ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

УДК 621.396.6

О.Б. Качалов, К.Ю. Плесовских, Н.П. Ямпурин

СИНТЕЗ УСТОЙЧИВОЙ МОДЕЛИ ДЛЯ РАСЧЕТА РАСХОДА НЕФТЕВОДОГАЗОВОГО ПОТОКА НА ОСНОВЕ УРАВНЕНИЯ РЕГРЕССИИ

Арзамасский политехнический институт (филиал) НГТУ им. Р.Е. Алексеева

Предлагается метод снижения числа обусловленности матрицы данных для получения устойчивой модели расхода нефтеводогазового потока.

Ключевые слова: многокомпонентный поток, регрессионное уравнение, математическая модель, число обусловленности, устойчивая модель.

В нефтедобывающей промышленности, как ни в какой другой отрасли, без анализа расходных параметров технологических процессов невозможно управление практически ни одним процессом, будь то бурение или цементирование скважины, добыча или транспортировка нефти, нагнетание воды в системах поддержания пластового давления и т.п. Измерение расхода добываемых флюидов из нефтяного пласта служит не только основой для решения задач оптимизации процесса извлечения нефти, но и информационным средством для создания систем контроля за разработкой месторождения [1].

Большую роль измерение расхода играет и во вспомогательных технологических процессах – от бурения скважины до сдачи подготовленной нефти в трубопродные транспортные системы.

Измерение расходных параметров всех компонент продукции нефтяной скважины является необходимым для решения двух важных задач:

1) оперативного контроля, в результате которого обнаруживаются внезапные и постепенные отказы в работе скважин и прискважинного оборудования;
2) контроля и регулирования разработки нефтяного месторождения [1].

Решение второй задачи предполагает и учет добываемой продукции, точность которого напрямую влияет на величину налоговой базы недропользователя. В связи с этим разработан, но до сих пор не введен в действие ГОСТ Р. 8.615- [2]. В ГОСТ Р. 8.615-2005 установлены пределы допускаемой основной относительной погрешности измерений массы нефти и зависимости от содержания воды по отдельной скважине от 2,5 до 15%.

Существует несколько способов измерения расхода нефти, однако их применение затрудняется сложным составом извлекаемой нефти.

Нефть из скважин всегда поступает с газом, поэтому ее относят к двухфазным средам. Как правило, продукт, выходящий из нефтяных скважин, состоит не только из нефти и газа, но и сопутствующей им воды, т.е. является трехкомпонентным.

Решение этой проблемы осуществляется двумя путями.

Первый предполагает разделение потока на отдельные компоненты с последующим прямым измерение каждого компонента в отдельности. Такой подход дает наиболее точные результаты, но самый дорогостоящий, длительный во времени, требует установки специального оборудования и потенциально небезопасен для экологии в силу возможной утечки нефти из измерительных установок (рис. 1, *а*).

Второй способ заключается в том, что поток беспрепятственно движется в нефтяной трубе, измерительная система (рис. 1, *б*) напрямую измеряет определенные параметры потока, а затем по полученным данным косвенным путем рассчитывает покомпонентный расход содержимого. Получение данных при этом возможно в режиме реального времени, замеры могут осуществляться бесконтактными способами, сохраняя при этом целостность трубопровода, что существенно снижает вероятность утечки нефти. Стоимость подобных измерительных систем значительно ниже, чем у сепарационных установок, применяемых в первом способе, но техническая реализация расчета расхода нефтеводогазового потока сама по себе при этом является сложной научно-технической задачей.

