

УДК 504.73+912+551.581.1

DOI: 10.14258/pbssm.2021018

Комплексное решение задач оценки растительных ресурсов методами ГИС и климатического моделирования

Complex solution on solving problems of estimating plant resources by GIS and climate model methods

Ваганов А. В.^{1,2}, Покалякин З. В.¹, Хворова Л. А.¹

Vaganov A. V.^{1,2}, Pokalyakin Z. V.¹, Khvorova L. A.¹

¹ Алтайский государственный университет, г. Барнаул, Россия.
E-mails: vaganov_vav@mail.ru; zpokalyakin@mail.ru; khvorovala@gmail.com

¹ Altai State University, Barnaul, Russia

² Сахалинский филиал Ботанического сада-института ДВО РАН, г. Южно-Сахалинск, Россия

² Sakhalin Branch of the Botanical Garden Institute of the Far Eastern Branch of the Russian Academy of Sciences, Yuzhno-Sakhalinsk, Russia

Реферат. В работе рассмотрены прикладные аспекты использования современных информационных технологий для точной оценки растительных ресурсов методами ГИС и климатического моделирования. Для наиболее эффективного достижения целей комплексного мониторинга и оценки растительных ресурсов авторами обсуждены и предложены ряд требований для исходных данных, факторов, влияющих на изменение ареала и результатов оценки растительных ресурсов. В качестве доступного бесплатного аналога методу коррекции пространственной неравномерности точек регистрации видов в SDMtoolbox (ArcGIS) нами предложен метод кластеризации DBSCAN, который реализован в Python библиотеке sklearn.

Ключевые слова. Большие данные, гербарий, ГИС, моделирование, ресурсная оценка, хозяйственно-ценные растения, DarwinCore, GBIF, Maxent, Phyton.

Summary. The paper considers the applied aspects of the use of modern information technologies for an accurate assessment of plant resources using GIS and climate model methods. For the most effective achievement of the goals of integrated monitoring and assessment of plant resources, the authors discussed and proposed a number of requirements for the initial data, factors affecting the change in the area and the results of the assessment of plant resources. As an available free analogue of the method for correcting the spatial unevenness of the points of registration of species in SDMtoolbox (ArcGIS), we proposed the DBSCAN clustering method, which is implemented in the Python library sklearn.

Kew words. Big data, Darwin Core, economically valuable plants, GBIF, GIS, Herbarium, Maxent, modeling, Phyton, resource assessment.

Объективная оценка растительных ресурсов возможна только при комплексном подходе, объединяющем различные прикладные и фундаментальные направления ботаники, математики, информационных технологий и возможности ГИС. Для крупной территории России именно богатая дикорастущая флора является стратегическим природным активом с высочайшим потенциалом хозяйственного значения. При этом знания о биоорганических компонентах растительного сырья, содержащихся и накапливаемых в них биологически активных веществах, имеют важное, но не решающее значение. Для решения задачи оценки растительных ресурсов необходима точная идентификация таксонов, что возможно только при определении материала систематиками-монографами, а также данные о точном (наличие координат) и наиболее полном (границы ареала) местонахождении видов на исследуемой территории. Для определения экономического вклада в решение задач продовольственной безопасности и развития биофармацевтических направлений важны оперативные мониторинговые данные, а также реальная и прогнозная оценка растительных ресурсов различных групп хозяйственно-ценных растений (лекарственных, пищевых и др.).

Установлено, что дикорастущая флора России представлена более чем 2250 видами лекарственных растений (по оценке на 2016 г.), используемых в народной и традиционной медицине (Растительные ресурсы России, 2016). Около ста видов из этого числа официально разрешены к использованию и входят в Государственный реестр. Наиболее ценные биоорганические соединения, содержащиеся в хозяйственно-ценных растениях, относятся к флавоноидам, кумаринам и дубильным веществам. Ввиду огромной территории России оценка запасов ценных растений представляется крайне затратным мероприятием. Однако современные глобальные геоинформационные технологии и возможности прогнозного моделирования стали все более широко использоваться в биологических науках для выявления закономерностей распространения и расчета потенциальных ареалов растений. При должном подходе данные методы можно применить для мониторинга и оценки растительных ресурсов хозяйственно-ценных растений.

