

УДК 630*:51

А.Н. Смольянов, А.И. Ревин, Н.Б. Старостюк

Смольянов Анатолий Николаевич родился в 1950 г., окончил в 1974 г. Воронежский лесотехнический институт, кандидат сельскохозяйственных наук, доцент кафедры лесной таксации и лесоустройства Воронежской государственной лесотехнической академии. Имеет 120 печатных работ в области динамики и прогнозирования лесистости, кадастровой оценки древостоев и моделирования экосистем.

E-mail: lestaks53@mail.ru



Ревин Алексей Иванович родился в 1959 г., окончил в 1985 г. Воронежский лесотехнический институт, кандидат сельскохозяйственных наук, доцент кафедры лесной таксации и лесоустройства Воронежской государственной лесотехнической академии. Имеет 50 печатных работ в области комплексной продуктивности насаждений.

E-mail: lestaks53@mail.ru



Старостюк Наталья Борисовна родилась в 1983 г., окончила в 2005 г. Воронежскую государственную лесотехническую академию, заместитель директора офиса коммерциализации инновационных проектов и разработок. Имеет 9 печатных работ в области охраны окружающей природной среды и особо охраняемых природных территорий.

E-mail: natder83@mail.ru



ИСПОЛЬЗОВАНИЕ РЕГРЕССИОННЫХ МЕТОДОВ В МОДЕЛИРОВАНИИ ЛЕСИСТОСТИ

Подробно проанализированы проблемы моделирования многомерных процессов в лесу, обсужден подход сокращения их размерности. На примере динамики процента лесистости Центрально-Черноземного региона предложена аппроксимация этого показателя с использованием сплайн-функции с изменяющимся узлом.

Ключевые слова: лесистость, регрессионный анализ, моделирование, полиномы, сокращение размерности, аппроксимация, интерполяция.

В лесоводственных исследованиях находит широкое применение классический регрессионный анализ. Характерное отличие множественных регрессий от других видов анализа состоит в том, что они дают аналитическое выражение для исходных соотношений и характеристик процесса, а это является важным условием последующего динамического моделирования.

Основная проблема при моделировании многомерных процессов, с которой приходится сталкиваться на этапе качественного анализа, – сокращение размерности массива данных, так как в общем случае многомерное пространство характеризуется отсутствием наглядности и трудно поддается непосредственному качественному анализу. В работах последних лет заслуживает внимание подход, который можно определить как метод последовательного сокращения размерности. На первом этапе проводится анализ парной связи зависимой переменной с определяющим фактором воздействия на фоне остальных учитываемых факторов. Эта связь выражается функцией с минимальным числом констант. Поскольку линейная связь является частным

случаем более общей нелинейной связи, то лучший результат дает использование нелинейных функций, которые можно с помощью простых операций привести к линейным.

С момента зарождения лесотаксационной науки подобные методы сокращения размерности широко практиковались с целью свести многомерную задачу к парным связям. В качестве фактора воздействия использовали среднюю высоту и возраст древостоя. Как упоминалось выше, метод последовательного сокращения размерности предполагает применение на первом этапе парной связи, описываемой нелинейной функцией с минимальным числом констант. В сравнении с обычными полиномами этому условию в большей степени удовлетворяют логарифмические полиномы первого и второго порядков, константы которых имеют определенную биологическую интерпретацию.

Известно, что изменение количества в процессе времени (роста) обычно происходит согласно уравнению

$$\frac{d_Y}{d_A} \frac{1}{Y} = a \frac{d_X}{d_A} \frac{1}{X}, \quad (1)$$

где X, Y – объемы, размеры;

A – время;

a – аллометрическая константа.

Интегрирование дифференциального уравнения (1) дает функцию

$$l_{ny} = l_{ne} + a l_{nx}, \text{ или } y = vx^a, \quad (2)$$

где v – константа начального процесса.

Первые попытки применить в количественной биологии данную зависимость и к процессам роста относятся к концу прошлого века.

Степенная функция известна в теоретической биологии и как функция параболического роста, делящая период большого роста древостоев на несколько возрастных этапов и аппроксимирующая таксационные показатели уравнением ступенчатого параболического порядка. При этом рассчитываются константы функции для каждого этапа. Не отрицая общебиологической концепции о наличии физиологически обусловленных точек перелома кривой, характеризующей рост организма, мы склонны интерпретировать логарифмический полином n -го порядка как «исправленную» параболическую функцию. Невозможность нахождения единственной модели по конечному числу экспериментальных данных составляет основной принцип теории множественности моделей.

В случае нелинейного поля из десяти точек, расчлененного на две последовательности по пять точек в каждой, при повышении степени полинома от одного до двух ошибка снижается, а по мере приближения степени к пяти будет стремиться к максимуму, так как кривые, аппроксимирующие ту и другую совокупности, проходят точно через соответствующие пять точек при максимальном расхождении параметров кривых. Наличие минимума внешнего критерия и определяет выбор оптимального уравнения.

В задачах интерполяции и при краткосрочных прогнозах удовлетворительные результаты могут быть получены и при классическом регрессионном анализе.

Модели, построенные по принципу самоорганизации, совершенно не поддаются содержательной интерпретации, и применять этот метод в «чистом» виде нужно с определенной долей осторожности. Нельзя не согласиться с заключением о невозможности выбора оптимального уравнения регрессии по тому или иному статистическому критерию. С учетом вышесказанного о принципе самоорганизации моделей ясно, что речь идет о внутренних критериях, оцененных по всей совокупности исходных данных. При многомерном подходе к биологическому объекту факторы, как правило, коррелируют между собой. При этом имеют место совмещенные эффекты двух или нескольких факторов, объясняющие, в частности, нелинейную часть их воздействия.

