

УДК 519.7 + 004.93

ДЕКОМПОЗИЦИЯ СПЕКТРАЛЬНЫХ ПРИЗНАКОВ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ НА ОСНОВЕ СОСТАВЛЯЮЩИХ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

© А. В. Лапко^{1,3}, В. А. Лапко^{1,3}, С. Т. Им^{2,3}¹Институт вычислительного моделирования СО РАН,
660036, г. Красноярск, Академгородок, 50, стр. 44²Институт леса им. В. Н. Сукачева СО РАН,
660036, г. Красноярск, Академгородок, 50, стр. 28³Сибирский государственный университет науки и технологий
им. академика М. Ф. Решетнева,
660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31
E-mail: lapko@ict.krasn.ru

Предложена методика декомпозиции области значений двумерных спектральных признаков по значениям составляющих их коэффициентов корреляции. Основа методики заключается в анализе произведения нормированных значений спектральных признаков. Особенность используемого показателя и вводимые пользователем пороги на его значения позволяют осуществить декомпозицию исходных статистических данных и картирование получаемых результатов. В отличие от традиционных методов предлагаемый подход обладает более высокой вычислительной эффективностью, что необходимо при обработке больших объемов статистических данных. Рассматриваются результаты применения методики при обработке данных дистанционного зондирования природного объекта.

Ключевые слова: декомпозиция статистических данных, автоматическая классификация, коэффициент корреляции, данные дистанционного зондирования, анализ спектральных данных.

DOI: 10.15372/AUT20250303

EDN: FZVCXO

Введение. На первом этапе обработки данных дистанционного зондирования используются алгоритмы автоматической классификации, которые предназначены для обнаружения компактных групп наблюдений в пространстве значений спектральных признаков исследуемых природных объектов [1–4]. Полученные результаты необходимы для дешифрирования исходных снимков с последующим оцениванием состояний природного объекта на основе алгоритмов распознавания образов [5]. При этом опускается объективный фактор — состояние природного объекта, соответствующего элементу земной поверхности, определяется не только значениями спектральных признаков, но и зависимостью между ними. Данный вывод подтверждает необходимость корректировки методики обработки данных дистанционного зондирования. Для частичного решения этой проблемы предложены методы корреляционного анализа, которые основаны на сегментации исходного снимка, вычислении коэффициентов корреляции и их анализе [6–8]. При этом не проводится анализ составляющих коэффициента корреляции, которые являются количественными характеристиками элементов земной поверхности.

В данной работе предлагается методика декомпозиции двумерных спектральных признаков по степени их зависимости в конкретных условиях на основе составляющих коэффициента корреляции. Рассматривается применение полученных результатов при обработке данных дистанционного зондирования природного объекта.

Декомпозиции двумерных спектральных признаков. Имеются данные дистанционного зондирования, представленные выборкой $V = (x_1^i, x_2^i, z_1^i, z_2^i, i = \overline{1, n})$, которая составлена из наблюдений спектральных признаков x_1, x_2 и соответствующих им координат (z_1, z_2) элементов земной поверхности. За основу анализа зависимости между x_1 и x_2 в конкретных условиях x_1^i, x_2^i примем составляющие коэффициента корреляции.

Представим коэффициент корреляции между случайными величинами x_1, x_2 в виде

$$R = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_1^i - \bar{x}_1}{\bar{\sigma}_1} \right) \left(\frac{x_2^i - \bar{x}_2}{\bar{\sigma}_2} \right) = \frac{1}{n-1} \sum_{i=1}^n r_1^i r_2^i. \quad (1)$$

Здесь $\bar{\sigma}_j = \left(\frac{1}{n-1} \sum_{i=1}^n (x_j^i - \bar{x}_j)^2 \right)^{1/2}$ — оценка среднего квадратического отклонения x_j ;

$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_j^i$ — среднее значение признака $x_j, j = 1, 2$.

Коэффициент корреляции (1) формируется на основе его составляющих $r_1^i, r_2^i, i = \overline{1, n}$, значения произведения которых характеризуют зависимость между случайными величинами x_1, x_2 в конкретных условиях x_1^i, x_2^i .

