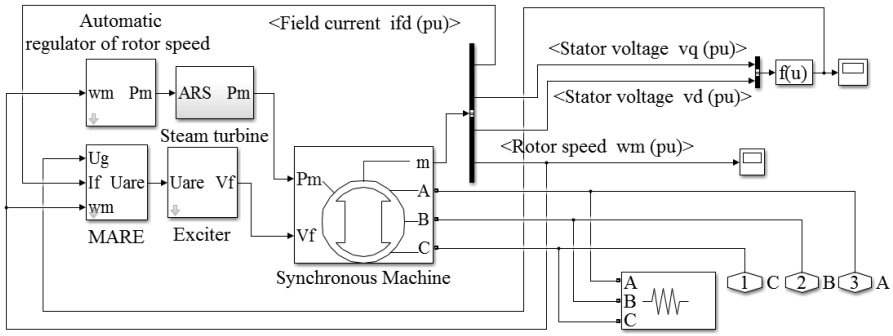


Рис. 2. Схема модели ТГУ в MATLAB

Fig. 2. Diagram of a turbine generator set (TGS) model in MATLAB

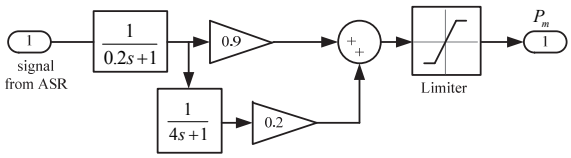


Рис. 3. Структурная схема модели паровой турбины

Fig. 3. Block diagram of a steam turbine model

Структурная схема модели АРВ, подробное описание которого приведено в работе [10], показана на рис. 4. Схема модели АРС приведена на рис. 5, где также представлено выражение для расчета постоянной времени ЛПЗ, которое было определено в соответствии с параметрами турбогенератора: $X_d = 2{,}34$ о. е., $E_q = 1{,}25$ о. е., $U_g = 1$ о. е., $T_{je} = 5{,}13$ с. С помощью блока Step имитировалось внешнее воздействие на прогнозирующее звено регулятора, изменяющее время прогноза.

Коэффициенты настройки АРВ и АРС определялись из практических соображений.

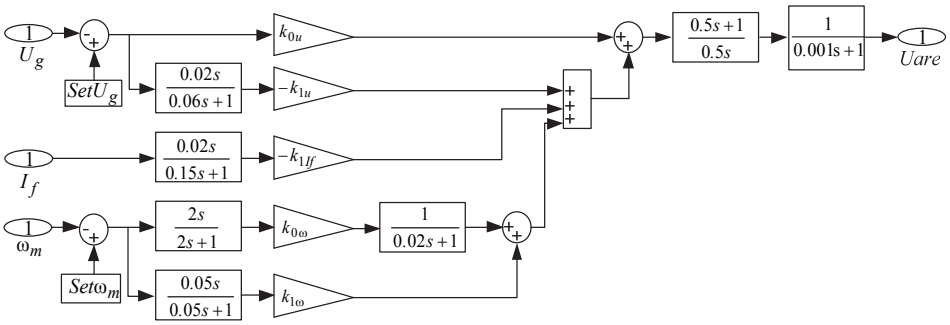


Рис. 4. Схема модели АРВ:

U_g – текущее напряжение; $SetU_g$ – заданное значение напряжения; I_f – ток возбуждения; ω_m – текущее значение частоты вращения ротора; $Set\omega_m$ – заданное значение частоты вращения; k_{0u}, k_{1u}, k_{1If}, $k_{0\omega}$, $k_{1\omega}$ – коэффициенты настройки АРВ

Fig. 4. Automatic voltage regulator (AVR) model diagram:

U_g is the current voltage; $SetU_g$ is a preset voltage value; I_f is the excitation current; ω_m is the current value of the rotor speed; $Set\omega_m$ is a set value of the rotor speed; k_{0u}, k_{1u}, k_{1If}, $k_{0\omega}$, $k_{1\omega}$ are tuning factors of AVR

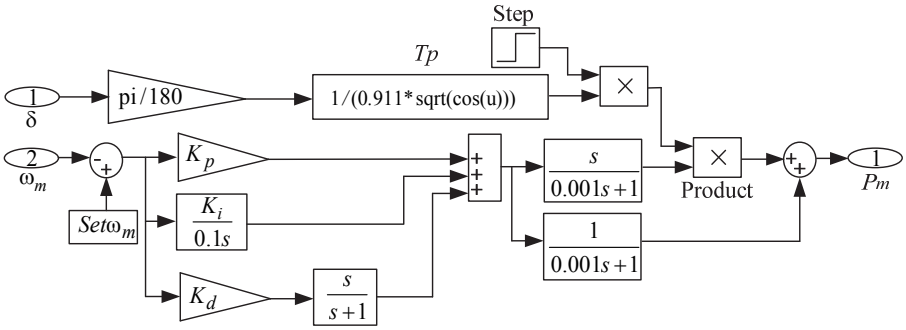


Рис. 5. Модель автопрогностического АРС:

K_p, K_i, K_d – коэффициенты настройки

Fig. 5. Auto-predictive ASR model:

K_p, K_i, K_d are tuning factors

Гидрогенератор моделировался по модели, представленной на рис. 2. Модель гидротурбины показана на рис. 6, где приведены передаточные функции и параметры главного сервомотора и гидротурбины [18].

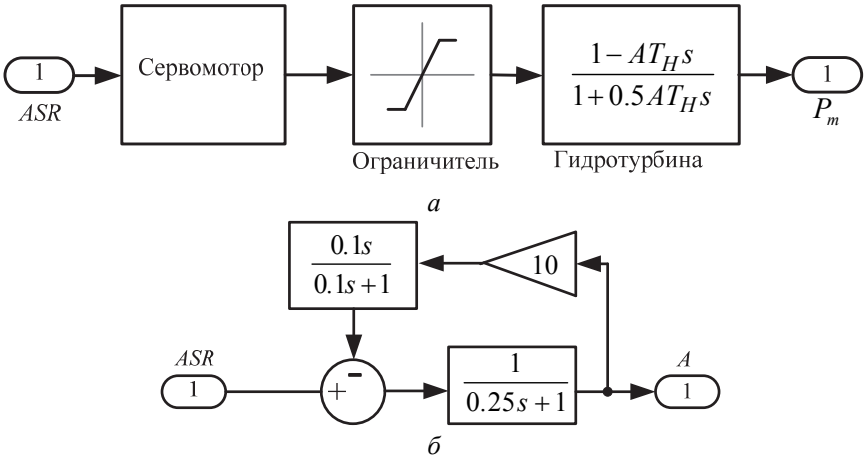


Рис. 6. Модель гидротурбины (*а*) и главного сервомотора (*б*):

T_H = 0.344 с – постоянная времени гидротурбины; A – положение открытия направляющего аппарата

Fig. 6. Model of the hydraulic turbine (*a*) and the main servo motor (*b*):

T_H = 0.344 s is the time constant of the hydraulic turbine; A is an opening position of the guide vane

При моделировании использовались следующие параметры гидрогенератора: $X_d = 2{,}84$ о. е., $E_q = 1{,}1$ о. е., $U_g = 1$ о. е., $T_{je} = 3{,}779$ с. В соответствии с этими параметрами постоянная времени ЛПЗ автопрогностического АРС определялась так:

$$T_p = \frac{1}{f_p(\delta)} = \frac{1}{0{,}903\sqrt{\cos\delta}}.$$

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Исследовались устойчивость и правильность работы линейных прогностических алгоритмов управления в случае несанкционированного изменения времени прогноза. В качестве возмущения рассматривалось кратковременное трехфазное короткое замыкание (КЗ). Исследования проводились в системе MATLAB на компьютерных моделях установок РГ с одним генератором и группой генераторов (рис. 7).

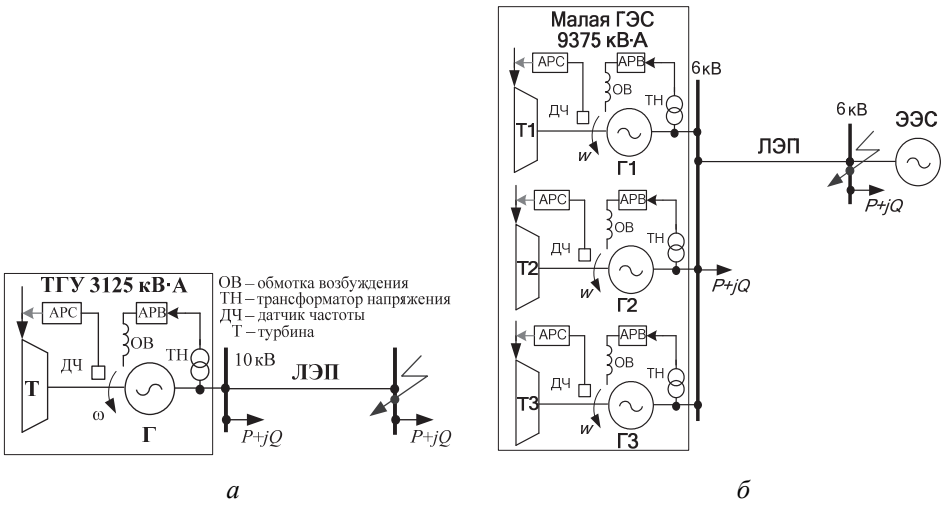


Рис. 7. Схемы СЭС:

а – изолированная СЭС с ТГУ; *б* – группа гидрогенераторов малой ГЭС, имеющей связь с ЭЭС

Fig. 7. Power supply system (PSS) diagrams:

a is an isolated PSS with TGS; *b* is a group of hydrogenerators of a small hydroelectric power station which is connected to the electric power system

Рассматривались следующие схемы и режимы работы СЭС с измененным в результате кибератаки временем прогноза АРС:

1) работа ТГУ на выделенную нагрузку (рис. 7, *а*) в нормальном режиме и при трехфазном КЗ у потребителя, отключающемся релейной защитой через 0,3 с;

2) работа малой ГЭС, состоящей из трех гидрогенераторов, связанных через линию с ЭЭС большой мощности (рис. 7, *б*) в тех же режимах.

В результате кибератаки постоянная времени прогнозирующего звена автоматического регулятора скорости принимала значения, умноженные на –100, –1, 0, 100, 10^6.

Результаты моделирования работы ТГУ на выделенную нагрузку (рис. 7, *а*) представлены в виде временных зависимостей скорости вращения ротора и напряжения генератора на рис. 8 и 9. В рассматриваемой изолированной СЭС частота напряжения и тока определяется скоростью вращения ротора генератора (рис. 8). По результатам моделирования можно сделать вывод о том, что увеличение, а также зануление постоянной времени прогнозирующего звена не выводит синхронный генератор из устойчивой работы даже при коротком

замыкании. Однако при $T_p = 0$ увеличивается колебательность и перерегулирование, повышается время переходного процесса по сравнению с режимами, когда T_p имеет нормальное или увеличенное значение (рис. 8). При отрицательной постоянной времени ЛПЗ наблюдается неправильная работа АРС, приводящая при КЗ к потере устойчивости и резкому снижению частоты вращения ротора генератора (характеристики *5* и *6* на рис. 8). Необходимо также отметить влияние отрицательных значений T_p на напряжение генератора, приводящее к появлению колебаний, которые после КЗ усиливаются (рис. 9, *б*). Таким образом, работа автопрогностического АРС с отрицательным значением постоянной времени линейного прогнозирующего звена недопустима.

Рис. 8. Временная зависимость скорости вращения ротора генератора при КЗ с измененной в результате кибератаки постоянной времени прогноза:

1 – $T_p = 0$; *2* – T_p рассчитывается автоматически и не изменяется; *3* – T_p увеличена в 100 раз; *4* – T_p увеличена в 10^6 раз; *5* – T_p умножена на –1; *6* – T_p умножена на –100

Fig. 8. Time dependence of the generator rotor speed at short circuit with the forecast time constant changed as a result of the cyber attack:

1 is $T_p = 0$; *2* – T_p is calculated automatically and does not change; *3* is T_p increased by 100 times; *4* is T_p increased by 10^6 times; *5* is T_p multiplied by –1; *6* is T_p multiplied by –100

Результаты моделирования при возмущениях для СЭС с группой гидрогенераторов, связанных с ЭЭС, представлены в виде временных зависимостей напряжения, скорости вращения и частоты на рис. 10–12. Изменение в результате кибератаки постоянной времени ЛПЗ одного из гидрогенераторов не привело к потере устойчивости при КЗ в системе; при этом нулевое и отрицательное значение T_p вызывало рост величин перерегулирования, колебательности и времени переходного процесса для напряжения, скорости ротора и частоты (характеристики *1* и *4* на рис. 10–12) по сравнению с автоматически рассчитанным значением T_p.

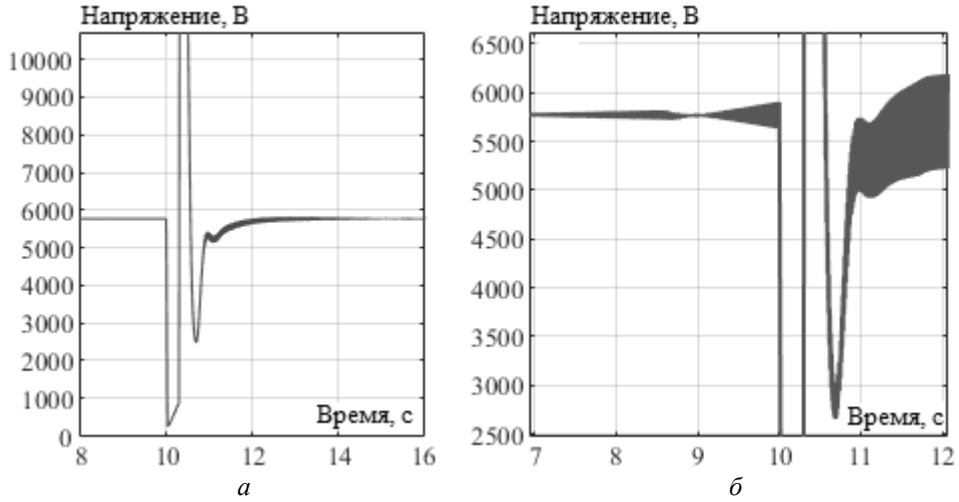


Рис. 9. Временная зависимость напряжения генератора при КЗ с измененной в результате кибератаки постоянной времени прогноза:

а – T_p увеличена до 10^6; *б* – T_p умножена на –100

Fig. 9. Time dependence of the generator voltage at short circuit with the forecast time constant changed as a result of a cyber attack:

a is T_p increased to 10^6; *b* is T_p multiplied by –100

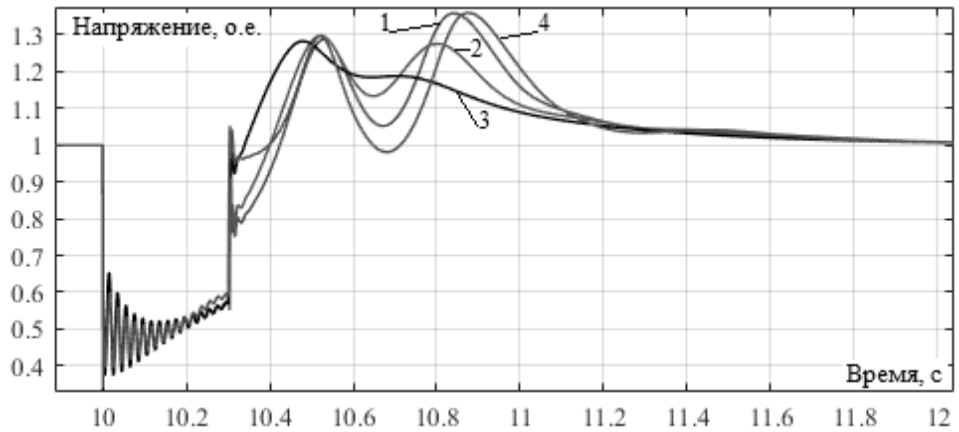


Рис. 10. Временные зависимости напряжений:

1 – $T_p = 0$; *2* – T_p рассчитывается автоматически и не изменяется; *3* – T_p увеличена в *n* раз, где *n* – число параллельно работающих генераторов; *4* – T_p умножена на –100

Fig. 10. Time dependences of voltages:

1 is $T_p = 0$; *2* is T_p is calculated automatically and does not change; *3* is T_p increased *n* times, where *n* is the number of parallel operating generators; *4* is T_p multiplied by –100

Результаты моделирования также показали, что лучшего эффекта регулирования удается достичь при увеличении постоянной времени ЛПЗ в *n* раз, где *n* – число параллельно работающих генераторов в группе (характеристики *3* на рис. 10–12): практически отсутствует колебательность, значительно уменьшается перерегулирование и время переходного процесса для напряже-

ния, скорости ротора и частоты в сети после возникновения и отключения КЗ; снижается провал частоты; управление скоростью вращения ротора становится более плавным.

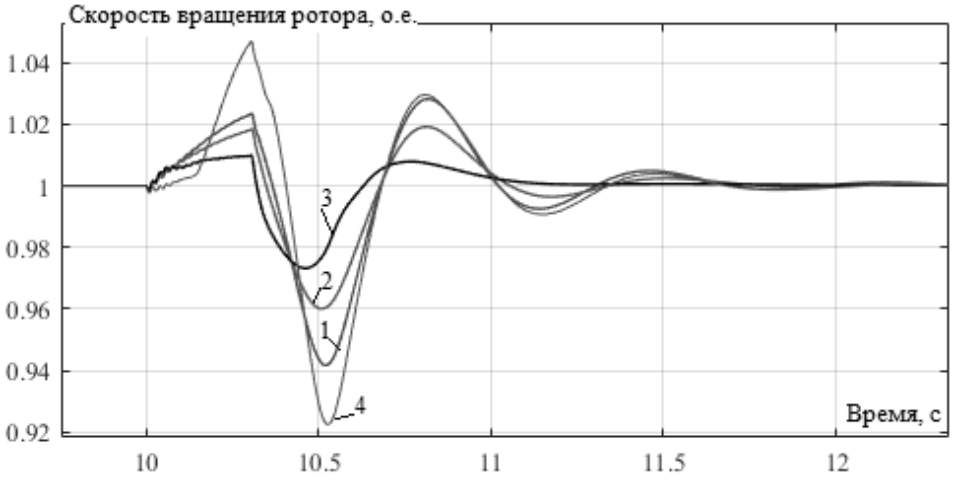


Рис. 11. Временные зависимости скорости вращения ротора: обозначения *1*, *2*, *3* и *4* аналогичны рис. 10

Fig. 11. Time dependences of the rotor speed: designations *1*, *2*, *3* and *4* are similar to Fig. 10

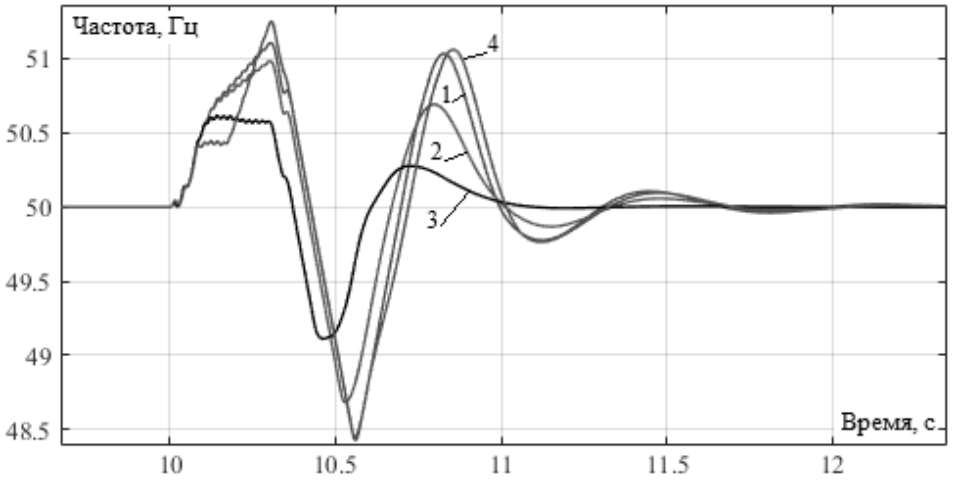


Рис. 12. Временные зависимости частоты: обозначения *1*, *2*, *3* и *4* аналогичны рис. 10

Fig. 12. Time dependences of frequency: designations *1*, *2*, *3* and *4* are similar to Fig. 10

Необходимо также отметить, что одновременное воздействие в результате кибератаки на постоянные времени прогнозирующих звеньев АРС отдельных гидрогенераторов приводит к возникновению низкочастотных колебаний параметров всех агрегатов, а также параметров системы электроснабжения. В качестве примера на рис. 13 представлено изменение частоты в ре-

зультате одновременного воздействия на ЛПЗ АРС гидрогенераторов. Для повышения надежности, равномерной загруженности и предотвращения возможности несанкционированного задания разных T_p для параллельно работающих синхронных генераторов предлагается использовать общую групповую систему определения постоянной времени ЛПЗ по углу нагрузки δ одного из генераторов, принимаемого в качестве ведущего. Аппаратное ограничение T_p позволит избежать негативных последствий от внешних воздействий, приводящих к снижению демпферных свойств и даже к потере устойчивости.

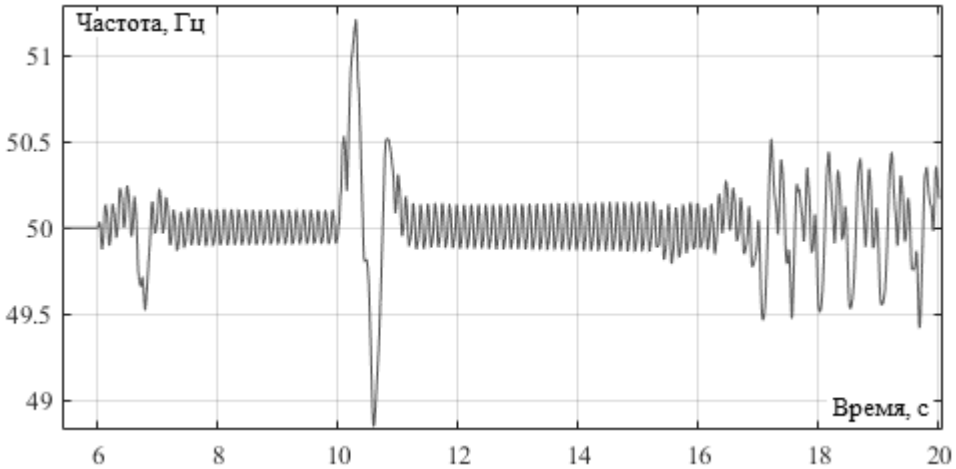


Рис. 13. Временная зависимость частоты при одновременном воздействии на все АРС: частота колебаний составляет 8…10 Гц

Fig. 13. Time dependence of frequency with simultaneous exposure to all ASR: vibration frequency is 8…10 Hz

ЗАКЛЮЧЕНИЕ

На основе результатов компьютерного моделирования можно сформулировать следующие выводы.

1. Увеличение или зануление постоянной времени прогнозирующего звена АРС не выводит синхронный генератор из устойчивой работы даже при последующем коротком замыкании. Однако при $T_p = 0$ увеличивается колебательность и перерегулирование, повышается время переходного процесса.

2. При отрицательном значении постоянной времени прогнозирующего звена наблюдается неправильная работа АРС, приводящая при КЗ к потере устойчивости и резкому снижению частоты генератора, работающего на выделенную нагрузку. Отрицательные значения T_p также вызывают колебания напряжения, которые после КЗ усиливаются. Таким образом, работа автопрогностического АРС с отрицательным значением постоянной времени прогноза недопустима.

3. Изменение в результате кибератаки постоянной времени ЛПЗ у АРС одного из параллельно работающих гидрогенераторов не приводит к потере

устойчивости при последующем коротком замыкании. Нулевое и отрицательное значения T_p приводят к повышению величины перерегулирования, колебательности и времени переходного процесса для напряжения, скорости и частоты в сети.

4. Для группы параллельно работающих генераторов наилучшего эффекта регулирования удается достичь при увеличении постоянной времени прогнозирующего звена в *n* раз, где *n* – число параллельно работающих генераторов. В этом случае отсутствует колебательность, уменьшается перерегулирование и время переходного процесса для напряжения, скорости ротора генератора и частоты в сети после возникновения и отключения КЗ, а также снижается провал частоты.

5. Одновременное несанкционированное воздействие на постоянные времени прогнозирующих звеньев АРС отдельных параллельно работающих генераторов может приводить к возникновению низкочастотных колебаний. Для повышения надежности, равномерной загруженности и предотвращения возможности несанкционированного задания разных постоянных времени предлагается использовать общую групповую систему определения этого параметра по углу нагрузки δ одного из генераторов, принимаемого в качестве ведущего.

6. Аппаратное ограничение диапазона изменения постоянной времени прогнозирующего звена АРС позволит избежать негативных последствий от внешних воздействий, приводящих к снижению демпферных свойств и даже к потере устойчивости при повреждениях в ЭЭС. Аппаратные ограничительные меры должны исключить возможность задания отрицательных или сильно завышенных значений T_p; при этом максимально возможное значение постоянной времени прогнозирующего звена АРС должно определяться с учетом индивидуальных особенностей и конкретных условий работы установок РГ.

СПИСОК ЛИТЕРАТУРЫ

1. *Barker Ph. P., Mello R.W.* Determining the impact of distributed generation on power systems. Pt. 1. Radial distribution systems // 2000 Power Engineering Society Summer Meeting, July 11–15. – Seattle, WA, USA, 2000. – P. 1645–1656. – DOI: 10.1109/PESS.2000.868775.

2. *Rugthaicharoencheep N., Auchariyamet S.* Technical and economic impacts of distributed generation on distribution system // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. – 2012. – Vol. 6. – P. 385–389. – DOI: 10.5281/zenodo.1327636.

3. *Voropai N.I., Stychinsky Z.A.* Renewable energy sources: theoretical foundations, technologies, technical characteristics, economics. – Magdeburg: Otto-von-Guericke-Universität, 2010. – 223 p.

4. *Magdi S.M., AL-Sunni F.M.* Control and optimization of distributed generation systems. – Cham: Springer International Publishing, 2015. – 578 p.

5. *Martínez Ceseña E.A., Capuder T., Mancarella P.* Flexible distributed multienergy generation system expansion planning under uncertainty // IEEE Transaction on Smart Grid. – 2016. – Vol. 7. – P. 348–357. – DOI: 10.1109/TSG.2015.2411392.

6. *Sikorski T., Rezmer J.* Distributed generation and its impact on power quality in low-voltage distribution networks // Power quality issues in distributed generation / ed. by Dr. J. Luszcz. – Rijeka: InTech, 2015. – DOI: 10.5772/61172.

7. *Раскин Е.М., Денисова Л.А., Мещеряков В.А.* Автоматизация проектирования системы регулирования с использованием генетического алгоритма оптимизации // Промышленные АСУ и контроллеры. – 2012. – № 7. – С. 8–14.

8. *Saad M.S., Jamaluddin H., Darus I.Z.M.* Implementations of PID-controller tuning using differential evolution and genetic algorithm // International Journal of Innovate Computing, Information and Control. – 2012. – Vol. 8, N 11. – P. 7761–7779.

9. *Chen Y., Ma Y., Yun W.* Application of improved genetic algorithm in PID controller parameters optimization // Telkomnika. – 2013. – Vol. 11, N 3. – P. 1524–1530.

10. *Bulatov Yu.N., Kryukov A.V.* Optimization of automatic regulator settings of the distributed generation plants on the basis of genetic algorithm // 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – Chelyabinsk, Russia, 2016. – P. 1–6. – DOI: 10.1109/ICIEAM.2016.7911456.

11. *Camacho E.F., Bordons C.* Model predictive control. – 2nd ed. – Springer, 2007. – 405 p.

12. *Пикина Г.А.* Принцип управления по прогнозу и возможность настройки систем регулирования одним параметром // Новое в российской электроэнергетике. – 2014. – № 3. – С. 5–13.

13. *Булатов Ю.Н., Крюков А.В., Нгуен В.Х.* Цифровые прогностические регуляторы установок распределенной генерации // Главный энергетик. – 2020. – № 1. – С. 14–21.

14. *Bulatov Yu.N., Kryukov A.V., Nguyen V.H.* Simulation of gas turbine power plants with voltage and speed prognostic regulators // International Russian Automation Conference (RusAutoCon). – Sochi, Russia, 2020. – P. 160–164. – DOI: 10.1109/RusAutoCon49822.2020.9208114.

15. *Bulatov Yu.N., Kryukov A.V., Nguyen V.H.* Automatic prognostic regulators of distributed generators // 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarFastCon). – Vladivostok, Russia, 2018. – P. 1–4. – DOI: 10.1109/FarEastCon.2018.8602718.

16. *Пикина Г.А., Кузнецов М.С.* Методы настройки прогностических типовых алгоритмов регулирования // Теплоэнергетика. – 2012. – № 2. – С. 64–68.

17. *Бушуев В.В, Лизалек Н.Н., Новиков Н.Л.* Динамические свойства энергосистем. – М.: Энергоатомиздат, 1995. – 320 с.

18. *Anderson P.M, Fouad A.A.* Power system control and stability. – 2nd ed. – IEEE Press, 2003. – 688 p.

Булатов Юрий Николаевич, кандидат технических наук, доцент, заведующий кафедрой электроэнергетики и электротехники Братского государственного университета. Основное направление научных исследований – моделирование и управление режимами электроэнергетических систем. Имеет более 150 публикаций. E-mail: bulatovyura@yandex.ru.

Крюков Андрей Васильевич, доктор технических наук, профессор Иркутского государственного университета путей сообщения, профессор Иркутского национального исследовательского технического университета. Основное направление научных исследований – моделирование и управление режимами электроэнергетических систем и систем электроснабжения железных дорог. Имеет более 600 публикаций. E-mail: and_kryukov@mail.ru.

Bulatov Yuri N., PhD (Eng.), associate professor, head of the department of electric power and electrical engineering of Bratsk State University. The main field of his scientific research is modeling and control of the modes of electric power systems. He has over 150 publications. E-mail: bulatovyura@yandex.ru.

Kryukov Andrey V., D.Sc. (Eng.), professor, Irkutsk State Transport University, Irkutsk National Research Technical University. The main field of his scientific research is modeling and control of the modes of electric power systems and railway power supply systems. He has over 600 publications. E-mail: and_kryukov@mail.ru.

DOI: 10.17212/2782-2001-2021-2-19-34

Study of cyber security of predictive control algorithms for distributed generation plants[*]

Yu.N. BULATOV[1,a], *A.V. KRYUKOV*[2,3,b]

[1] *Bratsk State University, 40 Makarenko Street, Bratsk, 665709, Russian Federation*
[2] *Irkutsk State Transport University, 15 Chernishevskiy Street, Irkutsk, 664074, Russian Federation*
[3] *Irkutsk National Research Technical University, 83 Lermontov Street, Irkutsk, 664074, Russian Federation*

[a] *bulatovyura@yandex.ru* [b] *and_kryukov@mail.ru*

Abstract

The power industry is currently actively developing the field related to the use of distributed generation plants located near the power receiving devices of consumers. At the same time, the introduction of distributed generation plants causes a lot of engineering problems which need solutions. One of them is the optimization of the settings of automatic voltage regulators (AVR) and speed regulators (ASR) of synchronous generators in all possible operating modes. This requires the use of complex models of power supply systems, distributed generation plants and their regulators, as well as labor-intensive calculations that take into account a large number of interrelated parameters. However, there is another approach based on the use of predictive controllers. In this case only one parameter is needed for linear predictive models.

The article describes a method for constructing and tuning the proposed predictive ASR synchronous generator, as well as computer models of distributed generation plants used in research. The purpose of the research was to determine cyber security of power supply systems equipped with various distributed generation plants with predictive speed controllers that can be implemented on the basis of the microprocessor technology. The studies were carried out in the MATLAB system using the Simulink and SymPowerSystems simulation packages on computer models of distributed generation plants with one turbine generator operating at a dedicated load, as well as a group of hydrogenerators connected to a high-power electric power system. The simulation results showed the effectiveness of the proposed predictive control algorithms, as well as the fact that their cyber security can be increased by introducing hardware restrictions on the range of changes in the time constant of the predictive link.

Keywords: distributed generation plants; synchronous generator; predictive control algorithms; cyber security; automatic voltage regulator; automatic speed regulator; linear predictive link; power supply system; computer modeling

REFERENCES

1. Barker Ph. P., Mello R.W. Determining the impact of distributed generation on power systems. Pt. 1. Radial distribution systems. *2000 Power Engineering Society Summer Meeting*, July 11–15, Seattle, WA, USA, 2000, pp. 1645–1656. DOI: 10.1109/PESS.2000.868775.

[*] *Received 26 January 2021.*

The research was carried out with financial support from a grant from the state assignment of the Ministry of Science and Higher Education of Russia (project No. 0667-2020-0039).

2. Rugthaicharoencheep N., Auchariyamet S. Technical and economic impacts of distributed generation on distribution system. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, 2012, vol. 6, pp. 385–389. DOI: 10.5281/zenodo.1327636.

3. Voropai N.I., Stychinsky Z.A. *Renewable energy sources: theoretical foundations, technologies, technical characteristics, economics*. Magdeburg, Otto-von-Guericke-Universität, 2010. 223 p.

4. Magdi S.M., AL-Sunni F.M. *Control and optimization of distributed generation systems*. Cham, Springer International Publishing, 2015. 578 p.

5. Martínez Ceseña E.A., Capuder T., Mancarella P. Flexible distributed multienergy generation system expansion planning under uncertainty. *IEEE Transaction on Smart Grid*, 2016, vol. 7, pp. 348–357. DOI: 10.1109/TSG.2015.2411392.

6. Sikorski T., Rezmer J. Distributed generation and its impact on power quality in low-voltage distribution networks. *Power quality issues in distributed generation*. Ed. by Dr. J. Luszcz. Rijeka, InTech, 2015. DOI: 10.5772/61172.

7. Raskin E.M., Denisova L.A., Meshcheryakov V.A. Avtomatizatsiya proektirovaniya sistemy re-gulirovaniya s ispol'zovaniem geneticheskogo algoritma optimizatsii [Automation of the design of the control system using the genetic optimization algorithm]. *Promyshlennye ASU i kontrollery = Industrial controllers ACS*, 2012, no. 7, pp. 8–14.

8. Saad M.S., Jamaluddin H., Darus I.Z.M. Implementations of PID-controller tuning using differential evolution and genetic algorithm. *International Journal of Innovate Computing, Information and Control*, 2012, vol. 8, no. 11, pp. 7761–7779.

9. Chen Y., Ma Y., Yun W. Application of improved genetic algorithm in PID controller parameters optimization. *Telkomnika*, 2013, vol. 11, no. 3, pp. 1524–1530.

10. Bulatov Yu.N., Kryukov A.V. Optimization of automatic regulator settings of the distributed generation plants on the basis of genetic algorithm. *2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, Chelyabinsk, Russia, 2016, pp. 1–6. DOI: 10.1109/ICIEAM.2016.7911456.

11. Camacho E.F., Bordons C. *Model predictive control*. 2nd ed. Springer, 2007. 405 p.

12. Pikina G.A. Printsip upravleniya po prognozu i vozmozhnost' nastroiki sistem regulirovaniya odnim parametrom [The principle of predictive control and the ability to configure control systems with one parameter]. *Novoe v rossiiskoi elektroenergetike = New in the Russian Electrical Power-Engineering*, 2014, no. 3, pp. 5–13.

13. Bulatov Yu.N., Kryukov A.V., Nguyen V.H. Tsifrovye prognosticheskie regulyatory ustanovok raspredelennoi generatsii [Digital prognostic regulators for distributed generation]. *Glavnyi energetik = Chief Power Engineer*, 2020, no. 1, pp. 14–21.

14. Bulatov Yu.N., Kryukov A.V., Nguyen V.H. Simulation of gas turbine power plants with voltage and speed prognostic regulators. *International Russian Automation Conference (RusAutoCon)*, Sochi, Russia, 2020, pp. 160–164. DOI: 10.1109/RusAutoCon49822.2020.9208114.

15. Bulatov Yu.N., Kryukov A.V., Nguyen V.H. Automatic prognostic regulators of distributed generators. *2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarFastCon)*, Vladivostok, Russia, 2018, pp. 1–4. DOI: 10.1109/FarEastCon.2018.8602718.

16. Pikina G.A., Kuznetsov M.S. Metody nastroiki prognosticheskikh tipovykh algoritmov regulirovaniya [Methods for setting predictive standard control algorithms]. *Teploenergetika = Thermal Engineering*, 2012, no. 2, pp. 64–68. (In Russian).

17. Bushuev V.V, Lizalek N.N., Novikov N.L. *Dinamicheskie svoistva energosistem* [Dynamic properties of power systems]. Moscow, Energoatomizdat, 1995. 320 p.

18. Anderson P.M, Fouad A.A. *Power system control and stability*. 2nd ed. IEEE Press, 2003. 688 p.

Для цитирования:

Булатов Ю.Н., Крюков А.В. Исследование кибербезопасности прогностических алгоритмов управления для установок распределенной генерации // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 19–34. – DOI: 10.17212/2782-2001-2021-2-19-34.

For citation:

Bulatov Yu.N., Kryukov A.V. Issledovanie kiberbezopasnosti prognosticheskikh algoritmov upravleniya dlya ustanovok raspredelennoi generatsii [Study of cyber security of predictive control algorithms for distributed generation plants]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 19–34. DOI: 10.17212/2782-2001-2021-2-19-34.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 35–46

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 35–46

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 004.051

DOI: 10.17212/2782-2001-2021-2-35-46

Структура и особенности программного комплекса для геофизических геометрических 3D-инверсий*

Д.В. ВАГИН

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

vagin@corp.nstu.ru

Рассматриваются структура и особенности программного комплекса для 3D-инверсий геофизических данных. Представленный программный комплекс ориентирован на решение прямых и обратных задач электроразведки и инженерной геофизики. Помимо параметров, определяющих физические свойства среды, программный комплекс позволяет восстанавливать и геометрические параметры геофизической модели: рельефы слоев и границы трехмерных включений. Включения могут иметь форму произвольных шестигранников или призм с основанием в виде многоугольника. Программный комплекс состоит из четырех основных подсистем: графического интерфейса, подсистем решения прямых и обратных задач и клиент-серверной части для выполнения расчетов на удаленных вычислительных узлах. Графический интерфейс состоит из ориентированных на специалиста-геофизика модулей пре- и постпроцессора, позволяющих описать задачу и представить результаты ее решения в удобных для пользователя терминах. Для решения прямых задач используется метод конечных элементов и технология разделения поля на нормальную и аномальную составляющие. При этом используются специальные методы дискретизации расчетной области, позволяющие учесть как сложную трехмерную структуру среды, так и наличие в расчетной области техногенных объектов (скважин). Для увеличения эффективности решения прямых задач используются несогласованные сетки с ячейками в виде произвольных шестигранников. Для геометрических параметров описаны способы эффективного вычисления производных (по этим параметрам), необходимых для решения обратных задач методом Гаусса–Ньютона. Главной идеей для эффективного вычисления производных является выделение влияния изменения значения параметра (используемого для вычисления значения обобщенной производной) на задачу. Описаны основные действия, выполняемые подсистемой решения обратных задач, и особенности, связанные с обработкой геометрических параметров.

---

* *Статья получена 10 декабря 2020 г.*

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ (НИЛ «Моделирование и обработка данных наукоемких технологий», проект № FSUN-2020-0012).

Ключевые слова: программные комплексы, 3D-инверсия, геометрические параметры, метод Гаусса–Ньютона, адаптивная регуляризация, вычислительная эффективность, обратные задачи геофизики, апробация на практических данных

ВВЕДЕНИЕ

Для решения трехмерных обратных задач геофизики различные исследователи пытаются применять самые разные методы и подходы. Их можно разделить на традиционные, основанные на классических методах минимизации функционалов [1–6], и новые, основанные, например, на использовании нейронных сетей [7]. Всплеск интереса к использованию нейронных сетей обусловлен возросшими вычислительными возможностями и доступностью многопроцессорных систем. Тем не менее пока неизвестно об успешном применении данных методов для решения реальных практических задач.

Таким образом, при выполнении инверсий на практике используются традиционные подходы, в которых исследуемая область обычно разбивается на непересекающиеся подобласти фиксированной геометрии, плотно заполняющие всю исследуемую область. Такие подобласти называют вокселями (voxels) или ячейками, а соответствующие решения обратных задач – воксель-инверсиями (voxel inversion) [3]. Результатом инверсии в этом случае является набор значений восстанавливаемых физических величин (электрическая проводимость, коэффициент поляризации, концентрация некоторого вещества и др.) в этих ячейках. Данный подход очень прост в реализации, но зачастую не позволяет четко определить границы подобластей с разными свойствами. Примером программного комплекса для решения обратных задач на основе воксель-инверсий является MTZ3D [Gribenko A.V., Zhdanov M.S.].

Очевидным решением этой проблемы является параметризация не только физических свойств, но и геометрии исследуемой области (parametric inversion) [3]. Проведенные инверсии показывают, что геометрическая параметризация действительно позволяет находить границы целевых объектов [6]. Это связано с тем, что выбранная параметризация геометрии исследуемой области уже является определенной регуляризацией обратной задачи. Поэтому одной из основных проблем решения обратных задач на основе геометрической параметризации является выбор конкретных геометрических параметров. Очевидный в одномерном случае (например, толщины слоев), уже в двумерном случае он требует достаточного профессионализма от выполняющего данную инверсию специалиста. В трехмерном случае эта задача становится еще более сложной, и при ее решении просто невозможно обойтись без компромиссов между удобством работы с программным комплексом человека, ограничениями, накладываемыми на решение выбранной параметризацией, сложностью программной реализации и ресурсоемкостью реализуемых процедур.

1. ПОСТАНОВКА ЗАДАЧИ

Чтобы программный комплекс, позволяющий выполнять 3D-инверсии на основе геометрической параметризации, мог успешно использоваться для выполнения практических интерпретаций, он должен удовлетворять нескольким требованиям. Первое требование – это наличие удобного и гибкого ин-

терфейса пользователя, позволяющего определять, какие элементы геометрии трехмерной модели будут являться искомыми параметрами обратной задачи. Второе требование – это контроль за тем, чтобы получаемые значения формальных геометрических параметров не приводили к противоречивой физической модели среды. Третьим важным требованием является гибкая и при этом автоматическая регуляризация для восстанавливаемых значений геометрических параметров.

Помимо указанных выше требований, принципиальную практическую применимость программного комплекса 3D-инверсий определяет общее время, затрачиваемое на решение обратной задачи. То есть программный комплекс должен эффективно понижать целевой функционал невязки экспериментальных и подобранных данных. И использование геометрической параметризации не должно значимым образом снижать эту эффективность.

Для того чтобы программный комплекс был удобен для практического применения, он должен описывать трехмерную модель среды в привычных для специалистов-геофизиков терминах. К ним можно отнести «слои» – большие пласты породы (как правило, осадочного происхождения), «объекты» – локальные подобласти со свойствами, отличными от свойств слоев, вмещающих эти подобласти, «рельефы» – трехмерные искривления слоев относительно горизонтальной плоскости, «контуры» – замкнутые границы проекций объектов сложной формы.

2. СТРУКТУРА ПРОГРАММНОГО КОМПЛЕКСА

Программный комплекс, реализующий геометрические 3D-инверсии, состоит из четырех основных подсистем.

1. *Интерфейс пользователя* (ИП). Включает в себя пре- и постпроцессор. Препроцессор позволяет в удобных для специалиста-геофизика терминах задавать модель изучаемой среды и указывать, какие скалярные и геометрические параметры этой модели являются искомыми для обратной задачи. Помимо модели среды, в препроцессоре также задаются тип задачи и связанные с ним параметры генераторных и приемных установок. Соответствующие данные препроцессор выгружает в виде файлов, которые идут на вход подсистем решения прямых и обратных задач. Препроцессор также позволяет загружать и редактировать конфигурационные файлы, содержащие параметры для других подсистем программного комплекса. Постпроцессор, в свою очередь, позволяет загружать результаты решения прямых и обратных задач и отображать их в виде графиков значений измеряемых величин или координаты или в виде распределения по площади в выбранной цветовой шкале. Часть окна препроцессора показана на рис. 1.

2. *Подсистема решения прямых задач* (ПРПЗ). Включает в себя набор исполняемых программных модулей, позволяющих решать прямые задачи на основе данных от препроцессора или подсистемы решения обратных задач. Для управления этой подсистемой используются соответствующие конфигурационные файлы, определяющие методы, используемые для решения прямых задач, число параллельных потоков и требуемую точность получаемых результатов. Модули подсистемы реализуют высокоэффективные методы моделирования прямых трехмерных задач [8–11].

3. *Подсистема решения обратных задач* (ПРОЗ). Включает в себя несколько программных модулей. Ядром данной подсистемы является программный модуль, реализующий метод Гаусса–Ньютона (МГН) для минимизации квадратичного функционала невязки экспериментальных и подобранных данных. Другими важными модулями данной подсистемы являются: модуль подготовки прямых задач (МППЗ), переводящий текущие значения искомых в обратной задаче параметров в данные (файлы) модели, идущие на вход подсистемы решения прямых задач; модуль генерации моделей для расчета производных по параметрам (МГМРПП), в зависимости от способа расчета производной для конкретного параметра может использовать либо модуль МППЗ, либо собственные процедуры; модуль адаптивной регуляризации (МАР), контролирующий, чтобы найденные приращения не выводили значения параметров за указанные локальные (на данном шаге инверсии) или глобальные ограничения.

4. *Клиент-серверная часть* (КСЧ). Включает в себя два исполняемых модуля: модуль клиента – подготавливает и отправляет данные (файлы) на удаленный вычислительный узел, и модуль сервера – передает полученные от клиента данные подсистеме решения прямых задач и после окончания вычислений возвращает клиенту полученные результаты. Сервер следит за работой ПРПЗ и в случае возникновения ошибки передает данные аварийного завершения клиенту. Клиент, в свою очередь, периодически опрашивает серверные модули на вычислительных узлах, на которые были отправлены задачи. Если серверный модуль на удаленном узле не отвечает в течение указанного временного интервала, то соответствующая задача отправляется на другой доступный вычислительный узел. При отправке задач на удаленные узлы клиент учитывает вычислительную мощность доступных узлов (количество вычислительных ядер и объем оперативной памяти) [12, 13].

