

В. В. Лапковский¹, В. Е. Куш¹*

Использование разнородных данных при моделировании свойств осадочных толщ методом сплайн-регрессии

¹ Институт нефтегазовой геологии и геофизики им. А. А. Трофимука СО РАН,
г. Новосибирск, Российская Федерация
* e-mail: LapkovskiiVV@ipgg.sbras.ru

Аннотация. Методы многомерной регрессии являются одними из основных при создании пространственных моделей распределения свойств в осадочных толщах. При этом зачастую приходится использовать разные группы данных, которые имеют разные погрешности и характеризуются разным объемом. Типичный пример – моделирование распределения пористости, для чего используются как лабораторные измерения по керну, так и оценки, полученные путем интерпретации каротажных данных, а также результаты сейсмической инверсии. Погрешности разных групп данных обычно можно оценить. И возникает вопрос – с какими относительными весами следует учитывать эти данные для создания наиболее достоверных моделей. Мы исследовали этот вопрос путем генерации «измерений» с разными погрешностями от заданных аналитических моделей с последующей оценкой точности аппроксимаций, получаемых с разными относительными весами данных. В качестве метода аппроксимации, позволяющего работать с многомерными данными больших объемов (big data), использован аппарат сплайн-регрессии. Выявленные зависимости относительных погрешностей данных и оптимальных весов в случае сплайн-регрессии оказались сходными для разных аналитических моделей. Представлен пример моделирования пористости методом сплайн-аппроксимации с разными весами исходной информации для одного из газонефтяных месторождений Западной Сибири. Дальнейшее улучшение полученной модели возможно после учета материалов акустической инверсии сейсмического куба для изучаемого объекта.

Ключевые слова: сплайн-регрессия, регрессия с весами, пористость, 3D модель, погрешности аппроксимации

V. V. Lapkovskiy¹, V. E. Kushch¹*

Using Heterogeneous Data in Modeling Properties of Sedimentary Complexes by Spline Regression Method

¹ Trofimuk Institute of Petroleum Geology and Geophysics SB RAS,
Novosibirsk, Russian Federation
* e-mail: LapkovskiiVV@ipgg.sbras.ru

Abstract. Multivariate regression methods are one of the main tools used in creating spatial models of property distribution in sedimentary deposits. In this case, it is often necessary to use different groups of data that have different errors and are characterized by different volumes. A typical example is modeling porosity distribution, which requires both laboratory measurements on core samples and estimates obtained through interpretation of well log data, as well as results of seismic inversion. The errors of different data groups can usually be estimated, leading to the question of how to weigh these data relative to each other to create the most accurate models. We investigated this question by generating "measurements" with different errors from given analytical models and then evaluating the accuracy of approximations obtained with different relative weights of the data. We used spline re-

gression as the approximation method, which allows for working with large volumes of multidimensional data (big data). The dependencies between relative data errors and optimal weights in spline regression were found to be similar for different analytical models. An example of porosity modeling using spline approximation with different weights of source information for one of the oil and gas fields in Western Siberia is presented. Further improvement of the obtained model is possible after taking into account the materials of acoustic inversion of the seismic cube for the studied object.

Keywords: spline regression, weighted regression, porosity, 3D model, approximation errors

Введение

Одними из основных в геологии и геофизике являются модели, в которых свойства осадочных толщ в пространстве оцениваются путем аппроксимации данных наблюдений (измерений). В статье рассмотрен вопрос о построении регрессионных моделей для закрытых территорий на основе измерений, имеющих разную точность. Характерным примером данной задачи может служить построение модели пространственного распределения пористости в осадочных толщах. Хотя, полученные нами результаты, справедливы и для моделирования других свойств.

