

УДК 528.8

КЛАССИФИКАЦИЯ СОСТОЯНИЙ УСЫХАНИЯ ЕЛИ ОБЫКНОВЕННОЙ НА ОСНОВЕ СПЕКТРОВ ОТРАЖЕНИЯ

А. О. МАРТИНОВ¹⁾

¹⁾Институт прикладных физических проблем им. А. Н. Севченко БГУ,
ул. Курчатова, 7, 220045, г. Минск, Беларусь

Представлен метод классификации спектров отражения еловой хвои разных категорий здоровья, и проведена оценка качества классификации. Выбраны такие категории здоровья елей, которые включают в себя начальные стадии усыхания, что играет существенную роль для раннего детектирования очагов заболевания, когда затруднена классификация по визуальным критериям средствами дистанционного зондирования Земли. Предложен алгоритм необучаемой классификации и визуализации спектральных данных на основе корреляционного и кластерного анализа. С помощью разработанного программного обеспечения исследованы полученные в лабораторных условиях спектры отражения еловой хвои и интерпретированы результаты. В ходе анализа различных сочетаний параметров внутри предложенного алгоритма, а также комбинаций отдельных компонент алгоритма с известными методами классификации определена наиболее эффективная комбинация параметров и методов классификации (проекция спектров в пространство главных компонент, устранение влияния первой главной компоненты на спектры, метрика связывания кластеров Уорда и стандартизированная евклидова метрика для вычисления спектрального расстояния) для детектирования разных стадий заболевания елей. Ее использование позволило повысить показатель качества классификации *F-score* для 2-й категории здоровья (наиболее важная категория для задачи обнаружения усыхания на ранних стадиях) до 70,59 %.

Ключевые слова: спектр отражения; еловая хвоя; усыхание; метод главных компонент; классификация.

Благодарность. Автор благодарит сотрудников отдела аэрокосмических исследований Института прикладных физических проблем имени А. Н. Севченко БГУ И. И. Бручковского, О. О. Силюк, Г. С. Литвиновича, А. А. Ломако, В. В. Станчика, С. И. Гуляеву за создание гониофотометра для измерения отражательных характеристик хвои, а также Л. В. Катковского за научную консультацию. Работа выполнена при поддержке Всемирной федерации ученых.

CLASSIFICATION OF NORWAY SPRUCE DRYING STATES ON THE BASIS OF REFLECTION SPECTRA

A. O. MARTINOV^a

^aA. N. Sevchenko Institute of Applied Physical Problems, Belarusian State University,
7 Kurčatava Street, Minsk 220045, Belarus

The article is devoted to the development of a method for classifying the reflection spectra of spruce needles of different health categories and assessing the quality of the classification. Such health categories of spruces have been chosen that include the initial stages of drying out, which is essential for early detection of disease foci, but makes it difficult to classify according to visual criteria by means of remote sensing of the Earth. An algorithm for untrained classification and visualisation of spectral data based on correlation and cluster analysis is proposed. The reflection spectra of spruce needles obtained under laboratory conditions were studied and the results were interpreted using the developed software. The analysis of various combinations of parameters within the proposed algorithm, as well as combinations of individual components of the algorithm with known classification methods, made it possible to determine the most effective combination of parameters and classification methods (projection of spectra into the space of principal components, elimination of the influence of the

Образец цитирования:

Мартинов АО. Классификация состояний усыхания ели обыкновенной на основе спектров отражения. *Журнал Белорусского государственного университета. Физика.* 2022; 3:26–38.
<https://doi.org/10.33581/2520-2243-2022-3-26-38>

For citation:

Martinov AO. Classification of Norway spruce drying states on the basis of reflection spectra. *Journal of the Belarusian State University. Physics.* 2022;3:26–38. Russian.
<https://doi.org/10.33581/2520-2243-2022-3-26-38>

Автор:

Антон Олегович Мартинов – кандидат физико-математических наук; научный сотрудник лаборатории оптико-физических измерений.

Author:

Anton O. Martinov, PhD (physics and mathematics); researcher at the laboratory of optical and physical measurements.
antonmartenov@gmail.com
<https://orcid.org/0000-0001-6871-1173>

first principal component on spectra, the Ward cluster linking metric and the standardised Euclidean metric for calculating the spectral distance) for detection of different stages of spruce disease. Its use to made it possible to increase the *F-score* classification quality indicator for the 2nd health category (the most important category for the task of detecting drying in the early stages) up to 70.59 %.

Keywords: reflection spectrum; spruce needles; drying; principal component analysis; classification.

Acknowledgements. The author thanks the employees of the department of aerospace research of the A. N. Sevchenko Institute of Applied Physical Problems of Belarusian State University I. I. Bruchkouski, V. A. Siliuk, H. S. Litvinovich, A. A. Lamaka, V. V. Stanchuk, S. I. Guliaeva for the creation of a goniophotometer for measuring the reflective spectra of needles, as well as L. V. Katkovsky for scientific advice. The work was supported by World Federation of Scientists.

Введение

В конце XX в. в странах Северного полушария началось крупномасштабное ухудшение санитарного состояния еловых лесов [1]. В последнее время все большее внимание уделяется естественным стрессовым факторам [2], к которым, например, относится заражение жуком-короедом. Для определения масштабов и динамики изменения санитарного состояния лесов необходимы их мониторинг и классификация на категории здоровья. Подобная классификация выполняется с помощью визуального наземного осмотра дерева [3]. Однако для масштабного мониторинга состояния лесов требуются иные подходы. Одним из таких подходов является дистанционное зондирование Земли (ДЗЗ). В ряде работ применяется классификация крон елей по RGB-изображениям и мультиспектральным изображениям, полученным посредством ДЗЗ [4–8]. Но, несмотря на использование нескольких категорий здоровья (в большинстве указанных работ в качестве них выбраны ели с зеленой кроной, усохшие ели без хвои и 1–2 промежуточных состояния с ярко выраженными признаками усыхания в виде рыжей кроны), классификация елей на такие категории здоровья как экспертными методами на основе изображений, так и программными средствами с применением различных алгоритмов не представляет особой сложности. Разработка компьютерных алгоритмов определения настолько разных категорий здоровья елей направлена на приближение точности автоматического классификатора к 100 % (в работе [4] удалось превысить точность 90 %). С точки зрения практического использования материалов лесопатологического мониторинга интерес представляет картирование лесных экосистем по иным категориям здоровья. В частности, широкое применение получила следующая шкала категорий здоровья хвойных деревьев: 1-я категория – без признаков ослабления (крона деревьев густая, хвоя зеленая, блестящая); 2-я категория – ослабленные (крона ажурная, хвоя зеленая или светло-зеленая, усыхание отдельных ветвей); 3-я категория – сильно ослабленные (крона сильно ажурная, хвоя бледно-зеленая или матовая, усыхание до 2/3 кроны); 4-я категория – усыхающие (крона сильно ажурная, хвоя желтоватая или желто-зеленая, осыпается); 5-я категория – свежий сухостой; 6-я категория – старый сухостой¹. Особый интерес представляют первые четыре категории. Определение усыхания на раннем этапе позволяет принять превентивные меры по предотвращению распространения заражения, например, жуком-короедом. Изображения первых четырех категорий здоровья деревьев, выделенных по результатам проведенных совместно с лесопатологом полевых измерений, представлены на рис. 1.



