

Развитие индексной парадигмы в дистанционном зондировании почвенно-растительного покрова

И. М. Михайленко, В. Н. Тимошин

*Агрофизический научно-исследовательский институт
Санкт-Петербург, 195220, Россия*

E-mails: ilya.mihailenko@yandex.ru, v.timoshin@yandex.ru

Цель работы — системный анализ и обобщение общепринятой индексной парадигмы использования данных дистанционного зондирования Земли (ДЗЗ) для оценивания состояния почвенно-растительного покрова. Установлено, что скалярная форма и отсутствие математической основы не позволяют использовать общепринятые вегетационные и подобные им индексы для оценивания векторов количественных показателей почвенно-растительного покрова. В то же время для принятия многих видов управляющих решений в земледелии актуально построение индексных изображений, отражающих такие неколичественные показатели, как виды культурных и сорных растений, наличие болезней растений, повреждений посевов и почв, физических и химических стрессов. По своему информационному содержанию оценивание таких неколичественных (счётных) состояний представляет собой процедуры распознавания образов или классов объектов почвенно-растительного комплекса. Субъективный эмпирический подход при выборе спектрального состава индексов и их комбинаций, используемый в настоящее время, не позволяет обеспечить достаточную надёжность таких процедур. Поэтому целью исследования в настоящей работе стала формализация процесса, благодаря чему можно полностью исключить эмпирический подход построения индексов и автоматизировать всю процедуру их формирования для любого числа и вида распознаваемых объектов. Основу формализации составляют алгоритмы оценивания и отбора информативности признаков с последующим построением индексных моделей, представляющих собой линейные решающие правила для распознавания классов. При этом признаками классов представляются спектральные поддиапазоны, на которые разбивается весь спектр данных ДЗЗ, а число информативных признаков выбирается из условия обеспечения требуемой надёжности распознавания всех наблюдаемых объектов (классов).

Ключевые слова: индексные модели, распознавание образов, информативные признаки, алгоритмы, информативность признаков, идентификация моделей

Одобрена к печати: 22.11.2022

DOI: 10.21046/2070-7401-2022-19-6-138-150

Введение

В настоящее время для работы со спектральной информацией при дистанционном зондировании Земли (ДЗЗ) часто прибегают к построению так называемых индексных изображений. На основе комбинации параметров отражения в определённых каналах, информативных для выделения исследуемого объекта, и расчёта по этим значениям «спектрального индекса» объекта строится изображение, соответствующее значению индекса в каждом пикселе, что и позволяет выделить исследуемый объект или оценить его состояние. Спектральные индексы, используемые для изучения и оценки состояния растительности, получили общепринятое название вегетационных индексов (ВИ, *англ.* vegetation index — VI) (Андреев и др., 1986; Baret et al., 1989; Crippen, 1990; Huete, 1988; Jackson, 1983; Jordan, 1969; Kaufman, Tanre, 1992; Kriegler, et al., 1969; Richardson, Everitt, 1992).

Вегетационный индекс — это безразмерный скалярный показатель, представляющий собой комбинацию параметров отражения в разных спектральных диапазонах данных дистанционного зондирования, показывающую состояние растительности на данном участке поверхности и в данном пикселе снимка (Pinty, Verstraete, 1991; Qi, Chehbouni, 1994; Richardson, Wiegand, 1977). В настоящее время существует более 160 вариантов вегетационных индексов. Они подбираются экспериментально (эмпирическим путём) исходя из известных особен-

ностей кривых спектральной отражательной способности растительности и почв (Кашкин, Сухинин, 2001).

В начальные годы оптического дистанционного зондирования, когда был доступен не-большой набор спектральных диапазонов, а вычислительная мощность была ограничена, расчёт большей части вегетационных индексов базировался на двух-трёх спектральных диапазонах. Чаще всего использовалось два наиболее стабильных (не зависящих от прочих факторов) участка кривой спектральной отражательной способности растений. На красную зону спектра (620–750 нм) приходится максимум поглощения солнечной радиации хлорофиллом, а на ближнюю инфракрасную зону (750–1300 нм) — максимальное отражение энергии клеточной структурой листа. Высокая фотосинтетическая активность (связанная, как правило, с большой биомассой растительности) ведёт к более низким значениям коэффициентов отражения в красной зоне спектра и большим значениям в ближней инфракрасной. Отношение этих показателей друг к другу позволяет чётко отделять растительность от прочих природных объектов (Андреев и др., 1986; Crippen, 1990; Leprieur et al., 1994; Lillesand, Kiefer, 1987).

С появлением фото- и гиперспектрометров открылись новые возможности для формирования комбинаций спектральных диапазонов и ввода новых видов вегетационных индексов. Так, при обработке космических снимков в программном комплексе ENVI (*англ.* Environment for Visualizing Images) имеется специальный калькулятор вегетационных индексов ENVI (калькулятор VI ENVI), который позволяет рассчитать 27 вегетационных индексов, используемых для оценки состояния растительности, содержания пигментов, азота, углерода, воды (Crippen, 1990).

Наиболее распространённый и часто используемый индекс для решения задач с применением количественных оценок растительного покрова — NDVI (*англ.* Normalized Difference Vegetation Index, нормализованный разностный индекс растительности), впервые описанный Б. В. Роуз (*англ.* B. W. Rouse) в 1973 г. Это простой численный показатель количества фотосинтетически активной биомассы, обычно называемый вегетационным индексом (Rouse et al., 1973). NDVI позволяет выявить проблемные зоны угнетённой растительности, давая возможность принимать наиболее верные в долгосрочной перспективе решения, направленные на повышение урожайности. Участки с различным состоянием растительности или объёмом зелёной фитомассы могут быть изображены разными цветами. При помощи статистической обработки карт NDVI помимо определения количества фитомассы можно также выделить площади посева различных сельскохозяйственных культур.

Следует отметить, что любые вегетационные индексы не дают абсолютных количественных показателей исследуемого свойства и их значения зависят от характеристик сенсора (ширина спектральных каналов, разрешения), условий съёмки, освещённости, состояния атмосферы. Они дают только относительные оценки свойств растительного покрова.

