

# ОБРАБОТКА ИЗОБРАЖЕНИЙ, РАСПОЗНАВАНИЕ ОБРАЗОВ

## Методика проверки гипотез о распределениях многомерных спектральных данных с использованием непараметрического алгоритма распознавания образов

А.В. Лапко<sup>1,2</sup>, В.А. Лапко<sup>1,2</sup>

<sup>1</sup> Институт вычислительного моделирования СО РАН, Россия, Красноярск,

<sup>2</sup> Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, Россия, Красноярск

### Аннотация

В работе рассматривается новая методика проверки гипотез о распределениях многомерных спектральных данных дистанционного зондирования. Предлагаемая методика основана на использовании непараметрических алгоритмов распознавания образов. Проверка гипотезы о тождественности двух законов распределений многомерных случайных величин заменяется на проверку гипотезы о равенстве ошибки распознавания образов значению 0,5. Применение данной методики позволяет исключить проблему декомпозиции области значений случайных величин на многомерные интервалы, которая свойственна для критерия Пирсона. Её эффективность подтверждается результатами проверки гипотез о распределении спектральных данных дистанционного зондирования лесных массивов. Проводится анализ законов распределения следующих состояний лесных массивов: темнохвойный лес, повреждённые и сухие древостои. Исходная информация получена по данным дистанционного зондирования территории юга Сибири с помощью аппарата Landsat по 6 спектральным каналам. Результаты исследований являются важными для формирования набора значимых спектральных признаков в задаче оценивания состояний лесных массивов.

**Ключевые слова:** проверка статистических гипотез, многомерные случайные величины, распознавание образов, ядерная оценка плотности вероятности, выбор коэффициента размытости, спектральные данные, дистанционное зондирование, состояния лесных массивов.

**Цитирование:** Лапко, А.В. Методика проверки гипотез о распределениях многомерных спектральных данных с использованием непараметрического алгоритма распознавания образов / А.В. Лапко, В.А. Лапко // Компьютерная оптика. – 2019. – Т. 43, № 2. – С. 238-244. – DOI: 10.18287/2412-6179-2019-43-2-238-244.

### Введение

Методы комплексного анализа данных дистанционного зондирования являются основой создания информационных систем для поддержки принятия решений в задачах оценивания состояний природных объектов [1, 2].

Для проверки гипотез о распределениях многомерных случайных величин используется критерий согласия Пирсона. Однако методика его формирования содержит трудно формализуемый этап разбиения области значений случайных величин на многомерные интервалы [3].

В статьях [4, 5] обосновывается возможность использования непараметрических алгоритмов распознавания образов, соответствующих критерию максимального правдоподобия, в задаче проверки статистических гипотез о распределениях случайных величин. Результаты использования предлагаемой методики сопоставимы с критерием Колмогорова–Смирнова для одномерных задач в условиях, когда количество элементов сравниваемых последовательностей случайных величин отличается незначительно. При различных объёмах случайных последовательностей наблюдается

снижение эффективности предлагаемой методики. Данный факт согласуется с результатами исследований [6], где показано значительное ухудшение аппроксимационных свойств непараметрической оценки уравнения разделяющей поверхности при увеличении степени неравномерности распределения элементов обучающей выборки между классами.

Цель работы – применение предложенной методики при решении задачи проверки гипотез о распределениях многомерных спектральных данных дистанционного зондирования в задачах исследования состояний лесных массивов.

### 1. Методика проверки гипотезы о распределениях многомерных спектральных данных дистанционного зондирования

Пусть  $X_1$  и  $X_2$  – две генеральные совокупности с произвольными законами распределения. Необходимо по независимым выборкам  $V_1 = (x^i, i = \overline{1, n_1})$  и  $V_2 = (x^i, i = \overline{1, n_2})$  многомерных спектральных данных  $x = (x_v, v = \overline{1, k})$ , извлечённым из данных генеральных совокупностей, проверить гипотезу

$$H_0 : F_1(x) \equiv F_2(x)$$

о тождественности их функций распределения  $F_1(x)$ ,  $F_2(x)$ .

Сформируем на основании  $V_1, V_2$  обучающую выборку  $V = (x^i, \delta(i), i = \overline{1, n})$ , где  $n = n_1 + n_2$ , а указания о принадлежности ситуации  $x^i$  к  $V_1$  либо  $V_2$  определяются соотношением

$$\delta(i) = \begin{cases} -1 \forall x^i \in V_1 \\ 1 \forall x^i \in V_2. \end{cases}$$

В этом случае элементы множеств  $V_1, V_2$  характеризуют состояние природных объектов и образуют классы  $\Omega_1, \Omega_2$ . Под классом  $\Omega_j$  будем понимать область определения плотности вероятности  $p_j(x) \neq 0, j = 1, 2$ .

