

Коросов А.В.¹, Платонова Е.А.²

ВЫЯВЛЕНИЕ ФИТОХОР С ОБУЧЕНИЕМ НА СМЕШАННЫХ ПОЛЕВЫХ И КОСМИЧЕСКИХ ДАННЫХ (НА ПРИМЕРЕ ПРИРОДНОЙ ТЕРРИТОРИИ БОТАНИЧЕСКОГО САДА ПетрГУ)

АННОТАЦИЯ

В работе предлагается методика совместной количественной обработки космических и полевых данных с целью выполнить тематическую классификацию пикселей спутникового снимка. В традиционном алгоритме дешифрирования классификация осуществляется отдельно от полевой верификации полученных кластеров. Предлагаемый алгоритм объединяет эти процедуры в рамках компонентного анализа. Поставлена задача выявить фитохор в растительном покрове природной территории Ботанического сада Петрозаводского государственного университета. Анализировали показатели 157 геоботанических описаний лесной растительности этой территории, в том числе: количество видов, число экземпляров деревьев четырех пород (сосны обыкновенной, ели финской, осины, березы повислой и пушистой), индекс теплообеспечения, индекс затененности и значения яркости космоснимков. Все показатели имеют количественную форму. Алгоритм анализа включает этап совмещения геоботанических описаний и яркостных характеристик тех пикселей снимков, которые покрывают площадки описаний. Совместный компонентный анализ значений яркости и геоботанических характеристик позволяет оценить вклад этих показателей в главные компоненты и выбрать по величине нагрузок, выбрать вариант фитохоры, идентифицируемый данной компонентой. Используя полученные нагрузки, рассчитываются значения компонент для всей территории. Отбор точек, соответствующих выделенным фитохорам позволяет определить узкий диапазон яркости пикселей и выполнить отбор пикселей, соответствующих искомым участкам. Фитохора «монодоминантные зеленомошные, преимущественно бруснично-черничные сосняки» занимает около 17 % из всего диапазона значений компоненты. С помощью процедуры классификации с обучением на основе метода главных компонент для территории Ботанического сада ПетрГУ получена контурная карта распространения данного типа сосняков.

КЛЮЧЕВЫЕ СЛОВА: космоснимок, классификация с обучением, компонентный анализ, фитохоры, ботанический сад ПетрГУ.

¹ Петрозаводский государственный университет, пр. Ленина, д. 33, 185940, Петрозаводск, Россия, *e-mail:* korosov@psu.karelia.ru

² Петрозаводский государственный университет, пр. Ленина, д. 33, 185940, Петрозаводск, Россия, *e-mail:* meles@sampo.ru

Andrey V. Korosov¹, Elena A. Platonova²

IDENTIFICATION OF PHYTOHORES WITH TRAINING ON MIXED FIELD AND SPACE DATA (USING THE EXAMPLE OF THE NATURAL TERRITORY OF THE BOTANICAL GARDEN OF PetrSU)

ABSTRACT

Joint quantitative processing of space and field data method has been proposed in order to perform thematic classification of satellite image pixels. Traditional decoding algorithm classification is carried out separately from the field verification of obtained clusters. The proposed algorithm combines these procedures in component analysis framework. The task is to identify phytohor in the vegetation cover of Petrozavodsk state University Botanical garden natural territory. The parameters of 157 geobotanical descriptions of forest vegetation of this territory were analyzed: the number of species, the number of specimens of four tree species (*Pinus sylvestris*, *Picea fennica*, *Populus tremula*, *Betula pubescens* and *Betula pendula*), heat supply index, shade index and brightness values of satellite images. Given parameters has quantitative form. The analysis algorithm includes the step of combining geobotanical descriptions and brightness characteristics of the image pixels that cover the description area. Jointed component analysis of brightness value and geobotanical characteristics allows evaluating the contribution of these parameters to the main components. By the size of the loads, phytohor variant is selected according to certain component. Using the resulting loads, component values are calculated for the entire territory. Point selection corresponding to the selected phytohors is allows to define a narrow range of pixel brightness and perform pixels selection corresponding to the desired areas. Phytohor "monodominant green-skinned, mainly cowberry-blueberry pines" occupies about 17 % of the entire range of component values. Using the classification procedure with training on the basis of the principal components method for the territory of PetrSU Botanical garden, a contour map of pine tree type distribution was obtained.

