

УДК 528.92

DOI: 10.33764/2411-1759-2021-26-4-65-72

ИНДЕКС ПОЛЕВОГО ОБСЛЕДОВАНИЯ ТЕРРИТОРИИ ПРИ СРЕДНЕМАСШТАБНОМ КАРТОГРАФИРОВАНИИ

Игорь Дмитриевич Махатков

Институт почвоведения и агрохимии СО РАН, 630090, Россия, г. Новосибирск, пр. Лаврентьева, 8/2, кандидат биологических наук, старший научный сотрудник лаборатории географии и генезиса почв, тел. (913)906-07-15, e-mail: makhatkov@issa-siberia.ru

Качество тематического картографирования во многом определяется репрезентативностью натурального обследования территории. В отличие от крупномасштабного картографирования, когда для проектирования точек обследования возможно использование приемов геостатистики, при среднемасштабном картографировании размещение точек обследования традиционно опирается на опыт, общие представления о тематическом разнообразии территории и имеющиеся картографические материалы. При численном пространственном моделировании тематических переменных используются косвенные признаки, в основном – спектральные изображения поверхности и производные цифровой модели рельефа. Пространственная экстраполяция точечной тематической информации так или иначе связана с использованием мер расстояний изученных и неизученных точек в пространстве косвенных переменных. Эту меру предлагается использовать в качестве численной меры качества обследования всего пространства картографирования. В статье приводятся численные эксперименты, иллюстрирующие поведение индекса. Предлагаются возможные пути использования индекса для контроля натурных работ с целью улучшения пространственной модели тематических переменных.

Ключевые слова: цифровая картография, натурные наблюдения, дистанционные данные, пространственное моделирование, тематическое картографирование, планирование наблюдений, контроль качества

Введение

Создание тематических карт связано с решением большого числа задач, в том числе с определением оптимального набора натуральных наблюдений. Основное требование к натурным исследованиям – их репрезентативность, максимально возможный охват разнообразия тематической информации территории, например разнообразия растительности или почвенного покрова [1, 2]. Традиционно в практике тематического картографирования выбор локализации натурального обследования территории опирается на опыт, общие представления о вероятном разнообразии тематической информации на территории и имеющиеся картографические материалы [3].

При создании цифровых моделей выбор мест и количества натуральных наблюдений определяется компромиссом между их репрезентативностью и точностью обследования с одной стороны, ограниченностью во времени, средствах и доступностью территории – с другой. В крупномасштабном карто-

графировании для проектирования натурального обследования привлекаются предварительные сведения, первичная выборка и средства геостатистики, позволяющие построить предварительную пространственную модель [4] и дополнить выборку, целенаправленно оптимизируя эту модель [5]. Такие подходы распространены, в том числе и в исследованиях почвенного покрова на небольших и сравнительно однородных пространствах, например в пределах пашни [6] или в процедуре кригинга с использованием дополнительных, косвенных признаков [7].

При среднемасштабном картографировании условие сравнительной однородности не соблюдается, что не позволяет использовать методы геостатистики. Создание пространственных моделей требует привлечение косвенных переменных, предположительно статистически связанных с тематическими [8]. В картографировании растительности и почвенного покрова это по большей части спектральные изображения поверхности и производные цифровой модели рельефа [9, 10].

В этом случае для рационального размещения точек натурного обследования эти косвенные переменные предлагается использовать для предварительной стратификации поверхности, и размещать точки обследования во всех стратах, равномерно или с учетом их площади [11, 12]. Учитывая, как правило, нелинейный характер связи тематических и косвенных переменных (будущих предикторов пространственной модели), очевидный недостаток такого подхода – сильная зависимость обучающей выборки и результирующей модели от предварительной стратификации [13]. Существенное ограничение оказывает и практическая доступность территории, что приходится учитывать при проектировании натурных работ [14].

В практике среднемасштабного картографирования почв и растительности натурные работы обычно планируются только в самых общих чертах, в основном исходя из общих представлений о пространственных закономерностях и доступности территории, хотя цифровые методы картографирования могли бы опираться и на численное выражение качества полевого обследования. Пространственные модели тематических переменных так или иначе опираются на статистическую связь тематических и косвенных переменных в точках обследования, а пространственная экстраполяция – на меры сходства изученных и неизученных точек косвенных переменных, т. е. на расстояния между точками в пространстве косвенных переменных [15, 16]. Например, если основой пространственного моделирования служат мультиспектральные снимки, то мерой сходства точек (пикселей изображения) будет расстояние в пространстве спектральной яркости.

