

АВТОМАТИЗАЦИЯ НЕФТЕГАЗОВЫХ ТЕХНОЛОГИЙ

Д.В. Московченко, А.Г. Бабушкин, А.А. Убайдулаев

РЕГРЕССИОННЫЙ АНАЛИЗ ПОСТУПЛЕНИЯ НЕФТИ В ПОВЕРХНОСТНЫЕ ВОДЫ (на примере месторождений ХМАО — ЮГРЫ)

Проведенный регрессионный анализ позволил провести оценку влияния различных техногенных факторов на содержание нефтяных углеводородов (НУВ) в водах рек и озер ХМАО — Югры. Содержание НУВ в максимальной степени зависит от площади земель, загрязненных за весь период эксплуатации месторождения, количества кустовых площадок и объема нефти, поступившей в результате аварий в окружающую среду. Регрессионная модель описывает около половины варьирования концентрации НУВ от совокупности влияющих переменных. Другая половина зависит от природного углеводородного «фона».

Регрессионный анализ, нефтяное загрязнение, гидрохимия, ХМАО — Югра.

Нефтяное загрязнение рек и озер является одной из наиболее актуальных экологических проблем Ханты-Мансийского автономного округа — Югры. Объективный анализ источников и последствий загрязнения возможен на основании математического описания гидрохимических процессов, что соответствует общей тенденции распространения идей и методов кибернетики на прикладные задачи управления процессами природопользования [8]. В последнее время разработаны общие методы математического моделирования нефтезагрязненности водотоков [9]. Существует несколько расчетных формул, основанных на экспоненциальной зависимости концентрации НУВ от площади загрязненных земель, которые позволяют провести количественную оценку миграции нефтяных углеводородов (НУВ) в поверхностные воды после аварийных разливов [10–12]. Однако такие факторы, как утечки из шламовых амбаров, из устья скважин, при этом не учитываются. Также нужно отметить, что содержание НУВ в поверхностных водах определяется не только техногенными факторами, но и зависит от природного фона [4]. Разделение природных и техногенных углеводородов при использовании традиционных методов химических анализов не проводится. Несмотря на появление ряда работ опытно-методического характера, в которых описаны возможные пути решения проблемы идентификации техногенной составляющей в общем составе НУВ с применением соответствующих химико-аналитических методик [3], до настоящего времени отсутствуют сведения о соотношении петрогенных, биогенных и техногенных углеводородов в поверхностных водах, донных отложениях, почвах Западной Сибири. Все это затрудняет объективный анализ экологической ситуации на участках нефтедобычи и путей ее оздоровления.

Одним из математических методов, позволяющих охарактеризовать элемент системы как функцию множества влияющих на него переменных, является регрессионный анализ — метод оценки зависимости случайной величины Y от переменных X_j ($j = 1, 2, \dots, k$), рассматриваемых как неслучайные величины, независимо от закона распределения X_j . Сущность регрессионного анализа составляет описание корреляционных связей, которые характеризуют изменение функции в зависимости от одного или нескольких аргументов [7].

Общую модель многомерной регрессии можно представить следующей формулой:

$$Y = a + \sum_{k \neq 1} \beta_{ik} (x_k - \mu_k),$$

где $\beta_{ik} = \frac{A(K(X_1, \dots, X_n))}{A(\sigma_{X_1, \dots, X_n}^2)}$ — коэффициенты уравнения регрессии при аргумен-

тах, определяемые из соотношения ковариационной матрицы $A(K(X_1, \dots, X_n))$ и матрицы дисперсий $A(\sigma_{X_i}^2)$, т.е. через соотношение вторых моментов многомерного распределения [7].

При анализе содержания нефти в водах рек и озер в качестве влияющих переменных можно использовать данные о факторах, которые вызывают загрязнение. Они довольно многочисленны, а степень их влияния (ранг) не поддается точной оценке и может меняться в зависимости от конкретной экологической ситуации. Поступление НУВ происходит из учтенных и неучтенных экологическими службами источников. Учитываются и отражаются в отчетных материалах сведения об аварийных разливах (их количество, причина, масса загрязнителей), однако поступление нефтяных компонентов происходят за счет миграции и рассеяния и при обычной, неаварийной, эксплуатации нефтепромысловых объектов [6] — из устья скважин, хранилищ топлива, шламовых амбаров. Они в большинстве случаев остаются вне рассмотрения при оценке миграции загрязнителей. Объем неучтенных поступлений косвенно зависит от количества технических объектов, которые являются потенциальными источниками загрязнения. Таким образом, решение задачи требует охвата максимального количества возможных факторов загрязнения — как данных об авариях на объектах инфраструктуры, так и об информации об уровне техногенной нагрузки — количестве техногенных объектов, являющихся потенциальными загрязнителями.

