

УДК 57.034; 519.254

СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ПРОСТРАНСТВЕННЫХ РАСПРЕДЕЛЕНИЙ В ДРЕВЕСНЫХ РАСТЕНИЯХ

©Кислов¹ Д.Е., Прилуцкий¹ А.Н., Фисенко² М.И.

¹Ботанический сад-институт ДВО РАН, Владивосток

²Уссурийская астрофизическая обсерватория ДВО РАН, с. Горнотаежное
a.priluckiy@mail.ru, kisl@mail.ru, mihail_fisenco@mail.ru

В статье представлен комплекс статистических подходов направленных на установление особенностей пространственной организации точечных мозаик, образуемых растительным покровом. Основное внимание работы ориентировано на проверку статистических гипотез о значимости особенностей пространственных распределений против гипотезы о случайном характере наблюдаемых распределений.

Ключевые слова: статистический анализ, пространственная организация растительного покрова.

Современные статистические подходы представляют собой важный исследовательский инструмент при изучении характеристик распределений растительности в сообществах и анализе их динамики. В зависимости от характера той или иной прикладной задачи среди используемых статистических подходов может быть выделен важный их комплекс, образующий так называемые методы анализа пространственно-распределенных данных.

Неоднородность пространственного размещения деревьев является одним из определяющих динамику факторов, который необходимо учитывать при исследовании лесных экосистем и управления ими. Структура древостоя определяет условия произрастания и непосредственно влияет на разнообразие биологических объектов лесного сообщества (Василевич, 1989; Fries et al. 1997; Angelstam, 1998). Поэтому описание особенностей пространственного размещения деревьев в лесной экосистеме является необходимым этапом при построении моделей его пространственно-временного развития.

Настоящая работа посвящена разработке статистических подходов, относящихся к области анализа пространственно-распределенных данных. Ретроспективный анализ проблематики исследований в этой области позволяет характеризовать данное направление статистически как довольно интенсивно развивающееся в последнее время, что во многом обусловлено развитием вычислительной техники. На заре формирования статистических методов формулировка эффективных статистических критериев для проверки гипотез о характере пространственных распределений представляло значительные трудности ввиду, прежде всего, сложности вычислительного плана. Проблема со-

стояла в том, что вводимые при исследовании пространственных данных статистики имели сложные распределения, что в свою очередь требовало осуществления процедур имитационного статистического моделирования. Использование последнего стало возможным только при соответствующем уровне развития доступного вычислительного потенциала.

Остановимся на основных этапах развития статистических подходов, используемых для анализа пространственно-распределенных данных, и дадим оценку предлагаемым в рамках данной работы методам в контексте современного развития предметной области.

История подходов, используемых для описания особенностей пространственного размещения объектов в экосистемах, позволяет выделить два этапа в их формировании. Первый этап характеризуется развитием методов, основанных в большей степени на базовых представлениях теории вероятности. К ним, в частности, относится подход, использующий критерий (Василевич, 1989), а также методы оценки особенностей пространственного размещения объектов, используемые в работах (Svedberg, 1922; Ashby, 1935; Bray, 1962; David, Moore, 1954).

Потребность в познании причин реализации конкретных размещений объектов привела к необходимости учета взаимных расстояний между ними, что совместно с ростом вычислительных возможностей определило предпосылки дальнейшего развития статистических методов анализа пространственно-распределенных данных (далее, говоря о статистических методах, будем иметь в виду только методы анализа пространственно-распределенных данных). Одним из первых шагов в этом направлении явились работы, в которых предложены подходы, допускающие интерпретации в свете представлений теории случайных процессов (Clark, Evance, 1954; Hopkins, 1954; Pielou, 1959; Ripley, 1979). Данными работами характеризуется начало следующего этапа в развитии статистических подходов, особенностью которых является наличие в них глубоких интерпретаций в рамках

теории случайных процессов (Ripley, 1979). На сегодняшний день эти методы активно используются при решении разнообразных научных и практических задач, связанных с анализом особенностей пространственного размещения объектов и хорошо представлены в тематическом программном обеспечении (R; Ripley, 2004).

Если первоначально предложенные методы были ориентированы на работу с однородными объектами, то современные их обобщения позволяют выполнять обработку объектов, принадлежащих разным группам, и, таким образом, способны учитывать возрастную структуру в сложении древостоя или особенности образования видовых ассоциаций в лесном сообществе. Основные вопросы, на которые дают ответы эти статистические подходы, – характеристика соответствия наблюдаемого пространственного распределения объектов некоторому гипотетическому (в частности, проверка гипотезы о случайности пространственных размещений объектов) либо установление факта наличия/отсутствия пространственных отношений в неоднородных совокупностях объектов.