Рис. 1. Внешний вид разных категорий здоровья елей:
а – 1-я категория; б – 2-я категория; в – 3-я категория; г – 4-я категория

Fig. 1. Appearance of different health categories of spruces:
а – 1st category; б – 2nd category; в – 3rd category; г – 4th category

¹Санитарные правила в лесах Российской Федерации : утв. Рослесхозом 18.05.1992 : в ред. от 20.01.1995. М. : ВНИИЦлес-ресурс, 1998. 25 с.

Как видно из рис. 1, первые три категории трудноразличимы визуально (часто определяются по другим признакам, например по повышенному выделению смолы на стволе дерева), только четвертая категория отличается наличием участков с рыжей хвоей. В связи с этим в данной работе было предложено анализировать спектры отражения еловой хвои, чтобы исключить возможную потерю информации в отдельных широких спектральных полосах мультиканального изображения.

Материалы и методы исследования

Классификации елей по летным данным должно предшествовать исследование, включающее в себя тщательный отбор образцов, их спектрометрирование, разработку методов и алгоритмов классификации и анализ результатов. Сначала путем наземных измерений с участием лесопатолога отбирались и классифицировались по степени усыхания ели. Затем с крон этих елей собирались образцы хвои. После этого проводилось спектрометрирование собранных образцов с помощью автоматизированного лабораторного гониофотометра, который позволяет измерять коэффициент отражения исследуемого образца со спектральным разрешением 1,5–2,9 нм в диапазоне 400–1050 нм и спектральным разрешением 10 нм в диапазоне 1000–2500 нм [9]. В результате лабораторных измерений были получены 356 спектров хвои ели обыкновенной в диапазоне 400–870 нм, каждому из которых присвоена одна из четырех категорий здоровья. Полученный набор спектров использован для разработки и апробации алгоритма классификации в данной работе.

Алгоритм классификации спектральных данных. Для анализа близких спектров отражения, какими являются спектры хвои елей, разработан алгоритм классификации на основе методов главных компонент, корреляционного и кластерного анализа, предлагающий наглядный способ визуализации классов. Структурная схема разработанного алгоритма классификации приведена на рис. 2. Этапы 4 и 7 обработки представлены на рис. 3.

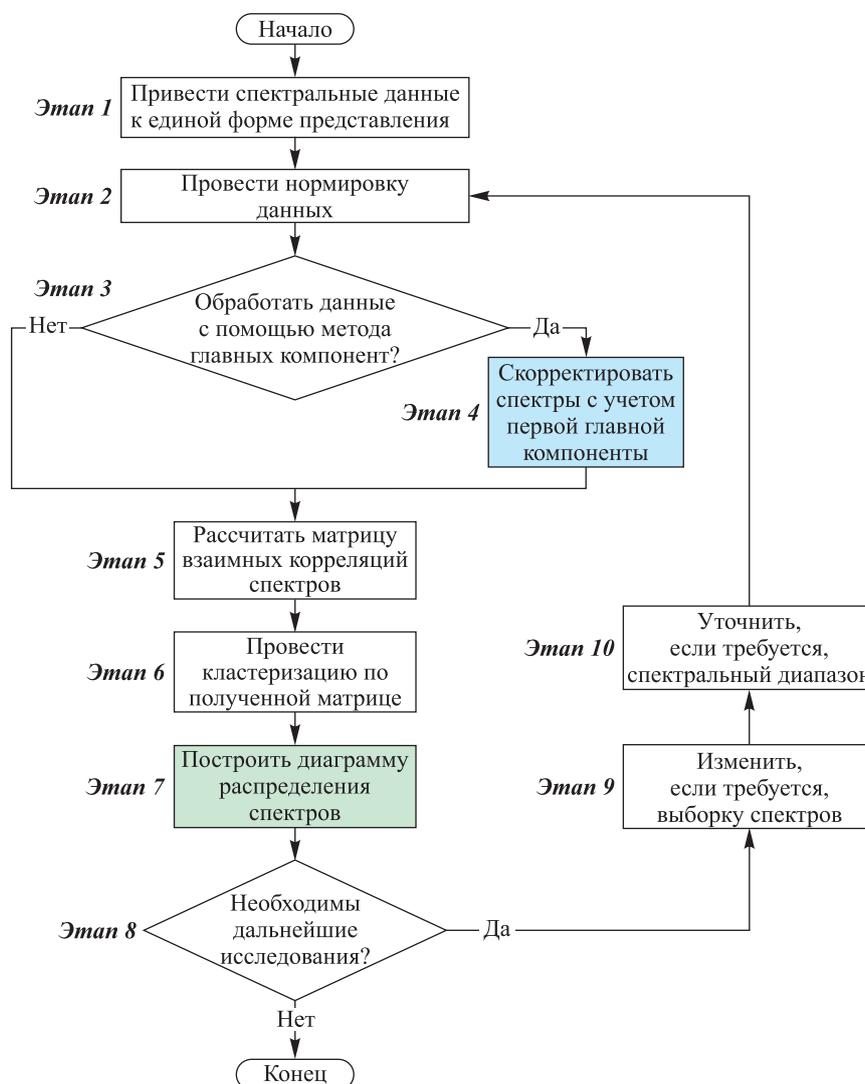


Рис. 2. Структурная схема разработанного алгоритма классификации
Fig. 2. Block diagram of the developed classification algorithm



Рис. 3. Структурные схемы процедур корректировки спектров с учетом первой главной компоненты (а) и построения диаграммы распределения спектров (б)

Fig. 3. Structural schemes of the procedures of correction of the spectra taking into account the first principal component (a) and spectrum distribution diagram formation (b)

Рассмотрим основные этапы алгоритма (этапы 1–8 на рис. 2).

