

ПРОГНОЗИРОВАНИЕ ПЕТРОФИЗИЧЕСКИХ ХАРАКТЕРИСТИК ОБЪЕКТОВ ЭКСПЛУАТАЦИИ ПХГ НА ОСНОВЕ МАРКОВСКОГО МНОГОФАКТОРНОГО МОДЕЛИРОВАНИЯ

М.Б. Матушкин, А.Г. Черников (ООО «Газпром ВНИИГАЗ»)

Для ведения мониторинга функционирования подземных газовых хранилищ (ПХГ) и оптимизации процессов их эксплуатации необходимы адекватные геолого-технологические модели объектов, существенной частью которых являются модели фильтрационно-емкостных свойств (ФЕС). Для получения оценок ФЕС объектов эксплуатации ПХГ необходимо решить задачу достоверного определения их по данным ГИС.

Для моделирования ПХГ ввиду больших объемов исходной информации, распределенной как в пространстве, так и во времени, целесообразно применять пакетную обработку данных. Эта технология позволяет включать в обработку и интерпретацию единовременно весь объем имеющейся, специально подготовленной (стандартизированной) исходной информации. Авторы предлагают программную реализацию этой технологии, разработанную на основе математического аппарата марковских процессов, выполненную в варианте системы марковского многофакторного прогнозирования – моделирования [2, 4].

Теория марковских процессов, разработанная в начале XX века русским математиком А.А. Марковым и глубоко проработанная применительно к наукам о Земле советским геологом и математиком А.Б. Вистелиусом [1], пока еще непросто и мало используется при обработке и интерпретации геолого-геофизических данных. Возможность применения марковской статистики в этой области основывается на нижеследующих рассуждениях.

Во множестве природных процессов (в том числе и геологических) наблюдается влияние предшествующих событий на последующие. То есть вероятность пребывания любой системы в определенном состоянии в заданный момент времени можно вывести из вероятностей предшествующих состояний. Именно эта зависимость отличает марковский процесс от серии независимых, случайных событий. Частным случаем марковского процесса является цепь Маркова – ее можно рассматривать как последовательность дискретных состояний во времени или пространстве, для которых вероятность перехода из одного состояния в другое зависит от предшествующего состояния. В геологии чередование слоев литотипов в стратиграфических последовательностях является примером подобного процесса.

Следовательно, изучая определенным образом марковские свойства последовательности произошедших событий, представляется возможным предсказать какое-либо событие, еще не произошедшее, но вероятность появления которого отлична от нуля. Термин «предсказание» употребляется здесь не в том смысле, что знание настоящего обеспечивает детерминированное предсказание будущего, оно лишь позволяет дать настолько хорошее вероятностное предсказание будущего, что это предсказание не может быть улучшено никакой дополнительной информацией, исходящей из прошлого. Впрочем, детерминированное предсказание также относится к марковскому предсказанию как частный случай. [1]

При изучении последовательности геотехнических событий можно рассматривать любые интервалы в качестве «прошлого» ($t - 1$) и «настоящего» (t) и тем самым воспользоваться марковским свойством для предсказания состояния системы на основе изменения состояний во времени (в разрезе) ее элементов. Аналогичный подход можно использовать для любых чередований событий, определенных на какой-либо шкале последовательностей [2, 3]. Так, например, комплекс геофизических методов можно представить в виде временной последовательности измерений значений переменных (ГК-ГГК-КС-АК-КМ), а значения зарегистрированных этим комплексом геофизических характеристик в конкретной точке разреза скважины можно записать в виде последовательности измеренных геофизических величин ($I_{\gamma}, I_{\gamma\gamma}, \rho_k, \Delta T, D_c$). В качестве базы проведения измерений принимается абстрактное «время». Построение марковской модели заключается в следующем: состояние (например, значения измеренных свойств) в последовательности обозначаем A_{it} ($t = 1, 2, 3, \dots, n$). Последовательность закодированных измерений A_{it} смещаем на 1 шаг и получаем

