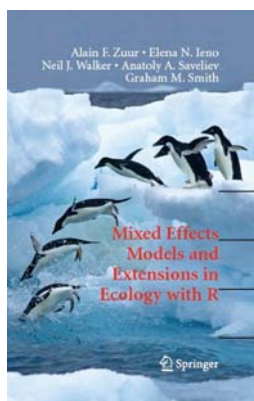


КРИТИКА И БИБЛИОГРАФИЯ



Zuur Alain F., Ieno Elena N., Walker Neil, Saveliev Anatoly A., Smith Graham M. Mixed Effects Models and Extensions in Ecology with R. – Berlin et al.: Springer Sci., 2009. – 574 p. (Ser.: Statistics for Biology and Health)

(Зуур А.Ф., Йено Е.Н., Уолкер Н., Савельев А.А., Смит Г.М. Модели со смешанными эффектами и их расширения в экологии [с использованием языка программирования R]. – 2009.)

Для многих естествоиспытателей применение статистических методов напоминает ныряние с головой в холодную воду, что прекрасно проиллюстрировано на обложке рецензируемой книги. Действительно, еще А.А. Любищев более 60 лет тому назад писал: «...на биофак, как и на гуманитарные факультеты, идут преимущественно по признаку совершенной невинности в математике» (20 августа 1950 г. [1, с. 59]). Сегодня эта «невинность в математике» подкрепляется и новой системой аттестации (пресловутый ЕГЭ): старшеклассники учатся избирательно и готовятся только к тем предметам, по которым они будут сдавать экзамены – выпускные в школе и вступительные в вузе. И если вступительные экзамены на «биофаке» – биология, химия, русский язык и литература – по минимуму, то математику приходится осваивать уже в вузе [32]. А это – не всегда просто...

Как отмечают сами авторы рецензируемой книги, она является продолжением («сиквелом») работы, которая вышла на два года раньше [3]. Это дало возможность расширить представления и показать особенности использования регрессионных моделей с применением языка программирования R для анализа экологических данных. R – язык программирования и среда для статистической обработки данных и работы с графикой (с включением математической символики), который был разработан в 1993 г. новозеландскими программистами Россом Айхэкой (Ross Ihaka) и Робертом Джентльменом (Robert Gentleman), что и объясняет название языка (первая буква их имен – R). Сегодня этот язык широко используется для статистического программного обеспечения и анализа данных и фактически стал стандартом для статистических программ. Это не только статистика в узком смысле слова, но и первичный анализ (графики, таблицы сопряженности и пр.), и продвинутое математическое моделирование.

Однако основная вычислительная мощь R лучше всего проявляется именно при статистическом анализе: от вычисления средних величин до вейвлет-преобразований временных рядов. Важным методологическим вопросом является определение характера зависимости между факторами и результативными показателями: функциональная она или стохастическая, прямая или обратная, прямолинейная или криволинейная и т. д. Здесь используются теоретико-статистические критерии, практический опыт, а также способы сравнения параллельных и динамических рядов, аналитических группировок исходной информации, графические методы и др. Все это, в той или иной степени, присутствует и в языке программирования R.

Во «Введении» авторы отвечают на ряд вопросов: Зачем эта книга? Как использовать её, если Вы – преподаватель? Что не сделано и почему? Как цитировать R и связанные с ним пакеты программ? Здесь особо следует отметить ответы на третий вопрос, так как они обезоруживают рецензента, который хотел бы видеть в книге не то, что есть, а то, с его точки зрения, что было бы необходимо в нее включить. Авторы честно признаются в определенном «техническом» упрощении подготовки данных для использования их в R-процедурах («это означает, что мы проводим читателя через тот же самый процесс подготовки данных, которым они должны были бы довести свои собственные данные до конца», с. 7), а также в сознательном отказе от излишней формализации и сокращении «математических деталей» в тексте (последнее преодолевается отсылками к ряду пособий по математической статистике).

Первая часть книги (главы 2–13, почти 60% текста) – это своеобразное логическое введение в регрессионное моделирование, в многочисленные методы и пакеты программ: GLM (generalized linear modeling), GAM (generalized additive modeling), GEE (generalized estimation equations), GLMM (generalized linear mixed models) и многие другие. Подробно обсуждаются ограничения линейного регрессионного анализа применительно к экологическим данным (анализ исходных данных на нормальность распределения, гетерогенность и пр.), учет в множественных линейных регрессионных моделях разного рода «нелинейностей», использование в регрессионном моделировании векторного и матричного анализа, пространственные и временные зависимости исходных данных и др. Иными словами, авторы демонстрируют почти весь спектр реальных проблем, с которыми сталкивается естествоиспытатель в мире живых существ и которые, зачастую, не могут быть решены традиционными методами.

Стиль изложения этих глав очень удобен*: почти все они начинаются с четких объяснений специальных проблем, далее следует, как уже отмечалось, «укороченное» описание математического аппарата и завершает все демонстрация работоспособности данного метода или приема обработки экологических данных. При этом все сопровождается подробными комментариями к каждому шагу обработки, что позволяет читателю пройти весь путь принятия статистически корректного решения и лучше понять ход рассуждений авторов «на языке» программирования R, лежащих в основе используемых статистических методов.

* Не удержусь, и процитирую нашу монографию [4], в которой изложение конкретных методов (в том числе и регрессионных) идет по аналогичной схеме: «формулировка задачи», «математический лист» и «результаты расчетов».

