

УДК: 330.15.12; 614.842/.847

JEL C53,Q53

В.Г. Ларионов, И.А. Сажин

Московский государственный технический университет им. Н.Э. Баумана, Москва,
Россия, Новосибирский государственный технический университет, Новосибирск,
Россия.

Анализ достоверности экологических данных Байкальского региона

Аннотация

Рассматриваются некоторые математические - статистические методики анализа экологических данных Байкальского региона произвольного вида. Представляется целесообразным при планировании экологических программ, направленных на сохранение озера Байкал и прилегающих территорий, диагностировать достоверность множества разнонаправленных данных, представленных в медиа пространстве. Часто важная информация отображена в виде данных произвольного вида. Предложены методики проверки достоверности выборок малого объема. Показана возможность формирования прогноза на основании преобразованных статистик лесных пожаров большого объема в Байкальском регионе. Представлен сравнительный анализ с известными результатами.

Ключевые слова: Экология, Байкал, данные, произвольный вид, достоверность, математические - статистические методики, прогноз.

V.G. Larionov, I.A. Sazhin

Bauman Moscow State Technical University, Moscow, Russia,
Novosibirsk State Technical University, Novosibirsk, Russia.

Analysis of the reliability of ecological data in Baikal region

Abstract

Some mathematical – statistical methods of analysis for ecological data in Baikal region of an arbitrary type are considered. It seems expedient when planning environmental programs aimed at preserving Lake Baikal and adjacent territories, to diagnose reliability of a variety of multidirectional data presented in media space. Often, important information is displayed in free form data. Methods for checking the reliability of small samples are proposed. The possibility of forming a forecast based on the transformed statistics of large – scale forest fires in the Baikal regions shown. A comparative analysis with known results is presented.

Keywords: Ecology, Baikal, data, arbitrary view, reliability, mathematical – statistical methods, forecast.

Необходимость сохранения флоры и фауны озера Байкал и прилегающих территорий обоснована во множестве исследований. В Государственном докладе о состоянии и об охране окружающей среды Иркутской области в 2014 году отражены проблемы экологии Байкальского региона. Реализуемый ряд природоохранных современных программ направлен на защиту и сохранение атмосферы, водных ресурсов, животного и растительного мира мировой жемчужины – региона озера Байкал. Следуя сложившейся практике, на первом этапе внедрения реально эффективных экологических программ, направленных на сохранение природы озера Байкал и прилегающих территорий, используются результаты прикладных исследований, в частности, выполненными методами математико-статистического анализа, основанными на достоверности множества эмпирических данных. Как правило, формирование любого природоохранного проекта на первом этапе связано со сбором соответствующих данных, методы систематизации и обработки которых хорошо известны. В рамках этого этапа реальные выборки должны удовлетворять ряду требований (репрезентативности, однородности) Возможны некоторые преобразования данных (группировка, упорядочение, сглаживание и др.). Точечные оценки должны быть состоятельными, несмещенными, эффективными, робастными. Анализ продуктов средств массовой информации (СМИ), которые связаны с экологическими проблемами Байкальского региона, показывает наличие сведений, которые можно назвать констатирующими, и данных, допускающих обработку математическим - статистическим аппаратом для формирования максимально достоверной динамической модели реального процесса. Достаточно часто весьма важная проблема оформлена в виде выборки произвольного вида при наличии аномалий и некоторых других особенностей. Распространены данные малого объема, обладающие достоверной информацией. Множество данных об экологии региона озера Байкал часто оформлены, мягко говоря, либо тенденциозно, либо слишком «приглажено». Представляется актуальным выявить достоверную часть изучаемой информации. Далее рассматривается несколько примеров, демонстрирующих разные стороны указанной проблемы и возможные способы ее решения. Известны основные показатели загрязнения окружающей среды в Республике Бурятия. Приведенные данные отображают наблюдаемые экологические показатели в течение 7 лет (1990 – 2004 г.г.) и 5 лет (2008 – 2012 г.г.). Прогнозы величин атмосферного загрязнения оказываются достоверными в ограниченном временном промежутке либо значительно усреднены по параметрам чистоты воздуха и местности. Несмотря на малые объемы выборок, составляющих 4-8 единиц, наблюдается наличие тренда, отсутствие аномальных наблюдений. Требуется проверка достоверности исходных данных с заданной доверительной вероятностью. В представленной работе рассматриваются методики формирования статистических – математических моделей, отображающих данные малых объемов, с классической точки зрения являющихся не репрезентативными выборками в силу их малого объема. Применение метода наименьших квадратов (МНК) в моделях, сформированных на базе вышеназванных данных нецелесообразно. С другой стороны, среди экологических данных БПТ встречается хронология крупномасштабных лесных пожаров с 1975 по 2014 г.г (работы Шубкина Р.Г.). Непосредственное применение методик многомерного регрессионного анализа в данном случае исключается из-за наличия аномальных наблюдений. Вероятно, в течение большого временного интервала на экологию БПТ повлиял антропогенный фактор. Следовательно, для достоверного анализа требуется выделить временной интервал с постоянными внешними воздействиями. В данной работе предложены процедуры преобразования, обработки данных большого объема. Показана возможность прогнозирования с помощью методик формирования регрессионных моделей и построения экстраполяционных многочленов разных порядков. Для обработки эмпирических данных малого объема результативно применение линейной вероятностной модели (linear probability models: LPM - модель), которая, в свою очередь, необходимо преобразовывается в logit – модель для устранения возможных недостатков и получения достоверного результата. Рассматривается пример применения робастной процедуры к выборкам малого объема. Известна экологическая информация в виде выборок малого объема с наличием аномалий. Для этого вида данных проводится проверка аномальности выборки ограниченного объема. Целесообразно применить робастный прием анализа: вычислить медиану четной выборки, выборочное среднее, в качестве несмещенного среднеквадратичного отклонения выбирается интерквартильная широта, вычисляется нормирующий коэффициент для устранения смещения оценок среднеквадратичного отклонения. Используя нормальный закон распределения находятся значения нормированной функции Лапласа. Делается вывод о аномальности наблюдения. Для оставшихся наблюдений применяется модели регрессионного анализа. При меньшем объеме выборки допустимо LPM-модель преобразовать в logit – модель, которая