смещенную последовательность A_{j-1} ($j = 1, 2, \dots, N$), для i и j идентичная индексация. Совместив две последовательности A_{ii} и A_{i-1} между точками t и $t-1$, получим вектор G_{ijk} переходных состояний A :

$$G_{ijk} = \{A_{ii} A_{j-1}\}; k(1, 2, 3, \dots, n).$$

Следующий этап – составление частотной матрицы $|ij|$ ($N \times N$). Элементы матрицы ij ($i, j = 1, 2, \dots, N$) равны числу пар (A_t, A_{t-1}) , ($t = 1, \dots, t-1$) таких, что $A_t = i$ и $A_{t-1} = j$, i – строка частотной матрицы показывает число переходов из i -го состояния в любое возможное состояние, а j -й столбец – число переходов из любого возможного состояния в $i-1$. Накапливая все реализации последовательностей событий, полученных путем наблюдения для исследуемого объекта (системы), получаем матрицу ($N \times N$) накопленных частот, которая впоследствии преобразуется в матрицу переходных вероятностей следующим образом:

$$P_{ij} = \frac{G_{ij}}{\sum G_{ij}}, (i = 1, \dots, N), (j = 1, \dots, N).$$

В знаменателе – сумма элементов i -й строки частотной матрицы. В матрице переходных вероятностей i -я строка показывает вероятности переходов из i -го состояния в любое возможное j -е состояние, а j -й столбец – вероятности переходов из любого возможного состояния в i -е. При $i = j$ показана вероятность продолжительности нахождения переменной в состоянии i . Сумма вероятностей в строке равна 1.

Полученная матрица переходных вероятностей представляет собой формализованное отображение основных категорий, характеризующих исследуемую систему – состав, строение, состояние. Так, например, для чередования литотипов в разрезе номер ячейки матрицы определяет литологический состав, распределение вероятностей по полю матрицы отражает строение, а изменение значений вероятностей переходов – состояние в интервале $i(t_0 \dots t_n)$. Полученная таким образом матрица переходных вероятностей должна быть идентифицирована по какому-либо свойству A , характеризующему изучаемый объект (систему).

Обратная задача – распознавание представляет собой отнесение предъявленного объекта по его формализованному описанию к одному из заданных классов (обучение с учителем). Математическая постановка задачи распознавания сводится к следующему. Пусть задано множество M состояний системы, а на этом множестве существует разбиение на конечное число подмножеств M (классов) Ω_i , $i = 1, 2, \dots, m$; $M = U(\Omega_i)$ $i = 1$. Разбиение определено не полностью. Задана лишь некоторая информация I_0 о классах Ω_i . Объекты ω задаются значениями некоторых признаков X_i , $i = 1, \dots, N$ (этот набор всегда один и тот же для всех объектов, рассматриваемых при решении определенной задачи). Совокупность значений признаков X_i определяет описание $I(\omega)$ объекта. В качестве признаков X_i нами принимается вектор марковской последовательности G_{ij} , определенный выше. Описание объекта $I(\omega) = A_{ij}(\omega)$ принимает значение из множества допустимых значений (таблица обучения).

Задача распознавания со стандартной информацией состоит в том, чтобы для данного объекта ω и набора классов $\Omega_1, \dots, \Omega_m$ по обучающей информации $I_0(\Omega_1, \dots, \Omega_m)$ о классах и описанию $I(\omega)$ вычислить значения вероятности отнесения объекта $\Omega?$ к одному из классов Ω_i [3].

При построении модели производится накопление выделенных классов Ω_i векторов марковской последовательности G_{ijk} . Таким образом, решение прямой задачи сводится к составлению образов интервальных значений прогнозируемого показателя. Каждый интервал значений Ω_i определяется последовательностью признаков (состояний), в качестве которых выступают формализованные наблюдаемые события.

Суммированием признаков в ячейках ij матрицы достигается тот эффект, что наиболее характерные (часто повторяющиеся) признаки получают наибольшие значения вероятности, а незначительные признаки (шумы), случайно распределенные по элементам матрицы, обладают малыми значениями вероятности.

Решение обратной задачи – прогноз параметра по марковской последовательности A_{ij} учитываемых событий B_i основывается на предположении о стационарности марковских цепей.