Этап 1. Переход к экспериментальным отсчетам спектров в общей системе спектральных каналов посредством интерполяции. Данные (коэффициенты отражения) хранятся в матрице, в столбцах которой располагаются номера спектров, а в строках – номера спектральных каналов.

Этап 2. Проведение нормировки спектров в целях устранения различий в абсолютных значениях спектральной выборки. Для этого из каждого спектра вычитается выборочное среднее этого спектра и результат делится на среднеквадратическое отклонение:

$$x'_{ij} = \frac{x_{ij} - \bar{X}_j}{\sigma(X_j)},$$

где $i = \overline{1, K}$ – номер спектрального канала; $j = \overline{1, N}$ – номер спектра; K – количество длин волн; N – количество спектров; \bar{X}_j – математическое ожидание j -го спектра; $\sigma(X_j)$ – среднеквадратическое отклонение для j -го спектра.

Этап 3. Принятие решения, требуется ли производить обработку данных с помощью метода главных компонент. Если в текущей выборке спектры имеют схожую форму (коэффициент корреляции более 0,9), то будет целесообразно ее провести.

Этап 4. Корректировка спектров с учетом первой главной компоненты (описание данного этапа см. ниже).

Этап 5. Расчет матрицы взаимных корреляций спектров. Так как у спектральных кривых основную роль играет форма кривой [10], то в качестве критерия сравнения спектров была выбрана корреляция. Коэффициент корреляции двух спектральных кривых X и Y рассчитывается по формуле

$$r_{XY} = \frac{\text{cov}_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^K (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^K (X_i - \bar{X})^2 \sum_{i=1}^K (Y_i - \bar{Y})^2}}, \quad \bar{X} = \frac{1}{K} \sum_{i=1}^K X_i, \quad \bar{Y} = \frac{1}{K} \sum_{i=1}^K Y_i,$$

где K – количество спектральных каналов; X и Y – сравниваемые спектры. Полученный коэффициент корреляции определяет сходство формы спектров. Для сравнения спектров строится матрица взаимных корреляций размера $N \times N$ с элементами r_{ij} – коэффициентами корреляции i -го и j -го спектров.

Этап 6. Проведение кластеризации по полученной матрице. Кластеризуются объекты со схожими спектральными кривыми, а значит, и физическими свойствами. Для кластеризации используется иерархический агломеративный метод [11]. Преимущество иерархических методов кластеризации заключается в их наглядности. Важным параметром кластеризации является выбор метрики связывания кластеров (*linkage*) и метрики определения степени сходства спектров (*distance*).

Этап 7. Построение диаграммы (визуализация) распределения спектров (описание данного этапа см. ниже).

Этап 8. Оценка результатов кластеризации и принятие решения о необходимости дальнейших исследований. Это позволяет провести классификацию и использовать полученные результаты для разграничения разных категорий здоровья. По расположению спектров на диаграмме можно сделать вывод о принадлежности неизвестного спектра к определенной категории здоровья елей.

Корректировка спектров с учетом первой главной компоненты (см. рис. 3, а). Метод главных компонент [12] преобразовывает систему зависимых между собой значений *длина волны – коэффициент отражения* к системе статистически независимых главных компонент, на которые со своими весовыми коэффициентами можно разложить каждый из исходных спектров. Можно предположить, что каждая главная компонента является отображением независимых свойств снимаемых объектов в исследуемой спектральной области. Первая главная компонента вносит максимальный вклад в объяснение дисперсии выборки.

Матрица главных компонент \mathbf{Z} строится на основе матрицы весовых коэффициентов $\mathbf{A} = \{a_{ij}\}, i = \overline{1, N}, j = \overline{1, N}$, учитывающих тесноту связи между исходными признаками \mathbf{X} (нормированными спектрами) и главными компонентами:

$$\mathbf{Z} = \mathbf{X}\mathbf{A}, \quad (1)$$

где матрица $\mathbf{X} = \{x_{ij}\}, i = \overline{1, K}, j = \overline{1, N}$ (K – количество длин волн; N – количество спектров). Элементы матрицы \mathbf{A} рассчитываются с использованием аппарата матричной алгебры.

Первый способ корректировки спектров состоит в том, что все спектры нормируются на первую главную компоненту. Происходит это следующим образом. Находится спектр первой главной компоненты Z_1 . Он нормируется от 0 до 1 ($Z_{1\text{norm}}$). Модифицированные спектры x' рассчитываются по формуле

$$x'_{ij} = x_{ij} (1 - Z_{1\text{norm}_i}),$$

где x_{ij} – значение в i -м канале j -го спектра; $Z_{1\text{norm}_i}$ – значения нормированной первой главной компоненты в i -м канале.

Второй способ корректировки спектров заключается в занулении столбца первой главной компоненты в матрице \mathbf{Z} и восстановлении по ней модифицированных спектров:

$$\mathbf{X}' = \mathbf{Z}'\mathbf{A}^{-1}, \quad (2)$$

где \mathbf{X}' – модифицированная матрица спектров; \mathbf{Z}' – модифицированная матрица главных компонент.

Построение диаграммы распределения спектров (см. рис. 3, б). По коэффициентам корреляции (1) определяются три наиболее различающихся спектра. Сначала в выражении (1) находится минимальный элемент матрицы $r_{s,t}$, s -й и t -й спектры являются двумя из трех искомого спектров. Для выбора третьего спектра вводится функция

$$A = r_{k,s} + r_{k,t}, \quad k = \overline{1, N}.$$

Номером третьего искомого спектра будет значение индекса k , при котором функция достигает своего минимального значения. Три найденных таким образом спектра используются для формирования трех осей прямоугольной системы координат. Для отображения в этой трехмерной системе координат точек, соответствующих всем остальным спектрам выборки, вдоль каждой из осей откладываются значения коэффициента корреляции текущего спектра с соответствующим осевым спектром. Таким образом, матрица взаимных корреляций определяет координаты спектров в указанной трехмерной системе координат. На расположение спектра влияет степень его сходства с тремя спектрами, выбранными для формирования осей.