Будем решать двувальтернативную задачу распознавания образов, используя байесовское решающее правило, соответствующее критерию максимального правдоподобия [7–10]

$$m(x) : \begin{cases} x \in \Omega_1, \text{ если } f_{12}(x) \leq 0 \\ x \in \Omega_2, \text{ если } f_{12}(x) > 0, \end{cases}$$

где уравнение разделяющей поверхности между классами  $f_{12}(x) = p_2(x) - p_1(x)$ ;  $p_j(x)$  – условная плотность вероятности распределения  $x$  в классах  $\Omega_j, j = 1, 2$ .

Этому решающему правилу соответствует ошибка распознавания образов

$$\rho = P_1\rho_1 + P_2\rho_2.$$

Здесь  $P_1, P_2$  – априорные вероятности классов  $\Omega_1, \Omega_2$ , а  $\rho_1, \rho_2$  – ошибки их классификации в пространстве признаков  $x$ .

Если  $\Omega_{12} = \Omega_1 \cap \Omega_2 = \emptyset$ , то вероятность ошибки распознавания образов  $\rho = 0$  и плотности вероятностей  $p_1(x), p_2(x)$  в соответствии с введенным определением класса отличаются. При сближении параметров плотностей вероятностей  $p_j(x), j = 1, 2$  область пересечения классов увеличивается, что сопровождается ростом вероятности ошибки распознавания образов. Если отмеченная выше тенденция сохраняется, то значения  $\rho_1$  и  $\rho_2$  стремятся к значению 0,5. Тогда нетрудно заметить справедливость утверждения  $\rho = 0,5$ , так как  $P_1 + P_2 = 1$ .

В этих условиях законы распределения данных в области определения классов, очевидно, совпадают. Поэтому появляется возможность перехода от задачи сравнения законов распределения многомерных спектральных данных к проверке гипотезы  $\bar{H}_0$  о равенстве вероятности ошибки оценивания состояний природных объектов значению 0,5.

Предлагаемая методика основывается на выполнении следующих действий.

По выборке  $V$  осуществить синтез непараметрического алгоритма распознавания образов, соответствующего критерию максимального правдоподобия [7, 8],

$$\bar{m}(x) : \begin{cases} x \in \Omega_1, \text{ если } \bar{f}_{12}(x) \leq 0 \\ x \in \Omega_2, \text{ если } \bar{f}_{12}(x) > 0. \end{cases} \quad (1)$$

При формировании оценки уравнения разделяющей поверхности

$$\bar{f}_{12}(x) = \bar{p}_2(x) - \bar{p}_1(x) \quad (2)$$

будем использовать непараметрические статистики

$$\bar{p}_1(x) = \frac{1}{n_1} \sum_{i \in I_1} \prod_{v=1}^k \frac{1}{c_v} \Phi\left(\frac{x_v - x_v^i}{c_v}\right),$$

$$\bar{p}_2(x) = \frac{1}{n_2} \sum_{i \in I_2} \prod_{v=1}^k \frac{1}{c_v} \Phi\left(\frac{x_v - x_v^i}{c_v}\right),$$

являющиеся оценками плотностей вероятности  $p_1(x), p_2(x)$  многомерной случайной величины  $x$  в классах  $\Omega_1, \Omega_2$  типа Розенблатта–Парзена [11–14]. Здесь  $I_j$  – множество номеров ситуаций из обучающей выборки, принадлежащих классу  $\Omega_j, j = 1, 2$ . Ядерные функции  $\Phi(u_v)$  удовлетворяют условиям:

$$\Phi(u_v) = \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty, \quad \int_{-\infty}^{+\infty} \Phi(u_v) du_v = 1,$$

$$\int_{-\infty}^{+\infty} u^m \Phi(u_v) du_v < \infty, \quad 0 \leq m < \infty, \quad v = \overline{1, k}.$$

Значения их коэффициентов размытости  $c_v$  ядерных функций убывают с ростом  $n_1, n_2$ .