KEYWORDS: cosmonaut, classification with training, component analysis, phytochores, botanical garden PetrSU.

ВВЕДЕНИЕ

Классификация пикселей космоснимка – типичная процедура дистанционного зондирования, для которой разработано множество эффективных алгоритмов [http://grass-wiki.osgeo.org/wiki/Image_classification]. Классификация с обучением предполагает, что у исследователя есть представление о том, каким изучаемым объектам соответствуют определенные области снимка. Эти области выделяют в качестве «эталонов», и все прочие пиксели снимка классифицируются по степени близости к эталонным сигнатурам [<http://gis-lab.info/qa/grass-ismap.html>]. Так получают одноканальную растровую карту с областями. После такого предварительного районирования следует процесс ее верификации – сбор полевых данных и проверка правильности отнесения природных объектов к той или иной группе на рассчитанном растре.

В описанной процедуре есть два момента, оптимизация которых, на наш взгляд, вполне актуальна и должна приветствоваться. Во-первых, процесс назначения эталонных участков есть акт квантирования непрерывной информации, поскольку и природные

¹ Petrozavodsk State University, Lenina str., 33, 185940, Petrozavodsk, Russia, e-mail: korosov@psu.karelia.ru

² Petrozavodsk State University, Lenina str., 33, 185940, Petrozavodsk, Russia, e-mail: meles@sampo.ru

объекты и их отображение на снимке есть континуумы (особенно это касается лесных территорий). Создавая этот артефакт, мы неизбежно теряем и искажаем часть информации, совершая ошибки и первого, и второго рода. По этой причине естественным было бы уйти от назначения эталонов и непосредственно включить имеющуюся априорную (полевую) информацию в процесс классификации. Эта идея воплощена в новых методах классификации, таких как байесовский метод и максент [Phillips, 2009]. Во-вторых, если полевая информация будет сразу же включена в расчеты, отпадает этап полевой верификации, пожалуй, самый трудоемкий и неблагодарный из всех шагов районирования. Его можно будет заменить расчетными процедурами, когда для первичной классификации используется один массив данных, а для его проверки – другой.

Опираясь на эти соображения, мы предлагаем (и в настоящем сообщении рассматриваем) процедуру совместной обработки космических и полевых данных (с помощью метода главных компонент) как способа классификации пикселей космоснимка. Поставлена задача выявления фитохор в растительном покрове Ботанического сада Петрозаводского госуниверситета. Эта территория является удобным полигоном для апробации методики, поскольку досконально изучена в геоботаническом отношении. В силу ограниченности объема публикации рассмотрена задача вычленения из растительного покрова одного типа фитохоры (в размерах таксационного выдела или фитоценоза).

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Природная территория Ботанического сада площадью более 630 га располагается в денудационно-тектоническом грядовом ландшафте, который характеризуется чередованием вытянутых в северо-западном направлении гряд, сложенных вулканогенно-осадочными породами протерозоя. Маломощные морены последнего верхневалдайского оледенения залегают прерывистым чехлом и местами перекрыты песчано-гравийными озерно-ледниковыми и озерными отложениями. Растительность территории представлена в основном лесами (80 %), в которых значительную долю составляют сосняки, меньшие площади занимают еловые и сосново-еловые леса. Геоботанические описания участков леса на исследуемой территории выполнены в 1996–1997 гг. [Платонова, Лантратова, 2001] (рис. 1).