Отличительной особенностью предлагаемых в настоящей работе подходов является наличие у них вполне определенных физических интерпретаций, позволяющих локально оценить характер динамики пространственно-временного развития древесного сообщества. Статистический критерий для проверки гипотез о специальном характере распределения объектов строится в условиях нулевой гипотезы о случайном (пуассоновском) характере пространственного распределения объектов.

Основные представления

В основе предлагаемых подходов лежит принцип локального изучения пространственных распределений точечных мозаик, нашедший в последнее время довольно широкое распространение при анализе пространственно-распределенных данных (Ripley, 2004). По типу устанавливаемых пространственных распределений рассматриваются три статистических подхода: 1) статистический подход, позволяющий установить преобладание кольцевых структур в наблюдаемом распределении объектов; 2) подход, позволяющий обнаружить элементы периодичности в наблюдаемом распределении; 3) подход, позволяющий выделить линейные структуры.

Решение задачи о статистической значимости во всех предлагаемых подходах осуществляется с использованием метода статистического имитационного моделирования в условиях нулевой гипотезы,

так называемого метода Монте-Карло (Ripley, 1979; Diggle, 1983; Ripley, 2004).

Рассмотрим математические аспекты предлагаемых подходов более подробно.

Исследование кольцевых структур

При индикации кольцевых структур рассматривается интегральный план растительного покрова, получаемый воображаемым наложением разбиения исходной пробной площади на квадраты одинакового размера. Зондирующее окно в этом случае имеет кольцевую форму (рис. 1). Внешний и внутренний радиусы кольца выбираются переменными, что обеспечивает индикацию кольцевых структур различного размера.

В качестве статистики, характеризующей степень размещения точечных объектов по кругу, используется значение H -функции, вычисленное согласно встречаемости объектов внутри секторов, получаемых разбиением исходного кольца на равное число частей. Альтернативой такой статистики является нормированная на максимальное значение H -функции статистика, значения которой уже варьируются от 0 до 1 (Прилуцкий и др., 2012). В случае максимального значения H -функции при текущих внутреннем и внешнем радиусах можно утверждать, что в наблюдаемом кольце объекты распределены равномерным образом, констатируя, таким образом, что имеет место кольцевая структура. Чувствительность метода определяется количеством секторов, формирующих разбиение зондирующего окна и используемых при количественной характеристике равномерности распределения объектов, оказавшихся внутри кольца. В частности, опыт вычислений показывает, что вполне приемлемым является количество секторов равным 15. Возможность индикации различных кольцевых структур обеспечивается блужданием зондирующего окна по интегральному плану (обозначенным на рис. 1 стрелочками).

При практическом исследовании участка пробной площади конечных размеров (например, прямоугольной формы), может использоваться процентное отношение обнаруженных кольцевых структур к общему числу используемых для зондирования пробной площади окон. В этом случае, задавшись определенным критическим значением для наблюдаемой статистики, представляется целесообразным вычислить при данных значениях внутреннего и внешнего радиуса указанное выше отношение, пробегая зондирующим окном по всей пробной площади. Таким образом, проделав вычисления для всех интересующих размеров колец

можно построить рисунок, характеризующий число обнаруженных кольцевых структур на пробной площади. Пример такой иллюстрации для участка лесного массива, расположенного на территории Горнотаежной станции ДВО РАН, приводится в работе А.Н.Прилуцкого с соавторами (Прилуцкий и др., 2012), где и дается проблемно-ориентированные интерпретации обнаруженных структур.

Обнаружение периодичности в пространственном размещении объектов

Методика обнаружения периодичности в размещении точечных объектов основывается на широко распространенной для задач подобного рода технологии дискретного преобразования Фурье (в работе для выполнения преобразования Фурье используется библиотека программ SciPy). Дискретное преобразование Фурье используется для выделения периодических гармоник как одно-, так и двумерных (например, при анализе изображений) сигналов и представляет собой удобное средство для анализа пространственного размещения объектов.

Суть предлагаемого подхода заключается в проецировании имеющихся данных на некоторое предварительным образом выбираемое направление с последующим применением процедуры дискретного преобразования Фурье к уже одномерному набору. Результатом такой процедуры является множество амплитуд разночастотных гармоник, которое и подлежит визуализации. Для оценки значимости факта обнаруженной периодичности предварительно (на основе метода статистических испытаний) строится критическая область критерия (Эфрон, 1990). Технология построения критических областей совпадает с традиционной для данной предметной области (Diggle, 1983; Ripley, 2004). Она реализуется на основе предварительно заданного числа статистических испытаний в методе имитационного моделирования и желаемого уровня значимости критерия. Выход за указанную критическую область интерпретируется как значимое отклонение от нулевой гипотезы (в качестве нулевой гипотезы при этом обычно выбирается предположение, что наблюдаемая структура является результатом пуассоновского точечного процесса, т. е. является случайным). Пример использования данной технологии с выводами при анализе горизонтальной структуры древостоя можно найти в работе (Прилуцкий и др., 2012).