Предлагаемая мера

Эту меру (расстояние в пространстве косвенных переменных) можно использовать и как самостоятельную величину для оценки качества обследования территории. Если тематическая информация выявлена для некоторого количества точек, то тематическая изученность территории может быть оценена как расстояние всех точек изображения до точек с выявленной тематической информацией

в пространстве переменных косвенной информации. Наиболее простой метрикой изученности будет евклидово расстояние или его производные. Например, в случае мультиспектральной съемки – расстояние каждого пикселя изображения до изученного в пространстве спектральных яркостей

$$R = \sqrt{\sum_{i=1}^k (P_i - S_i)^2},$$

где R – расстояние; k – количество спектральных каналов снимка (или в других случаях – количество переменных косвенной информации); P_i – значение i -го канала пикселя базового изображения; S_i – значение i -го канала изученного пикселя. Если изучено некоторое количество точек (пикселей), то мерой изученности каждого пикселя изображения будет расстояние до ближайшего изученного пикселя

$$R_{\min} = \min R(n, S_j),$$

где n – количество изученных пикселей; S_j – значения спектральных откликов j -го изученного пикселя. Полученные расстояния безразмерные, что неудобно при сравнении величины изученности разного набора данных, например по спектральным откликам поверхности и переменным цифровой модели рельефа. Удобнее использовать относительные значения расстояний, т. е. нормировать полученные значения расстояний по величине максимального расстояния двух точек косвенной информации

$$B_{\max} = \max B(P_i^m, P_j^m),$$

где B – расстояния между парами точками P_i^m и P_j^m косвенной информации; m – количество пикселей. Предварительные расчеты индекса показали, что квадратный корень из этой величины выразительнее, он и используется в дальнейшем

$$R_{\text{norm}} = \sqrt{R_{\min} / B_{\max}}.$$

Средняя величина индекса всех точек косвенной информации территории (R_{av}) может служить показателем общего качества обследования территории картографирования.

Численные эксперименты

Рассмотрим варианты расчета индекса обследования на примере фрагмента снимки Landsat 8 (<http://glovis.usgs.gov/>), в центральном секторе Сибирских Увалов (75.035 в. д. / 63.223 с. ш. – 75.519 в. д. / 63.391 с. ш.). Территория представлена в основном тремя контрастными типами ландшафта с ясными границами – северотаежными зональными лиственничными кустарничково-зеленомошными лесами [17] на подзолистых и глееподзолистых почвах [18], интразональными сосновыми кустарничково-лишайниковыми лесами [19] на песчаных подзолах иллювиально-железистых и комплексными верховыми болотами на торфяниках. Промежуточные варианты и пойменные ландшафты занимают по площади подчиненное положение.

Для обращения с данными использовался QGIS 2.8.6 (<http://www.qgis.org/>) с дополнительными плагинами, для вычислений – Python 2.7 (<https://www.python.org/>) с необходимыми библиотеками (<http://www.numpy.org/>, <http://scikit-learn.org/>, <http://www.gdal.org/>), и редактор PyCharm Community (<https://www.jetbrains.com/>).

Варианты расчета делались для массивов по 10 точек заведомо неполного обследования территории: 1 – только в сосновых лишайниковых лесах, 2 – только в зональных лиственничных зеленомошных лесах, 3 – только на облесенных верховых болотах, входящих в состав болотных комплексов.

Известно, что величины яркости спектральных каналов проявляют тесную взаимную корреляцию. Пиксели изображения в пространстве исходных значений яркости образуют сильно вытянутое облако (рис. 1, а).

Очевидно, расчет расстояний при этом может быть неэффективным. Для более корректной оценки расстояний целесообразнее использовать производные ортогональные переменные (рис. 2, б). В дальнейших численных экспериментах использовались значения главных компонент мультиспектральных изображений.

Чем больше информации о территории, тем эффективнее применение индекса обследования. Например, при использовании переменных только одного спектрального снимка самые большие значения индекса показали водные поверхности и техногенные нарушения (рис. 2, вариант 1). Для большей части территории индекс показывает лучшую тематическую изученность для участков, сходных с обследованными, но он оказался не очень выразительным. Если использовать сезонную серию снимков Landsat (значений главных компонент 18 каналов, из которых 4 оценивались как значимые), то индекс при тех же точках обследования более выразителен. Изображение индекса ясно показывает хорошую изученность только участков сосновых лесов.

