

УДК 519.2:550.834

Логинов Д.В., Лаврик С.А.ОАО «Центральная Геофизическая Экспедиция», Москва, Россия, loginovdv@gmail.com, dv_lavrik@cge.ru

НЕКОТОРЫЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ ИНФОРМАТИВНОГО НАБОРА СЕЙСМИЧЕСКИХ АТТРИБУТОВ ДЛЯ ПРОГНОЗИРОВАНИЯ СВОЙСТВ КОЛЛЕКТОРОВ

Рассматривается применение статистических и нейросетевых методов для определения информативного набора сейсмических атрибутов с целью эффективного решения задач прогнозирования фильтрационно-емкостных свойств коллектора. Приводятся их описание, методики применения, апробации и сравнительный анализ полученных результатов.

***Ключевые слова:** сравнительный анализ, сейсмический атрибут, метод главных компонент, метод главных факторов, алгоритм оценки значимости, карта прогнозного параметра, многослойная нейронная сеть, доверительный интервал.*

Введение

Проводится сравнительный анализ статистических и специального нейросетевого методов выбора информативных наборов сейсмических атрибутов. Сейсмические атрибуты используются для решения задач количественной оценки петрофизических свойств резервуаров и поиска сейсмофаций на заданных горизонтах. Основой решения таких задач является поиск зависимостей между прогнозируемыми петрофизическими параметрами и сейсмическими атрибутами.

Немалое количество потенциально возможных атрибутов, возможность расчета (усреднения) атрибутов во временных окнах различного размера и с различными смещениями относительно линии сейсмического горизонта, отсутствие четких критериев выбора атрибутов для различных геологических регионов и условий, небольшой опыт использования сейсмического атрибутного анализа, наличие многих потенциальных зависимостей от различных физических и иных факторов для отдельных сейсмических атрибутов, появление нежелательных перекрестных связей между отдельными сейсмическими атрибутами по мере увеличения их количества [Liu, Motoda, 2007 (chapter 2, pp.19-40)] – таков перечень причин, которые порождают большое количество альтернативных атрибутов и их модификаций.

Желание упростить расчеты и получить интерпретируемый результат заставляют применять для решения задач выбора информативных подмножеств атрибутов все более изощренные алгоритмы. Если на первых порах такие алгоритмы основывались на

статистических закономерностях данных (методы нахождения главных компонент, главных факторов, независимых компонент), то в настоящее время, при использовании техники нелинейной регрессии на основе нейронных сетей, могут и должны применяться методы самого аппарата прогнозирования (например, в рамках данной статьи, метод оценки значимости атрибутов на основе обученной нейронной сети).

Сравним результаты этих подходов, в том числе, и на основе реальных данных.

Постановка задачи

Задача выбора информативных сейсмических атрибутов относится к общему классу задач **выбора признаков (Feature Selection)**. Выбор существенных признаков – одна из важнейших и часто используемых технологий в предварительной обработке данных [Kalkomey, 1997; Piramuthu, 1999]. С помощью оптимального выбора из первоначального набора признаков можно удалить несущественные и избыточные данные. В общем случае можно сказать, что **выбор признаков** – это процесс, с помощью которого из множества исходных признаков отбирается наиболее информативная (конечно, в контексте решаемой задачи), существенная их часть.

Выбор подмножества существенных сейсмических атрибутов позволяет снизить вычислительные (и, соответственно, временные) затраты на расчет самих атрибутов, а также время расчета окончательного прогноза и, возможно, может способствовать нахождению новых зависимостей сейсмических атрибутов от петрофизических параметров, которые не учитывались ранее.

Здесь и далее под информативным набором сейсмических атрибутов будем понимать минимальный набор, который определяет многомерные зависимости от петрофизических параметров для эффективного решения задач прогнозирования свойств коллекторов с заданной точностью. Точность прогноза зависит как от качества сейсмического материала, данных ГИС, так и от геологических особенностей конкретных площадей и регионов.

Краткое описание используемых методов и алгоритмов

Для решения задачи выбора существенных признаков применяются различные методы и алгоритмы. Проведём сравнение результатов применения статистических методов, а именно метода главных компонент (Principal Component Analysis, PCA), одного из методов факторного анализа – метода главных факторов (Principal Factor Analysis, PFA) с алгоритмом оценки значимости на основе многослойной нейронной сети (МНС) для решения задачи выбора сейсмических атрибутов. Эти методы выбирались с целью сравнения линейных и нелинейных методов анализа и выбора сейсмических атрибутов.

Идея **метода главных компонент** состоит в преобразовании множества объясняющих переменных в новое множество попарно некоррелированных переменных, среди которых первая соответствует максимально возможной дисперсии, а вторая – максимально возможной дисперсии в подпространстве, ортогональном первому, и т.д. Другими словами, происходит переход от исходного набора признаков к небольшому числу вспомогательных переменных (которые либо отбираются из числа исходных, либо строятся по определенному правилу по совокупности исходных признаков), по которым впоследствии можно достаточно точно воспроизвести интересующие свойства анализируемого массива данных.

