

Еремеев И.С.,¹ Дичко А.О.²

1- Академия муниципального управления, Киев

2- Национальный технический университет «КПИ», Киев

ПОВЫШЕНИЕ ПРЕДСТАВИТЕЛЬНОСТИ ДАННЫХ БИОИНДИКАЦИИ

Для объективной оценки степени загрязнения окружающей среды предложено использовать метод лингвистических описаний, метод бутстрепа, а также метод оценки ареала загрязнений, учитывающий синергетический эффект различных факторов окружающей среды.

Наиболее простым и эффективным средством выявления загрязнений окружающей среды в мегаполисах и тенденций их динамики, в особенности в таких заповедных местах, как парки, зоны отдыха и т.п., является использование биоиндикаторов, которые не требуют специальных аппаратных средств для реализации, учитывают синергетический эффект и способны интегрировать малые воздействия на протяжении всего жизненного цикла или сезона. Причем в качестве биоиндикаторов возможно использование высших растений, применяемых в качестве декора, или же дикорастущих (например, клена, конского каштана, сосны и т.п.). Имеется довольно значительный ряд растений, которые однозначно реагируют на такие загрязнения среды, как, например, сернистые, фтористые, озон и т.п. путем изменения окраски листьев, прожилок листьев, появления белых или черных пятен, обесцвечивания по краям и между прожилками и т.д., причем в каждом заповеднике, парке или зоне отдыха можно отыскать группы деревьев, способные служить в качестве биоиндикаторов. Вопрос заключается в том, чтобы: (1) эти группы (виды) деревьев были достаточно представительны для данного ареала и дополняли друг друга; (2) их реакция на стрессы была вполне однозначной; (3) существовала объективная шкала оценок степени загрязнений; (4) объемы

выборки материала для оценок загрязнения среды являлись вполне представительными.

Первые два условия, как правило, не вызывают проблем. Что же касается объективной оценки степени загрязнений, то здесь можно прибегнуть к методу лингвистических описаний (ЛО), когда некоей усредненной «картинке» той или иной степени загрязнения можно сопоставить корень квадратный из относительной (в долях единицы) поверхности листа, пораженной поллютантом. Например, можно представить такие пять цифровых оценок (ЦО) (см. Рис. 1):

(а) – ЛО: «лист не поражен» (ЦО = 0,0);

(б) – ЛО: «лист незначительно поражен» (ЦО = 0,3);

(в) – ЛО: «лист средне поражен» (ЦО = 0,70);

(г) – ЛО: «лист сильно поражен» (ЦО = 0,87);

(д) – ЛО: «лист очень сильно поражен» (ЦО = 1,0).

Подобные оценки, конечно же, не являются достаточно однозначными, однако они вполне приемлемы для общего оценивания влияния загрязнений на тех или иных ареалах.

Но главная задача при оценивании качества среды – обеспечение представительных выборок для расчета статистических характеристик.

Здесь помощь следует искать в использовании методов *ресамплинга*, в частности, метода *бутстрепа*. Сущность этого метода состоит в том, чтобы путем статистических исследований методом Монте-Карло многократно генерировать повторные выборки из существующего эмпиричного распределения следующим образом:

➤ Берут конечную совокупность с n членов исходной выборки x_1, x_2, \dots, x_n , откуда на каждом шаге с n последовательных итераций с помощью генератора случайных чисел, равномерно распределенных на интервале $[1, n]$, «извлекается» для формирования новой выборки произвольный элемент x_k , который опять «возвращается» в исходную выборку (то есть может быть «извлечен» опять). Например, при $n=8$ первая (исходная) комбинация имеет вид $[x_1, x_2, x_3, x_4, x_5, x_5, x_7, x_8]$, вторая может иметь вид $[x_6, x_3, x_1, x_5, x_4, x_1, x_2, x_3]$ и т.д., то есть отдельные элементы могут повторяться.



(а)



(б)



(в)



(г)



(д)

Рис.1 - Лингвистическое описание (ЛО) зоны поражения листа конского каштана и его цифровая оценка (ЦО)

Таким образом, можно сформировать любое значительное количество бутстреп-выборок.

➤ Каждая последующая псевдовыборка, которая генерируется, будет придавать значения параметру, несколько отличающиеся от того, которое было получено во время расчетов для первичной (предыдущей) совокупности.

➤ Разброс значений параметров образующегося показателя дает возможность построения доверительных интервалов и получения других полезных выборочных параметров анализируемой величины.

Процедуры ресамплинга не требуют никакой априорной информации о законе распределения исследуемой случайной величины и поэтому могут рассматриваться как непараметрические. Они выполняют обработку разных фрагментов исходного массива эмпирических данных, как бы поворачивая их «разными гранями» и сопоставляя полученные таким образом результаты. Конечно, нельзя утверждать о полной корректности этого подхода, но, полагая его вполне законным, асимптотические преимущества ресамплинга в сравнении с классическими параметрическими тестами можно считать неоспоримыми. Значения параметров, построенные на основе размноженных псевдовыборок, не являются, строго говоря, независимыми, однако, при существенном увеличении n со значениями статистик, полученными во время ресамплинга, можно обращаться как с независимыми случайными величинами.