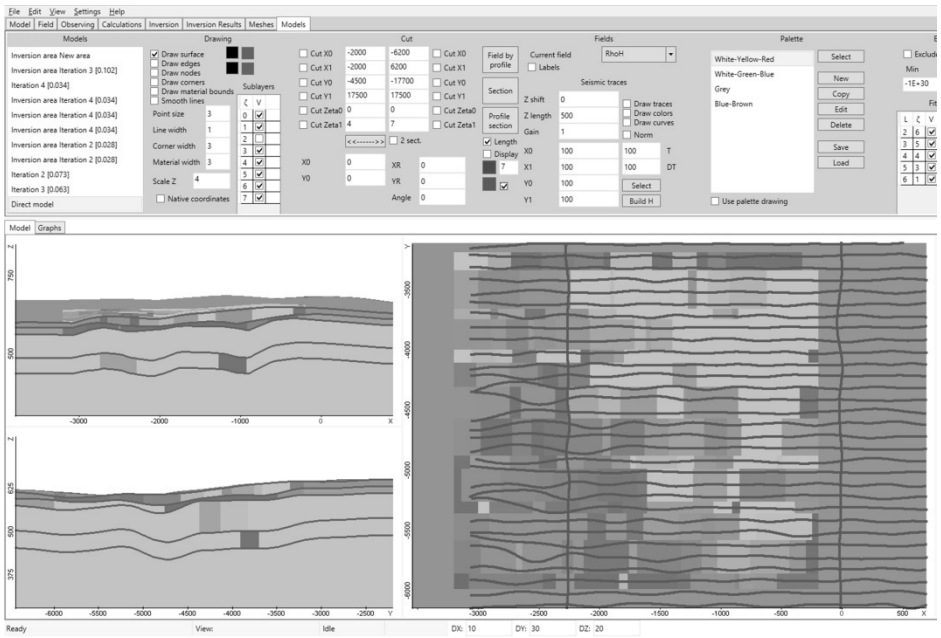


Рис. 1. Часть окна препроцессора

Fig. 1. Part of the preprocessor window

Общая структура описываемого программного комплекса представлена на рис. 2 (ВУ – вычислительный узел). Подсистемы ПРПЗ и ПРОЗ рассмотрим более подробно, так как именно от особенностей их реализации зависит вычислительная эффективность программного комплекса.

Основные вычислительные затраты в методе Гаусса–Ньютона связаны с необходимостью вычисления производных измеряемых величин по искомым параметрам. Для сложной трехмерной среды эти производные можно посчитать лишь численно:

$$\frac{df(p)}{dp_i} \approx \frac{\Delta f(p)_i}{\Delta p_i} = \frac{f(p+\Delta p_i) - f(p)}{\Delta p_i}, \quad i = 1 \ldots M. \tag{1}$$

Здесь M – это общее число искомых параметров; p – вектор значений искомых параметров, p_i – i-я компонента вектора p, а запись $p + \Delta p_i$ (вектор плюс скаляр) означает изменение на Δp_i i-й компоненты вектора p. Вычисление $f(p)$ означает решение прямой трехмерной задачи для модели, соответствующей вектору значений искомых параметров p.

Если моделируемый процесс нестационарный или нелинейный, то решение всех задач $f(p + \Delta p_i)$ может занимать много времени. Для эффективного расчета $\Delta f(p)_i$ подсистема ПРПЗ разделяет параметры на две категории.

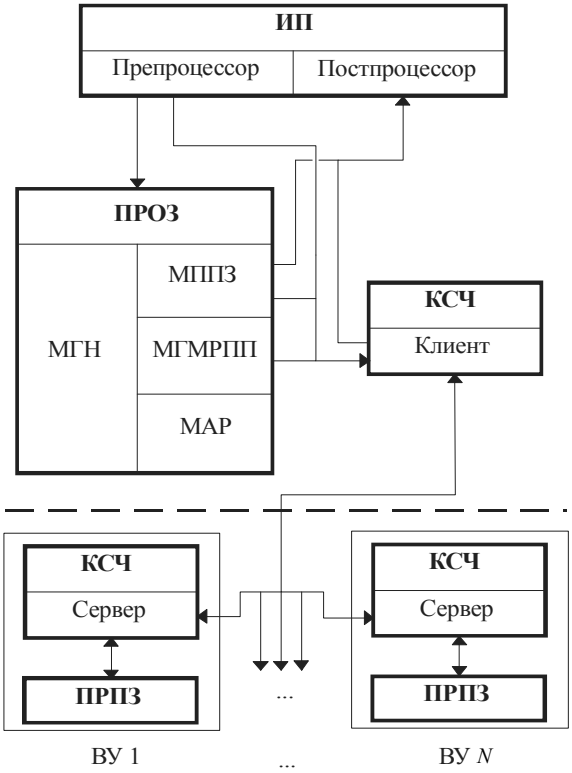


Рис. 2. Общая структура программного комплекса.
Стрелками показано направление движения данных

Fig. 2. General structure of the software package.
The direction of data flow is shown by arrows

К первой категории относятся параметры, для которых значение $\Delta f(p)_i$ с достаточной точностью можно вычислить путем решения специальной задачи. Такими параметрами являются, например, границы объектов. Приращение Δp_i в этом случае означает изменение сеточного оператора. С учетом того, что Δp_i, как правило, невелико (иначе будет плохой аппроксимация производной в (1)), изменение сеточного оператора может существенным образом сказаться на точности значения $\Delta f(p)_i$, вычисленного как разность задач $f(p+\Delta p_i)$ и $f(p)$.

Для конкретности предположим, что решается нестационарная электромагнитная задача в немагнитной среде в области времен, на которых токи смещения несущественны. Тогда описывающее задачу $f(p)$ уравнение имеет вид

$$\frac{1}{\mu_0}\operatorname{rot}\operatorname{rot}\vec{A}_p+\sigma_p\frac{d\vec{A}_p}{dt}=\vec{J}\,. \tag{2}$$

Здесь $\vec{A}_p$ – это векторный электромагнитный потенциал; μ_0 – магнитная проницаемость вакуума; $\vec{J}$ – сторонний ток (связанный с источником ЭМ поля), а функция σ_p описывает распределение удельной электрической проводимости.

В этом случае искомое значение $\Delta f(p)_i$ может быть найдено из решения задачи с уравнением

$$\frac{1}{\mu_0}\operatorname{rot}\operatorname{rot}\vec{A}_\Delta+\sigma_{\tilde{p}}\frac{d\vec{A}_\Delta}{dt}=(\sigma_p-\sigma_{\tilde{p}})\frac{d\vec{A}_p}{dt}\,, \tag{3}$$

где $\tilde{p}=p+\Delta p_i$, $\vec{A}_\Delta=\vec{A}_{\tilde{p}}-\vec{A}_p$. Так как $\Delta f(p)_i$ находится напрямую из $\vec{A}_\Delta$, то на него уже не оказывает влияние различие сеточных операторов для задач $f(\tilde{p})$ и $f(p)$. К тому же в задаче для уравнения (3) нет сторонних токов $\vec{J}$, что существенно понижает требования к подробности дискретизации расчетной области. Более того, правая часть (3) отлична от нуля только в небольшой (так как Δp_i невелико) подобласти. Таким образом, хотя задача (3) и остается трехмерной, но за счет пониженных требований к ее дискретизации она может быть решена с малыми вычислительными затратами.

Ко второй категории относятся параметры, для которых сложно сформулировать или решить задачу, подобную (3). Такими параметрами являются, например, параметры, определяющие геометрию границ сложных объектов. Например, это объекты, границы которых по латерали (в плане) представляются в виде некоторого многоугольника. В этом случае сетка для задач $f(p)$ и $f(\tilde{p})$ может быть построена без учета границ, определяемых соответствующими геометрическими параметрами [14]. Значения данных границ при этом используются для вычисления эффективных значений физических характеристик соответствующих ячеек сетки. В этом случае для вычисления $f(\tilde{p})-f(p)$ эти задачи могут решаться при весьма грубой дискретизации,

так как порождаемая этой дискретизацией систематическая ошибка в большой мере взаимовычитается.

Описание методов, используемых подсистемой ПРПЗ для решения различных задач, приведено в работах [8–11, 14].

Теперь рассмотрим подробнее подсистему ПРОЗ. Ее принципиальная схема представлена на рис. 3. Как видно из приведенной схемы, подсистема ПРОЗ выполняет девять основных действий.

1. *Фиксация части параметров*. Для улучшения сходимости инверсии на каждой ее итерации значения некоторых параметров могут быть зафиксированы. К таким параметрам можно отнести, например, границы слабоконтрастного к вмещающей среде объекта. То есть это параметры, значения производных по которым близки к нулю. Другой пример фиксации параметров – это выполнение специальной итерации коррекции значений параметров, определяющих физические свойства объектов и слоев, если в инверсии для этих же объектов и слоев присутствуют и геометрические параметры [15–18].

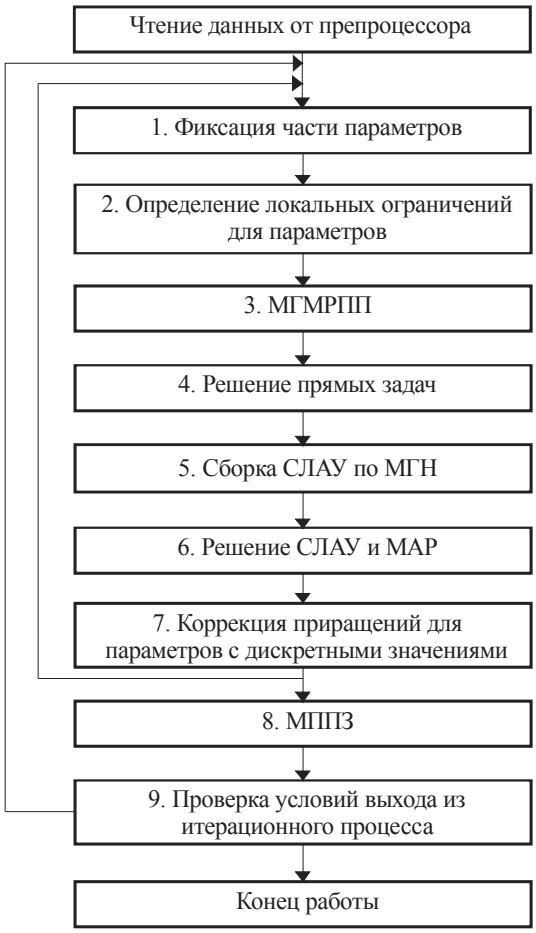


Рис. 3. Принципиальная схема подсистемы решения обратных задач

Fig. 3. Schematic diagram of the subsystem for solving inverse problems

2. *Определение локальных ограничений для параметров*. Помимо глобальных ограничений на возможные значения параметров, связанных с их физическим смыслом, на каждой итерации инверсии применяются дополнительные, более жесткие ограничения. Одной из причин наложения дополнительных ограничений является нелинейность исходного функционала. Более жесткие ограничения уменьшают получаемые на данной итерации значения приращений параметров, что делает сходимость итерационного процесса более равномерной, а получаемое в итоге решение – более регулярным [15–18]. Другой причиной дополнительных ограничений являются особенности используемой параметризации. Например, это требование выпуклости или сохранения подобия контуров, задающих сложные геометрические объекты.

3. *МГМРПП*. Сущность это модуля в достаточной мере раскрыта выше – в описании общей структуры комплекса.

4. *Решение прямых задач*. Это обращение к подсистеме ПРПЗ для получения решения прямой задачи для данного набора параметров и расчета значений производных по параметрам. При этом если в задаче есть удаленные друг от друга приемно-генераторные установки, то их наборы параметров, для которых в данных установках будут вычисляться производные по параметрам, могут не совпадать. Это связанно с тем, что влияние некоторых параметров может быть достаточно локальным и затрагивать только некоторые приемно-генераторные установки. В других установках его можно считать нулевым и не выполнять решение соответствующих прямых задач [6, 15–18].

5. *Сборка СЛАУ по МГН*. Матрица СЛАУ в методе Гаусса–Ньютона имеет плотный формат. Ее размер равен числу подбираемых параметров. При этом для каждого элемента матрицы суммируется множество слагаемых [6]. Для таких технологий, как, например, аэроразведка, число таких слагаемых может достигать миллиона. Поэтому данная процедура распараллеливается и может выполняться на вычислительных узлах.

6. *Решение СЛАУ и МАР*. Метод адаптивной регуляризации подразумевает подбор значений регуляризирующих параметров [6]. Увеличение значений регуляризирующих параметров приводит к уменьшенью получаемых приращений по искомым параметрам. Так как текущая модель удовлетворяет всем глобальным и локальным ограничением (что является обязательным требованием к стартовой модели), то этот процесс гарантированно конечен и на практике занимает порядка 10–30 итераций изменения параметров регуляризации и решения СЛАУ.

7. *Коррекция приращений для параметров с дискретными значениями*. Решение СЛАУ дает вещественные приращения к значениям параметров. Если же область определения параметра состоит из дискретного набора значений, то полученные вещественные значения необходимо округлить до ближайших значений из этого набора. После этих действий в случае наличия связанных физических и геометрических параметров для объектов или слоев можно зафиксировать соответствующие геометрические параметры и выполнить дополнительную итерацию коррекции значений параметров, определяющих физические свойства объектов и слоев (действия 1–6) [6].

8. *МППЗ*. Выполняет перевод значений подобранных параметров в файлы модели, которые могут быть загружены постпроцессором или использованы для решения прямой задачи.

9. *Проверка условий выхода из итерационного процесса*. Окончание итерационного процесса решения обратной задачи осуществляется при выполнении одного из двух условий: первое – достигнут требуемый уровень невязок экспериментальных и подобранных данных, второе – полученные приращения по параметрам близки к нулевым. Если инверсия останавливается по второму условию, то это означает близость полученной модели к оптимуму минимизируемого функционала в рамках заданной параметризации.

ЗАКЛЮЧЕНИЕ

В работе рассмотрена структура и особенности программного комплекса для 3D-инверсий геофизических данных. Главной особенностью данного программного комплекса является возможность использовать геометрическую параметризацию трехмерной модели среды. Основные сущности этой модели сформулированы в удобных для конечного пользователя (специалиста-геофизика) терминах и отражены в интерфейсе программного комплекса. Для решения обратных задач используется метод Гаусса–Ньютона. Для расчета производных по искомым параметрам используются специальные эффективные методы решения прямых задач. Эффективность решения обратных задач обеспечивается выбранными способами параметризации модели, использованием специальной регуляризации и применением дополнительной итерации коррекции значений негеометрических параметров.

Описанный в работе программный комплекс за счет используемых высокоэффективных подходов позволяет выполнять трехмерные геофизические инверсии реальных практических данных, что было подтверждено обработкой практических данных, полученных для различных технологий электроразведки в США, Восточной Сибири, на Дальнем Востоке, в Уральском регионе и в Казахстане, а также обработкой данных нефтедобычи в Республике Татарстан. Эффективность описанного в работе программного комплекса обусловлена использованием геометрической параметризации и высокоэффективными методами решения прямых задач.

СПИСОК ЛИТЕРАТУРЫ

1. A three-dimensional transient electromagnetic data inversion method based on a time-frequency transformation / Y. Yang, X. Wang, X. Liu, X. Mi, L. Mao // Applied Geophysics. – 2020. – Vol. 17, iss. 3. – P. 361–376.

2. *Um E., Kim J., Wilt M.* 3D borehole-to-surface and surface electromagnetic modeling and inversion in the presence of steel infrastructure // Geophysics. – 2020. – Vol. 85, iss. 5. – P. E139–E152.

3. 3D electromagnetic modelling and inversion: a case for open source / D.W. Oldenburg, L.J. Heagy, S. Kang, R. Cockett // Exploration Geophysics. – 2020. – Vol. 51, iss. 1. – P. 25–37.

4. *Tseng H., Lee K., Becker A.* 3D interpretation of electromagnetic data using a modified extended Born approximation // Geophysics. – 2003. – Vol. 68, iss. 1. – P. 127–137.

5. *Zhang Z.* 3D resistivity mapping of airborne EM data // Geophysics. – 2003. – Vol. 68, iss. 6. – P. 1896–1905.

6. Three-dimensional inversion of airborne data with applications for detecting elongated subvertical bodies overlapped by an inhomogeneous conductive layer with topography / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, D.S. Kiselev, A.M. Grif, Y.I. Koshkina, A.P. Sivenkova // Geophysical Prospecting. – 2020. – Vol. 68, iss. 7. – P. 2217–2253. – DOI: 10.1111/1365-2478.12979.

7. 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods / J. Guo, Y. Li, M.W. Jessell, J. Giraud, C. Li, L. Wu, F. Li, S. Liu // Computers and Geosciences. – 2021. – Vol. 149. – Art. 104701.

8. Finite element solution to 3-D airborne time-domain electromagnetic problems in complex geological media using non-conforming hexahedral meshes / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, D.S. Kiselev, Y.I. Koshkina // Journal of Applied Geophysics. – 2020. – Vol. 172. – Art. 103911. – DOI: 10.1016/j.jappgeo.2019.103911.

9. Finite-element solution to multidimensional multisource electromagnetic problems in the frequency domain using non-conforming meshes / Y.G. Soloveichik, M.G. Persova, P.A. Domnikov, Y.I. Koshkina, D.V. Vagin // Geophysical Journal International. – 2018. – Vol. 212, iss. 3. – P. 2159–2193. – DOI: 10.1093/gji/ggx530.

10. Numerical scheme for modelling the electromagnetic field in airborne electromagnetic survey taking into account follow currents in transmitter loop / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, Y.I. Koshkina, E.I. Simon // Актуальные проблемы электронного приборостроения (АПЭП-2018): труды 14 международной конференции, Новосибирск, 2–6 октября 2018 г. – Новосибирск, 2018. – Т. 1, ч. 4. – С. 216–221.

11. 3-D modelling of marine electromagnetic technologies taking into account induced polarization / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, Y.I. Koshkina, M.G. Tokareva // Актуальные проблемы электронного приборостроения (АПЭП-2018): труды 14 международной конференции, Новосибирск, 2–6 октября 2018 г. – Новосибирск, 2018. – Т. 1, ч. 4. – С. 222–225.

12. Parallel solution of 3D forward and inverse problems of airborne electromagnetic survey / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, D.S. Kiselev, O.S. Trubacheva, N.S. Kondratyeva // 14 International Forum on Strategic Technology (IFOST 2019). – Tomsk, Russia, 2019. – P. 193–196.

13. Intelligent scheduler for solution of forward and inverse geoelectrical problems / M.G. Persova, Y.G. Soloveichik, Y.I. Koshkina, O.S. Trubacheva, D.V. Vagin, P. Domnikov // 11 International Forum on Strategic Technology (IFOST 2016). – Novosibirsk, Russia, 2016. – Pt. 1. – P. 401–405. – DOI: 14.1109/IFOST.2016.7884139.

14. Oil production optimization based on the finite-element simulation of the multi-phase flow in porous media and inverse problem solution / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, A.M. Grif, I.I. Patrushev, A.S. Ovchinnikova // GeoBaikal 2020. – Irkutsk, Russia, 2020. – P. 1–6. – DOI: 10.3997/2214-4609.202052021.

15. An approach to the geometric 3D inversion of airborne EM data for detection and geometrization of local targets overlapped by laterally inhomogeneous layers / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, D.S. Kiselev, A.P. Sivenkova, Y.I. Koshkina // Engineering and Mining Geophysics 2020. – Perm, Russia, 2020. – P. 90. – DOI: 10.3997/2214-4609.202051090.

16. The approaches to geometric EM 3-D inversion with consideration for IP effect in survey with electric lines / M.G. Persova, Y.G. Soloveichik, O.S. Trubacheva, D.V. Vagin, D.S. Kiselev, A.P. Sivenkova, Y.I. Koshkina // GeoBaikal 2020. – Irkutsk, Russia, 2020. – P. 1–6. – DOI: 10.3997/2214-4609.202052022.

17. Approach to multidimensional geometric inversion of data obtained by multi-spacing TEM soundings / M.G. Persova, Y.G. Soloveichik, D.V. Vagin, Y.I. Koshkina // GeoBaikal 2016. – Irkutsk, Russia, 2016. – P. 202–206. – DOI: 10.3997/2214-4609.201601697.

18. Geometrical nonlinear 3D inversion of airborne time domain EM data / M.G. Persova, Y.G. Soloveichik, Y.I. Koshkina, D.V. Vagin, O.S. Trubacheva // Near Surface Geoscience 2016 – First Conference on Geophysics for Mineral Exploration and Mining. Extended abstract. – Barcelona, 2016. – DOI: 10.3997/2214-4609.201602114.

Вагин Денис Владимирович, кандидат технических наук, доцент кафедры прикладной математики Новосибирского государственного технического университета. Основное направление научных исследований – численное моделирование задач, описываемых уравнениями математической физики. Имеет более 80 печатных работ и учебных пособий. E-mail: vagin@corp.nstu.ru

Vagin Denis V., PhD (Eng.), associate professor, department of applied mathematics, NSTU. His research interests include numerical modeling of problems described by mathematical physics equations. He has more than 80 publications and teaching manuals. E-mail: vagin@corp.nstu.ru

DOI: 10.17212/2782-2001-2021-2-35-46

The structure and features of the software for geophysical geometrical 3D inversions[*]

D. VAGIN

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

vagin@corp.nstu.ru

Abstract

The structure and features of a software package for 3D inversion of geophysical data are considered. The presented software package is focused on solving direct and inverse problems of electrical exploration and engineering geophysics. In addition to the parameters that determine physical properties of the medium, the software package allows you to restore the geometry parameters of the geophysical model, namely layer reliefs and boundaries of three-dimensional inclusions. The inclusions can be in the form of arbitrary hexagons or prisms with a polygonal base. The software package consists of four main subsystems: an interface, subsystems for solving direct and inverse problems, and a client-server part for performing calculations on remote computing nodes. The graphical interface consists of geophysicist-oriented pre- and postprocessor modules that allow you to describe the problem and present the results of its solution in user-friendly terms. To solve direct problems, the finite element method and the technology for dividing the field into normal and anomalous components are used. At the same time, special methods of discretization of the computational domain are used, which make it possible to take into account both the complex three-dimensional structure of the environment and the presence of man-made objects (wells) in the computational domain. To increase the efficiency of solving direct problems, nonconforming grids with cells in the form of arbitrary hexahedrons are used. Methods for efficient calculation of derivatives (with respect to these parameters) necessary for solving inverse problems by the Gauss-Newton method are also described for the geometry parameters. The main idea for efficient derivatives computation is to identify the effect of changing the value of the parameter (used to compute the value of the generalized derivative) on the problem. The main actions performed by the subsystem for solving inverse problems and the features associated with the processing of geometry parameters are described.

Keywords: geophysics software, parametric inversion, 3D inversion, Gauss-Newton method, adaptive regularization, computational efficiency, inverse problems of geophysics, approbation on practical data

REFERENCES

1. Yang Y., Wang X., Liu X., Mi X., Mao L. A three-dimensional transient electromagnetic data inversion method based on a time-frequency transformation. *Applied Geophysics*, 2020, vol. 17, iss. 3, pp. 361–376.

2. Um E., Kim J., Wilt M. 3D borehole-to-surface and surface electromagnetic modeling and inversion in the presence of steel infrastructure. *Geophysics*, 2020, vol. 85, iss. 5, pp. E139–E152.

3. Oldenburg D.W., Heagy L.J., Kang S., Cockett R. 3D electromagnetic modelling and inversion: a case for open source. *Exploration Geophysics*, 2020, vol. 51, iss. 1, pp. 25–37.

4. Tseng H., Lee K., Becker A. 3D interpretation of electromagnetic data using a modified extended Born approximation. *Geophysics*, 2003, vol. 68, iss. 1, pp. 127–137.

5. Zhang Z. 3D resistivity mapping of airborne EM data. *Geophysics*, 2003, vol. 68, iss. 6, pp. 1896–1905.

6. Persova M.G., Soloveichik Y.G., Vagin D.V., Kiselev D.S., Grif A.M., Koshkina Y.I., Sivenkova A.P. Three-dimensional inversion of airborne data with applications for detecting elongated subvertical bodies overlapped by an inhomogeneous conductive layer with topography. *Geophysical Prospecting*, 2020, vol. 68, iss. 7, pp. 2217–2253. DOI: 10.1111/1365-2478.12979.

[*] *Received 10 December 2020.*

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Research Laboratory "Modeling and data processing of high technologies", the project code is FSUN-2020-0012).

7. Guo J., Li Y., Jessell M.W., Giraud J., Li C., Wu L., Li F., Liu S. 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods. *Computers and Geosciences*, 2021, vol. 149, art. 104701.

8. Persova M.G., Soloveichik Y.G., Vagin D.V., Kiselev D.S., Koshkina Y.I. Finite element solution to 3-D airborne time-domain electromagnetic problems in complex geological media using non-conforming hexahedral meshes. *Journal of Applied Geophysics*, 2020, vol. 172, art. 103911. DOI: 10.1016/j.jappgeo.2019.103911.

9. Soloveichik Y.G., Persova M.G., Domnikov P.A., Koshkina Y.I., Vagin D.V. Finite-element solution to multidimensional multisource electromagnetic problems in the frequency domain using non-conforming meshes. *Geophysical Journal International*, 2018, vol. 212, iss. 3, pp. 2159–2193. DOI: 10.1093/gji/ggx530.

10. Persova M.G., Soloveichik Y.G., Vagin D.V., Koshkina Y.I., Simon E.I. Numerical scheme for modelling the electromagnetic field in airborne electromagnetic survey taking into account follow currents in transmitter loop. *Aktual'nye problemy elektronnogo priborostroeniya, APEP-2018: trudy XIV mezhdunarodnoi konferentsii* [14th International Conference on Actual Problems of Electronic Instrument Engineering, APEIE-2014: Proceedings], Novosibirsk, October 2–6, 2018, vol. 1, pt. 4, pp. 216–221.

11. Persova M.G., Soloveichik Y.G., Vagin D.V., Koshkina Y.I., Tokareva M.G. 3-D modelling of marine electromagnetic technologies taking into account induced polarization. *Aktual'nye problemy elektronnogo priborostroeniya, APEP-2018: trudy XIV mezhdunarodnoi konferentsii* [14th International Conference on Actual Problems of Electronic Instrument Engineering, APEIE-2014: Proceedings], Novosibirsk, October 2–6, 2018, vol. 1, pt. 4, pp. 222–225.

12. Persova M.G., Soloveichik Y.G., Vagin D.V., Kiselev D.S., Trubacheva O.S., Kondratyeva N.S. Parallel solution of 3D forward and inverse problems of airborne electromagnetic survey. *14 International Forum on Strategic Technology (IFOST 2019)*, Tomsk, Russia, 2019, pp. 193–196.

13. Persova M., Soloveichik Y., Koshkina Y., Trubacheva O., Vagin D., Domnikov P. Intelligent scheduler for solution of forward and inverse geoelectrical problems. *11 International Forum on Strategic Technology (IFOST 2016*), Novosibirsk, Russia, 2016, pt. 1, pp. 401–405. DOI: 14.1109/IFOST.2016.7884139.

14. Persova M.G., Soloveichik Y.G., Vagin D.V., Grif A.M., Patrushev I.I., Ovchinnikova A.S. Oil production optimization based on the finite-element simulation of the multi-phase flow in porous media and inverse problem solution. *GeoBaikal 2020*, Irkutsk, Russia, 2020, pp. 1–6. DOI: 10.3997/2214-4609.202052021.

15. Persova M.G., Soloveichik Y.G., Vagin D.V., Kiselev D.S., Sivenkova A.P., Koshkina Y.I. An approach to the geometric 3D inversion of airborne EM data for detection and geometrization of local targets overlapped by laterally inhomogeneous layers. *Engineering and Mining Geophysics 2020*, Perm, Russia, 2020, p. 90. DOI: 10.3997/2214-4609.202051090.

16. Persova M.G., Soloveichik Y.G., Trubacheva O.S., Vagin D.V., Kiselev D.S., Sivenkova A.P., Koshkina Y.I. The approaches to geometric EM 3-D inversion with consideration for IP effect in survey with electric lines. *GeoBaikal 2020*, Irkutsk, Russia, 2020, pp. 1–6. DOI: 10.3997/2214-4609.202052022.

17. Persova M.G., Soloveichik Y.G., Vagin D.V., Koshkina Y.I. Approach to multidimensional geometric inversion of data obtained by multi-spacing TEM soundings. *GeoBaikal 2016*, Irkutsk, Russia, 2016, pp. 202–206. DOI: 10.3997/2214-4609.201601697.

18. Persova M.G., Soloveichik Y.G., Koshkina Y.I., Vagin D.V., Trubacheva O.S. Geometrical nonlinear 3D inversion of airborne time domain EM data. *Near Surface Geoscience 2016 – First Conference on Geophysics for Mineral Exploration and Mining*. Extended abstract, Barcelona, 2016. DOI: 10.3997/2214-4609.201602114.

Для цитирования:

Вагин Д.В. Структура и особенности программного комплекса для геофизических геометрических 3D-инверсий // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 35–46. – DOI: 10.17212/2782-2001-2021-2-35-46.

For citation:

Vagin D.V. Struktura i osobennosti programmnogo kompleksa dlya geofizicheskikh geometricheskikh 3D-inversii [The structure and features of the software for geophysical geometrical 3D inversions]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 35–46. DOI: 10.17212/2782-2001-2021-2-35-46.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 47–66

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 47–66

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 519.24 DOI: 10.17212/2782-2001-2021-2-47-66

Проблемы применения непараметрических критериев согласия в задачах обработки результатов измерений*

Б.Ю. ЛЕМЕШКО[a], С.Б. ЛЕМЕШКО[b]

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

[a] *lemeshko@ami.nstu.ru* [b] *skyer@mail.ru*

Утверждается, что в основе некорректного применения непараметрических критериев согласия в различных приложениях в большинстве случаев лежат две причины.

Первая причина заключается в том, что при проверке сложных гипотез и оценивании параметров закона по анализируемой выборке используют классические результаты, связанные с проверкой простых гипотез. При проверке сложных гипотез на распределения статистик критериев согласия влияют вид наблюдаемого закона $F(x, \theta)$, соответствующий проверяемой гипотезе, тип и число оцениваемых параметров, метод оценивания, в некоторых случаях – значение параметра формы. В работе показано влияние всех названных факторов на распределения статистик критериев. Подчеркивается, что пренебрежение при проверке сложных гипотез факта потери критерием свойства «свободы от распределения» приводит к увеличению вероятности ошибок 2-го рода. Показывается, что распределение статистики критерия, необходимое для формирования вывода о результатах проверки сложной гипотезы, может находиться с использованием имитационного моделирования в интерактивном режиме непосредственно в процессе проверки.

Вторая причина связана с наличием ошибок округления, которые могут существенно изменять распределения статистик критериев. В работе показано, что асимптотическими результатами при проверке простых и сложных гипотез можно пользоваться при ошибках округления Δ много меньше среднеквадратического отклонения σ закона распределения ошибок измерения и объемах выборок n, не превышающих некоторых максимальных значений. При объемах выборок больших, чем эти максимальные значения, реальные распределения статистик критериев отклоняются от асимптотических в сторону больших значений статистик. В таких ситуациях использование асимптотических распределений для формирования вывода о результатах проверки приводит к увеличению вероятностей ошибок 1-го рода (к отклонению справедливой проверяемой гипотезы). Показано, что при соизмеримости ошибок округления и распределения статистик критериев отклоняются от асимптотических распределений и при малых n. А с ростом n ситуация только усугубляется. В работе изменения распределений статистик под влиянием округлений демонстрируются при проверке как простых, так и сложных гипотез. Показано, что единственным выходом, обеспечивающим корректность выводов по применяемым критериям в таких не-

* *Статья получена 05 февраля 2021 г.*

стандартных условиях, является использование реальных распределений статистик. Эта задача может решаться в интерактивном режиме (в процессе проверки) и опираться на компьютерные технологии исследования и аппарат математической статистики.

Ключевые слова: проверка гипотез, непараметрические критерии согласия, простая гипотеза, сложная гипотеза, распределение статистики, ошибки округления, достигнутый уровень значимости, ошибка 1-го рода, ошибка 2-го рода, имитационное моделирование

ВВЕДЕНИЕ

Статистические методы, ориентированные на обнаружение и исследование скрытых закономерностей, в том числе вероятностных, содержащихся в рядах измерений, связанных с процессами в различных системах и окружающей среде, занимают заметное место в задачах анализа и обработки данных.

С момента зарождения в математической статистике накоплен огромный аппарат методов статистического анализа, предназначенных для решения тех или иных задач. Особенно стремительно этот аппарат развивался последние 120 лет. Если проанализировать сотни тысяч публикаций, касающихся статистических методов, то можно увидеть, что львиная доля из них посвящена разработке математического аппарата и значительно меньшая часть уделена использованию этого аппарата в приложениях (причем не всегда корректно применяемого). Можно обратить внимание и на то, что в некоторых областях, где полезность применения определенных статистических методов кажется очевидной, этого не делается.

Причин такого состояния много. Во-первых, потенциальному пользователю сложно ориентироваться во всем многообразии методов, о существовании большей части из которых он просто не слышал. Трудно отдать предпочтение какому-то методу или критерию из ряда, предназначенного для решения одной и той же задачи, из-за отсутствия четких рекомендаций, учитывающих свойства сравниваемых методов. Перспективные методы или критерии и полезные результаты остаются неизвестными, просто теряются в растущем потоке публикаций. Во-вторых, университетский курс теории вероятностей и математической статистики, читаемый для нематематических специальностей, абсолютно не ориентирован на применение соответствующих методов в приложениях. Поэтому будущему специалисту приходится впоследствии добывать необходимые знания и опыт самостоятельно. В-третьих, корректность применения соответствующего метода или статистического критерия обусловливается некоторыми стандартными предположениями, которые в условиях конкретного приложения могут не выполняться. В нестандартных условиях свойства критериев, как правило, меняются, и меняются существенно. Использование классических результатов в таких ситуациях может приводить к некорректности статистических выводов. В то же время это не исключает возможности применения конкретного критерия в нестандартных условиях. Для этого надо лишь знать свойства критерия в этих условиях.

В настоящей статье мы продемонстрируем, как меняются распределения статистик непараметрических критериев согласия в зависимости от вида проверяемой гипотезы (простой или сложной) и величины ошибок округления, а также покажем, что использование вычислительных технологий и статистического моделирования позволяет с успехом применять эти критерии в изменившихся условиях, обеспечивая корректность выводов.

1. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ СОГЛАСИЯ ПРИ ПРОВЕРКЕ ПРОСТЫХ ГИПОТЕЗ

Изначально все непараметрические критерии согласия предлагались для проверки простых гипотез вида H_0: $F(x) = F(x, \theta)$, где $F(x, \theta)$ – функция распределения вероятностей, с которой проверяют согласие наблюдаемой (упорядоченной) выборки $x_1, x_2, \ldots, x_n$ объемом n, а θ – известное значение параметра (скалярного или векторного). В процессе проверки вычисляется значение S^* статистики критерия S как некоторой функции от выборки и теоретического закона распределения $F(x, \theta)$. Далее, опираясь на асимптотическое (предельное) распределение $G(S|H_0)$ статистики критерия при справедливости H_0, вычисляют достигнутый уровень значимости $p_{value} = P\{S > S^*\} = 1 - G\left(S^*|H_0\right)$. Если p_{value} больше заданного α, гипотеза H_0 не отклоняется.

В критерии **Колмогорова** [1] рекомендуется использовать статистику с поправкой Большева в форме [2]

$$S_K = \sqrt{n} D_n + \frac{1}{6\sqrt{n}} = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{1}$$

где $D_n = \max\left(D_n^+, D_n^-\right)$, $D_n^+ = \max\limits_{1 \le i \le n}\left\{\frac{i}{n} - F(x_i, \theta)\right\}$, $D_n^- = \max\limits_{1 \le i \le n}\left\{F(x_i, \theta) - \frac{i-1}{n}\right\}$. При справедливости H_0 статистика (1) подчиняется распределению Колмогорова с функцией распределения $K(s) = \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 s^2}$.

Статистика критерия **Крамера–Мизеса–Смирнова** имеет вид [2]

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F(x_i, \theta) - \frac{2i-1}{2n}\right\}^2 \tag{2}$$

и при простой гипотезе подчиняется закону с функцией распределения $a1(s)$ [2].

В критерии **Андерсона–Дарлинга** [3, 4] статистика задается выражением

$$S_\Omega = -n - 2\sum_{i=1}^{n}\left\{\frac{2i-1}{2n}\ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n}\right)\ln(1 - F(x_i, \theta))\right\}, \tag{3}$$

и при простой H_0 подчиняется закону с функцией распределения $a2(s)$ [2]. В процессе применения этого критерия зачастую приходится сталкиваться с ситуациями, когда на интервале $[x_1, x_n]$ функция распределения $F(x_i, \theta)$ принимает значения 0 или 1. Это исключает возможность вычисления значения статистики (3) и одновременно сигнализирует о несправедливости H_0.

В критерии **Купера** [5] можно использовать статистику в виде [6]

$$V_n^{mod} = \sqrt{n}\left(D_n^+ + D_n^-\right) + \frac{1}{3\sqrt{n}}. \tag{4}$$

При справедливости H_0 асимптотическим распределением статистики (4) является функция распределения $G(s|H_0) = 1 - \sum_{m=1}^{\infty} 2(4m^2s^2 - 1)e^{-2m^2s^2}$ [7].

Статистика критерия **Ватсона** [8, 9] используется в форме

$$U_n^2 = \sum_{i=1}^{n}\left(F(x_i,\ \theta) - \frac{i-1/2}{n}\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n}F(x_i,\ \theta) - \frac{1}{2}\right)^2 + \frac{1}{12n} \tag{5}$$

и при простой H_0 подчиняется закону $G(s|H_0) = 1 - 2\sum_{m=1}^{\infty}(-1)^{m-1}e^{-2m^2\pi^2 s}$.

В критериях **Жанга** [10, 11] статистики задаются выражениями:

$$Z_K = \max_{1\le i\le n}\left(\left(i - \frac{1}{2}\right)\log\left\{\frac{i-1/2}{nF(x_i,\ \theta)}\right\} + \left(n - i + \frac{1}{2}\right)\log\left[\frac{n-i+1/2}{n\{1-F(x_i,\ \theta)\}}\right]\right), \tag{6}$$

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\log\{F(x_i,\ \theta)\}}{n-i+1/2} + \frac{\log\{1-F(x_i,\ \theta)\}}{i-1/2}\right], \tag{7}$$

$$Z_C = \sum_{i=1}^{n}\left[\log\left\{\frac{[F(x_i,\ \theta)]^{-1} - 1}{(n-1/2)/(i-3/4) - 1}\right\}\right]^2. \tag{8}$$

Эти критерии являются развитием критериев со статистиками (1)–(3). Как правило, критерии Жанга имеют преимущество в мощности по сравнению с критериями со статистиками (1)–(5) [6]. Распределения статистик (6)–(8) зависят от объемов выборок n, поэтому вычисление оценок p_{value} возможно только при использовании вычислительных технологий.

Относительно недавно предложен критерий, опирающийся на оценку информации **Кульбака–Лейблера** [12, 13], со статистикой

$$S_{KL} = -\frac{1}{n}\sum_{i=1}^{n}\ln\left\{\frac{n}{2m}[F(x_{i+m},\ \theta) - F(x_{i-m},\ \theta)]\right\}, \tag{9}$$

где $m \le n/2$; $x_i = x_1$, если $i < 1$; $x_i = x_n$, если $i > n$.

Критерии, использующие различные оценки энтропии или, как в данном случае, оценки информации Кульбака–Лейблера, как правило, демонстрируют высокую мощность. Однако в случае выборок с повторяющимися значениями их применение по понятным причинам может оказаться проблематичным. Распределения статистики (9), так же как и статистик критериев Жанга, зависят от объемов выборок n. Поэтому эффективное применение критерия с вычислением оценок p_{value} требует использования компьютерных технологий, базирующихся на методах имитационного (статистического) моделирования.

При проверке простых гипотез все перечисленные критерии являются «свободными от распределения», так как распределения статистик $G(S|H_0)$ не зависят от вида закона $F(x,\ \theta)$.

При малых объемах выборок распределения статистик $G(S_n|H_0)$ критериев могут несколько отличаться от предельных, что приходится учитывать на практике. Для критериев со статистиками (1)–(5) асимптотическими распределениями (вместо реальных $G(S_n|H_0)$), как правило, можно пользоваться при $n \geq 25...30$ [6].

2. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ СОГЛАСИЯ ПРИ ПРОВЕРКЕ СЛОЖНЫХ ГИПОТЕЗ

При проверке сложных гипотез, когда по той же самой выборке оценивают параметры наблюдаемого закона распределения вероятностей $F(x,\ \theta)$, все непараметрические критерии согласия теряют свойство «свободы от распределения». Впервые существование этой проблемы было обозначено в работе [14], которая послужила точкой отсчета для многочисленных попыток ее решения. Например, в работах [15, 16] с использованием статистического моделирования были построены таблицы критических значений для критерия Колмогорова при проверке сложных гипотез относительно нормального и экспоненциального законов. В [17] аналитическими методами аналогичная задача была решена для критерия Крамера–Мизеса–Смирнова. Решению таких же задач были посвящены работы [18, 19]. Жизнь подтвердила, что использование имитационного моделирования является более перспективным направлением при исследованиях распределений статистик критериев, применяемых при проверке различных гипотез.

В наших работах, опирающихся на компьютерные технологии исследований, были построены приближенные модели распределений статистик критериев Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга при проверке различных сложных гипотез относительно ряда законов, часто используемых в приложениях, а также была показана существенная зависимость распределений статистик от используемого метода оценивания параметров [20]. На базе этих результатов были разработаны рекомендации [21]. Последующие исследования и уточненные модели распределений статистик (1)–(5) непараметрических критериев [22–26] послужили основой руководства [6].

При проверке сложной гипотезы на распределение статистики $G(S|H_0)$ критерия влияет совокупность следующих факторов [6]: вид наблюдаемого закона распределения $F(x, \theta)$, соответствующего проверяемой гипотезе H_0; тип оцениваемого параметра и число оцениваемых параметров; используемый метод оценивания параметров [20]; в некоторых ситуациях – конкретное значение параметра (например, значения параметров формы гамма-распределения, бета-распределений, обобщенного нормального закона и других) [6, 25].

Влияние перечисленных факторов на распределения статистик непараметрических критериев согласия (без потери общности) продемонстрируем на статистике (2) критерия Крамера–Мизеса–Смирнова. На данном этапе мы не акцентируем внимания на возможном влиянии на распределения статистик $G(S|H_0)$ ошибок округления Δ.

На рис. 1 показано изменение распределения статистики (2) в зависимости от типа оцениваемого параметра (сдвига или масштаба), от числа оцениваемых параметров нормального закона в случае использования оценок максимального правдоподобия (ОМП), а также от метода оценивания. На рисунке показано также распределение статистики для случая оценивания двух параметров нормального закона в результате минимизации статистики (2) (MD-оценки). Как видим, наблюдается существенная зависимость от типа оцениваемого параметра и от используемого метода оценивания. При проверке согласия с другими законами, определяемыми также лишь параметрами сдвига и масштаба, картина будет аналогична. Однако распределения статистик, соответствующие тем же сложным гипотезам, будут отличаться от приведенных на рис. 1, так как распределения $G(S|H_0)$ зависят от вида закона $F(x,\theta)$. Построенные модели распределений статистик критериев при проверке различных сложных гипотез относительно ряда законов $F(x,\theta)$, которые наиболее часто используются в различных приложениях и для которых отсутствует зависимость от значения параметра (параметров) формы, приведены в [6, 27].

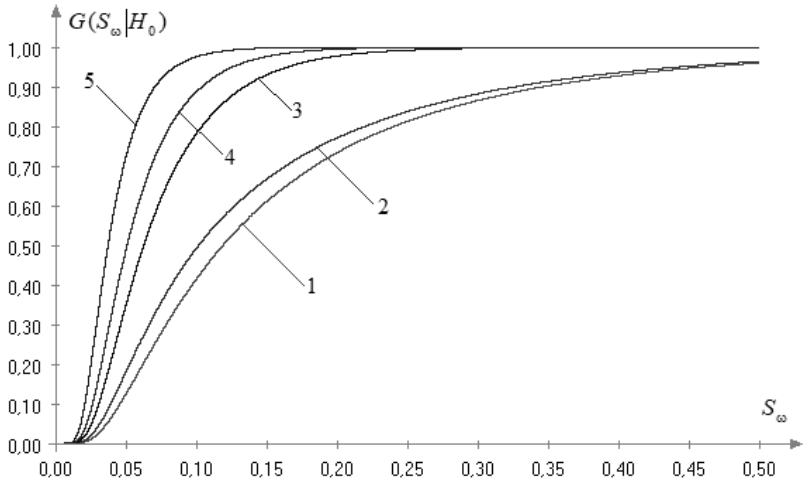


Рис. 1. Предельные распределения статистики (2) при справедливости простой и сложных гипотез о принадлежности выборки нормальному закону при отсутствии ошибок округления ($\Delta = 0$):

1 – распределение $a1(S)$; *2* – распределение статистики при вычислении только ОМП параметра масштаба; *3* – при вычислении только ОМП параметра сдвига; *4* – при вычислении ОМП параметров сдвига и масштаба; *5* – при вычислении MD-оценок параметров сдвига и масштаба

Fig. 1. Limiting distributions of statistics (2) with the validity of simple and composite hypotheses that the sample belongs to the normal law in the absence of round-off errors ($\Delta = 0$):

1 is distribution $a1(S)$; *2* is distribution of statistics when calculating only he MLE of the scale parameter; *3* – when calculating only the MLE of the shift parameter; *4* – when calculating the MLE of the shift and scale parameters; *5* – when calculating MD-estimates of shift and scale parameters

Влияние конкретного значения параметра формы на распределение статистики $G(S|H_0)$ при проверке сложной гипотезы продемонстрируем на обобщенном нормальном законе с плотностью $f(x)=\dfrac{\theta_2}{2\theta_1\Gamma(1/\theta_2)}\times$ $\times\exp\left\{-(|x-\theta_0|/\theta_1)^{\theta_2}\right\}$, где θ_2 – параметр формы. В частном случае при $\theta_2=2$ обобщенный нормальный закон совпадает с нормальным, а при $\theta_2=1$ – с распределением Лапласа. На рис. 2 показаны распределения $G(S|H_0)$ статистики (2) при проверке сложной гипотезы о принадлежности выборки обобщенному нормальному закону в случае оценивания всех трех параметров методом максимального правдоподобия в зависимости от значения θ_2. Следует отметить, что с ростом θ_2 распределение $G(S|H_0)$ сначала удаляется от $a1(S)$ (до $\theta_2\approx1.6$), а затем начинает сближаться с $a1(S)$.