Тогда статистику (2) можно представить выражением

$$\bar{f}_{12}(x) = \frac{1}{n} \sum_{i=1}^n \bar{\delta}(i) \prod_{v=1}^k \frac{1}{c_v} \Phi\left(\frac{x_v - x_v^i}{c_v}\right), \quad (3)$$

где

$$\bar{\delta}(i) = \begin{cases} -\bar{P}_1^{-1}, \text{ если } x^i \in \Omega_1 \\ \bar{P}_2^{-1}, \text{ если } x^i \in \Omega_2. \end{cases}$$

Здесь  $\bar{P}_1 = n_1/n, \bar{P}_2 = n_2/n$  – оценки априорных вероятностей классов  $\Omega_1, \Omega_2$ .

Выбор оптимальных значений  $\bar{c}_v, v = \overline{1, k}$  коэффициентов размытости  $c = (c_v, v = \overline{1, k})$  ядерных функций непараметрического решающего правила  $\bar{m}(x)$  осуществляется из условия минимума оценки вероятности ошибки распознавания образов

$$\bar{\rho}(c) = \frac{1}{n} \sum_{t=1}^n 1(\delta(t), \bar{\delta}(t)), \quad (4)$$

где индикаторная функция

$$1(\delta(t), \bar{\delta}(t)) = \begin{cases} 0, \text{ если } \delta(t) = \bar{\delta}(t) \\ 1, \text{ если } \delta(t) \neq \bar{\delta}(t). \end{cases}$$

Здесь  $\bar{\delta}(t)$  – «решение» алгоритма  $\bar{m}(x)$  (1) о принадлежности значений  $x^t$  к одному из классов  $\Omega_1, \Omega_2$ .

При вычислении  $\bar{\rho}(c)$  «решение»  $\bar{\delta}(t)$  алгоритма (1) определяется в соответствии со знаком статистики

$$\tilde{f}_{12}(x^i) = \frac{1}{n} \sum_{i \neq l} \bar{\delta}(i) \prod_{v=1}^k \frac{1}{c_v} \Phi\left(\frac{x_v^i - x_v^l}{c_v}\right).$$

Ситуация  $x^l$ , которая подаётся на контроль, исключается из процесса обучения алгоритма (1).

Оптимизацию непараметрического решающего правила (1) по коэффициентам размытости ядерных функций  $c_v, v = \overline{1, k}$ , можно упростить, если положить  $c_v = c \bar{\sigma}_v(j), v = \overline{1, k}$ , где  $\bar{\sigma}_v(j)$  – оценка среднеквадратического отклонения случайной величины  $x_v$  в классе  $\Omega_j$ . Данное утверждение является очевидным, так как большей длине интервала значений  $x_v$  соответствует больший коэффициент размытости  $c_v$  ядерной функции  $\Phi(u_v), v = \overline{1, k}$ . Подобный подход использовался при построении «быстрых» процедур оптимизации непараметрических оценок плотности вероятности ядерного типа [15 – 19].

В этом случае непараметрическая оценка уравнения разделяющей поверхности (3) запишется в виде

$$\bar{f}_{12}(x) = (n c^k)^{-1} \sum_{i=1}^n \bar{\delta}(i) \prod_{v=1}^k \frac{1}{c \bar{\sigma}_v^i} \Phi\left(\frac{x_v - x_v^i}{c \bar{\sigma}_v^i}\right), \quad (5)$$

где

$$\bar{\sigma}_v^i = \begin{cases} \bar{\sigma}_v(1), & \text{если } x^i \in \Omega_1 \\ \bar{\sigma}_v(2), & \text{если } x^i \in \Omega_2. \end{cases}$$

Значения оценок среднеквадратических отклонений  $\bar{\sigma}_v(j)$  определяются по статистическим данным  $x_v^i, i \in I_j$  из выборки  $V_j$

$$\bar{\sigma}_v(j) = \left( \frac{1}{n_j - 1} \sum_{i \in I_j} (x_v^i - \bar{x}_v)^2 \right)^{1/2}, \quad j = 1, 2.$$

Здесь  $\bar{x}_v$  – среднее значение случайной величины  $x_v$  в соответствующем классе.

Поэтому появляется возможность оптимизацию непараметрического алгоритма распознавания образов (1) проводить лишь по одному параметру  $c$  коэффициентов размытости ядерных функций из условия минимума оценки вероятности ошибки распознавания образов (4).