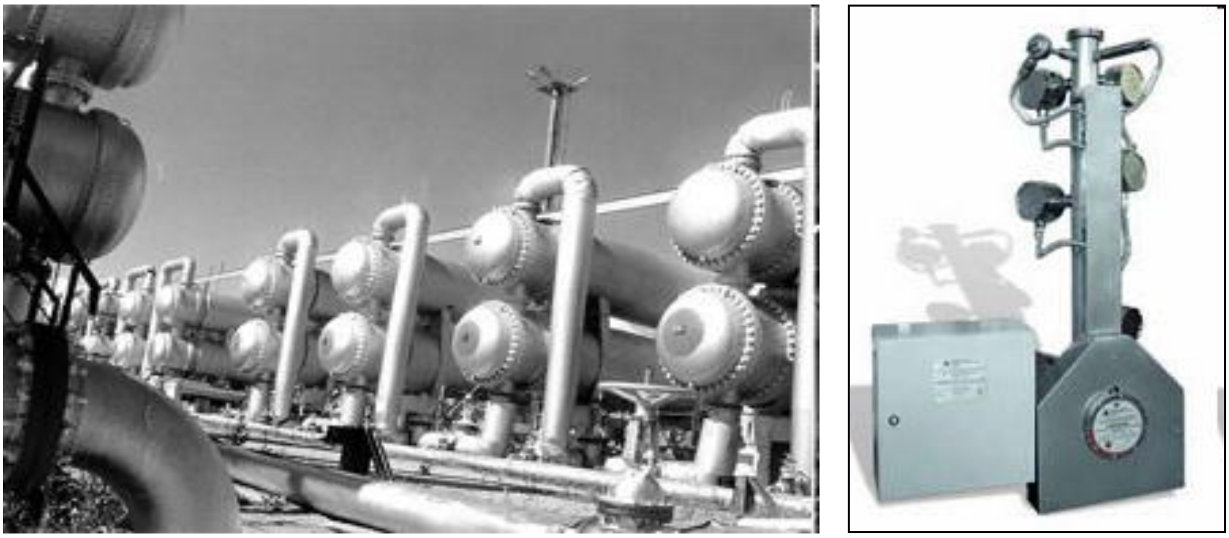

Рис. 1. Сепарационная установка (*а*) и измерительная система «Ультрафлоу» (*б*)

Дополнительные сложности в расчет вносит и тот факт, что структура двухфазного потока зависит от различных технологических, технических и других параметров: скорости потока, давления, диаметра трубопровода, его расположения в пространстве и процентного содержания той или другой фазы.

При успешном разделении калибровочных данных на отдельные режимы движения [3] потока, решение задачи сводится к построению многофакторной математической модели. И одним из способов построения при определенных условиях может являться нахождение уравнения регрессии.

Полученная модель при этом должна являться устойчивой, т.е. обеспечивать результаты расчетов, с отклонением от истинных значений в пределах диапазона, устанавливаемого ГОСТ Р. 8.615-2005.

Однако при построении математических моделей, основанных на уравнениях регрессии, возникает проблема мультиколлинеарности экспериментальных данных. Мультиколлинеарность проявляется в сильной корреляции между двумя или более признаками, что затрудняет оценивание параметров модели.

Мультиколлинеарность факторов является основной причиной того, что модель имеет значительные погрешности, а в некоторых случаях решаемая задача становится некорректно поставленной и требует специальных методов её решения.

Существует множество методик, позволяющих проверить устойчивость модели, но ни одна из них не является универсальной. Например, для проверки гипотезы об устойчивости результатов может быть использован критерий Уилкоксона, который служит для проверки того, относятся ли две выборки к одной и той же генеральной совокупности, т. е. обладают ли они одним и тем же статистическим признаком [4].

Критерий применим в тех случаях, когда признаки измерены, по крайней мере, в порядковой шкале. Целесообразно применять данный критерий, когда величина самих сдвигов варьируется в некотором диапазоне (10-15% от их величины). Это объясняется тем, что разброс значений сдвигов должен быть таким, чтобы появлялась возможность их ранжирования. В случае если сдвиги незначительно отличаются между собой, и принимают какие-то конечные значения, например. +1, -1 и 0, формальных препятствий к применению критерия нет, но ввиду большого числа одинаковых рангов ранжирование утрачивает смысл, и те же результаты проще получить с помощью критерия знаков.

Суть метода состоит в том, что сопоставляются абсолютные величины выраженности сдвигов в том или ином направлении. Для этого сначала все абсолютные величины сдвигов ранжируются, а потом суммируются ранги. Если сдвиги в ту или иную сторону происходят случайно, то и суммы их рангов окажутся примерно равны. Если же интенсивность сдвигов в одну сторону больше, то сумма рангов абсолютных значений сдвигов в противоположную сторону будет значительно ниже, чем это могло бы быть при случайных изменениях.