Под **факторным анализом** понимается совокупность методов, которые на основе реально существующих связей признаков (или объектов) позволяют выявлять обобщающие характеристики организационной структуры и механизма развития изучаемых явлений и процессов.

Дисперсия элементарных признаков в методах факторного анализа объясняется не в полном объеме, то есть признается, что часть дисперсии остается нераспознанной. Факторы обычно выделяются последовательно: первый, объясняющий наибольшую долю вариации элементарных признаков, затем второй, объясняющий меньшую, вторую после первого латентного фактора часть дисперсии, третий и т.д. Процесс выделения факторов может быть прерван на любом шаге, если принято решение о достаточности доли объясненной дисперсии элементарных признаков или с учетом интерпретируемости латентных факторов [Сошникова и др., 1999 (глава 7, стр. 333-400)].

Идея **алгоритма оценки значимости** [Дубровин, Субботин, 2002] на основе многослойной нейронной сети заключается в том, что если признак является избыточным для классификации набора примеров, то фиксация значения соответствующего входа сети не будет существенно ухудшать значение целевой функции обученной нейронной сети по сравнению с реальными значениями данного признака для того же набора примеров. Эта гипотеза может быть определена по-другому: чем более сильную значимость имеет признак, тем больше будет негативное воздействие на целевую функцию обученной нейронной сети, когда вход для этого признака фиксируется к среднему значению. Под *значимостью* здесь понимается коэффициент, который отражает степень реакции нейронной сети на фиксированный признак. Чем больше коэффициент значимости, тем лучше представлен диапазон данных этим признаком.

Методика применения рассматриваемых методов

Приведем общую методику применения представленных методов и алгоритмов для определения наиболее информативного набора сейсмических атрибутов (рис. 1) [Лаврик, Логинов, 2008].

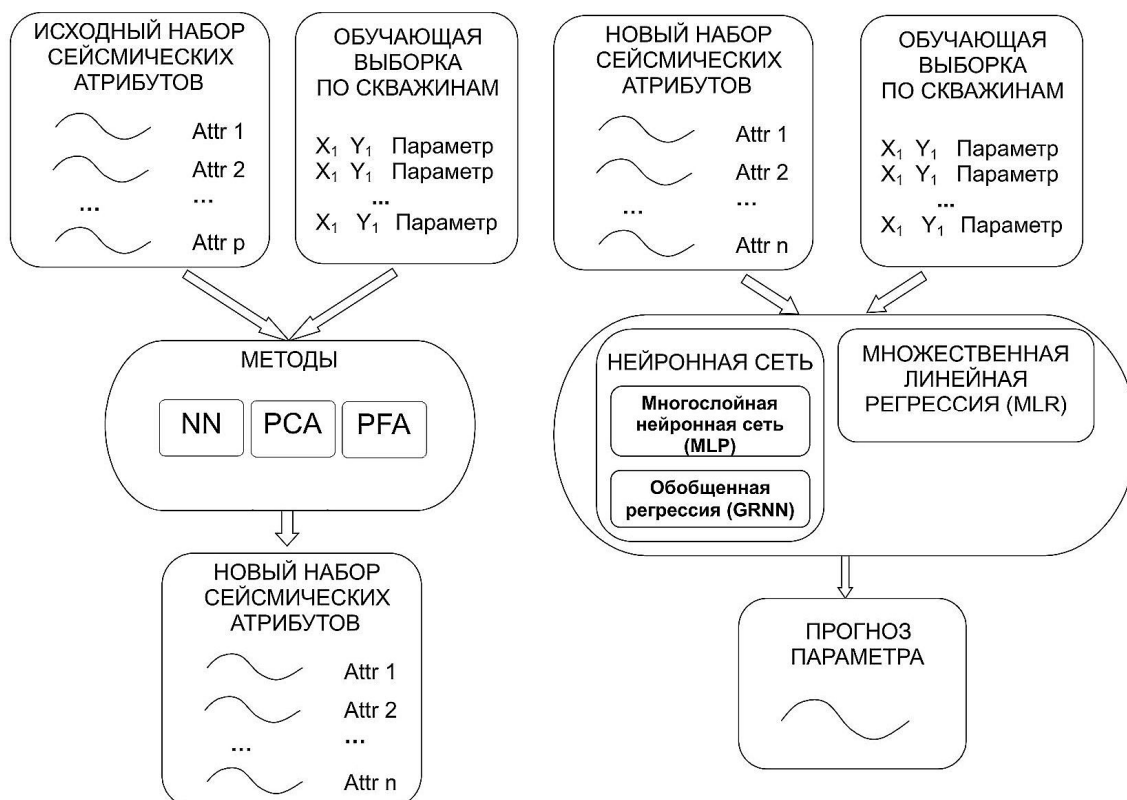


Рис. 1. Методика выбора информативных сейсмических атрибутов

NN – алгоритм оценки значимости на основе МНС, PCA – метод главных компонент, PFA – метод главных факторов.