представляется как ограниченная функция вида функции Лапласа. Следует заметить, что совпадение logit – модели и функции Лапласа проверяется по критерию согласия χ^2 - Пирсона. Условные вероятности заменяются относительными частотами выборок сгруппированных данных. Далее возможно применение метода взвешенных квадратов. В результате с необходимой статистической достоверностью можно дать заключение о целесообразности выбора значения объясняемой двоичной переменной. Кроме этого, нетрудно показать целесообразность применения критерия согласия χ^2 - Пирсона для определения степени достоверности выборок ограниченного объема, статистическое исследование которых с помощью других процедур весьма проблематично. Выборки ограниченного объема, представленные в виде логит и пробит – моделей имеют практически совпадающие функции распределения, при условии применения выравнивающего коэффициента равного 1.6, значение которого можно определяется при сравнении соответствующих рядов Тейлора-Маклорена. Оказывается зависимая переменная распределена по нормальному закону с нулевым математическим ожиданием и единичным среднеквадратичным отклонением. Для скорректированной статистики зависимого параметра целесообразно выдвинуть гипотезу о наличии нормального распределения, которая проверяется по критерию согласия χ^2 - Пирсона. Этот статистический прием определения степени достоверности выборки ограниченного объема применен к экологическим данным Байкальского региона, по-видимому, впервые. Основным допущением применения сформулированной выше методики является выполнение нулевой гипотезы о принадлежности рассматриваемых данных некоторой репрезентативной выборки большего объема генеральной совокупности. На практике это означает наличие соответствующих показателей за предшествующие периоды времени. В качестве исходных данных рассмотрены статистики количества ЗВ поступающих в атмосферу от автомобильного транспорта в % от общего количества ЗВ в Бурятии. Наблюдения охватывают промежуток в пять лет, т.е. являются выборкой малого объема, выявление статистически значимых закономерностей в этом случае практически затруднено. Производится перерасчет статистик с помощью линейного регрессионного приближения. Проведена корректировка относительно выборочного среднего. Полученные величины сравниваются с соответствующими величинами логит-модели, для которой проверяется гипотеза соответствия нормальному распределению. Применяется выравнивающий коэффициент равный 1.6 при проверке значимости данных по критерию согласия χ^2 - Пирсона. Устанавливается статистическая достоверность при заданном уровне значимости. Формируется вывод о тренде объемов выбросов загрязняющих веществ в атмосферу от автотранспорта в республике Бурятия. Большой объем данных о количестве пожаров, лесной площади пройденной огнем, средней площади пожаров и др. за значительный период времени (с 1975 по 2014 г.г) позволяет сформировать объективную картину влияния стихийного бедствия, которым является лесной пожар огромных лесных территорий, на современное состояние экологической ситуации Байкальского региона. В исследованиях Шубкина Р.Г. и др. на основании многовековой статистики лесных пожаров авторами реализованы процедуры корреляционного и спектрального анализа количества пожаров, Фурье и вейвлет анализ древесно-кольцевых хронологий, в результате – выявлены цикличности 3-4 - летние, 11-летние, 61-62 – летние. Установлена непосредственная зависимость количества пожаров от весенне – летнего увлажнения почв Байкальского региона, что позволяет улучшить качество прогнозов. Таким образом подтверждено 78% выявленных пожаров, а на период 2012 – 2015г.г. указана возможность двухлетней квазицикличности пожаров, т.е. возникновения ежегодных пожаров большой интенсивности. Достоверность прогнозов подтверждена реальными данными: значительное увеличение числа пожаров в 2011г. – 1863ед. против 830ед. в 2010г., в 2014 г.: количество пожаров - 2314ед., лесная площадь пройденная огнем – 302.3тыс.га, в 2015г.: лесная площадь пройденная огнем – 522тыс.га (данные РОССТАТ «irkutskstat.gks.ru» дата обращения 08.08.2017г., по данным работы [5]-2014год - 2234ед.пожаров, площадь пройденная огнем – 732.656тыс.га, 2015год – 1597ед. пожаров.). Развитие метода дендрохронологии, как инструмента прогнозирования лесопожарной обстановки в Байкальском регионе, показало свою эффективность. Разработанные методики позволили проводить долгосрочное прогнозирование. Представляется целесообразным для дополнения вышеназванных исследований и повышения достоверности прогнозов на краткосрочный период 1 - 3 года применить процедуры экстраполяции, регрессионные - ковариационные алгоритмы, статистические исследования сходимости сформированных моделей в зависимости от исходных данных. Формирование достоверных даже

