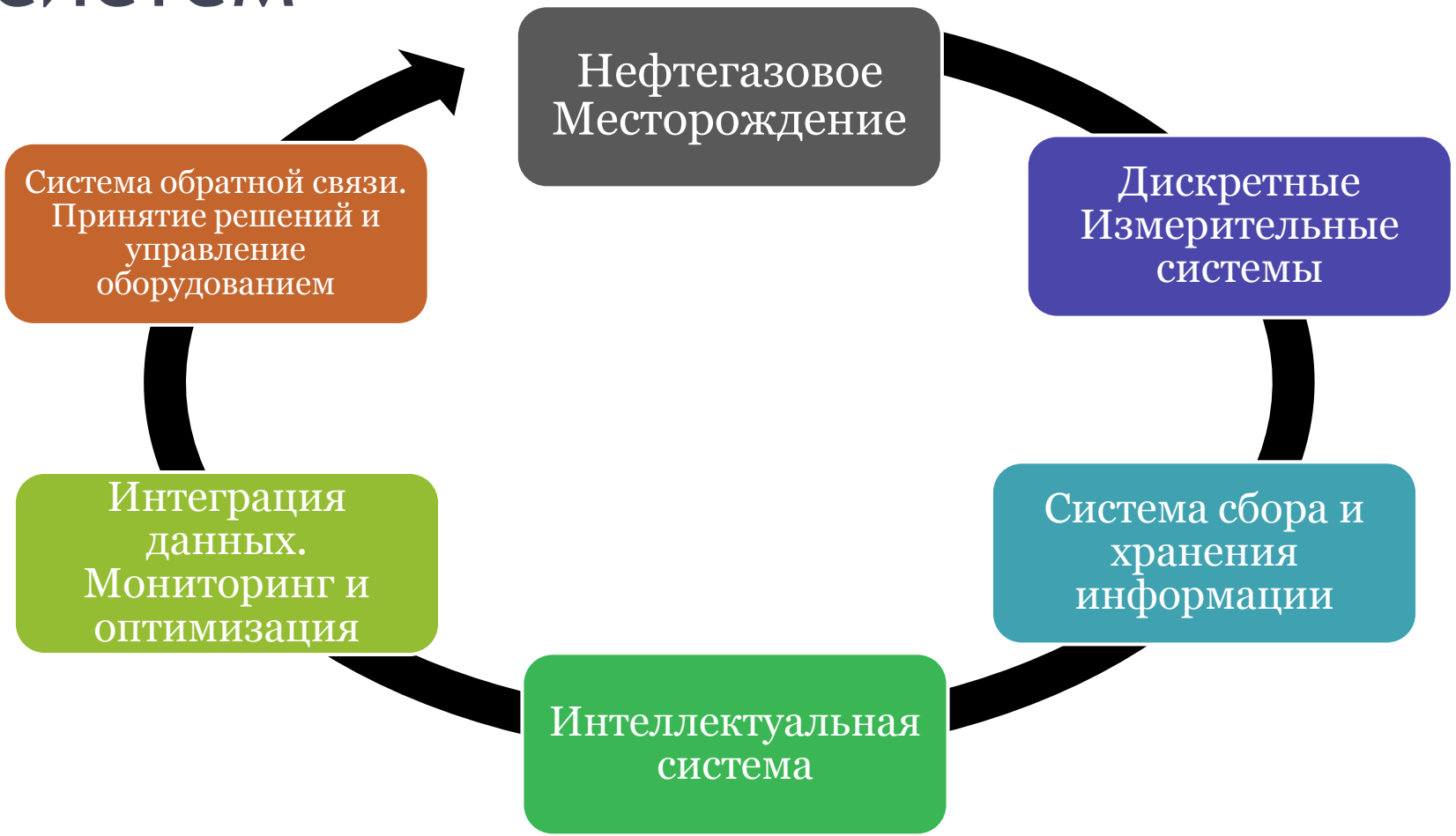
# Интеллектуальная система анализа геолого-промысловых данных

Докладчик: Татарников В.В.
Руководитель: Загоруйко Н. Г.

