

УДК 630*182.59

И. Ф. БУКША¹, В. П. ПАСТЕРНАК¹, О. А. ПРОСКУРНИН² *
**ПРИМЕНЕНИЕ НЕПАРАМЕТРИЧЕСКОГО РЕГРЕССИОННОГО АНАЛИЗА
ДЛЯ ВОССТАНОВЛЕНИЯ ПРОБЕЛОВ ВО ВРЕМЕННЫХ РЯДАХ ДАННЫХ
МОНИТОРИНГА ЛЕСОВ**

1. Украинский научно-исследовательский институт лесного хозяйства и агролесомелиорации им.

Г. Н. Высоцкого

2. Украинский научно-исследовательский институт экологических проблем

Обосновывается возможность использования непараметрического регрессионного анализа для восстановления недостающих данных наблюдений в системе экологического мониторинга лесов. Приводится пример оценки регрессионной зависимости дефолиации деревьев от дехромации для Луганской области по данным мониторинга лесов 2006 г. Адекватность регрессионной модели проверяется с помощью критерия Фишера, а также по данным независимых наблюдений 2007 г.

К л ю ч е в ы е с л о в а : мониторинг лесов, дефолиация, дехромация, регрессионная зависимость.

Лесные экосистемы характеризуются сложной структурно-функциональной организацией и как природные объекты, состоящие из множества взаимосвязанных элементов, сложно поддаются строгому (детерминированному) математическому моделированию. В этой связи, при исследованиях лесных экосистем актуальным является применение методов статистической обработки результатов наблюдений, особенно при анализе долговременных рядов данных, которые формируются при мониторинге лесов и, в частности – для восстановления недостающих данных наблюдений.

Согласно теории математической статистики, если x и y – одновременно измеряемые величины, характеризующие состояние природного объекта, и между ними существует стохастическая связь (в результате зависимости x и y от общих факторов либо из-за наличия между ними причинно-следственной связи), то статистические методы позволяют определять возможное значение одной величины по данным измерения другой. С этой целью рассматривается функция регрессии $y^*(x)$ – зависимость наиболее вероятного значения величины y от значения величины x . Решение регрессионной задачи заключается в статистической оценке функции $y^*(x)$ на основе ретроспективных данных натуральных наблюдений. Найденная зависимость позволит восстанавливать пропущенные значения либо обнаруживать недостоверность результатов анализов (т. е. проводить экспертизу данных).

Целью данной работы является демонстрация использования регрессионного анализа для восстановления пропущенных наблюдений при проведении мониторинга леса. В качестве примера определяли регрессионную зависимость дефолиации, т.е. – преждевременной потери листьев/хвои на деревьях (y) от дехромации – преждевременного изменения цвета листьев/хвои (x). Выбор данных показателей объясняется тем, что, во-первых, дефолиация и дехромация кроны являются базовыми показателями состояния насаждения [2], во-вторых – эти показатели связаны между собой, поскольку отражают влияние стресс-факторов на деревья.

Наиболее распространенным методом решения регрессионной задачи в настоящее время является метод наименьших квадратов. Однако проблема его практического использования заключается в том, что данный метод относится к группе параметрических методов статистики, для которых результат расчета зависит от того, насколько закон распределения рассматриваемых случайных величин близок к нормальному распределению [7]. На недостатки применения метода наименьших квадратов при решении задач экономико-экологического мониторинга в условиях случайного действия внешней среды указывают В. А. Тимофеев, Л. В. Левченко, А. Н. Куркин [6]. Особенно эти недостатки проявляются при сравнительно небольших объемах натуральных наблюдений, когда невозможно проверить

*© И. Ф. Букша, В. П. Пастернак, О. А. Проскурнин, 2008

гипотезу о законе распределения с достаточно высокой степенью достоверности. Поскольку реальные мониторинговые задачи обычно характеризуются сравнительно небольшим объемом выборки, а закономерность распределения определяемых параметров лесного мониторинга, в силу особенностей объектов наблюдения, как правило, существенно отличается от известных законов распределения (нормального, Фишера, Стьюдента, χ -квадрат и др.), указанный недостаток метода наименьших квадратов является существенным, что ограничивает возможности его использования в практике. Практическим выходом в данной ситуации является применение непараметрических статистических методов при анализе данных мониторинга лесов.