В стационарной марковской цепи P_{ij} – вероятность перехода системы из состояния A_i в состояние A_j в момент t – есть величина постоянная ($P_{ij} = \text{const}$). В нестационарной марковской цепи эти вероятности изменяются во времени и в пространстве. При стандартном описании объекта (постоянство числа переходов ($t = \text{const}$) и локализация цепи в пространстве признаков Ω_i) возможно допущение, что полученные матрицы переходных вероятностей обладают свойством стационарности:

$$P(Q_{ij})\Omega_i = \text{const}.$$

В этом случае имеется возможность численно определить вероятность перехода системы из состояния A_1 в состояние A_t за t шагов. Эта величина определяется произведением вероятностей перехода на каждом из шагов:

$$P_t = \prod_{i=1}^t P_{ij} = (1, \dots, t).$$

Таким образом, представляется возможным для каждой матрицы переходных вероятностей Q_k , относящейся к определенному значению Ω_k , рассчитать для последовательности G_{ijrk} вероятность перехода из состояния A_1 в состояние A_t . Принимая B_k полной группой событий ($P(B_k) = 1$), рассчитаем для вектора последовательности G_{ijrk} по набору матриц, относящихся к различным классам Ω , распределение вероятности по шкале показателя B . Полученные значения $P(\Omega)$ являются гипотезами. В качестве решения принимаются гипотезы с максимальными значениями вероятности:

$$W = \max P(\Omega_k).$$

Таким образом, прогнозирование включает два этапа:

1. Моделирование, заключающееся в создании марковского образа структурированной совокупности геолого-геофизических факторов, полученных в процессе геолого-геофизических исследований и соответствующих им параметров, полученных опытным путем, в процессе интерпретации или по результатам экспертных оценок.

2. Распознавание, заключающееся в том, что по структурированной совокупности полученных в результате наблюдения геолого-геофизических факторов дается вероятностный прогноз тех параметров, для которых прямое наблюдение или интерпретация являются невозможными или нецелесообразными по технологическим и (или) экономическим соображениям.

На первый взгляд, описанная система представляет собой одну из возможных реализаций байесовского подхода к решению задачи классификации, использующей марковские матрицы переходных вероятностей в виде предикторов. Однако существенной особенностью нашего алгоритма является учет структуры исследуемой системы.

В понятиях системологии марковская последовательность представляет собой реализацию системы в виде списка наблюдаемых переменных на базе, выраженной через абстрактное «время». Переменные выражены через их определенные (формализованные) значения, а их последовательность представляет собой вектор, идентифицированный меткой-параметром. Векторы переменных выполняются в виде графов. Графы подразделяются на простые и составные.

Простые графы бывают двух видов: открытые и замкнутые. Открытые графы представляют собой нормальную временную последовательность переменных. Примером такой последовательности служит порядок залегания литотипов в стратиграфическом разрезе:

$$L1 \Rightarrow L2 \Rightarrow L3 \Rightarrow \dots \Rightarrow Ln.$$

Замкнутые графы характеризуются тем, что они имеют на входе и выходе одну и ту же переменную. Примером замкнутой последовательности являются геофизические наблюдения, начинающиеся с замера на эталоне и им же завершающиеся:

$$I1 \Rightarrow I2 \Rightarrow I3 \Rightarrow \dots \Rightarrow I1.$$

Составные переменные представляют собой различные комбинации простых и замкнутых графов. Сущность составного вектора заключается в том, что он представляет собой не одну систему, а компо-

зицию, состоящую из систем, подсистем и подподсистем. В составных векторах возможны два варианта сочленения: система – система и система – подсистема. В первом варианте имеет место последовательное сочленение двух векторов. В этом случае желательно иметь между двумя последовательностями согласующую переменную. В качестве такой переменной может выступать фаза, общая для обеих систем, либо при невозможности ее определения вводится нейтральная переменная.