При смещении точек обследования меняется и общая картина качества обследования территории (рис. 3).

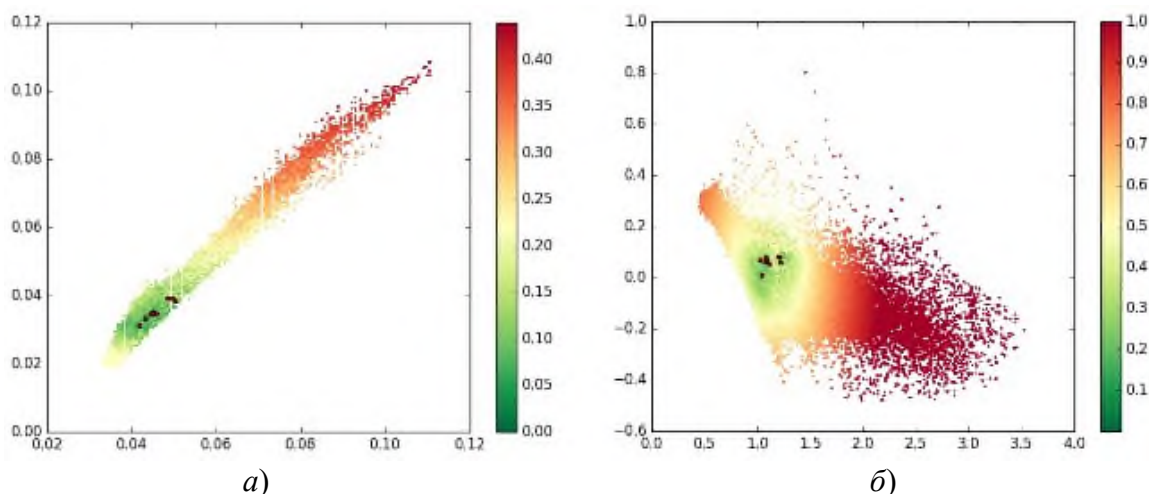


Рис. 1. Положение точек обследования сосновых лесов:

а) в пространстве 1 и 2 канала снимка Landsat; б) в пространстве значений 1 и 2 компоненты 6 каналов того же снимка

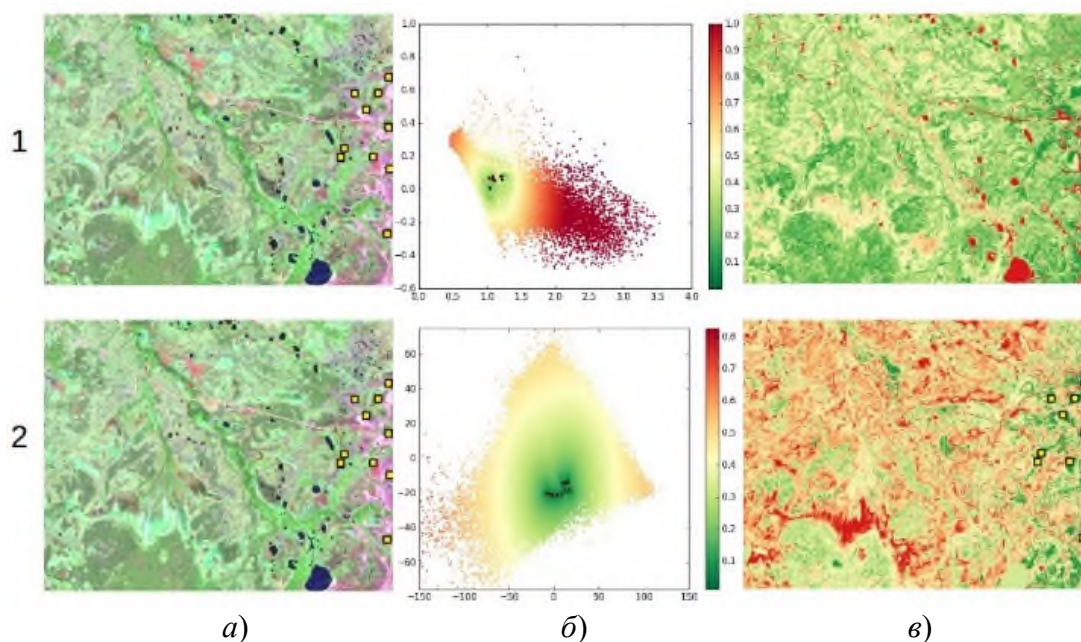


Рис. 2. Варианты расчета индекса обследования:

1 – использован один снимок (всего 6 каналов); 2 – использовано 3 снимка (всего 18 каналов); а) положение описаний на снимке Landsat 8 (желтые точки); б) положение изученных пикселей в пространстве 1 и 2 компонент спектральных откликов (черные точки); в) значение индекса изученности. Раскраска в б) и в) по значениям индекса

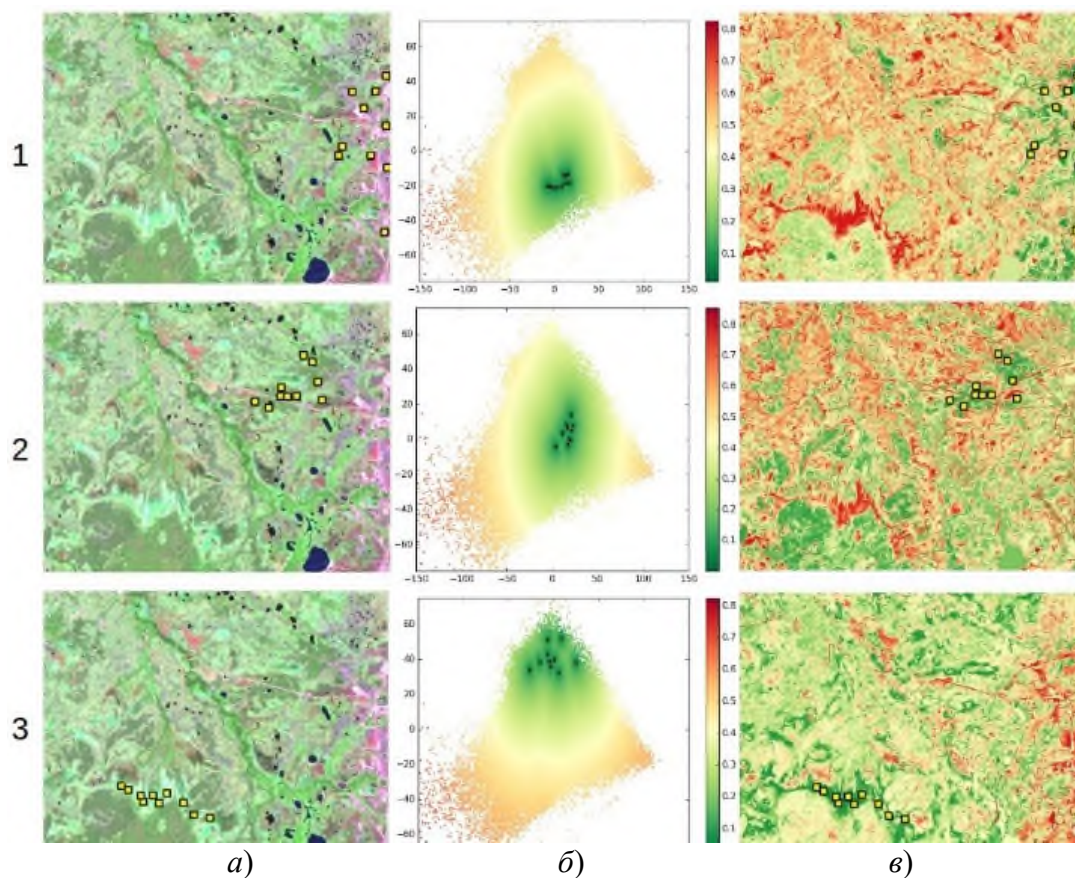


Рис. 3. Варианты расчета индекса изученности с использованием 3 снимков:

1 – изучены только сосновые леса; 2 – изучены только лиственные леса; 3 – изучены только верховые облесенные болота; а), б), в) как на рис. 2

В каждом случае неполного обследования изученные пиксели в точках обследования были сконцентрированы в ограниченной области переменных (рис. 3, б). Индексы обследования – расстояния для пикселей изображения с близкими значениями переменных в окрестностях точек обследования были наименьшими, и возрастали по мере роста расстояния в пространстве 4 компонент, с удалением от точек обследования. Визуализация значений индекса (рис. 3, в) показывает локализацию участков поверхности с разной степенью изученности. Во всех случаях индекс ясно показывает, какого типа поверхность изучена лучше и какого – хуже, что позволяет оценить качество обследования разных участков территории картографирования и корректировать ход натурных работ.