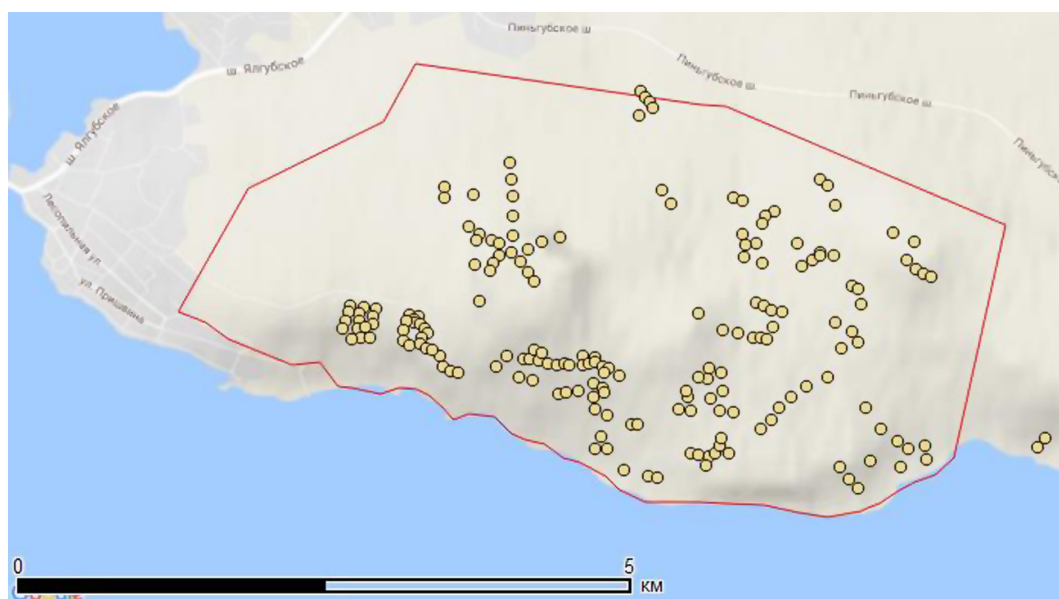


Рис. 1. Размещение точек описания на территории Ботанического сада ПетрГУ
Fig. 1. Placement of the sample plots on the territory of the Botanical Garden of PetrSU

Мы использовали следующие показатели по каждой пробной площадке (размером 0.01 га), полученные традиционными методами [Программа и методика..., 1974]:

- видовое богатство на участке (число видов, $bdiv$);
- число экземпляров деревьев четырех пород – сосны (**psy**), ели (**pab**), осины (**ptr**), березы повислой (**bpu**) и пушистой (**bpe**) для двух ярусов (древесного, кустарникового) (А, В);
- индекс теплообеспечения (**Tm**) (рассчитывался с использованием шкал Д.Н. Цыганова [1983] как средневзвешенный балл с учетом обилия видов на пробной площади);
- индекс затененности (**Lc**) (рассчитывался с использованием шкал Д.Н. Цыганова [1983] как средневзвешенный балл с учетом обилия видов на пробной площади).

Всего в анализ включено 157 описаний по 13 показателям.

Космоснимки были взяты с сайта [<https://landsatlook.usgs.gov/viewer.html>] для безоблачных дней зимы (1993.03), весны (1995.04), лета (1993.06) и осени (1994.09). Использовались трехканальные снимки. Всего файл со значениями яркости содержал 12 столбцов (X1993.03_1...X1994.09_3).

Процедура получения отклассифицированного растра включает несколько основных этапов. Сначала выполняется отбор тех пикселей снимков, которых покрывают области с геоботаническим описанием. При этом составляется матрица, совмещающая яркостные характеристики снимка и геоботанические описания в тех же точках пространства. Затем выполняется компонентный анализ зависимости каналов снимков от геоботанических характеристик, и выбираются те компоненты, которые надежно идентифицируют разные варианты растительности, или фитохоры; компоненты получают свое геоботаническое название. На третьем этапе рассчитываются значения выбранных компонент для всех пикселей снимка. Может быть получен растровый слой, выражающий континуальное распространение значений компонент, связанных с вариантами фитохор. Поскольку компоненты однозначно идентифицированы, полученная карта сразу обретает геоботанический смысл.

Подготовка геоботанических данных выполнялась в среде MS Excel. Чтение яркостных характеристик снимков, построение окончательных карт выполнялось в среде QGIS. Компонентный анализ и расчет значений компонент для построения растрового слоя выполнялся в среде R.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Последовательно рассмотрим алгоритм анализа по этапам.

Совмещение геоботанических описаний и характеристик снимков выполнялось с помощью процедуры чтения значений яркости со снимка (Extract...) – в таблицу для заранее нанесенных точек отбора проб. Характеристики со слоев космоснимков приписываются этим точкам из тех пикселей, над которыми они находятся.

Компонентный анализ основан на коррелированности изучаемых характеристик и выражает степень их взаимозависимости в форме «факторных нагрузок» – коэффициентов из диапазона ± 1 (табл.). При этом формируются группы более тесно взаимосвязанных показателей, выраженных как отдельные главные компоненты. Общая структура корреляций как бы «нарезается» на блоки, на отдельные группы более тесно коррелирующих переменных.