Проверить гипотезу  $\bar{H}_0: \rho = 0,5$ , используя методику доверительного оценивания ошибки распознавания образов  $\rho$  [3] либо критерий Колмогорова [20]. В первом случае гипотеза  $\bar{H}_0$  выполняется, если значение  $\rho = 0,5$  попадает в интервал

$$\rho_1, \rho_2 = \frac{\bar{\rho} + \frac{t_\alpha^2}{2n} \mp t_\alpha \sqrt{\frac{\bar{\rho}(1-\bar{\rho})}{n} + \frac{t_\alpha^2}{4n^2}}}{1 + \frac{t_\alpha^2}{n}}.$$

Здесь  $t_\alpha$  – значение аргумента функции Лапласа, при котором выполняется условие

$$2F(t_\alpha) = \alpha,$$

где  $\alpha$  – коэффициент доверия принадлежности  $\rho$  интервалу  $[\rho_1, \rho_2]$ .

При использовании критерия Колмогорова отклонение  $\bar{D}_{12} = |0,5 - \bar{\rho}(\bar{c})|$  сравнивается с его пороговым значением

$$D_\beta = \sqrt{-\ln\left(\frac{\beta}{2}\right) / (2n)}.$$

Здесь  $\beta$  – вероятность (риск) отвергнуть правильную гипотезу  $\bar{H}_0$ .

Если выполняется соотношение  $\bar{D}_{12} < D_\beta$ , то гипотеза  $\bar{H}_0$  справедлива и риск её отвергнуть равен  $\beta$ .

Анализ результатов вычислительных экспериментов показывает, что при значительном отличии объёмов  $n_1, n_2$  сравниваемых последовательностей  $V_1, V_2$  эффективность предлагаемого метода снижается.

Предлагаемая методика реализована в программном комплексе NSARSD v. 1,0 в среде Delphi. Его функциональные возможности позволяют дополнительно решать следующие задачи: оценивание достаточного полного набора количественных показателей законов распределений случайных величин  $x_v, v = \overline{1, k}$  в классах, оценивание принадлежности ситуации  $x$  к тому или иному классу и пространственное распределение состояний изучаемых объектов с последующей графической интерпретацией результатов классификации.

## 2. Проверка гипотез о распределениях многомерных спектральных данных дистанционного зондирования лесных массивов

Рассмотрим задачи проверки гипотезы о тождественности распределений многомерных спектральных данных дистанционного зондирования, соответствующих состояниям древостоев лесных массивов с использованием предложенной методики.

Исходная информация получена по данным дистанционного зондирования территории юга Красноярского края с помощью аппарата Landsat по шести спектральным каналам. Спектральные каналы имеют разрешение (размер одного пикселя) 30 м. Для оценивания состояний древостоев лесных массивов специалистами по дешифрированию космических снимков была сформирована обучающая выборка  $V = (x^i, \delta(i), i = \overline{1, n})$ , где  $\delta(i)$  – указания о принадлежности спектральных данных  $x^i = (x_v^i, v = \overline{1, 6})$  к

одному из состояний древостоев  $\Omega_j, j = \overline{1, 3}$  (темнохвойный лес, повреждённые (усыхающие) и сухие древостои) [21]. Количество ситуаций выборки  $V$ , принадлежащих состояниям  $\Omega_j, j = \overline{1, 3}$ , определялись значениями:  $n_1 = 968, n_2 = 198, n_3 = 570$ . Каждый элемент земной поверхности характеризовался шестью спектральными признаками  $x = (x_1, \dots, x_6)$ , которым свойственны следующие длины волн (наномет-

ры): 450–515 ( $x_1$ ), 525–600 ( $x_2$ ), 630–680 ( $x_3$ ), 845–885 ( $x_4$ ), 1560–1660 ( $x_5$ ), 2100–2300 ( $x_6$ ).

Проверялись гипотезы о тождественности законов распределения  $x = (x_v, v = \overline{1, 6})$  в состояниях  $\Omega_j$ ,  $j = \overline{1, 3}$ . Для каждого состояния  $\Omega_j$  соответствующая ему выборка  $V_j$  случайным образом разбивалась на две части  $V_j(1), V_j(2)$  равных объёмов. Проверялись гипотезы

$$H_0(j): F_{1j}(x) \equiv F_{2j}(x)$$

о тождественности функций распределений  $F_{1j}(x), F_{2j}(x)$ , представленных статистическими данными  $V_j(1), V_j(2)$ ,  $j = \overline{1, 3}$ . При этом значения  $\alpha = 0,95$ , а  $\beta = 0,05$ . Результаты вычислительных экспериментов представлены в табл. 1.

В таблице приведены средние значения параметров исследуемой методики, которые вычислялись по количеству результатов вычислительных экспериментов  $m = 50$ .