Включение подсистемы в состав системы осуществляется следующим образом. Система есть список переменных. Следовательно, включение в нее подсистемы является с точки зрения иерархии систем включением подподсистемы. Практически это выполняется таким образом, что в наборе переменных выбирается та переменная, которая назначается исследователем системой относительно внедряемой подсистемы. Определенная переменная разрывается (повторяется в векторе), и между ее повторенными значениями вставляется вектор, описывающий подсистему. В качестве примера можно привести вектор, описывающий пласт-коллектор в окружении вмещающих пород, используемый для определения пористости. Коллектор является подсистемой системы, представленной чередованием слоев: литотипы подошвы ($L_{под}$), пласта коллектора ($L_{кол}$) и покрышки ($L_{кр}$). Характеристики коллектора (подподсистема) представлены его геофизическими характеристиками – (методы ГИС). В этом случае вектор переменных будет иметь следующий вид:

$$[L_{под} \Rightarrow L_{кол} \Rightarrow \text{Методы ГИС (1...N)} \Rightarrow L_{кол} \Rightarrow L_{кр}] \Rightarrow \{K_{on}\}.$$

Реальный граф представляет собой целевую комбинацию установленных типов переменных, а оптимальный выбор его строения отражает структуру объекта исследований, что во многом определяет эффективность решения системных задач.

Разработанная на описанных выше принципах программа векторного марковского прогнозирования-моделирования (VecProgn) представляет собой систему, состоящую из блоков: управления (интерфейс), подготовки данных (очистки и нормализации), контроля массива данных, формализации переменных и параметров, моделирования, распознавания, исследования и оценки точности, визуализации переменных и результатов прогноза (рис. 1).

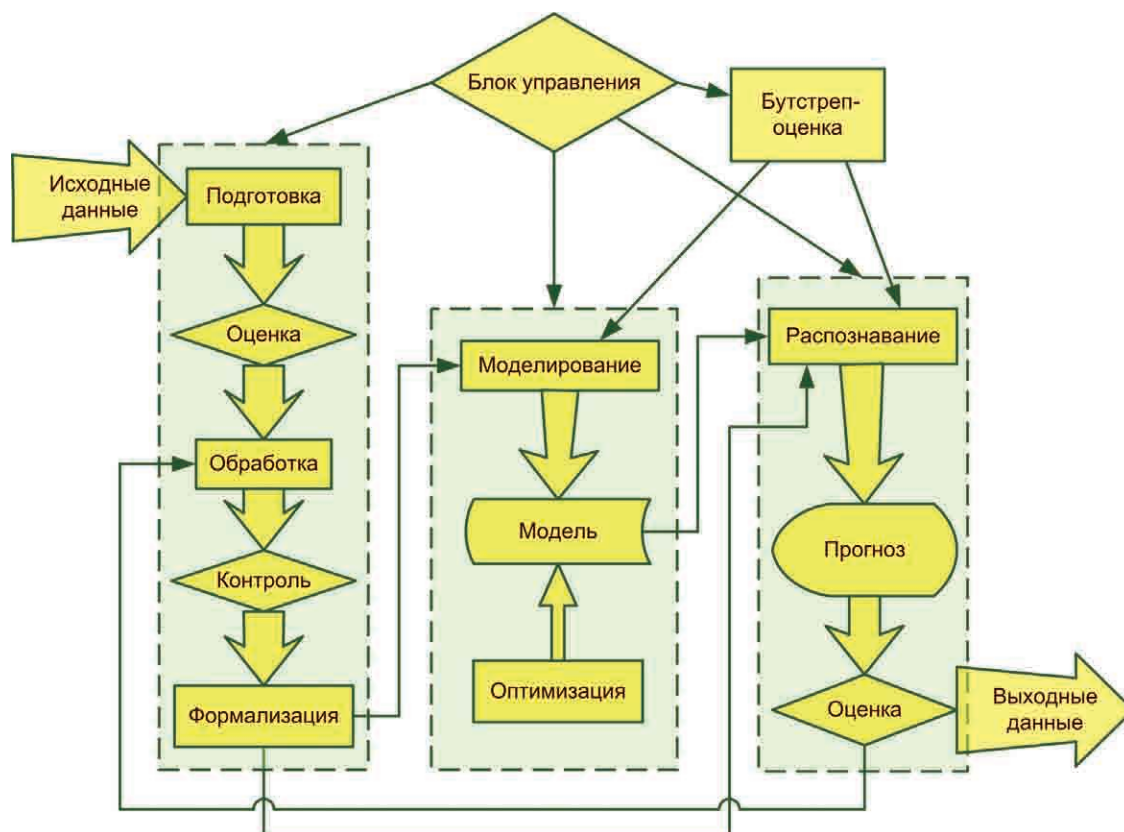


Рис. 1. Блок-схема формирования и обработки пакетных петрофизических моделей

Программа VecProgn написана в среде программирования Delphi, с использованием языка программирования Object Pascal.