Идентификация линейных структур

Обнаружение линейных структур осуществляется по тому же принципу, что и кольцевых. Задается зондирующее окно в форме полосы (рис. 2),

характеризующейся своей шириной и положением. В качестве меры линейной упорядоченности при идентификации линейных структур могут быть использованы хорошо известные показатели, характеризующие степень согласованности данных линейным регрессионным зависимостям. Таковым является коэффициент детерминации, вычисленный для данных, попавших в текущее зондирующее окно в форме полосы (Крамер, 1975).

Для решения этой задачи в качестве статистики также может использоваться сумма квадратов отклонений координат объектов от прямой линии, определяемой как решение задачи отыскания главных компонент (Hardle, 2007). После того, как меры для количественного описания линейной упорядоченности определены, вычислительный эксперимент может быть организован следующим образом. Очевидно, что положение каждого зондирующего окна в этом случае однозначно определяется двумя параметрами – координатой пересечения его оси симметрии (x) и, например, оси абсцисс – являющейся осью координат, в которых задан исходный набор данных и углом наклона (ϕ) оси симметрии к какой-либо оси базовой системы координат (для определенности, также оси абсцисс). Ширина зондирующего окна (h) определяет чувствительность метода и выбирается из проблемно-ориентированных соображений с учетом предполагаемых интерпретаций.

В зависимости от значений этих параметров зондирующее окно будет выделять из исходного набора данных подвыборки и тестировать их на линейную упорядоченность, присваивая каждому «окну» числовую характеристику – количественно выражающую степень линейной упорядоченности данных, попавших в него. Таким образом, осуществив программу вычислений при различных положениях зондирующего окна, оказывается возможным визуализировать получившуюся картину, совместив ее с исходным планом пробной площади. Значение статистики для каждой позиции окна можно представить в виде поверхности, либо ограничиться цветовыми тонами – более высоким значениям статистик, характеризующим большую линейную упорядоченность, задавать светлые тона, а более низким – соответственно более темные. Пример применения такого подхода для идентификации линейно-упорядоченных структур на пробных площадях, заложенной в лесах Горнотаежной станции ДВО РАН (Приморский край, Усурийский район), и иллюстрации статистик, построенных на основе коэффициента детерминации, даны в работе (Прилуцкий и др., 2012).

Важно отметить, что существующий статистический аппарат предлагает множество количественных мер, использование которых возможно при идентификации факта выраженной линейной упорядоченности объектов. При выполнении практические исследования были проведены сравнительными методами по использованию коэффициента детерминации и суммы квадратов отклонений в методе главных компонент как показателей выраженности линейной упорядоченности (Прилуцкий и др., 2012). Значимых качественных отличий между данными экспериментами обнаружено не было (за исключением единиц измерения шкалы статистики). Учитывая данный факт, на практике (цит. выше) нами был использован коэффициент детерминации как более распространенный в биолого-статистических исследованиях.

В заключение изложения основной части работы остановимся на аспектах практической реализации данных подходов. В настоящее время существует достаточно большое множество программных сред, предлагающих большой спектр методов статистического анализа пространственно-распределенных данных. Среди них следует отметить продукты компании ESRI (коммерческие, и довольно дорогие) и множество пользовательских библиотек, реализующих методы геостатистики в различных языках программирования (R, Python и др.). Реализация предложенных подходов может быть выполнена с использованием любой из этих статистических сред. При выполнении работы вычисления проводились с использованием интегрированной посредством языка программирования Python вычислительной среды на базе пакета научных вычислений SciPy (Scipy, 2011; Прилуцкий и др., 2012). При выборе среды реализации, заметим, имеет значение предоставляемый ею функционал, который должен соответствовать требованиям, предъявляемым к решению прикладной задачи и обеспечивать функции работы с пространственно-распределенными данными.

Заключение

Проблемно-ориентированные интерпретации особенностей распределения объектов, обнаруженных с помощью предлагаемых методов, зависят от специфики решаемых задач. Формулировка данных интерпретаций не являются предметом данной работы, ориентированной исключительно на описание статистических методик без указания области их приложения. Вместе с тем, опыт решения задач, связанных с анализом распределений в сообществах древесных растений, позволяет утверждать, что данные методы могут быть рекомен-

дованы для обнаружения автоволновых процессов, имеющих место пространственно-распределенных биологических системах (Васильев, 1987).