Табл. 1. Результаты вычислительных экспериментов

	Значения параметров предлагаемой методики					
	$\bar{\rho}(\bar{c})$	$\bar{c}$	$\rho_1$	$\rho_2$	$\bar{D}$	$D_\beta$
$\Omega_1$	0,478	2,44	0,447	0,509	0,022	0,044
$\Omega_2$	0,444	0,58	0,378	0,513	0,056	0,097
$\Omega_3$	0,477	0,87	0,437	0,518	0,023	0,057

В соответствии с данными табл. 1 законы распределения спектральных признаков в состояниях  $\Omega_j$ ,  $j = \overline{1, 3}$  являются тождественными. Результаты проверки гипотез законов распределения  $F_{1j}(x), F_{2j}(x)$ ,  $j = \overline{1, 3}$  с использованием методики доверительного оценивания  $\rho$  и критерия Колмогорова сопоставимы. Это утверждение подтверждается справедливостью соотношений  $\bar{D}_{12} < D_\beta$  и принадлежностью  $\bar{\rho}(\bar{c})$  доверительному интервалу  $\rho_1 \leq \bar{\rho}(\bar{c}) \leq \rho_2$  для всех состояний лесных массивов  $\Omega_j$ ,  $j = \overline{1, 3}$ . При этом оценка вероятности отвергнуть гипотезу  $H_0(1)$  для темнохвойного леса не превышает значения 0,02, а гипотезу  $H_0(2)$  для усыхающих древостоев – не более 0,05. Данный факт объясняется неоднородностью состояний древостоев  $\Omega_1, \Omega_2$ . Например, темнохвойный лес состоит из различных их типов (сосна, кедр, ель, лиственница, пихта), которые растут на территории с различным рельефом. Эти условия проявляются в значениях спектральных признаков элементов земной поверхности. Понятие усыхающие древостои в своём определении характеризуются их неоднородностью, что влияет на решения специалистов по дешифрированию лесных массивов. Гипотезы о тождественности законов распределения спектральных данных, соответствующих различным состояниям древостоев лесных массивов, не подтверждаются при реализации различных условий вычислительных экспериментов.

Дальнейшее развитие предлагаемой методики состоит в разработке её модификации на условия срав-

ниваемых выборок  $V_1, V_2$  многомерных случайных величин малого объёма, когда возникает необходимость сокращения затрат на дешифрирование исходной информации дистанционного зондирования. При малых выборках  $V_1, V_2$  традиционные непараметрические оценки плотности вероятности типа Розенблатта–Парзена  $\bar{p}_1(x), \bar{p}_2(x)$  и построенные на их основе ядерные оценки уравнений разделяющих поверхностей вида (3) характеризуются большой дисперсией, что сказывается на точности оценивания вероятности ошибки распознавания образов (4). Для этих условий целесообразно использовать непараметрические оценки плотностей вероятностей с неявно заданной формой ядерной функции [14]

$$\tilde{p}_j(x) = \left( n \prod_{v=1}^k 2\beta_v c_v \right)^{-1} \sum_{i \in I_j} \prod_{v=1}^k \int_{x_v - \beta_v}^{x_v + \beta_v} \Phi \left( \frac{u_v - x_v^i}{c_v} \right) du_v,$$

$j = \overline{1, 2}$  при оценивании уравнения разделяющей поверхности  $f_{12}(x) = p_2(x) - p_1(x)$  между классами  $\Omega_1, \Omega_2$ . Синтез статистик  $\tilde{p}_j(x)$  основан на сглаживании непараметрических оценок плотностей вероятности  $\bar{p}_j(x)$  путём использования оператора интегрирования в интервале  $[x_v \pm \beta_v, v = \overline{1, k}]$ .

### Заключение

Предложенная методика проверки гипотез о распределениях случайных величин обеспечивает обход проблемы декомпозиции области значений случайных величин на многомерные интервалы, которая свойственна для критерия Пирсона. Проверка гипотезы о тождественности двух законов распределений многомерных случайных величин заменена на проверку гипотезы о равенстве ошибки распознавания образов значению 0,5. Использование метода доверительного оценивания вероятности ошибки распознавания образов и критерия Колмогорова приводит к сопоставимым результатам. Эффективность предложенной методики подтверждается её применением при проверке гипотез о распределении спектральных данных дистанционного зондирования древостоев лесных массивов. Рассматриваемая методика может быть использована при формировании наборов значимых спектральных признаков в задаче оценивания состояний лесных массивов по степени их усыхания, что является важным в задачах принятия решений о пожароопасности лесных массивов.