Информационный анализ вегетационных и других спектральных индексов, перечень решаемых задач на их основе, а также быстро растущее число таких индексов показывают, что в настоящее время ещё не сложилась единая методика их формирования и спектрального состава используемых данных ДЗЗ. Отсутствие такой методики приводит к нехватке обоснования для введения новых индексов и их спектрального состава, а также не позволяет оценивать их эффективность при решении задач, в которых эти индексы используются.

В настоящей работе предлагается методика формирования вегетационных и подобных им индексов и оценки их эффективности. Методика апробирована на примере распознавания видов растений сельскохозяйственных культур и сорных растений.

Материалы и методы

Скалярная форма индексов, отсутствие физической размерности, независимо от сочетаний используемых спектральных каналов ДЗЗ, не позволяют решать задачи оценивания количественных показателей исследуемых свойств зондируемых растительных и почвенных покровов. Для оценивания этих показателей в соответствии с современной информационной теорией необходимы математические модели их зависимостей между собой и от основных

влияющих факторов, а также от связи параметров отражения средств ДЗЗ и оцениваемых показателей (Михайленко, 2011; Clevers, 1988; Crist, Cicone, 1984; Huete et al., 1985; Kauth, Thomas, 1976; Lillesand, Kiefer, 1987; Tu et al., 1998). При наличии таких моделей для оценивания векторов количественных показателей необходимы векторы параметров отражения (данных ДЗЗ) соответствующей размерности. Любые комбинации спектральных каналов в связи с отсутствием физической размерности индексов не отвечают этим требованиям. Для оценивания количественных показателей почвенно-растительных покровов необходимо сопоставление моделируемых и измеряемых количественных показателей, чего не позволяют вегетационные индексы.

В то же время данные индексы выступают индикаторами многих неколичественных состояний растительного и почвенного покрова, таких как стрессы различной природы, наличие болезней и вредителей, типы культурных растений и сорняков и многие другие. Именно в оценивании таких состояний были достигнуты наиболее значительные результаты использования различных индексов. Наличие пространственных картин оценок таких состояний — важная мониторинговая информация, позволяющая принимать предварительные управляющие решения начала и очередности технологических управлений, сева, уборки урожая и др. Поэтому повышение надёжности и достоверности таких оценок представляется актуальной проблемой. Ошибки в определении видов растений, сорняков, болезней растений и других неколичественных состояний могут привести к неадекватным управляющим решениям по выбору технологий и препаратов, началу и окончанию работ, а также к большим потерям урожая, используемых удобрений и средств защиты растений. Повышение надёжности и достоверности оценивания неколичественных состояний возможно только при наличии общей методики формирования индексов, исключающей произвол в выборе используемых спектральных диапазонов и нормирующей величины индексов по показателям надёжности оценок.

С точки зрения современной информационной теории оценивание неколичественных показателей (состояний) относится к задаче распознавания образов, а каждое из таких неколичественных состояний может рассматриваться как образ или класс (Дуда, Харт, 1976; Ту, Гонсалес, 1978). Как только мы сводим процедуру оценивания на основе индексов к задаче распознавания образов, то, согласно теории распознавания, сразу же возникнет проблема отбора информативных признаков, на основании которых решается эта задача. При использовании данных ДЗЗ признаками выступают отдельные градации (интервалы) спектра отражения средства ДЗЗ. Сама процедура отбора признаков практически не зависит от способа измерения информативности. Важно лишь, чтобы этот способ был одинаков для всех признаков (групп признаков), входящих в исходное их множество и участвующих в процедуре отбора.

Один из наиболее простых и эффективных методов оценивания информативности признаков — метод Шеннона (Шеннон, 1963). Он предлагает оценивать информативность как средневзвешенное количество информации, приходящееся на различные градации признака. Под информацией в теории информации понимают величину устранённой энтропии.

Информативность по методу Шеннона определяется следующим образом:

$$I(x_i) = 1 + \sum_{i=1}^G \left(P_i \cdot \sum_{k=1}^K P_{i,k} \cdot \log_k P_{i,k} \right), \quad (1)$$

где $I(x_i)$ — информативность признака x_i ; G — количество интервалов (градаций) признака; P_i — вероятность i -го интервала (градации) признака; K — количество классов признака; $P_{i,k}$ — вероятность появления i -го интервала признака в k -м классе.

Вероятность i -го интервала признака (P_i), рассчитывалась по формуле:

$$P_i = \frac{\sum_{k=1}^K m_{i,k}}{N}, \quad (2)$$

где $m_{i,k}$ — частота появления i -го интервала в k -м классе; K — количество классов признака; N — общее количество классов.

Вероятность появления i -го интервала признака в k -м классе (P_{i^*k}) рассчитывалась по формуле:

$$P_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^K m_{i,k}}, \quad (3)$$

где $m_{i,k}$ — частота появления i -го интервала в k -м классе; K — количество классов признака.

В формуле Шеннона (1) информативность представляет собой средневзвешенное количество информации, приходящееся на градации признака. При этом величина устранённой энтропии заключена в выражении под первым знаком суммирования.

После оценивания информативности всех градаций спектрального диапазона используемой системы ДЗЗ для облегчения дальнейшего отбора они выстраиваются по убыванию показателя информативности. Для отбора наиболее эффективных информативных признаков введём математическую модель индекса:

$$I_k = \sum_{i=1}^{n_k} b_{ik} x_{ik}^*, \quad (4)$$

или в векторно-матричной форме:

$$I_k = \mathbf{B}_k^T \mathbf{X}_k^*, \quad (5)$$

где b_{ik} — параметры модели индекса для k -го класса, объединённые в матрицу-строку \mathbf{B}_k^T ; x_{ik}^* — наиболее информативные признаки для k -го класса, обозначаемые индексом I , объединённые в векторы \mathbf{X}_k^* ; I_k — распознаваемые образы, обозначаемые последовательно числами 1, 2, 3, ..., K , служащими выходными величинами модели.

Параметры модели индекса (4), (5) оцениваются по обучающей выборке посредством рекуррентного метода наименьших квадратов РМНК (Эйкхофф, 1975):

$$\left. \begin{aligned} G_k(j) &= R_k(j-1)X_k(j), \\ z_k(j) &= X_k^T(j)G_k(j), \\ e_k(j) &= I_k(j+1) - B_k^T(j-1)X_k(j), \\ B_k(j) &= B_k(j-1) + \frac{1}{1+z_k(j)}G_k(j)e_k(j), \\ R_k(j) &= R_k(j-1) - \frac{1}{1+z_k(j)}G_k(j)G_k^T(j), \end{aligned} \right\} \quad (6)$$

здесь: $j = 1, 2, 3, \dots, j$ — номера записей в обучающей выборке; $e_k(j) = I_k(j+1) - B_k^T(j-1) \times X_k(j)$ — текущая ошибка обучения на j -м шаге процедуры; $G_k(j)$, $z_k(j)$, $R_k(j)$ — соответственно векторный, скалярный и матричный параметры алгоритма, обеспечивающие его сходимость.