Недостатки данного метода:

- малый объем выборки – от 5 до 50 элементов;
- нулевые сдвиги исключаются из рассмотрения;
- сдвиг в более часто встречающемся направлении принято считать «типичным» и наоборот [5].

Другим, менее известным способом, является оценка числа обусловленности cond матрицы плана эксперимента [6].

Число обусловленности является мерой чувствительности системы линейных уравнений к погрешностям задания вектора правых частей уравнений:

$$\mathrm{A}x = \mathrm{b}, \tag{1}$$

где A – определяемая матрица данных; b – вектор правых частей уравнений; x –результат решения системы.

Число обусловленности вычисляется по формуле

$$\mathrm{cond(A)} = \left\|\mathrm{A}^{-1}\right\| \left\|\mathrm{A}\right\|, \tag{2}$$

где $\|\mathrm{A}\|$ - норма матрицы A; $\left\|\mathrm{A}^{-1}\right\|$ - норма обратной матрицы A.

Чем больше число обусловленности cond, тем сильнее это воздействие и тем более неустойчив процесс нахождения решения. Хотя число обусловленности матрицы зависит от выбора нормы, если матрица хорошо обусловлена, то её число обусловленности будет мало при любом выборе нормы, а если она плохо обусловлена, то её число обусловленности будет велико при любом выборе нормы.

В работе [6] показано, что устойчивая модель имеет число обусловленности cond порядка 1. При cond от 1 до 10 модель имеет хорошую устойчивость, а от 10 до 100 - удовлетворительную устойчивость. При cond больше 100 модель имеет неудовлетворительную устойчивость.

В работе С. Радченко, П. Бабича получена устойчивая модель (число обусловленности cond = 1)[7]. Однако метод ее получения не описан, что не позволяет применять его на практике. В данной работе авторами предлагается альтернативный способ приведения числа обусловленности к единице, на основе которого синтезирована устойчивая модель для расчета расходов нефтеводогазового потока. Приведение матрицы к числу обусловленности, равному единице, осуществляется в три этапа.

На первом этапе проводится нормировка данных по формуле

$$\overline{Z} = \frac{Z - Z_S}{\sigma},$$

где значение признака $\overline{Z}$ – нормированное; Z_S – среднее; Z – абсолютное; σ – среднеквадратическое.

На втором этапе для нормированной матрицы строятся главные компоненты. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных. Соответствующие собственные числа этой матрицы равны дисперсиям проекций множества объектов на оси главных компонент.

Метод главных компонент осуществляет переход к новой системе координат $y_1,...,y_p$ в исходном пространстве признаков $x_1,...,x_p$, которая является системой ортонормированных линейных комбинаций.

В основе модели для выражения исходных признаков через компоненты лежит предположение о том, что число компонент равно числу исходных признаков:

$$\begin{aligned} X_1 &= a_{11}F_1 + ... + a_{1m}F_m \\ &\vdots \\ X_m &= a_{m1}F_1 + ... + a_{mm}F_m. \end{aligned} \tag{2}$$

Компоненты $F_1,...,F_m$ в модели (2) предполагаются независимыми стандартизованными показателями, распределенными по нормальному закону.

Очевидно, система уравнений (2) определяет здесь систему преобразования одних параметров в другие, поэтому каждое значение каждого исходного признака будет отображено в главных компонентах в соответствующей пропорции относительно других значений, определяемой весами $a_{11} – a_{mm}$.

Линейные комбинации выбираются таким образом, что среди всех возможных линейных нормированных комбинаций исходных признаков первая главная компонента $y_1(x)$ обладает наибольшей дисперсией. Геометрически это выглядит как ориентация новой координатной оси y_1 вдоль направления наибольшей вытянутости эллипсоида рассеивания объектов исследуемой выборки в пространстве признаков $x_1,...,x_p$.