Помимо стандартных блоков система содержит многофункциональный блок бутстреп-оценки модели, основное назначение которого – выполнять независимую оценку соответствия модели реальной исследуемой системе без привлечения дополнительных данных помимо используемых в обучающей выборке.

Бутстреп-метод отличается от традиционных методов тем, что предполагает многократную обработку различных частей одних и тех же данных, как бы их поворот (разными гранями), и сопоставление полученных таким образом результатов [5]. Суть метода заключается в том, что часть данных при обработке извлекается из исходной выборки. Оставшаяся часть используется для построения модели системы, а по исключенной из выборки группе производится прогноз и оценка точности модели, после чего она снова возвращается в исходную выборку. Затем следующая группа исключается из обучающей выборки, повторяется построение модели и ее оценка на исключенных данных. Так повторяется до тех пор, пока все данные не будут использованы для статистической оценки точности, которая определяется как матожидание среднего и среднеквадратического значений совокупности частных, независимых оценок. Такая оценка называется бутстреп-оценкой ожидаемой оценки подгонки (ООП). Повышение надежности оценки точности модели достигается многократным повторением ООП. Этот подход позволяет также оперативно осуществлять оптимизацию модели путем последовательного подбора состава переменных, структуры графов, уровней дискретизации переменных и параметров, а также других установочных данных. За оптимальную принимается модель с установочными параметрами, обеспечившими минимальные значения ООП.

Другой областью приложения бутстреп-метода является моделирование-прогнозирование в отсутствии или ненадежности опорных данных. В этом случае из совокупности переменных выбирается ведущая переменная, то есть согласно имеющимся экспертным оценкам, априори связанная с прогнозируемым параметром. Эта переменная разбивается на ряд интервалов (способами, предусмотренными в меню программы) и этим интервалам присваиваются экспертные значения: числа – для порядковых или классы – для номинальных шкал измерения. Трансформированная таким образом независимая переменная принимается в качестве оценки зависимого моделируемого параметра. Затем сгенерированный параметр сопоставляется с вектором предикторных (независимых) переменных, и с ними выполняется итерационный процесс бутстреп-подгонки. Суть его состоит в следующем. Так же, как описано выше, с последовательным исключением части данных из обучающей выборки ведется бутстреп-прогноз параметра. Прогноз сравнивается с исходными значениями и оценивается расхождение между ними. В следующую итерацию включается предыдущий прогноз в качестве «учителя», и каждый последующий прогноз обучается на предыдущем. При этом на каждой итерации происходит перераспределение значений параметра в многомерном фазовом пространстве. Итерационный процесс завершается, когда: 1) различие между предыдущим и последующим прогнозами станет меньше наперед заданной величины; 2) количество выполненных итераций достигнет заданного числа.

В результате проведенной подгонки оценки прогнозируемого параметра принимают значения, наиболее вероятные относительно предикторных переменных. На рис. 2 приводится рабочая таблица, содержащая пакетные данные предикторных переменных (геофизические и структурные показатели) и результаты прогнозирования – C_{2l} (объемная глинистость), K_{on} (открытая пористость) и ряд других параметров. Также представлены фрагменты кривых пакета ГИС (переменных) и результатов прогноза C_{2l} и K_{on} .

На приведенных (см. рис. 2) пакетных диаграммах на шкалу абсцисс вынесены значения глубин по скважинам и номера скважин.

По прогнозным значениям ФЕС коллекторов и покрышек с помощью программы марковского нечеткого 3D моделирования [3] рассчитываются модели-образы изменения глинистости, открытой пористости и проницаемости. Затем по профилям, проведенным через координаты заложения скважин, рассчитываются и визуализируются разрезы пласта-коллектора. На рис. 3 представлен пример построения модели проницаемости интервала рабочего горизонта ПХГ, пересекающего разрезы скважин по сложной траектории.