Для построения регрессионной модели были использованы материалы, поступающие от недропользователей в окружной департамент экологии в соответствии с региональными нормативно-правовыми документами, в том числе данные гидрохимического мониторинга и сведения об уровне техногенной нагрузки, которые были обобщены по лицензионным участкам нефтедобычи (ЛУ). В качестве условного допущения было принято, что в пределах ЛУ все техногенные объекты и аварийные разливы относятся к единому водосборному бассейну, охваченному пунктами мониторинговых наблюдений. Это в большинстве случаев соответствует реальности, поскольку расположение пунктов мониторинга в пределах лицензионных участков нацелено на максимально эффективное отражение влияния техногенных объектов, а их плотность зависит от показателей техногенной нагрузки. Достоверность использованной информационной базы подтверждается большой плотностью пунктов мониторинговых наблюдений, которая в настоящее время на территории ХМАО — Югры составляет 1 пункт на 10–20 км² [2], и количеством учтенных в анализе измерений содержания НУВ (более 9 тысяч).

При построении модели была обработана информация за один календарный год (январь–декабрь 2012 г.). Чтобы достоверно оценить влияние техногенеза и исключить влияние случайных вариаций, из общего массива данных были исключены участки, находящиеся на стадии разведочных работ, а также где влияние нефтедобычи на состав поверхностных вод проявляется слабо

и аварийных разливов не наблюдалось. В исследуемый период аварии происходили на 64 участках, их число в пределах ЛУ изменялось от 1 до 514. Как правило, наблюдались аварии с незначительными по объему разливами нефти — менее 1 тонны.

Исходными данными были следующие: в качестве зависимой переменной — среднее содержание НУВ (мг/дм^3) в поверхностных водах лицензионного участка; в качестве влияющих переменных — показатели аварийности и техногенной нагрузки (число аварий в пределах ЛУ; масса поступивших в окружающую среду загрязнителей, в том числе нефти и пластовых вод; площадь загрязненных земель в текущем году и в целом за время эксплуатации промысла; количество добывающих и разведочных скважин, шламовых амбаров, эксплуатационных кустов; протяженность межпромысловых трубопроводов и внутрипромысловых нефтесборных сетей). Обобщенные показатели параметров модели представлены в табл. 1.

Таблица 1

**Показатели техногенной нагрузки на лицензионных участках
распределенного фонда недр ХМАО — Югры
(участки, где наблюдались аварии)**

Показатель	min	max	Всего
Количество эксплуатационных скважин	15	7952	47497
Количество разведочных скважин	2	189	2742
Протяженность межпромысловых нефтепроводов, км	0	461,09	5690,76
Протяженность внутрипромысловых нефтесборных и выкидных линий, км	0	2078,602	15061,47
Количество кустовых площадок	1	1675	6790
Количество шламовых амбаров	0	284	878
Общее количество аварий	1	514	2883
Количество аварий на нефтепроводах	0	198	1262
Масса нефти и нефтепродуктов, поступивших в окружающую среду, тонн	0	46,034	264,578
Масса подтоварных (пластовых) вод, тонн	0	2169,015	4143,281
Площадь земель, загрязненных за все время эксплуатации промысла, га	0	1603,899	3511,929
Площадь земель, загрязненных в отчетном году, га	0	30,314	148,4838

Таким образом, матрица обрабатываемых данных содержала 13 переменных и 64 наблюдения. Обработка материалов проводилась с использованием пакета обработки данных Statistica 6.0.

Построение моделей регрессии в полной мере применимо только в том случае, если распределения близки к нормальным, а отношения линейны [7]. Однако проведенный анализ показал отсутствие нормального распределения практически для всех переменных. Поэтому на предварительном этапе было выполнено приведение данных к нормальному распределению, для чего выполнено преобразование Барлетта (извлечение квадратного корня при повторности варианта больше десяти) [5].

Поскольку регрессионный анализ основан на оценке корреляционных связей, на первом этапе был выполнен корреляционный анализ массива данных. Вычисление коэффициентов корреляции Спирмена показало отсутствие сильных зависимостей содержания НУВ от какого-либо отдельного фактора. Су-

существует слабая корреляционная зависимость между средним содержанием НУВ и общей площадью загрязненных земель ($R = 0,30$), площадью земель, загрязненных в текущем году ($R = 0,34$), количеством разведочных скважин ($R = 0,30$), числом шламовых амбаров ($R = 0,28$) и протяженностью межпромысловых нефтепроводов ($R = 0,38$). Для остальных показателей, в том числе для количества аварий и массы загрязняющих веществ, достоверной зависимости не выявлено.