Произведение $r = r_1 r_2$ нормированных значений x_1, x_2 является знакопеременной случайной величиной с плотностью вероятности $p(r)$. Пример закона распределения r представлен на рис. 1, который получен при обработке данных дистанционного зондирования с аппарата MODIS/Terra лесного массива на юге Красноярского края [5]. Спектральные признаки x_1 соответствуют длине волны (545–565 нм), а x_2 — (1230–1250 нм). При оценивании плотности вероятности $p(r)$ использовалась непараметрическая статистика Розенблатта — Парзена при объёме исходных статистических данных $n = 10201$ [9, 10], а выбор её коэффициента размытости осуществлялся на основе методики [11–13].

По результатам исследований ряду участков лесных массивов соответствовали подобные законы распределения параметра r . Их объединяет свойство асимметричности, нахождение моды в окрестности значений $r = 0$. Отличие плотностей вероятностей $p(r)$ наблюдается при больших положительных и отрицательных значениях r , что определяет усиление зависимости между спектральными признаками и является основой обнаружения особенностей исследуемого природного объекта.

С этих позиций определим методику декомпозиции исходных данных дистанционного зондирования при исследовании состояний природного объекта.

1. По исходным данным V сформировать выборку $V' = (x_1^i, x_2^i, i = \overline{1, n})$.

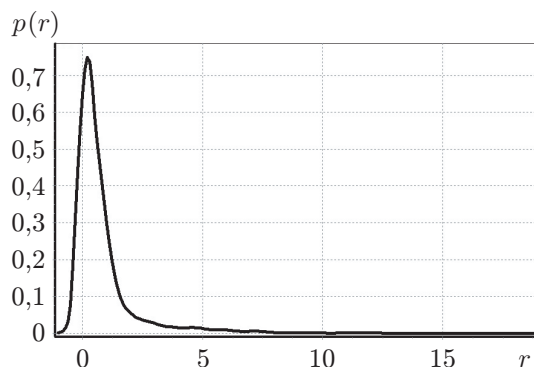


Рис. 1. Непараметрическая оценка плотности вероятности значений $r = r_1 r_2$

2. В соответствии с выражением (1) сформировать массив $V_r = (r^i, i = \overline{1, n})$ и осуществить синтез непараметрической оценки плотности вероятности $p(r)$ [9, 10]:

$$\bar{p}(r) = \frac{1}{nc} \sum_{i=1}^n \Phi\left(\frac{r - r^i}{c}\right), \quad (2)$$

где ядерные функции $\Phi(u)$ удовлетворяют следующим условиям:

$$\begin{aligned} \Phi(u) = \Phi(-u), \quad 0 \leq \Phi(u) < \infty, \quad \int_{-\infty}^{\infty} \Phi(u) du = 1, \\ \int_{-\infty}^{\infty} u^m \Phi(u) du < \infty, \quad 0 \leq m < \infty. \end{aligned}$$

Коэффициент размытости ядерных функций $c \rightarrow 0$ при $n \rightarrow \infty$.

3. Определить оценку оптимального коэффициента размытости ядерных функций статистики (2) из условия минимума по c критерия [13]:

$$\bar{W}(c) = \int_{-\infty}^{\infty} \bar{p}^2(r) dr - \frac{2}{n} \sum_{j=1}^n \bar{p}(r^j), \quad (3)$$

где

$$\bar{p}(r^j) = \frac{1}{nc} \sum_{i=1, i \neq j}^n \Phi\left(\frac{r^j - r^i}{c}\right).$$

Например, при ступенчатой ядерной функции

$$\Phi\left(\frac{r - r^i}{c}\right) = \begin{cases} 0,5, & \text{если } |r - r^i| \leq c; \\ 0, & \text{если } |r - r^i| > c. \end{cases}$$