Для комплексного подхода оценки растительных ресурсов исследуемой территории необходимо выполнить ряд требований. Первое – наиболее полная инвентаризация флоры с наличием экспертной идентификации таксонов и сведений с точными координатами мест сборов по всему ареалу исследуемых таксонов (особенно его границ). Второе – данные о пространственном распространении должны храниться в глобальных (не локальных офлайн базах данных) репозиториях в международных спецификациях и стандартах хранения и обмена данными. Третье – факторы, определяющие трансформацию ареала: климат, рельеф, биологические особенности таксона (конкуренция, стратегия размножения и адаптационная способность) должны быть измеримы и объективны. Четвертое – результаты оценки растительных ресурсов должны иметь пространственное и количественное выражение как в текущей ситуации, так и в прогнозе. По нашему мнению, единицей ресурсной оценки может служить только конкретный таксон с его биоорганическими свойствами, биологическими особенностями и исторически занятым местом в той или иной флоре, а также индивидуальным отношением к факторам, способствующим и препятствующим расселению («экологическая ниша»).

При картографировании разнообразия организмов учитывают общее количество видов в пределах изучаемой территории. Объектом картографирования может быть конкретная систематическая группа (География и мониторинг биоразнообразия, 2002). Авторы данной работы отмечают важную особенность данного метода, которая заключается в том, что за меру разнообразия в большинстве случаев принимается число видов. В ходе изучения биологического разнообразия в свое время создавалось множество подходов для разработки обзорных карт мира, показывающих сравнительные характеристики биоразнообразия (Малышев, 1975; Barthlott et al., 1999). Заложенные в отмеченных трудах основы по созданию карт, отражающих разнообразие организмов на планете Земля, где использован растровый метод (сеточные карты), картодиаграммы и изолинии, нашли выражение в современных глобальных агрегаторах данных о биоразнообразии с мощными ГИС-сервисами.

Наиболее крупным и признаваемым академическим сообществом в вопросах накопления, обмена и обработки данных о биоразнообразии планеты следует считать глобальную информационную систему по биоразнообразию (GBIF.org). GBIF является основой для облегчения оцифровки научных коллекций и для обеспечения совместимости пока еще неизвестного числа биологических баз данных, которые распространены по всему миру (Edwards et al., 2000). Накопление, хранение и обмен данными по биоразнообразию в GBIF производится в спецификации DarwinCore. Спецификация DarwinCore специально разработана международной группой «International Working Group on Taxonomic Databases For Plant Sciences (TDWG)» и представляет собой набор терминов, среди которых наиболее важными для оценки растительных ресурсов являются точные координаты. На данный момент в GBIF накоплено 1 689 176 163 наблюдений с точными геопривязками из общего числа более 1 700 000 000 записей (май 2021 г.). Спрос на данные растет экспоненциально. Записи о местонахождении живых организмов в GBIF использованы в более чем 5730 научных статей. Научные депозитарии GBIF, в число которых входят Гербарии, как правило, публикуют тип данных «OccurrenceData», которые содержат эткеточную информацию о нахождении того или иного вида в определенном месте и в определенное время, то есть предполагают наличие сведений о дате находки и географической привязке (координатах) места нахождения вида (или сбора образца) (Ваганов и др., 2019). Именно фактические места находок в GBIF по исследуемому таксону в коллекции, а также наблюдения в природе (iNaturalist, в некоторых случаях), следует рассматривать как исходные данные для реальной оценки ареала и, конечно же, в качестве основы для моделирования потенциального распространения.

В последние годы в зарубежной и отечественной литературе наиболее широкое применение для моделирования пространственного распространения видов (species distribution modeling) находит ме-