Разработанное программное обеспечение. Алгоритм классификации по спектральным данным был реализован на языке Matlab в виде программного обеспечения, позволяющего пользователю проводить анализ спектральных данных с помощью удобного графического интерфейса (рис. 4). Кроме разработанного алгоритма классификации в программном обеспечении имеется и другой функционал. В частности, вместо представления данных в осях коэффициентов корреляции добавлено представление спектров в проекционном пространстве первых трех главных компонент. Такое преобразование доступно как для исходных нормированных спектров, так и для модифицированных с помощью устранения влияния первой главной компоненты спектров. Стоит отметить, что главные компоненты в разработанном алгоритме и главные компоненты в данном проекционном пространстве – разные понятия. Разложение на главные компоненты в обоих случаях выполняется по одной и той же матрице спектров, но в одном из случаев она транспонированная, так как имеются принципиальные отличия в том,



Рис. 4. Пользовательский интерфейс разработанного программного обеспечения для анализа спектральных данных
 Fig. 4. User interface of the developed software for spectral data analysis

что именно (строки или столбцы) отвечает за признаки, а что отвечает за объекты. В зависимости от выбранных настроек классификация будет выполняться по разным критериям. При выборе системы координат, основанной на коэффициентах корреляции, классификация будет осуществляться по алгоритму, представленному на рис. 2. При выборе системы координат, основанной на первых трех главных компонентах, классификация будет выполняться по проекциям спектров на три главные компоненты. Кроме того, в разработанном программном обеспечении добавлена возможность проведения «чистой» классификации, которая осуществляется только по нормированным (этап 2 алгоритма) спектральным кривым, прошедшим или не прошедшим устранение влияния первой главной компоненты. Также реализован расчет вегетационного индекса NDVI [13].

Результаты и их обсуждение

Исходная выборка включала 356 спектров отражения еловой хвои в диапазоне 400–870 нм, каждому из которых была присвоена одна из четырех категорий здоровья. Если представить всю спектральную выборку в виде проекций на главные компоненты, то можно увидеть два выраженных кластера (рис. 5). После анализа данных было найдено объяснение полученной кластерной структуры: часть спектров зарегистрированы с использованием спектрометра в режиме сканирования, а остальные спектры – с применением спектрометра в другом, более быстром, режиме. Однако этот более быстрый режим работы спектрометра предполагает получение двух спектров в смежных диапазонах и их последующую сшивку. Сшивку спектров происходит неидеально и оставляет после себя порог в месте сшивки. Величина этого порога составляет примерно 7 % от максимального значения спектра. Из-за этой особенности спектров на рис. 5 произошла классификация спектральной выборки по методу получения спектров.

С учетом сказанного для дальнейшего анализа были выбраны 212 спектров, полученных спектрометром в режиме сканирования. Стоит отметить, что среди спектров 3-й категории здоровья оказались 7 спектров, которые предположительно (как показали исследования по классификации) принадлежат к трем другим категориям здоровья елей. Вполне возможно, что данные образцы были ошибочно классифицированы лесопатологом ввиду своей сильной схожести с другими классами, поэтому для уменьшения неопределенности классификации указанные 7 образцов 3-й категории здоровья исключены из рассматриваемой выборки. Также были исключены 2 спектра коры еловых ветвей. На рис. 6 представлены визуализация спектров хвои в осях коэффициентов корреляции после устранения влияния первой главной компоненты (см. рис. 6, а) и визуализация спектров хвои в осях главных компонент (см. рис. 6, б). Основные характеристики спектральной выборки (среднеквадратичное отклонение (СКО), средние спектры по классам) представлены на рис. 7.

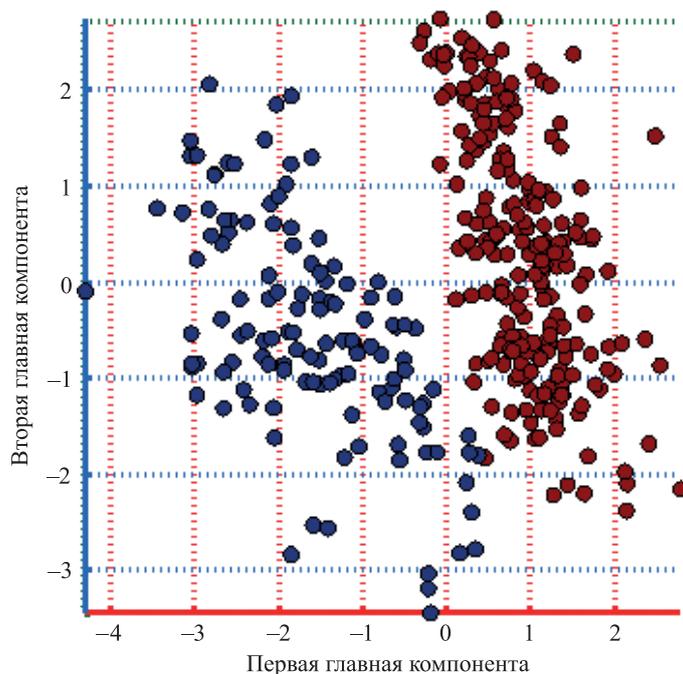


Рис. 5. Первичная классификация всей лабораторной спектральной выборки. Здесь и далее разные цвета соответствуют разным автоматически выделенным классам

Fig. 5. Primary classification of the entire laboratory spectral sample. Here and below different colours correspond to different automatically defined classes