1. Входными данными являются набор сейсмических атрибутов и обучающая выборка по скважинам (данные ГИС по скважинам).

2. Для этого набора сейсмических атрибутов применяется один из реализованных алгоритмов выбора информативных сейсмических атрибутов. В результате происходит уменьшение размерности исходного пространства сейсмических атрибутов и формируется новый набор сейсмических атрибутов.

3. Для нового набора сейсмических атрибутов применяются методы нейросетевого моделирования и множественной линейной регрессии для расчета карты прогнозного параметра и корреляционные зависимости прогнозного параметра от эффективной мощности, а также различные статистические данные.

4. Проводится сравнительный анализ предлагаемых методов. Сравниваются полученные результаты с данными, полученными по всему набору сейсмических атрибутов.

Сравнительный анализ проводится на основании следующих критериев:

- корреляционная зависимость прогнозного параметра от прогнозируемого петрофизического параметра;
- визуальная оценка полученных карт прогнозного параметра;
- оценка доверительного интервала.

5. Делаются выводы об эффективности применения реализованных алгоритмов определения информативного набора сейсмических атрибутов.

Результаты применения используемых методов

В качестве примера рассматривались 12 сейсмических атрибутов, рассчитанных в интервале продуктивного пласта (Plast1) на одной из площадей и использовались для расчета прогнозного петрофизического параметра – эффективной мощности. На рис. 2 представлена карта прогнозного параметра, вычисленная по всему набору сейсмических атрибутов с использованием многослойной нейронной сети (рис. 2А), и корреляционная зависимость прогнозного параметра от эффективной мощности (рис. 2Б). Все результаты, полученные с применением рассматриваемых методов, помимо сравнения между собой, также сопоставлялись с этими данными. Коэффициент корреляции между $N_{\text{эф}}$ и $N_{\text{эф}}^*$ по всем атрибутам составил 0,54.

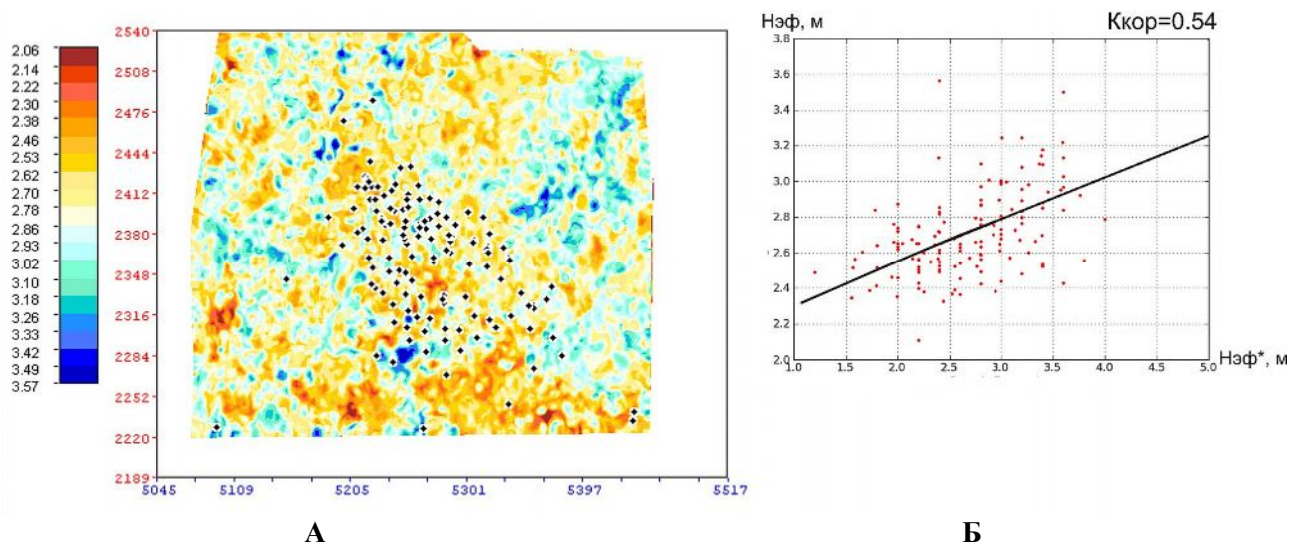


Рис. 2. Карта эффективной мощности $N_{\text{эф}}^*$, построенная с применением многослойной нейронной сети (MLP) с использованием всего набора сейсмических атрибутов (А). Кросс-плот $N_{\text{эф}}$ против $N_{\text{эф}}^*$ (Б)

Алгоритм оценки значимости

В результате применения этого алгоритма рассчитывались оценки значимости каждого атрибута. В табл. 1 приведены четыре сейсмических атрибута, отобранные в качестве информативных на основании полученных коэффициентов значимости. Следует отметить,

что коэффициенты значимости остальных сейсмических атрибутов оказались в пределах от 0.08 до 0.21.

