

Проскурнин О. А.,*НИУ «Украинский НИИ экологических проблем», г. Харьков,***Букша И.Ф., Пастернак В.П.***Украинский НИИ лесного хозяйства и агролесомелиорации им. Г.Н. Высоцкого, г. Харьков***АНАЛИЗ ИЗБЫТОЧНОСТИ КОНТРОЛЬНЫХ УЧАСТКОВ МОНИТОРИНГА ЛЕСА С ИСПОЛЬЗОВАНИЕМ МЕТОДА BOOTSTRAP**

Регулярные наблюдения за лесными насаждениями в той или иной форме проводятся с самого начала хозяйственной деятельности человека в лесу. Наличие информации о состоянии лесов необходимо для проведения научно обоснованных мероприятий по повышению устойчивости лесных насаждений и улучшению выполнения ими экологических функций (сохранения биоразнообразия, регулирования стока воды, защиты водоёмов от загрязнения и др.). Однако, развитие современных систем лесного мониторинга началось в Украине в 90-х годах в связи с необходимостью удовлетворения новых информационных потребностей, связанных с экологической защитой лесного хозяйства и стремлением к устойчивому (сбалансированному) лесопользованию [1, 2].

В Сумской области на данный момент действуют более 400 контрольных участков. Проведение необходимых исследований в рамках мониторинга требует значительных материальных, интеллектуальных и временных затрат. По этой причине является актуальным оптимизация расположения контрольных точек: выбор минимального количества контрольных участков, обеспечивающих сбор необходимой информации о состоянии древостоев.

Целью данной работы является анализ избыточности контрольных участков мониторинга леса. Избыточность рассматривается с точки зрения возможного дублирования собираемой информации из-за близкого расположения участков.

Исследование проводилось в 2 этапа. На 1-м этапе оценивалась корреляционная зависимость между состоянием деревьев на различных участках. Для этого отдельно по каждому показателю качества

рассчитывались выборочные коэффициенты корреляции (условно говоря, корреляция 1-го уровня):

$$r_{ij} = \frac{\sum_{g=2005}^{2012} (u_i^g - \bar{u}_i) \cdot (u_j^g - \bar{u}_j)}{\sqrt{\sum_{g=2005}^{2012} (u_i^g - \bar{u}_i)^2 \cdot \sum_{g=2005}^{2012} (u_j^g - \bar{u}_j)^2}}, \quad (1)$$

где u_i^g, u_j^g – средние значения показателей качества на участках i и j в год наблюдения g ; \bar{u}_i, \bar{u}_j – средние значения показателей за все годы наблюдений.

Рассчитанные величины $r_{i,j}$ характеризуют схожесть среднего состояния деревьев по каждому из рассматриваемых показателей на участках i и j . Проверка статистической значимости коэффициентов при этом не проверялась, поскольку на данном этапе исследования не ставилась задача оценки регрессионной зависимости между качеством деревьев на различных участках.

На 2-м этапе коэффициенты $r_{i,j}$ рассматривались как случайные величины и оценивалась корреляционная зависимость 2-го уровня $R(r_{i,j}, x_{i,j})$, где $x_{i,j}$ – расстояние между участками i и j . Отрицательное значение величины $R(r_{i,j}, x_{i,j})$ означает увеличение схожести среднего состояния деревьев с уменьшением расстояния между участками.

На данном этапе исследований определялась статистическая значимость корреляционной зависимости: проверялась нулевая гипотеза – равенство нулю истинного (по всей генеральной совокупности) коэффициента корреляции.

Проверка статистической значимости корреляции проводилась двумя способами. Традиционный подход основан на том факте, что величина

$$Z = \frac{R}{\sqrt{1-R^2}} \sqrt{N-2} \quad (2)$$

удовлетворяет t -распределению Стьюдента с $N-2$ степенями свободы [3].

Таким образом, вывод о статистической значимости делался исходя из выполнения условия

$$|R| > |R_{кр}|, \quad (3)$$

где $R_{кр}$ – критическое значение, равное решению уравнения

$$\frac{R_{кр}}{\sqrt{1-R_{кр}^2}} \sqrt{N-2} = t_{\alpha, N-2}, \quad (4)$$

где $t_{\alpha, N-2}$ – квантиль распределения Стьюдента; α – принятый уровень значимости.

Однако приведенный метод определения статистической значимости является надежным лишь при условии нормального распределения рассматриваемых случайных величин [4]. В данном же случае это условие заведомо не выполняется, поскольку случайная величина r_{ij} не является непрерывной и строго ограничена числами -1 и $+1$, что противоречит характеру нормального распределения.