Вторая главная компонента имеет наибольшую дисперсию среди всех оставшихся линейных преобразований, некоррелированных с первой главной компонентой. Она интерпретируется как направление наибольшей вытянутости эллипсоида рассеивания, перпендикулярное первой главной компоненте. Следующие главные компоненты определяются по аналогичной схеме [9, 10]. В результате формируется матрица, в которой дисперсия уменьшается с увеличением номера столбца.

На третьем этапе вновь проводится нормировка матрицы по указанной выше формуле, и в результате получается матрица с числом обусловленности равным единице.

Для того чтобы показать преимущества использования предложенного метода в практических расчетах, проведем сравнение ошибок моделей при изменении числа обусловленности. В расчетах использовались экспериментальные данные, полученные при калибровке прибора «Ультрафлоу» (табл. 1).

Таблица 1

Экспериментальные данные калибровки прибора «Ультрафлоу»

Дебит жидкости, м^3/сут	Сдвиг Доплера	Обводненность, %	Газонасыщенность, доли	Давление, МПа	Температура жидкости, °C
1	2	3	4	5	6
10.10	9150.06	19.98	0.834	0.355	26.5
30.19	6503.24	10.15	0.869	0.401	25.3
30.30	7423.88	10.30	0.908	0.403	25.1
30.03	8335.41	10.05	0.930	0.450	25.1

Окончание табл. 1

1	2	3	4	5	6
10.14	6646.71	20.95	0.983	0.502	25.3
30.04	10016.51	10.20	0.952	0.502	25.2
30.27	11827.39	8.19	0.964	0.513	24.5
30.12	12711.44	10.14	0.971	0.571	25.1
30.21	12374.20	28.98	0.971	0.577	24.1
30.35	10509.01	28.73	0.963	0.539	24.6
30.01	8528.85	29.6	0.952	0.497	25.1
30.39	6906.47	29.19	0.929	0.451	24.9
30.32	6234.30	29.25	0.908	0.418	25.4
10.64	5153.51	28.46	0.966	0.417	24.7
30.43	8959.76	51.15	0.963	0.500	24.8
30.30	11408.7	51.48	0.971	0.592	24.5

Построим на основе одного и того же регрессионного уравнения две различных модели – для данных с числом обусловленности, приведенным к единице, указанным выше способом, и данных с большим числом обусловленности.

Исходное регрессионное уравнение имеет вид

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_{13}x_1x_3 + b_{11}x_1^2 + b_{33}x_3^2, \tag{3}$$

где y – расход жидкости; x_1 – доплеровский сдвиг частоты; x_2 – обводненность жидкости; x_3 – газонасыщенность жидкости; x_4 – абсолютное давление в интервале замера; x_5 – температура в интервале замера.

Изменение числа обусловленности матрицы плана эксперимента осуществляется за счет изменения объема обучающей выборки (числа используемых при обучении экспериментальных точек).

На основе регрессионного уравнения с помощью модели 1(cond = 1) и модели 2 (cond = 55,9) были восстановлены значения расхода жидкости и проведено сравнение полученных значений с исходными. В табл. 2 приведены погрешности модели 1 и 2. Можно заметить, что погрешность второй модели существенно выше, чем у первой.

Таблица 2

Погрешности модели 1 (cond = 1) и модели 2 (cond = 55,9)

Модель 1, %	Модель 2, %
1	2
2,21	4,29
-0,72	0,01
-4,47	2,83
2,25	-7,59
0,22	1,23

Окончание табл. 2

1	2
-0,22	-4,24
-1,09	-9,7
-1,9	10,75
3,33	-6,55
2,25	-7,59
0,22	1,23
-0,72	0,01
-0,22	-4,24
-0,68	-0,91

Далее будем изменять число обусловленности матрицы исходных данных, по-прежнему приводя для первой модели матрицу к cond=1, а для второй - оставляя без изменения. Проследим, как увеличение числа обусловленности будет способствовать увеличению погрешности. В табл. 3 приведены отношения средних значений ошибок, среднеквадратических отклонений ошибок и максимальных абсолютных значений ошибок для моделей 1 и 2. Как видно из табл. 2 и 3 с увеличением числа обусловленности матрицы плана эксперимента ошибки регрессионной модели с cond > 10 существенно возрастают.