### Литература

1. **Афанасьев, А.А.** Гибридные методы автоматизированной идентификации изменений ландшафтного покрова по данным дистанционного зондирования Земли в условиях шумов / А.А. Афанасьев, А.В. Замятин // Компьютерная оптика. – 2017. – Т. 41, № 3. – С. 431-440. – DOI: 10.18287/2412-6179-2017-41-3-431-440.
2. **Васин, Ю.Г.** Распределённая СУБД для интегрированной обработки пространственных данных в ГИС / Ю.Г. Васин, Ю.В. Ясаков // Компьютерная оптика. –

2016. – Т. 40, № 6. – С. 919-928. – DOI: 10.18287/2412-6179-2016-40-6-919-928.
3. **Пугачёв, В.С.** Теория вероятностей и математическая статистика : учебное пособие / В.С. Пугачёв. – М.: Физматлит, 2002. – 496 с.
  4. **Лапко, А.В.** Непараметрические алгоритмы распознавания образов в задаче проверки статистической гипотезы о тождественности двух законов распределения случайных величин / А.В. Лапко, В.А. Лапко // Автометрия. – 2010. – Т. 46, № 6. – С. 47-53.
  5. **Лапко, А.В.** Сравнение эмпирической и теоретической функций распределения случайной величины на основе непараметрического классификатора / А.В. Лапко, В.А. Лапко // Автометрия. – 2012. – Т. 48, № 1. – С. 45-49.
  6. **Лапко, А.В.** Анализ асимптотических свойств непараметрической оценки уравнения разделяющей поверхности в двальтернативной задаче распознавания образов / А.В. Лапко, В.А. Лапко // Автометрия. – 2010. – Т. 46, № 3. – С. 48-53.
  7. Непараметрические системы классификации / А.В. Лапко, В.А. Лапко, М.И. Соколов, С.В. Ченцов. – Новосибирск: Наука, 2000. – 240 с.
  8. **Fukunaga, K.** Introduction to statistical pattern recognition / K. Fukunaga. – San Diego: Academic Press, 1990. – 614 p.
  9. **Theodoridis, S.** Pattern recognition / S. Theodoridis, K. Koutroumbas. – Burlington, MA: Academic Press, 2009. – 961 p.
  10. **Webb, A.R.** Statistical pattern recognition / A.R. Webb, K.D. Copesey. – Chichester: John Wiley & Sons, 2011. – 666 p.
  11. **Parzen, E.** On estimation of a probability density function and mode / E. Parzen // The Annals of Mathematical Statistics. – 1962. – Vol. 33, Issue 3. – P. 1065-1076. – DOI: 10.1214/aoms/1177704472.
  12. **Епанечников, В.А.** Непараметрическая оценка многомерной плотности вероятности / В.А. Епанечников // Теория вероятности и её применения. – 1969. – Т. 14, № 1. – С. 156-161.
  13. **Лапко, А.В.** Регрессионная оценка многомерной плотности вероятности и её свойства / А.В. Лапко, В.А. Лапко // Автометрия. – 2014. – Т. 50, № 2. – С. 50-56.
  14. **Лапко, А.В.** Непараметрическая оценка плотности вероятности парзеновского типа с неявно заданной формой ядерной функции / А.В. Лапко, В.А. Лапко // Измерительная техника. – 2016. – № 6. – С. 14-17.
  15. **Sheather, S.J.** Density estimation / S.J. Sheather // Statistical Science. – 2004. – Vol. 19, Issue 4. – P. 588-597. – DOI: 10.1214/088342304000000297.
  16. **Scott, D.W.** Multivariate density estimation: Theory, practice, and visualization / D.W. Scott. – New Jersey: John Wiley & Sons, 2015. – 384 p.
  17. **Chen, S.** Optimal bandwidth selection for kernel density functionals estimation / S. Chen // Journal of Probability and Statistics. – 2015. – Vol. 2015(1). – P. 1-21. – DOI: 10.1155/2015/242683.
  18. **Borrajo, M.I.** Bandwidth selection for kernel density estimation with length-biased data / M.I. Borrajo, W. González-Manteiga, M.D. Martínez-Miranda // Journal of Nonparametric Statistics. – 2017. – Vol. 29, Issue 3. – P. 636-668. – DOI: 10.1080/10485252.2017.1339309.
  19. **Лапко, А.В.** Быстрый алгоритм выбора коэффициентов размытости ядерных функций в непараметрической оценке плотности вероятности / А.В. Лапко, В.А. Лапко // Измерительная техника. – 2018. – № 6. – С. 16-20. – DOI: 10.32446/0368-1025it-2018-6-16-20.
  20. **Шаракшанэ, А.С.** Сложные системы / А.С. Шаракшанэ, И.Г. Железнов, В.А. Ивницкий. – М.: Высшая школа, 1977. – 248 с.
  21. **Kharuk, V.I.** Fir decline and mortality in the southern siberian mountains / V.I. Kharuk, S.T. Im, I.A. Petrov, M.L. Dvinskaya, E.V. Fedotova, K.J. Ranson // Regional Environmental Change. – 2017. – Vol. 17, Issue 3. – P. 803-812. – DOI: 10.1007/s10113-016-1073-5.