# Проблемы индустрии

- Падение добычи и рост издержек
- Усложнение географических условий
- Высокая степень неопределённости данных

# Этапы развития информационных систем

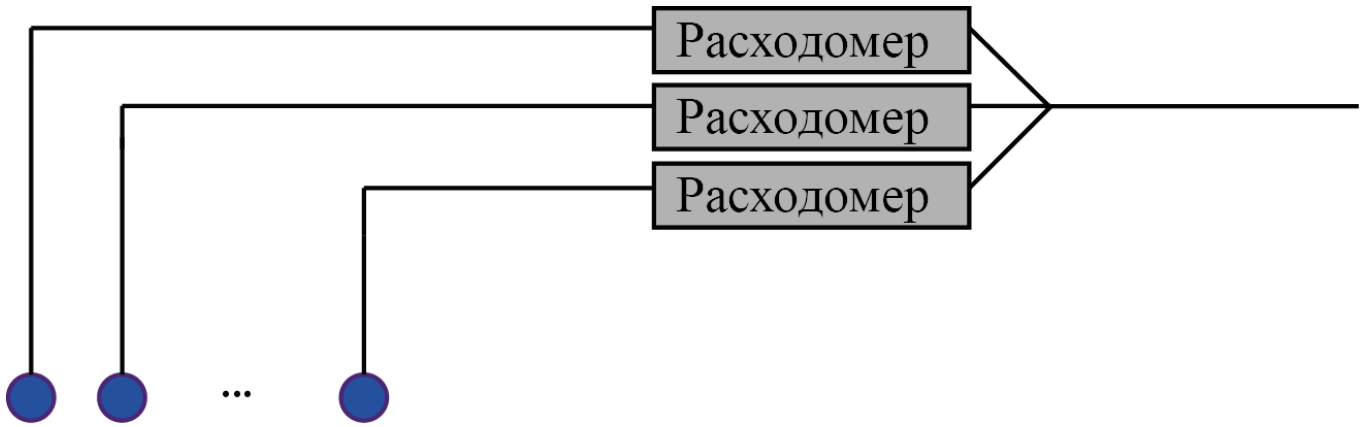


# Цель работы

Разработка интеллектуальной системы, обеспечивающей снижение степени неопределённости геолого-промысловых данных

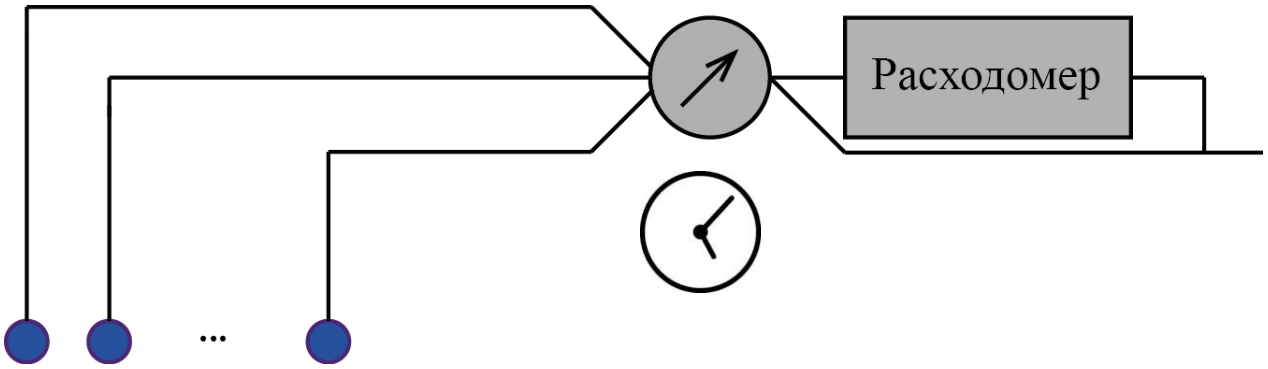
# Неопределённость данных?