Показатель общего качества обследования (R_{av}) меняется в зависимости от репрезентативности обследования (рис. 4). Например, для нашего случая, если 10 точек обследования сосредоточены не только с сосновых лесах (рис. 4, вариант 1), а во всех трех типах ландшафта (рис. 4, вариант 2), то среднее значение индекса для всей территории (R_{av}) снижается с 0,32 до 0,27. При увеличении точек обследования до 30 в тех же сообществах (рис. 2, вариант 3) среднее значение индекса хотя и снижается, но незначительно, до 0,26. В обоих случаях, и при 10, и при 30 точках обследования в трех типах ландшафта, участки с хорошей и плохой изученностью почти аналогичны (рис. 4, вариант 2 и 3), к плохо изученным участкам в этих двух вариантах относятся поймы местных рек и обширные плоскобугристые болота, где нет ни одной точки обследования.

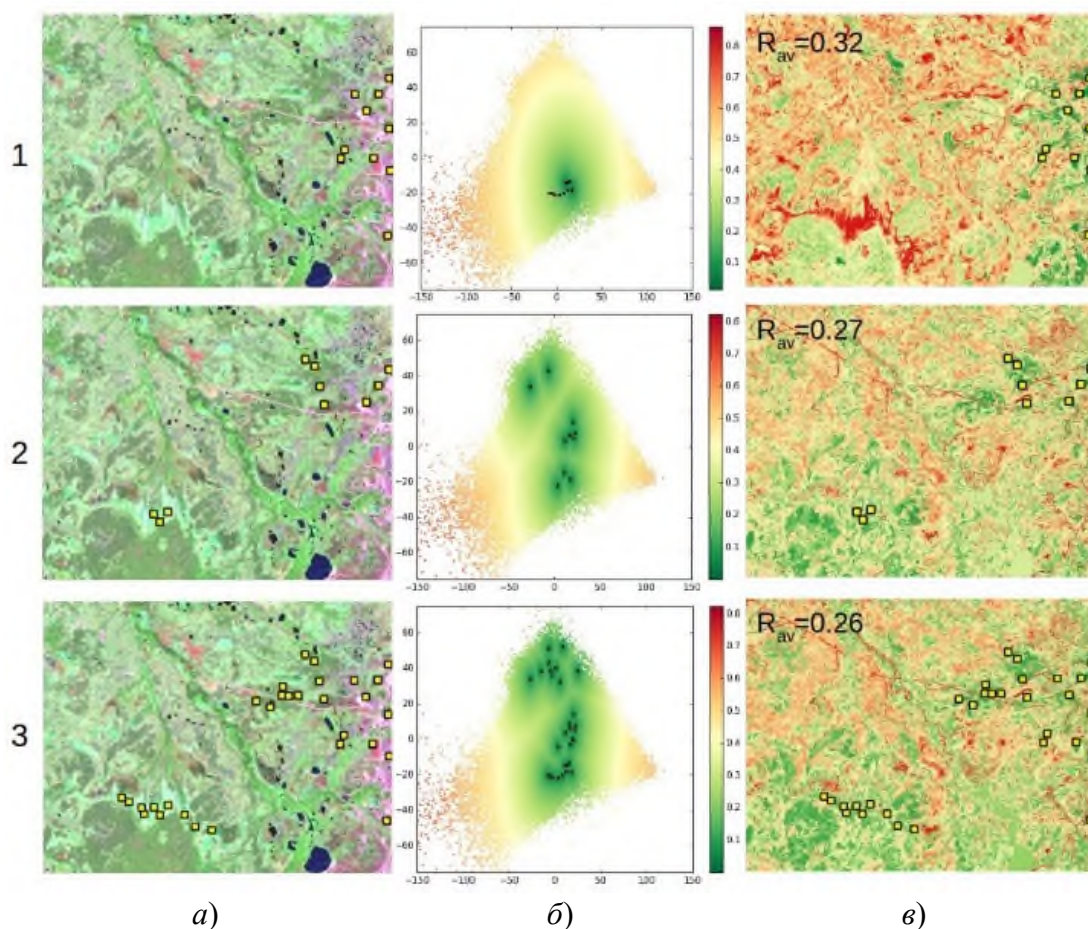


Рис. 4. Варианты расчета индекса изученности:

1 – изучены только сосновые кустарничково-лишайниковые леса; 2 – изучены только сосновые и лиственничные леса; 3 – изучены сосновые и лиственничные леса, и верховые облесенные болота; а), б), в) как на рис. 2

Заключение

В предложенном варианте расчета использовалось простое евклидово расстояние в пространстве базовых переменных. Возможно, более выразительной величиной будут другие метрики расстояний. Кроме того, для выразительности индекса, возможно, будет полезно отсеять самые большие значения индекса, связанные с экстремумами значениями спектральных яркостей или их производных небольшого количества пикселей изображения, а возможно – использование порогового значения индекса, делящего пиксели на «хорошо» и на «плохо» изученные.

Понятно, что индекс может рассчитываться для большого числа переменных в многомерном пространстве, которые невозможно оценить визуально. Кроме того, такие переменные могут сразу включать самую разнородную косвенную информацию о территории картографирования. Например, это могут быть значения главных компонент спектральных откликов поверхности, производные ЦМР, климатические переменные (для больших территорий) и т. д., предварительно нормализованные, разумеется. В таком случае индекс будет учитывать максимально возможную информацию о разнообразии поверхности. И наоборот, расчет индекса отдельно для разной косвенной информации может быть интересен для сопоставления ее информативности в конкретной картографической задаче.

В рассмотренном варианте расчета индекса предполагается, что точка обследования охватывает только один пиксель, т. е. только одну точку базовой информации, что оправдано, когда обследование ограничено небольшой площадью, которая намного меньше разрешения косвенной информации, и воспринимается как точечное обследо-

дование. В случае, когда площадь обследования превышает разрешение косвенной информации, т. е. включает несколько точек косвенной информации, индекс обследования, вероятно, может учитывать несколько метрик, например расстояние от средневзвешенного значения, медианы, квадратичного отклонения и т. д., значений изученных точек косвенной информации в одной точке обследования.

Понятно, что индекс может дать только приблизительное представление о натурном тематическом обследовании территории. Например, на предложенной территории напочвенный покров пойменных лесов с доминированием березы и сходным спектральным откликом может в разных участках поймы сильно отличаться, т. е. представлять разные типы растительности. Могут быть и обратные ситуации, когда участки со сходной растительностью обладают существенно различным спектральным откликом [13]. Все же, учитывая, что косвенная информация в дальнейшем будет основой пространственной экстраполяции тематической, использование индекса обследования может быть очень полезной и для корректировки натурального обследования территории, и для понимания качества экстраполяции тематической информации разных участков пространственной модели. Возможно использование индекса в два этапа, аналогично некоторым геостатистическим подходам в проектировании натурального обследования [4, 5]. На первом этапе – с использованием косвенной информации, на втором, после набора первичного объема точек обследования – с использованием предварительной пространственной модели тематической информации. На втором этапе индекс можно рассчитывать в пространстве тематических переменных, что даст более адекватную картину качества обследования.

Работа выполнена по государственному заданию Института почвоведения и агрохимии СО РАН.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Грибова С. И., Исаченко Т. И. Картирование растительности в съемочных масштабах. Полевая геоботаника. – Л.: Наука. Т. 4. 1972. – С. 137–330.
2. Franklin J. Mapping Species Distributions: Spatial Inference and Prediction. – Cambridge: Cambridge University Press, 2009. – 320 p.
3. Савин И. Ю., Жоголев А. В., Прудникова Е. Ю. Современные тренды и проблемы почвенной картографии // Почвоведение. – 2019. – № 5. – С. 517–528.