В модели (4), (5) и в алгоритме обучения (6) остаётся неясным, сколько наиболее информативных признаков должно быть отобрано для каждого распознаваемого класса, т. е. какова размерность n_k вектора \mathbf{X}_k^* .

Для этого сначала приведём решающее правило, посредством которого принимается решение об отнесении текущего вектора признаков $\mathbf{X}_k^*(j)$ к конкретному классу k^* :

$$I_k(j) = I_k^*, \text{ если } \left| I_k^* - \mathbf{B}_k^T \mathbf{X}_k^*(j) \right| \leq \delta, \quad (7)$$

где δ — заданный допуск на ошибку распознавания; как правило, $\delta = 0,05I_k^*$.

Принимая частоту попадания решений (7) в допуск в качестве оценки вероятности правильных решений:

$$\hat{P}_k = \frac{n_k}{\sum_{k=1}^K n_k}, \quad (8)$$

можно последовательно, начиная с первых двух признаков, вводить в порядке убывания наиболее информативные признаки, пока оценка вероятности (8) не достигнет заданного значения \hat{P}_k^* .

С учётом всех приведённых рассуждений общий алгоритм формирования моделей индексов включает в себя следующие этапы и шаги.

Этап 0. Формирование обучающей выборки. Заключается в фиксации спектров отражения K классов во всём рабочем диапазоне используемого средства ДЗЗ, который разбивается на G отдельных интервалов, обозначаемых переменными $i = 1, 2, \dots, G$. Для разделения информации о классах в естественных условиях выделяются отдельные зоны наблюдения с преимущественным содержанием объектов k -го класса, для агроценозов используются специально выделенные тестовые площадки с отдельными видами распознаваемых классов. Все данные фиксируются в базе данных (БД) с номерами записей $j = 1, 2, \dots, J$.

Этап 1. Оценивание информативности признаков. По формулам (1)–(3) формируется убывающая последовательность десяти наиболее информативных признаков.

Этап 2. Оценивание параметров моделей индексов и отбор информативных признаков:

Шаг 1. Принимается класс $k = 1$ и в модели индекса (5) принимается число информативных признаков $n_k = 2$ (первые два в убывающей последовательности).

Шаг 2. Посредством алгоритма (6) оцениваются параметры модели индекса (5) для текущего числа признаков n_k .

Шаг 3. Посредством решающего правила по контрольной выборке данных оценивают вероятность правильных решений P_k^* , которую сравнивают с заданным значением \hat{P}_k^* . Если $\hat{P}_k < P_k^*$, то принимается число информативных признаков $n_k = n_k + 1$ (добавляется следующий признак из убывающей последовательности) и осуществляется переход к шагу 1. Если $\hat{P}_k \geq P_k^*$, то принимается класс $k = 1 + 1$ и осуществляется переход к шагу 1. И так до последнего класса $k = K$.

Обсуждение результатов

В табл. 1, 2 представлены фрагменты данных спектров отражения различных растений для южных и северных широт Российской Федерации. Они существенно различаются по своим спектральным картинкам, где для растений северных широт значения параметров отражения распределены более равномерно по исследуемым спектральным диапазонам (признакам). При исследовании информативности признаков следует иметь в виду, что эмпирические оценки вероятности в формулах (1)–(3) зависят от выбранных пороговых значений признаков. При этом под пороговым значением подразумевается та величина признака, превышение которой учитывается при оценивании эмпирической вероятности. Как видно из диаграмм на рис. 1–6 (см. с. 143–144), выбор пороговых значений существенно отражается на оценках информативности. При малых значениях порогов эти оценки имеют большие значения и мало различаются между собой. При увеличении величины порогов оценки информативности уменьшаются вплоть до обнуления при увеличении степени различия между собой. В связи с тем, что информативность определяет последовательность ввода признаков в модели индексов (4), (5), следует выбрать компромиссные варианты, в которых нет нулевых информативностей и они достаточно различимы для отдельных признаков. В соответствии с этими соображениями для растений южных широт было выбрано пороговое значение, равное 2 % (см. рис. 1), а для растений северных широт — 6 % (см. рис. 2).

Отбор информативных признаков для построения моделей индексов в соответствии приведённым выше алгоритмом начинается с первых двух наиболее информативных каналов. Для растений южных широт это каналы: x_4 — 680–740 нм, x_5 — 800–1100 нм; для растений

северных широт: x_2 — 530–610 нм, x_4 — 680–740 нм. В табл. 3 приведены результаты оценки вероятности правильного распознавания классов для растений южных широт в зависимости от размерности вектора используемых признаков в модели индексов (5) для поля допуска $\pm 10\%$ от номера класса.

Таблица 1. Фрагмент спектральных данных некоторых растений южных широт Российской Федерации. Третий вегетационный период (III декада июля 2019 г.) (Данилов и др., 2020)