- Установить полный комплекс измерительных приборов на каждую скважину



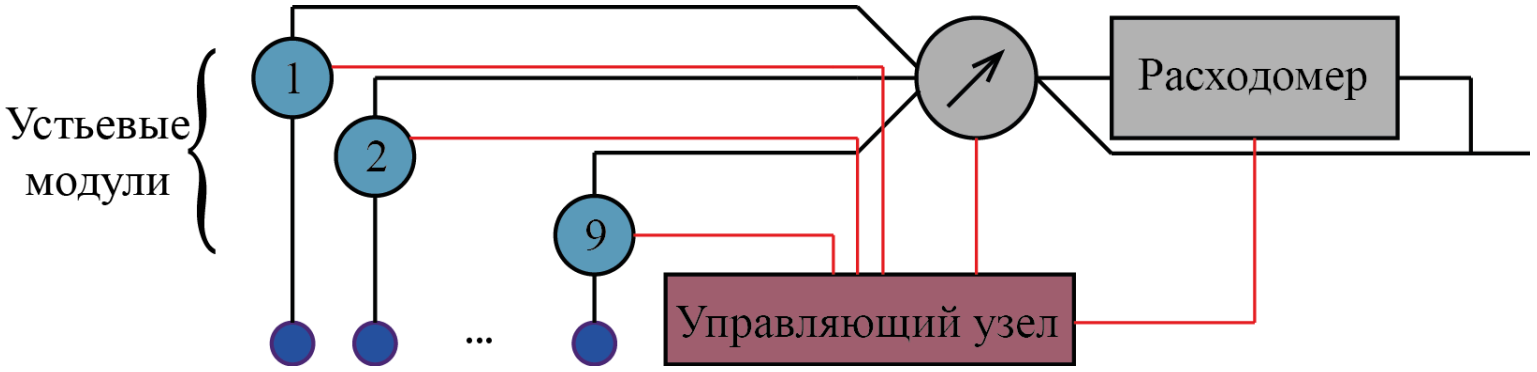
# Неопределённость данных

- Около 70% скважин в РФ низкодебитные
- Используется переключение по расписанию
- Дебит по скважинам расписывается вручную



# Неопределённость данных

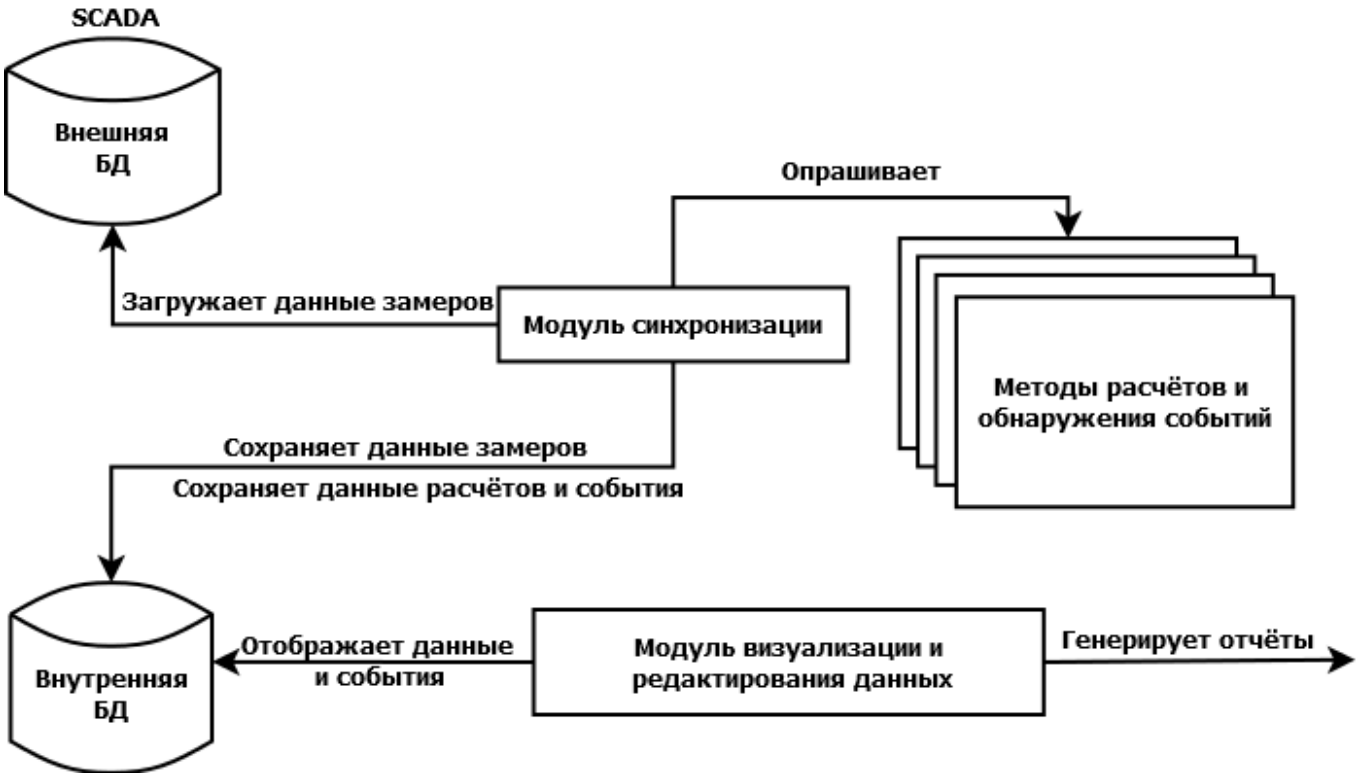
- Переключение по событиям
- Снижает потери информации
- Показания устьевых модулей содержат ошибки и пробелы