Таблица 1

Оценка значимости сейсмических атрибутов

Сейсмический атрибут	Значимость
Наиболее положительная кривизна	0.40
Углы падения	0.71
Мгновенная амплитуда по кубу интеграл	0.49
Наиболее отрицательная кривизна	0.41

На рис. 3 представлена карта прогнозного параметра (эффективной мощности – рис. 3А) и корреляционная зависимость прогнозного параметра от эффективной мощности, рассчитанные MLP (рис. 3Б). Эта карта вычислена по отобранному набору сейсмических атрибутов.

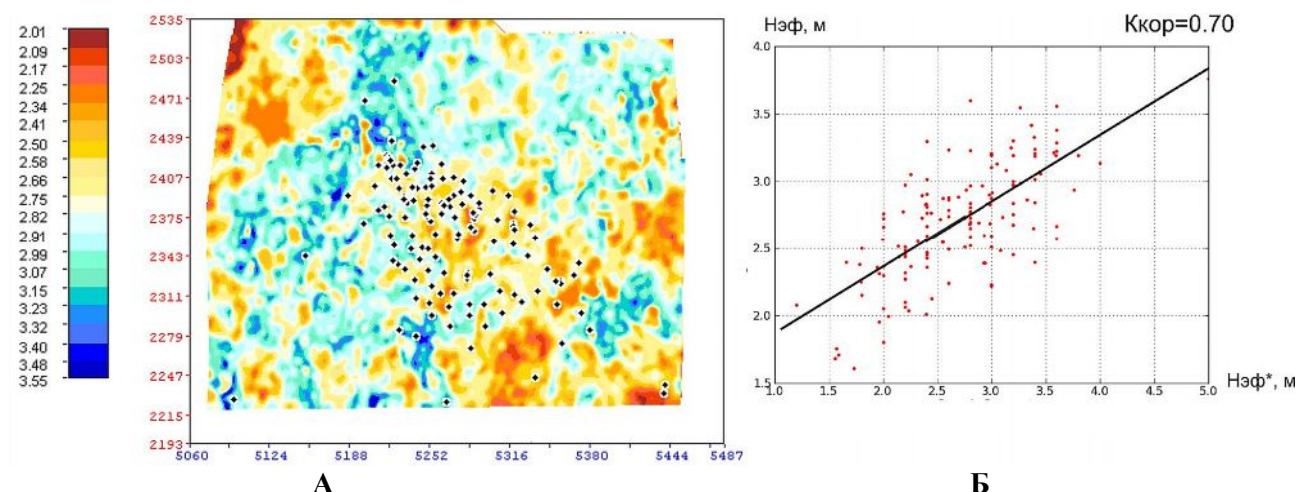


Рис. 3. Карта прогноза эффективной мощности Нэф*, построенная с применением многослойной нейронной сети (MLP) (А). Кросс-плот Нэф против Нэф* (Б)

Набор сейсмических атрибутов отобран алгоритмом оценки значимости.

Если сравнивать полученные результаты с результатами, полученными по всему набору атрибутов, то визуальный анализ прогнозных карт дает основание полагать, что выбор наиболее информативного набора сейсмических атрибутов с применением многослойных нейронных сетей привел к улучшению результата. Это подтверждается коэффициентом корреляции 0.70 и визуальным сравнением карт прогнозного параметра.

Статистические методы

Результаты применения статистических методов представлены в табл. 2, которая отражает процент дисперсии, объясненной каждой компонентой или фактором, и графическое изображение собственных значений (рис. 4).

Таблица 2

Выделенные компоненты и факторы

Компонента (фактор)	PCA		PFA	
	Общая дисперсия, %	Накопленная дисперсия, %	Общая дисперсия, %	Накопленная дисперсия, %
1	45.63	45.63	45.13	45.13
2	29.07	74.70	41.63	86.76
3	11.18	85.88	9.01	95.77
4	4.11	89.99	1.19	96.96
5	3.17	93.16	1.05	98.01
6	2.89	96.05	0.58	98.59
7	1.88	97.93	0.47	99.06
8	1.33	99.26	0.32	99.38
9	0.24	99.50	0.27	99.65
10	0.22	99.72	0.23	99.88
11	0.20	99.92	0.07	99.95
12	0.08	100.00	0.03	99.98