составляющие первого слагаемого в выражении (3) определяются значениями

$$\int_{-\infty}^{\infty} \Phi\left(\frac{r - r^i}{c}\right) \Phi\left(\frac{r - r^j}{c}\right) dr = \begin{cases} (2c - |r^i - r^j|)/4, & \text{если } |r^i - r^j| < 2c; \\ 0, & \text{если } |r^i - r^j| \geq 2c. \end{cases}$$

Для составляющих второго слагаемого в выражении (3) значения ядерной функции

$$\Phi\left(\frac{r^i - r^j}{c}\right) = \begin{cases} 0,5, & \text{если } |r^i - r^j| \leq c; \\ 0, & \text{если } |r^i - r^j| > c. \end{cases}$$

4. Исследуя особенности плотности вероятности $\bar{p}(r)$, вычислить её моду \tilde{r} и параметр λ , который характеризует симметричный фрагмент $\bar{p}(r)$ относительно \tilde{r} :

$$\int_{\tilde{r}-\lambda}^{\tilde{r}} \bar{p}(r) dr = \int_{\tilde{r}}^{\tilde{r}+\lambda} \bar{p}(r) dr.$$

5. Определить соответствие между интервалами значений параметров $r^i > \lambda$, $r^i < \lambda$ в выборке V_r и состояниями природного объекта с использованием результатов картирования по данным V .

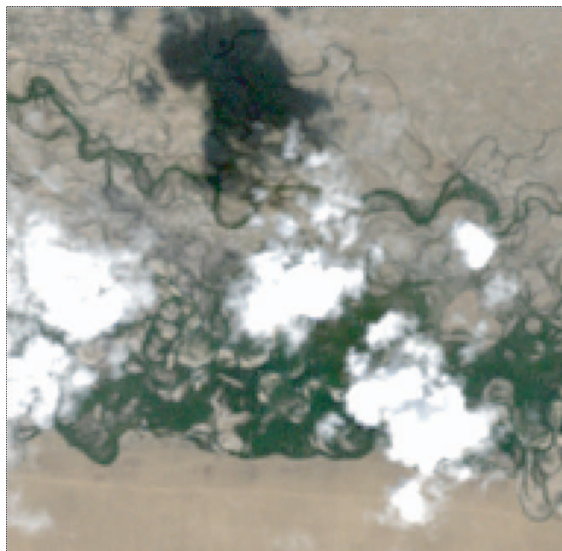


Рис. 2. Фрагмент съёмки с помощью спутника Landsat 7/ETM+ от 21 августа 2015 г. Отображение комбинации каналов x_3 , x_2 , x_1 в формате RGB

Дальнейшее развитие предлагаемой методики состоит во введении специалистом-экспертом новых порогов в интервалах $r > \lambda$, $r < \lambda$ и сопоставлении получаемых результатов с исходными данными V .

Анализ данных дистанционного зондирования. Рассмотрим применение предлагаемой методики при исследовании участка территории, покрытой травяно-кустарничковой растительностью вдоль речной сети в степной зоне восточной части Убсунурской котловины Республики Тыва. Съёмка участка произведена 21 августа 2015 г. спектрометрическим ETM+, установленным на спутнике Landsat 7 [14]. Данные съёмки получены с помощью геопортала Earth Explorer [15] из набора данных второго уровня обработки, включающего атмосферную коррекцию [16]. На снимке присутствуют типичные облака и тени, затрудняющие дешифрирование типов наземного покрова (рис. 2).

Задача выделения облаков и их удаления из снимков до сих пор является актуальной [17, 18]. Современные разрабатываемые алгоритмы детектирования облачности опираются на нейронные сети и вводимые пороговые значения [18–21]. Однако такие алгоритмы требуют использования значительных вычислительных ресурсов на подготовительных этапах расчётов либо при непосредственном детектировании облачности [22]. В случаях оперативной обработки больших массивов данных дистанционного зондирования требуется использование относительно быстрых алгоритмов выделения облаков с удовлетворительной точностью.

При анализе данных дистанционного зондирования каждый элемент земной поверхности задавался шестью спектральными признаками $x = (x_v, v = \overline{1, 5}, x_7)$, характеристика которых представлена в табл. 1.