Следует добавить, что тип данных, подлежащих ресамплингу для реализации алгоритма бутстрепа, не имеет значения. Это могут быть бинарные признаки: например, наличие/отсутствие (1/0) загрязнения листа, лингвистические описания или цифровые оценки (как на рис.1), прологарифмированные численности или частоты встречаемости видов в пробах и т.п.

Рандомизация дает, конечно, оценивание не самого выборочного параметра, а его имитационной модели при условии справедливости нулевой гипотезы (то есть для случайных композиций видов и отсутствия влияния на анализируемый параметр).

Имеет смысл рассмотреть два подхода для оценивания распространения загрязнений в границах определенного ареала. Первый предусматривает оценку максимального ареала, загрязняемого тем или иным токсином. Для этого определяется площадь $S_i(Z_j)$, на которой расположены отдельно деревья каждого из видов ($i = \overline{1, n}$), реагирующих в той или иной степени на конкретный загрязнитель Z_j ($j = \overline{1, m}$). При этом общая площадь ареала (Рис.2) определяется соответственно как

$$S_{z1} = \{ S_1(Z_1) \cup S_2(Z_1) \cup \dots \cup S_n(Z_1) \} = \max \{ S_1(Z_1), S_2(Z_1), \dots, S_n(Z_1) \}.$$

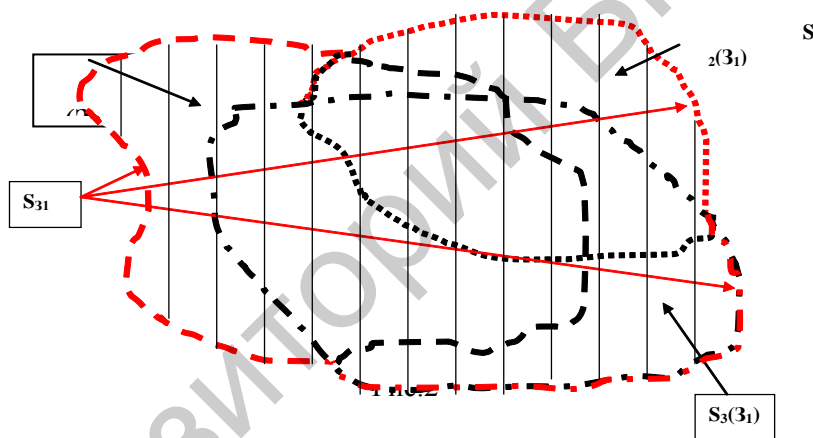


Рис.2 - Оценка максимального ареала, загрязняемого токсином

Неоднозначность индикации, влияние других загрязнений и т.п. уменьшают достоверность определения ареала конкретного загрязнения и потому стоит обратиться к подходу, который предполагает использование взаимоподтверждения разных биоиндикаторов относительно одного и того же загрязнителя. При этом определяется площадь ареала, общая для всех индикаторов Z_1 (т.е. район, где собирается наиболее достоверная информация относительно загрязнения Z_1):

$$S_{31} = \{S_1(Z_1) \cap S_2(Z_1) \cap \dots \cap S_n(Z_1)\} = \min\{S_1(Z_1), S_2(Z_1), \dots, S_n(Z_1)\},$$

которая характеризует площадь взаимного перекрытия всех отдельных площадей ареалов, отвечающих данным биоиндикаторов разных видов, характеризующих конкретный загрязнитель и учитывающих индивидуальные особенности каждого из видов биоиндикаторов, а также синергетический эффект от совместного действия разных факторов (Рис.3).

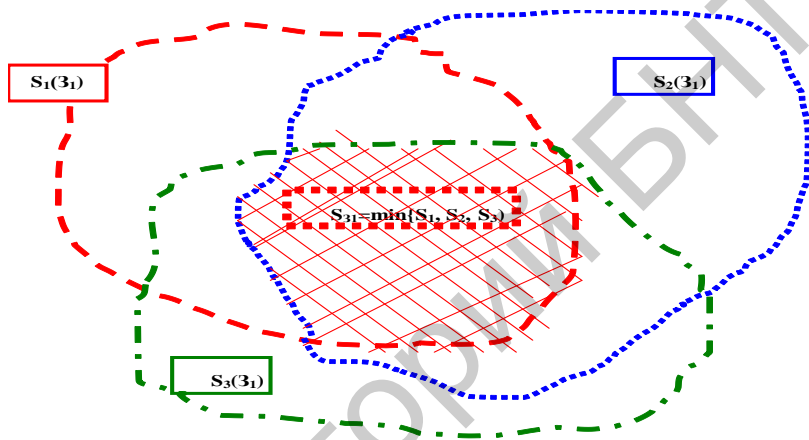


Рис.3 - Оценка максимального ареала, загрязняемого токсином с учетом синергетического эффекта

Таким образом, предложенная методика оценки ареала загрязнений обеспечивает получение наиболее представительной выборки данных биоиндикационных исследований, а также учесть синергетический эффект от совместного действия разных факторов окружающей среды.