краткосрочных прогнозов требует обработки множества разнонаправленной информации. В качестве исходной статистики рассматривается статистика пожаров с 1975 по 2014г.г. на территории Иркутской области. Наблюдается аномальность отдельных наблюдений. Особенностью приведенных статистических данных является большой размах выборок. Модели регрессионного анализа позволяют определить тренды количества пожаров, лесной площади пожаров, пройденной огнем, средней площади пожаров. Например, тренд количества пожаров – умеренно возрастающий, тренд лесной площади, пройденной огнем – слабо убывающий, тренд средней площади одного пожара – умеренно убывающий. Применено сглаживание методом скользящей средней. Аномальные и недостоверные (опечатки) данные убраны. В итоге рассмотрены две выборки объемами 26 и 31 единиц наблюдений. Исследуется влияние порядка регрессии на значение прогнозируемого параметра. Проведены стандартные процедуры регрессионного анализа. Формирование регрессионных моделей 1-ого, 2-ого, 3-его, 4-ого порядков выполнялись в среде «Mathcad-14» с использованием стандартных подпрограмм. Приведены результаты прогнозирования количества пожаров на следующий год. Получены следующие результаты: регрессионные модели 3 – его порядка: 202,64тыс.га, 4 – ого порядка: 325.317тыс.га, 5 – ого порядка: 467.4тыс.га. В качестве алгоритма экстраполирования выбрана методика, основанная на формировании многочленов Лагранжа разных степеней. Построены экстраполяционные полиномы от 2-ой до 5-ой степеней. Значения полиномов при экстраполировании количества пожаров на год вперед следующие: $L_2(x) = 1.918 \cdot 103 \text{ед.}$, $L_3(x) = 2.138 \cdot 103 \text{ед.}$, $L_4(x) = 2.096 \cdot 103 \text{ед.}$, $L_5(x) = 1.444 \cdot 103 \text{ед.}$ Проведенные вычисления показали существенную разнонаправленность результатов вычислений полиномами степеней выше 5-той. Сравнение данных экстраполирования и регрессионного анализа: наименьшая относительная погрешность $\delta = 1.53\%$. Вычисления показали недостаточную точность экстраполяции с помощью формул Гаусса и Ньютона и их модификаций применительно к статистикам таблицы 1.5, что объясняется спецификой исходных данных, в основном большим размахом. Относительная погрешность данных таблицы 1.6 и регрессионного анализа: $\delta = 9.421\%$, данных таблицы 1.6 и экстраполирования: $\delta = 0.10.8\%$. Полученные результаты демонстрируют, с одной стороны, целесообразность применения процедур экстраполирования совместно с регрессионными моделями, с другой, необходимость формирования моделей ковариационного анализа, учитывающих резкие изменения исходных данных, связанные с природными аномалиями, прогнозирование которых весьма непростая задача. В целом, целесообразно сформировать прогноз на основании нескольких методик: статистических данных, регрессионного анализа, методик экстраполирования. Анализ статистик лесной площади пройденной огнем, следуя методикам формирования регрессионных моделей разных степеней от первой до четвертой, процедурам экстраполирования полиномами Лагранжа порядков от первого до третьего позволяет выявить в данном примере соответствующие друг другу модели: регрессия третьего порядка и полином Лагранжа первого порядка, которые отличаются на 0.52%. Сравнение с опубликованными данными лесной площади пройденной огнем в 2015г. равной 522тыс.га дает следующие величины относительной погрешности регрессии и экстраполирования: 10.30%, 9.75%, соответственно. Следует заметить соответствие известным исследованиям.