Выборка $V = (x_1^i, x_2^i, x_3^i, x_4^i, x_5^i, x_7^i, i = \overline{1, 35907})$ формировалась по данным дистанционного зондирования территории (рис. 2).

Ставилась задача обнаружения соответствия между результатами декомпозиции пары спектральных признаков и идентифицируемыми природными объектами. По результатам вычислительного эксперимента установлены два признака x_4 и x_7 , которые соответствуют поставленной задаче. Плотность вероятности произведения $r = r_4 r_7$ составляющих оценки коэффициента корреляции случайных величин x_4 , x_7 , вычисляемых по выборке $V_{4,7} = (x_4^i, x_7^i, i = \overline{1, 35907})$, представлена на рис. 3.

Таблица 1

Характеристика используемых спектральных каналов аппарата Landsat 7

Спектральный канал	Диапазон спектра, нм	Разрешение, м/пиксель	Название
x_1	450–515	30	Голубой
x_2	525–605	30	Зелёный
x_3	630–690	30	Красный
x_4	775–900	30	Ближний инфракрасный
x_5	1550–1750	30	Средний (коротковолновый) инфракрасный
x_7	2090–2350	30	Средний (коротковолновый) инфракрасный

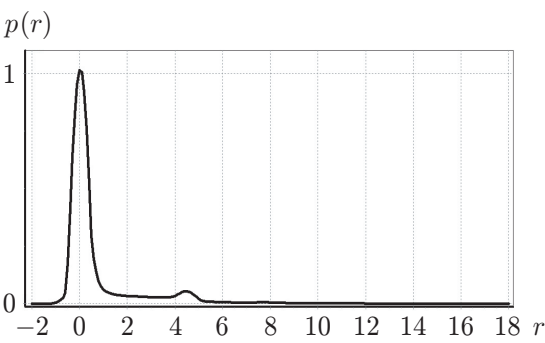


Рис. 3. Непараметрическая оценка плотности вероятности значений $r = r_4r_7$ при оценке их коэффициента корреляции $R = 0,896$

Таблица 2

Интервалы значений параметра $r = r_4r_7$ и соответствующие им области декомпозиции природного объекта

Области декомпозиции	Цвет областей на рис. 3	Интервалы значений параметра r	Оценки вероятностей принадлежности пикселей областям $\Omega_j, j = \overline{1,4}$
Ω_0	Чёрный	$-1,35 \leq r < 0$	0,394
Ω_1	Синий	$0 \leq r < 0,53$	0,364
Ω_2	Голубой	$0,53 \leq r < 1,48$	0,077
Ω_3	Зелёный	$1,48 \leq r < 2,42$	0,033
Ω_4	Жёлтый	$2,42 \leq r < 6,21$	0,105
Ω_5	Оранжевый	$6,71 \leq r < 10$	0,02
Ω_6	Красный	$10 \leq r \leq 17,57$	0,007

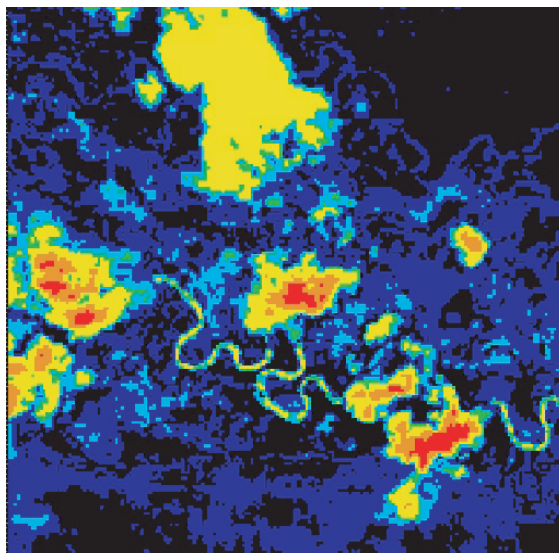


Рис. 4. Пространственное отображение интервалов декомпозиции параметра $r = r_4 r_7$

Для оценивания плотности вероятности $p(r)$ использовались статистика (2) и методика её оптимизации из условия минимума критерия (3). Интервалы значений параметров r и их характеристики приведены в табл. 2.