Наименование растений	Номера классов (k)	Спектральная плотность энергетической яркости (СПЭЯ), %				
		450–520 нм	530–610 нм	625–680 нм	680–740 нм	800–1100 нм
		x_1	x_2	x_3	x_4	x_5
Озимая пшеница (<i>T. aestivum</i> L.)	1	2,83±0,30	4,75±0,38	6,19±0,46	6,68±0,37	6,99±0,33
Яровая пшеница (<i>T. aestivum</i> L.)	2	2,81±0,24	4,78±0,31	6,26±0,32	6,32±0,12	6,1±0,38
Яровой ячмень (<i>H. vulgare</i> L.)	3	2,38±0,03	3,94±0,03	4,65±0,04	4,39±0,02	6,07±0,07
Кукуруза (<i>Z. mais</i> L.)	4	2,2±0,10	3,76±0,03	4,88±0,02	5,45±0,02	6,03±0,07
Подсолнечник (<i>H. annuus</i> L.)	5	1,7±0,12	3,42±0,24	2,96±0,28	4,6±0,19	5,25±0,08
Соя (<i>G. max</i> (L.))	6	0,96±0,01	1,67±0,02	1,11±0,02	3,77±0,11	5,95±0,17
Амброзия полыннолистная (<i>A. artemisiifolia</i> L.)	7	1,01±0,03	1,87±0,07	1,42±0,09	3,79±0,17	5,64±0,08
Дурнишник калифорнийский (<i>X. strumarium</i> L.)	8	0,38±0,03	0,96±0,01	0,45±0,01	2,78±0,03	4,98±0,07
Канатник Теофраста (<i>A. theophrastii</i> Medik.)	9	0,76±0,08	1,58±0,15	0,92±0,09	3,61±0,29	5,26±0,35
Марь белая (<i>C. album</i> L.)	10	0,84±0,03	2,04±0,04	1,1±0,03	3,96±0,06	4,91±0,07
Щирица запрокинутая (<i>A. retroflexus</i> L.)	11	1,02±0,05	1,5±0,06	1,02±0,06	3,14±0,05	5,14±0,08
Осот полевой (<i>S. arvensis</i> L.)	12	0,89±0,02	1,76±0,04	1,1±0,03	3,46±0,06	5,58±0,04
Лисохвост мыше-хвостиковидный (<i>A. myosuroides</i> Huds.)	13	1,21±0,09	1,99±0,10	1,74±0,17	3,8±0,09	5,87±0,13

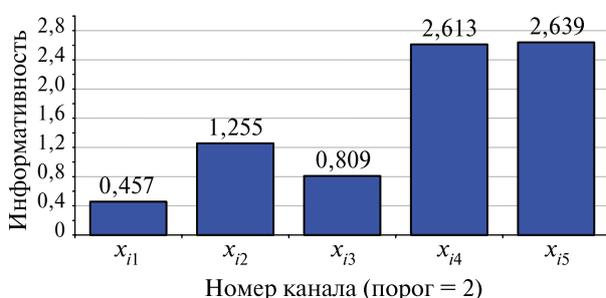


Рис. 1. Диаграмма информативности признаков для растений южных широт и порога чувствительности 2 % (формула Шеннона (1))

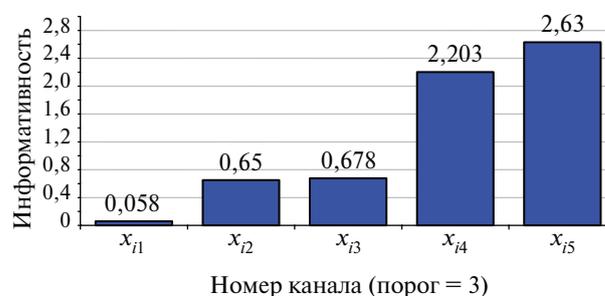


Рис. 2. Диаграмма информативности признаков для растений южных широт и порога чувствительности 3 % (формула Шеннона (1))

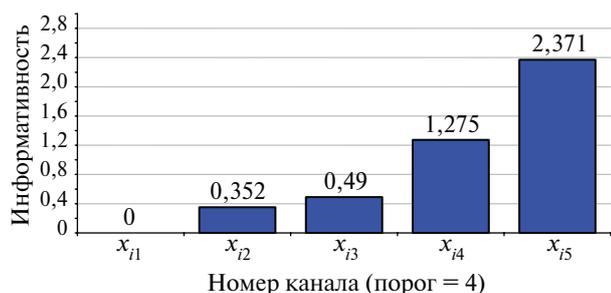


Рис. 3. Диаграмма информативности признаков для растений южных широт и порога чувствительности 4 % (формула Шеннона (1))

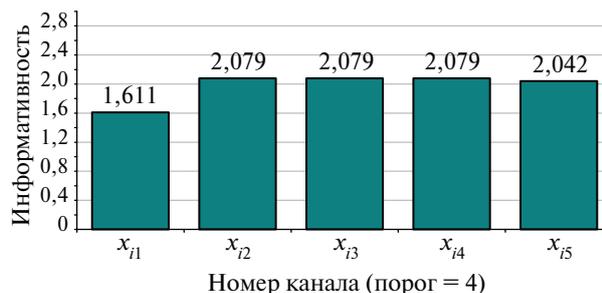


Рис. 4. Диаграмма информативности признаков для растений северных широт и порога чувствительности 4 % (формула Шеннона (1))

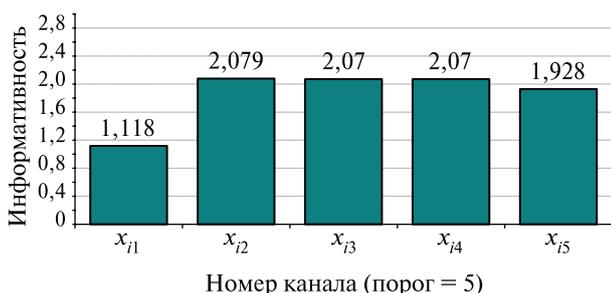


Рис. 5. Диаграмма информативности признаков для растений северных широт и порога чувствительности 5 % (формула Шеннона (1))

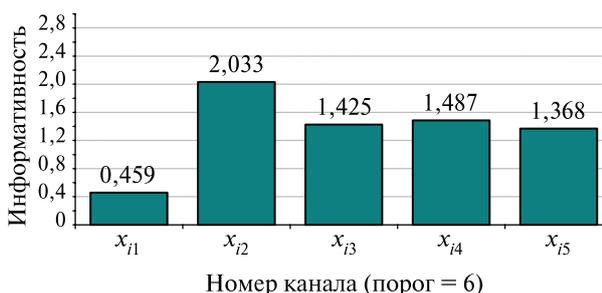


Рис. 6. Диаграмма информативности признаков для растений северных широт и порога чувствительности 6 % (формула Шеннона (1))

Таблица 2. Фрагмент спектральных данных некоторых растений северных широт Российской Федерации. Меньковская опытная станция Агрофизического научно-исследовательского института. Третий вегетационный период (III декада июля 2018 г.)