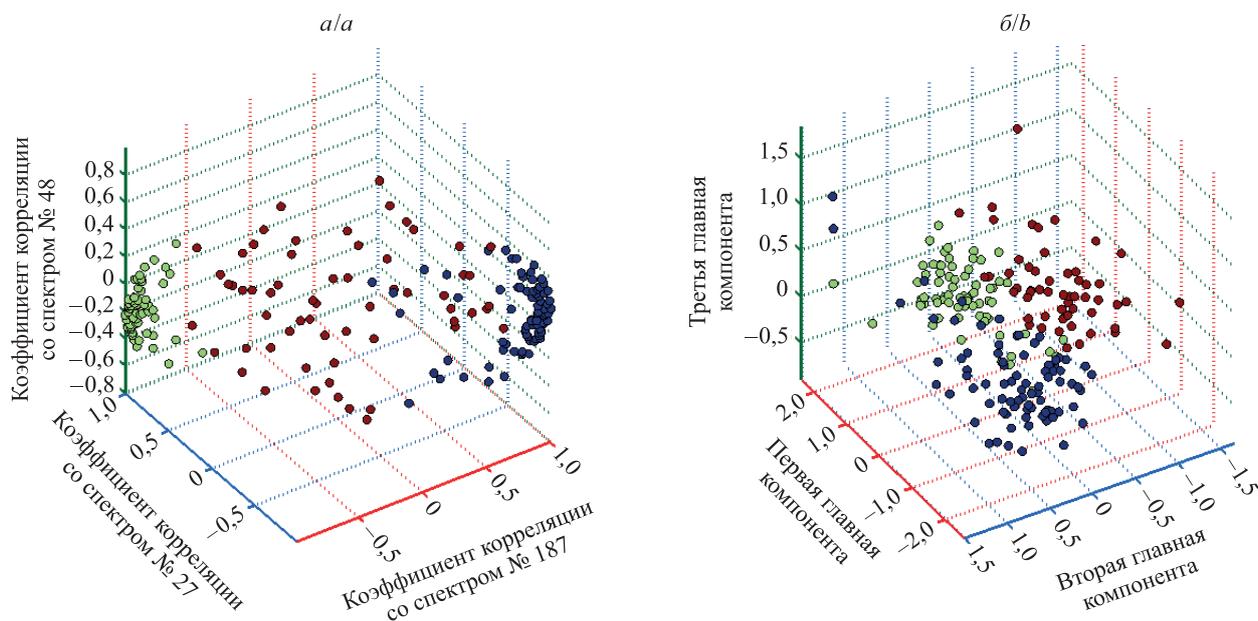


Рис. 6. Спектры хвои в осях коэффициентов корреляции с выполненным устранением влияния первой главной компоненты (а) и в осях главных компонент (б)

Fig. 6. Spectra of needles in the axes of the correlation coefficients with the elimination of the influence of first principal component (a) and in the axes of principal components (b)

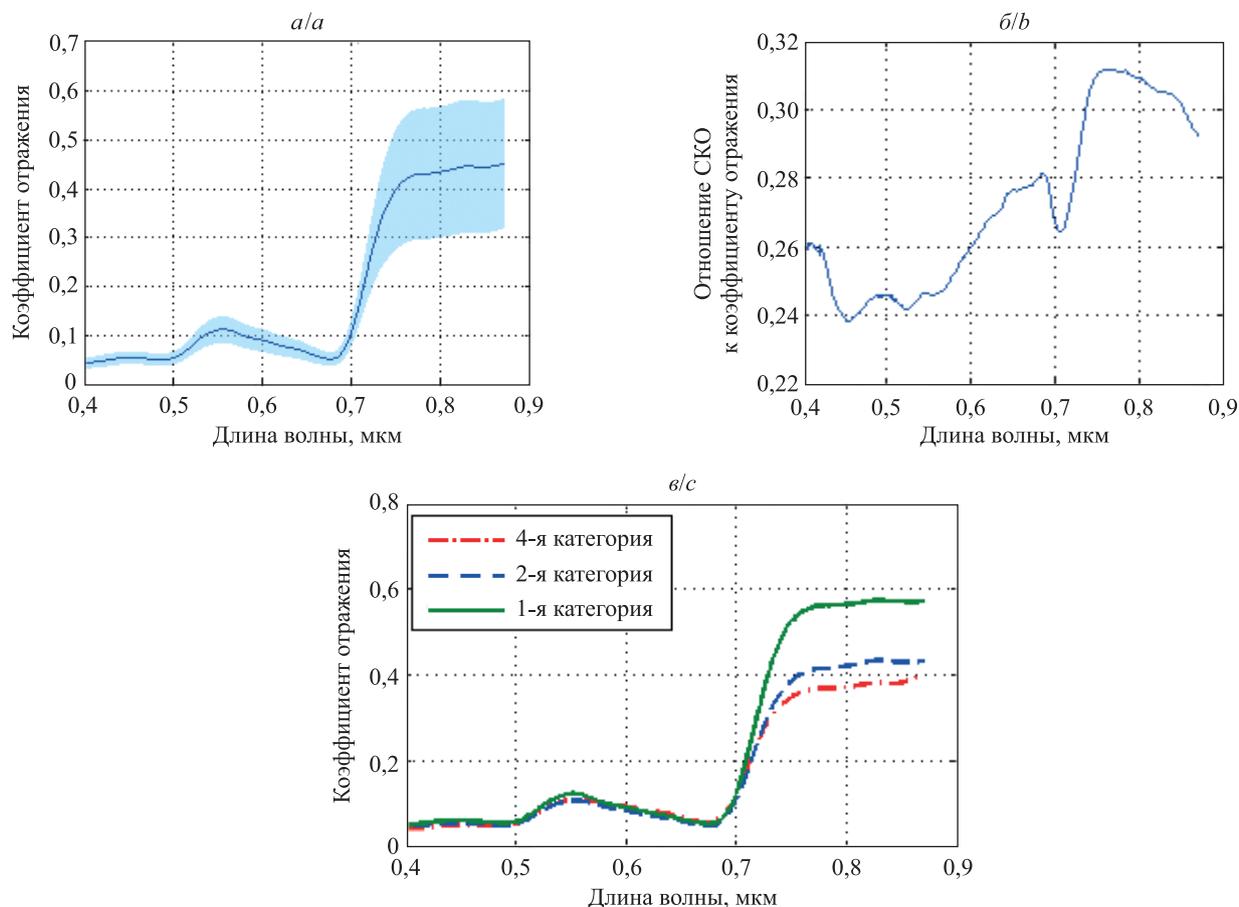


Рис. 7. График среднего спектра коэффициентов отражения 203 типичных образцов хвои (голубая область характеризует СКО) (а), график отношения СКО к коэффициентам отражения среднего спектра всех образцов (б) и графики средних спектров отражения трех категорий здоровья елей (в)

Fig. 7. Graph of the average spectrum of reflectances of 203 typical samples of needles (the blue area characterises the standard deviation) (a), graph of the ratio of the standard deviation to the reflection coefficients of the average spectrum of all samples (b) and graphs of mean reflectance spectra of three health categories of spruces (c)

По значениям СКО на рис. 7, а, и отношению СКО к коэффициентам отражения на рис. 7, б, можно сделать вывод, что максимальная изменчивость исследуемых спектров наблюдается в ближнем инфракрасном (ИК) диапазоне (0,75–0,87 мкм). В видимом же диапазоне наибольшее значение СКО соответствует зеленой области, а именно 0,55 мкм. Средние спектры разных категорий здоровья (см. рис. 7, в) демонстрируют максимальное различие в ближнем ИК-диапазоне, однако в этом же диапазоне каждая из категорий здоровья имеет наибольшее СКО (около 0,10–0,15). Само различие между значениями средних спектров в ближнем ИК-диапазоне принимает значение от 0,05 (для 2-й и 4-й категорий) до 0,17 (для 1-й и 4-й категорий), или 12 и 34 % соответственно. На основании этого можно предположить, что классы спектров хвои, определенные путем визуального осмотра, являются частично пересекающимися и трудноразличимыми.