Пространственное отображение интервалов декомпозиции параметра r приведено на рис. 4 при коэффициенте корреляции между спектральными признаками $R = 0,896$. Экспертный анализ рис. 4 показал, что в результате использования предлагаемой методики выделяются облачность, тени от облаков и полупрозрачная дымка на фоне растительного покрова. Наблюдается частичное перепутывание водного русла реки с тенями, что связано с их низкими значениями спектральных яркостей в анализируемых инфракрасных каналах.

На рис. 5, а приведено изображение, представляющее комбинацию использованных в анализе признаков x_4 и x_7 . Пространственное отображение интервалов декомпозиции параметра в условиях $r \geq 0,7$ приведено на рис. 5, б.

Белый цвет (область декомпозиции Ω_0) соответствует интервалу значений $r \in [-1,35; 0,7]$, которому, в свою очередь, соответствует оценка вероятности попадания пикселей в исследуемый интервал $\bar{P}_0 = 0,797$. Синим цветом обозначена область декомпозиции Ω_1 для условия $r \in [0,7; 1,54]$, при этом значение $\bar{P}_1 = 0,041$. Голубой цвет (Ω_2) на рис. 5 соответствует интервалу параметра $r \in [1,54; 2,39]$ при значении $\bar{P}_2 = 0,03$. Зелёным и жёлтым цветом обозначены области Ω_3 и Ω_4 , которые характеризуются интервалами $[2,39; 3,23]$, $[3,23; 5,76]$ параметра r и соответствующими им оценками вероятностей $\bar{P}_3 = 0,026$, $\bar{P}_4 = 0,076$. Для областей декомпозиции Ω_5 и Ω_6 характерен оранжевый цвет при $r \in [5,76; 9,13]$, $\bar{P}_5 = 0,02$ и красный при $r \in [9,13; 17,57]$, $\bar{P}_6 = 0,01$. Установлено, что тень от облака в основном характеризуется параметром $r \in [4; 5]$ (см. рис. 3).

Подобное пороговое значение $r \geq 0,7$ позволило визуально отделить облачность и тени от фоновой территории (см. рис. 5, б). Облачность (яркие объекты) достаточно хорошо отделилась от фона, однако тени (тёмные объекты) только частично дифференцировались от водной поверхности. Для уточнения выделения теней от облаков можно произвести комплексный анализ других комбинаций спектральных каналов с целью подбора оптимальных пороговых значений компонентов составляющих коэффициента корреляции для идентификации теней от облаков.