#### Сведения об авторах

**Лапко Александр Васильевич**, 1949 года рождения, в 1971 году окончил Фрунзенский политехнический институт по специальности «Автоматика и телемеханика», доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук, профессор кафедры космических средств и технологий Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Область научных интересов: непараметрическая статистика; распознавания образов и анализ изображений; моделирование и оптимизация неопределённых систем, дистанционное зондирование. E-mail: [lapko@icm.krasn.ru](mailto:lapko@icm.krasn.ru).

**Лапко Василий Александрович**, 1974 года рождения, в 1996 году окончил Красноярский государственный технический университет по специальности «Управление и информатика в технических системах», доктор технических наук, профессор, ведущий научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук, заведующий кафедрой космических средств и технологий Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Область научных интересов: непараметрическая статистика; распознавания образов и анализ изображений; моделирование неопределённых систем, дистанционное зондирование. E-mail: [valapko@yandex.ru](mailto:valapko@yandex.ru).

ГРНТИ: 28.23.15

Поступила в редакцию 16 октября 2018 г. Окончательный вариант – 30 января 2019 г.

## A technique for testing hypotheses for distributions of multidimensional spectral data using a nonparametric pattern recognition algorithm

A.V. Lapko<sup>1,2</sup>, V.A. Lapko<sup>1,2</sup>

<sup>1</sup>Institute of Computational Modeling of the Siberian Branch of the Russian Academy of Sciences, Russia, Krasnoyarsk,  
<sup>2</sup>Reshetnev Siberian State University of Science and Technology, Russia, Krasnoyarsk

### Abstract

The paper deals with a new method of testing hypotheses for the distribution of multidimensional remote sensing spectral data. The proposed technique is based on the use of nonparametric algorithms for pattern recognition. Testing the hypothesis of the identity of two laws of distributions of multidimensional random variables is replaced by testing a hypothesis stating that the pattern recognition error equals 0.5. The application of this technique allows doing without the decomposition of the random variable domain into multidimensional intervals, which is typical for the Pearson criterion. Its effectiveness is confirmed by the results of testing the hypotheses of the distribution of spectral data of remote sensing in forestry. The analysis of the distribution laws for the following types of forestry is carried out: dark coniferous forest, damaged and dry forest stands. The initial information was obtained from the southern Siberia remote sensing data using six spectral channels of Landsat. The results of the research form a basis for a set of significant spectral features when dealing with forest condition monitoring.

**Keywords:** testing a statistical hypothesis, multivariate random variables, pattern recognition, kernel density estimation, selecting bandwidth, spectral data, remote sensing, forest conditions.

**Citation:** Lapko AV, Lapko VA. A technique for testing hypotheses for distributions of multidimensional spectral data using a nonparametric pattern recognition algorithm. *Computer Optics* 2019; 43(2): 238-244. DOI: 10.18287/2412-6179-2019-43-2-238-244.