В силу наличия сомнения в корректности применения традиционного подхода к проверке нулевой гипотезы, корреляция также оценивалась методом bootstrap [5, 6], относящимся к группе непараметрических методов статистики (т.е. не требующими информации о законе распределения случайных величин). Суть метода bootstrap заключается в следующем. Для оценки некоторой величины T по выборке наблюдений $\{X_i\}$, $i = 1 \div N$, многократно составляется bootstrap-выборка $\{X^*_1, X^*_2, \dots, X^*_N\}$ из элементов начальной выборки. При этом некоторые элементы начальной выборки могут повторяться в каждой отдельной bootstrap-выборке, а некоторые вовсе не включаться. Для каждой bootstrap-выборки рассчитываются оценки $T^*(\{X^*_1, X^*_2, \dots, X^*_N\})$, по которым составляется вариационный ряд $\{T^*_1, T^*_2, \dots, T^*_B\}$, где B – количество bootstrap-операций. Искомой оценкой является среднее арифметическое ряда

$$\hat{T} = \frac{1}{B} \sum_{b=1}^B T^*_b, \quad (5)$$

а границами доверительного интервала – значения $\{T^*_b\}$ с порядковыми номерами $[B \cdot \alpha/2]$ и $[B \cdot (1-\alpha/2)]$.

Если в качестве выборки $\{X_i\}$ взято множество пар случайных чисел $\{x_i, y_i\}$, то методом bootstrap можно оценить коэффициент корреляции между величинами x и y .

Если общее количество участков равно M , то количество пар участков равно

$$N = C_M^2 = \frac{M \cdot (M-1)}{2!}, \quad (6)$$

где C_M^2 – обозначение числа сочетаний.

В силу большого количества пар участков было проведено разбиение Сумской области на отдельные подрегионы (табл. 1).

Таблица 1 – Рассматриваемые подрегионы Сумской области

Порода деревьев	Подрегион	Географические границы подрегиона	
		Долгота	Широта
дуб	№1	>34,1°	>50,6°
	№2	>34,1°	<50,6°
	№3	<34,1°	<50,9°
сосна	«Север»	–	>51,7°
	«Центр»	–	>51,1° и <51,7°
	«Юг»	–	<51,1°

В качестве показателей состояния деревьев рассматривались дефолиация (недостаток листвы) и дехромация (недостаточность цветности). Результат расчетов приведен в табл. 2. (Жирным шрифтом выделены статистически значимые значения для $\alpha = 0,05$.)

Как видно из табл. 2, в тех случаях, когда традиционный метод дает положительный результат – статистически значимую отрицательную корреляцию, такой же результат подтверждает и bootstrap-метод. В ряде случаев положительный результат дает лишь непараметрический метод.

Таблица 2 – Корреляционная зависимость 2-го уровня

Показатель	Подрегион	Количество участков M	Количество пар участков N (объем выборки)	Параметрический (традиционный) метод		Bootstrap-метод		
				Коэффициент корреляции R	Критическое значение $R_{кр}$	Коэффициент корреляции R	Доверительные границы	
							R^-	R^+
дуб								
дефолиация	№1	13	78	0,048	0,349	0,227	0,029	0,381
	№2	17	136	0,082	0,349	0,078	-0,094	0,242
	№3	7	21	0,037	0,433	-0,077	-0,679	0,466
дехромация	№1	13	78	0,093	0,349	0,068	-0,190	0,317
	№2	17	136	-0,461	0,349	-0,458	-0,599	-0,275
	№3	7	21	0,090	0,433	-0,197	-0,720	0,362
сосна								
дефолиация	«Север»	20	190	0,037	0,349	-0,473	-0,604	-0,340
	«Центр»	12	66	-0,315	0,349	-0,468	-0,651	-0,249
	«Юг»	14	153	-0,454	0,349	-0,452	-0,578	-0,326
дехромация	«Север»	20	190	0,057	0,349	-0,386	-0,516	-0,246
	«Центр»	12	66	-0,101	0,349	-0,228	-0,453	0,020
	«Юг»	14	153	-0,293	0,349	-0,282	-0,403	-0,161

Согласно расчетам методом bootstrap, зависимость близости результатов наблюдения от расстояния между контрольными участками сразу по обоим показателям имеет место только для породы сосна в северной и южной частях региона. При этом близость средних значений по одному показателю предполагает схожесть и по другому показателю. Это следует из статистической значимости коэффициента корреляции

$$\rho(r_{i,j}^1, r_{i,j}^2), \quad i, j \in K, \quad (7)$$

где K – множество контрольных участков сосны в подрегионе; $r_{i,j}^1, r_{i,j}^2$ – коэффициенты корреляции 1-го уровня соответственно для показателей дехромация и дефолиация.

Результат расчета величины ρ методом bootstrap следующие: в подрегионе «Север» $\rho = 0,397$ ($\rho^- = 0,270$; $\rho^+ = 0,524$), в подрегионе «Юг» $\rho = 0,294$ ($\rho^- = 0,131$; $\rho^+ = 0,457$).