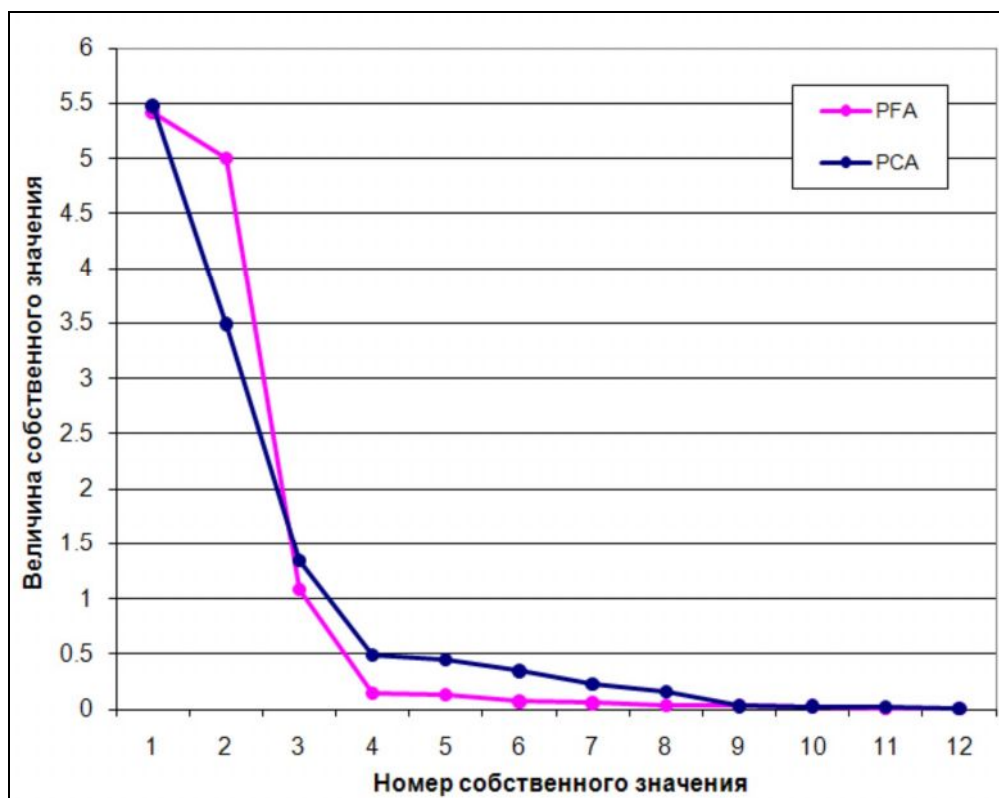


Рис. 4. График собственных значений

PCA – метод главных компонент, PFA – метод главных факторов.

Предварительный анализ этих результатов показывает, что для получения решения без существенного ухудшения размерность исходного пространства сейсмических атрибутов может быть понижена до 4-9 атрибутов.

На практике возникает важный вопрос: сколько главных компонент или факторов следует оставить? Решение его достаточно произвольно, поэтому обычно исследуются

несколько решений с большим или меньшим числом главных компонент (факторов), и затем выбирается одно наиболее «осмысленное».

Метод главных компонент

В результате применения метода главных компонент к исходному пространству сейсмических атрибутов строилось трансформированное пространство из 8 атрибутов. Их количество выбиралось на основании, как указано выше, проведенного исследования при количестве атрибутов от 4 до 9. Решение, полученное с использованием 8 атрибутов, оказалось лучше остальных. Например, для сравнения, коэффициент корреляции при 4-х атрибутах составил 0.46.

На рис. 5 представлена карта прогнозного параметра (эффективной мощности – рис. 5А) и корреляционная зависимость прогнозного параметра от эффективной мощности, рассчитанные по методу MLP (рис. 5Б).

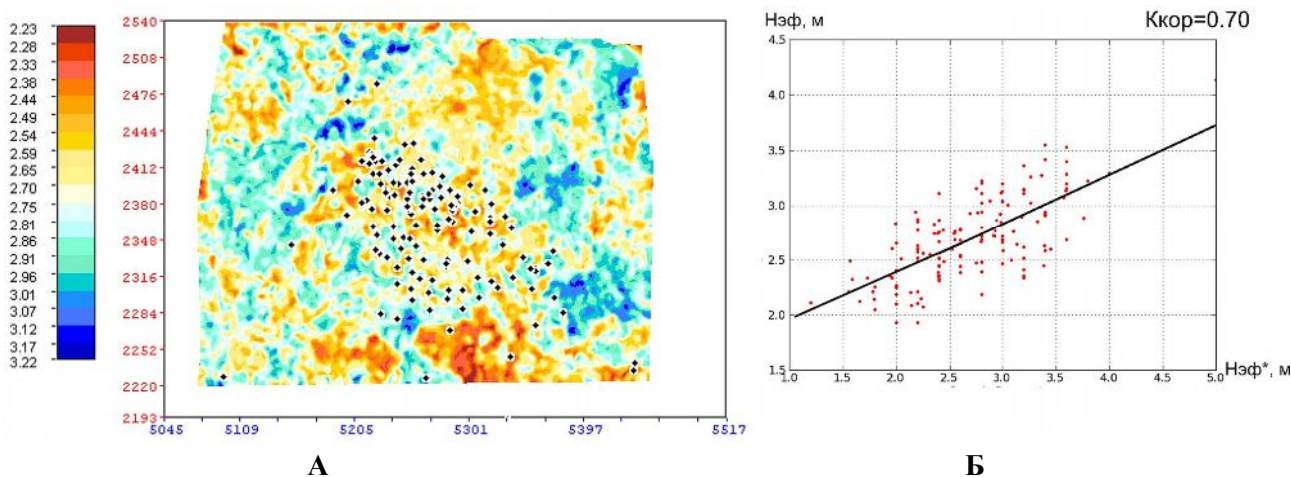


Рис. 5. Карта прогноза эффективной мощности $H_{эф}^*$, построенная с применением многослойной нейронной сети (MLP)(А). Кросс-плот $H_{эф}$ от $H_{эф}^*$ (Б)

Набор сейсмических атрибутов отобран методом главных компонент.