год максимальной энтропии (Guisan, Zimmermann, 2000; Anderson et al., 2003; Phillips et al., 2006; Phillips, Dudik, 2008; Олонова, Gao, 2014; Санданов, Найданов, 2015; Дудов, 2016; Корзников, 2019; Макунина и др., 2020), реализованный в программе MaxEnt (Maximum Entropy Species Distribution Modelling). Результатом работы алгоритма является модель пригодности местообитаний – карта с прогнозными вероятностями присутствия вида в каждой ячейке растра. Программный пакет Maxent позволяет строить вероятностную модель распространения анализируемого вида на основе данных о его присутствии в серии конкретных точек, не обязательно представляющих весь ареал вида. С помощью его можно оценивать комплексные взаимоотношения между видами и различными экологическими факторами, как непрерывными, так и категориальными. Потенциальный ареал опирается на климатические показатели и может быть определен с помощью биоклиматического моделирования, основанного на использовании ГИС-технологий. Под потенциальным ареалом Т. А. Работнов (1983) понимает область, где климатические условия благоприятны для произрастания вида. С. В. Дудов в своей работе (2016) напоминает о более конкретном термине, который характеризует потенциальный ареал таксона – «экологическая ниша». Данный термин ввел J. Grinnell (1917) как теоретическую основу для моделирования распространения видов в пространстве.

Климат – важнейший фактор, влияющий на формирование и распространение наземных экосистем (Дудов, 2016). Однако для достоверной модели потенциального ареала исследователи используют не только климатические, но топографические параметры, а также данные ДЗЗ. Отдельного внимания заслуживает работа, посвященная изучению видов лекарственных растений (на примере Кунгурской островной лесостепи Пермского края), приуроченных к различным типам почв и ботанико-географическим зонам (Турышев и др., 2015). Наиболее часто в практике биологических работ по прогнозным моделям задействуют биоклиматические параметры (BIO1-BIO19) и сведения о характере рельефа с проекта WorldClim (<https://www.worldclim.org>). Для более точной прогнозной оценки и подробной карты используют разрешение 2.5 arc-minutes (5 км на пиксель). В целях повышения точности получаемой модели определяют обучающую и тестовую выборки 75 % на 25 % соответственно. Состоятельность модели определяют индексом пригодности для прогнозного местообитания таксона (AUC). Как правило, оценка вклада каждой переменной (BIO1-BIO19) производится по тесту «jackknife» (Phillips et al., 2006). Результаты моделирования в программе MaxEnt можно отобразить в программе DIVA-GIS.

Анализ исходных данных на практике показал их пространственную неравномерность, которая заключается в том, что чаще данные собираются непосредственно вблизи инфраструктуры (населенные пункты, дороги и пр.). В связи с этим точки регистрации (ТР) сконцентрированы возле объектов антропогенной трансформации в большей степени нежели в «естественной части» ареала, следовательно неравномерно по всему исследуемому ареалу. Искусственное сгущение и разрежение ТР неизбежно приводит к смещению результатов исследования. Для того чтобы избежать этого эффекта, необходимо провести коррекцию пространственной неравномерности мест произрастания путем фильтрации и пространственного разделения данных. Для этого можно использовать программу SDMtoolbox (Brown, 2014), являющуюся приложением к программе ArcGIS 10.5.

В целях поиска доступного бесплатного аналога для коррекции пространственной неравномерности ТР нами в данном обзоре предложен метод кластеризации DBSCAN (Density Based Spatial Clustering of Applications with Noise), который реализован в Python библиотеке sklearn (<https://scikit-learn.org/https://scikit-learn.org/>). DBSCAN – это алгоритм кластеризации данных, который группирует тесно расположенные точки и обозначает как шум точки, которые далеко расположены. Это один из самых популярных методов кластеризации и хорошо зарекомендовавший себя в научно-исследовательской деятельности. DBSCAN имеет ряд преимуществ: 1) данный метод не требует изначально заданного количества кластеров; 2) кластеры могут быть произвольной формы; 3) различает понятия шум и выбросы; 4) для кластеризации необходимо задать два параметра: ϵ – максимальное расстояние между двумя соседними точками и minPts – минимальное количество точек в кластере (ϵ выбирается в зависимости от пространственного разрешения биоклиматических данных, а minPts равен единице). DBSCAN не чувствителен к порядку точек в данных. В итоге данный метод объединил точки, которые находятся в области с высокой плотностью, а точки в области с низкой плотностью оставил как есть (рис.). В результате данные стали менее неравномерны, что, к тому же, повысило точность работы модели по сравнению с программой SDMtoolbox.