4. Delmelle E. M. Spatial Sampling. Handbook of regional science. – Berlin : Springer, 2014. – P. 1385–1399.
5. Lark R. M., Marchant B. P. How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters? // *Geoderma*. – 2018. – Vol. 319. – P. 89–99.
6. Brus D. J., Gruijter J. J., Van Groenigen J. W. Designing spatial coverage samples by the k-means clustering algorithm // *Proceedings of the 8th International FZK/TNO Conference on contaminated soil*. – Gent (Belgium), 2003. – P. 504–509.
7. Brus D. J., Heuvelink G. B. M.. Optimization of sample patterns for universal kriging of environmental variables // *Geoderma*. – 2007. – Vol. 138. – P. 86–95.
8. Heuvelink G. B. M., Brus D. J., De Gruijter J. J. Optimization of sample configurations for digital mapping of soil properties with universal kriging // *Developments in Soil Science*. – 2007. – Vol. 31. – P. 137–151.
9. Сорокина Н. П. Методология составления крупномасштабных агроэкологически ориентированных почвенных карт. – М. : Почвенный институт им. В. В. Докучаева, 2006. – 160 с.
10. Симакова М. С., Рухович Д. И., Белобров В. П., Молчанов Э. Н., Чижилова Н. П., Гаврилова И. П., Герасимова М. И., Богданова М. Д. Руководство по среднemasштабному картографированию почв на основе ГИС. – М. : Почв. ин-т им. В. В. Докучаева, 2008. – 343 с.
11. Dobermann A., Simbahan G. C. Methodology for using secondary information in sampling optimisation for making fine-resolution maps of soil organic carbon // *Developments in Soil Science*. – 2007. – Vol. 31. Chapter 13. – P. 167–182.
12. Minasny B., McBratney A. B. Latin hypercube sampling as a tool for digital soil mapping // *Developments in Soil Science*. – 2007. – Vol. 31. – P. 153–166.
13. Rolecek J., Chytry M., Hajek M., Lvoncik S., Tichy L. Sampling design in large-scale vegetation Studies: do not sacrifice ecological thinking to Statistical purism! // *Folia Geobotanica*. – 2007. – No. 42. – P. 199–208.
14. Stumpf F., Schmidt K., Behrens T., Schonbrodt-Stitt S., Buzzo G., Dumperth C., Wadoux A., Xiang W., Scholten T. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping // *Journal of Plant Nutrition and Soil Science*. – 2016. – Vol. 179. – P. 499–509.
15. Савиных В. П., Цветков В. Я. Геоинформационный анализ данных дистанционного зондирования. – М. : Картгеоцентр – Геодезиздат, 2001. – 228 с.
16. Замятин А. В. Методы интеллектуального анализа данных в региональных системах аэрокосмического мониторинга // *Вестник СГУГиТ*. – 2016. – Вып. 4 (36). – С. 74–88.
17. Ильина И. С. и др. Растительный покров Западно-Сибирской равнины. – Новосибирск : Наука, 1985. – 249 с.
18. Смоленцев Б. А. Структура почвенного покрова Сибирских Увалов (северо-таежная подзона Западной Сибири). – Новосибирск : Изд-во СО РАН, 2002. – 118 с.
19. Хозяинова Н. В. Флора и растительность северной тайги Пуровского района Тюменской области (север Западной Сибири) // *Вестник экологии, лесоведения и ландшафтоведения*. – 2008. – Вып. 8. – С. 27–42.

Получено 06.04.2021

© И. Д. Махатков, 2021

INDEX OF FIELD SURVEY OF THE TERRITORY FOR MEDIUM-SCALE MAPPING

Igor D. Makhatkov

Institute of Soil Science and Agrochemistry SB RAS, 8/2, Prospect Lavrentiev St., Novosibirsk, 630090, Russia, Ph. D., Senior Researcher, Laboratory of Soil Geography and Genesis, phone: (913)906-07-15, e-mail: makhatkov@mail.ru

The quality of thematic mapping mostly depends on the representativeness of the field survey of the territory. In large-scale mapping geostatistical techniques can be used to design survey spatial sampling. In medium-scale mapping, the sampling scheme is traditionally based on experience, common views about the thematic diversity of the territory and available cartographic data. In the digital mapping of thematic variables secondary data is used. There are mainly remote sensing data and derivatives of a digital elevation model. Spatial extrapolation of thematic information of points is somehow connected with the use of distance measures between studied and unstudied points in the space of secondary variables. This measure is proposed to be used as a numerical measure of the survey quality of the entire mapping space. This article provides numerical experiments to illustrate the behavior of this index, as well as proposes possible ways of using the index to control fieldwork in order to improve the spatial model of thematic variables.

