

ВЫЯВЛЕНИЕ ОСНОВНЫХ СЕЗОННЫХ ЗАКОНОМЕРНОСТЕЙ ВРЕМЕННОГО ИЗМЕНЕНИЯ ЭКОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Черкашина Ю. А.

Научный руководитель: Гергет О. М., к.т.н., доцент

Национальный исследовательский Томский политехнический университет, 634050, Россия, г. Томск,
проспект Ленина, 30

E-mail: cherr999y@mail.ru

Экологическая обстановка в мире требует всестороннего и повседневного анализа состояния окружающей среды. Только на основании точных количественных данных можно принимать решения о проведении необходимых мероприятий, предупреждающих ухудшение состояния окружающей среды.

В настоящее время при решении экологических задач отходят от вербального описания и основываются на математических моделях и информационных технологиях. Одним из наиболее сложных и трудоемких процессов является выявление закономерностей из имеющихся массивов данных. Он не всегда заканчивается успешно, поскольку данные содержат разнотипную, противоречивую и неполную информацию.

Результатом хозяйственной или иной деятельности человека становится воздействие на все компоненты окружающей среды, в том числе и на земельные ресурсы. Негативные изменения могут проявляться в их загрязнении (порче), а конечном счете - в деградации. Одна из основных причин - отсутствие единой региональной системы сбора, вывоза, сортировки, использования, обезвреживания и размещения отходов различного происхождения.

Для анализа экологических данных использована система, которая позволяет выявить закономерности временного изменения экологических показателей. Она включает такие сервисы, как

- восстановление пропусков в данных;
- выявление наличия сезонности;
- выделение трендов во временных рядах;
- определение сезонной декомпозиции.

Рассмотрим подробнее каждый.

Восстановление пропусков в исходных данных

Пусть значения показателей известны в моменты времени t_j , $j=1, \dots, n$. Построим на интервале $[t_1, t_n]$ функцию $S_y(t)$, интерполирующую $y=f(t)$ так, что на каждом произвольном отрезке $[t_j, t_{j+1}]$, лежащем внутри интервала $[t_1, t_n]$, функция $S_y(t)$ являлась полиномом, а в узлах имела непрерывные производные.

В качестве $S_y(t)$ выбран сплайн третьей степени, который в узлах t_j имеет непрерывные 1-ю и 2-ю производные, и на каждом из отрезков $[t_j, t_{j+1}]$ принимает вид [2]:

$$S_y(t) = y_j + b_j(t-t_j) + c_j(t-t_j)^2 + d_j(t-t_j)^3,$$

где y_j – значение показателей в момент t_j ; b_j , c_j , d_j – коэффициенты, подлежащие определению, решением системы уравнений методом Гаусса.

Оценка качества восстановленного показателя осуществлялась с помощью нахождения коэффициента расхождения Тейлора:

$$B = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}} = \frac{\sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\sqrt{\sum_{i=1}^N y_i^2}}$$

где \hat{y}_i – предсказанное значение для y_i ; y_i – фактическое значение.

Выявление наличия сезонности

Оценка наличия сезонных ритмов во временных рядах осуществлялась на основе функции автокорреляции и её графического представления – коррелограммы. При помощи анализа коррелограммы можно выявить структуру ряда. Последовательность коэффициентов автокорреляции со смещениями 1, 2, 3 и т.д. называют автокорреляционной функцией, значения которой находятся в диапазоне $[-1; 1]$. Автокорреляционную функцию целесообразно использовать для выделения во временном ряде наличия трендовой и сезонной компонент.

Выделение трендов во временных рядах

Выявление наличия неслучайной составляющей сводилось к проверке гипотезы о неизменности среднего значения ряда с использованием критерия серий. При его использовании определяется медиана ряда, и образуются «серии» из плюсов и минусов по следующему правилу:

$$y_i = \begin{cases} +, & y_i > y_{med} \\ -, & y_i < y_{med} \end{cases}$$

Элементы временного ряда, равные y_{med} , в полученной последовательности не учитываются. Под «серией» понимается последовательность подряд идущих плюсов или подряд идущих минусов. Наличие неслучайной составляющей во временном ряде определяется из условия:

$$\begin{cases} v(n) > \left[\frac{1}{2} (n+2-1,96\sqrt{n-1}) \right], \\ K_{max} < [3,3(\lg n + 1)] \end{cases}$$

где $v(n)$ – общее число серий, K_{max} – длина наибольшей серии, $[]$ – целая часть от числа [3].