Наименование растений	Номера классов (k)	Спектральная плотность энергетической яркости (СПЭЯ), %				
		450–520 нм	530–610 нм	625–680 нм	680–740 нм	800–1100 нм
		x_1	x_2	x_3	x_4	x_5
Яровая пшеница (<i>T. aestivum</i> L.)	1	2,38±0,03	3,94±0,04	4,65±0,06	4,39±0,05	6,07±0,08
Пикульник обыкновенный (<i>Galeopsis tetrahit</i>)	2	6,67±0,12	7,24±0,14	6,20±0,16	6,00±0,14	5,80±0,13
Белая марь (<i>C. album</i> L.)	3	5,90±0,10	7,25±0,18	6,38±0,16	6,12±0,14	6,00±0,14
Полевой ярок (<i>Thlaspi arvense</i>)	4	0,96±0,01	1,67±0,02	1,11±0,03	3,77±0,06	5,95±0,10
Сумка пастушья (<i>Capsella bursa-pastoris</i>)	5	4,49±0,09	6,91±0,12	6,0±0,10	6,5±0,12	6,65±0,13
Звездчатка средняя (<i>Stellaria media</i> (L.))	6	5,35±0,11	7,72±0,12	7,08±0,11	7,84±0,09	7,60±0,08
Горец развесистый (<i>Persicaria lapathifolia</i>)	7	5,67±0,09	7,74±0,12	7,24±0,14	7,15±0,13	6,95±0,09
Дымянка аптечная (<i>Fumaria officinalis</i>)	8	7,12±0,15	8,96±0,18	8,24±0,19	8,12±0,15	7,42±0,14

Как видно из табл. 3, вероятность правильного распознавания классов для рассматриваемого случая не зависит от числа взятых информативных признаков. Поэтому для дальнейшего анализа мы используем модели индексов с первыми двумя признаками.

На рис. 7–10 приведены графики идентификации моделей индексов для первых и последних видов растений (классов), представленных в табл. 1, 2. Результаты идентификации даны в табл. 4, 5.

Таблица 3. Вероятность правильного распознавания классов в зависимости от числа используемых признаков.

Число и состав признаков							
2: x_4, x_4		3: x_4, x_4, x_2		4: x_4, x_4, x_2, x_3		5: x_4, x_4, x_2, x_3, x_1	
Вероятность правильной классификации							
1-й класс	14-й класс	1-й класс	14-й класс	1-й класс	14-й класс	1-й класс	14-й класс
0,95							

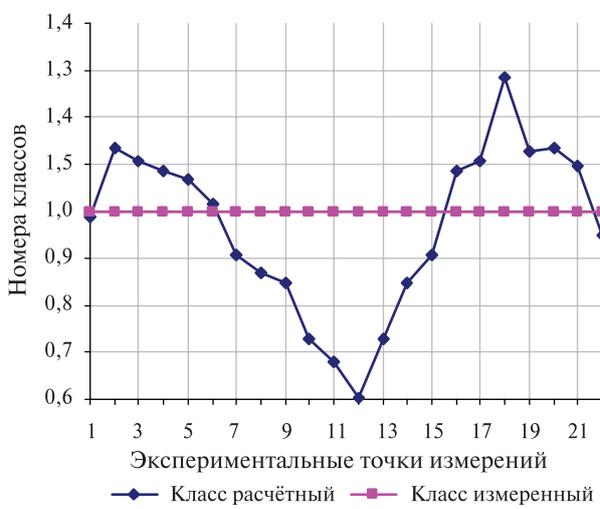


Рис. 7. Идентификация модели индекса 1-го класса растений южных широт (алгоритм (6))

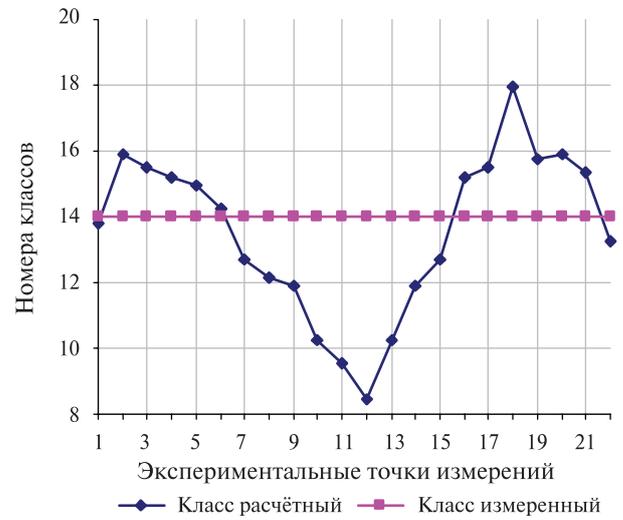


Рис. 8. Идентификация модели индекса 14-го класса растений южных широт (алгоритм (6))

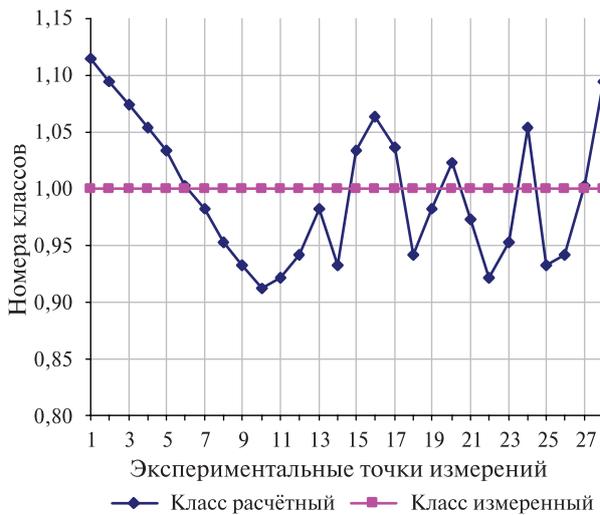


Рис. 9. Идентификация модели индекса 1-го класса растений северных широт (алгоритм (6))

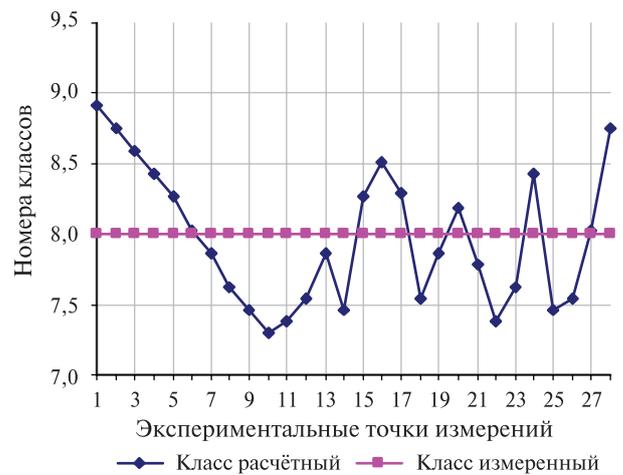


Рис. 10. Идентификация модели индекса 8-го класса растений северных широт (алгоритм (6))