# Задачи

- Обеспечить сбор и хранение, редактирование и визуализацию данных, создание отчётов
- Организовать обработку рассчитываемых параметров (выделение событий)
- Обеспечить предобработку данных измерений скважин (заполнение пробелов, исправление ошибок)

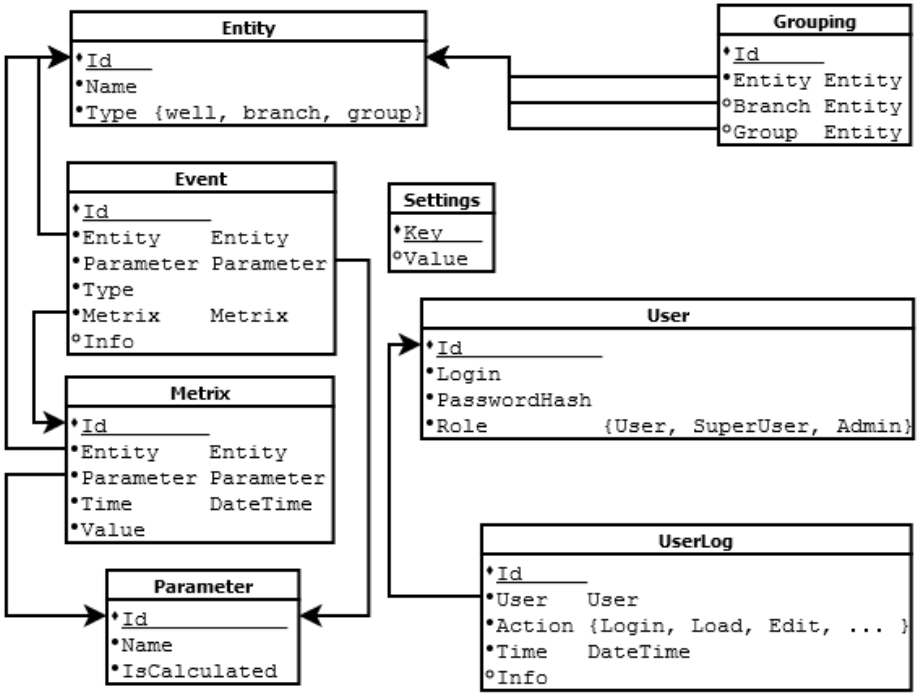
# Модульная организация ИС



# Данные

- Схема:
  - Скважины
  - Параметры
  - Измерения
  - События
  - Пользователи

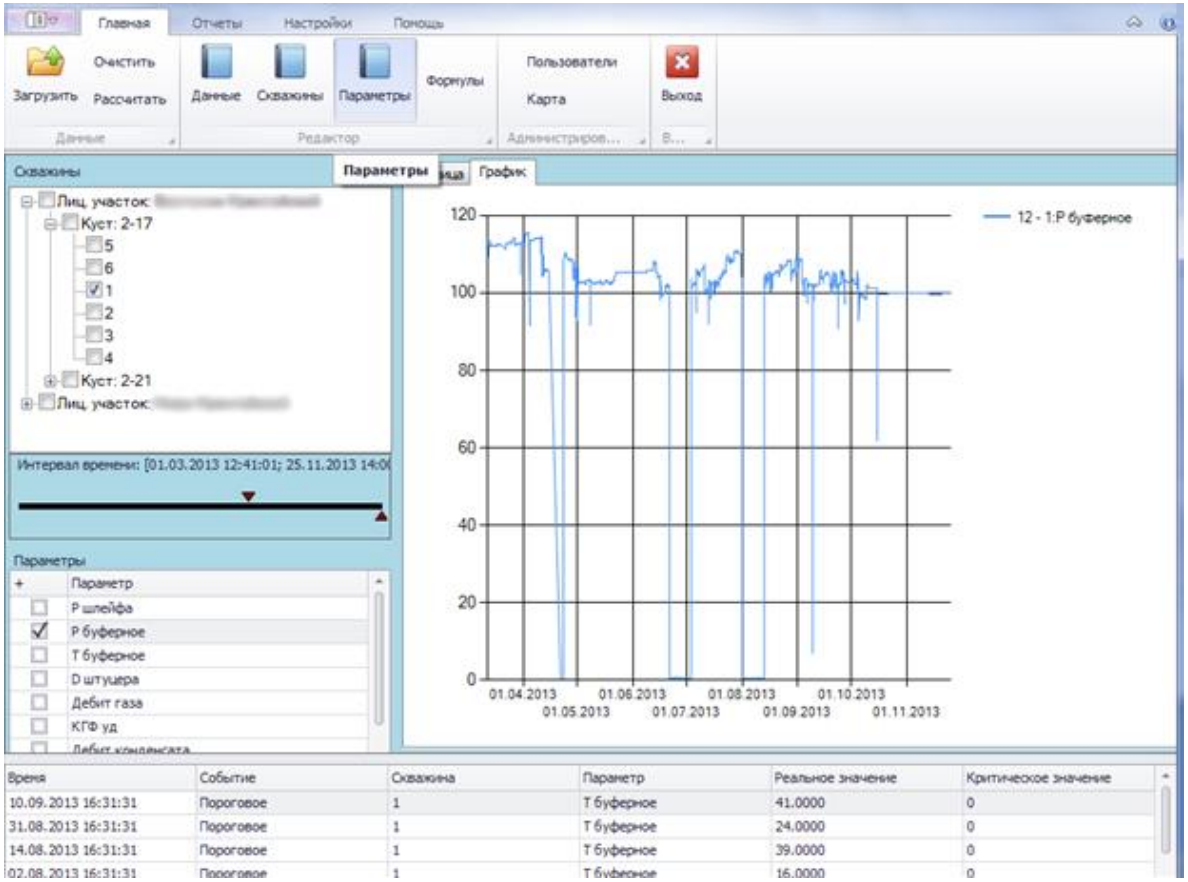
- Импорт данных:
  - Автоматический
  - Ручной