Для построения тренда использовались два метода: скользящих средних и экспоненциального сглаживания [3].

Метод скользящих средних заключается в следующем: 1) определяется количество наблюдений, входящих в интервал сглаживания; 2) вычисляется среднее значение наблюдений в интервале сглаживания по формуле:

$$\bar{y}_t = \frac{1}{m} \sum_{i=t-\frac{m-1}{2}}^{t+\frac{m-1}{2}} y_i$$

где m – количество наблюдений, входящих в интервал сглаживания. И так до тех пор, пока в интервал сглаживания не войдет последнее значение временного ряда.

В методе экспоненциального сглаживания каждое сглаженное значение рассчитывается путем сочетания предыдущего сглаженного значения и текущего значения временного ряда. В этом случае текущее значение временного ряда взвешивается с учётом сглаживающей константы:

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1}$$

где S_t – текущее сглаженное значение; y_t – текущее значение временного ряда; S_{t-1} – предыдущее сглаженное значение; α – сглаживающая константа, значение которой варьируется в диапазоне от 0 до 1.

Сервис оценки сезонной декомпозиции

Для определения сезонной составляющей разработан алгоритм сезонной декомпозиции.

1. Выделение тренда (метод скользящих средних).
2. Формирование сезонной компоненты (разность между исходным и сглаженным рядом).
3. Вычисление сезонной компоненты (среднее всех значений ряда, соответствующих данной точке сезонного интервала).

Характеристика исследуемых данных

Исследования проводились совместно с Департаментом природных ресурсов и охраны окружающей среды Томской области (далее Департамент). В основу исследования взяты данные по обращению с отходами производства и потребления на территории Томской области.

Сведения представляются в Департамент 1 раз в год, следующего за отчетным периодом.

Данные по обращению с отходами производства и потребления на территории Томской области брались за промежуток времени с 1998 по 2013 год. Было выбрано 7 районов Томской области, на которых проводились исследования. Это Асиновский, Каргасокский, Кожевниковский, Колпашевский, Томский районы, г. Томск и г.Стрежевой

Для исследования были использованы характеристики:

- Образовано отходов – образовано всего отходов в текущем году в районе;

- Количество предприятий – число предприятий района, которые предоставили отчетность по данным.

В результате анализа экологических данных были получены следующие результаты, представленные на рисунках 1-2. В результате исследования временных рядов на наличие периодичности и сезонности получено, что данные не являются периодичными и не содержат сезонности. Полученные результаты можно объяснить особенностью структуры данных.

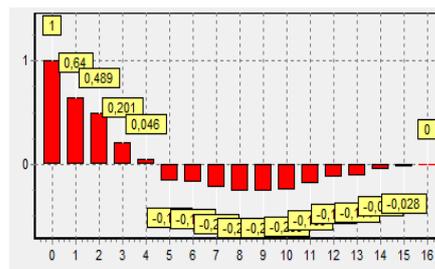


Рис. 1. Коррелограмма (нет сезонной составляющей)

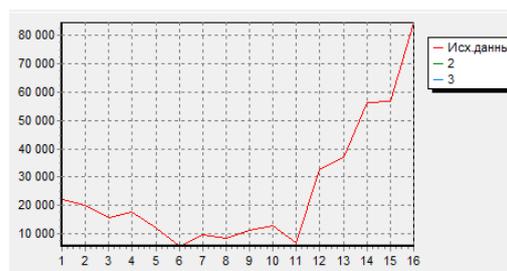


Рис. 2. Исходный ряд

Отсутствие в результатах проведенного исследования сезонности и периодичности свидетельствует о том, что основную массу отходов на исследуемых территориях составляют все-таки бытовые отходы (отходы потребления). Потребление населением товаров и услуг, в результате которых образуются отходы (упаковка, пластиковые бутылки, полиэтилен и др.) постоянны в течение года.

Что касается отходов производства, то отсутствие сезонности и периодичности может свидетельствовать о более-менее постоянном процессе производства на территории Томской области, т.е. сфера производства, развитая на территории области, в целом работает в круглогодичном режиме, без особых сезонных изменений.

Список использованных источников:

1. Бокс Дж., Дженкинс Г.М. Анализ временных рядов, прогноз и управление. М.:Мир, 1974. 406с.
2. Draper N., Smith H. Applied regression analysis. New York: Wiley, 1981. 693.
3. Гергет О.М., Милешин А.А., Информационные технологии выявления основных закономерностей временного изменения показателей биосистемы // Современные проблемы науки и образования, №6 2013.