Результаты классификации. Для оценки точности разбиения на классы предложено использовать численные характеристики качества классификации. Для каждого класса строится матрица ошибок, с помощью которой можно определить число предсказаний классификатора *true positive* (TP), *true negative* (TN), *false positive* (FP) и *false negative* (FN) [14]. На основе этих показателей рассчитываются следующие основные метрики оценки качества классификации: *accuracy*, *precision*, *recall* и *F-score*. Данные метрики принимают значения от 0 до 1 (или от 0 до 100 %). Наиболее полно описывающей качество классификации считается метрика *F-score*, которая представляет собой гармоническое среднее между метриками *precision* и *recall*.

Приведем комбинации отдельных компонент разработанного алгоритма с известными методами классификации [11–13], для которых выполнялись сравнения:

- PCA – do nothing – ward – euclidian;
- PCA – do nothing – ward – seuclidian;
- PCA – pure – do nothing – ward – euclidian;

- PCA – pure – do nothing – ward – seuclidian;
- PCA – pure – without 1st – ward – seuclidian – 700–869.5 nm;
- Corr – do nothing – ward – seuclidian;
- Corr – do nothing – ward – cityblock;
- Corr – do nothing – ward – correlation;
- Corr – without 1st – ward – seuclidian;
- Corr – without 1st – ward – cityblock;
- Corr – without 1st – ward – chebyshev;
- Corr – without 1st – ward – cosine;
- Corr – without 1st – ward – correlation;
- Corr – without 1st – ward – spearman;
- Corr – pure – without 1st – ward – euclidian;
- Corr – pure – without 1st – ward – seuclidian;
- Corr (only 3 values) – without 1st – ward – euclidian (классификация проводилась только по трем коэффициентам корреляции с осевыми спектрами);
- Index NDVI – ward – euclidian.

Для простоты восприятия используемые компоненты, метрики и методы включены в название. Метки «PCA» и «Corr» отвечают выбранному способу представления системы координат (главные компоненты и корреляции соответственно), а метки «do nothing» и «without 1st» показывают, устранено ли влияние первой главной компоненты (по формуле (2)) у спектров выборки. Метка «pure» означает, что была включена «чистая» классификация. Две последние метки отражают метрику связывания кластеров и метрику определения степени сходства спектров. Указанный диапазон длин волн или вегетационный индекс означает, что спектральный диапазон соответствующе изменен или что использована классификация по значению соответствующего индекса для каждого спектра.

Рассчитанные метрики для оценки точности результатов классификации с разными вариантами настроек представлены в таблице. Названия численных характеристик качества классификации указаны на английском языке с целью избежать искажения смысла при переводе общепринятых (в том числе и в русскоязычной литературе) англоязычных терминов.

Оценка точности классификации
Estimation of classification accuracy

Категория здоровья	Показатели качества классификации							
	TP	TN	FP	FN	Accuracy, %	Precision, %	Recall, %	F-score, %
PCA – do nothing – ward – euclidian								
1	42	130	21	8	85,57	66,67	84,00	74,34
2	44	105	28	24	74,13	61,11	64,71	62,86
4	62	114	4	21	87,56	93,94	74,70	83,22
PCA – do nothing – ward – seuclidian								
1	34	143	8	16	88,06	80,95	68,00	73,91
2	52	101	32	16	76,12	61,90	76,47	68,42
4	66	109	9	17	87,06	88,00	79,52	83,54
PCA – pure – do nothing – ward – euclidian								
1	42	133	18	8	87,06	70,00	84,00	76,36
2	48	106	27	20	76,62	64,00	70,59	67,13
4	62	114	4	21	87,56	93,94	74,70	83,22
PCA – pure – do nothing – ward – seuclidian								
1	49	123	28	1	85,57	63,64	98,00	77,17
2	37	115	18	31	75,62	67,27	54,41	60,16
4	64	113	5	19	88,06	92,75	77,11	84,21
PCA – pure – without 1 st – ward – seuclidian – 700–869.5 nm								
1	47	130	21	3	88,06	69,12	94,00	79,66
2	42	124	9	26	82,59	82,35	61,76	70,59
4	67	103	15	16	84,58	81,71	80,72	81,21

Продолжение таблицы
Continuation of the table

Категория здоровья	Показатели качества классификации							
	TP	TN	FP	FN	Accuracy, %	Precision, %	Recall, %	F-score, %
Corr – do nothing – ward – seuclidian								
1	40	134	17	10	86,57	70,18	80,00	74,77
2	48	106	27	20	76,62	64,00	70,59	67,13
4	64	113	5	19	88,06	92,75	77,11	84,21
Corr – do nothing – ward – cityblock								
1	41	132	19	9	86,07	68,33	82,00	74,55
2	46	104	29	22	74,63	61,33	67,65	64,34
4	62	114	4	21	87,56	93,94	74,70	83,22
Corr – do nothing – ward – correlation								
1	46	123	28	4	84,08	62,16	92,00	74,19
2	27	118	15	41	72,14	64,29	39,71	49,09
4	69	102	16	14	85,07	81,18	83,13	82,14
Corr – without 1 st – ward – seuclidian								
1	46	98	53	4	71,64	46,46	92,00	61,74
2	8	114	19	60	60,70	29,63	11,76	16,84
4	64	107	11	19	85,07	85,33	77,11	81,01
Corr – without 1 st – ward – cityblock								
1	43	114	37	7	78,11	53,75	86,00	66,15
2	29	105	28	39	66,67	50,88	42,65	46,40
4	61	115	3	22	87,56	95,31	73,49	82,99
Corr – without 1 st – ward – chebyshev								
1	43	97	54	7	69,65	44,33	86,00	58,50
2	8	113	20	60	60,20	28,57	11,76	16,67
4	64	106	12	19	84,58	84,21	77,11	80,50
Corr – without 1 st – ward – cosine								
1	46	106	45	4	75,62	50,55	92,00	65,25
2	7	124	9	61	65,17	43,75	10,29	16,67
4	71	95	23	12	82,59	75,53	85,54	80,23
Corr – without 1 st – ward – correlation								
1	3	131	20	47	66,67	13,04	6,00	8,22
2	49	83	50	19	65,67	49,49	72,06	58,68
4	66	105	13	17	85,07	83,54	79,52	81,48
Corr – without 1 st – ward – spearman								
1	46	95	56	4	70,15	45,10	92,00	60,53
2	2	125	8	66	63,18	20,00	2,94	5,13
4	71	100	18	12	85,07	79,78	85,54	82,56
Corr – pure – without 1 st – ward – euclidian								
1	42	133	18	8	87,06	70,00	84,00	76,36
2	48	106	27	20	76,62	64,00	70,59	67,13
4	62	114	4	21	87,56	93,94	74,70	83,22