С ростом объема данных для моделирования при ограниченности мощностей персональных компьютеров все большее развитие получают виртуальные лаборатории. Так, для прогнозного и реального биоклиматического моделирования сегодня находит популярность виртуальная лаборатория

«Biodiversity and Climate Change Virtual Laboratory, BCCVL» (Hallgren et al., 2016). В BCCVL можно выбрать все имеющиеся на данный момент алгоритмы SDM, включая MaxEnt.

После прогнозной оценки потенциального ареала вида важно оценить его площадь. Как в случае с DIVA-GIS, так и BCCVL, в результатах имеются только данные по состоятельности модели и карта с долей вероятности встречаемости вида. Для ресурсной оценки важно получить сведения о количественном значении площади потенциального ареала хозяйственно-ценного таксона. Для возможности объединить всю цепочку действий по ресурсной оценке нами в последующих исследованиях планируется разработать алгоритм и интерфейсное решение по оценке площади реального и прогнозного ареала исследуемого таксона с набором дополнительных сервисов.

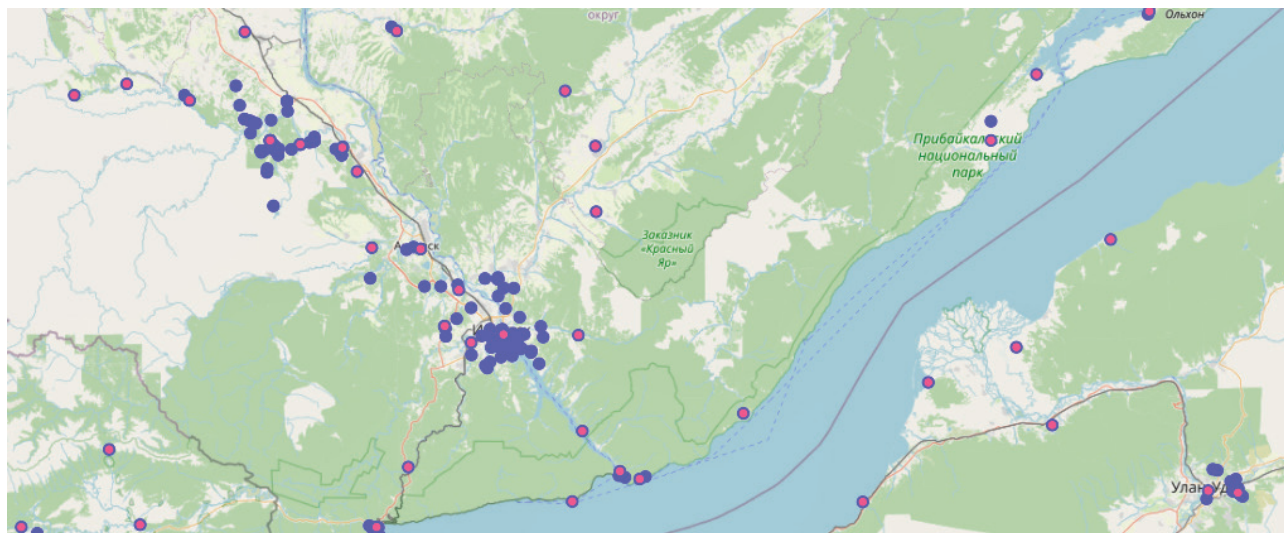


Рис. Результаты коррекции пространственной неравномерности данных точек регистрации иван-чая узколистного методом DBSCAN: синие точки – исходные, красные – после обработки (*Epilobium angustifolium L.*, $e = 10$, $minPts = 1$).

Благодарности. Работа выполнена при поддержке государственного задания Алтайского государственного университета (проект №FZMW-2020-0003) и Программы поддержки научно-педагогических работников ФГБОУ ВО «Алтайский государственный университет» (договор №11/21-ВГ от 23.04.2021).

ЛИТЕРАТУРА

Ваганов А. В., Шмаков А. И., Гудкова П. Д. Глобальные данные о фиторазнообразии Алтайской горной страны, представленные в мировых научных депозитариях // Проблемы ботаники Южной Сибири и Монголии, 2019. – Т. 18. – С. 222–227. DOI: 10.14258/pbssm.2019045

География и мониторинг биоразнообразия. / Колл. авторов. – М.: Изд-во Научного и учебно-методического центра, 2002. – 432 с.