Таблица 3

Зависимость отношений ошибок от числа обусловленности cond

cond	Отношение средних значений ошибок	Отношение среднеквадратических отклонений ошибок	Отношение максимальных абсолютных значений ошибок
7,1	0,9	1,1	1,2
39,8	24,3	4,2	4,1
55,9	44,5	2,8	3,2
63,6	42,4	4,4	3,7
71,3	71,6	5,0	9,7

Таблица 4

Погрешность модели 1 при изменении числа обусловленности

cond	Средняя погрешность $e \cdot 10^3$	Среднеквадратическое отклонение ошибок σ	Максимальная погрешность $\max(e)$
39,8	0,49	0,0196	0,0346
58,0	0,37	0,0225	0,0374
99,6	0,52	0,0220	0,0396
405,8	0,45	0,0205	0,0376

Данные о погрешности модели 1 при изменении числа обусловленности исходных данных, приведенные в табл. 4, свидетельствуют о том, что модель 1 имеет практически одинаковые погрешности при изменении cond от 39,8 до 405,8. Это указывает на то, что, даже имея исходные данные с большим числом обусловленности, можно добиться построения устойчивой регрессионной модели, если исходные данные подвергнуть обработке предложенным способом.

Проведенные исследования позволяют сделать следующие выводы:

1) при числе обусловленности cond > 10 модели, основанные на регрессионных уравнениях, дают значительные погрешности;

2) устойчивая модель, в которой матрица плана эксперимента с большим числом обусловленности преобразуется в матрицу с числом обусловленности cond = 1, имеет допустимую погрешность. Эта модель рекомендуется для практического использования при расчетах расхода нефтеводогазового потока.

Библиографический список

1. **Абрамов, Г.С.** Практическая расходометрия в нефтяной промышленности / Г.С. Абрамов, А.В. Барычев. – М.: ОАО «ВНИИОЭНГ», 2002. – 460 с.
2. ГСОЕИ. ГОСТ Р 8.615-2005. Измерение количества извлекаемой из недр нефти и нефтяного газа. Общие метрологические и технические требования. – М.: Стандартинформ, 2005.
3. **Второва, И.А.** Обработка многомерного сигнала на основе метода главных компонент / И.А. Второва, О.Б. Качалов, К.Ю. Плесовских // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева. 2010. №3(82). С. 21–26.
4. Вероятностные методы в инженерных задачах: справочник / А. Н. Лебедев [и др.]. – СПб.: Энергоатомиздат, 2000. – 333 с.
5. **Пригарин, С. М.** Методы численного моделирования случайных процессов и полей / С. М. Пригарин. – Новосибирск: Изд-во ИВМиМГ СО РАН, 2005. – 259 с.
6. **Дикусар, В.В.** Некоторые численные методы решения линейных алгебраических уравнений // Соросовский образовательный журнал. 1998. № 9. С. 111–120.
7. **Радченко, С.Г.** Устойчивые методы оценивания статистических моделей / С.Г. Радченко. – Киев: ГШ «Санспарель», 2005. – 504 с.
8. **Радченко, С.Г.** Анализ экспериментальных данных на основе использования многофакторных статистических математических моделей // Математические машины и системы. 2005. № 3. С. 102–115.
9. **Сизиков, В. С.** Устойчивые методы обработки результатов измерений : учеб. пособие / В.С. Сизиков. – СПб.: СпецЛит, 1999. – 240 с.
10. **Дюк, В.А.** Компьютерная психодиагностика / В.А. Дюк. – СПб., 1994.
11. Прикладная статистика. Классификация и снижение размерности / С.А. Айвазян [и др.]. – М.: Финансы и статистика, 1989. – 607 с.

Дата поступления
в редакцию 12. 07.2011

O.B. Kachalov, K.Y. Plesovskikh, N.P. Yampurin

SYNTHESIS OF SUSTAINABLE MODELS FOR CALCULATION OF FLOWOIL FLOWS BASED ON THE REGRESSION EQUATION

The paper proposes a method to reduce the number of a matrix of data to obtain a stable oil-water flow model of flow.

Key words: multi-thread regression equation, the mathematical model, condition number, stable model.