Л и т е р а т у р а

Василевич В.И. Статистические методы в геоботанике – Л.: Наука, 1969. – 232 с.

Васильев В.А., Романовский Ю.М., Яхно В.Г. Автоволновые процессы. /Под ред. Д.С. Чернавского. – М.: Наука, 1987. – 240 с.

Кобзарь А.И. Прикладная математическая статистика для инженеров и научных работников. – М.: Физматлит, 2003. – 853 с.

Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с.

Прилуцкий А.Н., Кислов Д.Е., Фисенко М.И. Механизмы упорядочивания горизонтальной структуры влажных дубняков // Бюллетень Ботанического сада-института ДВО РАН, [электронный журнал, 2012].

Свирижев Ю.М. Нелинейные волны, диссипативные структуры и катастрофы в экологии. – М.: Наука, 1987. – 368 с.

Эфрон Б. Нетрадиционные методы математической статистики. – М.: Финансы и статистика, 1988. – 263 с.

Angelstam P. Maintaining and restoring biodiversity in European boreal forests by developing natural disturbance regimes // *Journal of Vegetation Science*, 1998. Vol. 9. – P.593–602.

Ashby E. Quantitative analysis of vegetation // *Ann. Bot.*, 1935. Vol. 49, № 196. – P. 779–802.

Bray J.R. Use of non-area analytic data to determine species dispersions // *Ecology*, 1962. Vol. 43, № 2. – P. 328–333.

Clark P.J., Evans F.C. Distance to nearest neighbor as a measure of spatial relationships in populations // *Ecology*, 1954. Vol. 35. – P. 445–453.

David F.N., Moore P.G. Notes on contagious distributions in plant populations // *Ann. Bot.*, 1954. Vol. 18, № 69. – P. 47–53.

Diggle P.J. Statistical analysis of spatial point patterns. – London, UK, Academic Press, 1983. – 148 p.

Fries C., Johansson O., Petterson B., Simonsson P. Silvicultural models to maintain and restore natural stand structures in Swedish boreal forests // *Forest Ecology and Management*, 1997. Vol. 94. – P. 89–103.

Hardle W., Simar L. Applied multivariate statistical analysis. – Springer-Verlag, 2003. – 498 p.

Hopkins B. A new method for determining the type distribution of plant individuals // *Ann. Bot.*, 1954. Vol. 18, № 2. – P. 213–227.

Pielou E.C. The use of point-to-plant distances in the study of the pattern of plant populations // *Journal. Ecol.*, 1959. Vol. 47, № 3. – P. 607-613.

Python: Programming language // Электронный ресурс: <http://www.python.org/>

R: Programming language for statistical computations // Электронный ресурс: <http://www.r-project.org/>

Ripley B.D. Spatial statistics. – John Willey and son's, 2004. 252 p.

Ripley B.D. Tests of “randomness” for spatial point patterns // *Journal of the Royal Statistical Society*, 1979. В 41. P. 368–374.

Svedberg T. Ett Bidrag till de statistiska metodernas användning inom växtbiologien // *Svensk Botanisk Tidskrift*, 1922. Vol. 16. – P.1–8.

Scipy: Scientific Python // [Электронный ресурс, 2011] <http://scipy.org>

Wallenius T., Kuuluvainen T., Heikkilä R., Lindholm T. Spatial tree age structure and fire history in two old-growth forests in eastern Fennoscandia // *Silva Fennica*, 2002. Vol. 36, № 1. – P. 185–199.

Статья поступила в редакцию 25 марта 2012 г., принята к печати 20 апреля 2012г.

STATISTICAL APPROACHES FOR STUDYING SPATIAL POINT PATTERNS IN STANDS

Dmitry E. Kislov¹, Alexander N. Prilutsky¹, Mikhail I. Fisenko²

¹ - *Botanical garden-institute of the Far Eastern Branch of Russian Academy of Sciences, Vladivostok*

² - *Astrophysical station of the Far Eastern Branch of Russian Academy of Sciences, s. Gornotayezhnoe*

Key words: vegetation cover dynamics, spatial point pattern, Primorskiy krai, statistics

The article presents a set of statistical approaches to study features of spatial point patterns in vegetation. Basic work attention is devoted to verification of statistical hypotheses for significance of empirical spatial point patterns for stochastic nature of the ones.

II. 2. Bibl. 24.