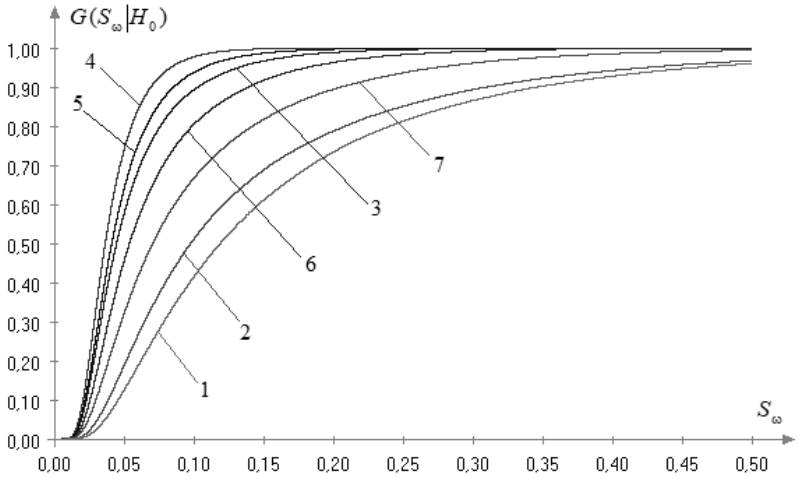


Рис. 2. Предельные распределения статистики (2) при справедливости сложной гипотезы о принадлежности выборки обобщенному нормальному закону при отсутствии ошибок округления ($\Delta = 0$) в зависимости от значения параметра формы θ_2 (при вычислении ОМП трех параметров закона):

1 – распределение $a1(S)$; *2* – при $\theta_2 = 0.5$; *3* – при $\theta_2 = 1$; *4* – при $\theta_2 = 1.6$; *5* – при $\theta_2 = 2$; *6* – при $\theta_2 = 3$; *7* – при $\theta_2 = 5$

Fig. 2. Limiting distributions of statistics (2) with the validity of the composite hypothesis that the sample belongs to the generalized normal law in the absence of round-off errors ($\Delta = 0$), depending on the value of the form parameter θ_2 (when calculating the MLE of 3 parameters of the law):

1 is distribution $a1(S)$; *2* – with $\theta_2 = 0.5$; *3* – with $\theta_2 = 1$; *4* – with $\theta_2 = 1.6$; *5* – with $\theta_2 = 2$; *6* – with $\theta_2 = 3$; *7* – with $\theta_2 = 5$

В подобных ситуациях (наличия зависимости $G(S|H_0)$ от значений параметра или параметров формы) исключается возможность заранее построить модель закона (асимптотического или предельного), так как значение пара-

метра мы узнаем только в процессе проверки гипотезы (при оценивании). Отсюда следует, что распределения статистик применяемых критериев должны находиться (моделироваться) в интерактивном режиме в ходе проводимого статистического анализа [28], а затем использоваться при формировании вывода по итогам проверки сложной гипотезы.

Реализация такого интерактивного режима требует наличия развитого программного обеспечения, позволяющего в целях ускорения распараллеливать процессы имитационного моделирования и привлекать доступные вычислительные ресурсы. В результате N имитационных экспериментов можно построить эмпирическое распределение статистики $G_N(S_n|H_0)$ (при n, соответствующем объему анализируемой выборки), по которому может быть найдена оценка p_{value}. В работе [29] такой режим исследования распределений статистик предусмотрен в случае применения критериев Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга, Купера, Ватсона, трех критериев Жанга и критерия Кульбака–Лейблера. При этом могут использоваться различные методы оценивания параметров. Для критериев Жанга и Кульбака–Лейблера такой режим используется и при проверке простых гипотез.

Следует признать, что появление рекомендаций [21] и руководства [6] не сильно изменило в лучшую сторону ситуацию с корректностью применения непараметрических критериев согласия при проверке сложных гипотез. Большинство ошибок применения непараметрических критериев согласия в приложениях (и в учебной литературе) по-прежнему связано с незнанием существующей проблемы.

Картина, представленная на рис. 1 и 2, показывает диапазон изменений распределений $G(S|H_0)$ при простой и сложных гипотезах и подчеркивает, что вследствие ошибок применения, не учитывающих сложности проверяемой гипотезы, существенно увеличивается вероятность ошибок 2-го рода.

Представленная картина не учитывает влияния ошибок округления, которые всегда присутствуют в реальных данных.

Такая картина сохраняется при ошибках округления $\Delta << \sigma$, где σ – среднеквадратичное отклонение ошибки измерения, и при таких максимальных объемах выборок $n_{\max}$, при которых $G(S_{n_{\max}}|H_0)$ (при данном Δ) еще практически не отличается от $G(S|H_0)$ [30]. Именно в таких ситуациях оказывается правомерным использование классических результатов, связанных с критериями проверки гипотез, и имеют силу результаты и рекомендации, представленные в руководстве [6] и в руководствах [31–34].

3. ВЛИЯНИЕ ОКРУГЛЕНИЯ НА СТАТИСТИЧЕСКИЕ ВЫВОДЫ

Любые измерения фиксируются с некоторой погрешностью округления. Что наличие ошибок округления Δ может как-то отражаться на результатах статистических выводов, очевидно давно. Например, о возможности проблем с применением критериев нормальности, являющихся следствием округления, отмечалось еще в работе [35]. В работах [36, 37] на примере критериев проверки гипотез о равенстве математического ожидания и дисперсии номинальным значениям, а также критериев Стьюдента об однородности средних и Фишера

об однородности дисперсий двух выборок было показано влияние ошибок округления на реальный уровень значимости, а также отмечено, что с их увеличением снижается мощность критериев. Авторами работы [38] при анализе множества выборок с повторяющимися наблюдениями было показано, что в такой ситуации критические значения распределений статистик непараметрических критериев согласия при проверке сложных гипотез относительно обобщенного распределения Парето отличаются от ранее полученных. Наконец, в наших работах [39–41] показано, как в условиях соизмеримости ошибок округления Δ и среднеквадратичного отклонения ошибок измерения σ изменяются распределения статистик различных статистических критериев.

В данном случае, опять же без потери общности, в условиях соизмеримости Δ и σ проиллюстрируем влияние Δ на распределения $G(S|H_0)$ статистики (2) критерия Крамера–Мизеса–Смирнова на примере проверки гипотезы о принадлежности выборок нормальному закону распределения.

На рис. 3 в условиях справедливости простой гипотезы H_0 показана зависимость распределения статистики (2) от величины $\Delta = w\sigma$ при $n = 50$. Очевидно, что отклонением $G(S_{50}|H_0)$ от $a1(S)$ можно пренебречь лишь при $\Delta < 0.1\sigma$.

На рис. 4, также при простой гипотезе H_0 и фиксированной ошибке округления $\Delta = 0.1\sigma$, показана зависимость $G(S_n|H_0)$ от объемов выборок *n*.

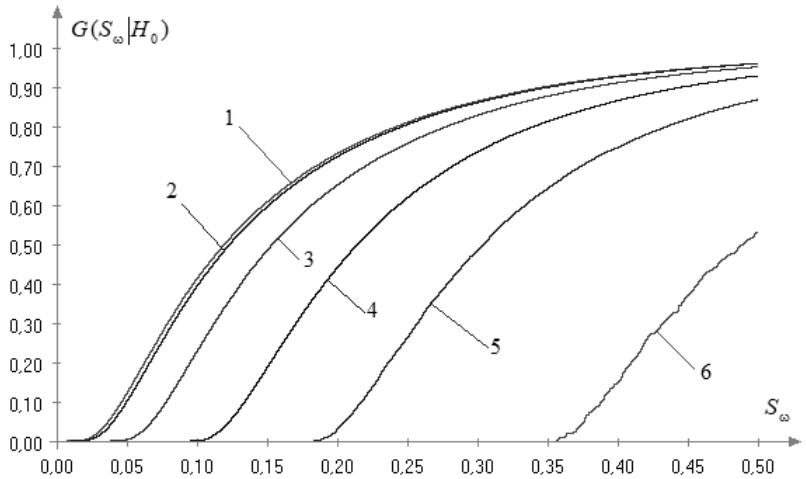


Рис. 3. Зависимость распределения статистики (2) от ошибки округления Δ при справедливости простой гипотезы H_0 о принадлежности выборки нормальному закону при $n = 50$:

1 – $a1(S)$ (при $\Delta = 0$); *2* – при $\Delta = 0.1\sigma$; *3* – при $\Delta = 0.3\sigma$; *4* – при $\Delta = 0.5\sigma$; *5* – при $\Delta = 0.7\sigma$; *6* – при $\Delta = \sigma$

Fig. 3. Dependence of the statistics (2) distribution on the round-off error Δ when the simple hypothesis H_0 that the sample belongs to the normal law is valid when $n = 50$:

1 – $a1(S)$ (with $\Delta = 0$); *2* – with $\Delta = 0.1\sigma$; *3* – with $\Delta = 0.3\sigma$; *4* – with $\Delta = 0.5\sigma$; *5* – with $\Delta = 0.7\sigma$; *6* – with $\Delta = \sigma$

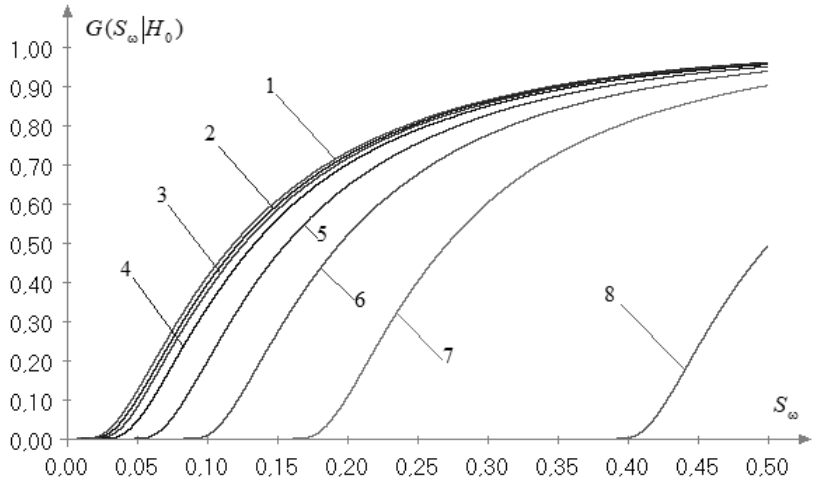


Рис. 4. Зависимость распределения статистики (2) от объема выборки n при справедливости простой гипотезы H_0 о принадлежности выборки нормальному закону при ошибке округления $\Delta = 0.1\sigma$:

1 – $a1(S)$; *2* – при $n = 50$; *3* – при $n = 100$; *4* – при $n = 200$; *5* – при $n = 500$; *6* – при $n = 1000$; *7* – при $n = 2000$; *8* – при $n = 5000$

Fig. 4. Dependence of the statistics (2) distribution on the sample size n with the validity of the simple hypothesis H_0 that the sample belongs to the normal law with a round-off error $\Delta = 0.1\sigma$:

1 is $a1(S)$; *2* – with $n = 50$; *3* – with $n = 100$; *4* – with $n = 200$; *5* – with $n = 500$; *6* – with $n = 1000$; *7* –with $n = 2000$; *8* – with $n = 5000$

Что вытекает из представленных результатов исследований?

Во-первых, можно видеть, что наличие ошибок округления приводит к появлению зависимости $G(S|H_0)$ от n.

Во-вторых, признание самого факта наличия округлений в данных исключает возможность использования предельного распределения $a1(S)$ в качестве распределения статистики (2) в условиях больших выборок.

В-третьих, в условиях соизмеримости Δ и σ и при относительно небольших объемах выборок распределения $G(S_n|H_0)$ статистики (2) могут значительно отличаться от $a1(S)$, что полностью исключает возможность при проверке гипотезы использовать классические результаты. С такими ситуациями часто сталкиваются в задачах анализа данных в биологии и медицине, где результаты измерений всегда имеют ограниченную точность, а в выборках оказывается много повторяющихся значений, что является признаком соизмеримости Δ и σ. Именно в этом заключается причина удивляющей специалистов «неработоспособности» многих статистических критериев, применяемых к подобным выборкам. Следует добавить, что с такой же ситуацией сталкиваются при анализе высокоточных измерений в технических приложениях, когда измерения осуществляются на пределе разрешающей способности используемой измерительной системы.

В-четвертых, проведенные исследования показали, что вследствие округлений потеря свойства «свободы от распределения» происходит и в

условиях проверки простых гипотез. В частности, при одном и том же соотношении $\Delta = w\sigma$ между Δ и σ наблюдаемого симметричного закона степень отклонения $G(S_n|H_0)$ от $a1(S)$ увеличивается в случае законов с более тяжелыми «хвостами» (по сравнению с нормальным законом).

В ситуации проверки сложной гипотезы при использовании ОМП для оценки двух параметров нормального закона мы имеем аналогичную картину влияния Δ на распределения статистики (2) (рис. 5 и 6).

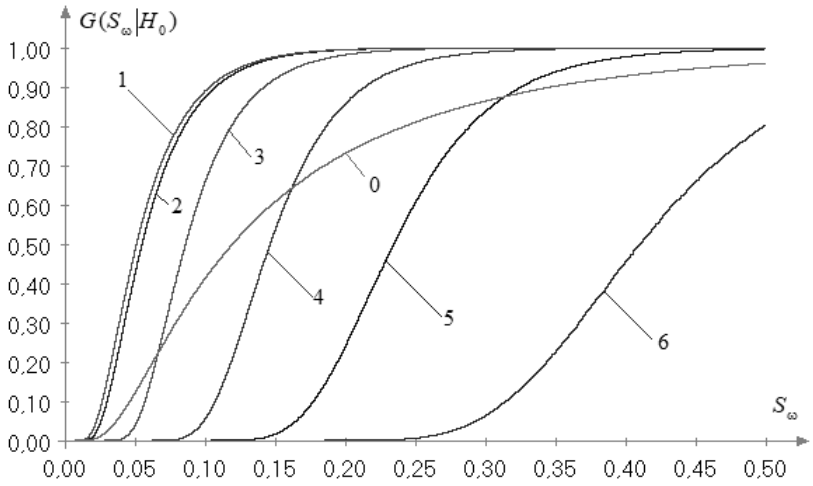


Рис. 5. Зависимость распределения статистики (2) от ошибки округления Δ при справедливости сложной гипотезы H_0 о принадлежности выборки нормальному закону (в случае ОМП) при $n = 50$:

0 – $a1(S)$; *1* – асимптотическое (при $\Delta = 0$); *2* – при $\Delta = 0.1\sigma$; *3* – при $\Delta = 0.3\sigma$; *4* – при $\Delta = 0.5\sigma$; *5* – при $\Delta = 0.7\sigma$; *6* – при $\Delta = \sigma$

Fig. 5. Dependence of the statistics (2) distribution on the round-off error Δ when the composite hypothesis H_0 that the sample belongs to the normal law is true (in the case of MLE) when $n = 50$:

0 – $a1(S)$; *1* is asymptotic distribution (with $\Delta = 0$); *2* – with $\Delta = 0.1\sigma$; *3* – with $\Delta = 0.3\sigma$; *4* – with $\Delta = 0.5\sigma$; *5* – with $\Delta = 0.7\sigma$; *6* – with $\Delta = \sigma$

На распределения статистик (6)–(9), которые и в отсутствие ошибок округления зависят от n, факторы, влияющие на распределения статистик непараметрических критериев согласия при проверке сложных гипотез, и ошибки округления воздействуют аналогичным образом.

Стандартным предположением, обусловливающим корректность выводов по применяемым непараметрическим критериям согласия со статистиками (1)–(9) с использованием известных классических результатов (предельных распределений или критических значений), касающихся проверки простых гипотез, в условиях $\Delta << \sigma$ является отсутствие зависимости $G(S|H_0)$ от Δ (при данном n).

В случае проверки сложных гипотез это предположение дополняется необходимостью знания модели предельного распределения статистики для соответствующей сложной гипотезы [6], которая чаще всего оказывается неизвестной.

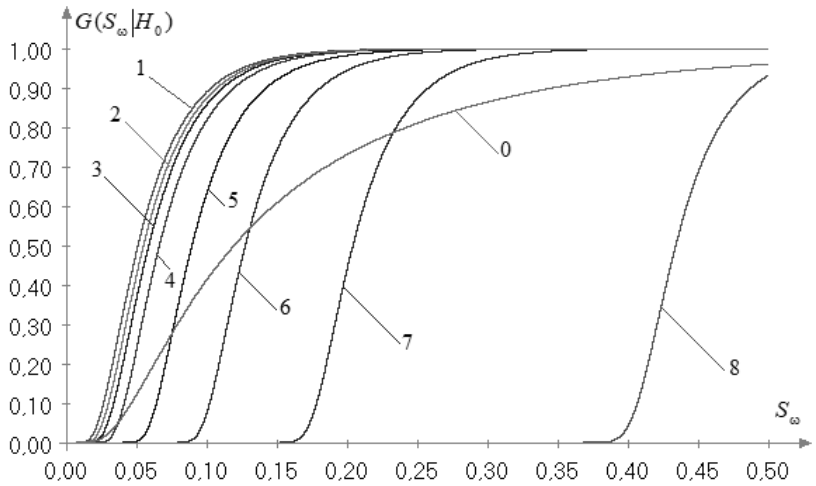


Рис. 6. Зависимость распределения статистики (2) от объема выборки n при справедливости сложной гипотезы H_0 о принадлежности выборки нормальному закону (в случае ОМП) при $\Delta = 0.1\sigma$:

0 – $a1(S)$; *1* – асимптотическое (при $\Delta = 0$); *2* – при $n = 50$; *3* – при $n = 100$; *4* – при $n = 200$; *5* – при $n = 500$; *6* – при $n = 1000$; *7* – при $n = 2000$; *8* – при $n = 5000$

Fig. 6. Dependence of the statistics (2) distribution on the sample size n with the validity of the composite hypothesis H_0 that the sample belongs to the normal law (in the case of MLE) with $\Delta = 0.1\sigma$:

0 – $a1(S)$; *1* is asymptotic distribution (at $\Delta = 0$); *2* – with $n = 50$; *3* – with $n = 100$; *4* – with $n = 200$; *5* – with $n = 500$; *6* – with $n = 1000$; *7* – with $n = 2000$; *8* – with $n = 5000$

Приходится констатировать, что в подавляющем числе случаев критерии применяются в условиях нарушения стандартных предположений.

4. ПРИМЕНЕНИЕ КРИТЕРИЕВ В НЕСТАНДАРТНЫХ УСЛОВИЯХ

Еще раз подчеркнем, что вывод о результатах проверки гипотезы, формируемый на основе оценки достигнутого уровня значимости $p_{value} = P\{S > S^*\} = 1 - G\left(S^* | H_0\right)$, более информативен и обоснован, чем принимаемый по результатам сравнения значения статистики S^*, вычисленной по анализируемой выборке, с критическим значением $S_{1-\alpha}$. Поэтому предпочтительней опираться на распределения статистик $G(S_n | H_0)$, а не на критические значения.

При проверке простых гипотез нарушение стандартных предположений, препятствующее возможности использования предельных распределений для вычисления p_{value}, может быть связано только с изменением $G(S_n | H_0)$ под влиянием ошибок округлений Δ.

При проверке сложных гипотез приходится учитывать зависимость $G(S_n|H_0)$ от вида $F(x,\theta)$, от числа и типа оцениваемых параметров, от используемого метода оценивания, от конкретных значений параметров формы, к которым добавляется влияние на $G(S_n|H_0)$ ошибок округления Δ. То есть за редким исключением [6], распределения статистик $G(S_n|H_0)$ оказываются неизвестными.

Для критериев со статистиками (6)–(9) распределения статистик $G(S_n|H_0)$ зависят от n и от всех факторов, связанных с проверкой сложных гипотез. При проверке как простых, так и сложных гипотез возможно влияние ошибок округления Δ. То есть по существу распределения $G(S_n|H_0)$ статистик (6)–(9) всегда оказываются неизвестными.

Вопрос об аналитических исследованиях вида $G(S_n|H_0)$ в таких разнообразных ситуациях даже не ставится как абсолютно бесперспективный.

А вот интерактивное исследование распределения $G(S_n|H_0)$ статистики критерия в ходе проверки конкретной гипотезы (простой или сложной, при фактических ошибках округления и объеме n анализируемой выборки) позволяет в результате статистического моделирования построить эмпирическое распределение статистики $G_N(S_n|H_0)$ и найти оценку p_{value}.

Такой интерактивный режим исследования распределений статистик предусмотрен в работе [29] при использовании рассмотренного множества критериев согласия. При этом в случае отсутствия модели распределения статистики используемого критерия при проверке соответствующей сложной гипотезы автоматически предлагается построить это распределение в результате моделирования. Если присутствуют опасения, что на результаты анализа могут влиять ошибки округления Δ, надо указать эту возможность и задать величину Δ, соответствующую округлениям в анализируемой выборке. Интерактивный режим моделирования $G_N(S_n|H_0)$ обеспечивает корректность выводов по применяемым критериям в условиях любых нарушений стандартных предположений.

Покажем, как меняются оценки p_{value}, соответствующие вычисленным значениям статистик (1)–(5), при учете влияния ошибок округления на распределения статистик критериев. Продемонстрируем это на примере проверки сложной гипотезы о принадлежности выборки объемом $n = 50$:

0.7 1.0 0.3 0.3 0.3 0.7 0.4 0.4 0.6 0.7 0.7 0.3 0.5 0.9 0.2
0.8 0.5 0.6 0.4 0.6 0.2 0.4 0.4 1.0 1.0 0.7 1.3 0.9 0.1 0.2
0.3 0.5 0.8 0.6 0.8 0.7 0.4 1.1 0.6 1.2 0.4 0.3 0.4 0.2 0.5
1.6 0.8 0.6 0.6 0.6

гамма-распределению с плотностью

$$f(x) = \left[\frac{x-\theta_2}{\theta_1}\right]^{\theta_0-1} \frac{e^{-(x-\theta_2)/\theta_1}}{\theta_1\Gamma(\theta_0)}.$$

Измерения фиксировались с ошибкой округления $\Delta = 0.1$. Параметры формы θ_0 и масштаба θ_1 оценивались методом максимального правдоподо-

бия. Параметр сдвига $\theta_2 = 0.0$ не оценивался. Распределения статистик $G(S_n|H_0)$ критериев в данном случае зависят от значения оценки $\hat{\theta}_0$. ОМП параметров, полученные по выборке, принимают значения: $\hat{\theta}_0 = 3.72296$, $\hat{\theta}_1 = 0.61699$. В таблице приведены значения статистик (1)–(5), рассматриваемых критериев и оценки p_{value}, соответствующие двум ситуациям: а) отсутствие ошибок округления (при $\Delta = 0$); б) наличие ошибок округления $\Delta = 0.1$.

Результаты проверки гипотезы по критериям со статистиками (1)–(5)

Results of hypothesis testing by criteria with statistics (1)–(5)

№ п/п	Критерий	Статистика	p_{value}	
			$\Delta = 0$	$\Delta = 0.1$
1	Колмогорова	0.776317	0.167	0.774
2	Крамера–Мизеса–Смирнова	0.064356	0.343	0.947
3	Андерсона–Дарлинга	0.360314	0.466	0.985
4	Купера	1.451192	0.075	0.744
5	Ватсона	0.063283	0.303	0.941

Как можно видеть, оценки p_{value}, учитывающие наличие округлений измерений порядка $\Delta = 0.1$, существенно отличаются от значений, не учитывающих ошибок округления.

ЗАКЛЮЧЕНИЕ

Таким образом, существует две основные причины, вследствие которых при идентификации закона распределения ошибок измерений с использованием непараметрических критериев согласия выводы могут оказаться некорректными.

Первая связана с использованием при проверке сложных гипотез классических результатов, соответствующих проверке простых гипотез, что приводит к увеличению вероятностей ошибок 2-го рода: при справедливости некоторой конкурирующей гипотезы проверяемая гипотеза H_0 не отклоняется.

Вторая причина связана с наличием ошибок округления, которые всегда сопровождают процесс измерений. При $\Delta << \sigma$ и n меньше некоторого $n_{\max}$, зависящего от n и σ, влиянием Δ на $G(S_n|H_0)$ можно пренебречь. Но при $n > n_{\max}$ реальное распределение статистики уже заметно отклоняется от асимптотического распределения, использование которого в этом случае приводит к увеличению вероятностей ошибок 1-го рода, т. е. к отклонению справедливой гипотезы H_0. При соизмеримости Δ и σ такая ситуация может иметь место и при малых объемах выборок, а с ростом n она будет только усугубляться.

Единственным выходом, способным обеспечивать корректность выводов по применяемым критериям в нестандартных условиях (при проверке

сложных гипотез и влиянии Δ на $G(S_n|H_0)$), является использование реальных распределений статистик критериев (имеющих место в этих нестандартных условиях). Эта задача должна решаться в интерактивном режиме (в процессе проверки) и опираться на компьютерные технологии исследования и аппарат математической статистики. Такие процедуры имитационного моделирования распределений статистик критериев могут быть реализованы в рамках специализированного программного обеспечения, предназначенного для анализа результатов измерений.

СПИСОК ЛИТЕРАТУРЫ

1. *Kolmogoroff A.N.* Sulla determinazione empirica di una legge di distribuzione // Giornale del Istituto Italiano degli Attuari. – 1933. – Vol. 4, N 1. – P. 83–91.

2. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. – М.: Наука, 1983. – 416 с.

3. *Anderson T.W., Darling D.A.* Asymptotic theory of certain “Goodness of fit” criteria based on stochastic processes // The Annals of Mathematical Statistics. – 1952. – Vol. 23. – P. 193–212.

4. *Anderson T.W., Darling D.A.* A test of goodness of fit // Journal of the American Statistical Association. – 1954. – Vol. 29. – P. 765–769.

5. *Kuiper N.H.* Tests concerning random points on a circle // Indagationes Mathematicae (Proceedings). – 1960. – Vol. 63. – P. 38–47.

6. *Лемешко Б.Ю.* Непараметрические критерии согласия: руководство по применению. – М.: Инфра-М, 2014. – 163 с. – DOI: 10.12737/11873.

7. *Stephens M.A.* The goodness-of-fit statistic V_N: distribution and significance points // Biometrika. – 1965. – Vol. 52, N 3–4. – P. 309–321.

8. *Watson G.S.* Goodness-of-fit tests on a circle. 1 // Biometrika. – 1961. – Vol. 48, N 1–2. – P. 109–114.

9. *Watson G.S.* Goodness-of-fit tests on a circle. 2 // Biometrika. – 1962. – Vol. 49, N 1–2. – P. 57–63.

10. *Zhang J.* Powerful goodness-of-fit and multi-sample tests: PhD Thesis / York University. – Toronto, 2001. – 113 p. – URL: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf (дата обращения 05.01.2021).

11. *Zhang J.* Powerful goodness-of-fit tests based on the likelihood ratio // Journal of the Royal Statistical Society: Series B. – 2002. – Vol. 64, N 2. – P. 281–294.

12. *Noughabi H.A, Arghami N.R.* General treatment of goodness of fit tests based on Kullback–Leibler information // Journal of Statistical Computation and Simulation. – 2013. – Vol. 83. – P. 1556–1569.

13. *Noughabi H.A.* A new estimator of Kullback–Leibler information and its application in goodness of fit tests // Journal of Statistical Computation and Simulation. – 2019. – Vol. 89, N 10. – P. 1914–1934.

14. *Kac M., Kiefer J., Wolfowitz J.* On tests of normality and other tests of goodness of fit based on distance methods // The Annals of Mathematical Statistics. – 1955. – Vol. 26. – P. 189–211.

15. *Lilliefors H.W.* On the Kolmogorov-Smirnov test for normality with mean and variance unknown // Journal of the American Statistical Association. – 1967. – Vol. 62. – P. 399–402.

16. *Lilliefors H.W.* On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown // Journal of the American Statistical Association. – 1969. – Vol. 64. – P. 387–389.

17. *Мартынов Г.В.* Критерии омега-квадрат. – М.: Наука, 1978. – 80 с.

18. *Тюрин Ю.Н.* О предельном распределении статистик Колмогорова–Смирнова для сложной гипотезы // Известия АН СССР. Серия математическая. – 1984. – Т. 48, № 6. – С. 1314–1343.

19. *Тюрин Ю.Н., Саввушкина Н.Е.* Критерии согласия для распределения Вейбулла–Гнеденко // Известия АН СССР. Серия: Техническая кибернетика. – 1984. – № 3. – С. 109–112.

20. *Лемешко Б.Ю., Постовалов С.Н.* О зависимости распределений статистик непараметрических критериев и их мощности от метода оценивания параметров // Заводская лаборатория. Диагностика материалов. – 2001. – Т. 67, № 7. – С. 62–71.

21. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. 2. Непараметрические критерии. – М.: Изд-во стандартов, 2002. – 64 с.

22. *Lemeshko B.Yu., Lemeshko S.B.* Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Pt. 1 // Measurement Techniques. – 2009. – Vol. 52, N 6. – P. 555–565.

23. *Lemeshko B.Yu., Lemeshko S.B.* Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Pt. 2 // Measurement Techniques. – 2009. – Vol. 52, N 8. – P. 799–812.

24. *Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N.* Statistic distribution models for some nonparametric goodness-of-fit tests in testing composite hypotheses // Communications in Statistics. Theory and Methods. – 2010. – Vol. 39, N 3. – P. 460–471.

25. *Lemeshko B.Yu., Lemeshko S.B.* Models of statistic distributions of nonparametric goodness-of-fit tests in composite hypotheses testing for double exponential law cases // Communications in Statistics. Theory and Methods. – 2011. – Vol. 40, N 16. – P. 2879–2892.

26. *Lemeshko B.Yu., Gorbunova A.A.* Application of nonparametric Kuiper and Watson tests of goodness-of-fit for composite hypotheses // Measurement Techniques. – 2013. – Vol. 56, N 9. – P. 965–973.

27. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск: Изд-во НГТУ, 2011. – 888 с.

28. *Lemeshko B.Yu.*, *Lemeshko S.B., Rogozhnikov A.P.* Interactive investigation of statistical regularities in testing composite hypotheses of goodness of fit // Statistical models and methods for reliability and survival analysis / ed. by V. Couallier, L. Gerville-Reache, C. Huber-Carol. – Hoboken, NJ: Wiley-ISTE, 2013. – Ch. 5. – P. 61–76.

29. Статистический анализ интервальных наблюдений одномерных непрерывных случайных величин «Интервальная статистика 5.4»: свидетельство о гос. регистрации программы для ЭВМ № 2018666213 / Лемешко Б.Ю., Лемешко С.Б., Блинов П.Ю., Веретельникова И.В., Новикова А.Ю. – Заявка № 2018663206; заявл. 22.11.2018; зарег. 13.12.2018. – URL: https://ami.nstu.ru/~headrd/ISW.htm (дата обращения: 22.01.2021).

30. *Лемешко Б.Ю. Лемешко С.Б., Семёнова М.А.* К вопросу статистического анализа больших данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2018. – № 44. – С. 40–49. – DOI: 10.17223/19988605/44/5.

31. *Лемешко Б.Ю., Блинов П.Ю.* Критерии проверки отклонения распределения от равномерного закона: руководство по применению. – М.: Инфра-М, 2015. – 183 с. – (Научная мысль. Математическая статистика). – DOI: 10.12737/11304.

32. *Лемешко Б.Ю.* Критерии проверки отклонения распределения от нормального закона: руководство по применению. – М.: Инфра-М, 2015. – 160 с. – (Научная мысль. Математическая статистика). – DOI: 10.12737/6086.

33. *Лемешко Б.Ю., Блинов П.Ю.* Критерии проверки отклонения от экспоненциального закона: руководство по применению. – М.: Инфра-М, 2021. – 352 с. – (Научная мысль. Математическая статистика). – DOI 10.12737/1097477.

34. *Лемешко Б.Ю.* Критерии проверки гипотез об однородности: руководство по применению. – 2-е изд., перераб. и доп. – М.: Инфра-М, 2021. – 248 с. – (Научная мысль. Математическая статистика). – DOI 10.12737/986695.

35. *Pearson E.S., D'Agostino R.B., Bowman K.O.* Tests for departure from normality: comparison of powers // Biometrika. – 1977. – Vol. 64. – P. 231–246. – DOI: 10.1093/biomet/64.2.427-a.

36. *Tricker A.R.* The effect of rounding on the significance level of certain normal test statistics // Journal of Applied Statistics. – 1990. – Vol. 17, N 1. – P. 31–38. – DOI: 10.1080/757582644.

37. *Tricker A.R.* The effect of rounding on the power level of certain normal test statistics // Journal of Applied Statistics. – 1990. – Vol. 17, N 2. – P. 219–228. – DOI: 10.1080/757582833.

38. *Deidda R., Puliga M.* Sensitivity of goodness-of-fit statistics to rainfall data rounding off // Physics and Chemistry of the Earth. – 2006. – Vol. 31. – P. 1240–1251. – DOI: 10.1016/j.pce.2006.04.041.

39. *Лемешко Б.Ю., Лемешко С.Б.* Влияние округления на свойства критериев проверки статистических гипотез // Автометрия. – 2020. – Т. 56, № 3. – С. 35–45. – DOI: 10.15372/AUT20200305.

40. *Лемешко Б.Ю., Лемешко С.Б.* О влиянии ошибок округления на распределения статистик критериев согласия // Вестник Томского государственного университета. Управление,

вычислительная техника и информатика. – 2020. – № 53. – С. 47–60. – DOI: 10.17223/19988605/53/5.

41. *Lemeshko B.Y., Lemeshko S.B.* About the effect of rounding on the properties of tests for testing statistical hypotheses // Journal of Physics: Conference Series. – 2021. – Vol. 1715. – P. 012063. – DOI: 10.1088/1742-6596/1715/1/012063.

Лемешко Борис Юрьевич, доктор технических наук, профессор, профессор кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление научных исследований – компьютерные технологии исследования статистических и вероятностных закономерностей. Имеет более 400 печатных работ, в том числе 20 монографий и учебных пособий. E-mail: lemeshko@ami.nstu.ru

Лемешко Станислав Борисович, кандидат технических наук, старший научный состудник Центра статистических технологий Новосибирского государственного технического университета. Основное направление научных исследований – компьютерные технологии моделирования и исследования статистических закономерностей. Имеет 67 печатных работ, в том числе одну монографию. E-mail: skyer@mail.ru

Lemeshko Boris Yu., D.Sc. (Eng,), professor, professor at the Department of Theoretical and Applied Informatics, Novosibirsk State Technical University. The main field of his scientific research is computer technologies to study statistical and probabilistic laws. He has more than 400 publications including 20 monographs and textbooks. E-mail: lemeshko@ami.nstu.ru

Lemeshko Stanislav B., PhD (Eng.), senior researcher in the Center for Statistical Technology, Novosibirsk State Technical University. The main field of his scientific research is computer modeling technologies and research of statistical laws. He has 67 publications including 1 monograph. E-mail: skyer@mail.ru

DOI: 10.17212/2782-2001-2021-2-47-66

Problems of nonparametric goodness-of-fit test application in tasks of measurement results processing[*]

B.Yu. LEMESHKO[a], *S.B. LEMESHKO*[b]

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

[a] *lemeshko@ami.nstu.ru* [b] *skyer@mail.ru*

Abstract

It is argued that in most cases two reasons underlie the incorrect application of nonparametric goodness-of-fit tests in various applications.

The first reason is that when testing composite hypotheses and evaluating the parameters of the law for the analyzed sample, classical results associated with testing simple hypotheses are used. When testing composite hypotheses, the distributions of goodness-of-fit statistics are influenced by the form of the observed law $F(x, \theta)$ corresponding to the hypothesis being tested, by the type and number of estimated parameters, by the estimation method, and in some cases by the value of the shape parameter. The paper shows the influence of all mentiomed factors on the distribution of test statistics. It is emphasized that, when testing composite hypotheses, the neglect, of the fact that the test has lost the property of "freedom from distribution" leads to an increase in the probability of the 2nd kind errors. It is shown that the distribution of the statistics of the test necessary for the formation of a conclusion about the results of testing a

[*] *Received 05 February 2021.*

composite hypothesis can be found using simulation in an interactive mode directly in the process of testing.

The second reason is associated with the presence of round-off errors which can significantly change the distributions of test statistics. The paper shows that asymptotic results when testing simple and composite hypotheses can be used with round -off errors Δ much less than the standard deviation σ of the distribution law of measurement errors and sample sizes n not exceeding some maximum values. For sample sizes larger than these maximum values, the real distributions of the test statistics deviate from asymptotic ones towards larger statistics values. In such situations, the use of asymptotic distributions to arrive at a conclusion about the test results leads to an increase in the probabilities of errors of the 1st kind (to the rejection of a valid hypothesis being tested). It is shown that when the round-off errors and σ are commensurable, the distributions of the test statistics deviate from the asymptotic distributions for small n. And as n grows, the situation only gets worse. In the paper, changes in the distributions of statistics under the influence of rounding are demonstrated both when testing both simple and composite hypotheses. It is shown that the only way out that ensures the correctness of conclusions according to the applied tests in such non-standard conditions is the use of real distributions of statistics. This task can be solved interactively (in the process of verification) and rely on computer research technologies and the apparatus of mathematical statistics.

Keywords: hypothesis testing, nonparametric goodness-of-fit tests, simple hypothesis, composite hypothesis, distribution of statistics, round-off errors, achieved significance level, error of the 1st kind, error of the 2nd kind, simulation

REFERENCES

1. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione. *Giornale del Istituto Italiano degli Attuari*, 1933, vol. 4, no. 1, pp. 83–91.

2. Bol'shev L.N., Smirnov N.V. *Tablitsy matematicheskoi statistiki* [Tables for mathematical statistics]. Moscow, Nauka Publ., 1983. 416 p.

3. Anderson T.W., Darling D.A. Asymptotic theory of certain “Goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 1952, vol. 23, pp. 193–212.

4. Anderson T.W., Darling D.A. A test of goodness of fit. *Journal of the American Statistical Association*, 1954, vol. 29, pp. 765–769.

5. Kuiper N.H. Tests concerning random points on a circle. *Indagationes Mathematicae (Proceedings)*, 1960, vol. 63, pp. 38–47.

6. Lemeshko B.Yu. *Neparametricheskie kriterii soglasiya: rukovodstvo po primeneniyu* [Nonparametric goodness-of-fit tests. Guide on the application]. Moscow, Infra-M, 2014. 163 p. DOI: 10.12737/11873.

7. Stephens M.A. The goodness-of-fit statistic V_N: distribution and significance points. *Biometrika*, 1965, vol. 52, no. 3–4, pp. 309–321.

8. Watson G.S. Goodness-of-fit tests on a circle. 1. *Biometrika*, 1961, vol. 48, no. 1–2, pp. 109–114.

9. Watson G.S. Goodness-of-fit tests on a circle. 2. *Biometrika*, 1962, vol. 49, no. 1–2, pp. 57–63.

10. Zhang J. *Powerful goodness-of-fit and multi-sample tests: PhD Thesis*. York University. Toronto, 2001. 113 p. Available at: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf (accessed 15.01.2021).

11. Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B*, 2002, vol. 64, no. 2, pp. 281–294.

12. Noughabi H.A, Arghami N.R. General treatment of goodness of fit tests based on Kullback–Leibler information. *Journal of Statistical Computation and Simulation*, 2013, vol. 83, pp. 1556–1569.

13. Noughabi H.A. A new estimator of Kullback–Leibler information and its application in goodness of fit tests. *Journal of Statistical Computation and Simulation*, 2019, vol. 89, no. 10, pp. 1914–1934.

14. Kac M., Kiefer J., Wolfowitz J. On tests of normality and other tests of goodness of fit based on distance methods. *The Annals of Mathematical Statistics*, 1955, vol. 26, pp. 189–211.

15. Lilliefors H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 1967, vol. 62, pp. 399–402.

16. Lilliefors H.W. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, 1969, vol. 64, pp. 387–389.

17. Martynov G.V. *Kriterii omega-kvadrat* [Omega-square criteria]. Moscow, Nauka Publ., 1978. 80 p.

18. Tyurin Yu.N. O predel'nom raspredelenii statistik Kolmogorova–Smirnova dlya slozhnoi gipotezy [On the limit distribution of Kolmogorov–Smirnov statistics for a composite hypothesis]. *Izvestiya AN SSSR. Seriya matematicheskaya = Mathematics of the USSR-Izvestiya*, 1984, vol. 48, no. 6, pp. 1314–1343. (In Russian).

19. Tyurin Yu.N., Savvushkina N.E. Kriterii soglasiya dlya raspredeleniya Veibulla–Gnedenko [Goodness of fit for the Weibull–Gnedenko distribution]. *Izvestiya AN SSSR. Seriya: Tekhnicheskaya kibernetika = Soviet journal of computer and systems sciences*, 1984, no. 3, pp. 109–112. (In Russian).

20. Lemeshko B.Yu., Postovalov S.N. O zavisimosti raspredelenii statistik neparametricheskikh kriteriev i ikh moshchnosti ot metoda otsenivaniya parametrov [On the dependence of the distributions of statistics of nonparametric tests and their power on the parameter estimation method]. *Zavodskaya laboratoriya. Diagnostika materialov = Industrial laboratory. Materials diagnostics*, 2001, vol. 67, no. 7. pp. 62–71. (In Russian).

21. R 50.1.037–2002. *Rekomendatsii po standartizatsii. Prikladnaya statistika. Pravila proverki soglasiya opytnogo raspredeleniya s teoreticheskim*. Ch. 2. *Neparametricheskie kriterii* [Recommendations for standardization. Applied statistics. Rules for checking the agreement of the experimental distribution with the theoretical one. Pt. 2. Nonparametric tests]. Moscow, Standards Publ., 2002. 64 p.

22. Lemeshko B.Yu., Lemeshko S.B. Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Pt. 1. *Measurement Techniques*, 2009, vol. 52, no. 6, pp. 555–565.

23. Lemeshko B.Yu., Lemeshko S.B. Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Pt. 2. *Measurement Techniques*, 2009, vol. 52, no. 8, pp. 799–812.

24. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N. Statistic distribution models for some nonparametric goodness-of-fit tests in testing composite hypotheses. *Communications in Statistics. Theory and Methods*, 2010, vol. 39, no. 3, pp. 460–471.

25. Lemeshko B.Yu., Lemeshko S.B. Models of statistic distributions of nonparametric goodness-of-fit tests in composite hypotheses testing for double exponential law cases. *Communications in Statistics. Theory and Methods*, 2011, vol. 40, no. 16, pp. 2879–2892.

26. Lemeshko B.Yu., Gorbunova A.A. Application of nonparametric Kuiper and Watson tests of goodness-of-fit for composite hypotheses. *Measurement Techniques*, 2013, vol. 56, no. 9, pp. 965–973.

27. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Chimitova E.V *Statisticheskii analiz dannykh, modelirovanie i issledovanie veroyatnostnykh zakonomernostei. Komp'yuternyi podkhod* [Statistical data analysis, simulation and study of probability regularities. computer approach]. Novosibirsk, NSTU Publ., 2011. 888 p.

28. Lemeshko B.Yu., Lemeshko S.B., Rogozhnikov A.P. Interactive investigation of statistical regularities in testing composite hypotheses of goodness of fit. *Statistical models and methods for reliability and survival analysis*. Ed. by V. Couallier, L. Gerville-Reache, C. Huber-Carol. Hoboken, NJ, Wiley-ISTE, 2013, ch. 5, pp. 61–76.

29. Lemeshko B.Yu., Lemeshko S.B., Blinov P.Yu., Veretel'nikova I.V., Novikova A.Yu. *Statisticheskii analiz interval'nykh nablyudenii odnomernykh nepreryvnykh sluchainykh velichin "Interval'naya statistika 5.4"* [Statistical analysis of interval observations of one-dimensional continuous random variables "Interval statistics 5.4"]. The certificate on official registration of the computer program. No. 2018666213, 2018. Available at: https://ami.nstu.ru/~headrd/ISW.htm. (accessed 22.01.2021).

30. Lemeshko B.Yu. Lemeshko S.B., Semenova M.A. K voprosu statisticheskogo analiza bol'shikh dannykh [To question of the statistical analysis of big data]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika = Tomsk State University Journal of Control and Computer Science*, 2018, no. 44. pp. 40–49. DOI: 10.17223/19988605/44/5.

31. Lemeshko B.Yu., Blinov P.Yu. *Kriterii proverki otkloneniya raspredeleniya ot ravnomernogo zakona: rukovodstvo po primeneniyu* [Tests for checking the deviation from uniform distribution law. Guide on the application]. Moscow, Infra-M Publ., 2015. 183 p. DOI: 10.12737/11304.

32. Lemeshko B.Yu. *Kriterii proverki otkloneniya raspredeleniya ot normal'nogo zakona: rukovodstvo po primeneniyu* [Tests for checking the deviation from normal distribution law. Guide on the application]. Moscow, Infra-M Publ., 2015. 160 p. DOI: 10.12737/6086.

33. Lemeshko B.Yu., Blinov P.Yu. *Kriterii proverki otkloneniya ot eksponentsial'nogo zakona: rukovodstvo po primeneniyu* [Tests for checking the deviation from exponential distribution law. Guide on the application]. Moscow, Infra-M Publ., 2021. 352 p. DOI 10.12737/1097477.

34. Lemeshko B.Yu. *Kriterii proverki gipotez ob odnorodnosti: rukovodstvo po primeneniyu* [Tests for homogeneity. Guide on the application]. 2nd ed., rev. and add. Moscow, Infra-M Publ., 2021. 248 p. DOI 10.12737/986695.

35. Pearson E.S., D'Agostino R.B., Bowman K.O. Tests for departure from normality: comparison of powers. *Biometrika*, 1977, vol. 64, pp. 231–246. DOI: 10.1093/biomet/64.2.427-a.

36. Tricker A.R. The effect of rounding on the significance level of certain normal test statistics. *Journal of Applied Statistics*, 1990, vol. 17, no. 1, pp. 31–38. DOI: 10.1080/757582644.

37. Tricker A.R. The effect of rounding on the power level of certain normal test statistics. *Journal of Applied Statistics*, 1990, vol. 17, no. 2, pp. 219–228. DOI: 10.1080/757582833.

38. Deidda R., Puliga M. Sensitivity of goodness-of-fit statistics to rainfall data rounding off. *Physics and Chemistry of the Earth*, 2006, vol. 31, pp. 1240–1251. DOI: 10.1016/j.pce.2006.04.041.

39. Lemeshko B.Yu., Lemeshko S.B. Effect of the roundoff on the properties of criteria for testing statistical hypotheses. *Optoelectronics, Instrumentation and Data Processing*, 2020, vol. 56, no. 3, pp. 35–45. DOI: 10.3103/S8756699020030103. Translated from *Avtometriya*, 2020, vol. 56, no. 3, pp. 35–45. DOI: 10.15372/AUT20200305.

40. Lemeshko B.Yu., Lemeshko S.B. O vliyanii oshibok okrugleniya na raspredeleniya statistik kriteriev soglasiya [About the influence of rounding errors on distributions of statistics of the goodness-of-fit tests]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika = Tomsk State University Journal of Control and Computer Science*, 2020, no. 53. pp. 47–60. DOI: 10.17223/19988605/53/5.