Существуют различные способы оценки пористости пород. Самый надежный, но и самый дорогой – лабораторное определение при исследовании керна скважин. Пористость, оцененная по керну, имеет минимальную погрешность, но как правило таких данных недостаточно для создания обоснованных пространственных моделей. Вторая группа данных – оценка пористости по зависимостям керн-ГИС (Геофизические Исследования Скважин). Здесь применяются различные алгоритмы, от достаточно простых, основанных на нахождении линейных регрессионных зависимостей, до процедур, связанных с искусственным интеллектом. Каротажные данные, обычно, существенно более представительные, чем лабораторные исследования керна, они имеются для всех скважин, в которых выполнен достаточный комплекс геофизических исследований и, как правило, характеризуют весь интервал продуктивных отложений. Но, вместе с тем, прогнозные оценки по каротажу менее надежны по сравнению с результатами анализа керна. Иногда погрешности оценки пористости по ГИС столь велики, что находятся за пределами разумных и естественных границ. Наконец, пористость целевых горизонтов может быть спрогнозирована путем инверсии волновых полей, и как представляется, эти оценки еще менее достоверны.

Задачу, которая рассматривается в данной статье, можно сформулировать следующим образом – если у нас есть разнородные оценки параметра, которые обладают разной погрешностью, и мы хотим использовать совокупность этих оценок при создании пространственной модели на основе методов многомерной регрессии то, как это сделать правильно?

Методы и материалы

Регрессионная модель предполагает, что некоторая переменная представляется как линейная комбинация базисных функций. Коэффициенты в уравнении регрессии определяются методом наименьших квадратов, в котором минимизируется выражение

$$\sum_{i=1}^N (\delta(x) - z_i)^2, \quad (1)$$

где z_i – наблюдаемые данные, а $\delta(x)$ – уравнение регрессии.

Обычно полагается, что все измерения равнозначны и они имеют одинаковый вес. Для случая, когда измерения учитываются с разными весами, система линейных уравнений в методе наименьших квадратов может быть изменена. Тогда вместо (1) минимизируется выражение

$$\sum_{i=1}^N w_i (\delta(x) - z_i)^2, \quad (2)$$

здесь w_i – вес, который приписан i -тому наблюдению.

Для получения моделей пористости нами были использованы алгоритмы сплайн-регрессии [1, 2]. Особенность этого решения состоит в том, что количество опорных узлов сплайна многократно меньше по сравнению с числом точек, в которых известны аппроксимируемые значения. В общем случае опорные узлы могут даже не являться подмножеством точек с исходными данными. Аналитически выражение сплайна представляется как [1]:

$$\delta(s) = \sum_{i=1}^N \lambda_i G(s, t_i) + \mu_0 + \sum_{k=1}^n \mu_k s_k, \quad (3)$$

где $G(s, t) = g(|s - t|)$ – радиальная функция, $s = (s_1, \dots, s_n)$, $t_i = (t_{i,1}, \dots, t_{i,n})$, $i = 1, \dots, N$ опорные узлы сплайна в n мерном пространстве R_n , $\mu_0 + \sum_{k=1}^n \mu_k s_k$ линейная компонента сплайна, λ_i – коэффициенты при радиальных функциях. Коэффициенты λ_i в случае сплайн-регрессии находятся методом наименьших квадратов. Сплайн-регрессия отличается от обычной многомерной регрессии тем, что в ней используются радиальные функции, связанные с набором опорных узлов, в то время как в обычной многомерной регрессии могут использоваться любые функции.

Вид сплайна зависит от выбора типа радиальных функций. Мы использовали хорошо зарекомендовавшие себя в различных задачах псевдокубические сплайны [1].

Еще раз уточним исследуемую проблему. Строится модель параметра (пористости), который задан разными группами измерений (керна, ГИС и др.). При этом имеется возможность оценить погрешности каждой группы данных. Наша задача - **задать веса для каждой группы измерений, чтобы получить наиболее качественную модель (с минимальными погрешностями)**. Соответствующего теоретического решения найти не удалось. Поэтому был выполнен ряд вычислительных экспериментов с аналитическими моделями. Мы использовали одномерные функции для генерации наборов данных, но как представляется, по-

лученные выводы не должны отличаться и при работе с моделями большей размерности. Вычислительные эксперименты были выполнены по следующей схеме:

- задавалась аналитическая модель (одномерная функция),
- на её основе генерировались две группы данных; немногочисленная группа с малой погрешностью, и многочисленная группа с большей погрешностью измерения,
- при разном соотношении весов этих групп решалась задача сплайн-регрессии и, определялось среднеквадратическое отклонение полученного приближения от исходной аналитической модели.