Категория здоровья	Показатели качества классификации							
	TP	TN	FP	FN	Accuracy, %	Precision, %	Recall, %	F-score, %
Corr – pure – without 1 st – ward – seuclidian								
1	48	123	28	2	85,07	63,16	96,00	76,19
2	37	115	18	31	75,62	67,27	54,41	60,16
4	64	112	6	19	87,56	91,43	77,11	83,66
Corr (only 3 values) – without 1 st – ward – euclidian								
1	45	112	39	5	78,11	53,57	90,00	67,16
2	25	108	25	43	66,17	50,00	36,76	42,37
4	62	113	5	21	87,06	92,54	74,70	82,67
Index NDVI – ward – euclidian								
1	24	134	17	26	78,61	58,54	48,00	52,75
2	44	103	30	24	73,13	59,46	64,71	61,97
4	74	106	12	9	89,55	86,05	89,16	87,57

По результатам оценки качества классификации наиболее успешной (по значениям показателя *F-score*) является комбинация PCA – pure – without 1st – ward – seuclidian – 700–869.5 nm. Устранение влияния первой главной компоненты у спектров, а также сужение спектрального интервала до области, в которой вариативность спектров выборки является максимальной, позволили достичь наибольших значений показателя *F-score*, в том числе для 2-й категории здоровья елей, что имеет важнейшее значение для практических задач детектирования лесопатологий. От точности определения начального этапа усыхания зависит скорость реагирования на зарождающийся очаг усыхания, а в случае, например, заражения жуком-короедом скорость реагирования чрезвычайно важна для предотвращения его распространения.

На рис. 8 показан результат классификации в осях главных компонент спектров, у которых выполнена предобработка по устранению влияния первой главной компоненты.

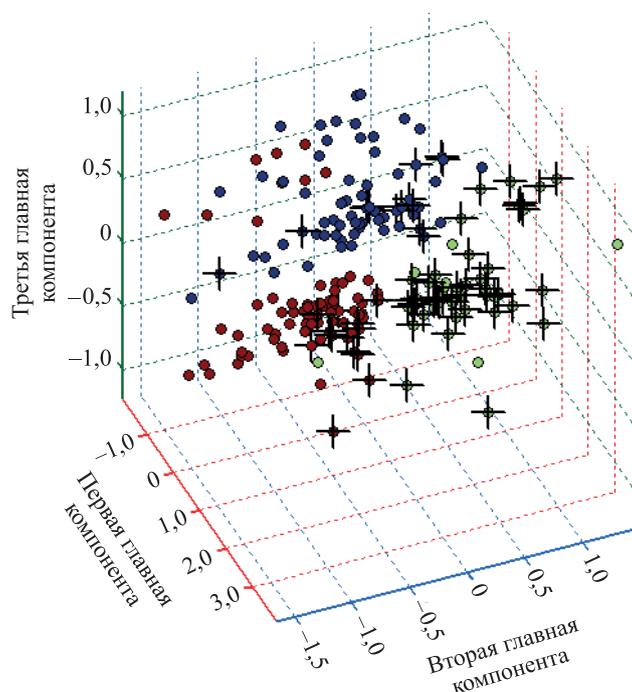


Рис. 8. Результат классификации PCA – pure – without 1st – ward – seuclidian – 700–869.5 nm. Синий цвет соответствует 1-й категории здоровья, зеленый – 2-й категории, красный – 4-й категории. Крестиками отмечены спектры, обозначенные лесопатологом как спектры елей 2-й категории здоровья
 Fig. 8. Result of the classification PCA – pure – without 1st – ward – seuclidian – 700–869.5 nm. Blue colour corresponds to the 1st health category, green – 2nd category, red – 4th category. The crosses mark the spectra marked by the forest pathologist as the spectra of the 2nd health category of spruces

Выбранные параметры классификации, в том числе сужение до ИК-диапазона спектрального интервала, позволили повысить точность модели классификации в определении всех трех категорий здоровья елей, а особенно 2-й категории (показатель F -score увеличился до 70,59 %), по сравнению с точностью, обеспечиваемой другими комбинациями параметров классификации. Показатель F -score определения 1-й и 4-й категорий здоровья превышает показатель F -score определения 2-й категории (79,66 и 81,21 % против 70,59 % соответственно). Такой результат объясняется явным различием 1-й и 4-й категорий, в то время как 2-я категория здоровья елей определяется уже более проблематично, а также возможностью ошибки интерпретации состояния здоровья лесопатологом.

Заключение

Разработан оригинальный алгоритм классификации спектральных данных, реализованный в виде программного обеспечения, которое дополнено известными методами классификации. С помощью этого программного обеспечения проведен анализ спектральных данных образцов хвои, выполнена классификация с разными комбинациями параметров и осуществлена оценка результатов классификации. На основе этой оценки определена такая комбинация параметров (проекция спектров в пространство главных компонент, устранение влияния первой главной компоненты на спектры, метрика связывания кластеров Уорда и стандартизированная евклидова метрика для вычисления спектрального расстояния), которая обеспечивает максимальное качество классификации среди исследованных комбинаций. Согласно оценке точности классификации добавление предложенного в работе метода по устранению влияния первой главной компоненты спектральной выборки позволило повысить показатель F -score для детектирования хвои 2-й категории здоровья с 60,16 до 70,59 %. Определение различия 1-й и 2-й категорий здоровья является очень важной задачей, так как идентификация деревьев с начальными признаками угнетения необходима для детектирования возникающих очагов усыхания еловых лесов на самых ранних стадиях. Анализ лабораторных данных с использованием наиболее эффективной комбинации параметров и методов классификации позволяет решить подобную задачу. Следующая задача, которая из нее вытекает, – анализ спектров, получаемых в процессе натуральных измерений с помощью беспилотного летательного аппарата.