41. Lemeshko B.Y., Lemeshko S.B. About the effect of rounding on the properties of tests for testing statistical hypotheses. *Journal of Physics: Conference Series*, 2021, vol. 1715, p. 012063. DOI: 10.1088/1742-6596/1715/1/012063.

Для цитирования:

Лемешко Б.Ю., Лемешко С.Б. Проблемы применения непараметрических критериев согласия в задачах обработки результатов измерений // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 47–66. – DOI: 10.17212/2782-2001-2021-2-47-66.

For citation:

Lemeshko B.Yu., Lemeshko S.B. Problemy primeneniya neparametricheskikh kriteriev soglasiya v zadachakh obrabotki rezul'tatov izmerenii [Problems of nonparametric goodness-of-fit test application in tasks of measurement results processing]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 47–66. DOI: 10.17212/2782-2001-2021-2-47-66.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 67–82

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 67–82

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 004.85 DOI: 10.17212/2782-2001-2021-2-67-82

Построение робастных нейронных сетей с различными функциями потерь*

М.А. СИВАК[a], В.С. ТИМОФЕЕВ[b]

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

[a] *pepelyaeva@ami.nstu.ru* [b] *v.timofeev@corp.nstu.ru*

В статье рассмотрена задача построения устойчивых модификаций нейронных сетей с использованием различных робастных функций потерь. Применение таких нейронных сетей целесообразно при работе с зашумленными данными и может служить альтернативой предварительной очистке выборки или усложнению архитектуры сети. Для корректной работы алгоритма обучения нейронной сети (алгоритма обратного распространения ошибки) требуется, чтобы функция потерь была непрерывно или бесконечно дифференцируема. В соответствии с этим ограничением было выбрано пять робастных функций потерь – Эндрюса, Уэлша, Хьюбера, Рамсея и Fair. Использование этих функций в алгоритме обратного распространения ошибки вместо квадратичной функции потерь позволило получить абсолютно новый класс нейронных сетей. Для исследования свойств построенных сетей проводился ряд вычислительных экспериментов при различной доле засоряющих наблюдений в выборке и различном числе эпох обучения. На первом этапе производилась настройка полученных сетей, то есть выбор таких значений внутренних параметров функции потерь, при которых достигается наивысшая точность работы нейронной сети. Для определения интервалов, на которых рассматривались значения параметров, а также шага разбиения этих интервалов проводилось предварительное исследование. Полученные на первом этапе результаты позволили дать рекомендации по выбору наилучших значений параметров для каждой из рассмотренных функций. На втором этапе проводилось сравнение точности работы построенных робастных сетей как между собой, так и с классической нейронной сетью. Анализ результатов показал, что использование робастного подхода позволяет получить значительный выигрыш в точности работы и в скорости обучения нейронной сети.

Ключевые слова: искусственная нейронная сеть, алгоритм обратного распространения ошибки, выбросы, вычислительный эксперимент, робастный подход, функция потерь, машинное обучение, задача классификации

---

* *Статья получена 18 декабря 2020 г.*

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90077.

ВВЕДЕНИЕ

Задача распознавания образов является одной из классических и наиболее распространенных задач машинного обучения, и зачастую для ее решения применяются искусственные нейронные сети (ИНС). Однако на практике исследователи неизбежно сталкиваются с тем, что имеющиеся данные содержат в себе некоторое количество нетипичных или засоряющих наблюдений (выбросов). Присутствие даже небольшого их числа может сильно сказаться на качестве работы нейронной сети с простой архитектурой [1–3].

Чтобы решить эту проблему, можно воспользоваться несколькими подходами. Самым очевидным вариантом является выполнение предобработки данных и исключение из них этих самых выбросов, однако такой подход может привести к потере важной информации. Кроме того, можно усложнить архитектуру ИНС, но в этом случае появятся дополнительные затраты на вычислительные ресурсы. Альтернативой этим двум методам может служить робастный подход [4, 5], использование которого позволит учесть нетипичные наблюдения и при необходимости снизить их негативное влияние в ходе обучения нейронной сети. В этом случае предполагается введение робастной функции потерь в алгоритме обучения нейронной сети вместо обычно используемой квадратичной функции. В рамках настоящей статьи рассматриваются несколько таких модификаций нейронной сети, полученных за счет использования различных робастных функций потерь.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим в общем виде задачу обучения нейронной сети. Для начала пусть имеется ИНС с одним скрытым слоем, ее архитектура представлена на рис. 1.

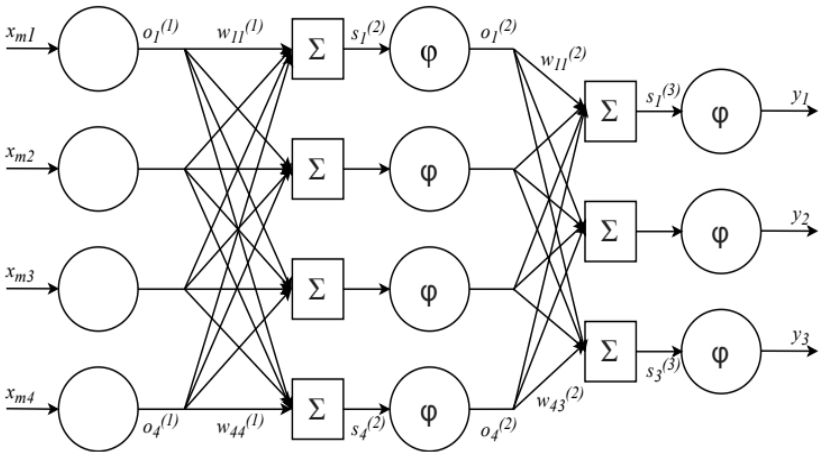


Рис. 1. Искусственная нейронная сеть с одним скрытым слоем

Fig. 1. An artificial neural network with one hidden layer

На этом рисунке использованы следующие обозначения: x_{mi}, $i = 1,..., 4$ – значения, которые подаются на вход сети; y_1, y_2, y_3 – значения нейронов

выходного слоя сети; $w_{ij}^{(1)}$, $i, j = 1,..., 4$, $w_{jk}^{(2)}$, $k = 1,..., 3$ – веса на ребрах между нейронами соседних слоев (первого и второго, второго и третьего соответственно); $s_j^{(2)}$, $s_k^{(3)}$ – входные значения нейронов на втором и третьем слоях соответственно; $o_i^{(1)}, o_j^{(2)}$ – выходные значения нейронов на первом и втором слоях соответственно, φ – функция активации. В рамках настоящей работы в качестве функции активации $\varphi = \varphi(z)$ используется сигмоида [6]

$$\varphi(z) = \frac{1}{1+e^z}.$$

Приведенные выше обозначения без труда можно обобщить на случай ИНС с более сложной архитектурой: x_{mi}, $i = 1,...,T$ – значения, которые подаются на вход сети; y_k, $k = 1,...,Y$ – значения на выходном слое нейронной сети; $w_{ij}^{(n-1)}$, $i = 1,..., l^{(n-1)}$, $j = 1,..., l^{(n)}$ – вес между j-м нейроном слоя n и i-м нейроном слоя n – 1 ($l^{(n)}$ – количество нейронов на слое n); $o_j^{(n)}$ – выходное значение j-го нейрона на слое n, определяющееся следующим образом:

$$o_j^{(n)} = \varphi\left(s_j^{(n)}\right), \tag{1}$$

где $s_j^{(n)}$ – входное значение j-го нейрона на слое n. Если нейрон находится на входном слое сети, то для него будет справедливо

$$s_i^{(1)} = o_i^{(1)} = x_{mi}.$$

В противном случае входное значение нейрона на слое n определяется через выходные значения нейронов на слое n – 1 как

$$s_j^{(n)} = \sum_{i=1}^{l^{(n-1)}} w_{ij}^{(n-1)} o_i^{(n-1)}, \quad n = 2,\ 3,...,\ N, \tag{2}$$

где N – количество слоев нейронной сети. Значения на выходном слое сети будут вычисляться в соответствии с выражением

$$y_k = \varphi\left(s_k^{(N)}\right). \tag{3}$$

Наиболее распространенным алгоритмом обучения ИНС является алгоритм обратного распространения ошибки, подробное описание которого можно найти в [6]. Обучение нейронной сети заключается в решении задачи оптимизации вида

$$E = \sum_{j=1}^{l^{(N)}} f(t_j, y_j) \to \min_{w_{ij}^{(1)},...,w_{ij}^{(N-1)}}, \tag{4}$$

где E – суммарная функция потерь; t_j – требуемый ответ на j-м нейроне выходного слоя сети. Обычно при обучении нейронной сети используется квадратичная функция потерь

$$f(t_j, y_j) = \frac{1}{2}(y_j - t_j)^2 . \tag{5}$$

Далее будем называть классической нейронную сеть, в алгоритме обучения которой используется данная функция.

Однако квадратичная функция потерь является неустойчивой к выбросам, вследствие чего в рамках настоящей работы ставится задача при использовании различных робастных функций потерь получить устойчивые варианты алгоритма обучения ИНС, а также провести сравнительный анализ работы построенных робастных сетей. Кроме того, для каждой из исследуемых функций потерь необходимо определить, при каких значениях их внутренних параметров точность работы сети будет наилучшей.

2. РОБАСТНЫЙ АЛГОРИТМ ОБУЧЕНИЯ

Для решения задачи оптимизации (4) требуется вычислить производную функции потерь по весам нейронной сети. На основе (1) и (3) можно записать следующее цепное правило для вычисления частной производной суммарной функции потерь [6]:

$$E'^{(n)}_{ji} = \frac{\partial E}{\partial w^{(n-1)}_{ij}} = \frac{\partial E}{\partial o^{(n)}_j} \frac{\partial o^{(n)}_j}{\partial \mathrm{s}^{(n)}_j} \frac{\partial \mathrm{s}^{(n)}_j}{\partial w^{(n-1)}_{ij}} .$$

Выполнив несложные преобразования с учетом (1)–(3) и (5), данную производную можно записать в виде

$$E'^{(n)}_{ji} = \delta^{(n)}_j o^{(n-1)}_i ,$$

где $\delta^{(n)}_j$ будет вычисляться следующим образом:

$$\delta^{(n)}_j = \frac{\partial E}{\partial o^{(n)}_j} \frac{\partial o^{(n)}_j}{\partial s^{(n)}_j} = \begin{cases} \dfrac{\partial f(y_j, t_j)}{\partial y_j} y_j (1 - y_j), & n = N, \\ \left(\sum\limits_{k=1}^{l^{(n+1)}} w^{(n)}_{jk} \delta^{(n+1)}_k \right) \varphi\left(\mathrm{s}^{(n)}_j\right)\left(1 - \varphi(\mathrm{s}^{(n)}_j\right) & \text{иначе.} \end{cases} \tag{6}$$

Алгоритм обратного распространения ошибки накладывает ограничение на используемую функцию потерь: она должна быть непрерывно или бесконечно дифференцируема. С учетом этого ограничения было выбрано пять робастных функций потерь [7–11], анализ которых представлен в работе [12]. Эти функции и их частные производные по выходу j-го нейрона на выходном слое приводятся в табл. 1 (здесь и далее β – параметр робастной функции потерь, принимающий неотрицательные значения).

Таблица 1

Table 1

Робастные функции потерь и их производные

Robust loss functions and their derivates

Функция потерь	Производная функции потерь
Хьюбера	
$\begin{cases} \frac{1}{2}(y_j - t_j)^2, & \lvert y_j - t_j \rvert \le \beta \\ \beta \lvert y_j - t_j \rvert - \frac{1}{2}\beta^2, & \lvert y_j - t_j \rvert > \beta \end{cases}$	$\begin{cases} y_j - t_j, & \lvert y_j - t_j \rvert \le \beta \\ -\beta, & y_j - t_j < -\beta \\ \beta, & y_j - t_j > \beta \end{cases}$
Эндрюса	
$\begin{cases} \beta(1 - \cos\frac{(y_j - t_j)}{\beta}), & \lvert y_j - t_j \rvert < \pi\beta \\ 2\beta, & \lvert y_j - t_j \rvert \ge \pi\beta \end{cases}$	$\begin{cases} \sin\frac{(y_j - t_j)}{\beta}, & \lvert y_j - t_j \rvert < \pi\beta \\ 0, & \lvert y_j - t_j \rvert \ge \pi\beta \end{cases}$
Рамсея	
$\frac{1 - (1 + \beta \lvert y_j - t_j \rvert)\exp(-\beta \lvert y_j - t_j \rvert)}{\beta^2}$	$(y_j - t_j)\exp\{-\beta \lvert y_j - t_j \rvert\}$
Fair	
$\beta^2\left(\frac{\lvert y_j - t_j \rvert}{\beta} - \ln\left(1 + \frac{\lvert y_j - t_j \rvert}{\beta}\right)\right)$	$\frac{y_j - t_j}{(1 + \lvert y_j - t_j \rvert / \beta)}$
Уэлша	
$1 - \exp\left(-\frac{1}{2}\left(\frac{y_j - t_j}{\beta}\right)^2\right)$	$\frac{1}{\beta^2}(y_j - t_j)\exp\left(-\frac{1}{2}\left(\frac{y_j - t_j}{\beta}\right)^2\right)$

Заменим (5) на одну из робастных функций потерь из табл. 1, обозначив ее как $f_R(t_j, y_j)$. Тогда (6) примет вид

$$\delta_j^{(n)} = \frac{\partial E}{\partial o_j^{(n)}} \frac{\partial o_j^{(n)}}{\partial s_j^{(n)}} = \begin{cases} \frac{\partial f_R(y_j, t_j)}{\partial y_j} y_j(1 - y_j), & n = N, \\ \left(\sum_{k=1}^{l^{(n+1)}} w_{jk}^{(n)} \delta_k^{(n+1)}\right) \varphi\left(s_j^{(n)}\right)\left(1 - \varphi\left(s_j^{(n)}\right)\right) & \text{иначе.} \end{cases}$$

Таким образом, проводя обучение ИНС с использованием этих функций потерь, получим класс совершенно новых нейронных сетей. Далее приводятся результаты исследований их свойств.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Сравнение точности работы полученных ИНС проводилось на наборе данных «Ирисы Фишера» [13], который включает в себя 150 объектов X_m, m = 1,…,150, в равных долях принадлежащих трем различным классам y_k,

k = 1,…,3, и описываемых четырьмя признаками x_{mi}, i = 1,…,4. Выбранный набор данных является сравнительно небольшим – за счет этого появляется возможность отслеживать закономерности в данных и анализировать результат работы сети.

При проведении исследований описанный набор данных делился на обучающую и тестовую выборки. Обучающая выборка $L = \{X_1,..., X_{|L|}\}$ – это набор объектов, используемый для обучения нейронной сети. В данном случае в нее вошло 80 % данных исходного набора ($|L|$ = 120 объектов – объем обучающей выборки). Оставшиеся 20 % объектов вошли в тестовую выборку $D = \{X_{|L|+1},..., X_{|X|}\}$, которая использовалась для оценки эффективности работы нейронной сети ($|D|$ = 30 объектов – объем тестовой выборки).

В данном случае все классы являются сбалансированными, поэтому для оценки точности работы построенных нейронных сетей использовалась только метрика *accuracy* [14] – отношение числа правильно классифицированных объектов тестовой выборки к объему тестовой выборки:

$$\alpha = \frac{D_{corr}}{|D|} 100\ \%,$$

где D_{corr} – количество объектов, отнесенных к правильному классу.

На рис. 2 для каждой пары признаков приводятся диаграммы рассеивания, позволяющие наглядно увидеть влияние признаков на взаимное расположение классов.

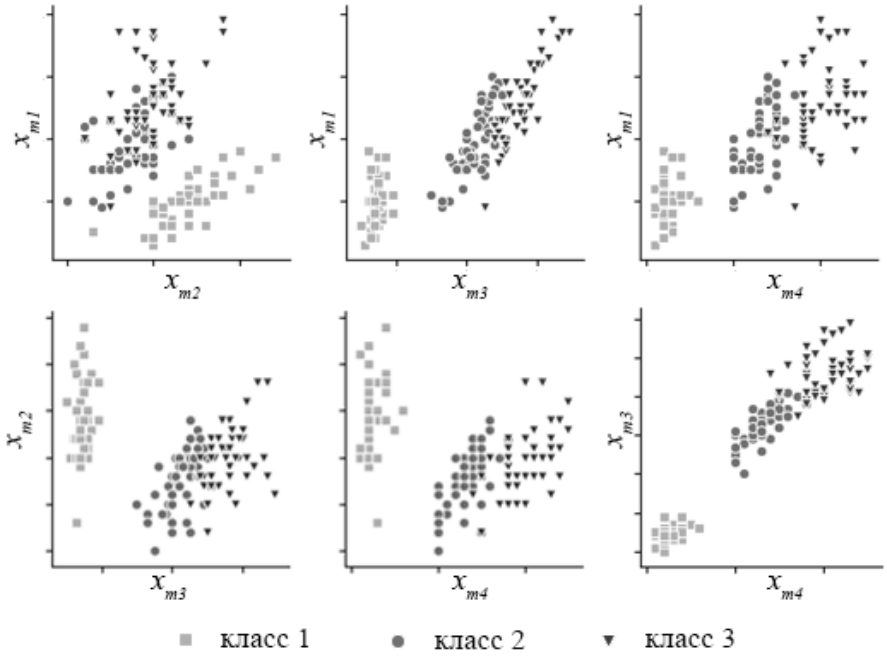


Рис. 2. Попарные графики признаков объектов набора «Ирисы Фишера»

Fig. 2. Pair plots of the Fisher Iris observations attributes

Из рис. 2 видно, что наиболее важными с точки зрения различия объектов являются третий и четвертый признаки, поэтому зашумление обучающей выборки выполнялось по ним:

$$\tilde{x}_{mi} = x_{mi} + \varepsilon_{mi}, \quad i = 3,\ 4,$$

где ε_{mi} – случайные, независимые, одинаково распределенные ошибки. Эти ошибки имеют следующую функцию распределения:

$$F_i(x) = (1-\lambda)F_1(x,\ 0,\ \sigma_{i1}) + \lambda F_2(x,\ 0,\ \sigma_{i2}), \quad i = 3,\ 4,$$

где $F_j(x,\ 0,\ \sigma_{ij}),\ j = 1,\ 2,$ – функция нормального распределения с нулевым математическим ожиданием и дисперсией σ_{ij}^2, $\lambda \in [0,\ 1]$ – параметр смеси, играющий роль доли засоряющих наблюдений.

В рамках настоящей работы полагалось, что $\sigma_{i1}^2 < \sigma_{i2}^2$, причем задавались не сами значения дисперсий σ_{i1}^2 и σ_{i2}^2, а соответствующие им значения уровня шума. Уровень шума [15] определялся как

$$\rho_{ij} = \frac{\sigma_{ij}}{c} 100\ \%,$$

где c^2 – дисперсия незашумленной выборки.

При проведении исследований для третьего признака дисперсия σ_{31}^2 соответствовала уровню шума $\rho_{31} = 30\ \%$, а дисперсия σ_{32}^2 – уровню шума $\rho_{32} = 120\ \%$; для четвертого признака дисперсия σ_{41}^2 соответствовала уровню шума $\rho_{41} = 40\ \%$, а дисперсия σ_{42}^2 – уровню шума $\rho_{42} = 150\ \%$. Значение λ доли засоряющих наблюдений в обучающей выборке варьировалось в пределах от 0,05 до 0,40 с шагом в 0,05.

На первом этапе исследований необходимо было выполнить настройку полученных робастных нейронных сетей, то есть определить наилучшие значения параметра робастных функций потерь β. Иными словами, это такие значения, при которых точность работы сети была бы наилучшей; значения параметра варьировались на различных интервалах $(\beta_{\min},\ \beta_{\max}]$.

Для определения этих интервалов было проведено предварительное исследование. Для всех функций из табл. 1 были зафиксированы левая граница интервала $\beta_{\min} = 0{,}00$, правая граница $\beta_{\max} = 15{,}00$ и шаг разбиения 0,10. Значение доли засоряющих наблюдений было зафиксировано в точке $\lambda = 0{,}25$. Для каждого значения β на этом интервале проводилось по 10 вычислительных экспериментов, в ходе которых фиксировалось значение метрики α на 500 эпох обучения. Полученные результаты затем усреднялись.

Проведенное исследование позволило определить правую границу интервала – такое значение β, при котором точность работы сети перестает значимо изменяться. Кроме того, для функций потерь Хьюбера и Fair был скорректирован шаг разбиения интервала. В первом случае шаг разбиения 0,10 оказался большим для выбранного интервала, вследствие этого не всегда

удавалось найти наилучшее значение параметра. Во втором случае данный шаг разбиения, наоборот, оказался слишком маленьким – такое разбиение привело бы к значительному увеличению вычислительных затрат при поиске наилучших значений β. Полученные в результате предварительного исследования границы и шаг разбиения интервалов приводятся в табл. 2.

Таблица 2

Table 2

Интервалы, на которых рассматривались значения параметра робастных функций потерь

Ranges of the robust loss function parameter values

Функция потерь	β_{min}	β_{max}	Шаг разбиения
Эндрюса	0,00	5,00	0,10
Уэлша	0,00	8,00	0,10
Хьюбера	0,00	1,00	0,05
Рамсея	0,00	5,00	0,10
Fair	0,00	10,00	0,50

Далее исследования проводились следующим образом. Для всех обозначенных выше значений доли засоряющих наблюдений при различном числе эпох обучения нейронной сети (50, 100 и далее до 1000 с шагом в 100 эпох) фиксировалось значение метрики α, полученное для каждого из значений параметра β на интервалах из табл. 2. Соотношение количества эпох обучения и значения метрики α позволяет судить о скорости обучения ИНС (очевидно, что чем меньше количество эпох и чем больше при этом значение α, тем быстрее обучается нейронная сеть).

Для всех робастных сетей выполнялось по 100 вычислительных экспериментов, результаты которых затем усреднялись. После этого для каждой исследуемой ИНС были выбраны наилучшие значения параметра функции потерь β. В табл. 3 приводятся эти значения при различной доле выбросов на 100, 300, 500 и 1000 эпох обучения.

Таблица 3

Table 3

Значения параметра робастной функции потерь, при которых достигается наилучшая точность работы сетей

Values of the robust loss functions parameters that maximize the accuracy of neural networks

λ	Число эпох	Функция потерь				
		Эндрюса	Уэлша	Хьюбера	Рамсея	Fair
0,05	100	1,5	1,1	0,55	0,6	1,5
	300	5,0	2,3	0,5	0,8	1,5
	500	5,0	3,1	0,45	1,4	0,5
	1000	5,0	3,5	0,3	1,7	0,1
0,10	100	1,6	1,2	0,55	0,6	1,0
	300	5,0	2,6	0,25	1,6	0,5

Окончание табл. 3

End of Tab. 3

λ	Число эпох	Функция потерь				
		Эндрюса	Уэлша	Хьюбера	Рамсея	Fair
	500	5,0	3,2	0,25	1,7	0,5
	1000	5,0	3,9	0,2	1,8	0,1
0,15	100	1,7	1,2	0,45	0,8	1,0
	300	4,9	2,4	0,3	1,4	0,5
	500	5,0	3,2	0,25	1,7	0,5
	1000	5,0	3,7	0,25	1,7	0,1
0,20	100	4,8	2,1	0,4	1,1	0,5
	300	5,0	2,5	0,25	1,6	0,5
	500	5,0	3,2	0,2	1,7	0,5
	1000	5,0	4,1	0,25	1,7	0,1
0,25	100	2,3	1,5	0,4	0,9	0,5
	300	5,0	2,5	0,25	1,6	0,5
	500	5,0	2,9	0,25	1,7	0,1
	1000	5,0	3,7	0,3	1,7	0,1
0,30	100	2,4	1,5	0,45	0,9	0,5
	300	4,7	2,2	0,25	1,6	0,5
	500	5,0	2,7	0,25	1,6	0,5
	1000	5,0	4,4	0,35	1,4	0,5
0,35	100	2,9	1,6	0,4	1,0	0,5
	300	5,0	2,4	0,25	1,6	0,5
	500	5,0	3,1	0,25	1,7	0,5
	1000	5,0	3,7	0,25	1,5	0,5
0,40	100	2,5	0,5	0,4	1,0	0,5
	300	5,0	2,4	0,25	1,6	0,5
	500	5,0	2,8	0,25	1,6	0,5
	1000	5,0	4,6	0,35	1,5	0,1

Приведенные результаты показывают, что практически у всех функций потерь, за исключением функции Хьюбера, явная зависимость между значением параметра β и долей засоряющих наблюдений фактически не прослеживается. Для функции потерь Хьюбера при сравнительно небольшой доле выбросов ($\lambda \le 0{,}15$) характерно убывание значения β с ростом числа эпох обучения. При большей доле выбросов значение параметра с ростом числа эпох начинает колебаться либо не изменяется ($\lambda = 0{,}35$).

Для остальных рассмотренных функций наблюдается зависимость значения параметра от числа эпох обучения. Так, в случае функций потерь Эндрюса и Уэлша с ростом числа эпох значение параметра, при котором достигается наилучшая точность работы сети, возрастает. Для нейронной сети с функцией потерь Рамсея при доле выбросов $\lambda \le 0{,}25$ наилучшее значение β возрастает с ростом числа эпох. При большей доле выбросов и числе эпох обучения до 500 оно так же возрастает, а при числе эпох от 501 до 1000, наоборот, начинает убывать. В случае функции потерь Fair наилучшее значе-

ние параметра с ростом числа эпох обучения чаще всего убывает (либо не изменяется при $\lambda = 0{,}30$ и $\lambda = 0{,}35$).

Зависимость наилучшего значения параметра β от числа эпох обучения особенно ярко выражена для функции потерь Уэлша, что продемонстрировано на рис. 3, где показана зависимость точности работы сети от значения параметра β при различном числе эпох для доли засоряющих наблюдений 0,25.

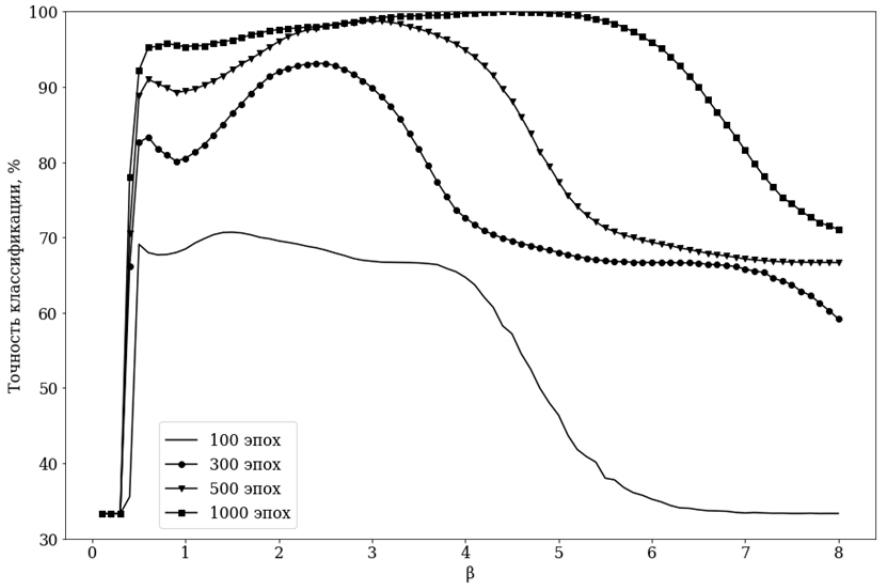


Рис. 3. Зависимость наилучшего значения параметра функции потерь Уэлша от числа эпох обучения

Fig. 3. Dependency of the optimal value of the Welsch loss function parameter on the learning epoch number

На основе полученных результатов можно дать следующие рекомендации относительно выбора значений параметра β для исследуемых нейронных сетей:

- функция потерь Эндрюса: до 100 эпох обучения – на отрезке [1,5; 2,9], более 100 эпох – на отрезке [4,7; 5,0];
- функция потерь Уэлша: до 300 эпох обучения – на отрезке [0,5; 2,1], от 301 до 500 эпох обучения – на отрезке [2,3; 3,2], более 500 эпох обучения – на отрезке [3,5; 4,6];
- функция потерь Хьюбера: при малой доле выбросов ($\lambda \le 0{,}05$) или при малом числе эпох (до 100) – на отрезке [0,4; 0,55], при числе эпох больше 100 или при большей доле выбросов ($\lambda \ge 0{,}10$) – на отрезке [0,2; 0,35];
- функция потерь Рамсея: до 100 эпох обучения – на отрезке [0,6; 1,0], больше 100 эпох – на отрезке [0,8; 1,7];
- функция потерь Fair: до 300 эпох обучения – на отрезке [0,5; 1,5], более 300 эпох обучения – на отрезке [0,1; 0,5].

На втором этапе исследований проводилось сравнение точности работы полученных робастных ИНС между собой, а также с классической сетью. Для этого рассматривались значения метрики α, полученные при наилучших

значениях параметра β. В табл. 4 приводятся результаты для всех обозначенных ранее значений доли засоряющих наблюдений на отметках в 100, 300 и 500 эпох обучения. При большем числе эпох точность работы всех сетей резко возрастала и в отдельных случаях достигала значений 98…99 %, что однозначно говорит о переобучении. При сравнении точности работы построенных сетей рассматривать такие случаи нецелесообразно.

Таблица 4

Table 4

Точность работы нейронных сетей с различными функциями потерь при разных значениях доли выбросов

Accuracy of neural networks with different loss functions for various outlier fraction values

λ	Число эпох	Функция потерь					
		Эндрюса	Уэлша	Хьюбера	Рамсея	Fair	Квадратичная
0,05	100	76,6	77,0	76,7	77,1	76,9	75,0
	300	91,2	91,3	90,7	90,4	90,4	89,2
	500	96,6	96,8	94,3	95,0	95,1	91,4
0,10	100	72,3	72,3	71,9	72,2	72,2	71,6
	300	89,4	89,8	83,2	83,9	83,2	77,8
	500	92,7	95,6	91,3	91,6	89,8	83,5
0,15	100	71,0	71,3	71,5	71,5	71,5	69,8
	300	86,2	86,6	84,8	84,6	84,4	77,8
	500	91,0	94,3	90,6	90,2	88,8	82,4
0,20	100	68,2	68,0	67,7	68,0	68,0	67,2
	300	79,5	79,9	78,6	78,7	77,9	72,6
	500	88,7	91,2	89,1	88,9	86,0	78,9
0,25	100	70,7	70,7	69,3	70,0	70,1	67,3
	300	92,7	93,1	87,3	88,3	87,1	76,6
	500	97,1	98,7	94,0	94,7	94,0	85,2
0,30	100	69,0	69,1	67,3	68,1	68,4	66,3
	300	87,4	87,6	85,1	86,3	85,5	76,4
	500	93,9	95,5	92,2	92,7	91,7	84,8
0,35	100	69,6	69,6	67,7	68,7	68,9	65,6
	300	88,5	89,5	84,1	84,2	82,4	74,7
	500	92,5	96,0	90,8	91,3	90,0	81,8
0,40	100	69,3	69,4	69,0	69,2	69,1	67,0
	300	86,3	87,0	84,9	85,7	82,8	74,3
	500	93,7	95,4	90,7	91,8	89,9	80,3

Из табл. 4 видно, что при сравнительно малом числе эпох (100) точность работы, а следовательно, и скорость обучения робастных нейронных сетей сопоставимы с показателями классической нейронной сети. Кроме того, можно заметить, что при малой доле засоряющих наблюдений (λ = 0,05) даже с увеличением числа эпох до 300 робастные нейронные сети работают незна-

чительно точнее классической (на 1,6 % в среднем), а если увеличить число эпох до 500, то робастные сети начинают работать точнее в среднем на 4,2 %.

Однако с возрастанием доли засоряющих наблюдений до 0,10 уже на 300 эпохах нейронные сети, при построении которых используются робастные функции потерь, обучаются быстрее и показывают точность в среднем на 8,1 % выше, чем классическая нейронная сеть. При работе с сильно зашумленными выборками (доля выбросов $\lambda = 0{,}35$ и $\lambda = 0{,}40$) робастные сети работают точнее в среднем на 11 % и при 300 эпохах обучения, и при 500.

Для большей наглядности на рис. 4 приводятся графики точности работы для всех исследуемых нейронных сетей на 500 эпох обучения при различной доле засоряющих наблюдений. Данные графики позволяют сделать вывод о том, что все робастные сети дают значительный выигрыш в точности по сравнению с классической.

Нетрудно заметить, что нейронная сеть, в алгоритме обучения которой использовалась функция потерь Уэлша, работает точнее остальных НС при всех значениях доли засоряющих наблюдений. По сравнению с классической ИНС при $\lambda = 0{,}25$ выигрыш в точности на 500 эпох обучения составляет 13,5 %, а при сильном зашумлении выборки (при $\lambda = 0{,}35$ и $\lambda = 0{,}40$) выигрыш составляет 14,2 % и 15,1 % соответственно.

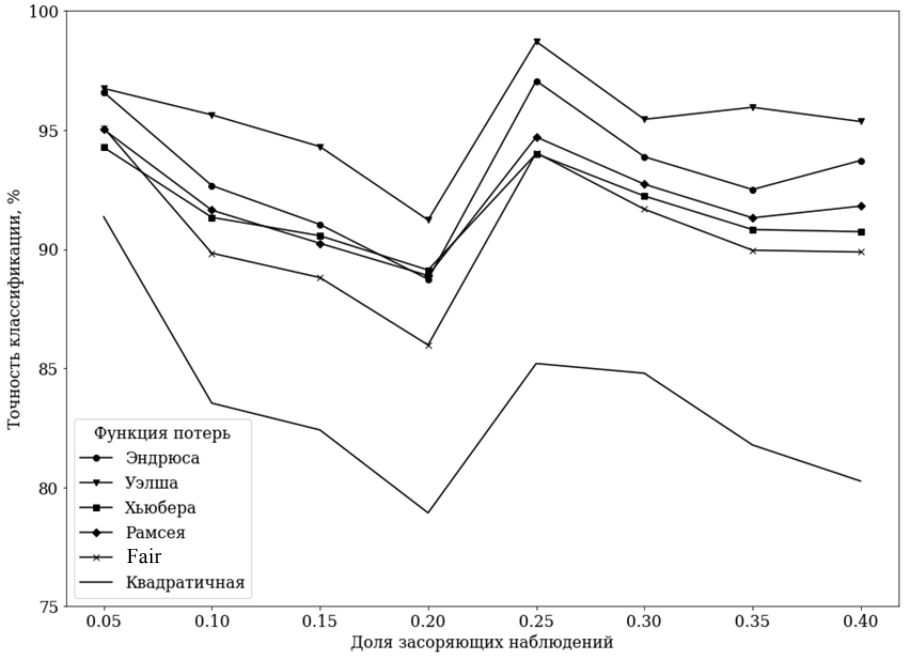


Рис. 4. Точность работы исследуемых НС при различной доле выбросов на 500 эпох

Fig. 4. Accuracy of the investigated neural networks for various outlier fraction values and for 500 epochs

Кроме того, видно, что нейронная сеть, построенная с использованием функции потерь Fair, при значениях λ от 0,10 до 0,25 работает хуже других робастных сетей, но с ростом доли выбросов точность ее работы становится сопоставима с сетями, в алгоритме обучения которых использовались функ-

ции Рамсея и Хьюбера. Нейронная сеть с функцией потерь Эндрюса, наоборот, в большинстве случаев работает несколько лучше нейронных сетей, построенных на основе функций Рамсея, Хьюбера и Fair, и лишь при значении $\lambda = 0{,}20$ точность ее работы становится сопоставима с точностью работы перечисленных выше трех робастных сетей.

Помимо всего прочего, полученные результаты позволяют отчетливо увидеть ухудшение точности работы всех ИНС – и классической, и робастных – при доле засоряющих наблюдений $\lambda = 0{,}20$, а также резкое улучшение точности при доле выбросов $\lambda = 0{,}25$ и последующее ее ухудшение с ростом значения доли выбросов. Этот парадокс можно объяснить конфигурацией выборок, а именно тем, насколько различные классы объектов удалены друг от друга, если за расстояние между классами принять расстояние между центрами масс точек, соответствующих объектам каждого класса. Поскольку зашумление производилось только по 3-му и 4-му признакам объектов, то точки, соответствующие объектам, рассматривались в двумерном евклидовом пространстве: первая координата точки соответствовала значению третьего признака, вторая – значению четвертого. Значение этого расстояния было получено для каждой из ста выборок при значениях доли выбросов $\lambda = 0{,}20$ и $\lambda = 0{,}25$. Вычисленные значения расстояния затем усреднялись. Полученные результаты представлены в табл. 5.

Таблица 5

Table 5

Расстояния между классами объектов при различной доле засоряющих наблюдений

Distances between object classes for various outlier fraction values

λ	Расстояние между классами		
	1-й и 2-й классы	2-й и 3-й классы	1-й и 3-й классы
0,20	3,098	1,399	4,484
0,25	2,946	1,686	4,624

Нетрудно заметить, что с ростом доли выбросов расстояние между первым и вторым классами уменьшилось на 4,9 %, а между первым и третьим классами увеличилось на 3 %. В то же время расстояние между вторым и третьим классами увеличилось на 20,5 %. Исследования показали, что при $\lambda = 0{,}20$ ИНС гораздо чаще «путают» объекты этих классов между собой, чем при $\lambda = 0{,}25$, что и приводит к такому резкому перепаду в точности работы сетей. Фактически ситуация, когда значение доли засоряющих наблюдений равно 0,20, в данном случае эквивалентна большей степени зашумления выборки.

ЗАКЛЮЧЕНИЕ

При выполнении исследований были получены пять робастных нейронных сетей с использованием различных функций потерь (Эндрюса, Уэлша, Рамсея, Хьюбера и Fair). Были исследованы свойства построенных ИНС.

Для этого был выполнен ряд вычислительных экспериментов, в ходе которых рассматривались различные значения доли выбросов, а значение точности работы нейронных сетей фиксировалось при различном числе эпох обучения.

Проведенные исследования позволили сформировать рекомендации по выбору наилучшего значения параметра β для каждой из рассмотренных функций потерь. Кроме того, полученные результаты показали, что при достаточно малой доле выбросов точность работы робастных нейронных сетей сопоставима с классической ИНС. Однако с ростом доли засоряющих наблюдений в выборке применение робастного подхода позволяет добиться значительного выигрыша как в скорости обучения нейронной сети, так и в точности ее работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Ланкин Ю.П., Басканова Т.Ф., Лобова Т.И.* Нейросетевой анализ сложноорганизованных экологических данных // Современные проблемы науки и образования. – 2012. – № 4. – URL: https://www.science-education.ru/ru/article/view?id=6754 (дата обращения: 01.12.2020).

2. *Манжула В.Г., Федяшов Д.С.* Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Фундаментальные исследования. – 2011. – № 4. – С. 108–115. – URL: https://www.fundamental-research.ru/ru/article/view?id=21239 (дата обращения: 05.12.2020).

3. Глубокие нейросети (Ч. 1). Подготовка данных. – URL: https://www.mql5.com/ru/articles/3486 (дата обращения: 26.11.2020).

4. *Fan J., Gijbels I.* Local polynomial modelling and its applications. – London: Chapman & Hall, 1996. – 360 p. – DOI: 10.1201/9780203748725.

5. *Huber J.P.* Robust statistics. – 2nd ed. – Hoboken, NJ: Wiley, 2009. – 370 p. – DOI: 10.1002/9780470434697.

6. *Bishop C.* Neural networks for pattern recognition. – New York: Oxford University Press, 1995. – 502 p.

7. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: исследование зависимостей. – М.: Финансы и статистика, 1985. – 448 с.

8. *Вучков И., Бояджиева Л., Солаков Е.* Прикладной линейный регрессионный анализ. – М.: Финансы и статистика, 1987. – 240 с.

9. *Демиденко Е.З.* Линейная и нелинейная регрессии. – М.: Финансы и статистика, 1981. – 304 с.

10. *Black M.J., Rangarajan A.* On the unification of line processes, outlier rejection, and robust statistics with applications in early vision // International Journal of Computer Vision. – 1996. – Vol. 19. – P. 57–91. – DOI: 10.1007/BF00131148.

11. Fair loss: margin-aware reinforcement learning for deep face recognition / B. Liu, W. Deng, Y. Zhong, M. Wang, J. Hu, X. Tao, Y. Huang // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – Seoul, Korea (South), 2019. – P. 10051–10060. – DOI: 10.1109/ICCV.2019.01015.

12. *Сивак М.А.* Исследование применимости робастных функций потерь в нейронных сетях // Сборник научных трудов НГТУ. – 2020. – № 4 (99). – С. 50–58. – DOI: 10.17212/2307-6879-2020-4-50-58.

13. UCI Machine Learning Repository. – URL: http://www.ics.uci.edu/ mlearn/MLRepository.html (accessed: 03.05.2021).

14. Classification: Accuracy. – URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy (accessed: 03.05.2021).

15. *Ивахненко А.Г., Степашко В.С.* Помехоустойчивость моделирования. – Киев: Наукова думка, 1985. – 216 с.

Сивак Мария Алексеевна, аспирант кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление научных исследований – прикладная математическая статистика, машинное обучение, обработка текстов на естественном языке. E-mail: pepelyaeva@ami.nstu.ru

Тимофеев Владимир Семенович, доктор технических наук, профессор кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление научных исследований – разработка и исследование устойчивых методов и алгоритмов анализа многофакторных объектов, в том числе с использованием непараметрической статистики. E-mail: v.timofeev@corp.nstu.ru

Sivak Maria A., a postgraduate student at the department of theoretical and applied computer science, Novosibirsk State Technical University. The focus of her research is applied mathematical statistics, machine learning, and natural language processing. E-mail: pepelyaeva@ami.nstu.ru

Timofeev Vladimir S., D.Sc. (Eng.), professor of the department of theoretical and applied computer science, Novosibirsk State Technical University. The focus of his research is developing and investigating robust methods and algorithms for a multifactor objects analysis using distribution-free statistics techniques. E-mail: v.timofeev@corp.nstu.ru

DOI: 10.17212/2782-2001-2021-2-67-82

Building robust neural networks using different loss functions[*]

M.A. SIVAK[1], *V.S. TIMOFEEV*[2]

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

[a] *pepelyaeva@ami.nstu.ru* [b] *v.timofeev@corp.nstu.ru*

Abstract

The paper considers the problem of building robust neural networks using different robust loss functions. Applying such neural networks is reasonably when working with noisy data, and it can serve as an alternative to data preprocessing and to making neural network architecture more complex. In order to work adequately, the error back-propagation algorithm requires a loss function to be continuously or two-times differentiable. According to this requirement, two five robust loss functions were chosen (Andrews, Welsch, Huber, Ramsey and Fair). Using the above-mentioned functions in the error back-propagation algorithm instead of the quadratic one allows obtaining an entirely new class of neural networks. For investigating the properties of the built networks a number of computational experiments were carried out. Different values of outliers' fraction and various numbers of epochs were considered. The first step included adjusting the obtained neural networks, which lead to choosing such values of internal loss function parameters that resulted in achieving the highest accuracy of a neural network. To determine the ranges of parameter values, a preliminary study was pursued. The results of the first stage allowed giving recommendations on choosing the best parameter values for each of the loss functions under study. The second stage dealt with comparing the investigated robust networks with each other and with the classical one. The analysis of the results shows that using the robust technique leads to a significant increase in neural network accuracy and in a learning rate.

Keywords: artificial neural network, error back-propagation algorithm, outliers, computational experiment, robust technique, loss function, machine learning, classification problem

[*] *Received 18 December 2020.*

REFERENCES

1. Lankin Yu.P., Baskanova T.F., Lobova T.I. Neirosetevoi analiz slozhnoorganizovannykh ekologicheskikh dannykh [Neural network analysis of complicated ecological data]. *Sovremennye problemy nauki i obrazovaniya = Modern problems of science and education*, 2012, no. 4. Available at: https://www.scienceeducation.ru/ru/article/view?id=6754 (accessed 01.12.2020).

2. Manzhula V.G., Fedyashov D.S. Neironnye seti Kokhonena i nechetkie neironnye seti v intellektual'nom analize dannykh [Kohonen neural networks and fuzzy neural networks in data mining]. *Fundamental'nye issledovaniya = Fundamental research*, 2011, no. 4, pp. 108–115. Available at: https://www.fundamentalresearch.ru/ru/article/view?id=21239 (accessed 05.12.2020).

3. *Glubokie neiroseti* (Ch. 1). *Podgotovka dannykh* [Deep neural networks (Pt. 1). Data preprocessing]. Available at: https://www.mql5.com/ru/articles/3486 (accessed 26.11.2020).

4. Fan J., Gijbels I. *Local polynomial modelling and its applications*. London, Chapman & Hall, 1996. 360 p. DOI: 10.1201/9780203748725.

5. Huber J.P. *Robust statistics*. 2nd ed. Hoboken, NJ, Wiley, 2009. 370 p. DOI: 10.1002/9780470434697.

6. Bishop C. *Neural networks for pattern recognition*. New York, Oxford University Press, 1995. 502 p.

7. Aivazyan S.A., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika: issledovanie zavisimostei* [Applied statistics: dependency investigation]. Moscow, Finansy i statistika Publ., 1985. 448 p.

8. Vuchkov I., Boyadzhieva L., Solakov E. *Prikladnoi lineinyi regressionnyi analiz* [Applied linear regression analyzis]. Moscow, Finansy i statistika Publ., 1987. 240 p.

9. Demidenko E.Z. *Lineinaya i nelineinaya regressii* [Linear and non-linear regressions]. Moscow, Finansy i statistika Publ., 1981. 304 p.

10. Black M.J., Rangarajan A. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 1996, vol. 19, pp. 57–91. DOI: 10.1007/BF00131148.

11. Liu B., Deng W., Zhong Y., Wang M., Hu J., Tao X., Huang Y. Fair loss: margin-aware reinforcement learning for deep face recognition. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 10051–10060. DOI: 10.1109/ICCV.2019.01015.

12. Sivak M.A. Issledovanie primenimosti robastnykh funktsii poter' v neironnykh setyakh [The research on using robust loss functions for neural networks]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Transaction of scientific papers of the Novosibirsk state technical university*, 2020, no. 4 (99), pp. 50–58. DOI: 10.17212/2307-6879-2020-4-50-58.

13. *UCI Machine Learning Repository*. Available at: http://www.ics.uci.edu/ mlearn/ MLRepository.html (accessed 03.05.2021).

14. *Classification: Accuracy*. Available at: https://developers.google.com/machine-learning/ crash-course/classification/accuracy (accessed 03.05.2021).