Типичная ситуация показана на рисунке 1.

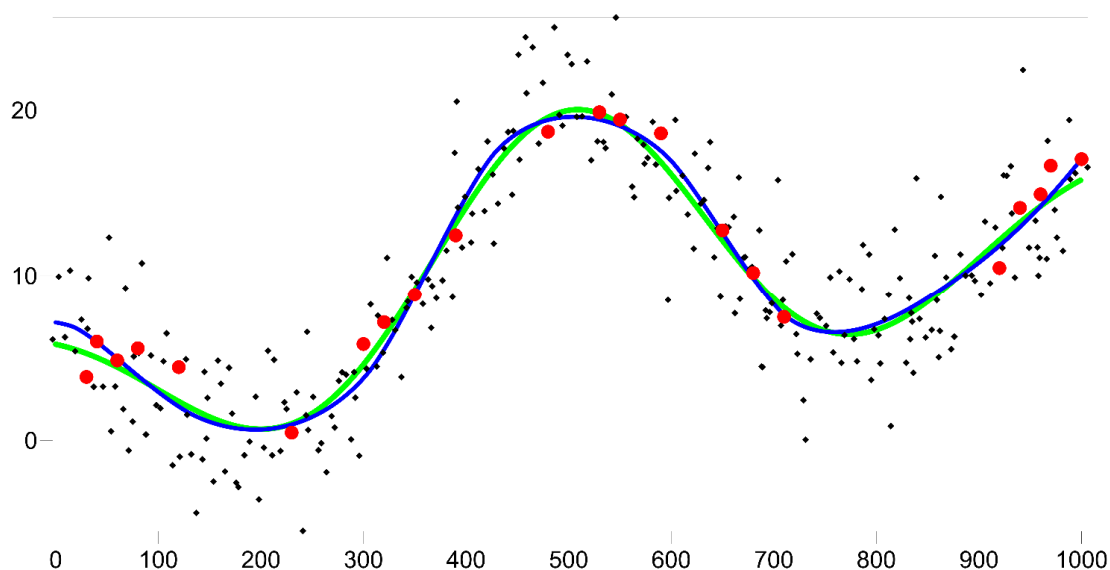


Рис. 1. Теоретическая модель приближаемой величины, её представление группами данных и полученное приближение сплайн-регрессией. Зеленый график – модель приближаемой величины, синий график – результат сплайн-регрессии, красные точки – группа исходных данных с наименьшей погрешностью, черные точки – группа данных с большей погрешностью.

На рисунке 2 показаны вариограммы для самой модели и разных групп данных. Эти вариограммы позволяют оценить погрешность тех или иных данных. Дисперсия соответствующих наборов данных равна «эффекту самородка» вариограммы (значение вариограммы при нулевом смещении). В приведенном примере, группа относительно качественных, но не многочисленных, данных имеет стандартное отклонение – 1, а группа данных с большей погрешностью имеет стандартное отклонение – 3.

На рисунке 3 показаны погрешности аппроксимации в зависимости от отношений стандартных ошибок групп данных (ось x) и отношения весов, с которыми аппроксимируются эти данные (ось y). Выделяется изогнутая область низких значений погрешности аппроксимации, которая для каждого отношения погреш-

ности в группах данных имеет один минимум. Положение этого минимума позволяет оценить оптимальный вес для набора менее точных данных, который обеспечивает наилучшую аппроксимацию.