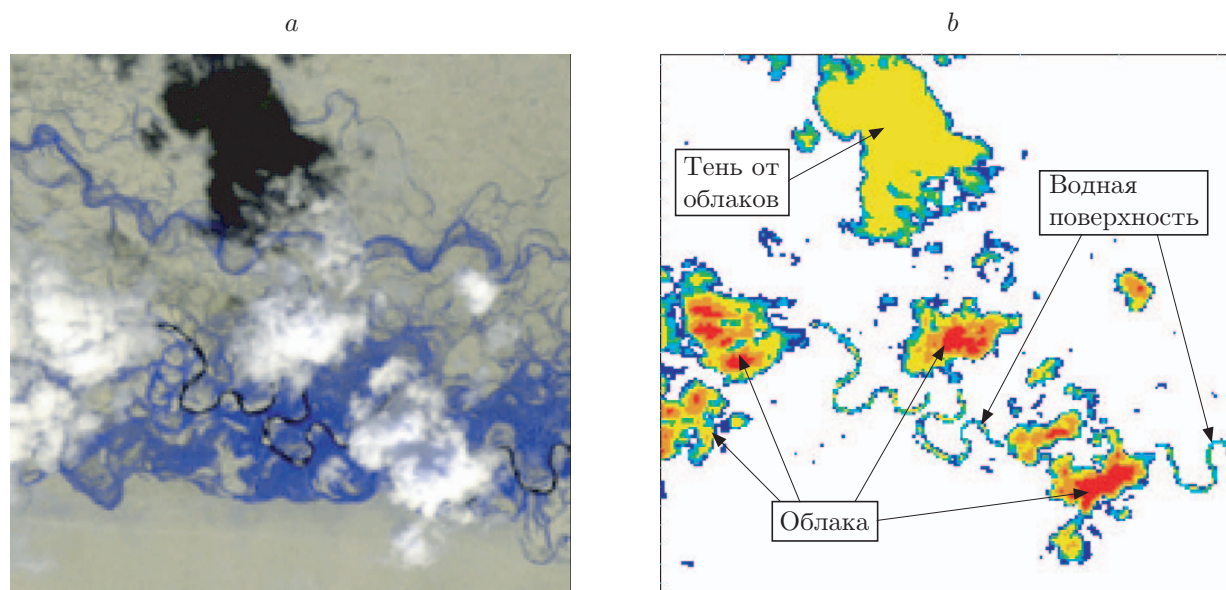


Рис. 5. Фрагмент съёмки с помощью спутника Landsat 7/ETM+ от 21 августа 2015 г. (RGB: x_7, x_7, x_4) (a), пространственное отображение интервалов декомпозиции параметра $r = r_4 r_7 \geq 0,7$ (b)

Заключение. Состояние природных объектов по данным дистанционного зондирования определяется произведением нормированных значений составляющих коэффициента корреляции двумерных спектральных признаков. Этот параметр характеризует зависимость между спектральными признаками в конкретных условиях. При анализе данных дистанционного зондирования значение введённого параметра является знакопеременным и отличается количественно, что позволяет обоснованно осуществлять декомпозицию данных дистанционного зондирования. Полученные выводы подтверждаются применением предлагаемой методики декомпозиции данных дистанционного зондирования при анализе состояний травяно-кустарничковой растительности вдоль речной сети в степной зоне при наличии облачности и теней, затрудняющих дешифрирование типов наземного покрова. В этих условиях применение предлагаемого подхода позволяет эффективно решить проблему выделения облаков по сравнению с традиционными методами, требующими использования значительных вычислительных ресурсов в условиях больших объёмов данных дистанционного зондирования.

Дальнейшее развитие предлагаемого метода состоит в его обобщении на многомерные условия и сопоставлении получаемых результатов при его применении в задаче оценивания состояний природных объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Борзов С. М., Потатуркин О. И. Выбор информативной системы признаков при классификации сельскохозяйственных культур по гиперспектральным данным // Автометрия. 2020. **56**, № 4. С. 134–144. DOI: 10.15372/AUT20200414.
2. Шипко В. В., Борзов С. М. Исследование эффективности классификации гиперспектральных данных при ограничениях на разрядность квантования, количество спектральных каналов и пространственное разрешение // Автометрия. 2022. **58**, № 3. С. 79–87. DOI: 10.15372/AUT20220309.