Дудов С. В. Моделирование распространения видов по данным рельефа и дистанционного зондирования на примере сосудистых растений нижнего горного пояса хр. Тукурингра (Зейский заповедник, Амурская область) // Журнал общей биологии, 2016. – Т. 77, № 1. – С. 16–28.

Корзников К. А. Климатическое моделирование местообитания *Kalopanax septemlobus* и *Phellodendron amurense* var. *sachalinense* в островном секторе Дальнего Востока России // Известия РАН. Серия биологическая, 2019. – № 6. – С. 648–657. DOI: 10.1134/s1062359019040083

Макунина Н. И., Егорова А. В., Писаренко О. Ю. Построение потенциальных ареалов растительных сообществ с целью ботанико-географического районирования (на примере лесов Тувы) // Сибирский экологический журнал, 2020. – № 4. – С. 517–524. DOI: 10.1134/s1995425520040095

Мальшиев Л. И. Количественный анализ флоры: пространственное разнообразие, уровень видового богатства и репрезентативность участков обследования // Бот. журн., 1975. – Т. 60, № 11. – С. 1537–1550.

Олонова М. В., Gao X. Потенциальные возможности распространения адвентивного растения *Poa compressa* L. в Сибири // Вестник Томского государственного университета. Биология, 2014. – № 4(28). – С. 56–69. DOI: 10.17223/19988591/28/4

Работнов Т. А. Фитоценология. 2-е изд. – М.: Изд-во МГУ, 1983.– 296 с.

Растительные ресурсы России: дикорастущие цветковые растения, их компонентный состав и биологическая активность / РАН, отв. ред. А. Л. Буданцев. – СПб.-М.: Тов-во науч. изд. КМК, 2008–2016. Т. 1–7.

Санданов Д. В., Найданов Б. Б. Пространственное моделирование ареалов Восточно-азиатских видов растений: современное состояние и динамика под влиянием климатических изменений // Растительный мир Азиатской России, 2015. – № 3(19). – С. 30–35.

Турьишев А. Ю., Согрина А. Н., Яковлев А. Б. Комплексная оценка популяций дикорастущих лекарственных растений в пределах Кунгурской островной лесостепи // Современные проблемы науки и образования, 2015. – № 4. – С. 571–580.

Anderson R. P., Lew D., Peterson A. T. Evaluating predictive models of species' distributions: criteria for selecting models // Ecological Modelling, 2003. – Vol. 162. – P. 211–232. DOI: 10.1016/s0304-3800(02)00349-6

Barthlott W., Biedinger N., Braun G., Feig F., Kier G., Mutke J. Terminological and Methodological Aspects of the Mapping and Analysis of the Global Biodiversity // Acta Bot. Fennica, 1999. – Vol. 162. – P. 103–110.

Brown J. L. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses // Methods in Ecology and Evolution, 2014. – Vol. 5, № 7. – P. 694–700.

Edwards J. L., Lane M. A., Nielsen E. S. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop // Science, 2000. – Vol. 289. – P. 2312–2314. DOI: 10.1126/science.289.5488.2312

Grinnell J. The niche-relationships of the California Thrasher // The Auk., 1917. – Vol. 34. – P. 427–433. DOI: 10.2307/4072271

Guisan A., Zimmermann N. E. Predictive habitat distribution models in ecology // Ecological Modelling, 2000. – Vol. 135. – P. 147–186. DOI: 10.1016/s0304-3800(00)00354-9

Hallgren W., Beaumont L., Bowness A., Chambers L., Graham E., Holewa, H., Laffan S., Mackey B., Nix H., Price J., Vanderwal J., Warren R., Weis G. The Biodiversity and Climate Change Virtual Laboratory: Where ecology meets big data // Environmental Modelling and Software, 2016. – № 76. – P. 182–186. DOI: 10.1016/j.envsoft.2015.10.025

Phillips S. J., Anderson R. P., Schapire R. E. Maximum entropy modeling of species geographic distributions // Ecological Modelling, 2006. – Vol. 190. – P. 231–259. DOI: 10.1016/j.ecolmodel.2005.03.026

Phillips S. J., Dudik M. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation // Ecography, 2008. – Vol. 31. – P. 161–175. DOI: 10.1111/j.0906-7590.2008.5203.x