В биометрических исследованиях часто используют уравнение множественной регрессии

$$y = a_0 x_1^a x_2^{a_2} \dots x_n^{a_n}, \quad (3)$$

где a_0, \dots, a_n – константы.

Данная множественная статистическая модель представляет в биологических исследованиях такое же исключение, как линейная парная связь среди множества нелинейных зависимостей, поскольку предполагается, что константы уравнения парной связи с одним из признаков никак не коррелируют с другим признаком. Это идеальный случай для множественного регрессионного анализа, но в реальной ситуации такой вариант можно допустить лишь с определенным приближением.

Ввод в уравнение факторов x и x^2 гарантирует их взаимную корреляцию, так и в множественной связи имеется подобная ситуация при вводе x_1 и $x_1 x_2$. Поэтому стремление обеспечить при описании закономерности математическую корректность в известной мере приходит в противоречие с необходимостью адекватного описания.

Резюмируя сказанное, необходимо отметить, что если выполнять все статистические предписания как залог корректности многофакторной модели, то от применения множественного статистического моделирования придется отказаться. По-видимому, при аппроксимации опытных данных практический смысл имеет оценка регрессии по соотношению остаточной и общей дисперсий, по стандартной ошибке с указанием пределов применимости регрессий и достоверности констант по критериям Стьюдента или Фишера. Требуемую точность необходимо соотносить с возможностью реализовать ее на практике, так как бессмысленно доводить модель до точности 5 %, если исходные данные получены с ошибкой 10 % и выше.

При моделировании динамики и прогнозе лесистости нами применялось логическое обоснование аргументов с приведением их к нелинейному полиному на основе предварительного графического анализа, что, хотя и не представляется бесспорным, но, по крайней мере, недалеко от истины.

Этот метод обеспечивает большую компактность модели и, исходя из биологических и таксационных соображений, дает предварительное общее представление о взаимосвязи переменных величин, число которых обычно превышает 4-5. Большое число аргументов не приводит к существенному увеличению точности модели, но делает ее более громоздкой и снижает критерии значимости констант.

Для выравнивания каждого показателя обычно подбирают соответствующую эмпирическую кривую. Как отмечалось выше, универсальной функции, позволяющей с большой точностью аппроксимировать все таксационные показатели на разных возрастных этапах и лесистость как их производную, до сих пор не существует. Возможны различные варианты.

Так, для аппроксимации динамики возраста и процента лесистости отдельных областей ЦЧР (более чем за 250-летний период времени) анализировали пригодность сплайн-функции с изменяющимся узлом, которая имеет следующий вид*:

$$Y = \epsilon_0 + \epsilon_1 A + \epsilon_2 A^2 + \epsilon_3 A^3 + \epsilon_4 (A - C_0)^3, \quad (4)$$

где Y – исследуемый таксационный признак (лесистость), %;

$\epsilon_0, \dots, \epsilon_4$ – линейные параметры функций;

A – возраст;

C_0 – нелинейный параметр;

$$A - C_0 = \begin{cases} A - C_0, & \text{если } A \geq C_0; \\ 0, & \text{если } A \leq C_0. \end{cases}$$

Данная функция – частный случай кусочной аппроксимации. Кривая процента лесистости при выравнивании разбивается на две части. В области $0 \leq A \leq C_0$ сплайн-функция является параболой

$$Y = \epsilon_0 + \epsilon_1 A + \epsilon_2 A^2 + \epsilon_3 A^3, \quad (5)$$

а в области $C_0 \leq A \leq C_{\max}$ она дополнена новым параметром ϵ_4 и имеет следующий вид:

$$Y = \epsilon_0 + \epsilon_1 A + \epsilon_2 A^2 + \epsilon_3 A^3 + \epsilon_4 (A - C_0)^3. \quad (6)$$

Нелинейный параметр C_0 в данной функции является возрастом «стыковки» двух парабол и может быть определен интеграционным методом. Линейные параметры $\epsilon_0, \dots, \epsilon_4$ находят методом наименьших квадратов при фиксированных значениях C_0 . Сплайн-функция была проверена нами для выравнивания нестыкующихся показателей уровня лесистости.

Показатели лесистости используются на основе материалов учета лесного фонда РФ, что подтверждает единый методический подход при ее определении.

* Успенский В.В., Мироненко А.В. Моделирование динамики сосновых древостоев с использованием сплайн-функций // Лесн. журн. – 1994. – № 4. – С. 65–69. – (Изв. высш. учеб. заведений).

Разработка данного вопроса тесно связана с будущим искусственно созданных лесных насаждений на территории ЦЧР. Применение сплайн-функций следует считать пробным этапом в моделировании динамики лесистости.

A.N. Smolianov, A.I. Revin, N.B. Starostyuk

Use of Regression Methods in Forest Land Simulation

Simulation problems of multivariate processes in the forest are analyzed in detail, the approach of reducing their dimension is discussed. Based on the example of the forest-land percentage dynamics of the Central Chernozem Region the approximation of this indicator is offered by using spline function with a changeable block.

Keywords: forest-land percentage, regression analysis, simulation, polynomials, reduction of dimension, approximation, interpolation.