В табл. 2 приведены показатели модели множественной регрессии, отражающей влияние независимых переменных, характеризующих различные факторы техногенной нагрузки, на зависимую (среднее содержание НУВ в поверхностных водах лицензионного участка).

Таблица 2

Основные показатели регрессионного анализа общего массива данных

Показатель	Значение
Коэффициент корреляции R	0,76
Коэффициент детерминации R^2	0,58
Исправленный коэффициент детерминации R^2_{adj}	0,48
Критерий Фишера $F(12,51)$	5,90
Уровень значимости модели p	0,000003
Оценка средней квадратической ошибки модели	0,0375

Коэффициент детерминации R^2 показывает степень соответствия значений регрессионной модели к наблюдаемым значениям y_j . При величине коэффициента $R^2 = 0,58$ и исправленного коэффициента детерминации $R^2_{adj} = 0,48$ примерно половина вариации содержания НУВ объясняется влиянием факторов техногенеза. Другая половина определяется действием неучтенных в регрессионной модели факторов. Такое соотношение требует проверки адекватности модели, которая выполняется на основании дисперсионного анализа, а мерой проверки является значение критерия Фишера

Проведенная проверка показала, что при $\alpha = 0,95$ и $n(12,51)$ границей доверительного интервала является $F\alpha = 2,30$. Поэтому на основании вычисленного критерия Фишера можно констатировать, что вклад регрессионной модели в описание варьирования является статистически значимым.

Качество регрессионной модели также оценивается по распределению остатков (разница между расчетным и реальным значениями), которое должно подчиняться нормальному закону. На рисунке показана гистограмма распределения остатков, которая дает основания для вывода о нормальном распределении и, следовательно, об удовлетворительном качестве модели.

Очевидно, что не все включенные в модель переменные оказывают равнозначное влияние на зависимую величину (содержание НУВ). Силу влияния каждой переменной характеризует коэффициент Beta, изменяющийся от 0 до 1, а значимость влияния переменных оценивается с помощью t -критерия Стьюдента. Показатели влияния переменных приведены в табл. 3. Исходя из величины коэффициента Beta, максимальная положительная зависимость содержания НУВ наблюдается от таких показателей, как количество аварий на нефтепроводах, количество кустовых площадок, масса разлитой нефти и площадь земель, загрязненных за все время эксплуатации промысла. Однако «вклад» числа аварийных случаев на нефтепроводах по t -критерию статистически незначим, а значимое влияние имеют такие показатели, как количество кустовых площадок, протяженность межпромысловых нефтепроводов, масса разлитой

нефти и площадь земель, загрязненных за все время эксплуатации промысла. Несоответствие силы влияния и значимости некоторых факторов определяется тем, что ряд из них связан сильными корреляционными связями между собой (мультиколлинеарность).

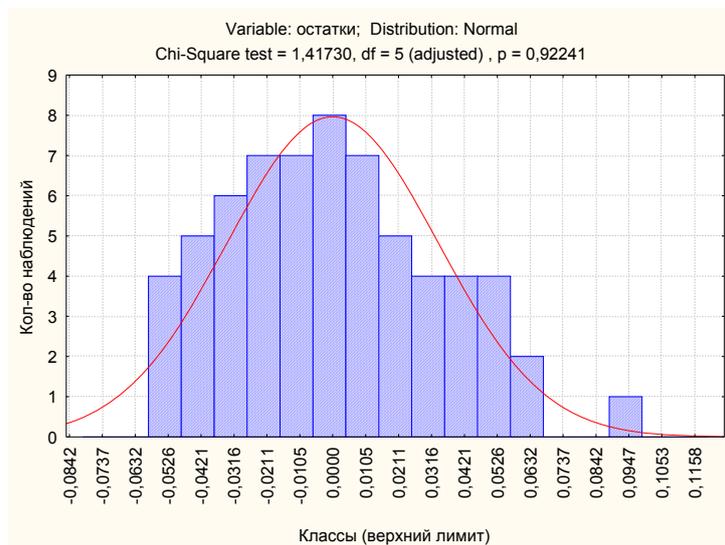


Рис. Оценка нормальности распределения остатков регрессионной модели

Для определения влияния на результирующий показатель отдельных переменных необходимо определение частных коэффициентов корреляции, которые показывают корреляцию каждого аргумента с переменной вне зависимости от ее собственной корреляции с другими переменными, которые косвенно могут увеличивать или уменьшать значение коэффициента корреляции, рассчитываемого просто по варьированию функции и аргумента [7]. В соответствии с этими оценками частная корреляция максимальна для таких показателей, как протяженность межпромысловых трубопроводов, масса нефти, попавшей в окружающую среду при авариях и общая площадь загрязненных земель (табл. 3). Чувствительность (толеранс), которая показывает степень некоррелируемости независимых переменных [7] при больших значениях, может исказить значение коэффициента регрессии, и ее не рекомендуется включать в модель. В нашем случае максимальное значение имеет показатель протяженности межпромысловых трубопроводов.