15. Ivakhnenko A.G., Stepashko V.S. *Pomekhoustoichivost' modelirovaniya* [Noise-resistant modeling]. Kiev, Naukova dumka Publ., 1985. 216 p.

Для цитирования:

Сивак М.А., Тимофеев В.С. Построение робастных нейронных сетей с различными функциями потерь // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 67–82. – DOI: 10.17212/2782-2001-2021-2-67-82.

For citation:

Sivak M.A., Timofeev V.S. Postroenie robastnykh neironnykh setei s razlichnymi funktsiyami poter' [Building robust neural networks using different loss functions]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 67–82. DOI: 10.17212/2782-2001-2021-2-67-82.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 83–94

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 83–94

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 81+81.322+519.25 DOI: 10.17212/2782-2001-2021-2-83-94

Тестирование γ-классификатора, настроенного на распознавание языков произведений на основе латинского алфавита*

З.Д. УСМАНОВ[1,a], А.А. КОСИМОВ[2,b]

[1] *734063, Республика Таджикистан, г. Душанбе, пр. Айни, 299/1, Институт математики им. А. Джураева НАН РТ*
[2] *734042, г. Душанбе, пр. Акад. Раджабовых, 10, Таджикский технический университет имени академика М.С. Осими*

[a] *zafar-usmanov@rambler.ru* [b] *abdunabi_kbtut@mail.ru*

В статье на примере модельной коллекции из десяти текстов на пяти языках (английском, немецком, испанском, итальянском и французском) с использованием латинской графики устанавливается применимость γ-классификатора для автоматического распознавания языка произведения на основе частотности общих 26 латинских алфавитных букв. Математическая модель γ-классификатора представляется в виде триады. Ее первым компонентом является цифровой портрет (ЦП) текста – распределение в тексте частотности буквенных униграмм; вторым компонентом служит формула для вычисления расстояний между ЦП текстов и третьим – алгоритм машинного обучения, реализующий гипотезу «однородности» произведений, написанных на одном языке, и «неоднородности» произведений, написанных на разных языках. Настройка алгоритма, использующего таблицу парных расстояний между всеми произведениями модельной коллекции, заключалась в определении оптимального значения вещественного параметра γ, для которого минимизируется ошибка нарушения гипотезы «однородности». Обученный на текстах модельной коллекции γ-классификатор показал высокую, 100 %-ю точность в распознавании языков произведений. Для тестирования классификатора были выбраны дополнительно шесть случайных текстов, из которых пять на тех же языках, что и тексты модельной коллекции. Методом ближайшего (по расстоянию) соседа все новые тексты подтвердили свою однородность с соответствующими парами одноязычных произведений. Шестой текст на румынском языке показал свою неоднородность по отношению ко всем элементам коллекции. Вместе с тем проявил близость по минимальным расстояниям, прежде всего, к двум текстам на испанском языке и затем и к двум произведениям на итальянском языке.

Ключевые слова: текст, язык, латинская графика, алфавит, частотность униграмм, цифровой портрет текста, гипотеза однородности, классификатор, обучение, распознавание языков, тестирование классификатора, оценка эффективности

* *Статья получена 02 декабря 2020 г.*

ВВЕДЕНИЕ

В наше время письменность на основе латинского алфавита получила широкое распространение среди романской, германской, славянской, финно-угорской, тюркской, семитской и иранской групп языков, среди стран Индокитая, Зондского архипелага и Филиппин, Африки (южнее Сахары), Америки, Австралии и Океании [1]. За исключением современного английского, для большинства других языков латинский алфавит из 26 букв (a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z) оказался недостаточным, в связи с чем для отражения фонетических особенностей тех или иных языковых систем к базовой латинской графике были добавлены различные диакритические знаки, лигатуры и другие модификации букв.

Задача, решением которой будем заниматься в настоящей работе, состоит в том, возможно ли обойтись только лишь 26 латинским буквами для автоматического распознавания языка, на котором написана та или иная печатная продукция.

В качестве экспериментального материала, на котором развертывается наше исследование, выбрана небольшая коллекция ***C*** из десяти произведений (текстов), среди которых

на английском языке **(En)**:

У. Шекспир. Romeo and Juliet (Ромео и Джульетта, **en_1**, 25832 *слов*),

М. Твен. A Connecticut Yankee in King Arthur's Court (Янки из Коннектикута при дворе короля Артура, **en_2**, 117257 *слов*);

на немецком языке **(De)**:

Г. Пиз. Schiff ohne Mannschaft (Корабль без экипажа, **de_1**, 59695 *слов*),

Г. Диана. Das flammende Kreuz: Roman (Пылающий Крест: Роман, **de_2**, 70104 *слов*);

на испанском языке **(Es)**:

Д.Дж. Генрих. El ocaso de la magia (Сумерки магии, **es_1**, 73300 *слов*);

В.Ф. Альберто. Oceano (Океан, **es_2**, 103596 *слов*);

на итальянском языке **(It)**:

Г. Эд. Elminster: la nascita di un mago (Эльминстер: рождение волшебника, **it_1**, 127087 *слов*);

С. Роберт. Il paradosso del passato (Парадокс прошлого, **it_2**, 69697 *слов*);

на французском языке **(Fr)**:

С. Жорж. Lavinia (Лавиния, **fr_1**, 13151 *слов*);

Б. Мишель. Les Nymphéas noirs (Черные водяные лилии, **fr_2**, 108137 *слов*).

Отметим, что сведения о текстах содержат имена авторов, названия их произведений в оригинале и в переводе на русский язык, а также сокращенные обозначения произведений совместно с их размерами, определяемыми количеством слов. Особенность коллекции в том, что она охватывает всего лишь 5 европейских языков, и все ее тексты на основе латинской графики с использованием дополнительных специфических символов: четырех символов ä, ö, ß, ü – в немецком (**de**), одного символа ñ – в испанском (**es**), десяти символов à, è, é, ì, í, î, ò, ó, ù, ú – в итальянском (**it**) и четырнадцати символов â, à, ç, é, ê, è, ë, î, ï, ô, û, ù, ü, ÿ – во французском (**fr**) языках.

Приступая к решению поставленной задачи, отметим, что в качестве исследовательского инструмента мы будем использовать *математическую*

триаду в составе цифрового портрета текстов, представляемых распределениями частотности 26 латинских букв, формулы для вычисления расстояний между текстом и алгоритмом для выявления однородных текстов [2]. Упомянутая триада с момента своего появления в 2017 году применялась, прежде всего, при распознавании авторства для различных вариантов ЦП текстов [3–13]. В дополнение к сказанному уместно отметить, что в моногра-фии [14] представлен обширный обзор работ по идентификации авторов текста на основе разнообразных цифровых портретов текстов и применяемых методов классификации.

1. ЦИФРОВОЙ ПОРТРЕТ ПРОИЗВЕДЕНИЙ

В качестве учетных элементов для описания произведений взяты указанные для всех пяти языков 26 латинских букв.

Определение 1. *Цифровым портретом (ЦП) текста будем называть распределение в нем частотности 26 букв.*

ЦП текста T записывается в табличном виде:

$$\begin{array}{lllll} N: & 1 & 2 & \ldots & 26 \\ P: & p_1 & p_2 & \cdots & p_{26}, \end{array} \tag{1}$$

в котором первая строка – номера букв, расположенных в алфавитном порядке, а вторая – относительные частотности букв в тексте *T*, причем $\sum_{k=1}^{26} p_k = 1$.

Одновременно с (1) цифровой портрет представляется также в виде дискретной функции

$$F(s) = \sum_{k=1}^{S} p_k \qquad (s = 1, \ldots, 26). \tag{2}$$

2. РАССТОЯНИЯ МЕЖДУ ЦИФРОВЫМИ ПОРТРЕТАМИ ТЕКСТОВ

Пусть T_1, T_2 – произвольная пара текстов, характеризуемых на основе единого алфавита, и

$$F^{(\alpha)}(s) = \sum_{k=1}^{S} p_k^{(\alpha)} \tag{3}$$

– соответствующие им ЦП, представленные дискретными функциями, $\alpha = 1, 2,$ и $(s = 1, \ldots, 26)$.

Определение 2. *Расстоянием между текстами* T_1 *и* T_2 *называется положительное число* $\rho(T_1, T_2)$, *определяемое формулой*

$$\rho(T_1, T_2) = \sqrt{\frac{26}{2}} \max_S \left| F^{(1)}(s) - F^{(2)}(s) \right|. \tag{4}$$

3. ГИПОТЕЗА *H* «ОДНОРОДНОСТИ» ПРОИЗВЕДЕНИЙ

Она привлекается для того, чтобы выделить характерную особенность текстов, предназначенную для построения математической модели распознавания языка произведений. Ее мы формулируем в следующем виде.

ГИПОТЕЗА *H*. *Произведения, написанные на одном языке, «однородны», а на разных языках – «не однородны».*

Говоря об «однородности» произведений (текстов), мы имеем в виду их похожесть, одинаковость, сходность, однотипность, родственность и т. п.

4. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ *H*-ГИПОТЕЗЫ

Пусть γ – некоторое положительное число.

Определение 3. *Тексты называются γ-однородными (написанными на одном языке), если*

$$\rho(T_1, T_2) \le \gamma, \tag{5}$$

и γ-неоднородными (написанными на разных языках), если

$$\rho(T_1, T_2) > \gamma. \tag{6}$$

Неравенства (5) и (6) являются математической интерпретацией (моделью) гипотезы *H*. Это значит, что в дальнейшем мы приступаем к распознаванию языков произведений с помощью математического аппарата, названного γ-классификатором [2].

Определение 4. *γ-классификатор – это зависящий от одного вещественного параметра γ алгоритм принятия решения об отнесении пары текстов T_1 и T_2 к одному или двум разным языкам.*

Очевидно, что от значения γ зависит однородность или неоднородность любой пары текстов, следовательно, и степень выполнимости гипотезы. Принадлежность двух текстов одному языку в рамках математической модели означает справедливость неравенства (5), а двум разным языкам – справедливость неравенства (6). Гипотеза *H* может нарушаться для каких-то пар текстов одного и того же языка в случае, когда вместо неравенства (5) имеет место неравенство (6), а также в случае, когда какие-то два текста на разных языках удовлетворяют неравенству (5) вместо того, чтобы выполнялось неравенство (6).

Пусть $\tau = \tau(\gamma)$ – суммарное количество нарушений гипотезы *H* одновременно в двух случаях: при невыполнении неравенства «однородности» в случае двух текстов, принадлежащих одному языку, и невыполнении неравенства «неоднородности» в случае двух текстов, принадлежащих разным языкам. Тогда для фиксированного γ показатель выполнения гипотезы будет определяться величиной π, задаваемой формулой

$$\pi = 1 - \tau(\gamma) / L,$$

где L – число взаимных расстояний между всеми парами текстов из коллекции $\boldsymbol{C}$ (в нашем случае $L = C_{10}^2 = 45$. Из этой формулы следует, что π может

принимать значения из отрезка [0, 1], причем $\pi = 0$, если $\tau = L(= 45)$, и $\pi = 1$, если $\tau = 0$. В первом случае гипотезу *H* следует признать непригодной, а во втором – полностью согласованной с обучающей выборкой.

В связи с тем, что эффективность γ-классификатора зависит от значения параметра γ, представляет интерес найти такое его значение, при котором π принимает максимальное значение. Именно в этом и заключается суть настройки γ-классификатора на данных обучающей выборки. Если такая настройка будет приемлемой, то можно говорить о решении задачи обучения γ-классификатора и его предрасположенности к распознаванию языков произведений самых разнообразных коллекций.

5. ИТОГОВЫЕ РЕЗУЛЬТАТЫ НА ПРИМЕРЕ КОЛЛЕКЦИИ *C*

В процессе настройки выполняются следующие операции:

– предобработка экспериментальных данных путем удаления из всех произведений коллекции дополнительных сверх латинского алфавита букв;

– вычисления цифровых портретов (1) (частотности 26 латинских букв) для всех десяти произведений модельной коллекции ***C***;

– вычисления по формулам (2), (3) и (4) сорока пяти парных расстояний $\rho(T_1, T_2)$ между произведениями коллекции ***C*** (результаты расчетов приведены в табл. 1);

Таблица 1

Table 1

Расстояния между текстами коллекции *C*

Distances between texts of the collection *C*

Тексты		En		De		Es		It		Fr	
		en_1	en_2	de_1	de_2	es_1	es_2	it_1	it_2	fr_1	fr_2
En	en_1										
	en_2	0.0832									
De	de_1	0.3949	0.3281								
	de_2	0.3817	0.3148	0.0287							
Es	es_1	0.3606	0.3030	0.2845	0.2963						
	es_2	0.3471	0.2895	0.3077	0.2945	0.0450					
It	it_1	0.2486	0.2302	0.2677	0.2579	0.1950	0.1814				
	it_2	0.2426	0.2243	0.2988	0.2928	0.2086	0.1951	0.0378			
Fr	fr_1	0.1354	0.1945	0.3982	0.3849	0.3205	0.2941	0.2628	0.2691		
	fr_2	0.1480	0.1833	0.4038	0.3920	0.3260	0.2995	0.2776	0.2773	0.0299	

– вычисление с помощью алгоритма настройки γ-классификатора оптимального интервала значений γ, для которого величина $\tau = \tau(\gamma)$ суммарного числа случаев нарушения гипотезы *H* достигает минимального значения, и, следовательно, величина π показателя выполнения гипотезы *H* принимает максимальное значение.

По данным табл. 1 получены следующие результаты:

– совокупность всех пар расстояний размещается на отрезке [0.0287, 0.4038], при этом минимальное расстояние реализуется между двумя произведениями **de_1** и **de_2** на немецком языке, а максимальное – между **de_1** на немецком и **fr_2** на французском языках;

– оптимальный полуинтервал значений γ оказывается в пределах

$$\gamma^{\text{опт}} \in [0.0833; 0.1353); \tag{7}$$

в соответствии с определением 3 это значит, что если расстояние $\rho(T_1, T_2)$ между двумя текстами не превосходит значения $\gamma^{\text{опт}}$ из указанного полуинтервала, то пара текстов принадлежит одному и тому же языку (соответствующие расстояния в таблице помечены серым цветом); если же превосходит, то принадлежит разным языкам (соответствующие расстояния оставлены непомеченными);

– отметим, что для всех (без исключений) произведений коллекции ***C*** полностью подтвердилась гипотеза *H* и ее математическая интерпретация в виде определения 3, и потому получено

$$\tau = \tau_{\min} = 0,$$

т. е. ни одно из неравенств (5) и (6) не было нарушено;

– вследствие этого показатель эффективности предложенной в настоящей работе математической модели распознавания языка произведений оказался равным

$$\pi = \pi_{\max} = 1.$$

6. ТЕСТИРОВАНИЕ

Итак, настройка (обучение) γ-классификатора на данных модельной коллекции текстов ***C*** прошла успешно. Для тестирования классификатора выбраны случайным образом 6 текстов:

на английском языке **(En)**:

Дж. Лондон. The Call of the Wild (Зов предков) (Text_En, 31763 *слов*);

на немецком языке **(De)**:

М. Вилли. Die seltsamen Reisen des Marco Polo (Странные путешествия Марко Поло) (Text_De, 126607 *слов*);

на испанском языке **(Es)**:

Д. Арне. Misterioso (Таинственный) (Text_Es, 106835 *слов*);

на итальянском языке **(It)**:

Ш. Боб. Sfida al cielo (Вызов небу) (Text_It, 101154 *слов*);

на французском языке **(Fr)**:

К.С. Доминикович. Fantôme (Призрак) (Text_Fr, 46089 *слов*);

на румынском языке **(Ro)**:

Т.Р. Руэл. Întoarcerea regelui (Возвращение короля) (Text_Ro, 146266 *слов*).

Отметим, что сведения относительно выбранных произведений описаны по той же схеме, что и для элементов коллекции ***C***. В дополнение к предыдущему отметим, что в румынском языке (**ro**) латинский алфавит расширен на 5 символов, ă, â, î, ș. ț.

Для шести произведений, предназначенных для тестирования, построены цифровые портреты (1) и затем для каждого из них по формулам (2)–(4) вычислены расстояния до десяти объектов коллекции ***C***. Соответствующие значения записаны в ячейках табл. 2, расположенных на пересечениях строк и столбцов. Там же серым цветом выделены пары ячеек с минимальными расстояниями между тестируемыми и содержащимися в коллекции произведениями.

Таблица 2

Table 2

Расстояния между текстами коллекции *C* и шестью случайно выбранными тестируемыми произведениями

Distances between the texts of the collection *C* and the 6 random tested works

Тексты		Text_En	Text_De	Text_Es	Text_It	Text_Fr	Text_Ro
En	en_1	0.1592	0.4069	0.3235	0.2378	0.1477	0.2084
	en_2	0.0857	0.3400	0.2659	0.2194	0.1905	0.1760
De	de_1	0.2599	0.0305	0.2659	0.2866	0.4235	0.2723
	de_2	0.2467	0.0489	0.2526	0.2734	0.4103	0.2663
Es	es_1	0.2674	0.3010	0.0552	0.1874	0.3250	0.1707
	es_2	0.2538	0.3197	0.0430	0.1738	0.2985	0.1440
It	it_1	0.1987	0.2882	0.1579	0.0330	0.3050	0.1565
	it_2	0.2365	0.3260	0.1715	0.0281	0.3047	0.1563
Fr	fr_1	0.2802	0.4101	0.2712	0.2501	0.0460	0.1933
	fr_2	0.2690	0.4158	0.2767	0.2604	0.0448	0.2033

Полученные результаты показывают, что ближайшими соседями [15, 16] первых пяти произведений оказались как раз однородные с ними по языку пары текстов исходной коллекции. Что касается текста на румынском языке

(Text_Ro), то все его расстояния до десяти коллекционных текстов превзошли максимальное значения $\gamma^{опт}$ (см. формулу (7)). Следовательно, как и ожидалось, для Text_Ro в коллекции не оказалось ни одного однородного объекта. Интересно, однако, отметить, что γ-классификатор указал в качестве ее ближайших соседей два произведения es_1 и es_2 на испанском и два произведения it_1 и it_2 на итальянском языках.

ЗАКЛЮЧЕНИЕ

Итак, γ-классификатор с фиксированным значением $\gamma = \gamma^{опт}$ на случайных выборках текстов с цифровыми портретами на основе частотности 26 базовых латинских букв подтвердил 100 %-ю статистическую способность к распознаванию языков произведений.

Таким образом, математическая триада в составе цифрового портрета текстов, представляемых распределениями частотности 26 латинских букв, формул (1)–(3) для вычисления расстояний между текстами и алгоритмами для выявления однородных текстов оказалась подходящей для эффективного решения поставленной задачи.

Авторы выражают уверенность в том, что увеличение объема исходной коллекции текстов не станет препятствием для успешного применения γ-классификатора не только для распознавания языков, но также и для самых разнообразных однородностей текстовых документов.

СПИСОК ЛИТЕРАТУРЫ

1. Список латинских букв // Википедия. – URL: https://ru.wikipedia.org/wiki/Список_латинских_букв (дата обращения: 13.11.2020).

2. *Усманов З.Д.* Алгоритм настройки кластеризатора дискретных случайных величин // Доклады Академии наук Республики Таджикистан. – 2017. – Т. 60, № 9. – С. 392–397.

3. *Усманов З.Д., Косимов А.А.* О распознавании авторства таджикского текста // Доклады Академии наук Республики Таджикистан. – 2016. – Т. 59, № 3–4. – С. 114–119.

4. *Косимов А.А.* Оценка эффективности использования биграмм при идентификации текста // Доклады Академии наук Республики Таджикистан. – 2017. – Т. 60, № 5–6. – С. 224–229.

5. *Косимов А.А.* Оценка эффективности использования триграмм при идентификации текста // Известия Академии наук Республики Таджикистан. Отделение физико-математических, химических, геологических и технических наук. – 2017. – № 1 (166). – С. 51–57.

6. *Каримов А.А.* О цифровом портрете текстовой информации // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2019. – № 1 (45). – С. 7–10.

7. *Каюмов М.М.* О цифровом портрете текстовой информации, основанном на частотности знаков пунктуации // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2019. – № 1 (45). – С. 20–23.

8. *Каюмов М.М.* О распознавании автора текста на основе частотности $\alpha\beta$-кодов словоформ // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2020. – № 2 (50). – С. 29–36.

9. *Ашурова Ш.Н.* Оценка эффективности использования словесных биграмм при идентификации текста // Материалы международной научно-практической конференции ТУТ «Роль ИКТ в инновационном развитии экономики Республики Таджикистан». – Душанбе: Баҳманрӯд, 2017. – С. 292–297.

10. *Ашурова Ш.Н.* Оценка эффективности использования словесных триграмм при идентификации текста // Вестник Технологического университета Таджикистана. – 2017. – № 4 (31). – С. 51–58.

11. *Ашурова Ш.Н., Тошхуджаев Х.А.* О распознавании автора текста на основе частотности словесных биграмм // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2020. – № 2 (50). – С. 57–61.

12. *Бахтеев К.С.* О применимости укороченных цифровых портретов для идентификации автора текста // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2020. – № 2 (50). – С. 25–28.

13. *Бахтеев К.С.* О распознавании авторства по усеченным цифровым портретам текста // Известия Академии наук Республики Таджикистан. Отделение физико-математических, химических, геологических и технических наук. – 2018. – № 4 (173). – С. 82–92.

14. *Романов А.С., Шелупанов А.А., Мещеряков Р.В.* Разработка и исследование математических моделей, методик и программных средств информационных процессов при идентификации автора текста. – Томск: В-Спектр, 2011. – 188 с.

15. *Воронцов К.В.* Математические методы обучения по прецедентам. – URL: http://www.ccas.ru/voron (дата обращения: 01.11.2020).

16. *Дьяконов А.Г.* Анализ данных, обучение по прецедентам, логические игры, системы WEKA, RapidMiner и MatLab (Практикум на ЭВМ кафедры математических методов прогнозирования): учебное пособие. – М.: ВМК МГУ им. М.В. Ломоносова, 2010. – 278 с.

Усманов Зафар Джураевич, доктор физико-математических наук, профессор, академик НАН РТ, Институт математики им. А. Джураева НАН РТ. E-mail: zafar-usmanov@rambler.ru

Косимов Абдунаби Абдурауфович, кандидат технических наук, старший преподаватель кафедры АСУ факультета ИКТ, Таджикский технический университет имени академика М.С. Осими. E-mail: abdunabi_kbtut@mail.ru

Usmanov Zafar Dzh., Doctor of Physical and Mathematical Sciences, professor, Academician of the National Academy of Sciences of the Republic of Tajikistan, Institute of Mathematics. A. Dzhuraeva National Academy of Sciences of the Republic of Tajikistan. E-mail: zafar-usmanov@rambler.ru

Kosimov Abdunabi A., PhD (Eng.), senior lecturer at the department of ACS, Tajik Technical University named after academician M.S. Osimi. E-mail: abdunabi_kbtut@mail.ru

DOI: 10.17212/2782-2001-2021-2-83-94

Testing the γ-classifier adapted to recognize the languages of works based on the Latin alphabet[*]

Z.D. USMANOV[1,a], *A.A. KOSIMOV*[2,b]

[1] *734063, Republic of Tajikistan, Dushanbe, 299/1 Aini Prospekt, Academician of NAS RT; A. Juraev Institute of Mathematics, the National Academy of Sciences of Tajikistan*
[2] *734042, Dushanbe, 10 Acad. Radjabovykh Prospekt, Tajik Technical University named after acad. M.S. Osimi*

[a] *zafar-usmanov@rambler.ru* [b] *abdunabi_kbtut@mail.ru*

Abstract

Using the example of a model collection of 10 texts in five languages (English, German, Spanish, Italian, and French) using Latin graphics, the article establishes the applicability of the γ-classifier for automatic recognition of the language of a work based on the frequency of 26 common Latin alphabetic letters. The mathematical model of the γ-classifier is represented as a triad. Its first component is a digital portrait (DP) of the text - the distribution of the frequency of alphabetic unigrams in the text; the second component is formulas for calculating the distances between the DP texts and the third is a machine learning algorithm that implements the hypothesis of "homogeneity" of works written in one language and "heterogeneity" of works written in different languages. The tuning of the algorithm using a table of paired distances between all products of the model collection consisted in determining an optimal value of the real parameter γ, for which the error of violation of the "homogeneity" hypothesis is minimized. The γ-classifier trained on the texts of the model collection showed a high, 100% accuracy in recognizing the languages of the works. For testing the classifier, an additional six random texts were selected, of which five were in the same languages as the texts of the model collection. By the method of the nearest (in terms of distance) neighbor, all new texts confirmed their homogeneity with the corresponding pairs of monolingual works. The sixth text in Romanian showed its heterogeneity in relation to all elements of the collection. At the same time, it showed closeness in minimum distances, first of all, to two texts in Spanish and then to two works in Italian.

Key words: text, language, Latin graphics, alphabet, frequency of unigrams, digital portrait of text, hypothesis of homogeneity, classifier, teaching, language recognition, testing of the classifier, performance evaluation

REFERENCES

1. List of Latin letters. *Wikipedia*. (In Russian). Available at: https://ru.wikipedia.org/wiki/Список_латинских_букв (accessed 13.11.2020).

2. Usmanov Z.D. Algoritm nastroiki klasterizatora diskretnykh sluchainykh velichin [Tuning the algorithm of the clasifier of discrete random variables]. *Doklady Akademii nauk Respubliki Tadzhikistan = Reports of the Academy of Sciences of the Republic of Tajikistan*, 2017, vol. 60, no. 9, pp. 392–397.

3. Usmanov Z.D., Kosimov A.A. O raspoznavanii avtorstva tadzhikskogo teksta [On authorship identification of a text written in Tajik]. *Doklady Akademii nauk Respubliki Tadzhikistan = Reports of the Academy of Sciences of the Republic of Tajikistan*, 2016, vol. 59, no. 3–4, pp. 114–119.

[*] *Received 10 December 2020.*

4. Kosimov A.A. Otsenka effektivnosti ispol'zovaniya bigramm pri identifikatsii teksta [Evaluation of the efficiency of using bigrams in text identification]. *Doklady Akademii nauk Respubliki Tadzhikistan = Reports of the Academy of Sciences of the Republic of Tajikistan*, 2017, vol. 60, no. 5–6, pp. 224–229.

5. Kosimov A.A. Otsenka effektivnosti ispol'zovaniya trigramm pri identifikatsii teksta [Evaluation of trigramm use efficiency for a text identification]. *Izvestiya Akademii nauk Respubliki Tadzhikistan. Otdelenie fiziko-matematicheskikh, khimicheskikh, geologicheskikh i tekhnicheskikh nauk = News of the Academy of Sciences of the Republic of Tajikistan. Department of physical, mathematical, chemical, geological and technical sciences*, 2017, no. 1 (166), pp. 51–57.

6. Karimov A.A. O tsifrovom portrete tekstovoi informatsii [About digital portrait of text information]. *Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii = Polytechnic Bulletin. Series: Intelligence. Innovation. Investments*, 2019, no. 1 (45), pp. 7–10.

7. Kayumov M.M. O tsifrovom portrete tekstovoi informatsii, osnovannom na chastotnosti znakov punktuatsii [About digital portrait text information based on the frequency of punctuation]. *Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii = Polytechnic Bulletin. Series: Intelligence. Innovation. Investments*, 2019, no. 1 (45), pp. 20–23.

8. Kayumov M.M. O raspoznavanii avtora teksta na osnove chastotnosti αβ-kodov slovoform [Identifying the author's text based on frequency αβ – formal codes]. *Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii = Polytechnic Bulletin. Series: Intelligence. Innovation. Investments*, 2020, no. 2 (50), pp. 29–36.

9. Ashurova Sh.N. [Assessment of the effectiveness of the use of verbal bigrams in the identification of text]. *Materialy mezhdunarodnoi nauchno-prakticheskoi konferentsii TUT "Rol' IKT v innovatsionnom razvitii ekonomiki Respubliki Tadzhikistan"* [Materials of the international scientific-practical conference HER "The role of ICT in the innovative development of the economy of the Republic of Tajikistan"]. Dushanbe, Bahmanrud Publ., 2017, pp. 292–297. (In Russian).

10. Ashurova Sh.N. Otsenka effektivnosti ispol'zovaniya slovesnykh trigramm pri identifikatsii teksta [Evaluation of the effectiveness of the use of verbal trigrams in text identification]. *Vestnik Tekhnologicheskogo universiteta Tadzhikistana = Bulletin of the Technological University of Tajikistan*, 2017, no. 4 (31), pp. 51–58. (In Russian).

11. Ashurova Sh.N., Toshkhudzhaev Kh.A. O raspoznavanii avtora teksta na osnove chastotnosti slovesnykh bigramm [On recognition of the author of the text based on the frequency of verbal bigrams]. *Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii = Polytechnic Bulletin. Series: Intelligence. Innovation. Investments*, 2020, no. 2 (50), pp. 57–61.

12. Bakhteev K.S. O primenimosti ukorochennykh tsifrovykh portretov dlya identifikatsii avtora teksta [On the applicability of shortened digital portraits to identify the author of the text]. *Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii = Polytechnic Bulletin. Series: Intelligence. Innovation. Investments*, 2020, no. 2 (50), pp. 25–28.

13. Bakhteev K.S. O raspoznavanii avtorstva po usechennym tsifrovym portretam teksta [On the recognition of authorship by truncated digital portraits of text]. *Izvestiya Akademii nauk Respubliki Tadzhikistan. Otdelenie fiziko-matematicheskikh, khimicheskikh, geologicheskikh i tekhnicheskikh nauk = News of the Academy of Sciences of the Republic of Tajikistan. Department of physical, mathematical, chemical, geological and technical sciences*, 2018, no. 4 (173), pp. 82–92.

14. Romanov A.S., Shelupanov A.A., Meshcheryakov R.V. *Razrabotka i issledovanie matematicheskikh modelei, metodik i programmnykh sredstv informatsionnykh protsessov pri identifikatsii avtora teksta* [Development and research of mathematical models, methods and software for information processes in the identification of the author of the text]. Tomsk, V-Spektr Publ., 2011. 188 p.

15. Vorontsov K.V. *Matematicheskie metody obucheniya po pretsedentam* [Mathematical methods of teaching by precedents]. Available at: http://www.ccas.ru/voron (accessed 01.11.2020).

16. D'yakonov A.G. *Analiz dannykh, obuchenie po pretsedentam, logicheskie igry, sistemy WEKA, RapidMiner i MatLab (Praktikum na EVM kafedry matematicheskikh metodov prognozirovaniya)* [Data analysis, training on precedents, logic games, WEKA, RapidMiner and MatLab sys-

tems (Workshop on the computer of the Department of Mathematical Forecasting Methods)]. Moscow, Faculty of CMC, MSU Publ., 2010. 278 p.

Для цитирования:

Усманов З.Д., Косимов А.А. Тестирование γ-классификатора, настроенного на распознавание языков произведений на основе латинского алфавита // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 83–94. – DOI: 10.17212/2782-2001-2021-2-83-94.

For citation:

Usmanov Z.D., Kosimov A.A. Testirovanie γ-klassifikatora, nastroennogo na raspoznavanie yazykov proizvedenii na osnove latinskogo alfavita [Testing the γ-classifier adapted to recognize the languages of works based on the Latin alphabet]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 83–94. DOI: 10.17212/2782-2001-2021-2-83-94.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 95–120

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 95–120

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 519.873 DOI: 10.17212/2782-2001-2021-2-95-120

Модель минимизации дефицита мощности электроэнергетических систем с учетом ограничений по контролируемым сечениям*

Д.В. ЯКУБОВСКИЙ[a], Д.С. КРУПЕНЁВ[b], Д.А. БОЯРКИН[c]

664033, Иркутская область, г. Иркутск, ул. Лермонтова, 130, Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук

[a] *yakubovskii.dmit@mail.ru* [b] *krupenev@isem.irk.ru* [c] *boyarkin_denis@mail.ru*

Устойчивая тенденция к развитию электроэнергетических систем ведет к их постоянному укрупнению и усложнению, появляются новые способы их контроля. В связи с этим существующие модели и комплексы оценки надежности могут работать неполноценно и неэффективно с точки зрения адекватности получаемых результатов. Для оценки текущего состояния существующих моделей и комплексов был проведен обзор и анализ отечественных и зарубежных программно-вычислительных комплексов. В частности, рассматривались входящие в них математические модели минимизации дефицита мощности. В основе настоящей работы рассматривается задача модификации математических моделей минимизации дефицита мощности, используемых при оценке балансовой надежности электроэнергетических систем одного из рассматриваемых комплексов. В качестве модификации математических моделей предлагается отказаться от существующего способа учета пропускных способностей линий и использовать корректный учет максимально допустимого перетока активной мощности в контролируемых сечениях. Отраженная в работе экспериментальная часть касается тестирования вариантов моделей минимизации дефицита мощности, а также предлагаемых модификаций на различных системах, в том числе состоящих из трех и семи зон надежности с вариативным количеством контролируемых сечений и входящих в них линий электропередачи.

По результатам работы предлагаемые модификации показали свою работоспособность и могут быть в дальнейшем использованы, также авторами были получены наиболее адекватные результаты с позиции физических законов функционирования электроэнергетических систем за счет модели минимизации дефицита мощности с квадратичными потерями, которая учитывает ограничения передачи мощности по контролируемым сечениям.

Ключевые слова: электроэнергетическая система, надежность, математическая модель, минимизация дефицита мощности, максимальный допустимый переток, квадратичная модель, контролируемые сечения, балансовая надежность

---

* *Статья получена 19 февраля 2021 г.*

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-08-00550 А.

ВВЕДЕНИЕ

Обеспечение требуемого уровня надежности электроэнергетических систем (ЭЭС) является одной из актуальных задач энергетики и, в свою очередь, способствует сбалансированному развитию экономики. Из-за постоянных изменений в сторону укрупнения и усложнения ЭЭС раннее планирование, своевременная корректировка изменений и резервирование ее элементов являются основными направлениями обеспечения надежности. Совокупность всех мер напрямую влияет на снижение ущерба из-за отказов энергетического оборудования и, как следствие, ограничения поставки электроэнергии потребителям. Однако мероприятия, связанные с формированием дополнительных резервов мощности и заменой устаревших или аварийных элементов системы, являются затратными и требуют квалифицированной оценки и объективного обоснования их проведения. Для решения представленной задачи необходимо проводить оценку балансовой надежности перспективных схем ЭЭС и на её основе корректировать планы развития. Результатом такой оценки является набор показателей балансовой надежности (ПБН) ЭЭС, которые также имеют экономическую интерпретацию и в дальнейшем подвергаются анализу.

К числу ПБН относят: вероятность бездефицитной работы; математическое ожидание (м.о.) дефицита мощности; м.о. недоотпуска электроэнергии как в зонах надежности (ЗН) – элементов расчетной модели ЭЭС, представляющий собой часть ЭЭС с набором генерирующих агрегатов и нагрузкой, ограничения по пропускной способности линий электропередачи внутри которой отсутствуют, так и системы в целом; вероятностно взвешенные двойственные оценки достаточности генерирующих мощностей и сетевой инфраструктуры. В международной практике используются близкие по смыслу ПБН, такие как ожидаемая величина недопоставленной энергии (Expected Unserved Energy, EUE); вероятность возникновения дефицита мощности (Loss of Load Probability, LOLP); среднее число дней дефицита мощности в год (Loss of Load Expectation, LOLE); среднее число часов дефицита мощности в год (Loss of Load Hours, LOLH).

Наиболее используемым в мировой практике методом для оценки БН является метод статистических испытаний (метод Монте-Карло) [1–8]. Метод Монте-Карло состоит из трех основных этапов. На первом этапе происходит формирование набора псевдослучайных состояний системы. На втором этапе для каждого сформированного псевдослучайного состояния системы происходит моделирование каждого такого состояния и вычисление дефицитов мощности у потребителей, а также фиксируются события, влияющие на дефицит мощности (например, максимальная загрузка генерирующих источников в зонах надёжности и линий электропередачи, входящих в межзонные связи). Происходящие в соседние моменты времени случайные события могут быть как зависимыми, так и независимыми. На третьем этапе по результатам статистической обработки определяются показатели балансовой надежности и другие характеристики, которые представляют интерес для дальнейшего анализа.

Отечественные и зарубежные программные комплексы оценки балансовой надежности ЭЭС [3–4, 8–26], работа которых основана на методе Монте-Карло, используют различные математические модели и преследуют разные цели – как технические, так и экономические. Несмотря на это, большинство

из них применяют линейные или линеаризованные в процессе решения модели, основанные на транспортной задаче потокораспределения. Однако следует отметить, что существующие линейные модели вносят достаточно существенную погрешность при определении дефицита мощности. В работе [1] было показано, что наиболее адекватной является постановка в нелинейной форме, где потери в линиях электропередачи имеют квадратичную зависимость от передаваемой мощности. В некоторых существующих отечественных решениях уже используются нелинейные модели с учетом квадратичных потерь, однако данные модели являются несовершенными и характеризуются излишним набором переменных и ограничений, которые в анализируемых схемах большой размерности могут приводить к искажению результата и необоснованно высокому времени счета.

Еще одним недостатком существующих моделей минимизации дефицита мощности (МДМ) является то, что в них ограничения на перетоки мощности должны задаваться по межзонным связям. Как правило, данные о пропускной способности каждой межзонной связи отсутствуют, а для анализа и управления используются контролируемые сечения с заданной характеристикой максимально допустимых перетоков активной мощности (МДП). На практике трудно получить полное соответствие межзонных связей и контролируемых сечений. Это обстоятельство необходимо учитывать при разработке математических моделей МДМ.

Таким образом, в представленной работе рассматриваются существующие зарубежные и отечественные варианты моделей, поднимается вопрос адекватности используемых моделей, а также представляются варианты их модификаций, соответствующие современным условиям функционирования ЭЭС.

Статья состоит из трех разделов. В первом разделе рассматриваются и анализируются существующие постановки задачи МДМ. Во втором разделе представлена модификация модели МДМ, в которой учтены особенности учета контролируемых сечений ЭЭС. В третьем разделе представлены результаты экспериментальных исследований предлагаемой модификации модели МДМ ЭЭС.

1. ПОСТАНОВКА ЗАДАЧИ

Второй этап методики оценки балансовой надежности ЭЭС методом Монте-Карло заключается в моделировании случайных состояний ЭЭС. Обычно в условиях оценки балансовой надежности решается транспортная задача, заключающаяся в нахождении потокораспределения при заданных параметрах сети, генерирующих мощностей и уровней потребления мощности в зонах надежности. Как в отечественных, так и в зарубежных программно-вычислительных комплексах (ПВК) применяется множество различных постановок задачи МДМ. Следует учесть, что с точки зрения расчета ПБН в первую очередь должна рассматриваться задача минимизации дефицита мощности с оптимальным и физически адекватным распределением генерации между источниками и потребителями. Однако практически все зарубежные ПВК в качестве основной рассматривают задачу минимизации различных затрат, т. е. упор делается на экономические критерии, что также влияет на последующее потокораспределение.

1.1. МОДЕЛИ МДМ ЭЛЕКТРОЭНЕРГЕТИЧЕСКИХ СИСТЕМ В ЗАРУБЕЖНЫХ И ОТЕЧЕСТВЕННЫХ ПРОГРАММНО-ВЫЧИСЛИТЕЛЬНЫХ КОМПЛЕКСАХ

Модель ANTARES [9, 10, 13, 14] – это линейная модель, учитывающая множество различных параметров, а для работы с ЭЭС используется неориентированный граф. Целевая функция направлена на минимизацию суммы затрат различных видов мощности. Балансовые ограничения этой модели основаны на первом законе Кирхгофа, где учитываются нагрузка потребителей, генерация мощности и потоки мощности. Рассматриваемая модель включает и экономическую составляющую, т. е. при определении дефицита мощности учитываются затраты на производство мощности и ее передачу.

Модель Transmission Reliability Evaluation of Large-Scale Systems (TRELSS), применяемая в одноименном программном комплексе, в настоящее время полностью перенесена в программный комплекс Transmission Contingency Analysis Reliability Evaluation (TransCARE) [8, 15–17], основой которого является моделирование с помощью марковских процессов, а вычисление показателей надежности происходит несколькими разными способами. Один из них основан на мерах, ориентированных исключительно на системные проблемы компонентов сети, такие как перегрузка, нарушения или отклонения напряжения и разделение сети. Этот подход называется «подходом к системным проблемам» и обеспечивает информацию по показателям частоты, продолжительности и серьезности системных проблем. Стоит обратить внимание, что этот подход не учитывает возможность исправления проблем с помощью реакции системы и / или действий оператора. Следовательно, этот подход дает пессимистический взгляд на надежность и является индикатором наихудшего сценария. Другой подход, названный «подходом возможностей», предоставляет набор индикаторов отключения нагрузки в качестве показателя ненадежности ЭЭС. Целью данного подхода является оценка объема нагрузки, которую необходимо отключить, если проблемы в ЭЭС сохраняются даже после принятия корректирующих действий после отказа энергетического оборудования. Все используемые показатели надежности, такие как ожидаемая неиспользованная энергия (EUE), вероятность, частота и продолжительность отключения нагрузки, вычисляются для каждого нагрузочного узла, а также для системы в целом. Подход не учитывает время отклика корректирующих действий.

Модель Siemens PTI PSS/E TPLAN [8, 9, 18–21] предоставляет возможность расчета нескольких целевых функций, например, минимизация затрат на топливо, минимизация потерь активной мощности, минимизация потерь реактивной мощности, минимизация или максимизация передачи активной мощности и др. Ввиду того, что программное обеспечение является закрытым, отсутствует возможность полноценного знакомства с моделью, как, например, в программном комплексе Antares, однако существующие брошюры и упоминания о комплексе ясно дают понять, что частично используется линейная модель, а вышеописанные цели могут быть скомбинированы пользователем самостоятельно.

Программный комплекс CORAL [8, 9, 22] разработки PSR включает в себя такие модели, как SDDP, OPTGEN/NETPLAN, последняя отвечает за расчеты, связанные с планированием расширения производства / передачи с

ограничениями надежности поставок. За моделирование работы системы отвечает надстройка Coral, цель задается как максимум покрытия нагрузки (Expected Power Not Supplied (EPNS)), т. е. как максимальный ожидаемый дефицит мощности. Модель OptGen так же, как и модель Antares, учитывает гидро- и теплоэнергетические агрегаты ЭЭС, в том числе эксплуатационные затраты этих агрегатов.

Программный комплекс Grid Reliability and Adequacy Risk Evaluator (GRARE) [23] оценивает надежность и экономические эксплуатационные возможности с помощью метода Монте-Карло. Данный инструмент предназначен для анализа полной модели системы (линии, генераторы, трансформаторы и т. д.). Математическая модель, как и большинство вышеописанных комплексных моделей, учитывает эксплуатационные затраты для гидро- и теплогенерирующих агрегатов ЭЭС. Поддерживаются модели до 5000 узлов. Для расчета перетоков мощности и оценки уровня напряжения по модели постоянного тока применяется алгоритм Зауэра (Sauer). При моделировании учитывается удовлетворение спроса на энергию при минимальных затратах, максимизации доходов от производства энергии и оптимального потокораспределения с учетом пропускных способностей связей. Для решения задачи потокораспределения применяются Flow- или NTC-подходы. Уровень надежности ЭЭС определяется с помощью показателей ENS, LOLE, LOLP. Производство возобновляемых источников рассчитывается путем случайного отбора, исходя из статистически обработанных данных по производительности этих станций.

Программный комплекс PLEXOS [23-24], разрабатываемый Energy Exemplar, представляет собой сложный инструмент моделирования энергосистем. Он использует смешанные целочисленные методы оптимизации для определения минимальных затрат и решения задач диспетчеризации для удовлетворения спроса при соблюдении технико-экономических ограничений на генерирующих агрегатах. Расширенное смешанное целочисленное программирование (MIP) является основным алгоритмом моделирования и оптимизации. Задачи удовлетворения спроса на энергию при минимальных затратах, максимизации доходов от производства энергии и оптимального потокораспределения формулируются в линейном виде (после линеаризации) с целочисленными переменными, раскрывающими состояние генератора в сети.

В программном комплексе Multi-Area Reliability Simulation (MARS) [8, 25, 26] разработки General Electric уровень надежности оцениваемой системы определяется по показателям LOLE, LOEE, частота и длительность аварийных отключений, необходимость инициации процедур в экстренных ситуациях. Для решения поставленных задач используется транспортная модель, с помощью которой выполняется анализ потоков мощности, соединяющих энергозоны системы. Данный комплекс получил развитие в виде еще одной модели Multi-area Power System (MAPS) [26], предназначенной для решения задачи определения наименьших затрат, рассматриваемой и в других вышеупомянутых комплексах.

В отечественных программных комплексах применяются как линейные модели потокораспределения, так и нелинейные. Из линейных можно обратить внимание на модель МДМ, представленную в работе [4]. Целевая функция направлена на минимизацию разности мощности спроса нагрузки и обес-

печения потребности с учетом коэффициента, отражающего стоимостные показатели ограничений потребителей. В балансовых ограничениях учитывается разница обеспеченной потребности и используемой генерирующей мощности для каждой зоны надежности, где также принимается в расчет сумма перетоков, связанных с данной зоной, потери мощности не учитываются. При этом подобная линейная постановка имеет недостаток, заключающийся в возникающей неоднозначности в распределении дефицита мощности по отдельным зонам надежности. Устранение неоднозначности требует второго этапа решения задачи минимизации системного дефицита мощности. Для второго этапа меняется целевая функция, стоимостной коэффициент не применяется, балансовые ограничения не меняются. Значения переменных для различных по глубине дефицитов мощности разделенных частей объединения ЭЭС определяются по принципу пропорционального ограничения потребителей (принцип пропорциональности).

В работе [7] представлено несколько моделей, направленных на решение задачи минимизации дефицита мощности. В первую очередь рассматривается модель минимума суммарных по всем зонам электроэнергетической системы затрат, связанных с ограничением нагрузки, и затрат на производство электроэнергии. Эта модель характеризуется наличием ограничений на текущую величину нагрузки и генерации, мощности, передаваемой в зону по электрической сети (сетевые инъекции), а также на перетоки мощности по межсистемным связям. Соблюдаемый баланс мощностей соответствует первому закону Кирхгофа и определяется условием равенства нулю суммы инъекций и потерь мощности. Потоки мощности пропорциональны инъекциям, что частично учитывает второй закон Кирхгофа, а также в модели используются диагональная матрица проводимостей. Воздействие варьируемых переменных – генерации и нагрузки в какой-либо отдельной зоне на суммарный баланс мощности в системе – осуществляется через сетевые инъекции. Решение задачи оптимального распределения небаланса мощности узла между генерацией и ограничением нагрузки, которое можно рассматривать как дополнительную генерацию, не представляет каких-либо проблем и известно, как распределение по критерию равенства относительных приростов затрат на производство электроэнергии. Одним из вариантов решения данной задачи является квадратичное программирование, вследствие чего появляется необходимость в линейной форме представления всех ограничений – таким образом автор формулирует линейную задачу. Также автором рассматривается вариант данной модели с учетом затрат на генерацию, что говорит о возможности регулировки учета экономических показателей при расчетах. Еще в одной модели применяются сетевые инъекции в качестве управляющих переменных, при этом реализуется критерий минимума суммарных эксплуатационных затрат. Таким образом, целевая функция принимает вид квадратичной формы. Также автор рассматривает вариант линейной модели с учетом пропорциональности вынужденных ограничений нагрузки и генерации, где учитывается критерий оптимальности распределения дефицита мощности пропорционально нагрузкам. Данная модель представлена с учетом критерия минимума суммы квадратов отклонений от единого для всех зон идеального коэффициента, полученного без учета сетевых ограничений – таким образом задача сформулирована как нелинейная. Еще один вариант развития модели –

применение критерия минимального отклонения от идеальных значений, что в конечном итоге позволяет использовать в качестве целевой функции сумму взвешенных квадратов невязок.