Таблица 4. Результаты идентификации моделей индексов для растений южных широт

№ класса	Значения параметров моделей		Среднеквадратическая ошибка
	b_1	b_2	
1	0,07175	0,09685	0,037
2	0,15193	0,19765	0,074
3	0,33588	0,37310	0,112
4	0,34083	0,48952	0,149
5	0,51534	0,68969	0,186
6	0,58559	1,30939	0,223
7	0,72659	1,46233	0,261
8	0,87877	2,52582	0,298
9	0,90715	2,05981	0,335
10	1,01760	1,96592	0,372
11	1,34589	2,80647	0,410
12	1,38494	2,71301	0,447
13	1,38991	2,64499	0,484
14	1,03964	1,67757	0,521

Таблица 5. Результаты идентификации моделей индексов для растений северных широт

№ класса	Значения параметров моделей		Среднеквадратическая ошибка
	b_1	b_2	
1	0,08456	0,08713	0,011
2	0,16562	0,16874	0,023
3	0,22744	0,20778	0,034
4	0,31158	0,30478	0,045
5	0,32315	0,33304	0,057
6	0,42514	0,43709	0,068
7	0,43551	0,47565	0,079
8	0,65740	0,61787	0,090

Графики на *рис. 7–10* и *табл. 4, 5* показывают, что выбор наиболее информативных признаков и настройка параметров моделей индексов обеспечивают высокую надёжность распознавания классов растений. Такими возможностями все используемые в настоящее время вегетационные и другие индексы не обладают. Для сохранения этих свойств системы распознавания при изменениях спектральных характеристик растений необходима повторная идентификация параметров моделей индексов при смене фенологических фаз растений. Такая методика может быть распространена и на другие объекты почвенно-растительного покрова с другими наборами спектров отражения.

В заключение следует заметить, что предлагаемая методика применима только в попиксельной обработке данных ДЗЗ, когда предполагается, что в отдельном пикселе преобладает объект одного вида.

Заключение

Скалярная форма, отсутствие модельной и математической основы, а также обоснования спектрального состава не позволяют использовать современные вегетационные и другие по-

добные им индексы для оценивания количественных параметров состояния почвенно-растительного покрова. В то же время оценивание большого числа таких неколичественных состояний, как вид растительности, болезни растений, стрессы физического и химического характера, повреждения и др., представляется чрезвычайно актуальной проблемой, решение которой позволяет принимать управляющие решения различного рода. По своему информационному содержанию оценивание неколичественных (счётных) состояний относится к процедурам распознавания образов или классов. В свою очередь, отправным пунктом реализации процедуры распознавания образов становится выявление наиболее информативных признаков, в пространстве которых могут быть построены решающие правила распознавания. При использовании средств оптического дистанционного зондирования Земли (ДЗЗ) такими признаками служат отдельные спектральные поддиапазоны, на которые разбивается весь спектральный диапазон, используемый средствами ДЗЗ. В этом случае индексами выступают линейные решающие правила (ЛПР), число которых равно числу классов, а сами ЛПР представляют собой взвешенную линейную комбинацию наиболее информативных признаков, количество которых в каждом ЛПР выбирается из условия обеспечения требуемой надёжности процедуры распознавания. При таком подходе полностью исключается эмпирический выбор спектрального состава индексов, а вся процедура их формирования может быть формализована и автоматизирована.

Литература

1. Андреев Г. Г., Бочаров В. П., Сазонов Н. В., Чабан Л. Н. Автоматизация тематической обработки космических снимков при оценке состояния сельскохозяйственных культур // Исслед. Земли из космоса. 1986. № 5. С. 95–102.
2. Данилов Р. Ю., Кремнева О. Ю., Исмаилов В. Я., Третьяков В. А., Ризванов А. А., Кривошеин В. В., Пачкин А. А. Общая методика и результаты наземных гиперспектральных исследований сезонного изменения отражательных свойств посевов сельскохозяйственных культур и отдельных видов сорных растений // Современные проблемы дистанционного зондирования Земли из космоса. 2020. Т. 17. № 1. С. 113–127. DOI: 10.21046/2070-7401-2020-17-1-113-127.
3. Дуда Р., Харт П. Распознавание образов и анализ сцены: пер. с англ. Г. Г. Вайнштейна, А. М. Васюковского / под ред. В. Л. Стефанюка. М.: Мир, 1976. 507 с.
4. Кашкин В. Б., Сухинин А. И. Дистанционное зондирование Земли из космоса. Цифровая обработка изображений. М.: Логос, 2001. 264 с.
5. Михайленко И. М. Основные задачи оценки состояния посевной и почвенной среды по данным космического зондирования // Экологические системы и приборы. 2011. № 8. С. 17–24.
6. Ту Дж., Гонсалес Р. Принципы распознавания образов: пер. с англ. И. Б. Гуревича / под ред. Ю. И. Журавлева. М.: Мир, 1978. 411 с.
7. Шеннон К. Работы по теории информации и кибернетике: пер. с англ. / с предисл. А. Н. Колмогорова; под ред. Р. Л. Добрушина, О. Б. Лупанова. М.: Изд-во иностр. лит-ры, 1963. 829 с.
8. Эйкхофф П. Основы идентификации систем управления. Оценка параметров и состояний: пер. с англ. В. А. Лотцкого, А. С. Манделя / под ред. Н. С. Райбмана. М.: Мир, 1975. 683 с.
9. Baret F., Guyot G., Major D. TSAVI: A vegetation index which minimizes soil brightness effects on LAI or APAR estimation // 12th Canadian Symp. Remote Sensing and Remote Sensing Symp. (IGARSS'89). 10–14 July 1989, Vancouver, Canada. 1989. P. 1355–1358. DOI: 10.1109/IGARSS.1989.576128.
10. Clevers J. G. P. W. The derivation of a simplified reflectance model for the estimation of leaf area index // Remote Sensing of Environment. 1988. V. 35. P. 53–70. DOI: 10.1016/0034-4257(88)90041-7.
11. Crippen R. E. Calculating the Vegetation Index Faster // Remote Sensing of Environment. 1990. V. 34. P. 71–73. DOI: 10.1016/0034-4257(90)90085-z.
12. Crist E. P., Cicone R. C. Application of the tasseled cap concept to simulated thematic mapper data // Photogrammetric Engineering and Remote Sensing. 1984. V. 50. P. 343–352.
13. Huete A. R. A Soil-Adjusted Vegetation Index (SAVI) // Remote Sensing of Environment. 1988. V. 25. P. 295–309. DOI: 10.1016/0034-4257(88)90106-X.
14. Huete A. R., Jackson R. D., Post D. F. Spectral response of a plant canopy with different soil backgrounds // Remote Sensing of Environment. 1985. V. 17. P. 37–53. DOI: 10.1016/0034-4257(85)90111-7.
15. Jackson R. D. Spectral indices in n-space // Remote Sensing of Environment. 1983. V. 13. P. 409–421. DOI: 10.1016/0034-4257(83)90010-X.