Высокие значения нагрузок в одной компоненте говорят о тесной корреляции показателей. Доля от общей корреляции, учтенная данной компонентой, выражается значением дисперсии; самые важные корреляции учитываются первыми компонентами. В нашем случае первая главная компонента с дисперсией 2,6² (учитывает 25 % общей корреляции) включает практически только характеристики космоснимков, анализируя совпадение контрастов почти на всех слоях. С небольшой нагрузкой в компоненту входит первый ярус ели, что нам не интересно.

Факторные нагрузки (Loadings) и стандартные отклонения (S) в четырех первых компонентах (Comp.1...Comp.4) для показателей смешанной выборки
 Factor Loadings and Standard Deviations (S) in the first four principal components
 (Comp.1 ... Comp.4) for mixed sample

Нагрузки (Loadings)				
Показатель	Comp. 1	Comp. 2	Comp. 3	Comp. 4
X1993.03_1	-0,26	-0,12	0,05	0,27
X1993.03_2	-0,34	0,12	-0,10	-0,02
X1993.03_3	-0,32	0,21	-0,11	0,01
X1993.06_1	-0,28	-0,11	0,18	-0,24
X1993.06_2	-0,30	-0,06	0,07	-0,28
X1993.06_3	-0,12	-0,20	0,33	0,13
X1994.09_1	-0,33	-0,09	-0,11	0,01
X1994.09_2	-0,31	-0,16	-0,13	0,14
X1994.09_3	-0,17	-0,32	0,14	0,17
X1995.04_1	-0,12	-0,15	0,18	0,42
X1995.04_2	-0,34	0,14	0,00	-0,06
X1995.04_3	-0,32	0,22	0,04	-0,14
vid_nas	-0,14	-0,28	-0,10	-0,02
Apsy	0,00	0,40	0,05	-0,09
Bpsy	-0,07	0,22	0,10	0,37
Apab	0,12	-0,32	-0,16	-0,03
Bpab	0,06	-0,06	0,12	0,31
Aptr	-0,02	-0,23	-0,03	0,05
Bptr	0,02	-0,15	-0,14	-0,12
Abpe	0,02	-0,06	-0,01	0,17
Bbpe	0,00	0,04	-0,04	0,23
Abpu	0,01	-0,18	0,41	-0,26
Bbpu	-0,01	-0,06	0,49	-0,25
Tm	-0,14	-0,08	-0,44	-0,07
Lc	0,07	-0,36	-0,24	-0,22
S	2,60	1,67	1,36	1,32

Вторая компонента с дисперсией $1,7^2$ (12 %) имеет следующие переменные с максимальными факторными нагрузками: +0,4 доля сосны, -0,3 доля ели, -0,2 доля осины, -0,17 доля березы, -0,35 затененность; из показателей снимков высокую нагрузку несут осенние: -0,3 для 1994.09 и зимний +0,21 для 1993.03, что свидетельствует об их обратной коррелированности. Интерпретировать это можно таким образом: компонента характеризует изменчивость доли в древостое сосны, осины и березы, но с противоположными знаками – чем больше сосны, тем меньше осины и березы, и наоборот. Яркость зимних снимков с этим трендом совпадает, весенних – противоположна. Компонентный анализ использует факторные нагрузки как коэффициенты для расчета значений главных компонент. В нашем случае получается, что чем больше на площадке сосны (знак +) и меньше ели, осины, березы (знак -), тем большее значение индекса получит данное описание, т. е. тем правее на диаграмме расположится соответствующая ему точка.

Отбор точек, соответствующих соснякам, выполняется с помощью таблицы исходных данных, таблицы главных компонент и списка номеров проб. Расчеты показали, что в правой области распределения значений второй компоненты выделяется небольшая

группа с максимальными значениями PC2 > 2 (рис. 2). Анализ таблицы первичных данных показал, что для этих точек характерно доминирование сосны и повышенная освещенность, значит, именно они являются представителями искомым лесов. Это монодоминантные сосняки (на денудационных террасах южной экспозиции с близким залеганием и выходом на поверхность кристаллических пород) зеленомошные, преимущественно бруснично-черничные и вересковые, в небольшой доле черничные.