Вторая часть работы (главы 14–23) посвящена весьма подробному описанию конкретных экологических примеров, которые проанализированы с помощью предлагаемых методов. Только одно перечисление этих десяти экологических систем и объектов дает представление об охвате проблем, решаемых с помощью R-моделирования:

- оценка трендов изменений численности арктических птиц (два вида пингвинов [*Aptenodytes forsteri* и *Pygoscelis adeliae*] и буревестник [*Daption capense*]) в зависимости от изменений климата;
- влияние на характеристики сельскохозяйственных почв крупномасштабных дренажных работ в Шотландии;
- отрицательные биномиальные распределения в анализе гибели на дорогах амфибий (на юге Португалии);
- смешанное регрессионное моделирование распределения глубоководных морских биолюминесцирующих организмов;
- аналогичное моделирование временных (сезонных и многолетних) изменений морского фитопланктона;
- аналогичное моделирование воздействия инфекционного заболевания (AFB – American Foulbrood) на личинки медоносных пчел;
- определение возраста животных шести видов семейства китовых;
- использование GLMM для описания специального распределения коал (*Phascolarctos cinereus*) в фрагментарных ландшафтах восточной Австралии;
- сравнение ряда R-моделей применительно к описанию популяции барсука обыкновенного (*Meles meles*) в юго-западной Англии;
- временные корреляции обилия обыкновенного тюленя (*Phoca vitulina*) в заливе Бантрни и реке Кенмэр в Ирландии.

Эти примеры описаны очень доступно, можно сказать, даже научно-популярно. Акцент на эвристических основах процессов моделируемых экосистем, а не только (и не сколько) на технике собственно моделирования, помогает читателю понять, «почему» это делается именно «так», а не вслепую следовать за простым списком рекомендаций «как» это делать, используя язык R. Иными словами, эти разделы будут полезны как практикующим экологам и аспирантам, которые владеют лишь азами моделирования, так и модельерам-профессионалам.

Завершает монографию приложение («Некоторые предварительные сведения о линейном регрессионном анализе с совокупностью примеров моделирования»), список литературы (209 наименований, из которых почти 70% опубликовано после 2000 г.) и обобщенный предметный и авторский указатель.

Подводя итог обсуждению этой книги, следует с удовлетворением констатировать, что весь спектр формальных описаний инструментов программного обеспечения R, графических иллюстраций, практических примеров, все это очень качественно позволяет авторам объяснить и продемонстрировать эффективную работоспособность самых важных многомерных статистических методов для анализа экологической информации. Более того, эта книга способствует лучшему пониманию количественных (статистических) методов, применяемых авторами к реальным экосистемам. При этом не следует забывать, что чаще всего эмпирико-статистические модели (к которым относятся и регрессионные

уравнения) являются «сырьем» и обоснованием подходов к построению моделей других типов (в первую очередь, имитационных). И здесь, как нельзя, кстати, подходят слова В.И. Василевича [5, с. 218], которыми он заканчивает свою монографию: «Не следует думать, что тот математический аппарат, который в настоящее время используют геоботаники (*и экологи*. – Г.Р.), дает наилучшее решение. С одной стороны, его необходимо углублять и дополнять, совершенствовать уже существующие методы, а с другой стороны, необходимо искать принципиально новые пути решения проблем. Авторитет старых методов так же опасен, как и авторитет старых теорий».

Приходится констатировать, что за последние 50 лет развития статистических методов в экологии успехов значительно меньше, чем за весь вековой период становления этих методов. Наверное, это вполне объяснимо. Некоторые методы «устоялись» и стали обязательными (прямой градиентный анализ или рассмотренный в рецензируемой монографии язык R**), другие остаются экзотическими и используются в основном авторами (распознавание образов или самоорганизация при прогнозировании). Развитие количественных методов пошло в направлении разработки математических моделей и создания разного рода баз знаний и данных. Этот процесс – не простой, и ожидать здесь быстрых успехов не приходится.

Литература

1. Любищев А.А. Мысли о многом / Сост. О.П. Орлицкая. – Ульяновск: Изд-во Ульян. гос. пед. ун-та, 1997. – 272 с.
2. Розенберг Г.С. А.А. Любищев о применении математики и биометрии в биологии (с экологическими комментариями и пояснениями) // Теоретические проблемы экологии и эволюции (Пятые Любищевские чтения). – Тольятти: Кассандра, 2010. – С. 5–12.
3. Zuur A.F., Ieno E.N., Smith G.M. Analysing Ecological Data. – Berlin et al.: Springer Sci., 2007. – 680 p. (Ser.: Statistics for Biology and Health).
4. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы, критерии, решения: в 2 кн. – М.: Наука, 2005. – Кн. 1. – 281 с.; Кн. 2. – 337 с.
5. Василевич В.И. Статистические методы в геоботанике. – Л.: Наука, 1969. – 232 с.
6. Савельев А.А., Мухарамова С.С., Пилюгин А.Г. Основные понятия языка R. – Казань: Изд-во Казан. ун-та, 2007. – 36 с.
7. Савельев А.А., Мухарамова С.С., Пилюгин А.Г. Использование языка R для статистической обработки данных. – Казань: Изд-во Казан. ун-та, 2007. – 30 с.

Г.С. Розенберг,
Институт экологии Волжского
бассейна РАН, г. Тольятти

** Этот процесс наблюдается и в отечественной экологии, чему существенно способствует активная просветительская деятельность одного из авторов рецензируемой монографии – профессора теперь уже Казанского (Приволжского) федерального университета А.А. Савельева (см., например, [6, 7]).