Библиографические ссылки

1. Иванов ВП, Глазун ИН, Шелухо ВП, Смирнов СИ, Нартов ДИ. Усыхание еловых лесов – проблема регионов. В: *Лес XXI века. Тезисы докладов Международной практической конференции; 20–24 октября 2005 г.; Брянск, Россия*. Брянск: Брянский государственный технический университет; 2005. с. 42.
2. Манько ЮИ, Гладкова ГА. Массовое усыхание пихтово-еловых лесов на российском Дальнем Востоке: основные итоги изучения. *Комаровские чтения*. 2003;49:131–171.
3. Звягинцев ВБ, Блинцов АИ, Козел АВ, Кухта ВН, Сазонов АА, Середич МО и др. *Защита леса*. Минск: Белорусский государственный технологический университет; 2019. 164 с.
4. Safonova A, Tabik S, Alcaraz-Segura D, Rubtsov A, Maglinets Yu, Herrera F. Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*. 2019;11(6):643. DOI: 10.3390/rs11060643.
5. Санников ПЮ, Андреев ДН, Бузмаков СА. Выявление и анализ сухостоя при помощи беспилотного летательного аппарата. *Современные проблемы дистанционного зондирования Земли из космоса*. 2018;15(3):103–113. DOI: 10.21046/2070-7401-2018-15-3-103-113.
6. Девятова НВ, Ершов ДВ. Съемка MODIS/TERRA в мониторинге вспышек массового размножения очагов насекомых-вредителей. *Современные проблемы дистанционного зондирования Земли из космоса*. 2005;2(2):262–266.
7. White JC, Wulder MA, Brooks D, Reich R, Wheate RD. Detection of red attack stage mountain pine beetle infestation with high spatial resolution satellite imagery. *Remote Sensing of Environment*. 2005;96(3–4):340–351. DOI: 10.1016/j.rse.2005.03.007.
8. Waser LT, Kuchler M, Jütte K, Stampfer T. Evaluating the potential of WorldView-2 data to classify tree species and different levels of ash mortality. *Remote Sensing*. 2014;6(5):4515–4545. DOI: 10.3390/rs6054515.
9. Bruchkouski II, Siliuk VA, Litvinovich HS, Lamaka AA, Stanchuk VV, Guliaeva SI. Goniophotometer for measurements of spectral reflectance coefficients and transmission spectra. *Journal of Applied Spectroscopy*. 2021;88(2):363–369. DOI: 10.1007/s10812-021-01183-7.
10. Розова СС. *Классификационная проблема в современной науке*. Сычева ЛС, редактор. Новосибирск: Наука; 1986. 224 с.
11. Лагутин БМ. *Наглядная математическая статистика*. Москва: Бином. Лаборатория знаний; 2007. 472 с.
12. Jolliffe IT. *Principal component analysis*. 2nd edition. New York: Springer-Verlag; 2002. XXIX, 488 p. (Springer series in statistics). DOI: 10.1007/b98835.
13. Spruce JP, Hicke JA, Hargrove WW, Grulke NE, Meddens AJH. Use of MODIS NDVI products to map tree mortality levels in forests affected by mountain pine beetle outbreaks. *Forests*. 2019;10(9):811. DOI: 10.3390/f10090811.
14. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861–874. DOI: 10.1016/j.patrec.2005.10.010.

References

1. Ivanov VP, Glazun IN, Shelukho VP, Smirnov SI, Nartov DI. [Drying of spruce forests – a problem of the regions]. In: *Les XXI veka. Tezisy dokladov Mezhdunarodnoi prakticheskoi konferentsii; 20–24 oktyabrya 2005 g.; Bryansk, Rossiya* [Forest of the 21st century. Abstracts of the International practical conference; 2005 October 20–24; Bryansk, Russia]. Bryansk: Bryansk State Technical University; 2005. p. 42. Russian.

2. Man'ko YuI, Gladkova GA. [Mass drying out of fir-spruce forests in the Russian Far East: the main results of the study]. *V. L. Komarov Memorial Lectures*. 2003;49:131–171. Russian.
3. Zvyagintsev VB, Blintsov AI, Kozel AV, Kukhta VN, Sazonov AA, Seredich MO, et al. *Zashchita lesa* [Forest protection]. Minsk: Belarusian State Technological University; 2019. 164 p. Russian.
4. Safonova A, Tabik S, Alcaraz-Segura D, Rubtsov A, Maglinets Yu, Herrera F. Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*. 2019;11(6):643. DOI: 10.3390/rs11060643.
5. Sannikov PYu, Andreev DN, Buzmakov SA. Identification and analysis of deadwood using an unmanned aerial vehicle. *Current Problems in Remote Sensing of the Earth from Space*. 2018;15(3):103–113. Russian. DOI: 10.21046/2070-7401-2018-15-3-103-113.
6. Devyatova NV, Ershov DV. [MODIS/TERRA survey in monitoring outbreaks of mass reproduction of foci of insect pests]. *Current Problems in Remote Sensing of the Earth from Space*. 2005;2(2):262–266. Russian.
7. White JC, Wulder MA, Brooks D, Reich R, Wheate RD. Detection of red attack stage mountain pine beetle infestation with high spatial resolution satellite imagery. *Remote Sensing of Environment*. 2005;96(3–4):340–351. DOI: 10.1016/j.rse.2005.03.007.
8. Waser LT, Küchler M, Jütte K, Stampfer T. Evaluating the potential of WorldView-2 data to classify tree species and different levels of ash mortality. *Remote Sensing*. 2014;6(5):4515–4545. DOI: 10.3390/rs6054515.
9. Bruchkouski II, Siliuk VA, Litvinovich HS, Lamaka AA, Stanchuk VV, Guliaeva SI. Goniophotometer for measurements of spectral reflectance coefficients and transmission spectra. *Journal of Applied Spectroscopy*. 2021;88(2):363–369. DOI: 10.1007/s10812-021-01183-7.
10. Rozova SS. *Klassifikatsionnaya problema v sovremennoi nauke* [Classification problem in modern science]. Sycheva LS, editor. Novosibirsk: Nauka; 1986. 224 p. Russian.
11. Lagutin BM. *Naglyadnaya matematicheskaya statistika* [Visual mathematical statistics]. Moscow: Binom. Laboratoriya znaniy; 2007. 472 p. Russian.
12. Jolliffe IT. *Principal component analysis*. 2nd edition. New York: Springer-Verlag; 2002. XXIX, 488 p. (Springer series in statistics). DOI: 10.1007/b98835.
13. Spruce JP, Hicke JA, Hargrove WW, Grulke NE, Meddens AJH. Use of MODIS NDVI products to map tree mortality levels in forests affected by mountain pine beetle outbreaks. *Forests*. 2019;10(9):811. DOI: 10.3390/f10090811.
14. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861–874. DOI: 10.1016/j.patrec.2005.10.010.

Получена 04.05.2022 / исправлена 24.05.2022 / принята 24.05.2022.
Received 04.05.2022 / revised 24.05.2022 / accepted 24.05.2022.