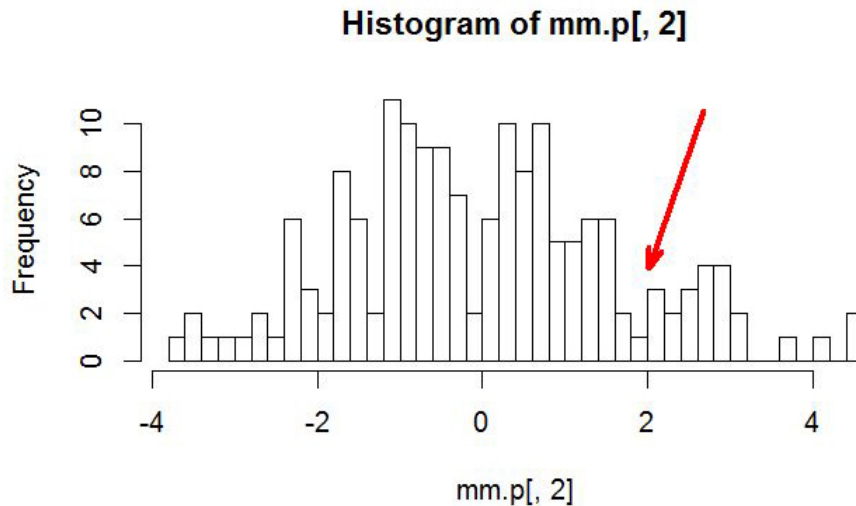


Рис. 2. Распределение значений второй главной компоненты
 Fig. 2. Distribution of values of the second main component

Определение диапазона яркости пикселей, соответствующих искомым фитохорам, выполняется в среде R. Для этого рассчитываем значения компонент, используя только нагрузки для показателей яркости (-0,12287941, 0,11977884, 0,20752337, -0,11143326, -0,05962177, -0,20098435, -0,09322922, -0,15653191, -0,31572369, -0,15108658, 0,13909984, 0,22394727). Зная номера проб, ранее отнесенных к сосновым лесам, нетрудно увидеть, что им соответствуют значения второй компоненты из диапазона -4...9 (рис. 3).

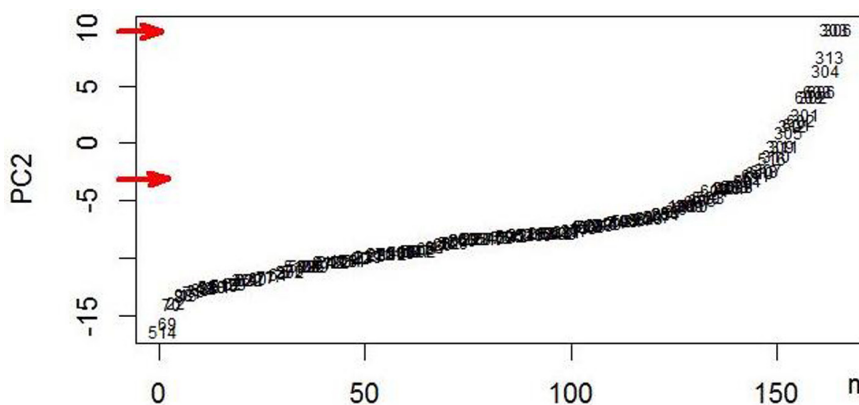


Рис. 3. Расположение проб из сосняков на оси второй главной компоненты
 Fig. 3. Location of samples from pine forests on the axis of the second main component

Отбор пикселей, соответствующих соснякам, на всем снимке. Для этого с помощью калькулятора растра создается новое растровое изображение, яркость пикселей на котором рассчитывается с использованием факторных нагрузок для всех каналов. Гистограмма показывает, что искомые фитохоры занимают на оси яркостей небольшой участок (рис. 4).