16. *Jordan C.F.* Derivation of leaf area index from quality of light on the forest floor // *Ecology*. 1969. V. 50. P. 663–666. DOI: 10.2307/1936256.
17. *Kaufman Y.J., Tanre D.* Atmospherically resistant vegetation index (ARVI) for EOS-MODIS// *IEEE Trans. Geoscience and Remote Sensing*. 1992. V. 30. Iss. 2. P. 261–270. DOI: 10.1109/36.134076.
18. *Kauth R.J., Thomas G.S.* The tasseled cap-A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat // *Proc. Symp. Machine Processing of Remotely Sensed Data*. 29 June – 1 July 1976. Paper 159. P. 4B-41–4B-51.
19. *Kriegler F.J., Malia W.A., Nalepka R.F., Richardson W.* Preprocessing transformations and their effects on multispectral recognition // *Proc. 6th Intern. Symp. Remote Sensing of Environment*. Ann Arbor, MI: Univ. Michigan, 1969. P. 97–131.
20. *Leprieur C., Verstraete M.M., Pinty B., Chehbouni A.* NOAA/AVHRR Vegetation Indices: Suitability for Monitoring Fractional Vegetation Cover of the Terrestrial Biosphere // *Physical Measurements and Signatures in Remote Sensing: Proc. 6th ISPRS Symp.* 1994. P. 1103–1110.
21. *Lillesand T.M., Kiefer R.W.* *Remote Sensing and Image Interpretation*. 2nd ed. N.Y.; Chichester; Brisbane; Toronto; Singapore: John Wiley and Sons, 1987. 721 p. DOI: 10.1080/01431168708954824.
22. *Pinty B., Verstraete M.M.* GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites // *Plant Ecology*. 1991. V. 101. P. 15–20. DOI: 10.1007/BF0031911.
23. *Qi J., Chehbouni A., Huete A.R., Ker Y.H.* Modified Soil Adjusted Vegetation Index (MSAVI) // *Remote Sensing of Environment*. 1994. V. 48. P. 119–126. DOI: 10.1016/0034-4257(94)90134-1.
24. *Richardson A.J., Everitt J.H.* Using spectral vegetation indices to estimate rangeland productivity // *Geocarto Intern.* 1992. V. 1. P. 63–69. DOI: 10.1080/10106049209354353.
25. *Richardson A.J., Wiegand C.L.* Distinguishing vegetation from soil background information // *Photogrammetric Engineering and Remote Sensing*. 1977. V. 43. P. 1541–1552.
26. *Rouse J.W., Haas R.H., Schell J.A., Deering D.W.* Monitoring vegetation systems in the great plains with ERTS // *Third ERTS-1 Symp. NASA, Goddard Space Flight Center*, 1973. V. 1. P. 309–317.
27. *Tu T.-M., Chen Ch.-H., Wu J.-L., Chang Ch.-I.* A Fast Two-Stage Classification Method for High Dimensional Remote Sensing Data // *IEEE Trans. Geoscience and Remote Sensing*. 1998. V. 36. No. 1. P. 182–191. DOI: 10.1109/36.655328.

Development of the index paradigm in remote sensing of soil cover

I. M. Mikhailenko, V. N. Timoshin

Agrophysical Research Institute, Saint Petersburg 195220, Russia
E-mails: ilya.mihailenko@yandex.ru, v.timoshin@yandex.ru

The aim of the work is the systematic analysis and generalization of the conventional index paradigm of using Earth remote sensing data to assess the state of the soil and vegetation cover. It has been established that the scalar form and the lack of a mathematical basis do not allow the use of conventional vegetation and the similar indices for estimating the vectors of quantitative indicators of the soil and vegetation cover. At the same time, for making many types of management decisions in agriculture, it is important to build index images which reflect such qualitative indicators as types of cultivated and weed plants, the presence of plant diseases, damage of crops and soils, physical and chemical stresses. In terms of informational content, the evaluation of such qualitative states is a procedure for recognizing patterns or classes of soil-and-vegetation complex objects. The subjective empirical approach in choosing the spectral composition of the indices of their combinations, which is currently used, does not currently allow for sufficient reliability of such procedures. Therefore, the purpose of the study present is to formalize the process, which enables to completely exclude the empirical approach of constructing indices and automate the entire procedure for their formation for any number and types of recognizable objects. The basis of formalization is the algorithms for evaluating and selecting the information content of features, followed by the construction of index models, which are linear decision rules for class recognition. The attributes of the classes are the spectral subranges into which the entire spectrum of remote sensing data is divided. The number of informative features is selected from the condition for ensuring the required reliability of recognition of all observed objects (classes).