Обращает на себя внимание отрицательная связь между общим количеством аварий и зависимой переменной, из чего можно сделать парадоксальный вывод о снижении содержания НУВ в поверхностных водах при росте числа аварий. Корреляционный анализ показал практически полное отсутствие связи между этими величинами (ранговый коэффициент корреляции Спирмена $R = 0,11$). Вероятно, отрицательное значение в модели регрессии обусловлено включением в статистику всех случаев аварийности, в том числе не связанных с разливами нефти аварий на водоводах (они в исследуемом году составили 47 % аварийных случаев).

Таблица 3

**Оценка влияния показателей техногенеза
на содержание НУВ в поверхностных водах**

Переменная	Стандартизи- рованный коэффициент Beta	Ошибка Std.Err. of Beta	Частная корреля- ция	Полу- частная корре- ляция	Чувстви- тельность	Частный коэффициент детерми- нации R^2	Коэффици- ент уравнения регрессии B	Ошибка коэффи- циента регрессии B	Критерий Стьюдента $t(51)$	Уровень значи- мо- сти p -level
Константа Intercept	—	—					0,132	0,014	9,13	0,00
Количество эксплуатационных скважин	-0,57	0,31	-0,251	-0,168	0,086	0,914	-0,002	0,001	-1,85	0,07
Количество разведочных скважин	0,30	0,18	0,220	0,146	0,245	0,755	0,006	0,004	1,61	0,11
Протяженность межпромысловых нефтепроводов	0,33	0,12	0,362	0,252	0,577	0,423	0,004	0,001	2,78	0,01
Протяженность внутрипромысловых нефтесборных и выкидных линий	-0,49	0,28	-0,235	-0,156	0,102	0,898	-0,003	0,002	-1,73	0,09
Количество кустовых площадок	0,70	0,36	0,266	0,178	0,064	0,936	0,006	0,003	2,00	0,05
Количество шламовых амбаров	-0,02	0,19	-0,011	-0,007	0,220	0,780	0,000	0,004	-0,08	0,94
Общее количество аварий	-0,49	0,44	-0,153	-0,101	0,043	0,957	-0,005	0,005	-1,11	0,27
Количество аварий на нефтепроводах	0,71	0,46	0,212	0,140	0,040	0,960	0,011	0,007	1,55	0,13
Масса нефти и нефтепродуктов, поступивших в окружающую среду:	0,41	0,18	0,306	0,208	0,257	0,743	0,014	0,006	2,29	0,03
Масса подтоварных (пластовых) вод	-0,29	0,19	-0,211	-0,140	0,232	0,768	-0,002	0,001	-1,54	0,13
Общая площадь нефтезагрязненных земель	0,57	0,23	0,329	0,225	0,155	0,845	0,005	0,002	2,49	0,02
Площадь территории, загрязненной в отчетном году	0,29	0,19	0,211	0,140	0,227	0,773	0,013	0,008	1,54	0,13

Примечание. Выделены показатели, достоверно влияющие на зависимую переменную.

Оптимальное построение регрессионной модели требует исключения малозначущих факторов. Переходя к меньшему числу признаков, но с более высоким информационным содержанием, мы получаем адекватное представление о связях в системе в пространстве меньшей размерности [1]. Рассмотрение показателей регрессионной модели (табл. 3) свидетельствует, что ряд переменных практически не оказывает влияния на содержание НУВ в поверхностных водах. Так, в регрессионной модели фактически отсутствует влияние такой переменной, как количество шламовых амбаров, что свидетельствует о незначительности утечек при соблюдении природоохранных регламентов (амбары должны иметь гидроизоляцию, препятствующую распространению их содержимого). Средства программного пакета Statistica 6.0 позволяют провести исключение переменных, вклад которых по частной корреляции невелик и исключение которых не снижает качества модели, методом «шаг назад». Использование этой опции выявило возможность отбросить из модели такие показатели, как количество эксплуатационных скважин, общее число аварий и протяженность внутрипромысловых нефтесборных линий. Низкий вклад этих показателей объясним с учетом факторов экологического риска: загрязнение от эксплуатационных скважин менее выражено, чем от разведочных; в статистику общего числа аварий входят и те аварии, которые не сопровождаются разливами нефти и подтоварных вод.