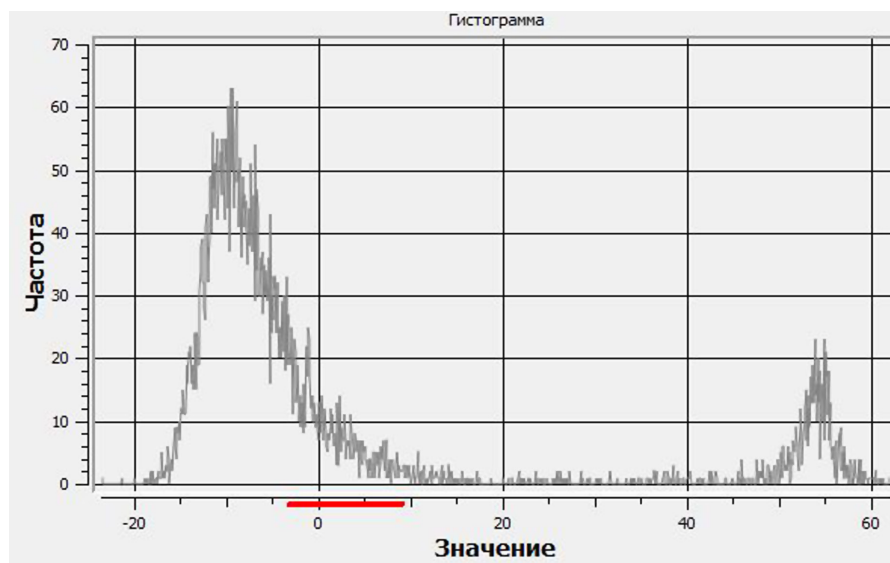


Рис. 4. Гистограмма для синтетического слоя второй компоненты
Fig. 4. Histogram for the synthetic layer of the second component

Далее с помощью калькулятора раstra из изображения вырезаются области пикселей, в которых соответствуют условию «rs1 3345@1» > - 4 and «rs1 3345@1» < 10; полученный слой со значениями 0 и 1 (пиксели из заданного диапазон), отображают сосняки (рис. 5).

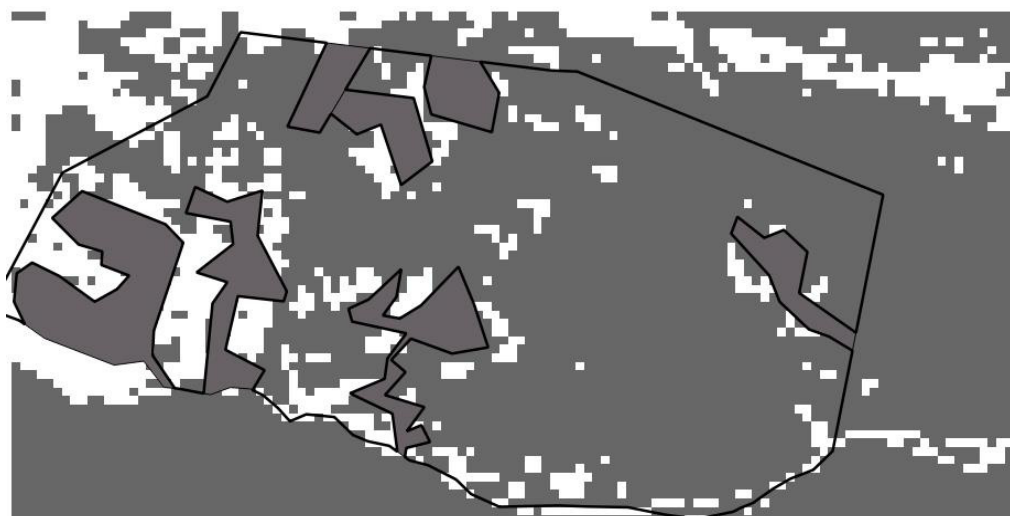


Рис. 5. Распространение выделенного варианта сосняков
в Ботаническом саду ПетрГУ (белая заливка)
Fig. 5. Distribution of identified pine phytohores
in the Botanical Garden of PetrSU (white filling)

Возможен другой вариант анализа – если привести значения яркостей к диапазону 0–1, то такой слой будет выражать метрику «вероятность быть сосняком искомого варианта для данного участка территории».