Keywords: index models, pattern recognition, informative features, algorithms, informative features, model identification

Accepted: 22.11.2022

DOI: 10.21046/2070-7401-2022-19-6-138-150

References

1. Andreev G. G., Bocharov V. P., Sazonov N. V., Chaban L. N., Automation of thematic processing of space images in assessing the state of agricultural crops, *Issledovanie Zemli iz kosmosa*, 1986, No. 5, pp. 95–102 (in Russian).
2. Danilov R. Yu., Kremneva O. Yu., Ismailov V. Ya., Tretyakov V. A., Rizvanov A. A., Krivoshein V. V., Pachkin A. A., General methodology and results of ground-based hyperspectral studies of seasonal changes in the reflective properties of agricultural crops and certain types of weeds, *Sovremennye proplemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2020, Vol. 17, No. 1, pp. 113–127 (in Russian), DOI: 10.21046/2070-7401-2020-17-1-113-127.
3. Duda R., Hart P., *Pattern Classification and Scene Analysis*, New York; London; Sydney; Toronto: John Wiley and Sons, 1973.
4. Kashkin V. B., Sukhinin A. I., *Distantsionnoe zondirovanie Zemli iz kosmosa. Tsifrovaya obrabotka izobrazhenii* (Remote sensing of the Earth from space. Digital image processing), Moscow: Logos, 2001, 264 p. (in Russian).
5. Mikhaylenko I. M., Main tasks evaluation crop condition and soil emissions on this space sensing, *Ekologicheskie sistemy i pribory*, 2011, No. 8, pp. 17–24 (in Russian).
6. Tou J. T., Gonzalez R. C., *Pattern Recognition Principles*, Reading, Massachusetts; London; Amsterdam; Don Mills, Ontario; Sydney; Tokyo: Addison-Wesley Publ. Company, 1974. 378 p.
7. Shannon K., *Works on Information Theory and Cybernetics*, R. L. Dobrushin, O. B. Lupanov (eds.), Moscow: Publ. House of Foreign Literature, 1963, 829 p. (in Russian).
8. Eickhoff P., *System Identification: Parameters and State Estimation*, London: Wiley-Interscience, 1974, 555 p.
9. Baret F., Guyot G., Major D., TSAVI: A vegetation index which minimizes soil brightness effects on LAI or APAR estimation, *12th Canadian Symp. Remote Sensing and Remote Sensing Symp. (IGARSS'89)*, 10–14 July 1989, Vancouver, Canada, pp. 1355–1358, DOI: 10.1109/IGARSS.1989.576128.
10. Clevers J. G. P. W., The derivation of a simplified reflectance model for the estimation of leaf area index, *Remote Sensing of Environment*, 1988, Vol. 35, pp. 53–70, DOI: 10.1016/0034-4257(88)90041-7.
11. Crippen R. E., Calculating the Vegetation Index Faster, *Remote Sensing of Environment*, 1990, Vol. 34, pp. 71–73, DOI: 10.1016/0034-4257(90)90085-z.
12. Crist E. P., Cicone R. C., Application of the tasseled cap concept to simulated thematic mapper data, *Photogrammetric Engineering and Remote Sensing*, 1984, Vol. 50, pp. 343–352.
13. Huete A. R., A Soil-Adjusted Vegetation Index (SAVI), *Remote Sensing of Environment*, 1988, Vol. 25, pp. 295–309, DOI: 10.1016/0034-4257(88)90106-X.
14. Huete A. R., Jackson R. D., Post D. F., Spectral response of a plant canopy with different soil backgrounds, *Remote Sensing of Environment*, 1985, Vol. 17, pp. 37–53, DOI: 10.1016/0034-4257(85)90111-7.
15. Jackson R. D., Spectral indices in n-space, *Remote Sensing of Environment*, 1983, Vol. 13, pp. 409–421, DOI: 10.1016/0034-4257(83)90010-X.
16. Jordan C. F., Derivation of leaf area index from quality of light on the forest floor, *Ecology*, 1969, Vol. 50, pp. 663–666, DOI: 10.2307/1936256.
17. Kaufman Y. J., Tanre D., Atmospherically resistant vegetation index (ARVI) for EOS-MODIS, *IEEE Trans. Geoscience and Remote Sensing*, 1992, Vol. 30, Issue 2, pp. 261–270, DOI: 10.1109/36.134076.
18. Kauth R. J., Thomas G. S., The tasseled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat, *Proc. Symp. Machine Processing of Remotely Sensed Data*, 29 June – 1 July 1976, Paper 159, pp. 4B-41–4B-51.
19. Kriegler F. J., Malia, W. A., Nalepka R. F., Richardson W., Preprocessing transformations and their effects on multispectral recognition, *Proc. 6th Intern. Symp. Remote Sensing of Environment*, Ann Arbor. MI: Univ. Michigan, 1969, pp. 97–131.
20. Leprieux C., Verstraete M. M., Pinty B., Chehbouni A., NOAA/AVHRR Vegetation Indices: Suitability for Monitoring Fractional Vegetation Cover of the Terrestrial Biosphere, *Physical Measurements and Signatures in Remote Sensing: Proc. 6th ISPRS Symp.*, 1994, pp. 1103–1110.
21. Lillesand T. M., Kiefer R. W., *Remote Sensing and Image Interpretation, 2nd ed.*, New York; Chichester; Brisbane; Toronto; Singapore: John Wiley and Sons, 1987, 721 p., DOI: 10.1080/01431168708954824.

22. Pinty B., Verstraete M. M., GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites, *Plant Ecology*, 1991, Vol. 101, pp. 15–20, DOI: 10.1007/BF0031911.
23. Qi J., Chehbouni A., Huete A. R., Ker Y. H., Modified Soil Adjusted Vegetation Index (MSAVI), *Remote Sensing of Environment*, 1994, Vol. 48, pp. 119–126, DOI: 10.1016/0034-4257(94)90134-1.
24. Richardson A. J., Everitt J. H., Using spectral vegetation indices to estimate rangeland productivity, *Geocarto Intern*, 1992, Vol. 1, pp. 63–69, DOI: 10.1080/10106049209354353.
25. Richardson A. J., Wiegand C. L., Distinguishing vegetation from soil background information, *Photogrammetric Engineering and Remote Sensing*, 1977, Vol. 43, pp. 1541–1552.
26. Rouse J. W., Haas R. H., Schell J. A., Deering D. W., Monitoring vegetation systems in the great plains with ERTS, *Third ERTS Symp.*, NASA, Goddard Space Flight Center, 1973, Vol. 1, pp. 309–317.
27. Tu T.-M., Chen Ch.-H., Wu J.-L., Chang Ch.-I., A Fast Two-Stage Classification Method for High Dimensional Remote Sensing Data, *IEEE Trans. Geoscience and Remote Sensing*, 1998, Vol. 36, No. 1, pp. 182–191, DOI: 10.1109/36.655328.