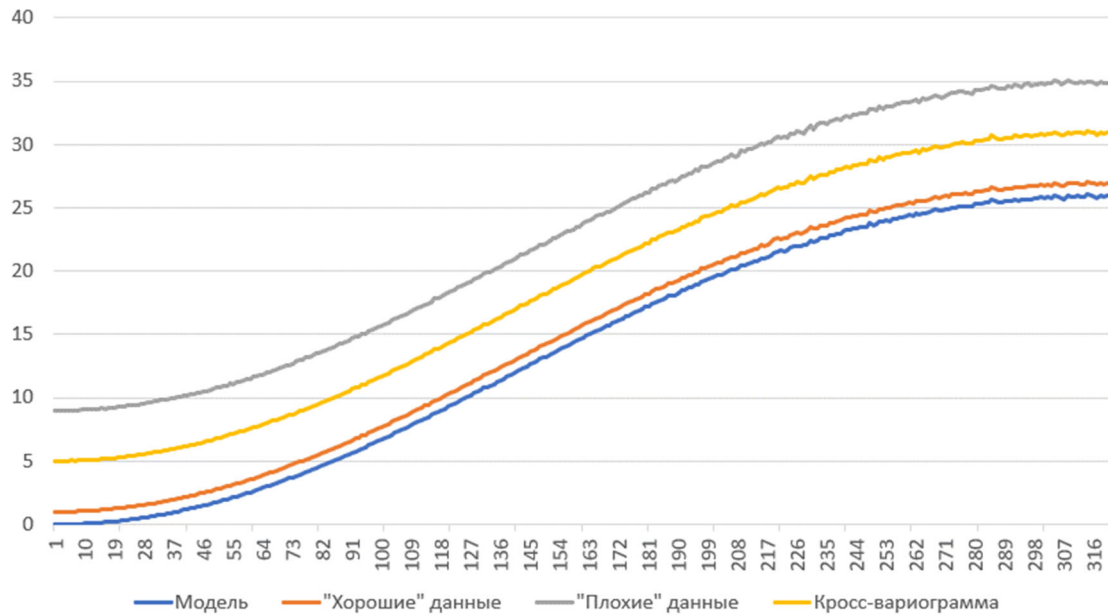


Рис. 2. Вариограммы модели и разных групп данных

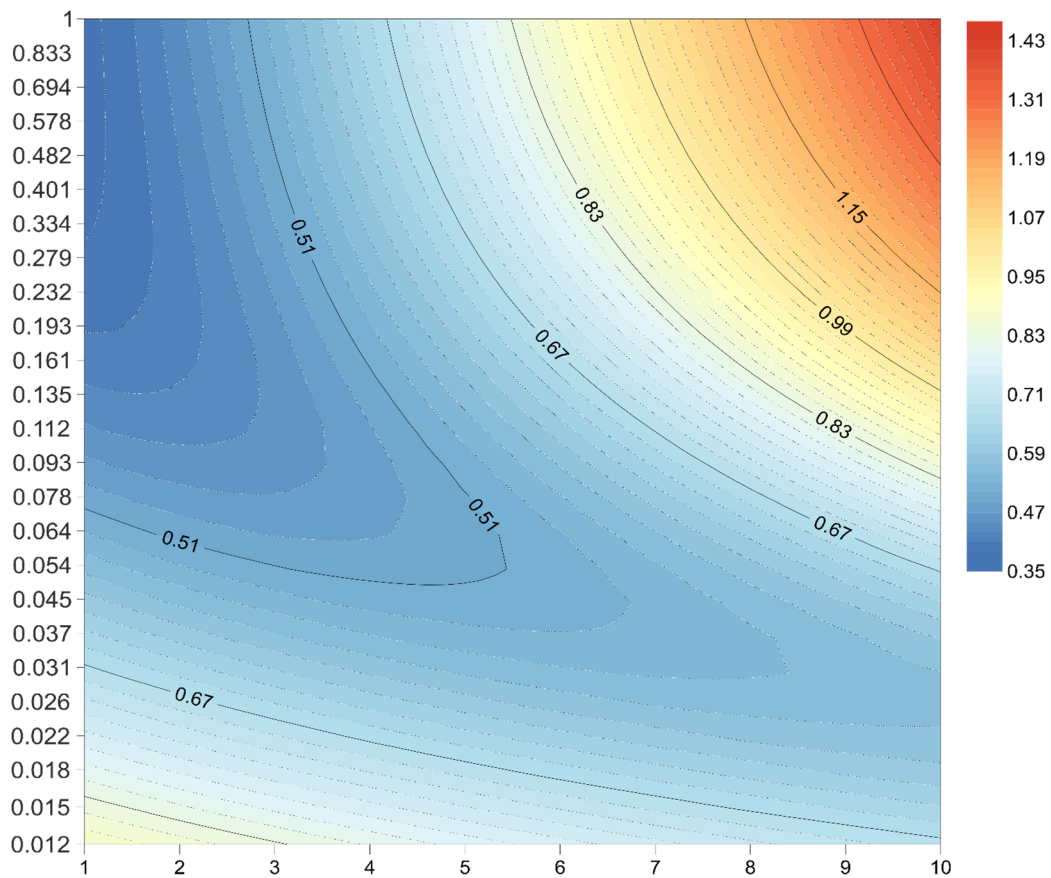


Рис. 3. Карта погрешностей аппроксимации в зависимости от отношений стандартных ошибок данных (ось x) и весов «вспомогательных данных» (ось y)

Кроме приведенной на рисунке 1 аналитической модели мы выполнили вычислительные эксперименты с другими аналитическими функциями. Принципиально, картина не меняется в зависимости от соотношения объемов «точных» и «грубых» данных, а также от отношений погрешностей в группах.

Результаты моделирования пространственного распределения пористости для продуктивных отложений васюганской свиты

Полученные зависимости были использованы для построения модели пористости продуктивных отложений одного из месторождений Западной Сибири. В качестве исходных данных использовались лабораторные измерения пористости по керну, а также прогнозные значения по ГИС в скважинах. У вариограммы пористости по керну значение «эффекта самородка» равно 3.71. Это означает, что среднеквадратическая погрешность лабораторных определений может быть оценена как 1.3 %. Погрешность оценки пористости по ГИС была определена как 2.6 %. Последняя величина получена путем вычисления дисперсии отклонений прогнозных значений по ГИС и лабораторных измерений (она составила 8.46). Если допустить, что погрешность лабораторных измерений и прогноза по ГИС являются независимыми случайными величинами, то стандартное отклонение погрешности оценок по ГИС вычисляется как $\sqrt{8,46 - 3,71} = 2,6$.

В данном примере, погрешность по ГИС оказалась ровно вдвое больше, чем у лабораторных измерений. Учитывая результаты, изложенные в предыдущем разделе (рисунок 3), оптимальным соотношением весов лабораторных измерений и прогнозных оценок было установлено как 1.0 к 0.15. С этими весами построена трехмерная модель распределения пористости васюганской свиты одного из месторождений Западной Сибири. Поскольку объем исходных данных составляет несколько десятков тысяч точек, то в качестве метода аппроксимации был использован аппарат сплайн-регрессии (3). Два сечения полученной трехмерной модели показаны на рисунке 4. Также здесь изолиниями представлена подошва осадочного чехла и точки, в которых имеются определения пористости. Горизонтальные размеры представленной модели составляют 19 на 17 км. Вертикальный размер модели – 700 метров (масштаб увеличен).

Заключение

Полученные результаты показали, что при построении регрессионных моделей с одновременным использованием данных, имеющих разную погрешность следует устанавливать веса для менее точных данных по закону, напоминающему гиперболу. Для применения этого подхода необходимо получить оценки погрешностей каждой группы используемых данных. Это, может быть сделано на основе построения вариограмм, кросс-вариограмм, либо путем вычисления дисперсий производных из набора данных случайных величин.