В работе [1] теоретически обоснован и предложен метод непараметрического регрессионного анализа, в основе которого лежит метод статистических испытаний (Монте-Карло). Метод позволяет по данным выборочных наблюдений $\{x_j, y_j\}$ объема n находить оценку регрессионной зависимости $f(x) \approx y^*(x)$ в виде элемента функционального евклидова пространства с ортогональным базисом $\{\Theta_i(x)\}$:

$$f(x) = \sum_{i=0}^M \gamma_i \Theta_i(x)$$

где γ_i – коэффициенты разложения, рассматриваемые в данной задаче как параметры регрессии, M – порядок искомой функции, определяемый вместе с γ_i .

Оценка параметров γ_i включает также проверку нулевой гипотезы с заданным уровнем значимости. Кроме того, определяются границы доверительной полосы:

$$f^-(x) = \hat{f}(x) - 2\sigma f(x),$$

$$f^+(x) = \hat{f}(x) + 2\sigma f(x),$$

где $\sigma f(x)$ – среднеквадратичное отклонение величины $f(x)$.

С целью дальнейшего представления результата в виде степенного ряда (наиболее удобной формы для анализа результата) в качестве базиса $\{\Theta_i\}$ целесообразно выбирать одну из полных полиномиальных систем [5].

Оценку регрессионной зависимости дефолиации от дехромации проводили по данным мониторинга лесов Луганской обл. за 2006 г. для дубовых насаждений. Лесоводственно-таксационная характеристика насаждений приведена в табл. 1.

Таблица 1

Таксационная характеристика насаждений на участках мониторинга

№ участка	Состав пород	Возраст, лет	Тип леса	Полнота	Бонитет
192141	7ЯО 3ДО	79	D ₁ -бркпД	0,7	III
203443	6ДО 4СО	82	B ₂ -дС	0,5	IV
203540	9ДО 1ВЗМ +ЯЗ	42	D ₁ -бркпД	0,7	III
203650	10ДО	53	D ₀ -кпД	0,7	IV
215341	10ДО	41	D ₁ -бркпД	0,7	III
226551	8ДО 2КЛО	113	D ₁ -бркпД	0,6	III
226733	10ДО	58	B ₁ -дС	0,5	III
226763	10ДО	94	D ₁ -бркпД	0,6	IV
227001	10ДО +ЯО	85	D ₂ -клД	0,7	II
238532	8ДО 2ЯО	93	D ₂ -клД	0,6	III
238902	10ДО	104	D ₂ -бркпДп	0,6	III
250540	10ДО	53	D ₀ -кпД	0,7	V
250551	6Ос 4ДО	53	D ₃ -клД	0,6	III
250650	8ДО 2ЯО	57	D ₂ -клД	0,7	IV
274931	6ДО 3ЯО 1ЛпД	58	D ₂ -клД	0,7	III

Определение дефолиации и дехромации проводили специалисты Харьковской государственной лесоустроительной экспедиции ПО «Укргослеспроект» в соответствии с методическими рекомендациями, разработанными в УкрНИИЛХА [4].

Для расчета брали средние значения показателей дефолиации и дехромации для исследуемых участков. При этом с целью снижения фактора случайности исключались из рассмотрения участки с количеством деревьев менее 10. Исходные данные для построения регрессионной модели представлены в табл. 2.

Таблица 2

Средние значения дефолиации и дехромации по участкам мониторинга лесов в Луганской области за 2006 г.

№ участка	Количество деревьев, шт.	Средняя дехромация, X	Средняя дефолиация, Y
192141	17	13,82	19,12
203443	12	7,08	8,58
203540	18	6,11	13,17
203650	24	6,88	23,00
215341	24	8,33	8,38
226551	12	19,58	20,08
226733	24	15,21	26,75
226763	23	18,04	32,52
227001	12	1,25	16,00
238532	21	16,75	17,30
238902	24	7,08	11,50
250540	24	1,88	8,42
250551	10	0,00	3,50
250650	21	3,33	7,57
274931	17	4,00	4,90

Результатом решения регрессионной задачи непараметрическим методом является следующая модель:

$$f = 4,86 + 1,012 x$$

Таким образом, имеет место статистически значимая линейная регрессионная зависимость между дефолиацией и дехромацией древостоя. На рис. 1 представлен график оцененной зависимости.