Статистическая значимость коэффициентов R и ρ дает основание для нахождения регрессионной зависимости между состоянием сосны на близко расположенных участках. В качестве примера построена регрессионная модель для участков 402403 и 402442. (Коэффициенты корреляции $r_{402403, 402442}$ для показателей дефолиация и дехромация равны соответственно 0,763 и 0,976; расстояние между участками – 6 км.) В силу отсутствия информации о законе распределения рассматриваемых случайных величин, расчет производился непараметрическим методом регрессионного анализа, основанным на методе статистических испытаний (метод Монте-Карло) [7]. Результат расчета приведен на рис. 1.

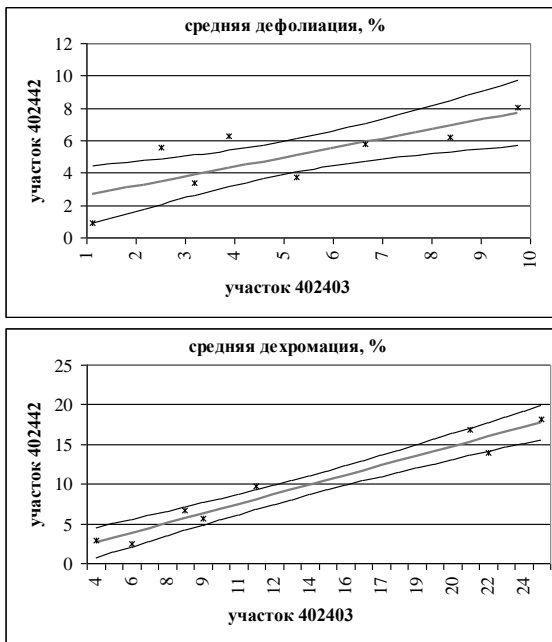


Рис. 1. Регрессионная зависимость между качеством деревьев на контрольных участках

Таким образом, регрессионная модель позволяет по информации о среднем значении дехромации или дефолиации на участке 402403 определять вероятное значение соответствующего показателя на участке 402442.

Выводы

В Сумской области по породе дуб отсутствует статистически значимая корреляционная зависимость между данными наблюдений на различных контрольных участках и расстоянием между участками. По породе сосна указанная зависимость имеет место на севере и юге области. Наличие данной корреляции позволяет построить регрессионную зависимость между осредненными качественными характеристиками деревьев на близкорасположенных участках, что, в свою очередь, позволяет сократить количество участков наблюдения.

Направлениями дальнейших исследований, связанных с оптимальным расположением контрольных участков наблюдения, являются:

- дальнейшее накопление и обработка данных по состоянию древостоя, про-

верка адекватности регрессионных моделей состояния деревьев на независимых данных (т.е. на данных наблюдений, которые не были использованы при построении модели);

- выработка рекомендаций по допустимому уровню погрешности при замене реальных измерений на оценочные путем использования регрессионных моделей.

ЛИТЕРАТУРА:

1. Ткач В.П., Букша И.Ф. Леса и лесное хозяйство Украины: проблемы научного обеспечения устойчивого развития // Materialele Simpozinului internațional «Dezvoltarea durabilă a sectoru lui forestier – noi obiective și priorități». – Chisinau, 17-19 noiembrie 2011. – P. 95 – 101.
2. Пастернак В. П. Обґрунтування напрямів діяльності у секторі землекористування та лісового господарства для забезпечення виконання вимог Кіотського протоколу в Україні / В. П. Пастернак, І. Ф. Букша // Науковий вісник Національного аграрного університету. – 2007. – Вип. 106. – С. 124–134.
3. Справочник по прикладной статистике. / Под ред. Э.Ллойда, У.Ледермана. Т. 2. – М.: Финансы и статистика, 1990. – 528 с.
4. Худсон Д. Статистика для физиков. – М.: Мир, 1970. – 152 с.
5. A Bootstrap-Regression Procedure to Capture Unit Specific Effects in Data Envelopment Analysis. [Электронный ресурс]. Режим доступа: <http://www.econ.uconn.edu/working/2004-15.pdf>.
6. Проскурнин О.А. Использование метода bootstrap для оценки влияния сброса сточных вод на качество воды водного объекта / О.А. Проскурнин, С.А. Смирнова // Проблеми охорони навколишнього природного середовища та екологічної безпеки: Зб. наук. пр. УкрНДІЕП – Харків: ВД «Райдер», 2015. – № 37 – С. 163-171.
7. Проскурнін О. А. Прогнозування впливу техногенного забруднення на довкілля методом непараметричного регресійного аналізу: автор. дис. на здобуття наук. ступеня канд. техн. наук: спец. 21.06.01 «Екологічна безпека» / О.А. Проскурнін. – Харків, 2007. – 22 с.