### References

- [1] Afanasyev AA, Zamyatin AV. Hybrid methods for automatic landscape change detection in noisy data environment. *Computer Optics* 2017; 41(3): 431-440. DOI: 10.18287/2412-6179-2017-41-3-431-440.
- [2] Vasin YuG, Yasakov YuV. Distributed database management system for integrated processing of spatial data in a GIS. *Computer Optics* 2016; 40(6): 919-928. DOI: 10.18287/2412-6179-2016-40-6-919-928.
- [3] Pugachev VS. Probability theory and mathematical statistics: textbook [In Russian]. Moscow: "Fizmatlit" Publisher; 2002.
- [4] Lapko AV, Lapko VA. Nonparametric algorithms of pattern recognition in the problem of testing a statistical hypothesis on identity of two distribution laws of random variables. *Optoelectronics, Instrumentation and Data Processing* 2010; 46(6): 545-550. DOI: 10.3103/S8756699011060069.
- [5] Lapko AV, Lapko VA. Comparison of empirical and theoretical distribution functions of a random variable on the basis of a nonparametric classifier. *Optoelectronics, Instrumentation and Data Processing* 2012; 48(1): 37-41. DOI: 10.3103/S8756699012010050.
- [6] Lapko AV, Lapko VA. Analysis of asymptotic properties of nonparametric estimate of the equation of the separation surface in a two-alternative problem of pattern recognition. *Optoelectronics, Instrumentation and Data Processing* 2010; 46(3): 243-247. DOI: 10.3103/S8756699010030064.
- [7] Lapko AV, Lapko VA, Sokolov MI, Chentsov SV. Nonparametric classification systems [In Russian]. Novosibirsk: "Nauka" Publisher; 2000.
- [8] Fukunaga K. Introduction to statistical pattern recognition. San Diego: Academic Press; 1990.
- [9] Theodoridis S, Koutroumbas K. Pattern recognition. Burlington, MA: Academic Press; 2009.
- [10] Webb AR, Copsey KD. Statistical pattern recognition. Chichester: John Wiley & Sons, 2011.
- [11] Parzen E. On estimation of a probability density function and mode. *Ann Math Statistic* 1962; 33(3): 1065-1076. DOI: 10.1214/aoms/1177704472.
- [12] Epanechnikov VA. Nonparametric estimation of multidimensional probability density [In Russian]. *Theory of Probability and its Applications* 1969; 14(1): 156-161.
- [13] Lapko AV, Lapko VA. Regression estimate of the multidimensional probability density and its properties. *Optoelectronics, Instrumentation and Data Processing* 2014; 50(2): 148-153. DOI: 10.3103/S875669901402006X.
- [14] Lapko AV, Lapko VA. Nonparametric estimate of a parzen-type probability density with an implicitly specified form of the kernel. *Measurement Techniques* 2016; 59(6): 571-576. DOI: 10.1007/s11018-016-1010-5.
- [15] Sheather SJ. Density estimation. *Statistical Science* 2004; 19(4): 588-597. DOI: 10.1214/088342304000000297.
- [16] Scott DW. Multivariate density estimation: Theory, practice, and visualization. New Jersey: John Wiley & Sons, Inc; 2015.
- [17] Chen S. Optimal bandwidth selection for kernel density functionals estimation. *Journal of Probability and Statistics* 2015; 2015(1): 1-21. DOI: 10.1155/2015/242683.
- [18] Borrajo MI, González-Manteiga W, Martínez-Miranda MD. Bandwidth selection for kernel density estimation with length-biased data. *Journal of Nonparametric Statistics* 2017; 29(3): 636-668. DOI: 10.1080/10485252.2017.1339309.
- [19] Lapko AV, Lapko VA. Fast algorithm for choosing kernel function blur coefficients in a nonparametric probability density estimate. *Measurement Techniques* 2018; 61(6): 540-545. DOI: 10.1007/s11018-018-1463-9.
- [20] Sharakshaneh AS, Zheleznov IG, Ivnitiskij VA. Complex system [In Russian]. Moscow: "Vysshaya shkola" Publisher; 1977.
- [21] Kharuk VI, Im ST, Petrov IA, Dvinskaya ML, Fedotova EV, Ranson KJ. Fir decline and mortality in the southern siberian mountains. *Regional Environmental Change* 2017; 17(3): 803-812. DOI 10.1007/s10113-016-1073-5.

---

*Authors' information*

**Alexander Vasilievich Lapko** (1949 b.), graduated from Frunze Polytechnic Institute on speciality “Automation and telemechanics” in 1971. Doctor of Science in technology, professor, honored worker of science of the Russian Federation, chief researcher of the Institute of Computational Modeling of the Siberian branch of the Russian Academy of Sciences; Professor of Space Facilities and Technologies department of the Reshetnev Siberian State University of Science and Technology. Research interests: nonparametric statistics; pattern recognition and image analysis; modeling and optimization of uncertain systems; remote sensing. E-mail: [lapko@icm.krasn.ru](mailto:lapko@icm.krasn.ru).

**Vasily Aleksandrovich Lapko** (1974 b.), graduated from Krasnoyarsk State Technical University on speciality “Management and Informatics in Technical Systems” in 1996. Doctor of Science in Technology, professor, leading researcher at the Institute of Computational Modeling of the Siberian branch of the Russian Academy of Sciences; Head of Space Facilities and Technologies department of the Reshetnev Siberian State University of Science and Technology. Research interests: nonparametric statistics; pattern recognition and image analysis; modeling of uncertain systems; remote sensing. E-mail: [valapko@yandex.ru](mailto:valapko@yandex.ru).

---

*Received October 16, 2018. The final version – January 30, 2019.*

---