Если сравнивать полученные результаты с результатами, полученными по всему набору атрибутов, то визуальный анализ полученной прогнозной карты и исходной прогнозной карты дает основание полагать, что уменьшение исходного пространства до 8 атрибутов не привело к ухудшению исходного результата. При этом коэффициент корреляции существенно увеличился.

Факторный анализ

В результате применения метода главных факторов к исходному пространству сейсмических атрибутов построилось трансформированное пространство из 8 атрибутов. Аналогично методу главных компонент, проводились исследования при количестве

атрибутов от 4 до 9. В данном случае, наилучшее решение оказалось при использовании 8 атрибутов. В качестве сравнения, коэффициент корреляции при 4-х атрибутах составил 0.42.

На рис. 6 представлена карта прогнозного параметра (эффективной мощности – рис. 6А) и корреляционная зависимость прогнозного параметра от эффективной мощности, рассчитанные MLP (рис. 6Б).

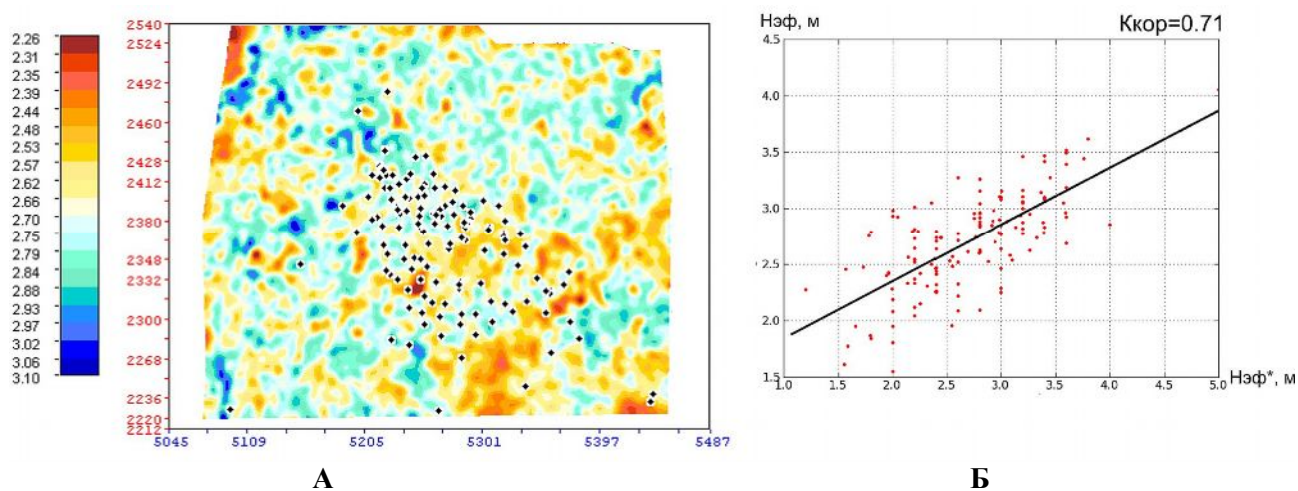


Рис. 6. Карта прогноза эффективной мощности $H_{эф}^*$, построенная с применением многослойной нейронной сети (MLP)(А). Кросс-плот $H_{эф}$ против $H_{эф}^*$ (Б)

Набор сейсмических атрибутов отобран методом главных факторов.

Если сравнивать полученные результаты с результатами, полученными по всему набору атрибутов, то визуальный анализ полученной прогнозной карты и исходной прогнозной карты дает основание полагать, что уменьшение исходного пространства до 8 атрибутов не привело к ухудшению исходного результата. При этом коэффициент корреляции существенно увеличился до значения 0.71.

Оценка доверительного интервала

Оценка доверительного интервала выбирается по таблицам процентных точек распределения Стьюдента при заранее выбранной доверительной вероятности $1-\alpha$. В данном случае доверительная вероятность равна 0.95 (или 95%). Отметим, что доверительный интервал (допустимая ошибка) – это допустимое отклонение наблюдаемых значений от истинных.

Из анализа представленных карт доверительных интервалов (рис. 7) видно, что прогнозы параметра $H_{эф}^*$, на основе наборов атрибутов, отобранных рассматриваемыми методами, оказались лучше, то есть более устойчивыми и достоверными, чем прогноз по всему набору атрибутов. Об этом свидетельствуют интервалы значений на шкалах. В случае исходных данных он значительно больше. А чем меньше длина доверительного интервала, тем точнее оценка.