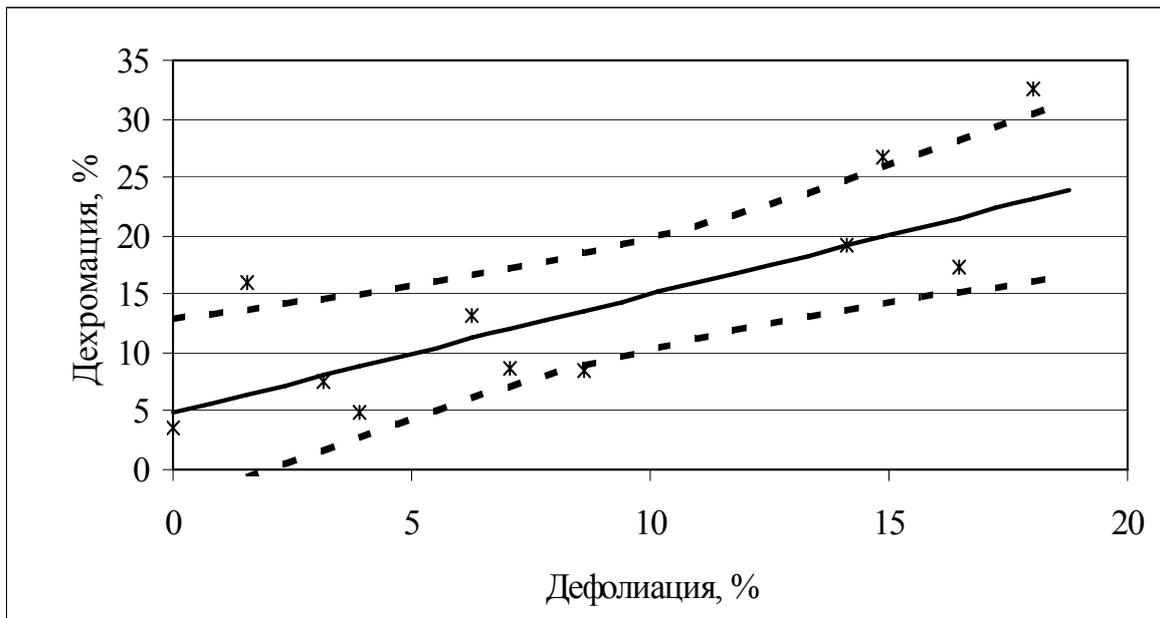


Рис. 1 – Регрессионная зависимость дефолиации древостоя от дехромации (жирная линия – функция $f(x)$, пунктир – доверительные границы $f - f_{довер}$ и $f + f_{довер}$)

Адекватность полученной регрессионной модели была проверена с помощью критерия Фишера [3], который заключается в сравнении расчетной величины:

$$F_{расч} = F_{m,n-m-1} = \frac{(n-m-1) \sum_{i=1}^n (f_i - y_{cp})^2}{m \cdot \sum_{i=1}^n (f_i - y_i)^2}$$

где m – количество независимых переменных модели ($m = 1$), y_{cp} – выборочное среднее величины y ,
с критическим значением $F_{крит.}$, равным квантилю F-распределения для принятого уровня значимости α и количества степеней свободы m и $n-m-1$.

Если $F_{крит.} < F_{расч.}$, то полученную регрессионную модель можно считать адекватной экспериментальным данным с уровнем надежности $P = 1-\alpha$.

Согласно расчету, $F_{расч.} = 125,20 > F_{крит.} = 4,41$. Таким образом, полученная модель адекватна экспериментальным данным.

Работоспособность регрессионной модели была проверена на независимых (неиспользованных при построении модели) данных за 2007 г. на примере участков с наибольшим количеством деревьев (табл. 3).

Таблица 3

Результат проверки регрессионной модели на различных участках

№ участка	Количество деревьев	Средняя погрешность $ y-f $, %
203650	24	12,5
215341	24	6,9
226733	24	7,43
238902	24	10,7
250540	24	4,16

Как видно из табл. 3, средняя величина расхождения между замеренной и прогнозируемой дефолиацией является относительно небольшой.

Таким образом, полученные регрессионные модели позволяют определять диапазон наиболее вероятных значений дефолиации деревьев по известному значению дехромации, что, в свою очередь, позволяет использовать модели для восстановления недостающих данных мониторинга лесов.