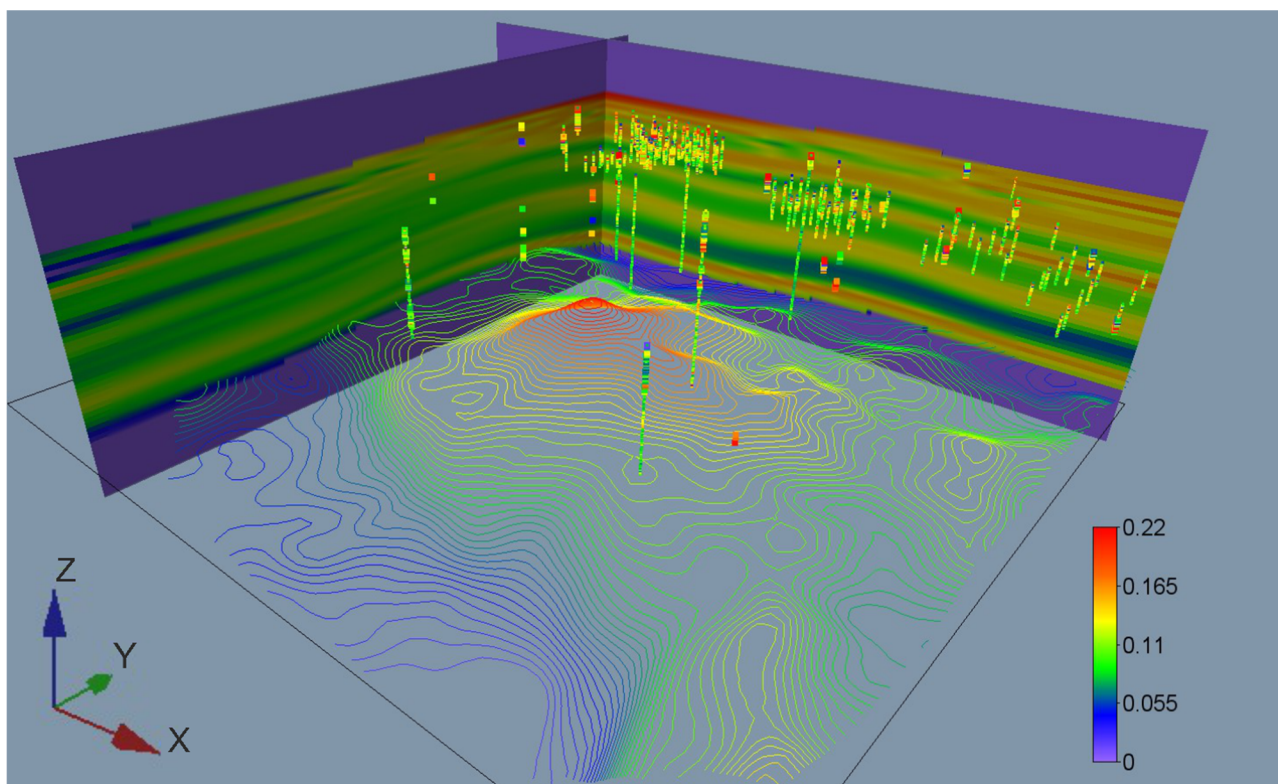


Рис. 4. Визуализация 3D-модели пористости в вертикальных сечениях

Благодарности

Работа выполнена в рамках научной темы FWZZ-2022-0009 «Цифровые геолого-геофизические модели и оценка перспектив нефтегазоносности осадочных бассейнов Арктической зоны Сибири и республики Саха (Якутия); усовершенствование геолого-геофизических методов исследований» Государственной программы ФНИ.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Рожено А.И. Теория и алгоритмы вариационной сплайн-аппроксимации / А.И. Рожено; отв. ред. А.М. Мацокин. – Новосибирск: ИВМиМГ СО РАН, 2005 (ЗАО РИЦ Прайскурьер). – 243 с.
2. Рожено А.И. Сравнение радиальных базисных функций // Сибирский журнал вычислительной математики. – 2018. – Т. 21. – № 3. – С. 273–292.

© В. В. Лапковский, В. Е. Куц, 2023