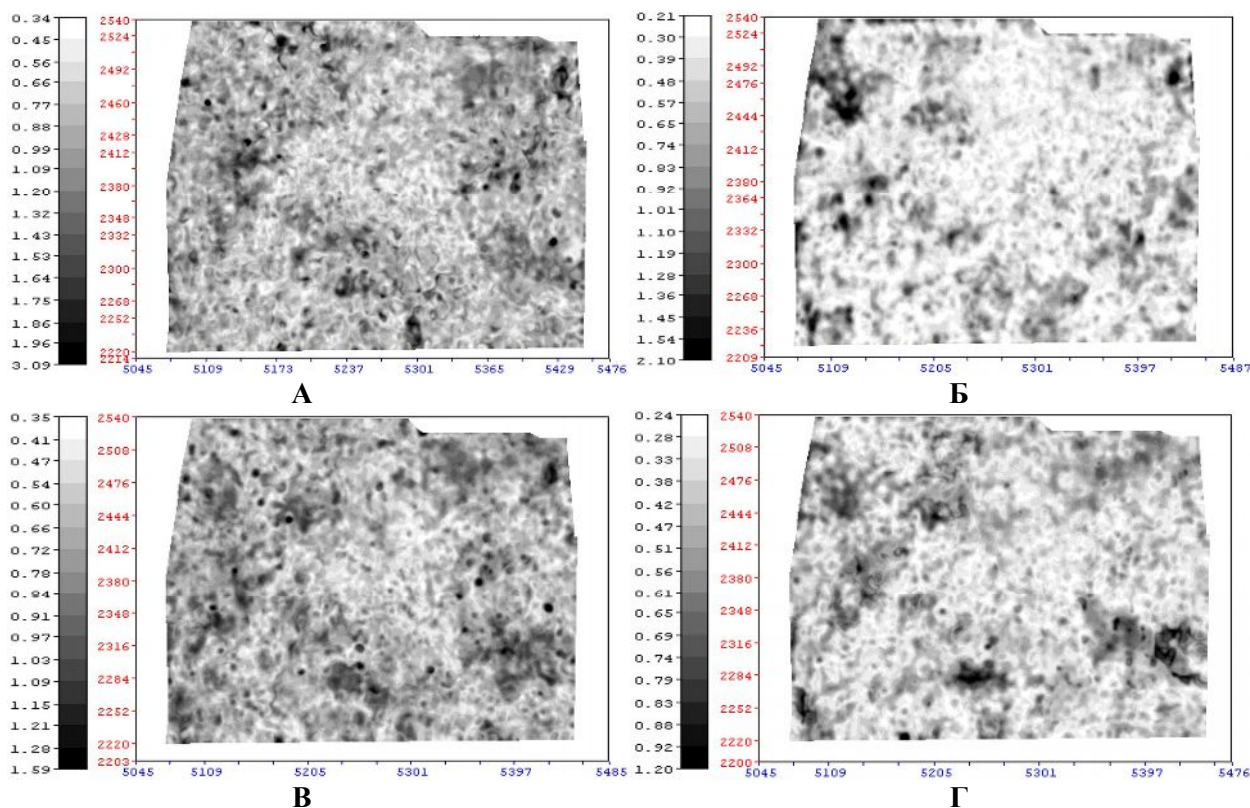


Рис. 7. Карты доверительных интервалов

по всему набору сейсмических атрибутов (А), по наборам сейсмических атрибутов отобранных: нейронной сетью (Б), методом главных компонент (В) и методом главных факторов (Г)

Сравнение результатов используемых методов

Сравнивая результаты, полученные рассмотренными методами, между собой, видно, что они оказались сопоставимы. Коэффициенты корреляции во всех случаях почти одинаковы (0.70, 0.70, 0.71). При сравнении полученных прогнозных карт видно, что на карте, основанной на наборе атрибутов, отобранных нейросетевым методом, более четко выделены области повышенной и пониженной эффективной мощности. Однако, прогнозныe карты на основе наборов атрибутов, отобранных статистическими методами, являются немногим более устойчивыми. Об этом свидетельствуют доверительные интервалы (рис. 7Б, В, Г). В случае применения статистических методов этот интервал меньше.

В приведенном примере заметно существенное различие в количестве используемых атрибутов, выбранных по статистическим и нейросетевому методам. Применение нейросетевого метода обычно позволяет значительно снизить количество существенных атрибутов. Это говорит о том, что этот алгоритм более чувствителен к нелинейности.

Заключение

Подводя итог, можно сделать вывод, что использование рассмотренных методов для выбранной площади позволило существенно снизить количество сейсмических атрибутов. С

помощью метода главных компонент и факторного анализа удалось уменьшить размерность исходного пространства сейсмических атрибутов с 12 до 8, алгоритма оценки значимости по нейросетевой модели – с 12 до 4.

В большинстве случаев, несмотря на схожие результаты, эффективнее использовать алгоритм оценки значимости по нейросетевой модели, поскольку применение этого метода обычно позволяет значительно снизить количество существенных атрибутов, и он более чувствителен к нелинейности.