Выводы. Метод непараметрического регрессионного анализа позволяет по относительно небольшому объему данных натуральных измерений оценивать регрессионную зависимость между базовыми показателями мониторинга лесов – дефолиацией и дехромацией крон деревьев. Полученные регрессионные модели могут быть использованы для восстановления недостающих данных лесного мониторинга.

Для повышения точности регрессионных моделей рекомендуется производить стратифицированные выборки данных для решения регрессионной задачи с учетом лесоводственных и таксационных характеристик древостоя, например – возраста, типов леса, географического положения участков, полноты, сомкнутости и т.п. Также в более углубленных исследованиях можно применять многомерные регрессионные модели для повышения точности статистических оценок.

СПИСОК ЛИТЕРАТУРЫ

1. Баранник В. А., Проскурнин О. А. Применение метода статистических испытаний в регрессионном анализе данных экологических исследований // Экологическая, техногенная безопасность и социальный прогресс // Вест. ХИСП. – 2001. – Вып. 1. – С. 9 – 21.
2. Букша И. Ф., Пастернак В. П., Волкова Р. Е., Мешкова Т. С., Одинцов А. В. Особенности функционирования и развития мониторинга лесов в Украине // Екологічний менеджмент як складова частина сталого розвитку: Зб.наук.пр. Донецького держ.ун-ту управління. – Сер. «Державне управління». – Донецьк: ДДУУ, 2004. – Т. 5, вип. 36 – С. 127 – 134.
3. Кафаров В. В. Методы кибернетики в химии и химической технологии. – М: Химия, 1971. – 180 с.

4. Методичні рекомендації з моніторингу лісів України І рівня. – Х.: УкрНДЦЛГА, 2001. – 33 с.
5. Суєтин П. К. Классические ортогональные многочлены. – М.: Наука, 1976. – С. 100 – 120.
6. Тимофеев В. А., Левченко Л. В., Куркин А. Н. Об одной процедуре экономико-экологического мониторинга в условиях случайного действия внешней среды // Управління розвитком: збірник наукових статей ХНЕУ. – М-ли Міжнародної науково-практичної конференції «Стратегії ІТ-технологій в освіті, економіці та екології» (15 – 16 листопада 2007 р., Харків). – Х.: ХНЕУ, 2007. – № 7. – С. 101 – 102.
7. Худсон Д. Статистика для физиков. – М.: Мир, 1970. – 152 с.

Buksha I. F.¹, Pasternak V. P.¹, Proskurnin O. A.²

USING OF UNPARAMETRIC REGRESSION ANALYSES FOR RENEWAL LACKING DATA IN TIME SERIES OF FOREST MONITORING

1. *Ukrainian Research Institute of Forestry and Forest Melioration named after G. M. Vysotsky*

2. *Ukrainian Research Institute of Ecological Problems*

Possibilities of unparametric regression analyses using for renewal lacking data of observation in system of ecological forest monitoring are validated. Regression relation between defoliation and discoloration for Luhansk region on monitoring data 2006 is exemplified. Adequacy of regression model is testified by Fisher criterion and independent data for 2007.

К e y w o r d s : forest monitoring, defoliation, discoloration, regression.

Букша І. Ф.¹, Пастернак В. П.¹, Проскурнін О. А.²

ЗАСТОСУВАННЯ НЕПАРАМЕТРИЧНОГО РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ ВІДНОВЛЕННЯ ПРОБІЛІВ У ЧАСОВИХ РЯДАХ ДАНИХ МОНІТОРИНГУ ЛІСІВ

1. *Український науково-дослідний інститут лісового господарства та агролісомеліорації ім. Г. М. Висоцького*

2. *Український науково-дослідний інститут екологічних проблем*

Обґрунтовується можливість використання непараметричного регресійного аналізу для відновлення даних спостережень, яких бракує при проведенні екологічного моніторингу лісів. Наведено приклад оцінки регресійної залежності дефоліації дерев від дехромації для Луганської області за даними моніторингу лісів 2006 р. Адекватність регресійної моделі перевіряється за допомогою критерію Фішера, а також за даними незалежних спостережень 2007 р.

К л ю ч о в і с л о в а : моніторинг лісів, дефоліація, дехромація, регресійна залежність.

Одержано редколегією 24.10.2007 р.