В ИСЭМ СО РАН на протяжении длительного времени разрабатываются ПВК оценки балансовой надежности, это ПВК «КОРАЛ», «ПОТОК-3» и «ЯНТАРЬ» [3] которые включают в себя оригинальные математические модели, направленные на выявление минимума дефицита мощности. В настоящее время для оценки балансовой надежности применяется ПВК «Надежность», а также проводятся его постоянные улучшения. Изначально в разрабатываемых ПВК использовалась модель, основанная на потоковых алгоритмах и включающая в себя балансовые уравнения, а также модель, основанная на уравнениях сети в идеализации по постоянному току. Позднее стала применяться двухкритериальная модель оценки дефицита мощности ЭЭС, которая сводилась к классической задаче о максимальном потоке в сетях и была определена как линейная. Целевая функция обозначалась как минимум дефицита мощности, а балансовые ограничения включали в себя разницу мощности в самой зоне, а также входящий и исходящий переток мощности. В данной модели учитывался только 1-й закон Кирхгофа, а потери при этом не учитывались. Для распределения дефицита по всем потенциально дефицитным зонам надежности примерно пропорционально нагрузкам этих зон надежности был внедрен второй этап оптимизации по другой целевой функции. Таким образом была сформирована двухэтапная модель, которая впоследствии получила две модифицированные целевые функции. Далее были сформированы модели, учитывающие потери мощности на межузловых связях. Так появилась модель с потерями в ЛЭП, выражающимися в виде линейной функции от объемов передаваемой мощности. Данная модель была дополнительно модифицирована, а именно: были изменены балансовые линейные ограничения на нелинейные ограничения равенства. Также была сформирована аналогичная модель, но учитывающая особенности работы ЭЭС в условиях оптового рынка. Несмотря на обилие моделей, все они работали с учетом потокораспределения и наличия исходной информации о пропускной способности связей. На сегодняшний момент такая информация может быть получена не в полном объеме или может отсутствовать вовсе, а для учета сетевых ограничений используются МДП контролируемых сечений (КС).

На основании проведенного анализа математических моделей минимизации дефицита мощности, используемых в отечественных и зарубежных ПВК оценки балансовой надежности ЭЭС, можно сделать вывод, что актуальной задачей для развития модели является учет специфики задания корректных сетевых ограничений, соответствующих контролируемым сечениям ЭЭС.

1.2. МОДЕЛИ МДМ ЭЛЕКТРОЭНЕРГЕТИЧЕСКИХ СИСТЕМ В ЗАРУБЕЖНЫХ И ОТЕЧЕСТВЕННЫХ ПРОГРАММНО-ВЫЧИСЛИТЕЛЬНЫХ КОМПЛЕКСАХ

В ПВК «Надежность» [2], который в настоящее время развивается в ИСЭМ СО РАН, используют линейные и нелинейные модели минимизации дефицита мощности. Задача минимизации дефицита мощности (МДМ) формулируется следующим образом: для известных значений работоспособных

генераторных мощностей, требуемых уровней потребления мощности, ограничений на передачу мощности по межзонным связям и коэффициентов потерь мощности в линиях электропередачи, необходимо определить минимальное значение дефицита мощности в ЭЭС.

Математически линейная задача формулируется следующим образом:

$$\sum_{i=1}^{n}(\overline{y}_i - y_i) \to \min_{y,x,z}, \tag{1}$$

при соблюдении балансовых ограничений

$$x_i - y_i + \sum_{j=1}^{n}(1-a_{ji})z_{ji} - \sum_{j=1}^{n} z_{ij} = 0, \tag{2}$$

а также ограничений на оптимизируемые переменные:

$$0 \le y_i \le \overline{y}_i, \ \ i = 1,\ldots, n, \tag{3}$$

$$0 \le x_i \le \overline{x}_i, \ \ i = 1,\ldots, n, \tag{4}$$

$$0 \le z_{ij} \le \overline{z}_{ij}, \ \ i = 1,\ldots, n, \ \ j = 1,\ldots, n, \ \ i \ne j, \tag{5}$$

где x_i – используемая мощность в зоне надежности i (МВт); $\overline{x}_i$ – располагаемая мощность в зоне надежности i (МВт); y_i – фактически потребляемая мощность в зоне надежности i (МВт); $\overline{y}_i$ – максимальная потребляемая мощность в зоне надежности i (МВт); z_{ij} – поток мощности из зоны надежности i в j (МВт); $\overline{z}_{ij}$ – суммарная пропускная способность ЛЭП между зонами надежности i и j (МВт); z_{ji} – поток мощности из зоны надежности j в i (МВт); $\overline{z}_{ij}$ – суммарная пропускная способность ЛЭП между зонами надежности j и i (МВт); a_{ji} – коэффициенты удельных потерь мощности при ее передаче из зоны надежности j в i, $j \ne i$, $i = 1,\ldots, n$, $j = 1,\ldots, n$.

Модель (1)–(5) является распространенной моделью потокораспределения в области оценки балансовой надежности. Эта модель представляет собой транспортную задачу. Для решения представленной задачи оптимизации ввиду ее относительной простоты наиболее часто используется симплекс-метод и двойственный симплекс-метод в их разных вариациях.

1.3. МОДЕЛЬ МДМ ЭЛЕКТРОЭНЕРГЕТИЧЕСКИХ СИСТЕМ С КВАДРАТИЧНЫМИ ПОТЕРЯМИ

Модель МДМ ЭЭС с линейными потерями имеет допущения, касающиеся неполного учета потерь при перетоках мощности. Так, в [1] существует обоснованный вывод, что модель, где потери мощности зависят от квадрата передаваемой мощности, является более адекватной моделью, близкой по физическому смыслу к реальной работе ЭЭС. Для этого в модели (1)–(5) ис-

пользуются модифицированные балансовые ограничения, где ограничения вида (2) заменены на ограничения, представленные ниже:

$$x_i - y_i + \sum_{j=1}^{n}(1 - a_{ji}z_{ji})z_{ji} - \sum_{j=1}^{n} z_{ij} = 0, \;\; i = 1,\ldots, n. \qquad (6)$$

Таким образом, задачу МДМ можно представить в виде задачи нелинейного программирования. При исследовании моделей (1)–(5) и (1), (3)–(6) выявлено, что эти модели имеют ряд недостатков, в частности, в виде неверного потокораспределения, проявляющегося в одновременных встречных потоках мощности по одной межзонной связи.

1.4. МОДЕЛЬ МДМ С УЧЕТОМ КВАДРАТИЧНЫХ ПОТЕРЬ МОЩНОСТИ И ОГРАНИЧЕНИЕМ ПЕРЕМЕННЫХ

В качестве улучшения существующей постановки задачи МДМ с квадратичными потерями рассматривается модель, представленная в [27], где были проведены улучшения с целью сокращения количества оптимизируемых переменных по перетокам, т. е. вместо использования двух переменных, обозначающих перетоки мощности по межзонным связям для каждого направления, используется одна переменная, что, в свою очередь, избавляет от проблемы наличия перетоков мощности в разные стороны по одной и той же связи. В качестве целевой функции используется выражение (1), а ограничения (3) принимают следующий вид:

$$\underline{z}_j \le z_j \le \overline{z}_j, \;\; j = 1,\ldots, m, \qquad (7)$$

балансовые ограничения принимают вид

$$x_i - y_i + \sum_{j=1}^{m} t_{ij} z_j - \sum_{j=1}^{m} [\tilde{\alpha}_{ij}(z_j)](z_j)^2 \ge 0, \;\; i = 1,\ldots, n, \qquad (8)$$

где m – число связей между узлами; t_{ij} – элементы матрицы связей размера $n \times m$,

$$t_{ij} = \begin{cases} -1, \text{ если связь начинается в зоне надежности } j, \\ 1, \text{ если связь оканчивается в зоне надежности } j, \\ 0, \text{ если связь с зоной надежности отсутствует.} \end{cases} \qquad (9)$$

При этом функции $\tilde{\alpha}_{ij}(z_j)$ определяются следующим образом:

$$\tilde{\alpha}_{ij}(z_j) = \begin{cases} \alpha_j, \text{ если } t_{ij} z_j > 0, \\ 0, \text{ если } t_{ij} z_j \le 0, \end{cases} \;\; i = 1,\ldots, n, \; j = 1,\ldots, m, \qquad (10)$$

$$\underline{z}_j = -\overline{z}_j \quad \text{для всех } j, \qquad (11)$$

В тех случаях, когда полученное в результате оптимизации значение $z_j < 0$, необходимо считать что переток направлен в обратную сторону отно-

сительно заданного направления обозначенного в матрице *t*. Однако данная модель не может быть полноценно использована для расчетов ввиду наличия ограничения (11), где пропускная способность связи задается с одним и тем же значением для прямого и обратного направления. Такой способ задания перетоков не совсем корректен по причине применения различных ограничений в прямом и обратном направлениях. Таким образом необходимо избавиться от ограничения (11), тем самым позволив использовать отличные пропускные способности связей по встречным направлениям. Ограничение (8) из-за перехода от равенства к неравенству не отвечает физике процесса и приводит к появлению неестественно больших значений генераторных мощностей, что фактически невозможно в реальных ЭЭС. Однако несмотря на наличие данных недостатков необходимо отметить, что снижение количества оптимизируемых переменных положительно сказывается на скорости и объеме расчетов.

2. МОДИФИКАЦИЯ МОДЕЛИ МИНИМИЗАЦИИ ДЕФИЦИТА МОЩНОСТИ ЭЛЕКТРОЭНЕРГЕТИЧЕСКИХ СИСТЕМ С КВАДРАТИЧНЫМИ ПОТЕРЯМИ

Представленные модели (1)–(5) и (1), (3)–(6), а также (1), (7)–(11) имеют как минусы, так и плюсы своей постановки. В настоящей работе предлагается внедрить и экспериментально проверить модификации, направленные на решение существующих проблем.

2.1. РЕШЕНИЕ ПРОБЛЕМЫ ВСТРЕЧНЫХ ПЕРЕТОКОВ МОЩНОСТИ ПО МЕЖЗОННЫМ СВЯЗЯМ

Получаемые оптимальные значения с помощью моделей (1)–(5) и (1), (3)–(6) не обладают необходимой адекватностью с позиции физики процесса потокораспределения. В решении этой задачи могут присутствовать встречные перетоки мощности по межзонным связям. Также трудно найти единственное решение из-за наличия плато или множества оптимальных решений.

Результаты работы моделей (1)–(5) и (1), (3)–(6) при оценке балансовой надежности могут привести к искажению значений показателей надежности. Для устранения этой проблемы было сформулировано дополнительное ограничение для перетоков мощности, широко известное в оптимизации как «обнуление переменной», в данном случае ограничение представлено в следующем виде:

$$z_{ji} * z_{ij} = 0, \quad i = 1,\ldots,n, \quad j = 1,\ldots,n. \tag{12}$$

Таким образом, ограничение (12) преобразует рассматриваемую задачу в корректную с точки зрения моделирования работы перетоков мощности между ЗН и меняет поведение модели на более правильное с точки зрения физики процесса потокораспределения и формирует модель минимизации дефицита мощности с квадратичными потерями (1), (3)–(6), (12), далее МДМ_1.

2.2. РЕШЕНИЕ ПРОБЛЕМЫ НЕВЕРНОГО ПОТОКОРАСПРЕДЕЛЕНИЯ МОЩНОСТИ

Экспериментальные исследования показали, что, несмотря на модификацию нелинейной модели и ее представление в виде (1), (3)–(7), ввиду наличия множества оптимальных решений существует неверное потокораспределение мощности. Иначе говоря, потокораспределение не является оптимальным, однако, исходя из содержательной постановки задачи, требуется найти не только минимум дефицита мощности, но и корректное потокораспределение. Некорректное потокораспределение заключается в появлении кольцевых перетоков мощности.

Для решения этой проблемы был предложен ряд модификаций модели, описанных в работе [2], в первую очередь путем изменения балансовых ограничений (6), – с ограничений типа равенств на ограничения типа неравенства. Данное ограничение предусматривало переход модели оценки дефицита мощности к виду задачи выпуклого программирования:

$$x_i - y_i + \sum_{j=1}^{n}(1 - a_{ji}z_{ji})z_{ji} - \sum_{j=1}^{n} z_{ij} \geq 0, \quad i = 1,\ldots,n. \tag{13}$$

Были проведены экспериментальные исследования данной модификации модели. В итоге было выявлено, что минимум дефицита мощности совпадает со значением модели МДМ$_1$. При этом физически неверное распределение перетоков мощности осталось, однако повысилась загрузка генераторных мощностей, что физически невозможно, так как излишки генераторной мощности заперты и нет потребителей, которые могут удовлетворить такому предложению.

В качестве решения возникшей проблемы предложено использовать второй этап оптимизации. Так, на первом этапе проводится оптимизация выше обозначенной модели (1), (3)–(5), (12), (13), такой подход обеспечит выпуклое множество допустимых решений. Далее полученные оптимальные решения относительно переменной y_i необходимо зафиксировать и обозначить новую переменную как $\widetilde{y_i}$. После этого перейти ко второму этапу решения, сформировав новую целевую функцию, которая предполагает собой минимизацию суммы квадратов перетоков мощности:

$$\sum_{i=1}^{n} z_{ji}^2 \to \min, \tag{14}$$

а также модифицировать текущие балансовые ограничения (13) на балансовые ограничения, представленные ниже:

$$x_i - \widetilde{y_i} + \sum_{j=1}^{n}(1 - a_{ji}z_{ji})z_{ji} - \sum_{j=1}^{n} z_{ij} = 0, \quad i = 1,\ldots,n. \tag{15}$$

Данный подход был проверен в рамках экспериментов. Он включал в себя поочередную двухэтапную оптимизацию и взаимодействие двух моделей МДМ – с балансовыми ограничениями неравенствами (1), (3)–(5), (12), (13) и модели минимизации евклидовой нормы по перетокам (9), (3)–(5), (12), (13). Полученные результаты показали работоспособность такого подхода [2].

2.3. МОДЕЛИРОВАНИЕ КОНТРОЛИРУЕМЫХ СЕЧЕНИЙ

В текущих условиях функционирования и развития ЭЭС данные о пропускной способности межзонных связей зачастую полностью или частично отсутствуют. В то же время для контроля передачи мощности между зонами надежности ЭЭС используются контролируемые сечения с заданной характеристикой МДП. Такие КС включают в себя до нескольких ветвей (ЛЭП) с обозначенным направлением перетока мощности. Таким образом, использование представленных выше моделей без внесения дополнительных изменений становится невозможным, в связи с чем предлагается рассмотреть необходимые корректировки и дополнения для постановки математической задачи с квадратичными потерями.

В первую очередь необходимо обозначить матрицу S контролируемых сечений, в которой обозначены присутствие ветвей в КС с учетом их направлений. Размерность матрицы $l \times m$, где l – число КС, а m – число ветвей, элементы матрицы обозначим как cs_{kf}:

$$cs_{kf} = \begin{cases} 1, \text{ если ветвь присутствует в контролируемом сечении,} \\ 0, \text{ если ветвь отсутствует в контролируемом сечении.} \end{cases} \quad (16)$$

Каждое контролируемое сечение имеет МДП в прямом и обратном направлениях, для их хранения необходима матрица M размерностью $l \times 2$, где первый элемент (md_{k1}) в строке содержит значения прямого МДП, второй элемент (md_{k2}) – обратного МДП. Для работы модели также потребуется внести ограничения по КС для прямого и обратного МДП:

$$\sum_{f=1}^{m} cs_{kf} z_f^{\text{прям}} \le md_{k1}; \;\; k = 1,\ldots, l, \quad (17)$$

$$\sum_{f=1}^{m} cs_{kf} z_f^{\text{обр}} \le md_{k2}; \;\; k = 1,\ldots, l, \quad (18)$$

где $z_f^{\text{прям}}$ – прямой переток мощности z_{ij} (МВт), а $z_f^{\text{обр}}$ – обратный переток мощности (МВт), т. е. z_{ji} определяются при формировании матрицы S. Таким образом, существующая модель (1), (3)–(6), (12) должна быть преобразована в модель (1), (3)–(6), (12), (16)–(18) для устранения проблемы двусторонней загрузки контролируемых сечений. Проблема была решена путем введения дополнительных ограничений на контролируемые сечения:

$$\left(\sum_{f=1}^{m} cs_{kf} z_f^{\text{прям}}\right)\left(\sum_{f=1}^{m} cs_{kf} z_f^{\text{обр}}\right) = 0; \;\; i = 1,\ldots,n, \;\; j = 1,\ldots,n, \quad (19)$$

Это ограничение, помимо физически корректного распределения потоков мощности по сечениям в одном из направлений, позволяет также изъять из модели ограничение (12) ввиду дублирования действий по направлению

перетоков мощности. Итогом вышеописанных дополнений и изменений является следующая модель минимизации дефицита мощности с квадратичными потерями и с учетом контролируемых сечений: (1), (3)–(6), (16)–(19), далее МДМ$_2$.

2.4. МОДИФИЦИРОВАННАЯ ПОСТАНОВКА ЗАДАЧИ МИНИМИЗАЦИИ ДЕФИЦИТА МОЩНОСТИ

Учитывая то, что постановка (1), (7)–(11) имеет ряд преимуществ, а именно отсутствие встречных перетоков и снижение количества переменных, предлагается внести изменения, позволяющие убрать вероятную неоднозначность решения, а также сделать возможным использование данной модели в рамках расчетов с использованием реальных данных и ограничений по контролируемым сечениям. Предлагается рассмотреть следующую постановку математической задачи: целевая функция (1), а также верхние и нижние ограничения на переменные генерации и нагрузки (3) и (4), матрица, определяющая направление потоков мощности по связям (9), а также функции определения коэффициента потерь $\tilde{\alpha}_{ij}(z_j)$ (10) остаются без изменений. Для исключения возможности появления запертой генерации балансовые ограничения (8) заменяются следующими ограничениями типа равенства:

$$x_i - y_i + \sum_{j=1}^{m} t_{ij} z_j - \sum_{j=1}^{m} [\tilde{\alpha}_{ij}(z_j)](z_j)^2 = 0, \quad i = 1,\ldots,n. \tag{20}$$

Ограничение (11) не позволяет использовать разные пропускные способности перетоков мощности в прямом и обратном направлении и противоречит физике процесса работы ЭЭС. Таким образом, формируется модифицированная модель минимизации дефицита мощности следующего вида: (1), (3), (4), (9), (10), (20), далее МДМ$_3$, которая является идентичной модели (1), (3)–(6), (12). Для учета контролируемых сечений в полученной модели, необходимо воспользоваться матрицей КС – S, элементы которой необходимо изменить, поэтому вместо элементов (16) необходимо использовать следующие:

$$cs_{kf} = \begin{cases} 1, \text{ если связь с прямым направлением относится к КС,} \\ -1, \text{ если связь с обратным направлением относится к КС,} \\ 0, \text{ если связь отсутствует в контролируемом сечении.} \end{cases} \tag{21}$$

Помимо преобразований элементов матрицы S, также необходимо заменить ограничения контролируемых сечений, приведенных в (17) и (18), на ограничения вида:

$$\text{если } (z_f) \geq 0, \text{ то } \sum_{f=1}^{m} (cs_{kf} z_f) \leq md_{k1}; \quad k = 1,\ldots,l; \tag{22}$$

$$\text{если } (z_f) < 0, \text{ то } \sum_{f=1}^{m} (cs_{kf} z_f) \geq md_{k2}; \quad k = 1,\ldots,l, \tag{23}$$

где z_f – переток мощности z_j, а его направление определяется в матрице *S*. Таким образом формируется модифицированная модель минимизации дефицита мощности с учетом контролируемых сечений, которая примет следующий вид: (1), (3), (4), (9), (10), (20)–(23), далее МДМ_4.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ РАЗЛИЧНЫХ МОДЕЛЕЙ МИНИМИЗАЦИИ ДЕФИЦИТА МОЩНОСТИ

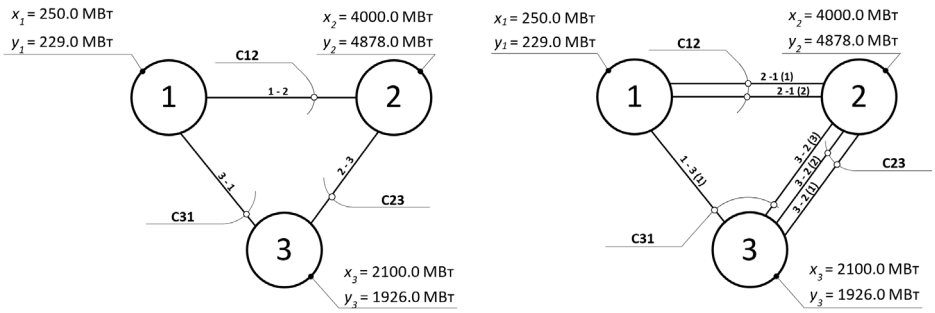
Экспериментальные расчеты проводились для систем различной конфигурации с разными начальными параметрами. Были проанализированы результаты работы ранее рассмотренных моделей: МДМ_1, МДМ_2 и МДМ_3. Модифицированная модель МДМ_4 не принимала участия в эксперименте ввиду невозможности полноценно воспроизвести ограничения (22) и (23) и провести расчеты, так как целевая функция и ее градиент с учетом метода штрафных функций не может быть вычислена. Попытки определить формализованные ограничения не привели к успеху ввиду наличия разных знаков переменных перетоков, а также пересекающихся и противоречивых ограничений для разных контролируемых сечений. Таким образом, модифицированная модель МДМ_4 не может корректно работать с контролируемыми сечениями без ввода дополнительных переменных контроля направления перетоков.

Всего было проведено четыре этапа экспериментов, где для исследований было подготовлено три разных вида расчетов: расчет № 1 – для указанных исходных данных; расчет № 2, в котором из всех исходных данных меняются только удельные коэффициенты потерь на ЛЭП, все они приравниваются к значению 0,1; расчет № 3, в котором используются исходные данные с измененными характеристиками МДП КС.

На первом этапе экспериментальные исследования проводились для трехзонных систем, представленных на рис. 1. Схема ТС_1 состоит из трех зон надежности и трех ЛЭП. Схема ТС_2 состоит из трех зон надежности и шести ЛЭП, распределенных по трем разным контролируемым сечениям. Расчеты выполнялись в программном комплексе GAMS, который используется в некоторых отечественных и зарубежных ПВК оценки балансовой надежности.

На первом этапе серия расчетов проводилась для ТС_1, исходные параметры для которой указаны на рис. 1, а также в табл. 1. В данном случае в каждое КС входит одна ЛЭП, а величина МДП КС совпадает с пропускной способностью входящей в КС ЛЭП. Исходя из результатов, полученных в расчете № 1, работа моделей МДМ_1 и МДМ_3 имеет идентичную логику распределения генерации по зонам надежности и загрузки контролируемых сечений. Основным результатом принято считать найденный дефицит мощности, в данном случае для модели МДМ_1 минимум дефицита мощности составляет 718,734 МВт, для модели МДМ_3 – 718,543 МВт. Разница составила 0,191 МВт, что входит в пределы погрешности исходных данных до 5 МВт,

однако для достижения меньшего дефицита моделью $МДМ_3$ были задействованы большие объемы генераторных мощностей 3-й зоны надежности и большей загрузки сечения С31.



$ТС_1$ $ТС_2$

Рис. 1. Схемы анализируемых систем с тремя зонами надежности и с разным количеством ЛЭП в контролируемых сечениях, где С – контролируемое сечение

Fig. 1. Schemas of the analyzed systems with three reliability zones and different number of power lines in controlled sections, where C is the controlled section

Таблица 1

Table 1

Дополнительные исходные данные для $ТС_1$

Additional initial data for $ТС_1$

Удельный коэффициент потерь ЛЭП		Обозначение ЛЭП		
		1–2	2–3	3–1
		0,0006033	0,0000113	0,0001729
МДП КС (МВт)	Прямое	30	150	20
	Обратное	45	130	40

Для подтверждения эффективности модели $МДМ_3$ с точки зрения найденного дефицита мощности был проведен расчет № 2, в котором удельные коэффициенты потерь для каждой из связей были установлены на уровне 0,1. В результате этого минимум дефицита мощности для расчета моделью $МДМ_1$ составил 873,0 МВт, а для модели $МДМ_3$ – 838,192 МВт. Таким образом, разница решений равна 31,808 МВт. Расчет № 3 включал в себя использование оригинальных исходных данных с измененными характеристиками пропускных способностей сечений, где были найдены идентичные дефициты мощности с разницей в 0,001 МВт.

В результате работы моделей $МДМ_1$ и $МДМ_3$ преимущество в поиске наименьших значений дефицита мощности показала модель $МДМ_3$. Ввиду

того, что обе модели являются идентичными, дополнительно были проведены расчеты для определения факторов, влияющих на различие конечных результатов. В модели $МДМ_1$ используются балансовые ограничения равенства (6), в котором используются входящие и выходящие перетоки мощности, в том числе учитываются квадратичные потери для всех примыкающих к зоне надежности ЛЭП, что, в свою очередь, влияет на результат минимизации. В модели $МДМ_3$ квадратичные потери учитываются только для ЛЭП, направленных в конкретную зону надежности, в связи с этим потери при минимизации учитываются иначе. Для подтверждения данного вывода целевая функция модели $МДМ_3$ была заменена на следующую целевую функцию:

$$\sum_{i=1}^{n}(\overline{y}_i - y_i) + \sum_{j=1}^{m}[\tilde{\alpha}_{ij}(z_j)](z_j)^2 \to \min_{x,y,z}. \qquad (24)$$

В результате была получена модель (24), (3), (4), (9), (10), (16)–(19), для нее были проведены расчеты № 1, № 2 и № 3, в результате были получены полностью идентичные с моделью $МДМ_1$ результаты расчетов. Это может говорить о том, что квадратичные потери в модели $МДМ_3$ учитываются неверно.

Для проверки корректности оптимизированных значений переменных полученные результаты расчетов № 1 и № 2 были напрямую подставлены в модели $МДМ_1$ и $МДМ_3$, после этого были рассчитаны значения балансовых ограничений (БО).

Ввиду специфики хранения чисел в памяти ЭВМ, а также стандартных погрешностей методов, реализованных в программном обеспечении GAMS, полученные БО находятся в пределах нуля, но не всегда ему равны. Из табл. 2 видно, как изменяется порядок погрешности в зависимости от расчета. Для модели $МДМ_1$ наибольшей погрешностью можно считать результат расчета балансовых ограничений 2-й зоны надежности в расчете № 2, равный 4,98456 МВт, что входит в пределы погрешности по входным данным. Для модели $МДМ_3$ порядок погрешности превышен в БО для 3-го узла. Полученное значение, равное 141,6724 МВт, не входит в пределы погрешности исходных данных и не может считаться верным. По итогу работы моделей $МДМ_1$ и $МДМ_3$ можно заключить, что с помощью модели $МДМ_1$ можно получить более точные результаты, в то же время модель $МДМ_3$ показывает лучшие значения дефицита мощности, однако с большей погрешностью расчетов, что недопустимо при оценке балансовой надежности.

На втором этапе экспериментов серия расчетов осуществлялась для $ТС_2$, где повторяющиеся исходные параметры полностью совпадают с предыдущей системой и указаны на рис. 1. В табл. 3 указаны дополнительные исходные данные. В результате расчета № 1 для модели $МДМ_2$ дефицит составил 811,148 МВт. Изменения исходных данных для расчета № 2 также затронули значения удельных коэффициентов потерь, они были приравнены к 0,1. В результате расчета № 2 были получены дефициты на уровне 865,5 МВт для $МДМ_2$. Также проводился расчет № 3, где были изменены характеристики

пропускных способностей сечений. В результате этого для модели $МДМ_2$ было получено значение 822,045 МВт.

Таблица 2

Table 2

Результаты подстановки полученных значений в балансовые ограничения

Results of using the obtained values in the balance constraints

Балансовые ограничения	Расчет № 1		Расчет № 2	
	$МДМ_1$	$МДМ_3$	$МДМ_1$	$МДМ_3$
1-й узел	–5,7E-05	0	0,0479	0
2-й узел	0,0004	0,0003	4,9846	2,4849
3-й узел	1,24E-13	8,22E-05	–8,72635E-14	141,6724

Таблица 3

Table 3

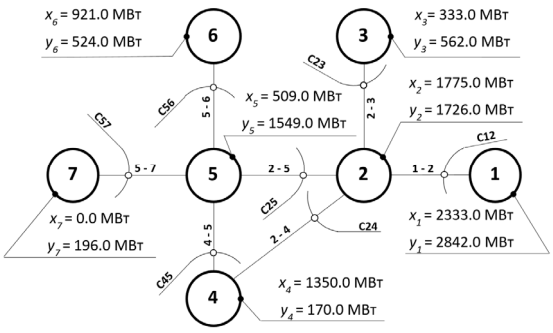
Дополнительные исходные данные для $ТС_2$

Additional initial data for $ТС_2$

Удельный коэффициент потерь ЛЭП		Обозначение ЛЭП		
		2–1(1); 2–1 (2)	3–2(1); 3–2 (2); 3–2 (3)	1–3
		0,0006033	0,0000113	0,0001729
МДП КС (МВт)	Прямое	188	15	40
	Обратное	175	10	36

На третьем и четвертом этапах экспериментальных исследований рассматривалась система, состоящая из семи зон надежности и семи линий электропередач, такая схема уже использовалась ранее в работах [1, 3, 27]. Система также рассматривалась в двух вариантах: в первом использовались линии, входящие в одноименные КС, что позволило сравнить эффективность поиска минимума дефицита мощности для моделей $МДМ_1$ и $МДМ_3$ на системе большего размера; во втором варианте использовались контролируемые сечения, включающие в себя как одну, так и несколько линий.

Эксперимент третьего этапа включал в себя расчеты для $ТС_3$. Как видно из схемы на рис. 2, некоторые узлы являются избыточными, другие – дефицитными по балансам мощности. Контролируемые сечения С23 и С25 имеют разные пропускные способности для прямого и обратного направлений. Более подробное описание приведено в табл. 4.



TC_3

Рис. 2. Схема анализируемой системы с семью зонами надежности и семью ЛЭП, входящими в контролируемые сечения

Fig. 2. Diagram of the analyzed system with seven reliability zones and seven power lines, included into the controlled sections

Таблица 4

Table 4

Начальные параметры ЛЭП и контролируемых сечений для TC_3

Initial parameters of the PTL and controlled sections for TS_3

Наименование контролируемого сечения	Номера зон надежности, примыкающих к ЛЭП	Удельные коэффициенты потерь на ЛЭП	Пропускные способности сечений в направлениях	
			прямое (МВт)	обратное (МВт)
C12	1–2	0,0004035	360	360
C23	2–3	0,0000216	150	100
C24	2–4	0,0001829	200	200
C25	2–5	0,0002255	800	1000
C45	4–5	0,0000114	1200	1200
C56	5–6	0,0005221	300	300
C57	5–7	0,0003116	150	150

По результатам расчетов системы с учетом моделей $МДМ_1$ и $МДМ_3$ направление потокораспределения полностью идентичное. Задействуются те же самые линии, однако различаются некоторые из значений загрузки. Результаты полученного расчета № 1 представлены ниже.

Таблица 5

Table 5

Результаты расчета № 1 моделей $МДМ_1$ и $МДМ_3$ для $ТС_3$

Results of the first calculation of models MDM_1 and MDM_3 for TS_3

Требуемая генерация		Обеспеченная нагрузка		Загрузка ЛЭП		
$МДМ_1$ (МВт)	$МДМ_3$ (МВт)	$МДМ_1$ (МВт)	$МДМ_3$ (МВт)	Номера зон надежности, примыкающих к ЛЭП	$МДМ_1$ (МВт)	$МДМ_3$ (МВт)
2333,00	2333,00	2451,45	2492,97	1–2	–124,74	–159,97
1775,00	1775,00	1726,00	1726,00	2–3	150,00	150,00
333,00	333,00	482,50	482,50	2–4	–154,51	–178,45
1350,00	1350,00	170,00	170,00	2–5	–76,93	–92,86
509,00	509,00	1549,00	1549,00	4–5	1025,48	995,72
824,00	870,98	524,00	524,00	5–6	–300,00	–300,00
0	0	142,98	142,98	5–7	150,00	150,00

Исходя из полученных результатов расчета № 1 (табл. 5), основное различие заключается в значении минимума дефицита мощности. Так, для модели $МДМ_1$ оно составляет 523,056 МВт, а для $МДМ_3$ – 481,537. В предыдущих расчетах ощутимая разница в 31,808 МВт найденного значения минимума дефицита мощности была только при расчетах с учетом удельных коэффициентов потерь, равных 0,1. В данном случае при расчетах с использованием удельных коэффициентов потерь, не превышающих 0,0001, разница полученных дефицитов составляет 41,519 МВт, что не может входить в пределы погрешности исходных данных до 5 МВт. По результатам расчета № 2 минимум дефицита мощности для $МДМ_1$ был определен как 1961,5 МВт, а для $МДМ_3$ – 1875,128 МВт. Разница составила 86,372 МВт, что также выходит за пределы погрешностей. Для расчета № 3 была изменена пропускная способность ЛЭП 4–5 до уровня 300 МВт. В результате этого были получены следующие значения дефицита мощности: $МДМ_1$ – 1182,003 МВт, $МДМ_3$ – 1127,891 МВт; разница составила 54,112 МВт.

Для проверки работоспособности моделей с учетом контролируемых сечений была проведена иная серия расчетов для четвертого этапа. В данном случае $ТС_4$ включала от одной до трех линий на одно контролируемое сечение, как это показано на рис. 3.

Данная схема идентична по уровню доступной генераторной мощности и обеспеченной нагрузки в зонах надежности, а начальные параметры для ЛЭП, обозначенные в табл. 4, также используются в данном расчете. Изменения коснулись содержания контролируемых сечений и их МДП по направлениям, которые описаны в табл. 6.

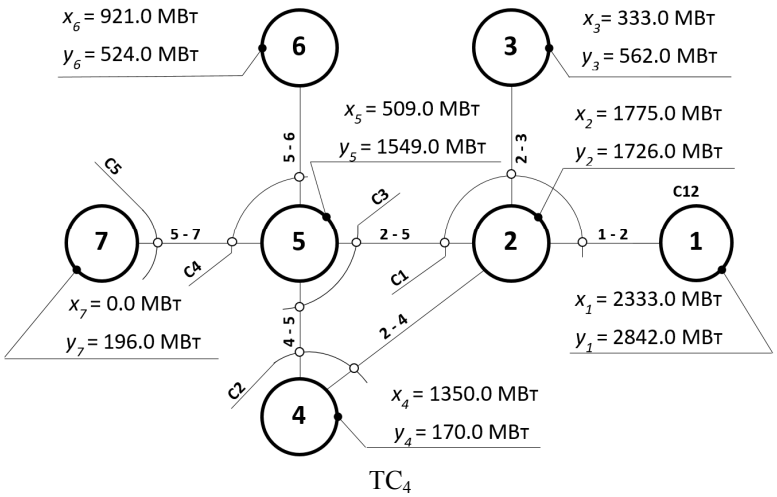


Рис. 3. Схема анализируемой системы с семью зонами надежности и семью ЛЭП входящих в пять контролируемых сечений

Fig. 3. Diagram of the analyzed system with seven reliability zones and seven power lines, included into the five controlled sections

В результате работы модели $МДМ_2$ в рамках расчета № 1 был определен дефицит мощности, равный 766,137 МВт. Распределение мощности по ЛЭП отличалось от результатов работы моделей $МДМ_1$ и $МДМ_3$, в данном случае были загружены всего три линии. Генерация в 6-й зоне надежности ниже по сравнению с расчетами для схемы $ТС_3$.

Таблица 6

Table 6

Начальные параметры контролируемых сечений и ЛЭП для $ТС_4$

Initial parameters of controlled sections and power lines for TS_4

Наименование контролируемого сечения	Номера зон надежности, примыкающих к ЛЭП	МДП сечений по направлениям	
		Прямое (МВт)	Обратное (МВт)
C1	1–2	1310	1460
	2–3		
	2–5		
C2	2–4	1400	1400
	4–5		
C3	4–5	2000	2200
	2–5		
C4	5–7	450	450
	5–6		
C5	5–7	150	150

Как видно из табл. 7, при работе модели $МДМ_2$ с $ТС_4$ наблюдаются отличия, а именно из семи линий задействуется только три, при этом идентично загружаются генераторные мощности шести из семи зон надежности, нагрузка также покрывается по-разному. В данном случае на распределение мощности повлияли направления связей внутри сечений, линии внутри не могут быть расставлены произвольно, их направления определяются экспертно.

Таблица 7

Table7

Результаты расчетов моделей $МДМ_1$ для $ТС_3$ и $МДМ_2$ для $ТС_4$

Results of calculating models MDM_1 for TS_3 and MDM_2 for TS_4

Требуемая генерация		Обеспеченная нагрузка		Загрузка ЛЭП		
$МДМ_1$ (МВт)	$МДМ_2$ (МВт)	$МДМ_1$ (МВт)	$МДМ_2$ (МВт)	Номера зон надежности, примыкающих к ЛЭП	$МДМ_1$ (МВт)	$МДМ_2$ (МВт)
2333,00	2333,00	2451,45	2381,03	1–2	–124,74	–49,00
1775,00	1775,00	1726,00	1726,00	2–3	150,00	0
333,00	333,00	482,50	333,00	2–4	–154,51	0
1350,00	1350,00	170,00	170,00	2–5	–76,93	0
509,00	509,00	1549,00	1549,00	4–5	1025,48	1180,00
824,00	524,00	524,000	524,00	5–6	–300,00	0
0	0	142,98	119,83	5–7	150,00	124,68

Для подтверждения был проведен отдельный расчет, в котором направления перетоков по ЛЭП были выбраны такими же, как и в полученном решении расчета № 1 для $МДМ_1$. Дефицит, полученный с помощью модели $МДМ_2$ и экспертно заданными направлениями перетоков по линиям в контролируемых сечениях, равен 473,171 МВт. В потокораспределении, в свою очередь, были задействованы идентичные ЛЭП, но с другими значениями.

Подводя итог, необходимо отметить, что по результатам проведенных экспериментов модифицированная модель минимизации дефицита мощности $МДМ_3$ является работоспособной. Данная модель использует меньшее количество ограничений равенств и меньшее количество оптимизируемых переменных, находит лучшие значения минимума дефицита мощности в системе, однако имеет погрешности, выходящие за пределы ошибки в исходных данных. На данном этапе эта модель не способна учитывать контролируемые сечения, вследствие чего не может быть использована для расчетов в программных комплексах, предназначенных для оценки балансовой надежности. В свою очередь, модель минимизации дефицита мощности с квадратичными потерями и учетом контролируемых сечений отвечает современным требованиям расчета дефицита мощности с учетом контролируемых сечений.

ЗАКЛЮЧЕНИЕ

В рамках проведенной работы были проанализированы математические модели, предназначенные для расчетов потокораспределения и используемые в отечественных и зарубежных программных комплексах оценки балансовой надежности ЭЭС. В ходе работы подверглись анализу математические модели из таких программных комплексов, как ANTARES, TRELSS, TransCARE, Siemens PTI PSS/E TPLAN, CORAL (включая SDDP, OPTGEN/NETPLAN), GRARE, PLEXOS, MARS, MAPS, «Орион», «Янтарь», «Надежность». В некоторых из перечисленных комплексах используется более одной модели с набором различных целевых функций. Были экспериментально исследованы модели МДМ_1, МДМ_2 и МДМ_3 на системах разной размерности и топологии. Основным отличием анализируемых моделей от существующих постановок является учет сетевых ограничений, соответствующих контролируемым сечениям ЭЭС, а не ЛЭП. Экспериментальные исследования показали, что модель МДМ_3, находит лучшие значения минимума дефицита мощности по сравнению с МДМ_1, однако ввиду наличия погрешностей, выходящих за пределы ошибки по входным данным, а также невозможности внедрения дополнительных ограничений, формирующих контролируемые сечения, модель МДМ_3 не может в дальнейшем использоваться для оценки балансовой надежности. В то же время сформированная модель МДМ_2 справилась с поставленной задачей поиска минимума дефицита мощности с учетом контролируемых сечений с заданной характеристикой МДП.

СПИСОК ЛИТЕРАТУРЫ

1. *Ковалев Г.Ф., Лебедева Л.М.* Надежность систем электроэнергетики / отв. ред. Н.И. Воропай. – Новосибирск: Наука, 2015. – 224 с.

2. *Iakubovskii D.V., Krupenev D.S., Boyarkin D.A.* An analysis of shortage minimization models to assess power system adequacy // Energy Systems Research. – 2018. – Vol. 1, N 3. – P. 25–32. – DOI: 10.25729/esr.2018.03.0003J.

3. *Ковалев Г.Ф., Лебедева Л.М.* Модель оценки надежности электроэнергетических систем при долгосрочном планировании их работы // Электричество. – 2000. – № 11. – С. 17–24.

4. *Чукреев Ю.Я.* Модели обеспечения надежности электроэнергетических систем. – Сыктывкар: Коми НЦ УрО РАН, 1995. – 176 с.

5. *Billinton R., Li W.* Reliability assessment of electric power systems using Monte Carlo methods. – New York: Springer, 1994. – 352 p.

6. *Li W.* Probabilistic transmission system planning. – Hoboken, NJ: Wiley-IEEE Press, 2011. – 376 p.

7. *Обоскалов В.П.* Алгоритмические аспекты расчета вероятностных показателей дефицита мощности в задаче балансовой надежности ОЭС // Известия Российской академии наук. Энергетика. – 2020. – № 2. – С. 59–74. – DOI: 10.31857/S0002331020010094.

8. Working Group 601 of Study Committee C4. Review of the current status of tools and techniques for risk-based and probabilistic planning in power systems / International Conference on Large High Voltage Electric Systems, March 2010. – CIGRE, 2010.

9. Systematic mapping of power system models: Expert survey, EUR 28875 EN / R. Fernandez Blanco Carramolino, F. Careri, K. Kavvadias, I. Hidalgo Gonzalez, A. Zucker, E. Peteves. – Luxembourg: Publications Office of the European Union, 2017. – DOI: 10.2760/422399.

10. Assessment of underlying capacity mechanism studies for Greece: EUR 28611 EN / G. Antonopoulos, S. Chondrogiannis, K. Kanellopoulos, I. Papaioannou, A. Spisto, T. Efthimiadis, G. Fulli. – Luxembourg: Publications Office of the European Union, 2017. – DOI: 10.2760/51331.

11. *Bertoldi O., Scalcino S., Salvaderi L.* Adequacy avaluation: an application of ENEL's SICRET program to New Brunswick power system: WG 38/03/01 / CIGRE Symposium "Electric Power System Reliability". – Montreal; Canada, 1991.

12. Composite power system reliability analysis application to the New Brunswick Power Corporation System: the draft report of the CIGRE Symposium on Electric Power Systems Reliability. – Montreal; Canada, 1991.

13. RTE Antares. Antares Optimization problems formulation. – URL: https://antares.rte-france.com (accessed: 13.01.2021).

14. RTE Antares. Modelling of flow-based domains in Antares for adequacy studies. – URL: https://antares.rte-france.com/?page_id=19&lang=en (accessed: 04.05.2021).

15. A Study on probabilistic risk assessment for transmission and other resource planning / A. Gaikwad, S. Agarwal, K. Carden, N. Wintermantel, S. Meliopoulos, M. Kumbale / Electric Power Research Institute for EISPC and NARUC. – EPRI, 2015. – (NARUC-2013-RFP027-DE0316).

16. *Urrego A.L.* A novel method for the approximation of risk of Blackout in operational conditions: Doctoral thesis. – Paris, 2016.

17. Survey of tools for risk assessment of cascading outages / M. Papic, K. Bell, Y. Chen, I. Dobson, L. Fonte, E. Haq, P. Hines, D. Kirschen, X. Luo, S.S. Miller, N. Samaan, M. Vaiman, M. Varghese, P. Zhang // 2011 IEEE Power and Energy Society General Meeting. – Detroit, MI, 2011.

18. *Hong Y.-Y., Lee L.-H.* Reliability assessment of generation and transmission systems using fault-tree analysis // Energy Conversion and Management. – 2009. – Vol. 50. – P. 2810–2817.

19. Siemens AG and Siemens Industry, Inc. Model Management Module for PSS®E. – URL: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed: 15.12.2020).

20. Siemens AG and Siemens Industry, Inc. Optimal Power Flow PSS®E. – URL: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed: 21.01.2021).

21. Siemens AG and Siemens Industry, Inc. PSS®E. High-performance Transmission Planning and Operations Software for the Power Industry. – Art. EMDG-B10125-00-7600--PSSE Brochure. – URL: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed: 23.12.2020).

22. PSR – Energy consulting and analytics. OPTGEN User Manual, Version 7.4. – Rio de Janeiro: PSR, 2019.

23. ENTSO-E. Mid-term adequacy forecast 2018. Appendix 1: Methodology and detailed results. – URL: https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/MAF_2018_Methodology_and_Detailed_Results.pdf (accessed: 04.05.20210).

24. PLEXOS Market Simulation Software. – URL: https://energyexemplar.com/solutions/plexos/ (accessed: 01.02.2021).

25. *Chu K.* MARS. Multi-area reliability simulation. EOP – on demand feature. – General Electric Company, 2014.

26. *Jirutitijaroen P., Singh C.* Reliability and cost trade-off in multi-area power system generation expansion using dynamic programming and global decomposition // IEEE Transactions on Power Systems. – 2006. – Vol. 21, iss. 3. – P. 1432–1441.

27. *Зоркальцев В.И., Пержабинский С.М.* Модель оптимизации дефицита мощности электроэнергетической системы // Управление большими системами. – 2010. – № 30-1. – С. 300–318.