Оценки выбранного параметра (эффективной мощности $N_{эф*}$), полученные с применением наборов выбранных атрибутов лучше, чем результаты прогноза по всему исходному набору атрибутов. Это подтверждается оценками силы корреляционных зависимостей, визуальным анализом полученных карт прогнозного параметра и оценками доверительных интервалов.

Представленные результаты показывают, что рассмотренные методы могут эффективно применяться для уменьшения количества сейсмических атрибутов без потери существенной информации. Проверка представленной методики на других площадях подтверждает сделанные выводы [Лаврик, Логинов, 2008а; Лаврик, Логинов, 2008б; Лаврик, Логинов, 2009].

Все рассмотренные методы реализованы в программном комплексе SeisProN (ОАО «ЦГЭ») и применяются для решения реальных задач прогнозирования петрофизических параметров.

Литература

Дубровин В.И., Субботин С.А. Оценка значимости признаков на основе многослойных нейронных сетей в задачах диагностики и распознавания //Информатика и системы управления, 2002. - № 1(3). – С. 66-72.

Лаврик С.А. Результаты анализа эффективности и применимости статистических методов для определения информативного набора сейсмических атрибутов //Технологии сейсморазведки, 2009. - №1. – С. 36-44.

Лаврик С.А., Логинов Д.В. Оценка значимости сейсмических атрибутов //Интеллектуальные системы и технологии: сб. науч. трудов в 15 т. – М.: МИФИ, 2008. – Т. 10. - С. 212-214.

Лаврик С.А., Логинов Д.В. Оценка значимости сейсмических атрибутов на основе многослойных нейронных сетей //Геонауки – от новых идей к новым открытиям: труды 3-ей международной конференции и выставки. - СПб, 2008. - 106 с.

Лаврик С.А., Логинов Д.В. Применение статистических методов для определения информативного набора сейсмических атрибутов //Геомодель-2008: труды конференции. - Геленджик, 2008.

Сошникова Л.А., Тамашевич В.Н., Уебе Г., М. Шеффер. Многомерный статистический анализ в экономике: учеб. пособие для вузов /Под ред. проф. В.Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.

Liu H., Motoda H.(Eds). Computational methods of Feature Selection. - Chapman & Hall. - CRC, 2007. - 419 p.

Kalkomey C. T. Potential risks when using seismic attributes as predictors of reservoir properties //The Leading Edge, March, 1997. - v. 16, no. 3. – p. 247-251.

Piramuthu S. The Hausdorff Distance Measure for Feature Selection in Learning Applications //Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, 1999.

Рецензент: Лухминский Б.Е., доктор физико-математических наук.

Loginov D.V., Lavrik S.A.

Central Geophysical Expedition JSC, Moscow, Russia, loginovdv@gmail.com, dv_lavrik@cge.ru

SOME METHODS OF DETERMINING AN INFORMATIVE SET OF SEISMIC ATTRIBUTES FOR FORECASTING RESERVOIR PROPERTIES

The application of statistical and neural network methods for determining an informative set of seismic attributes for efficient solving the problems of forecasting the filtration-capacity properties of reservoir rocks is considered. The description of these methods, the technique of their application and approbation and also a comparative analysis of the received results are given.

Key words: *comparative analysis, seismic attribute, method of basic components, method of basic factors, algorithm of importance estimation, map of forecast parameter, multi-layer neural network, trust interval.*

References

Dubrovin V.I., Subbotin S.A. Ocenka značimosti priznakov na osnove mnogoslojnyh nejronnyh setej v zadačah diagnostiki i raspoznavaniâ //Informatika i sistemy upravleniâ, 2002. - # 1(3). – S. 66-72.

Lavrik S.A. Rezul'taty analiza èffektivnosti i primenimosti statističeskikh metodov dlâ opredeleniâ informativnogo nabora sejsmičeskikh atributov //Tehnologii sejsmorazvedki, 2009. - #1. – S. 36-44.

Lavrik S.A., Loginov D.V. Ocenka značimosti sejsmičeskikh atributov //Intellectual'nye sistemy i tehnologii: sb. nauč. trudov v 15 t. – М.: MIFI, 2008. – Т. 10. - S. 212-214.

Lavrik S.A., Loginov D.V. Ocenka značimosti sejsmičeskikh atributov na osnove mnogoslojnyh nejronnyh setej //Geonauki – ot novyh idej k novym otkrytiâ: trudy 3-ej meždunarodnoj konferencii i vystavki. - SPb, 2008. - 106 s.

Lavrik S.A., Loginov D.V. Primenenie statističeskikh metodov dlâ opredeleniâ informativnogo nabora sejsmičeskikh atributov //Geomodel'-2008: trudy konferencii. - Gelendžik, 2008.

Sošnikova L.A., Tamaševič V.N., Uebe G., M. Šefffer. Mnogomernyj statističeskij analiz v èkonomike: učeb. posobie dlâ vuzov /Pod red. prof. V.N. Tamaševiča. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.