Таким образом, содержание НУВ в поверхностных водах в максимальной степени зависит от площади земель, загрязненных за весь период эксплуатации месторождения, количества кустовых площадок и объема нефти, поступившей в результате аварий в окружающую среду. Физически это хорошо объяснимо: поступление НУВ в поверхностные воды происходит из разливов, общая площадь которых отражает суммарное загрязнение за многолетний период. Кустовые площадки, на которых зачастую наблюдаются утечки из скважин, и погребенные в толще насыпного основания разливы являются источником дополнительного загрязнения. Показателем, который по всем критериям (силе влияния, значимости, частному коэффициенту корреляции, чувствительности) должен быть включен в модель, является общая площадь нефтезагрязненных земель. Это подтверждает правомерность использования такого показателя, как площадь замазученности, для подсчета модуля смыва нефти [10]. Следует также обратить внимание на тот факт, что модель описывает около половины варьирования концентрации НУВ от совокупности значений влияющих переменных (исправленный коэффициент детерминации $R^2_{adj} = 0,48$), то есть существуют реально действующие факторы, не описанные моделью. Очевидно, что неучтенными факторами являются природные источники, формирующие углеводородный «фон» территории.

ЛИТЕРАТУРА

1. Аренс Х., Лейтер Ю. Многомерный дисперсионный анализ. М.: Финансы и статистика, 1985. 230 с.
2. Бабушкин А.Г., Московченко Д.В., Пикунов С.В. Гидрохимический мониторинг поверхностных вод Ханты-Мансийского автономного округа — Югры // Новосибирск: Наука, 2007. 152 с.
3. Кульков М.Г., Артамонов В.Ю., Коржов Ю.В., Углев В.В. Индивидуальные органические соединения нефти как индикаторы техногенного нефтяного загрязнения водной среды // Известия Томского политехнического университета. 2010. Т. 317. № 1. С. 196–200.

4. Кузнецов А.Н., Федоров, Ю.А. Нефтяные компоненты в устьевой области р. Дон и в Азовском море (результаты многолетних исследований) // Водные ресурсы, 2014, т. 41, № 1, С. 49–59.
5. Любичев А.А. Дисперсионный анализ в биологии. М.: Изд-во Моск. ун-та, 1986. 200 с.
6. Паничева Л.П., Моисеенко Т.И., Кремлева Т.А., Волкова С.С. Биохимическая трансформация нефтяных углеводородов в водах Западной Сибири // Вестник ТюмГУ, 2012. № 12. С.38–48.
7. Пузаченко Ю.Г. Математические методы в географических и экологических исследованиях. М.: Изд. Центр «Академия», 2004. 416 с.
8. Соловьев И.Г. Концепции и методы кибернетики в задачах природопользования // Вестник кибернетики, 2010, № 9. С. 4–12.
9. Соловьев И.Г., Пикиноров П.В., Шмелева Т.А. Математические методы контроля нефтезагрязненности водотоков на месторождениях с длительными периодами ледостава. I. Модель // Криосфера Земли, 2008, т. XII, № 2. С. 25–31.
10. Соромотин А.В. Техногенная трансформация природных экосистем таежной зоны в процессе нефтегазодобычи (на примере Тюменской области) // Автореф. дис. ... д.б.н. Тюмень, 2007. 47 с.
11. Хорошавин В.Ю. Загрязнение нефтепродуктами малых рек бассейна Пура рассредоточенными источниками // Геоэкологические проблемы Тюменского региона. Вып. 1. Тюмень: Вектор Бук, 2004. С. 105–130.
12. Хорошавин В.Ю. Прогноз формирования качества речных вод под влиянием рассредоточенных источников нефтепродуктов // Вестник ТюмГУ, 2010. № 7. С. 5–19. С. 153–161.

Moskovchenko D.V., Babuschkin A.G., Ubaydullaev A.A.

*Regression analysis of the oil input into the surface waters
(an example of KHMАО — Yugra deposits)*

Regression analysis has allowed to assess the impact of various anthropogenic factors on the content of total petroleum hydrocarbons (TPH) in the waters of the rivers and lakes in Khanty-Mansy Autonomous Okrug — Yugra. The content of TPH to the maximum extent depends on the area of land contaminated during the whole period of exploitation, the number of well pads and quantity of oil received as a result of accidents in the environment. The regression model describes about half of the variation of the concentration TPH the aggregate of influencing variables. The other half depends on the natural hydrocarbon «background».

Regression analysis, oil pollution, hydrochemistry, Khanty-Mansy Autonomous Okrug-Yugra.