Keywords: digital cartography, field survey, remote sensing data, spatial modeling, thematic mapping, sampling design, quality control

REFERENCES

1. Gribova, S. I., & Isachenko, T. I. (1972). *Kartirovaniye rastitel'nosti v s'yemochnykh masshtabakh. Poleyaya geobotanika: T. 4 [Vegetation mapping at survey scale. Field geobotany: Vol. 4]* (pp. 137–330). Leningrad: Nauka Publ. [in Russian].
2. Franklin, J. (2009). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press, 320 p.
3. Savin, I. Yu., Zhogolev, A. V., & Prudnikova, Ye. Yu. (2019). Modern trends and problems of soil cartography. *Pochvovedeniye [Eurasian Soil Science]*, 5, 517–528 [in Russian]
4. Delmelle, E. M. (2014). *Spatial Sampling. Handbook of regional science* (pp. 1385–1399). Berlin: Springer.
5. Lark, R. M., & Marchant, B. P. (2018). How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters? *Geoderma*, 319, 89–99.
6. Brus, D. J., Gruijter, J. J., & Van Groenigen, J. W. (2003). Designing spatial coverage samples by the k-means clustering algorithm. *Proceedings of the 8th International FZK/TNO Conference on Contaminated Soil* (pp. 504–509). Gent (Belgium).
7. Brus D. J., & Heuvelink G. B. M. (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138, 86–95.
8. Heuvelink, G. B. M., Brus, D. J., & De Gruijter, J. J. (2007). Optimization of sample configurations for digital mapping of soil properties with universal kriging. *Developments in Soil Science*, 31, 137–151.
9. Sorokina, N. P. (2006). *Metodologiya sostavleniya krupnomasshtabnykh agroekologicheskii oriyentirovannykh pochvennykh kart [Methodology for compiling large-scale agroecologically oriented soil maps]*. Moscow: Soil Institute. V. V. Dokuchaeva Publ., 160 p. [in Russian].
10. Simakova, M. S., Rukhovich, D. I., Belobrov, V.P., Molchanov, E. N., Chizhikova, N. P., Gavrilova, I. P., Gerasimova, M. I., & Bogdanova, M. D. (2008). *Rukovodstvo po srednemashtabnomu kartografirovaniyu pochv na osnove GIS A guide to medium-scale GIS-based soil mapping*. Moscow: Soil Institute. V. V. Dokuchaeva Publ., 343 p. [in Russian].
11. Dobermann, A., & Simbahan, G. C. (2007). Methodology for using secondary information in sampling optimisation for making fine-resolution maps of soil organic carbon. *Developments in Soil Science*, 31(13), 167–182.
12. Minasny, B., & McBratney, A. B. (2007). Latin hypercube sampling as a tool for digital soil mapping. *Developments in Soil Science*, 31, 153–166.
13. Rolecek, J., Chytry, M., Hajek, M., Lvoncik, S., & Tichy, L. (2007). Sampling design in large-scale vegetation Studies: do not sacrifice ecological thinking to Statistical purism! *Folia Geobotanica*, 42, 199–208.
14. Stumpf, F., Schmidt, K., Behrens, T., Schonbrodt-Stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., & Scholten, T. (2016). Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. *Journal of Plant Nutrition and Soil Science*, 179, 499–509.
15. Savinykh, V. P., & Tsvetkov, V. Ya. (2001). *Geoinformatsionnyy analiz dannykh distantsionnogo zondirovaniya [Geoinformation analysis of remote sensing data]*. Moscow: Kartgeocenter–Geodezizdat Publ., 228 p. [in Russian]
16. Zamyatin, A. V. (2016). Data mining methods in regional aerospace monitoring systems. *Vestnik SGUGiT [Vestnik SSUGT]*, 4(36), 74–88 [in Russian]
17. Il'ina, I. S., & et al. (1985). *Rastitel'nyy pokrov Zapadno-Sibirskoy ravniny [Vegetation cover of the West Siberian Plain]*. Novosibirsk: Nauka Publ., 249 p. [in Russian].
18. Smolentsev, B. A. (2002). *Struktura pochvennogo pokrova Sibirskikh Uvalov [The structure of the soil cover of the Siberian Uvaly]*. Novosibirsk: SB RAS Publ., 118 p. [in Russian].
19. Khozyainova, N. V. (2008). Flora and vegetation of the northern taiga of the Purovsky district of the Tyumen region (north of Western Siberia). *Vestnik ekologii, lesovedeniya i landshaftovedeniya [Bulletin of Ecology, Forestry and Landscape Science]*, 8, 27–42 [in Russian].

Получено 06.04.2021

© I. D. Makhatkov, 2021