3. **Лапко А. В., Лапко В. А., Им С. Т. и др.** Непараметрический алгоритм выделения классов, соответствующих одномодальным фрагментам плотности вероятности многомерных случайных величин // *Автометрия*. 2019. **55**, № 3. С. 22–30. DOI: 10.15372/AUT20190303.
4. **Тубольцев В. П., Лапко А. В., Лапко В. А.** Модифицированный непараметрический алгоритм автоматической классификации статистических данных большого объёма и его применение // *Искусственный интеллект и принятие решений*. 2023. Вып. 4. С. 49–57. DOI: 10.14357/20718594230405.
5. **Лапко А. В., Лапко В. А.** Непараметрические алгоритмы оценивания состояний природных объектов // *Автометрия*. 2018. **54**, № 5. С. 33–39. DOI: 10.15372/AUT20180504.
6. **Сазонов Д. С.** Корреляционный анализ экспериментальных дистанционных измерений и моделей микроволнового излучения взволнованной водной поверхности // *Исследование Земли из космоса*. 2017. Вып. 3. С. 53–64. DOI: 10.7868/S020596141703006X.
7. **Lu Z., Mingsheng L., Limin Y., Hui L.** Remote Sensing Change Detection Based on Canonical Correlation Analysis and Contextual Bayes Decision // *Photogrammetric Eng. & Remote Sens.* 2007. **73**, N 3. P. 311–318.
8. **Im J., Jensen J. R., Tullis J. A.** Object-based change detection using correlation image analysis and image segmentation // *Int. Journ. Remote Sens.* 2008. **29**, Iss. 2. P. 399–423. DOI: 10.1080/01431160601075582.
9. **Parzen E.** On Estimation of a Probability Density Function and Mode // *Ann. Math. Statist.* 1962. **33**, N 3. P. 1065–1076.
10. **Епанечников В. А.** Непараметрическая оценка многомерной плотности вероятности // *Теория вероятностей и её применения*. 1969. **14**, вып. 1. С. 156–161.
11. **Rudemo M.** Empirical Choice of Histograms and Kernel Density Estimators // *Scandinavian Journ. Statist.* 1982. **9**, N 2. P. 65–78.
12. **Hall P.** Large Sample Optimality of Least Squares Cross-Validation in Density Estimation // *Ann. Statist.* 1983. **11**, N 4. P. 1156–1174.
13. **Лапко А. В., Лапко В. А.** Анализ методов оптимизации непараметрической оценки плотности вероятности по коэффициенту размытости ядерных функций // *Измерительная техника*. 2017. **66**, № 6. С. 3–8.
14. **Landsat 7.** End of an era: Historic Landsat 7 mission takes final images. URL: <https://www.usgs.gov/landsat-missions/landsat-7> (дата обращения: 06.05.2024).
15. **EarthExplorer.** USGS: Science for a changing world. URL: <https://earthexplorer.usgs.gov> (дата обращения: 06.05.2024).
16. **USGS EROS Archive — Landsat Archives — Landsat 7 ETM Plus Collection 2 Level-2 Science Products.** URL: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-7-etm-plus-collection-2-level-2> (дата обращения: 06.05.2024).
17. **Бочаров Д. А., Николаев Д. П., Павлова М. А., Тимофеев В. А.** Алгоритм детекции и компенсации теней от облаков на мультиспектральных спутниковых снимках для местностей сельскохозяйственных угодий // *Информационные процессы*. 2021. **21**, № 4. С. 295–312. DOI: 10.53921/18195822.2021.21.4.295.
18. **He Q., Sun X., Yan Z., Fu K.** DABNet: Deformable Contextual and Boundary-Weighted Network for Cloud Detection in Remote Sensing Images // *IEEE Trans. Geosci. and Remote Sens.* 2021. **60**. 5601216.
19. **Hu K., Zhang D., Xia M. et. al.** LCDNet: Light-Weighted Cloud Detection Network for High-Resolution Remote Sensing Images // *IEEE Journ. Selected Topics in Appl. Earth Observations and Remote Sens.* 2022. **15**. P. 4809–4823. DOI: 10.1109/jstars.2022.3181303.
20. **Dong J., Wang Y., Yang Y. et. al.** MCDNet: Multilevel cloud detection network for remote sensing images based on dual-perspective change-guided and multi-scale feature

- fusion // Int. Journ. Appl. Earth Observation and Geoinformation. 2024. **129**. 103820. DOI: 10.1016/j.jag.2024.103820.
21. **Shang H., Letu H., Xu R. et. al.** A hybrid cloud detection and cloud phase classification algorithm using classic threshold-based tests and extra randomized tree model // Remote Sens. Environment. 2024. **302**. 113957. DOI: 10.1016/j.rse.2023.113957.
22. **Chai D., Huang J., Wu M. et. al.** Remote sensing image cloud detection using a shallow convolutional neural network // ISPRS Journ. Photogrammetry and Remote Sens. 2024. **209**. P. 66–84. DOI: 10.1016/j.isprsjprs.2024.01.026.

Поступила в редакцию 06.05.2024

После доработки 18.02.2025

Принята к публикации 29.04.2025