Якубовский Дмитрий Викторович, младший научный сотрудник лаборатории надежности топливо- и энергоснабжения № 32, Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук. Основное направление научных исследований – математическое моделирование, оптимизация, параллельные вычисления. Имеет более 20 печатных работ. E-mail: yakubovskii.dmit@mail.ru

Крупенёв Дмитрий Сергеевич, кандидат технических наук, заведующий лабораторией надежности топливо- и энергоснабжения № 32, Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук. Основное направление научных исследований – надежность систем энергетики, оценка балансовой и режимной надежности электроэнергетических систем, методы оптимизации. Имеет более 45 печатных работ. E-mail: krupenev@isem.irk.ru

Бояркин Денис Александрович, младший научный сотрудник лаборатории надежности топливо- и энергоснабжения № 32, Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук. Основное направление научных исследований – машинное обучение, имитационное моделирование, вероятностное моделирование. Имеет более 20 печатных работ. E-mail: boyarkin_denis@mail.ru

Iakubovsky Dmitry V., junior researcher in the laboratory "Reliability of fuel and energy supply" No. 32, the Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences. His research interests are currently focused on mathematical modeling, optimization methods and parallel computing. He has more than 20 publications. E-mail: yakubovskii.dmit@mail.ru

Krupenev Dmitry S., PhD (Eng.), head of the laboratory "Reliability of fuel and energy supply" No. 32, Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences. His research interests are currently focused on the reliability of energy systems, assessment of the adequacy and operating reliability of electric power systems, and optimization methods. He has more than 45 publications. E-mail: krupenev@isem.irk.ru

Boyarkin Denis A., junior researcher of the laboratory "Reliability of fuel and energy supply" No. 32, Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences. His research interests are currently focused on machine learning, simulation modeling, and probabilistic modeling. He has more than 20 publications. E-mail: boyarkin_denis@mail.ru

DOI: 10.17212/2782-2001-2021-2-95-120

A minimization model of the power shortage of electric power systems with regard to restrictions on controlled sections[*]

D.Ya. IAKUBOVSKY[a], *D.S. KRUPENEV*[b], *D.A. BOYARKIN*[c]

Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences, 130 Lermontov Street, Irkutsk, 664033, Russian Federation

[a] *yakubovskii.dmit@mail.ru* [b] *krupenev@isem.irk.ru* [c] *boyarkin_denis@mail.ru*

Abstract

A steady trend towards the development of electric power systems leads to their continuous enlargement and sophistication. As a result, new ways of their control appear. In this regard, the existing models and complexes for adequacy assessment may work inadequately and ineffectively in terms of the obtained results adequacy. To assess the current state of the existing models and complexes, we reviewed and analyzed the domestic and foreign software and computer systems. In particular, we considered mathematical models of minimizing the power shortage. This work is based on the problem of modifying mathematical models of minimizing the power shortage used in adequacy assessment of the electric power systems of one of the complexes under consideration. As a modification of mathematical models, it is proposed to exclude the existing method of using the line capacities and start use correct accounting for the maximum permissible active power flow in controlled sections. The experimental part reflected in the paper concerns the testing of options for models to minimize the power shortage, as well as the proposed modifications on various systems, including those consisting of three and seven reliability zones with a variable number of controlled sections and power lines included in them.

The results of the study have shown that the proposed modifications are efficient and can be used in the future. The authors also obtained the most adequate results in terms of the physical laws of electric power system operation due to the model of minimizing the power shortage with quadratic losses which takes into account the limitations of power transmission over controlled sections.

[*] *Received 19 February 2021.*
The reported study was funded by RFBR, project number 20-08-00550 A.

Keywords: electric power system, reliability, mathematical model, minimization of power shortage, maximum permissible flow, quadratic model, controlled sections, adequacy assessment

REFERENCES

1. Kovalev G.F., Lebedeva L.M. *Nadezhnost' sistem elektroenergetiki* [Reliability of electric power systems]. Novosibirsk, Nauka Publ., 2015. 224 p.

2. Iakubovskii D.V., Krupenev D.S., Boyarkin D.A. An analysis of shortage minimization models to assess power system adequacy. *Energy Systems Research*, 2018, vol. 1, no. 3, pp. 25–32. DOI: 10.25729/esr.2018.03.0003J.

3. Kovalev G.F., Lebedeva L.M. Model' otsenki nadezhnosti elektroenergeticheskikh sistem pri dolgosrochnom planirovanii ikh raboty [Model for assessing the reliability of power systems in the long-term planning of their work]. *Elektrichestvo = Electrical Technology Russia*, 2000, no. 11, pp. 17–24.

4. Chukreev Yu.Ya. *Modeli obespecheniya nadezhnosti elektroenergeticheskikh sistem* [Models of power system reliability support]. Syktyvkar, Komi NTs UrO RAN Publ., 1995. 176 p.

5. Billinton R., Li W. *Reliability assessment of electric power systems using Monte Carlo methods*. New York, Springer, 1994. 352 p.

6. Li W. *Probabilistic transmission system planning*. Hoboken, NJ, Wiley-IEEE Press, 2011. 376 p.

7. Oboskalov V.P. Algoritmicheskie aspekty rascheta veroyatnostnykh pokazatelei defitsita moshchnosti v zadache balansovoi nadezhnosti OES [Algorithmic aspects of probabilistic power shortage indices calculation in the interconnected power system adequacy problem]. *Izvestiya Rossiiskoi akademii nauk. Energetika = Proceedings of the Russian Academy of Sciences. Power Engineering*, 2020, no. 2, pp. 59–74. DOI: 10.31857/S0002331020010094.

8. Working Group 601 of Study Committee C4. *Review of the current status of tools and techniques for risk-based and probabilistic planning in power systems*. International Conference on Large High Voltage Electric Systems, March 2010.

9. Fernandez Blanco Carramolino R., Careri F., Kavvadias K., Hidalgo Gonzalez I., Zucker A., Peteves E. *Systematic mapping of power system models*: Expert survey, EUR 28875 EN. Luxembourg, Publications Office of the European Union, 2017. DOI: 10.2760/422399.

10. Antonopoulos G., Chondrogiannis S., Kanellopoulos K., Papaioannou I., Spisto A., Efthimiadis T., Fulli G. *Assessment of underlying capacity mechanism studies for Greece*: EUR 28611 EN. Luxembourg, Publications Office of the European Union, 2017. DOI: 10.2760/51331.

11. Bertoldi O., Scalcino S., Salvaderi L. *Adequacy avaluation: an application of ENEL's SICRET program to New Brunswick power system*: WG 38/03/01. Montreal, CIGRE Symposium "Electric Power System Reliability". Montreal, Canada, 1991.

12. *Composite power system reliability analysis application to the New Brunswick Power Corporation System*: the draft report of the CIGRE Symposium on Electric Power Systems Reliability, Montreal; Canada, 1991.

13. RTE Antares. *Antares Optimization problems formulation*. Available at: https://antares.rte-france.com (accessed 13.01.2021).

14. RTE Antares. *Modelling of flow-based domains in Antares for adequacy studies*. Available at: https://antares.rte-france.com/?page_id=19&lang=en (accessed 04.05.2021).

15. Gaikwad A., Agarwal S., Carden K., Wintermantel N., Meliopoulos S., Kumbale M. *A study on probabilistic risk assessment for transmission and other resource planning*. Electric Power Research Institute for EISPC and NARUC. EPRI, 2015. (NARUC-2013-RFP027-DE0316).

16. Urrego A.L. *A novel method for the approximation of risk of Blackout in operational conditions*. Doctoral thesis. Paris, 2016.

17. Papic M., Bell K., Chen Y., Dobson I., Fonte L., Haq E., Hines P., Kirschen D., Luo X., Miller S.S., Samaan N., Vaiman M., Varghese M., Zhang P. Survey of tools for risk assessment of cascading outages. *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, 2011.

18. Hong Y.-Y., Lee L.-H. Reliability assessment of generation and transmission systems using fault-tree analysis. *Energy Conversion and Management*, 2009, vol. 50, pp. 2810–2817.

19. Siemens AG and Siemens Industry, Inc. *Model Management Module for PSS®E*. Available at: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed 15.12.2020).

20. Siemens AG and Siemens Industry, Inc. *Optimal Power Flow PSS®E*. Available at: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed 21.01.2021).

21. Siemens AG and Siemens Industry, Inc. *PSS®E. High-performance Transmission Planning and Operations Software for the Power Industry*. Art. EMDG-B10125-00-7600--PSSE Brochure. Available at: https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/pss-software/pss-e.html (accessed 23.12.2020).

22. PSR – Energy consulting and analytics. *OPTGEN User Manual*, *Version 7.4*. Rio de Janeiro, PSR, 2019.

23. ENTSO-E. *Mid-term adequacy forecast 2018. Appendix 1: Methodology and detailed results*. Available at: https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/MAF_2018_Methodology_and_Detailed_Results.pdf (accessed 04.05.2021).

24. *PLEXOS Market Simulation Software*. Available at: https://energyexemplar.com/solutions/plexos/ (accessed 01.02.2021).

25. Chu K. *MARS. Multi-area reliability simulation. EOP – on demand feature*. General Electric Company, 2014.

26. Jirutitijaroen P., Singh C. Reliability and cost trade-off in multi-area power system generation expansion using dynamic programming and global decomposition. *IEEE Transactions on Power Systems*, 2006, vol. 21, iss. 3, pp. 1432–1441.

27. Zorkaltsev V.I., Perzhabinsky S.M. Model' optimizatsii defitsita moshchnosti elektroenergeticheskoi sistemy [Model of power shortage optimization in electric power system]. *Upravlenie bol'shimi sistemami = Large-Scale Systems Control*, 2010, no. 30-1, pp. 300–318.

Для цитирования:

Якубовский Д.В., Крупенёв Д.С., Бояркин Д.А. Модель минимизации дефицита мощности электроэнергетических систем с учетом ограничений по контролируемым сечениям // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 95–120. – DOI: 10.17212/2782-2001-2021-2-95-120.

For citation:

Yakubovsky D.V., Krupenev D.S., Boyarkin D.A. Model' minimizatsii defitsita moshchnosti elektroenergeticheskikh sistem s uchetom ogranichenii po kontroliruemym secheniyam [A minimization model of the power shortage of electric power systems with regard to restrictions on controlled sections]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 95–120. DOI: 10.17212/2782-2001-2021-2-95-120.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 121–134

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 121–134

ПРИБОРОСТРОЕНИЕ, МЕТРОЛОГИЯ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ ПРИБОРЫ И СИСТЕМЫ

INSTRUMENT MAKING, METROLOGY AND INFORMATION MEASUREMENT DEVICES AND SYSTEMS

УДК 621.3.088.7+621.373.52 DOI: 10.17212/2782-2001-2021-2-121-134

Реализация метода адаптивного квантования в измерительных генераторах с цифровым управлением*

М.М. БАБИЧЕВ

630073, РФ, г. Новосибирск, пр. К. Маркса, 20, Новосибирский государственный технический университет

babichev@corp.nstu.ru

Измерительные генераторы с цифровым управлением, в частности, генераторы (калибраторы) фиктивной мощности, применяемые для поверки счетчиков электроэнергии, содержат цифро-аналоговый преобразователь (ЦАП), преобразующий отсчеты формируемого сигнала в напряжение. Отсчеты сигнала хранятся в памяти генератора. В связи с дискретизацией (квантованием) во времени и по уровню отсчетов сигнала в ЦАП возникает методическая погрешность дискретности (шум квантования). Относительное значение шума квантования зависит от амплитуды формируемого сигнала (сравнительно опорного напряжения ЦАП): чем больше амплитуда, тем больше старших разрядов ЦАП принимает участие в процессе преобразования и тем меньше относительное значение шума. В генераторах, где амплитуда выходного сигнала меняется в широких пределах (большой динамический диапазон) путем изменения цифровых отсчетов сигнала, шум квантования при малых амплитудах сигнала может стать недопустимо большим. Такая ситуация может наблюдаться в генераторе фиктивной мощности. В нем выходной ток меняется в сотни раз, поскольку погрешность поверяемого счетчика электроэнергии нормируется в широком диапазоне протекающего через него тока.

Предлагается новый алгоритм формирования отсчетов синусоидального сигнала в измерительных генераторах с цифровым управлением, называемый адаптивным квантованием. Адаптивное квантование способно существенно улучшить один из выбранных параметров сигнала (так называемый критерий оптимальности), например, снизить погрешность воспроизведения первой гармоники или уменьшить значение высших гармонических составляющих. Кроме того, предлагаемый алгоритм уменьшает зависимость выбранного параметра от частоты отсчетов и от количества используемых разрядов ЦАП, что позволяет расширить динамический диапазон генератора (в канале тока) без использования дополнительных усилителей с программируемым коэффициентом усиления. Исследования, проведенные при помощи компьютерного моделирования, подтвердили работоспособность алгоритма адаптивного квантования.

* *Статья получена 24 декабря 2020 г.*

Ключевые слова: счетчик электроэнергии, калибратор фиктивной мощности, измерительный генератор, цифро-аналоговый преобразователь, погрешность квантования, адаптивное квантование, критерий оптимальности, гармоники, коэффициент гармоник, динамический диапазон

ВВЕДЕНИЕ

Современные статические (т. е. электронные, а не индукционные) счетчики электрической энергии классов точности 0.5S и 0.2S работают в широком диапазоне изменения тока, протекающего через них, а соответственно, и в широком диапазоне мощности нагрузки, включенной на выходе счетчика. Согласно ГОСТу, основная инструментальная погрешность счетчика нормируется при изменении тока от $0.01I_{\text{ном}}$ до $I_{\text{макс}}$ [1, с. 4]. С учетом того, что номинальный ток $I_{\text{ном}}$ обычно выбирается равным 1/2 … 2/3 максимального тока $I_{\text{макс}}$, кратность тока, протекающего через счетчик, достигает 200. Также ГОСТ задает стартовый ток, равный $0.001I_{\text{ном}}$ [1, с. 7], это повышает кратность тока счетчика до 2000.

Для поверки счетчиков электрической энергии используют поверочные установки, в состав которых входят измерительные генераторы – калибраторы фиктивной мощности (power calibrators) [2, с. 9], далее ГФМ. ГФМ состоит из двух синхронизированных каналов: источника переменного напряжения и источника переменного тока. В качестве примера уже существующего и широко использующегося на производстве ГФМ можно привести КФМ-06 [3]. ГФМ должны обеспечивать изменение выходного тока в широких пределах, соответствующих приведенной выше кратности тока счетчика.

В генераторах сигналов с цифровым управлением (генераторах сигналов произвольной формы, или AWG [12, с. 17]) применяются цифро-аналоговые преобразователи (ЦАП), преобразующие цифровые коды, записанные в памяти генератора, в отсчеты аналогового сигнала, который затем усиливается / ослабляется, фильтруется и подается на выход генератора. Увеличить динамический диапазон генератора можно изменением кодов, подаваемых на вход ЦАП (чтобы уменьшить сигнал в 10 раз, уменьшают значение цифрового кода в 10 раз, округляя его, если нужно), либо управлением усилителями с программируемым коэффициентом усиления (PGA) [15], переключая обмотки в выходных трансформаторах тока / напряжения, и иными способами.

Способ с изменением кодов проще в реализации и позволяет упростить и удешевить схему генератора. Однако в этом случае при уменьшении выходного сигнала снижается разрешающая способность ЦАП (уменьшается количество двоичных разрядов, работающих в процессе преобразования), что увеличивает относительное значение шума квантования в выходном сигнале. Так, при шести двоичных разрядах младший разряд имеет значение 1.56 % от полной шкалы ЦАП, а при 16 разрядах – только 0.0015 % [4, p. 2.13]. Прочие способы усложняют конструкцию ГФМ и вносят дополнительные инструментальные погрешности за счет PGA или ключей. В настоящей статье рассматривается вопрос уменьшения шума квантования в широком диапазоне выходных сигналов при формировании сигнала с помощью ЦАП без использования PGA или переключаемых обмоток.

1. ПОСТАНОВКА ЗАДАЧИ

Сигнал на выходе ЦАП является ступенчатой функцией, имеющей конечное число уровней напряжения (уровней квантов). Переход от одного уровня к другому осуществляется в фиксированные (дискретные) моменты времени. Таким образом, сигнал на выходе ЦАП подвергается округлению и по уровню, и по времени.

В отличие от непрерывного («аналогового») сигнала, получаемый с помощью ЦАП сигнал содержит погрешность квантования, которая по своему происхождению является методической, то есть зависит не от инструментальных погрешностей используемого ЦАП, а от его количества двоичных разрядов, частоты отсчетов и от используемого способа квантования (округления) отсчетов выходного сигнала. Как было обнаружено в ходе моделирования процесса цифро-аналогового преобразования кодов, описывающих синусоидальный сигнал, погрешность квантования также существенно меняется при изменении количества отсчетов на период формируемого сигнала даже на единицу [9].

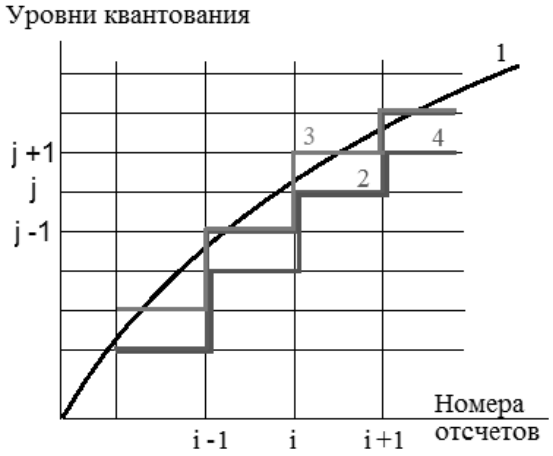


Рис. 1. Иллюстрация основных способов округления сигнала по уровню

Fig. 1. Illustration of the main methods of signal quantization by level

В ЦАП, по аналогии с АЦП, могут использоваться три способа [5] округления по напряжению отсчетов выходного сигнала (рис. 1): до соседних верхних (график *3*), до соседних нижних (график *4*) и до ближайших (график *2*) уровней квантования, относящихся к значениям сигнала без округления *1*, который нужно получить в идеале. Поскольку при округлении до ближайших уровней шум квантования минимален (не превышает по модулю 1/2 уровня кванта), этот способ округления в настоящее время используется чаще всего. Он взят за основу предлагаемого в данной статье алгоритма.

При цифро-аналоговом преобразовании кодов сигнала, форма которого заранее неизвестна (например, аудиосигнала) ЦАП просто преобразует имеющиеся цифровые коды сигнала в соответствующие им уровни аналогового сигнала, осуществляя округление сигнала единственным по сути доступным

способом. Однако в цифровых генераторах, формирующих известные сигналы (например, синусоиду заданной амплитуды и частоты), создание кодов генерируемого сигнала и запись их в память генератора осуществляются заранее. Это позволяет округлять формируемый сигнал разными способами, то есть приводить каждый код формируемого сигнала к любому уровню кванта (относительно уровня идеального неокругленного сигнала): к верхнему $j + 1$, к нижнему j, к ближайшему j (для i-го отсчета на рис. 1), а возможны и другие варианты. Руководствуясь некоторыми соображениями, о них речь пойдет ниже. Назовем такой рациональный процесс априорного формирования кодов сигнала, впоследствии подаваемых на ЦАП, **адаптивным квантованием** сигнала.

Следует заметить, что термин «адаптивное квантование» уже давно известен. Под ним обычно понимают квантование в АЦП с переменным шагом по времени или по уровню (нелинейное квантование). Например, если известно, что преобразуемый сигнал большую часть времени имеет уровень, находящийся в определенном достаточно узком диапазоне (плотность вероятностей сигнала максимальна в диапазоне), то в этом диапазоне сетка уровней квантов может быть более частой, чем в других диапазонах. Таким образом, без увеличения количества двоичных разрядов АЦП / ЦАП можно более точно описывать форму сигнала, снизив шум квантования, либо уменьшить количество передаваемых (сохраняемых) данных. Примеры применения адаптивного квантования: измерения в области технической защиты информации, касающиеся разборчивости речевого сигнала [6], сжатие изображений, получаемых при радиолокации [13], сжатие аудиосигналов [14].

В настоящей статье под адаптивным квантованием подразумевается только цифро-аналоговое преобразование с постоянным шагом по времени и по уровню кодов заранее известного сигнала. Эти коды до подачи в ЦАП подвергаются модификации (адаптации или адаптивному квантованию) с той целью, чтобы аналоговый сигнал на выходе ЦАП с модифицированными кодами был лучше (по некоторому критерию) сигнала с немодифицированными кодами. При адаптации кодов выходного сигнала алгоритм адаптивного квантования допускает смещение отдельных отсчетов выходного сигнала на целое число уровней квантования вверх или вниз относительно уровней отсчетов немодифицированного (неадаптированного) сигнала.

Возникает вопрос: может ли адаптивное квантование быть полезным на практике? Чтобы дать на него ответ, зададимся критерием оптимальности квантования. Это может быть любой измеряемый и выражаемый в числовой форме параметр формируемого сигнала, который нужно оптимизировать. Для генератора синусоидального сигнала критерием оптимальности может быть погрешность воспроизведения основной (первой) гармоники, значение N-й высшей гармонической составляющей сигнала ($N \geq 2$), среднее квадратическое значение нескольких высших гармонических составляющих и так далее. В случае, если критерием выбрана погрешность основной гармоники или значение высшей гармонической составляющей, улучшение критерия оптимальности означает уменьшение модуля его значения.

Далее следует провести эксперимент и определить, насколько адаптивное квантование способно улучшить данный критерий. Поскольку шум квантования является методической погрешностью, целесообразно экспериментировать с моделью «идеального ЦАП», который не будет искажать резуль-

таты эксперимента своими инструментальными погрешностями. Таким образом, для данных исследований целесообразно использовать компьютерное моделирование.

2. АЛГОРИТМ АДАПТИВНОГО КВАНТОВАНИЯ

Алгоритм адаптивного квантования сигнала описан в патенте [7]. Он состоит из нескольких последовательных этапов.

1. Производится округление отсчетов генерируемого сигнала до ближайших уровней квантования (график *2* на рис. 1), и такой сигнал считается исходным.

2. Вычисляется и запоминается значение критерия оптимальности исходного сигнала. Переходят к произвольному отсчету сигнала i, с которого начнется оптимизация (поскольку сигнал периодический, начинать оптимизацию можно с любого его отсчета).

3. На текущем отсчете сигнала направление округления изменяют на противоположное. Это значит, что если на i-м отсчете сигнала график *2* был округлен до уровня j и был выше графика *1*, то устанавливается уровень $j - 1$; если же график *2* был ниже графика *1*, то устанавливается уровень $j + 1$.

4. Снова вычисляют значение критерия оптимальности сигнала. Если полученное значение лучше ранее запомненного, то изменение направления округления сигнала сохраняют, значение полученного критерия оптимальности также запоминают. Если же полученный критерий не лучше исходного, направление округления текущего отсчета сигнала и значение критерия оптимальности оставляют прежними.

5. Переходят к следующему отсчету сигнала, повторяют п. 3 и 4 (если дошли до последнего отсчета сигнала, то переходят к первому), и так далее, вплоть до того момента, когда окажется, что любое изменение направления округления любого отсчета сигнала не улучшает значение критерия оптимальности. На этом адаптивное квантование сигнала завершается.

Алгоритм адаптивного квантования, по сути, можно отнести к генетическим алгоритмам оптимизации [8]. Поиск оптимального варианта осуществляется путем перебора на каждом шагу нескольких возможных вариантов («мутаций») сигнала и «селекцией» наилучшего варианта. Далее «гены» выбранного варианта сигнала сохраняются и используются для формирования дальнейших «мутаций», из которых снова осуществляется «селекция», и так до того момента, когда любые последующие «мутации» будут получаться не лучше предыдущих. На этом процесс завершают.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Предложенный способ адаптивного квантования реализован в виде компьютерной модели идеального N-разрядного ЦАП, формирующего периодический дискретный сигнал, имеющий M отсчетов на период, в котором шум квантования определяется только процессами дискретизации и квантования [4, p. 2.37], а не инструментальными погрешностями преобразователя.

На начальном этапе моделирования процесса адаптивного квантования был сформирован дискретный синусоидальный сигнал с малым, для наглядности, количеством уровней квантов, соответствующим 4-разрядному (включая знаковый разряд) ЦАП. Количество отсчетов на период сигнала также невелико – 16 (рис. 2). Белые столбцы на графике – отсчеты сигнала, соответствующие кодам, рассчитанным так, чтобы они были ближе всего к соответствующим отсчетам идеальной неокругленной синусоиды (округление до ближайших уровней квантов), т. е. исходный сигнал. Черные столбцы – отсчеты исходного сигнала, подвергнутого адаптивному квантованию по такому критерию оптимальности, как минимум погрешности основной гармоники. Видно, что в процессе адаптации программа модифицировала отсчеты номер 3, 4, 8 и 16, поменяв направление округления как в большую, так и в меньшую стороны.

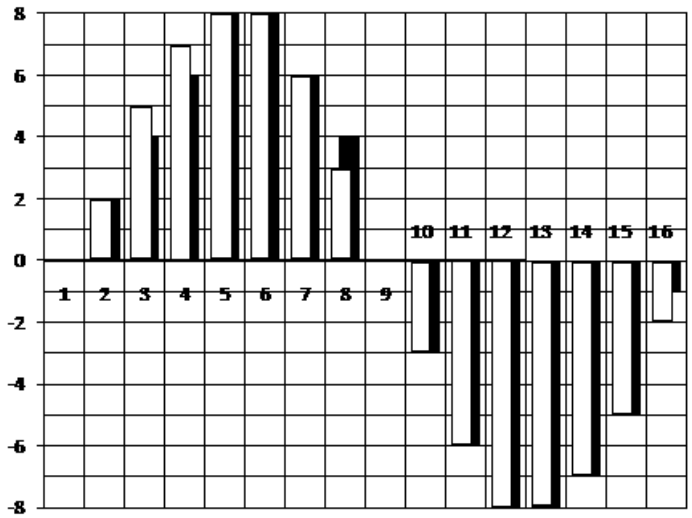
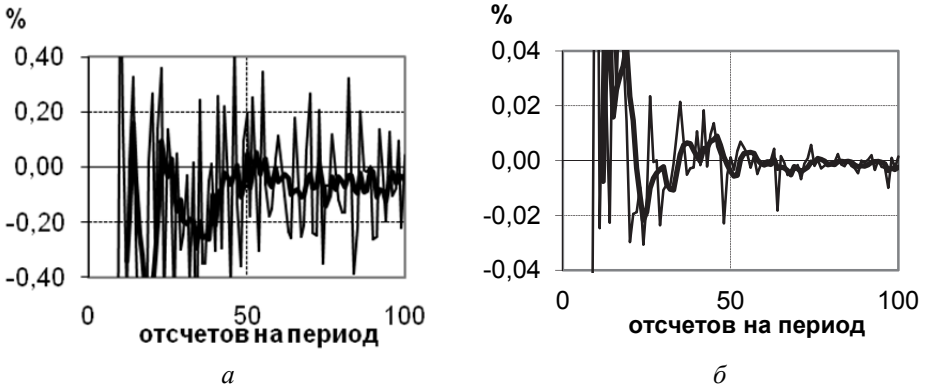


Рис. 2. Пример реализации алгоритма адаптивного квантования

Fig. 2. An example of the implementation of the adaptive quantization algorithm

Многочисленные эксперименты с программной моделью алгоритма адаптивного квантования показали, что «адаптация» сигнала обычно требует 2-3 перебора всех отсчетов сигнала, но в некоторых случаях процесс оптимизации затягивается до пяти проходов.

Рисунок 3 показывает зависимость относительной погрешности воспроизведения основной гармоники (выраженной в процентах от значения основной гармоники) от числа дискретных отсчетов на периоде для сигнала, формируемого идеальным ЦАП с шестью двоичными разрядами. При этом график на рис. 3, *а* показывает кривую погрешности для сигнала, сформированного из кодов, рассчитанных с округлением до ближайших уровней квантов. График на рис. 3, *б* демонстрирует тот же результат для сигнала, сформированного путем адаптивного квантования по критерию погрешности основной гармоники. Для выявления интегральной зависимости погрешностей от количества отсчетов на графиках сформирована жирная линия – скользящее среднее по пяти отсчетам.



а *б*

Рис. 3. Адаптивное квантование с оптимизацией погрешности основной гармоники

Fig. 3. Adaptive quantization with optimization of the fundamental error

Следует обратить внимание, что вертикальный масштаб на рис. 3, *б* в 10 раз меньше, чем на рис. 3, *а*. Без использования адаптивного квантования (рис. 3, *а*) погрешность основной гармоники сильно зависит от частоты дискретизации сигнала. Малейшее изменение частоты формируемого сигнала за счет изменения количества отсчетов на период сигнала (при неизменном шаге дискретизации), например с 47 до 48, зачастую вызывает существенное изменение погрешности, в том числе смену ее знака, как было отмечено еще в работе [9]. Используя метод адаптивного квантования, можно значительно уменьшить амплитуду колебаний графика погрешности, как видно на рис. 3, *б*.

Вторым важным преимуществом адаптивного квантования является то, что оно снижает погрешность воспроизведения основной гармоники сигнала более чем на порядок при условии, что отсчетов на период не менее 25 – эффективность алгоритма повышается с ростом количества отсчетов. Это можно объяснить тем, что при большем количестве отсчетов увеличивается количество вариантов, которые алгоритм может перебирать в процессе оптимизации.

Было обнаружено, что многократное уменьшение погрешности воспроизведения основной гармоники может достигаться ценой ухудшения некоторых других параметров сигнала, прежде всего гармонических составляющих. На рис. 4 показана часть спектра (гармоники со 2-й по 10-ю) двух сигналов, сформированных из кодов, округленных до ближайших уровней квантов (косые штриховые линии) и адаптированных согласно критерию погрешности основной гармоники (горизонтальные штриховые линии). Такая форма спектра соответствует ЦАП с шестью разрядами, частота дискретизации соответствует 50 точкам на период синусоиды. Масштаб вертикальной оси на рис. 4 соответствует процентам от значения основной гармоники.

В некоторых случаях бывает целесообразно уменьшить значение одной из высших гармонических составляющих сигнала на низкой частоте (2-я, 3-я, 5-я, 7-я, 9-я гармоники), если она относительно велика, что позволит упростить фильтрацию сигнала с выхода ЦАП и снизить коэффициент гармоник сигнала на выходе генератора [10]. Коэффициент гармоник в измерительных генераторах нормирован, в том числе в генераторах фиктивной мощности,

используемых для поверки счетчиков электрической энергии, где он не должен превышать 5 % [2, с. 9]. На рис. 5 показано применение адаптивного квантования, для которого в качестве критерия оптимальности выбрано значение 3-й гармоники. По вертикальной оси – значение 3-й гармоники сигнала (в процентах от амплитуды 1-й гармоники), ЦАП 12-разрядный. На рис. 5, *а* показано округление до ближайшего уровня, на рис. 5, *б* – адаптивное.

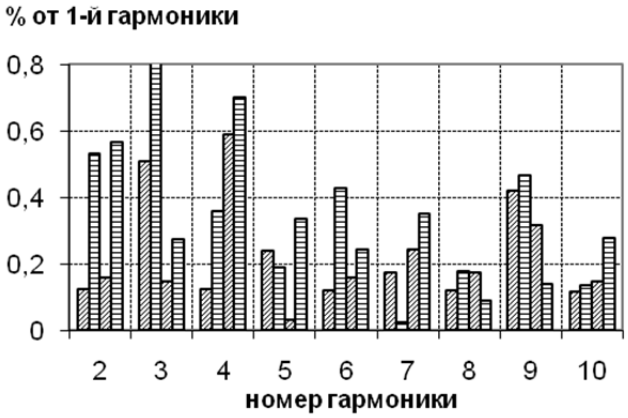
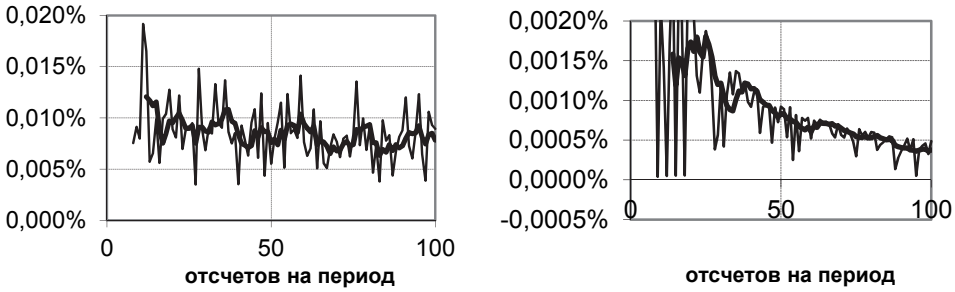


Рис. 4. Сравнение гармонических составляющих двух сигналов: обычного и адаптированного по минимуму погрешности 1-й гармоники

Fig. 4. Comparison of harmonic components of two signals: conventional and adapted for the minimum error of the fundamental



а *б*

Рис. 5. Адаптивное квантование с оптимизацией 3-й гармоники

Fig. 5. Adaptive quantization with the 3rd harmonic optimization

На рис. 6 показана часть спектра синусоиды, воспроизводимой ЦАП при округлении отсчетов до ближайших уровней (наклонная штриховка) и с адаптацией (горизонтальная штриховка) по минимуму 3-й гармоники. ЦАП имеет 12 разрядов, период синусоиды содержит 55 отсчетов. Ценой увеличения 2-й, 4-й, 5-й… гармоник достигается уменьшение значения 3-й гармоники в несколько раз.

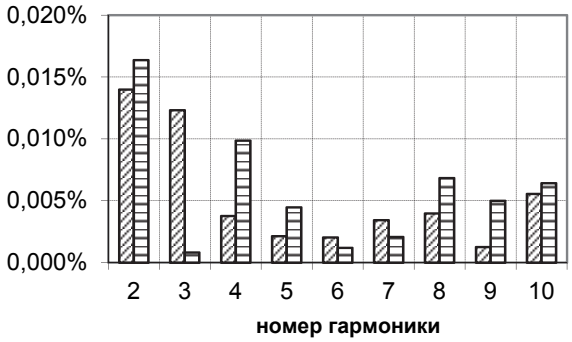


Рис. 6. Сравнение гармонических составляющих двух сигналов: обычного и адаптированного по минимуму 3-й гармоники

Fig. 6. Comparison of harmonic components of two signals: conventional and adapted by the minimum of the 3rd harmonic

На рис. 7 показано отношение между значениями 3-й гармоники без адаптивного квантования и с использованием адаптивного квантования, подавляющего эту гармонику. Этот пример относится к 12-разрядному ЦАП. Для лучшего понимания закономерностей на графике приведено скользящее среднее (по пяти точкам) для отношения – жирная кривая. Можно сделать вывод, что подавление 3-й гармоники улучшается с ростом количества отсчетов на период. Ослабление 3-й гармоники не менее четырех раз при 40 отсчетах на период и не менее восьми раз при 80 и более отсчетах на период.

Рис. 7. Уменьшение значения 3-й гармоники при адаптации сигнала (критерий оптимальности – 3-я гармоника)

Fig. 7. Reduction the value of the 3rd harmonic decreases by adaptive quantization (the criterion of optimality is the 3rd harmonic)

Метод адаптивного квантования позволяет также уменьшать одновременно две или три гармоники сигнала. Так, на рис. 8 изображен пример (55 отсчетов на период, 9 разрядов ЦАП) спектров синусоиды, квантованной до ближайшего уровня (наклонная штриховка), и адаптивно с подавлением 2-й

и 3-й гармоник (горизонтальная штриховка). В качестве критерия оптимальности использовался минимум корня из суммы квадратов 2-й и 3-й гармоник сигнала.

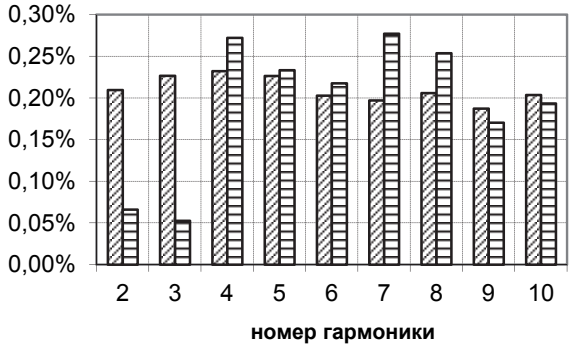


Рис. 8. Сравнение гармонических составляющих двух сигналов: обычного и адаптированного по минимуму корня из суммы квадратов 2-й и 3-й гармоник

Fig. 8. Comparison of harmonic components of two signals: normal and adapted by the minimum of the root of sum of squares of the 2nd and 3rd harmonics

На рис. 9 показано, во сколько раз в зависимости от числа отсчетов на период сигнала метод адаптивного квантования минимизирует более сложный критерий оптимальности – корень из суммы квадратов двух гармоник сигнала (2-й и 3-й). Было обнаружено, что одновременное подавление двух гармоник производится с меньшей эффективностью, чем ослабление одной гармоники. Однако при 60 и более отсчетах на период всё же удается достичь суммарного подавления двух гармоник в 4–8 раз, тогда как при меньшем количестве отсчетов на период существенного ослабления двух гармоник не получается.

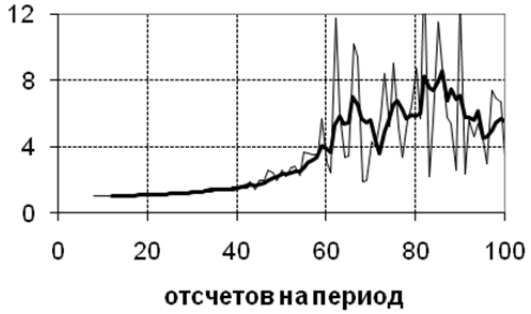


Рис. 9. Уменьшение корня из суммы квадратов 2-й и 3-й гармоник при адаптации сигнала (критерий оптимальности – 2-я и 3-я гармоники)

Fig. 9. Reduction the root of the sum of the squares of the 2nd and 3rd harmonics decreases when adapting the signal (the optimality criterion is the 2nd and 3rd harmonics)

ЗАКЛЮЧЕНИЕ

Предлагаемый алгоритм адаптивного квантования, описанный в виде способа округления кодов функции [7] и испытанный на компьютерной модели, применим для формирования заранее известных периодических сигналов в генераторах произвольных сигналов, например, генераторах фиктивной мощности с цифровым управлением (на основе ЦАП). Метод адаптивного квантования с последовательным подбором уровней квантов на каждом отсчете сигнала улучшает так называемый критерий оптимальности (оптимизируемый параметр). В некоторых случаях критерий улучшается на порядок и более. После этого коды «адаптированного» сигнала записываются в память генератора и в дальнейшем им воспроизводятся.

В качестве критериев оптимальности можно привести минимум погрешности основной гармоники синусоиды либо минимум одной или двух гармонических составляющих синусоиды по выбору. Минимизация 2-й, 3-й, 4-й, 5-й… гармонических составляющих позволяет улучшить фильтрацию сигнала на выходе ЦАП простым ФНЧ низкого порядка (использование фильтров высокого порядка, вносящих значительные амплитудные и фазовые погрешности, в генераторах фиктивной мощности нежелательно).

Следует отметить, что улучшение выбранного критерия оптимальности неизбежно сопровождается ухудшением некоторых иных свойств сигнала. Так, адаптивное квантование отсчетов синусоиды для 10-разрядного ЦАП позволяет в несколько раз уменьшить относительную погрешность воспроизведения основной гармоники (при условии, что на один период синусоиды приходится не менее 40 отсчетов сигнала) ценой некоторого увеличения коэффициента гармоник.

Дополнительным эффектом от применения адаптивного квантования в генераторах сигналов является уменьшение зависимости оптимизируемого параметра от числа дискретных отсчетов на период сигнала. Это позволяет менять в небольших пределах частоту формируемого сигнала (либо частоту дискретизации сигнала той же частоты) без значительного ухудшения оптимизируемого параметра.

Все это позволяет, в частности, расширить диапазон амплитуд выходного сигнала генератора, обеспечиваемый только сменой кодов в ЦАП, до 1000 без использования управляемых усилителей PGA. Для этого нужно оптимизировать отсчеты синусоиды, как минимум, для 6-разрядного ЦАП, что будет соответствовать минимальному уровню сигнала. При этом будут задействованы 6 младших разрядов 16-разрядного ЦАП. Максимальный (в 1000 раз больше) уровень сигнала при этом будет получаться при использовании всех 16 разрядов ЦАП ($2^{10} = 1024 \approx 1000$), а такая разрядность легко достижима на практике.

СПИСОК ЛИТЕРАТУРЫ

1. ГОСТ 31819.22–2012. Статические счетчики активной энергии классов точности 0,2 S и 0,5 S. – М.: Стандартинформ, 2013. – 15 с.

2. ГОСТ 8.584–2004. Счетчики статические активной электрической энергии переменного тока. Методика поверки: введ. 2005–01–01. – М.: Изд-во стандартов, 2004. – 14 с.

3. *Пасынков Ю.А., Савиных М.А.* Технологический контроль метрологических характеристик счетчиков электроэнергии на производстве // Научный вестник НГТУ. – 2018. – № 1 (70). – С. 177–188. – DOI: 10.17212/1814-1196-2018-1-177-188.

4. Analog-digital conversion / ed. by W. Kester. – Analog Devices Inc., 2004. – ISBN 0-916550-27-3.

5. Основы метрологии и электрические измерения: учебник для вузов / Б.Я. Авдеев, Е.М. Антонюк, Е.М. Душин и др.; под ред. Е.М. Душина. – 6-е изд., перераб. и доп. – Л.: Энергоатомиздат, 1987. – 480 с.

6. *Продеус А.Н., Дидковский В.С., Гавриленко А.В.* Расчет и измерение разборчивости речи при малых отношениях сигнал–шум. Ч. 1. Корректное измерение функции распределения речевого сигнала // Электроника и связь. – 2007. – № S1: Тематический выпуск «Проблемы электроники». – С. 137–141.

7. Патент № 2420799 Российская Федерация, МПК G 06 F 17/7. Способ округления кодов функции / М.М. Бабичев, Ю.А. Пасынков. – Заявл. 27.05.2009; опубл. 10.06.2011.

8. *Mitchell M.* An introduction to genetic algorithm. – Cambridge, MA: MIT Press, 1996. – 205 p.

9. *Babichev M.M., Pasynkov Yu.A.* Error estimators of discreteness while gaging the RMS of periodic signals // Актуальные проблемы электронного приборостроения (АПЭП-2014): труды XII международной конференции, 2–4 октября 2014 г.: в 7 т. – Новосибирск: Изд-во НГТУ, 2014. – Т. 1. – С. 286–289.

10. *Вакулин А.А., Беляков В.В., Бабичев М.М.* Улучшение фильтрации синусоидального сигнала на выходе измерительного генератора путем выбора критерия оптимальности адаптивного квантования // Наука. Технологии. Инновации: сборник научных трудов, Новосибирск, 3–7 декабря 2018 г. – Новосибирск: Изд-во НГТУ, 2018. – Ч. 1. – С. 10–12. – ISBN 978-5-7782-3734-6.

11. *Babichev M.M., Vakulin A.A.* Features of the application of adaptive quantization in generators with digital control // Актуальные проблемы электронного приборостроения (АПЭП-2018): труды 14 международной конференции, Новосибирск, 2–6 октября 2018 г. – Новосибирск, 2018. – Т. 1, ч. 2. – С. 18–21. – ISBN 978-5-7782-3614-1.

12. Генераторы сигналов от А до Я: учебное пособие. – Tectronix, 2008. – URL: https://doc.platan.ru/docs/pdf/generatory-signalov.pdf (дата обращения: 12.12.2020).

13. *Нестеров И.М.* Влияние сжатия данных на качество радиолокационных изображений // Журнал радиоэлектроники. – 2016. – № 8. – URL: http://jre.cplire.ru/jre/aug16/5/text.html (дата обращения: 25.11.2020).

14. *Поров А.В.* Адаптивное скалярное квантование спектральных коэффициентов для систем сжатия аудиосигналов: дис. … канд. техн. наук: 05.13.01 / А.В. Поров. – СПб., 2009. – 136 с.

15. *Шерри А*. Усилитель с программируемым коэффициентом усиления (PGA) в Сигма-Дельта АЦП: AN-610R. Руководство по применению. – 2011. – URL: https://www.analog.com/media/ru/technical-documentation/application-notes/41175119575886AN610R.pdf (дата обращения: 10.11.2020).

Бабичев Михаил Михайлович, старший преподаватель кафедры защиты информации Новосибирского государственного технического университета. Основное направление научных исследований – измерения интегральных характеристик периодических сигналов, измерения в электроэнергетике, способы уменьшения погрешности дискретности в генераторах фиктивной мощности. Является автором 27 публикаций, включая 2 учебных пособия, одну статью в ведущих научных отечественных и зарубежных издательствах; 4 публикации, входящие в систему Scopus, одного патента на способ. E-mail: babichev@corp.nstu.ru

Babichev Mikhail M., lecturer at the Department of Information Security in the Novosibirsk State Technical University. The main areas of his scientific interests are measurements of integral characteristics of periodic signals, measurements in the electric power industry, and methods of reducing the discreteness error in power calibrators. He is the author of 27 publications, including 5 papers in leading scientific domestic and foreign journals among them 4 publications in the Scopus data base system, as well as 2 textbooks and 1 patent for the method. E-mail: babichev@corp.nstu.ru

DOI: 10.17212/2782-2001-2021-2-121-134

Implementation of the adaptive quantization method in digitally controlled measuring generators *

M.M. BABICHEV

Novosibirsk State Tehcnical University, 20 K Marx Prospekt, Novosibirsk, 630119, Russian Federation

babichev@corp.nstu.ru

Abstract

Measuring generators with digital control, in particular power calibrators, used to calibrate electricity meters, contain a digital-to-analog converter (DAC) that converts codes of the generated signal into voltage. Signal codes are stored in the generator memory. A truncation discreteness error (quantization noise) arises caused by sampling (quantization) in time and by the level of signal samples in the DAC. A relative value of the quantization noise depends on the amplitude of the generated signal (relative to the reference voltage of the DAC): the larger the amplitude, the more significant bits of the DAC are involved in the conversion process, and the less the relative value of the noise. In generators, where the amplitude of the output signal changes over a wide range (high dynamic range) by changing the digital samples of the signal, the quantization noise at low signal amplitudes can become unacceptably large. This situation occurs in power calibrators where the output current changes hundreds of times since the error of the verified electricity meter is normalized in a wide range of current flowing through it.