Проблема совмещения геоботанических описаний и характеристик снимков – не тривиальная задача, поскольку положение точки описания в общем смысле не совпадает с пикселем, имеющим те же координаты, в силу ошибок определения координат и точности привязки снимка. Следовательно, конкретному описанию фактически будет

соответствовать пиксель из некоторой области, которую, например, можно ограничить точностью 100 м, т. е. полем с размером 3*3 пикселя вокруг точки отбора. Это означает, что отдельное полевое описание нужно приписать не одному, а девяти ближайшим пикселям. Техническое решение таково: формируется сеть точек с шагом 30 м, из которых отбираются только те, что размещены вокруг точек отбора проб на расстоянии 50 м; атрибуты от точек описания передаются точкам гнезд. Характеристики со слоев космоснимков приписываются всем этим точкам.

Мы не рассматривали существенный этап – верификацию полученного геоизображения. Она может состоять в повторении рассмотренных процедур с *разными исходными наборами опорных точек* (пикселей, снабженных описаниями), получением серии растровых слоев и их обобщении. Вариантов обобщения может быть несколько (оценка средней, медианной или модальной яркости каждого пикселя...), в том числе расчет единственной главной компоненты по всем полученным картам, которая может рассматриваться как многомерный фильтр [Ефимов и др., 1988]. Более того, на основе исходных данных можно проводить разные варианты ресамплинга [Шитиков, Рзенберг, 2013] – выполнять множественные построения искомого растрового слоя с использованием только части информации, и получить на этой основе устойчивые статистические характеристики районирования.

К сожалению, это можно будет сделать, располагая более репрезентативным материалом, в частности, получив описания с других частей исследуемой территории. Размещение растительности на территории очень неравномерно, снимки получаются контрастными и самые сильные корреляции должны были бы проявляться по трендам – луг-лес. Пока описаний открытых территорий мало и поэтому указанный тренд точно оценить не удалось.

ВЫВОДЫ

С помощью процедуры классификации с обучением на основе метода главных компонент для территории Ботанического сада ПетрГУ получены два результата: традиционная контурная карта распространения одного из различных вариантов сосняков и континуальная карта распределения вероятности присутствия фитохоры данного типа. В целом полученная карта соответствует геоботанической карте, построенной в 2000 г. [<http://hortus.karelia.ru/>], поэтому представленный метод можно рассматривать как перспективный для районирования фитохор разного типа.

СПИСОК ЛИТЕРАТУРЫ

1. Ефимов В.М., Галактионов Ю.К., Шушпанова Н.Ф. Анализ и прогноз временных рядов методом главных компонент. Новосибирск: Наука, 1988. 70 с.
2. Платонова Е.А., Лантратова А.С. Разнообразие растительного покрова заповедной территории Ботанического сада ПетрГУ // Hortus botanicus. 2001. Т. 1. С. 42–51.
3. Программа и методика биогеоэкологических исследований / Отв. ред. Н.В. Дылис. М.: Наука, 1974. 403 с.
4. Цыганов Д.Н. Фитоиндикация экологических режимов в подзоне хвойно-широколиственных лесов. М., 1983. 196 с.
5. Шитиков В.К., Розенберг Г.С. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R. Тольятти: Кассандра, 2013. 314 с.
6. Phillips S.J. A Brief Tutorial on Maxent // Network of Conservation Educators and Practitioners. Center for Biodiversity and Conservation, American Museum of Natural History. Lessons in Conservation. 2009. V. 3. P. 108–135.

REFERENCES

1. *Cyganov D.N.* Phytoindication of ecological regimes in the subzone of coniferous-broadleaf forests. M., 1983. 196 p. (in Russian).
2. *Efimov V.M., Galaktionov Yu.K., Shushpanova N.F.* Analysis and prediction of time series by the principal component method. Novosibirsk: Nauka, 1988. 70 p. (in Russian).
3. *Phillips S.J.* A Brief Tutorial on Maxent, Network of Conservation Educators and Practitioners. Center for Biodiversity and Conservation, American Museum of Natural History. Lessons in Conservation. 2009. V. 3. P. 108–135 (in Russian).
4. *Platonova E.A., Lantratova A.S.* Variety of vegetative cover of the protected area of the Botanical Garden of PetrSU, Hortus botanicus. 2001. T. 1. P. 42–51 (in Russian).
5. Program and methodology of biogeocenological research. Otv. red. N.V. Dylip. M.: Nauka, 1974. 403 p. (in Russian).
6. *Shitikov V.K., Rozenberg G.S.* Randomization and bootstrap: statistical analysis in biology and ecology using R. Tol'yatti: Cassandra, 2013. 314 p. (in Russian).