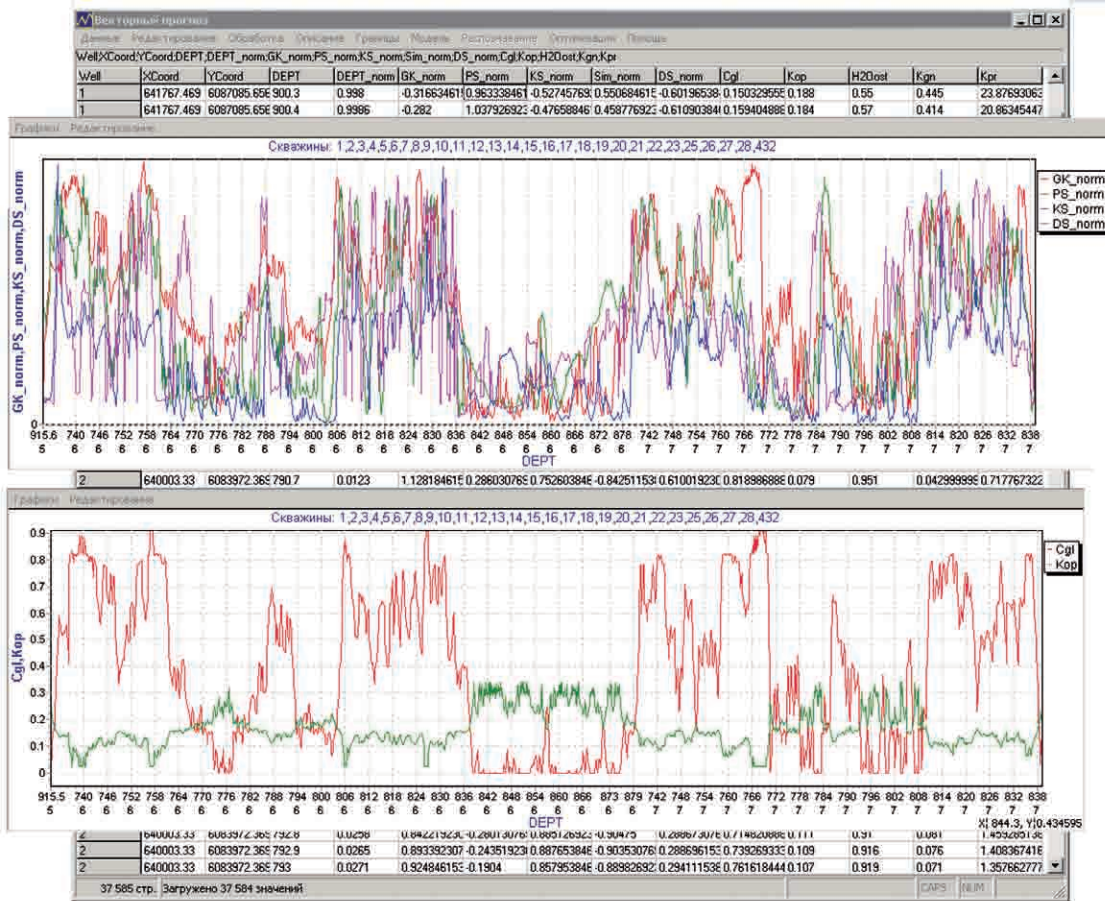


Рис. 2. Пример прогноза $C_{гп}$, $K_{оп}$ в программе VecProgn

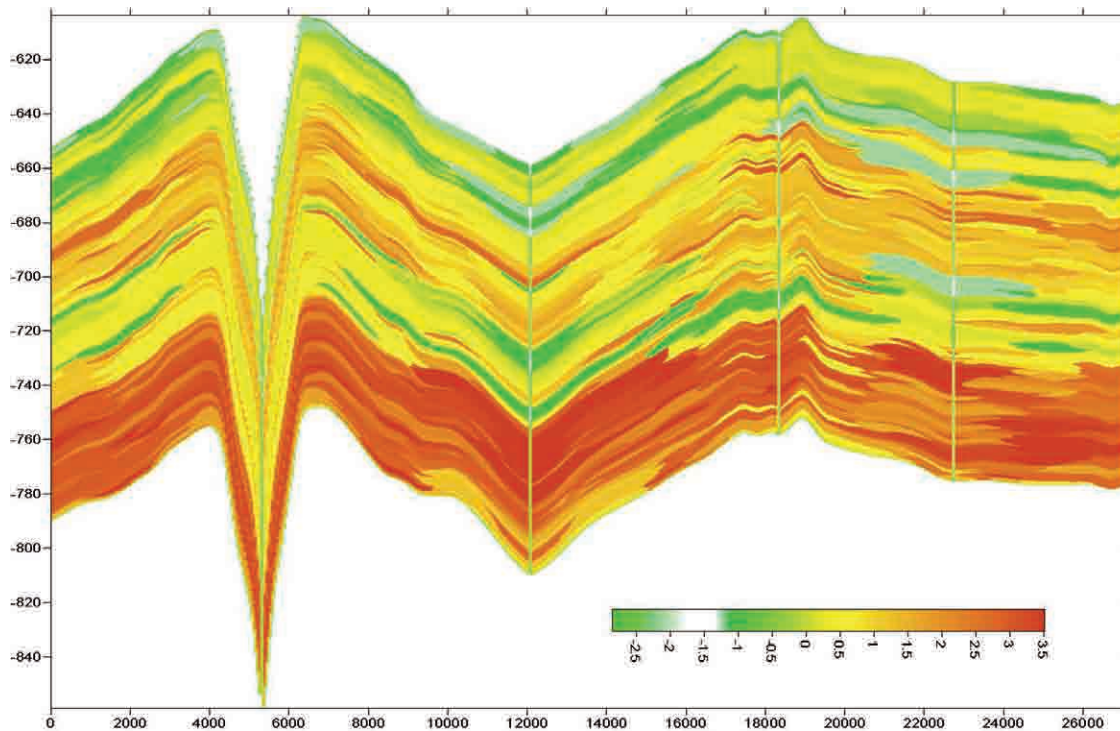


Рис. 3. Строение пласта-коллектора по заданной линии профиля, полученное по методике 3D моделирования на основе математического аппарата нелинейной марковской статистики. Градации цвета соответствуют изменению значений логарифма проницаемости, мД

Таким образом, использование системы марковского векторного прогнозирования и нетрадиционных статистических процедур обработки данных позволяет осуществлять пакетную обработку больших числовых массивов, оптимизировать параметры исследуемой системной модели и при этом экономно использовать данные, добываемые в процессе исследований. Описанная система обработки геолого-геофизических данных в настоящее время успешно используется при создании постоянно действующих геолого-технологических моделей подземных газовых хранилищ.

Список литературы

1. *Вистелиус А.Б.* Основы математической геологии / А.Б. Вистелиус. – Л.: Наука, 1980. – 389 с.
2. *Гриб Н.Н.* Методологические основы системного исследования массива горных пород / Н.Н. Гриб, А.В. Самохин, А.Г. Черников. – Якутск: Изд-во ЯНЦ СО РАН, 2000. – 104 с.
3. *Исхаков А.Я.* Моделирование изменчивости свойств породного массива на основе нечетких марковских последовательностей / А.Я. Исхаков, М.Б. Матушкин, Р.Г. Темиргалеев, А.Г. Черников // Подземное хранение газа. Полвека в России: опыт и перспективы: сб. науч. тр. – М.: ВНИИГАЗ, 2008. – С. 255–265.
4. *Черников А.Г.* Состояние и возможности геофизики при решении задач количественной оценки параметров метаноносности угольных пластов / А.Г. Черников // Газовая промышленность. – 2010. – № 7. – С. 22–25.
5. *Эфрон Э.Б.* Нетрадиционные методы многомерного статистического анализа: пер. с англ. / Э.Б. Эфрон. – М.: Финансы и статистика, 1988. – 263 с.