A new algorithm for generating samples of a sinusoidal signal in measuring generators with digital control called adaptive quantization is proposed. Adaptive quantization can significantly improve one of the selected signal parameters (the so-called optimality criterion), for example, reduce the error in reproduction of the first harmonic, or reduce the value of higher harmonic components. In addition, the proposed algorithm reduces the dependence of the selected parameter on the sampling frequency and on the number of DAC bits used, which makes it possible to expand the dynamic range of the generator (in the current channel) without using additional amplifiers with programmable gain (PGA). Studies carried out using computer simulation have confirmed the efficiency of the adaptive quantization algorithm.

Keywords: electricity meter, power calibrator, measuring generator, digital-to-analog converter, quantization error, adaptive quantization, optimality criterion, harmonics, THD, dynamic range

REFERENCES

1. GOST 31819.22–2012. *Staticheskie schetchiki aktivnoi energii klassov tochnosti 0,2S i 0,5S* [State standard 31819.22–2012. Static meters for active energy (classes 0,2S и 0,5S)]. Moscow, Standartinform Publ., 2013. 15 p.

2. GOST 8.584–2004. *Schetchiki staticheskie aktivnoi elektricheskoi energii peremennogo toka. Metodika poverki* [Alternating current static meters for active energy. Methods of verification]. Moscow, Standards Publ., 2004. 14 p.

3. Pasinkov Y.A., Savinykh M.A. Tekhnologicheskii kontrol' metrologicheskikh kharakteristik schetchikov elektroenergii na proizvodstve [Technological control of metrological characteristics of electricity meters during production]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Science bulletin of the Novosibirsk state technical university*, 2018, no. 1 (70), pp. 177–188. DOI: 10.17212/1814-1196-2018-1-177-188.

4. Kester W., ed. *Analog-digital conversion*. Analog Devices Inc., 2004. ISBN 0-916550-27-3.

5. Dushin E.M., ed. *Osnovy metrologii i elektricheskie izmereniya* [Fundamentals of metrology and electrical measurements]. 6th ed., rev. and add. Leningrad, Energoatomizdat Publ., 1987. 480 p.

6. Prodeus A.N., Didkovskii V.S., Gavrilenko A.V. Raschet i izmerenie razborchivosti rechi pri malykh otnosheniyakh signal-shum. Ch. 1. Korrektnoe izmerenie funktsii raspredeleniya rechevogo

* *Received 24 December 2020.*

signala [Calculation and measurement of speech intelligibility at low signal-to-noise ratios. Pt. 1. Correct measurement of the speech signal distribution function]. *Elektronika i svyaz' = Electronics and communication*, 2007, no. S1, pp. 137–141.

7. Babichev M.M., Pasynkov Yu.A. *Sposob okrugleniya kodov funktsii* [Method of rounding function codes]. Patent RF, no. 2420799, 2011.

8. Mitchell M. *An introduction to genetic algorithm*. Cambridge, MA, MIT Press, 1996. 205 p.

9. Babichev M.M., Pasynkov Yu.A. Error estimators of discreteness while gaging the RMS of periodic signals. *Aktual'nye problemy elektronnogo priborostroeniya, APEP-2014: trudy XII mezhdunarodnoi konferentsii* [12th International Conference on Actual Problems of Electronic Instrument Engineering, APEIE-2014: Proceedings], Novosibirsk, Russia, 2–4 October 2014, vol. 1, pp. 286–289.

10. Vakulin A.A., Belyakov V.V., Babichev M.M. [Improving the filtering of a sinusoidal signal at the output of the measuring generator by choosing the criterion of optimality of adaptive quantization]. *Nauka. Tekhnologii. Innovatsii*: sbornik nauchnykh trudov [Science. Technologies. Innovations: collection of scientific papers]. Novosibirsk, December 3–7, 2018, pt. 1, pp. 10–12. ISBN 978-5-7782-3734-6.(In Russian).

11. Babichev M.M., Vakulin A.A. Features of the application of adaptive quantization in generators with digital control. *Aktual'nye problemy elektronnogo priborostroeniya, APEP-2018: trudy XIV mezhdunarodnoi konferentsii* [14th International Conference on Actual Problems of Electronic Instrument Engineering, APEIE-2014: Proceedings], Novosibirsk, October 2–6, 2018, vol. 1, pt. 2, pp. 18–21. ISBN 978-5-7782-3614-1.

12. *Generatory signalov ot A do Ya* [Signal generators from A to Z]. Tectronix, 2008. Available at: https://doc.platan.ru/docs/pdf/generatory-signalov.pdf (accessed 12.12.2020).

13. Nesterov I.M. Vliyanie szhatiya dannykh na kachestvo radiolokatsionnykh izobrazhenii [Influence of data compression on the quality of radar images]. *Zhurnal radioelektroniki = Journal of Radio Electronics*, 2016, no. 8. Available at: http://jre.cplire.ru/jre/aug16/5/text.html (accessed 25.11.2020).

14. Porov A.V. *Adaptivnoe skalyarnoe kvantovanie spektral'nykh koeffitsientov dlya sistem szhatiya audio signalov*. Diss. kand. tekhn. nauk [Adaptive scalar quantization of spectral coefficients for audio signal compression systems. PhD eng. sci. diss.]. St. Petersburg, 2009. 136 p.

15. Sherri A. The PGA on Sigma-Delta ADCs: AN-610. Application Note. (In Russian). Available at: https://www.analog.com/media/en/technical-documentation/application-notes/AN-610.pdf (accessed 10.11.2020).

Для цитирования:

Бабичев М.М. Применение адаптивного квантования в измерительных генераторах с цифровым управлением // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 121–134. – DOI: 10.17212/2782-2001-2021-2-121-134.

For citation:

Babichev M.M. Primeneniye adaptivnogo kvantovaniya v izmeritel'nykh generatorakh s tsifrovym upravleniyem [Implementation of the adaptive quantization method in digitally con-trolled measuring generators]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 121–134. DOI: 10.17212/2782-2001-2021-2-121-134.

ISSN 2782-2001
Системы анализа
и обработки данных
том 82, № 2, 2021, с. 135–145

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 82, No. 2, 2021, pp. 135–145

ПРИБОРОСТРОЕНИЕ, МЕТРОЛОГИЯ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ ПРИБОРЫ И СИСТЕМЫ

INSTRUMENT MAKING, METROLOGY AND INFORMATION MEASUREMENT DEVICES AND SYSTEMS

УДК 004.021 DOI: 10.17212/2782-2001-2021-2-135-145

Разработка аппаратно-программного комплекса анализа и коррекции речи*

Д.В. БОРОВИКОВА[a], О.В. ГРИШИН[b], А.В. НЕНЬКО[c], А.В. ЮПАШЕВСКИЙ[d], А.С. КАЗЬМИНА[e], А.В. МАРКОВ[f], К.А. МЕЦЛЕР[g]

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

[a] *borovikova@corp.nstu.ru* [b] *8asteroid@gmail.com*
[c] *nenkoanastasia507@gmail.com* [d] *up1997ant@gmail.com*
[e] *kazmina.anny@gmail.com* [f] *avm@332lab.ru* [g] *kostya.metsler@gmail.com*

В последние годы наблюдается резкое увеличение числа лиц, страдающих функциональными нарушениями голосовой функции, как правило, обусловленными психоэмоциональным стрессом. Подобные расстройства приносят в жизнь человека существенный дискомфорт: они снижают способность к коммуникации и социальной адаптации, что, в свою очередь, усиливает психоэмоциональную нагрузку. В результате функциональные нарушения закрепляются по механизму порочного круга и могут трансформироваться в патологию речевого аппарата. Основным способом диагностики остается экспертная оценка, которая напрямую зависит от профессиональных навыков специалиста по работе с голосом. В связи с этим актуальным представляется вопрос разработки таких систем диагностики голосоречевых нарушений, которые позволили бы провести объективную оценку на основе обработки голосоречевых характеристик, а также вовремя выявить нарушение и предотвратить развитие патологии. Такие методы и системы могут быть полезны как для диагностики, так и для контроля эффективности голосовой терапии. Существующие методики аппаратной диагностики до сих пор не нашли своего применения на практике из-за несоответствия результатам экспертной оценки. В настоящей работе предлагается новая концепция аппаратно-программного комплекса анализа голоса на основе акустических характеристик по совокупности гармоник голосового сигнала. Разработан ком-плекс VASA (Voice and Speech Analysing system), обеспечивающий автоматический анализ значений амплитуд первых 16 гармоник. Выполненные на трех добровольцах испытания показали высокий уровень воспроизводимости и повторяемости (в пределах 10 % < %R&R < 30 %), достаточный для проведения сравнительных исследований у здоровых людей и лиц с функ-циональными нарушениями речевого аппарата.

Ключевые слова: фониатрия, дисфония, джиттер, шиммер, цифровая обработка сигнала, акустический анализ, объективный анализ, R&R анализ, автоматизированный комплекс

* *Статья получена 14 января 2021 г.*

ВВЕДЕНИЕ

Функциональные расстройства речи или голоса – это группа нарушений центрального и периферического характера, обусловленных изменением функции без органических повреждений голосового аппарата. Значительную роль в функциональных расстройствах играют анатомические резонаторы, участвующие в формировании голосовых звуков [1]. Резонаторы формируют специфический тембральный оттенок голоса – обертона или «призвуки», которые образуются за счет перераспределения целого ряда гармоник и являются добавочными к основному тону звука [2].

В настоящее время оценка функционального состояния голосового аппарата в России и за рубежом до сих пор основывается на экспертной оценке специалистов по работе с голосом – фонопедов [3]. Фонопед субъективно на слух определяет степень «включения» того или иного резонатора в процессе фонации и отражает это в виде стандартизованной шкалы [4, 5]. На этом основании делается заключение о наличии в голосе того или иного типа функционального расстройства. Такая экспертная диагностика полностью зависит от профессионализма специалистов.

Вместе с этим существует несколько подходов объективной диагностики функциональных расстройств. В основе большинства из них лежит акустический анализ голоса с последующей оценкой шумовой составляющей, изменения частоты основного тона (ЧОТ) и определения общих спектральных характеристик звукового сигнала (формантный и гармонический состав) [6, 7].

Одним из наиболее распространенных методов комплексной оценки качества голосового сигнала является расчет коэффициента степени выраженности дисфонии (DSI – dysphonia severity index) [8]. Расчет DSI проводится по четырем параметрам, оценивающим разные характеристики голосового сигнала: максимальное время фонации (*MPT*, с) [9], наивысшая частота (F_{high}, Гц), наименьшая интенсивность (I_{low}, дБ (А)), джиттер (абсолютный) (jitter, %), который рассчитывается как среднее значение разницы ЧОТ между двумя соседними периодами акустической волны, выраженное в процентах [10, 11]. Эмпирическим путем была рассчитана формула для вычисления *DSI*:

$$DSI = 0.13MPT + 0.0053F_{0\ high} - 0.26I_{low} - 1.18Jitter + 12.4\,. \qquad (1)$$

Однако предлагаемые границы голосовой дисфункции (*DSI* от –5 до +5), по мнению ряда исследователей и фониаторов, не имеют диагностического значения [12, 13]. Возможно, причина заключается в том, что методика расчета DSI рассматривает изменение только одной гармоники – частоты основного тона. Как было отмечено, функциональные нарушения характеризуются изменениями в голосе, отражающимися в виде обертонов, интенсивность которых увеличивается либо уменьшается в зависимости от типа нарушения. В связи с этим цель настоящей работы – разработать аппаратно-программный комплекс для анализа совокупности гармоник голосового сигнала.

1. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Предлагаемый аппаратно-программный комплекс анализа голоса и речи состоит из нескольких модулей VASA (Voice and Speech Analysing system). Блок-схема разрабатываемого устройства приведена на рис. 1.

Рис. 1. Блок-схема программно-аппаратного комплекса анализа голоса

Fig. 1. Block diagram of the software and hardware complex for voice analysis

Комплекс состоит из нескольких модулей. Модуль акустического анализа голоса представляет собой систему, показанную на рис. 2. Микрофон помещается на расстоянии 15 см от ротовой полости обследуемого. Такое расстояние остается неизменным на протяжении всей диагностической процедуры. Тестирование заключается в произнесении определенных звуков и фраз. Сигнал с микрофона поступает на специализированную аудиокарту и подается на вход концентратора.

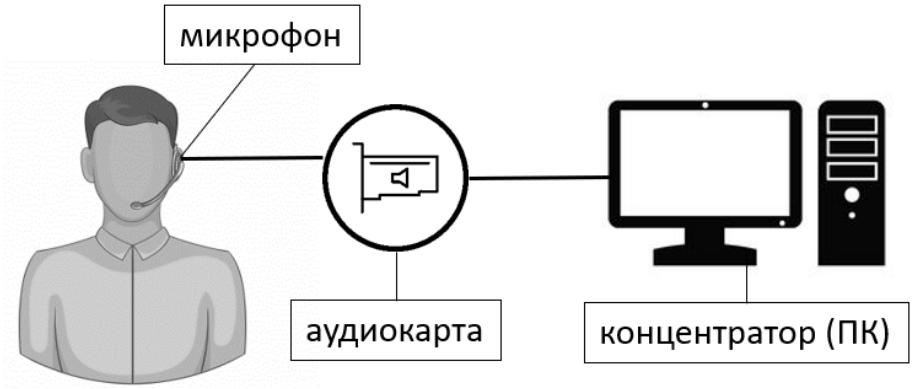


Рис. 2. Модуль акустического анализа голоса

Fig. 2. An acoustic voice analysis module

В этой версии аппаратно-программного комплекса использовались микрофон Shure WH30XLR и профессиональная аудиокарта фирмы M-Audio Fast Track Ultra, которая позволяет получить высококачественную оцифровку

звукового сигнала (разрядность ЦАП/АЦП: 24 бит/24 бит; максимальная частота ЦАП/АЦП: 96 кГц/96 кГц). Выбранная модель микрофона обеспечивает высокий уровень изоляции от посторонних источников звука, имеет плавную частотную характеристику, изолирующий демпфер для снижения механических шумов (кардиоидный; диапазон частот 40…20 000 Гц; выходное сопротивление 150 Ом). В качестве концентратора в предлагаемой системе используется персональный компьютер, однако в последующих версиях может быть использован мобильный телефон или планшет.

Специально разработанное программное обеспечение, установленное на концентраторе, путем цифровой обработки сигнала рассчитывает ряд акустических параметров голосового сигнала, коррелирующих с различными типами нарушений голоса. С помощью разработанного ПО рассчитываются и анализируются следующие диагностические параметры.

• Джиттер (*Jitter*) – среднее значение разницы ЧОТ между двумя соседними периодами [11]:

$$Jitter(absolute) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i+1}\right|, \qquad (2)$$

где T – длина периода; N – число таких периодов. Значение джиттера в голосе без нарушений не должно превышать 0,5 %.

• Шиммер – количественная оценка изменения амплитуд двух соседних периодов (дБ) [11]:

$$Shimmer(dB) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log\left(\frac{A_{i+1}}{A_i}\right)\right|, \qquad (3)$$

где A – амплитуда в выбранном периоде; N – число таких периодов.

• Отношение гармоника–шум – доля гармонической составляющей сигнала (фонационной) по отношению к шумовой составляющей:

$$HNR = 10\log 10\left(\frac{I_1}{I_2}\right), \qquad (4)$$

где I_1 – интенсивность гармонической составляющей; I_2 – интенсивность шумовой составляющей [14].

• Коэффициент степени выраженности дисфонии – комплексный параметр оценки качества голоса [8]:

$$DSI = 0.13MPT + 0.0053F_{0\ high} - 0.26I_{low} - 1.18Jitter + 12.4, \qquad (5)$$

где MPT – максимальное время фонации (с); F_{high} – наивысшая частота (Гц); I_{low} – наименьшая интенсивность (дБ(А)); $jitter$ – джиттер абсолютный (%).

• Коэффициент звонкости голоса – отношение интенсивности сигнала в области высокой певческой форманты (ВПФ) к общей интенсивности сигнала [15]:

$$K_{зв} = \frac{I_1}{I_2}100\ \%, \qquad (6)$$

где I_1 – интенсивность сигнала в области ВПФ; I_2 – общая интенсивность гласного.

Графическое отображение звукового профиля, а также значения всех рассчитываемых акустических параметров выводятся на монитор концентратора в удобном для пользователя виде. Пример окна акустического анализа голоса показан на рис. 3.

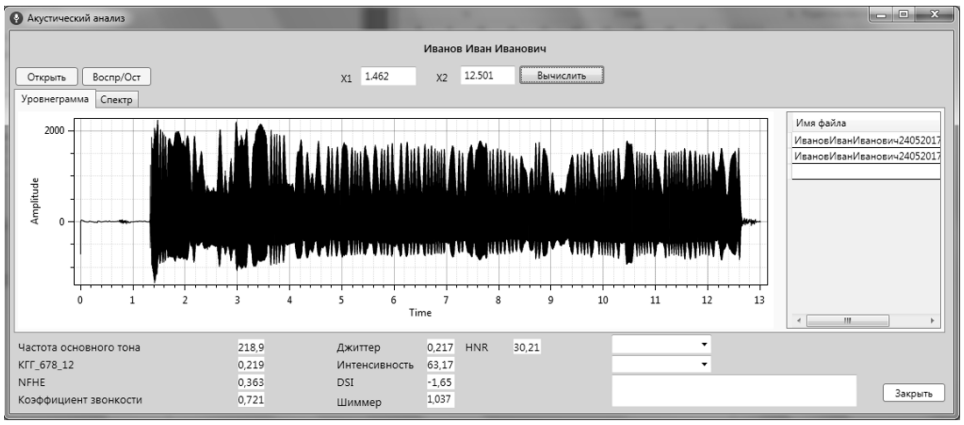


Рис. 3. Пример окна акустического анализа голоса

Fig. 3. Example of an acoustic voice analysis window

Разработанный аппаратно-программный комплекс предоставляет пользователю возможность выбора определенного фрагмента для анализа, а также автоматически производит расчет значений амплитуд первых 16 гармоник, что позволяет исследовать различные их комбинации в зависимости от типа анализируемого нарушения фонации так же, как это делает эксперт на слух, оценивая степень выраженности обертонов.

Данные с модулей агрегируются на концентраторе (персональном компьютере). Специально разработанное нами программное обеспечение обрабатывает полученные данные и выводит их в удобном для пользователя виде.

Отличительной особенностью комплекса являются разработанные алгоритмы, которые на основе первичных данных позволяют получить информацию о наличии в голосе того или иного типа нарушения и степени его выраженности при помощи комплексного акустического анализа сигнала.

В результате комплексного акустического анализа специалист получает информацию о наличии у говорящего того или иного типа нарушения фонации и степени ее выраженности с учетом индивидуальных особенностей пациента, а также результатов аудитивной и психологической оценки.

2. РЕЗУЛЬТАТЫ ОБСУЖДЕНИЙ

В рамках проверки комплекса на *n* (кол-во) добровольцах была проведена оценка воспроизводимости и повторяемости результатов при помощи R&R анализа (Repeatability & Reproducibility). На начальном этапе произво-

дилась проверка работы модуля акустического анализа голосового сигнала. Было проведено трехкратное измерение стандартных параметров акустического анализа *Jitter*, *Shimmer*, *HNR*, $K_{зв}$ и у всех обследуемых лиц двумя независимыми операторами. Исследование проводилось в шумоизолированном помещении. На момент обследования добровольцы не жаловались на наличие каких-либо заболеваний, влияющих на качество фонации, и не страдали нарушениями голосового аппарата. Расчет параметров производился для длительного гласного звука |а|.

Анализ воспроизводимости и повторяемости проводился методом R&R при помощи STATISTICA v10. Графики анализа повторяемости и воспроизводимости для пяти акустических параметров представлены на рис. 4.

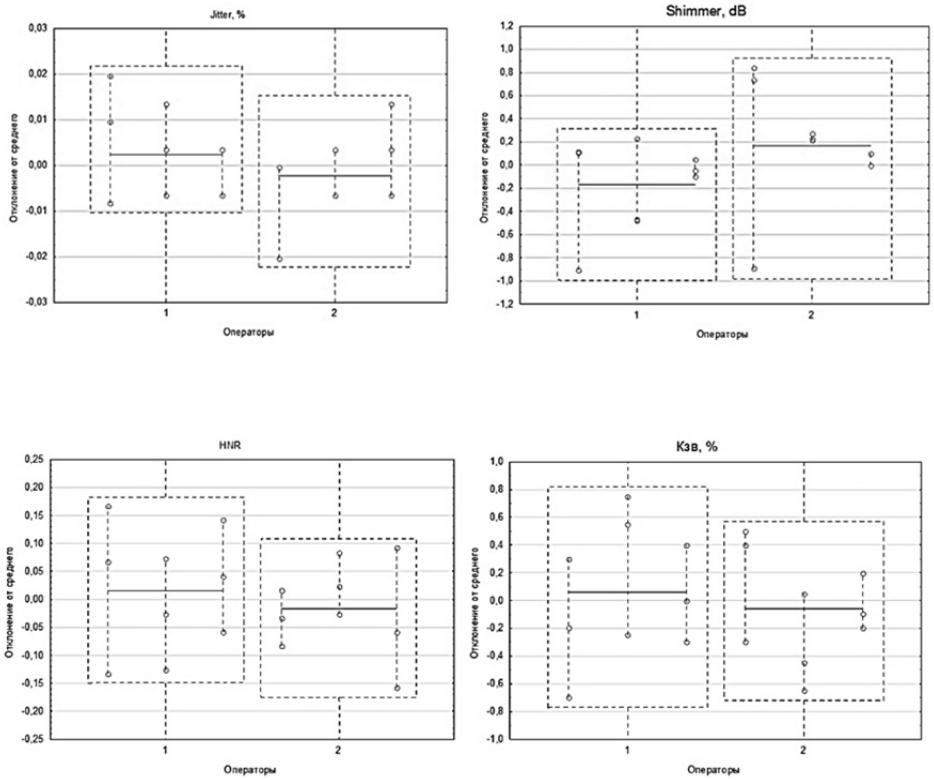


Рис. 4. Графики анализа воспроизводимости и повторяемости результатов

Fig. 4. Graphics for analyzing the reproducibility and repeatability of results

Результаты оценки дисперсии методом размахов представлены в таблице. Согласно полученным данным разрабатываемая система обладает высокими показателями воспроизводимости и повторяемости результатов. Однако следует учитывать, что различия в полученных данных могут обосновываться индивидуальными характеристиками голоса исследуемых.

Оценка дисперсии методом размахов.
Оценка повторяемости и воспроизводимости

Estimation of variance by the sweep method.
Assessment of repeatability & reproducibility

Параметр	Оценен. сигма	Оценен. дисперсии	% от R & R	% от общей
Повторяемость				
Jitter	0,01	0,0001	100,00	25,85
Shimmer	0,36	0,13	76,34	40,93
HNR	0,02	0,000237	2,29	1,034
$K_{зв}$, %	0,44	0,20	100,009	2,96
КГГ	0,13	0,016	52,78	23,77
Воспроизводимость				
Jitter	0,00	0,00	0,00	0,00
Shimmer	0,20	0,04	23,66	12,69
HNR	0,10	0,01	97,72	44,21
$K_{зв}$, %	0,00	0,00	0,00	0,00
КГГ	0,12	0,01	47,22	21,26

ЗАКЛЮЧЕНИЕ

В настоящей работе на основе акустических характеристик по совокупности гармоник голосового сигнала разработан новый аппаратно-программный комплекс VASA для объективной диагностики функциональных нарушений голосовой функции и контроля эффективности терапевтической коррекции. Комплекс позволяет проводить автоматический расчет значений амплитуд первых 16 гармоник, что обеспечивает поиск объективных критериев для выявления различных типов нарушения фонации. Выполненные испытания на трех добровольцах показали высокий уровень воспроизводимости и повторяемости (в пределах 10 % < %R&R < 30 % для трех показателей (см. таблицу)), достаточный для проведения сравнительных исследований в группе здоровых и лиц с функциональными нарушениями речевого аппарата.

СПИСОК ЛИТЕРАТУРЫ

1. *Василенко Ю.С.* Голос. Фониатрические аспекты. – М.: Дипак, 2013. – 396 с.
2. Музыкальная энциклопедия. В 6 т. Т. 3 / гл. ред. Ю.В. Келдыш. – М.: Советская энциклопедия, 1976. – 1104 с.

3. *Fujiki R.B., Thibeault S.L.* Examining relationships between GRBAS ratings and acoustic, aerodynamic and patient-reported voice measures in adults with voice disorders // Journal of Voice. – 2021. – DOI: 10.1016/j.jvoice.2021.02.007.

4. Acoustic characteristics of the voice for Brazilian Portuguese speakers across the life span / E.A. Spazzapan, E.M. Gradim Fabron, L.C. Berti, E.F.B. Chagas, V.C. de Castro Marino // Journal of Voice. – 2020. – DOI: 10.1016/j.jvoice.2020.09.019.

5. Changes in acoustic aspects of vocal function in children after adenotonsillectomy / F.F. Brkic, D.T. Liu, N.J. Campion, M. Leonhard, S. Altumbabic, M. Korlatovic, A. Kaider, J. Kabil-Hamidovic, F. Brkic, E. Vyskocil // Journal of Voice. – 2020. – DOI: 10.1016/j.jvoice.2020.06.022.

6. Acoustic analysis of normal voice patterns in Italian adults by using praat / C. Gorris, A. Ricci Maccarini, F. Vanoni, M. Poggioli, R. Vaschetto, M. Garzaro, P. Aluffi Valletti // Journal of Voice. – 2020. – Vol. 34 (6). – P. 961.e9–961.e18. – DOI: 10.1016/j.jvoice.2019.04.016.

7. *Soumya M., Narasimhan S.V.* Correlation between subjective and objective parameters of voice in elderly male speakers / M. Soumya, S.V. Narasimhan // Journal of Voice. – 2020. – DOI: 10.1016/j.jvoice.2020.10.006.

8. Relation between dysphonia severity index (DSI) and consensus auditory-perceptual evaluation of voice (CAPE-V) / E. Ataee, H. Khoramshahi, E. Naderifar, E.M. Dastoorpour // Journal of Voice. – 2020. – DOI: 10.1016/j.jvoice.2020.06.030.

9. *Sobol M., Sielska-Badurek E.M.* The dysphonia severity index (DSI)-normative values. Systematic review and meta-analysis // Journal of Voice. – 2020. – Vol. 63. – P. 951–956.

10. *Brockmann-Bauser M., Bohlender J.E., Mehta D.D.* Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders // Journal of Voice. – 2018. – Vol. 32. – P. 162–168.

11. *Cataldo E., Soize C.* Voice signals produced with jitter through a stochastic one-mass mechanical model // Journal of Voice. – 2017. – Vol. 31. – P. 111.e9–111.e18.

12. The Influence of gender and age on the acoustic voice quality index and dysphonia severity index: a normative study / B. Barsties V. Latoszek, N. Ulozaitė-Stanienė, Y. Maryn, T. Petrauskas, V. Uloza // Journal of Voice. – 2019. – Vol. 33 (3). – P. 340–345.

13. Validation of the dysphonia severity index in the Dr. speech program / H. Kim, S. Gao, B. Yi, R. Shi, Q. Wan, Z. Huang // Journal of Voice. – 2019. – Vol. 33 (6). – P. 948.e23–948.e29. – DOI: 10.1016/j.jvoice.2019.08.011.

14. *Yanagihara N.* Significance of harmonic changes and noise components in hoarseness // Journal of Speech and Hearing Research. – 1967. – Vol. 10. – P. 531–541.

15. *Морозов В.П.* Биофизические основы вокальной речи. – Л.: Наука, 1977. – 232 с.

Боровикова Дарья Владимировна, ассистент кафедры электронных приборов Новосибирского государственного технического университета. Основное направление научных исследований – определение методики объективизации параметров голосоречевого сигнала. Является автором более 15 научных статей. E-mail: borovikova@corp.nstu.ru

Гришин Олег Витальевич, доктор медицинских наук, главный научный сотрудник НИИ физиологии и фундаментальной медицины, доцент кафедры внутренних болезней медицинского факультета Новосибирского государственного университета, профессор кафедры электронных приборов Новосибирского государственного технического университета, руководитель департамента перспективных исследований ЗАО СЦФБ. Автор более 120 научных статей. Основное направление научных исследований – медицинская электроника. E-mail: 8asteroid@gmail.com

Ненько Анастасия Викторовна, студентка 4-го курса кафедры электронных приборов Новосибирского государственного технического университета. Основное направление научных исследований – детектирование психоэмоциональных нарушений по акустическому анализу голосоречевого сигнала. E-mail: nenkoanastasia507@gmail.com

Юпашевский Антон Витальевич, аспирант Новосибирского государственного технического университета по профилю «Фотоника, приборостроение, оптические и биотехнические системы и технологии». Основное направление научных исследований – разработка оборудования для биомониторинговых систем и оборудования для образовательных целей. E-mail: up1997ant@gmail.com

Казьмина Анна Сергеевна, аспирант Новосибирского государственного технического университета по профилю «Фотоника, приборостроение, оптические и биотехни-

ческие системы и технологии». Автор 10 научных статей. Основное направление научных исследований – разработка оборудования для биомониторинговых систем, разработка систем на программируемых логических интегральных схемах. E-mail: kazmina.anny@gmail.com

Марков Артём Владимирович, старший преподаватель кафедры электронных приборов Новосибирского государственного технического университета. Основное направление научных исследований – разработка оборудования для биомониторинговых систем и оборудования для образовательных целей. Является автором более 20 научных статей. E-mail: avm@332lab.ru

Мецлер Константин Александрович, аспирант Новосибирского государственного технического университета по профилю «Фотоника, приборостроение, оптические и биотехнические системы и технологии». E-mail: kostya.metsler@gmail.com

Borovikova Daria V., a graduate from the Department of Electronic Devices in the Novosibirsk State Technical University, assistant lecturer at the Department of Electronic Devices in the Novosibirsk State Technical University. The main field of her scientific research is determining objectification technique parameters of the voice signal. She is the author of more than 15 scientific articles. E-mail: borovikova@corp.nstu.ru

Grishin Oleg V., Doctor of Medical Sciences, Chief Researcher of the Research Institute of Physiology and Fundamental Medicine, associate professor at the Department of Internal Diseases of the Medical Faculty in Novosibirsk State University; professor at the Department of Electronic Devices in the Novosibirsk State Technical University, Head of the Department of Advanced Research of JSC SCFB. He is the author of more than 120 research papers. The main area of his scientific research is medical electronics. E-mail: 8asteroid@gmail.com

Nenko Anastasia V., a 4th year student at the Department of Electronic Devices in the Novosibirsk State Technical University. The main field of her scientific research is the detection of psychoemotional state by an acoustic analysis of the voice-speech signal. E-mail: nenkoanastasia507@gmail.com

Yupashevsky Anton V., a graduate from the Department of Electronic Devices in the Novosibirsk State Technical University, a post-graduate student at the Novosibirsk State Technical University in the profile "Photonics, instrumentation, optical and biotechnical systems and technologies". The main field of his scientific research is the development of equipment for biomonitoring systems and equipment for educational purposes. E-mail: up1997ant@gmail.com

Kazmina Anna S., a graduate from the Department of Electronic Devices in the Novosibirsk State Technical University, a post-graduate student at the Novosibirsk State Technical University in the profile "Photonics, instrumentation, optical and biotechnical systems and technologies". She is the author of 10 scientific articles. The main field of her scientific research is the development of equipment for biomonitoring systems and the development of systems based on programmable logic integrated circuits. E-mail: kazmina.anny@gmail.com

Markov Artem V., a graduate from the Department of Electronic Devices in the Novosibirsk State Technical University, senior lecturer at the Department of Electronic Devices in the Novosibirsk State Technical University. The main field of his scientific research is the development of equipment for biomonitoring systems and equipment for educational purposes. He is the author of more than 20 research paper. E-mail: avm@332lab.ru

Metsler Konstantin A., a graduate from the Department of Electronic Devices in the Novosibirsk State Technical University, a post-graduate student at the Novosibirsk State Technical University in the field of photonics, instrument making, optical and biotechnical systems and technologies. E-mail: kostya.metsler@gmail.com

DOI: 10.17212/2782-2001-2021-2-135-145

Development of a hardware and software complex for speech analysis and correction[*]

D.V. BOROVIKOVA[a], O.V. GRISHIN[b], A.V. NENKO[c], A.V. YUPASHEVSKY[d], A.S. KAZMINA[e], A.V. MARKOV[f], K.A. METSLER[g]

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

[a] *borovikova@corp.nstu.ru* [b] *8asteroid@gmail.com*
[c] *nenkoanastasia507@gmail.com* [d] *up1997ant@gmail.com*
[e] *kazmina.anny@gmail.com* [f] *avm@332lab.ru* [g] *kostya.metsler@gmail.com*

Abstract

In recent years, there has been a dramatic increase in the number of people suffering from functional disorders of voice, usually caused by a psychoemotional stress. Such disorders bring significant discomfort to a person's life as they reduce their communication and social adaptation capacitty, which in turn increases the psychoemotional load. As a result, functional disorders are fixed by the vicious circle mechanism o and can be transformed into the pathology of the speech apparatus. The main method of diagnosis remains expert assessment, which directly depends on the professional skills of a specialist in working with voice. In this connection, the issue of developing such systems for diagnosing voice-speech disorders that would allow for an objective assessment based on the processing of voice-speech characteristics, as well as to identify the violation in time and prevent the development of pathology, is relevant. Such methods and systems can be useful both for diagnostics and for monitoring the effectiveness of voice therapy. The existing methods of hardware diagnostics have not yet found their application in practice due to their inconsistency with the results of expert evaluation. In this paper, we propose a new concept of hardware and software complex for the analysis of voice based on acoustic characteristics of a set of harmonics of the voice signal. A VASA (Voice and Speech Analyzing system) complex has been developed that provides an automatic analysis of the amplitudes of the first 16 harmonics. The tests performed on three volunteers showed a high level of reproducibility and repeatability (within 10 % < %R&R < 30 %), sufficient for conducting comparative studies on healthy people and people with functional speech disorders.

Keywords: phoniatrics, dysphonia, jitter, shimmer, digital signal processing, acoustic analysis, objective analysis, R&R analysis, automated complex

REFERENCES

1. Vasilenko Yu.S. *Golos. Foniatricheskie aspekty* [The voice. Phoniatric aspects]. Moscow, Dipak Publ., 2013. 396 p.

2. Keldysh Yu.V., ed. *Muzykal'naya entsiklopediya*. V 6 t. T. 3 [The musical encyclopedia. In 6 vol. Vol. 3]. Moscow, Sovetskaya entsiklopediya Publ., 1976. 1104 p.

3. Fujiki R.B., Thibeault S.L. Examining Relationships Between GRBAS Ratings and Acoustic, Aerodynamic and Patient-Reported Voice Measures in Adults With Voice Disorders. *Journal of Voice*, 2021. DOI: 10.1016/j.jvoice.2021.02.007.

4. Spazzapan E.A., Gradim Fabron E.M., Berti L.C., Chagas E.F.B., Castro Marino V.C. de. Acoustic characteristics of the voice for Brazilian Portuguese speakers across the life span. *Journal of Voice*, 2020. DOI: 10.1016/j.jvoice.2020.09.019.

5. Brkic F.F., Liu D.T., Campion N.J., Leonhard M., Altumbabic S., Korlatovic M., Kaider A., Kabil-Hamidovic J., Brkic F., Vyskocil E. Changes in acoustic aspects of vocal function in children after adenotonsillectomy. *Journal of Voice*, 2020. DOI: 10.1016/j.jvoice.2020.06.022.

6. Gorris C., Ricci Maccarini A., Vanoni F., Poggioli M., Vaschetto R., Garzaro M., Aluffi Valletti P. Acoustic analysis of normal voice patterns in Italian adults by using praat. *Journal of Voice*, 2020, vol. 34 (6), pp. 961.e9–961.e18. DOI: 10.1016/j.jvoice.2019.04.016.

[*] *Received 14 January 2021.*

7. Soumya M., Narasimhan S.V. Correlation between subjective and objective parameters of voice in elderly male speakers. *Journal of Voice*, 2020. DOI: 10.1016/j.jvoice.2020.10.006.

8. Ataee E., Khoramshahi H., Naderifar E., Dastoorpour E.M. Relation between dysphonia severity index (DSI) and consensus auditory-perceptual evaluation of voice (CAPE-V). *Journal of Voice*, 2020. DOI: 10.1016/j.jvoice.2020.06.030.

9. Sobol M., Sielska-Badurek E.M. The dysphonia severity index (DSI)-normative values. Systematic review and meta-analysis. *Journal of Voice*, 2020, vol. 63, pp. 951–956.

10. Brockmann-Bauser M., Bohlender J.E., Mehta D.D. Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders. *Journal of Voice*, 2018, vol. 32, pp. 162–168.

11. Cataldo E., Soize C. Voice signals produced with jitter through a stochastic one-mass mechanical model. *Journal of Voice*, 2017, vol. 31, pp. 111.e9–111.e18.

12. Barsties V. Latoszek B., Ulozaitė-Stanienė N., Maryn Y., Petrauskas T., Uloza V. The influence of gender and age on the acoustic voice quality index and dysphonia severity index: a normative study. *Journal of Voice*, 2019, vol. 33 (3), pp. 340–345.

13. Kim H., Gao S., Yi B., Shi R., Wan Q., Huang Z. Validation of the dysphonia severity index in the Dr. speech program. *Journal of Voice*, 2019, vol. 33 (6), pp. 948.e23–948.e29. DOI: 10.1016/j.jvoice.2019.08.011.

14. *Yanagihara N.* Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research*, 1967, vol. 10, pp. 531–541.

15. Morozov V.P. *Biofizicheskie osnovy vokal'noi rechi* [Biophysical foundations of vocal speech]. Leningrad, Nauka Publ., 1977. 232 p.

Для цитирования:

Разработка аппаратно-программного комплекса анализа и коррекции речи / Д.В. Боровикова, О.В. Гришин, А.В. Ненько, А.В. Юпашевский, А.С. Казьмина, А.В. Марков, К.А. Мецлер // Системы анализа и обработки данных. – 2021. – № 2 (82). – С. 135–145. – DOI: 10.17212/2782-2001-2021-2-135-145.

For citation:

Borovikova D.V., Grishin O.V., Nenko A.V., Yupashevsky A., Kazmina A.S., Markov A.V., Metsler K.A. Razrabotka apparatno-programmnogo kompleksa analiza i korrektsii rechi [Development of a hardware and software complex for speech analysis and correction]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 135–145. DOI: 10.17212/2782-2001-2021-2-135-145.

Правила для авторов

Журнал «Научный вестник НГТУ» основан в 1995 году на базе Новосибирского государственного технического университета. Печатная версия журнала зарегистрирована в Министерстве РФ по делам печати, телевещания и средств массовых коммуникаций в 2000 году. Свидетельство о регистрации ПИ № 77-1599 от 10.02.2000 года. Периодичность выхода издания – один раз в три месяца (четыре номера в год). В 2021 году был переименован в «Системы анализа и обработки данных». Научно-технические статьи, направленные в адрес журнала, проходят рецензирование и редактирование.

В журнале «Системы анализа и обработки данных» публикуются оригинальные статьи по следующим группам специальностей:

05.11.00 – Приборостроение, метрология и информационно-измерительные приборы и системы;

05.11.07 – Оптические и оптико-электронные приборы и комплексы;

05.11.16 – Информационно-измерительные и управляющие системы (по отраслям);

05.11.17 – Приборы, системы и изделия медицинского назначения;

05.13.00 – Информатика, вычислительная техника и управление;

05.13.01 – Системный анализ, управление и обработка информации (по отраслям);

05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей;

05.13.17 – Теоретические основы информатики;

05.13.18 – Математическое моделирование, численные методы и комплексы программ.

Разделы журнала: приборостроение, метрология и информационно-измерительные приборы и системы; информатика, вычислительная техника и управление.

Рекомендуемый объем статьи 10–16 страниц. Статьи объемом, не превышающим 8 страниц, могут быть размещены в разделе «Краткие сообщения». Возможна публикация дискуссионного материала в разделе «Дискуссии». В информационном разделе возможно размещение сообщений о состоявшихся и планируемых конференциях, семинарах и симпозиумах, сведений о факультетах и специальностях, научных разработках и научных школах. Статьи обзорного характера, как правило, не публикуются.

Website: http://journals.nstu.ru/vestnik

Условия приема статей

В редакцию журнала представляются следующие материалы.

1. Статья, подготовленная в соответствии с правилами оформления, – печатная версия, два экземпляра, подписанных авторами.
2. Контактная информация (телефоны рабочий и сотовый, адреса электронной почты, место работы, адрес места работы, должность, ученая степень, ученое звание автора, ORCID) – печатная версия, два экземпляра.
3. Описание статьи для базы данных «Российский индекс научного цитирования (РИНЦ)», подготовленное в соответствии с правилами оформления, – печатная версия, один экземпляр.
4. Лицензионный договор, заполненный и подписанный, отправляется со всем пакетом документов; сканированный заполненный договор записывается на диск.
5. Электронная версия статьи, контактной информации, описания статьи для базы данных РИНЦ и сканированный лицензионный договор – в отдельных файлах на CD.
6. Экспертное заключение о возможности опубликования.

Плата за публикацию рукописей не взимается.

Требования к оформлению статьи

Формат оригинала **– А4**. Шрифт **– Times New Roman**. Текст набирается в русифицированном редакторе **Microsoft Word**. Параметры страницы (см):

левое поле – 3,3 верхнее поле – 4,0

правое поле – 3,3 нижнее поле – 3,9.

текстовое поле 22 × 14,4

Межстрочные интервалы одинарные. Перед УДК оставляется четыре пустые строки размером 12. УДК набирается прописными буквами кеглем 8. После УДК перед заголовком оставляется пустая строка размером 8. Название статьи набирается кеглем 14 полужирным строчным шрифтом с отступом слева 1,8 см, без переносов слов, с выравниванием влево. После заголовка оставляется пустая строка размером 12. Фамилии автора – прописным шрифтом кеглем 8. После фамилий оставляется пустая строка размером 7. Указатель сноски * после названия статьи устанавливается в верхнем индексе. В сноске дата поступления статьи указывается курсивным шрифтом кеглем 8. После ФИО авторов статьи набирается аннотация кеглем 8 с отступом первой строки на 0,8 см, межстрочный интервал с множителем 1,1. После аннотации оставляется пустая строка размером 4. **Ключевые слова** – кеглем 8.

Заголовки разделов обязательны (например, **Введение**, **Постановка задачи**, **Методы анализа**, **Основной результат**, **Выводы** или **Заключение**), набираются прописным полужирным прямым шрифтом кеглем 8 с отступом 1,8 см. Перед ними оставляется пустая строка размером 20, после них – пустая строка размером 10. Основной текст – кеглем 10.

Подзаголовки – прописным полужирным шрифтом кеглем 8 с отступом 1,8 см. Перед ними и после них – пустая строка размером 10.

Рисунки представляются в **Corel Draw 11** или **Microsoft Word (RUS)**. Рисунки в других редакторах не принимаются. Слово *Рис*. и его номер выделяются курсивом. Рисунки должны быть черно-белыми, в описании не должно быть ссылок на цвет.

Рисунки и таблицы нумеруются арабскими цифрами по порядку упоминания их в тексте. После номера должно следовать название **на русском и английском языках**. Подрисуночные подписи и названия таблиц – кеглем 9.

Все графы в таблицах должны иметь заголовки и разделяться вертикальными линиями. В головке таблиц по возможности указываются буквенные обозначения параметров и их единицы измерения.

Переменные в тексте набираются курсивным шрифтом кеглем 10. При написании формул использовать MathType версии не выше 6.0C. Формулы набирать по центру кеглем 11, надстрочные и подстрочные индексы – кеглем 9, знаки – кеглем 14 (латинские буквы – курсивным шрифтом, русские и греческие – прямым, цифры – прямым).

Например:

$$A(s) = \sum_{i=0}^{n^2+3} A_i s \lambda^{(i+1)^2} + \alpha .$$

Между текстом и формулами, а также между формулами оставляются пустые строки размером 4. Формулы не сжимать. Нумерацию формул указывать справа кеглем 10. Например:

$$D_f = D_c D + N_c N. \qquad (1)$$

Нумерация страниц осуществляется следующим образом: Вставка / Номера страниц...; установить **положение** «вверху страницы», **выравнивание** – «снаружи», без нумерации первой страницы.

Колонтитулы устанавливаются вверху (с отступом от верхнего края страницы 3 см). Чтобы колонтитулы на четной, нечетной и первой страницах отличались, нужно установить: Файл / Параметры страницы / Разметка; установить **разные четные и нечетные** и **первая отличается**.

Колонтитулы на четных страницах. Вид / Колонтитулы; набираются курсивным прописным шрифтом кеглем 8. Фамилия автора с инициалами впереди у правого края без отступа. Если авторов несколько, то указываются только два первых, далее «и др.».

Колонтитулы на нечетных страницах. Названия статьи – с прописной буквы курсивным шрифтом кеглем 8. Номера страниц – кеглем 10.

Перед списком литературы пустая строка размером 20. Заголовок «Список литературы» – кеглем 8 прописным полужирным шрифтом с отступом 1,8 см. После списка литературы пустая строка размером 10. Порядковый номер – с красной строки кеглем 8 с интервалом 1,1. Фамилии и инициалы – кеглем 8 полужирным строчным шрифтом, название статьи (книги) – светлым тем же шрифтом. Оформление библиографического списка по ГОСТ Р.7.0.5–2008. Библиографическая ссылка. Общие требования и правила составления. – М.: Стандартинформ, 2008. – 19 с.

После списка литературы идет пустая строка размером 24. Далее приводятся краткие сведения об авторах кеглем 8 с интервалом 1,1 и отступом от левого края 1 см. Фамилия, имя, отчество – с красной строки курсивом. Последующий текст (ученая степень, ученое звание, членство в академиях и т. д.; основное направление научных исследований; количество имеющихся публикаций, в том числе монографий, учебных пособий; e-mail, рабочий телефон) – прямым шрифтом. **Информация предоставляется на русском и английском языках**.

Далее дается перевод на английский язык фамилии, имени, отчества автора и названия статьи, а также реферата и ключевых слов (обязательно).

Исправления, вносимые автором, при доработке статьи **выделять** желтым цветом.

Требования к оформлению РИНЦ

В описании статьи должна быть указана следующая информация.

1. Номера страниц, на которых размещена статья в сборнике.
2. Коды: УДК, и / или ББК, и / или DOI, и / или других классификационных индексов или систем регистрации.
3. Название статьи.
4. Фамилия, имя, отчество авторов.
5. Место работы каждого автора в именительном падеже. Указать должность автора по основному месту работы.
6. Контактная информация (почтовый адрес места работы, номер телефона, адрес электронной почты – обязательно).
7. Аннотация (не менее 320 и не более 640 знаков).
8. Ключевые слова: каждое слово или словосочетание отделяется от другого запятой (10–20 слов).
9. Пункты 3–8 на английском языке.
10. Список литературы.

Образец оформления приведен на сайте http://journal.nstu.ru/vestnik.