# Редактирование и визуализация

- Работает с внутренней БД
- Ручная загрузка данных
- Ручное редактирование данных, скважин и параметров
- Визуализация данных
- Отображение событий
- Формирование отчётов

# Редактирование и визуализация



# Методы выделения событий

- Реализованы в виде подключаемых модулей

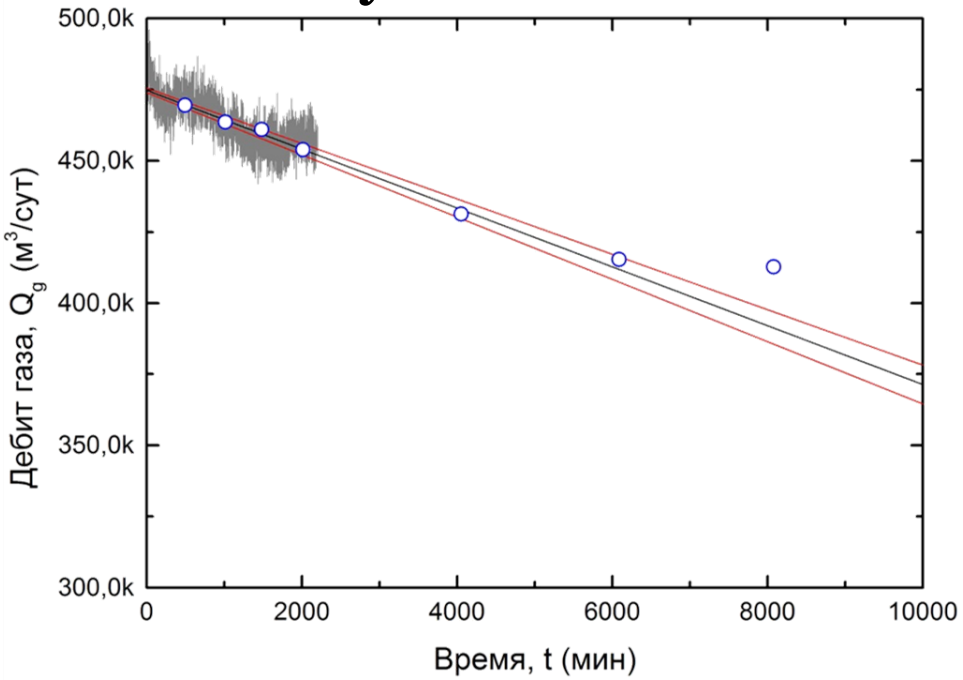
```
public interface EventChecker
{
    string Name { get; }
    string Help { get; }
    string Version { get; }

    …

    Event[] findEvents(DateTime d, DataHelper helper);
}
```

# Методы выделения событий

- Скачок значения выше заданного порога
- Обнаружена ошибка или пробел
- Доверительный конус



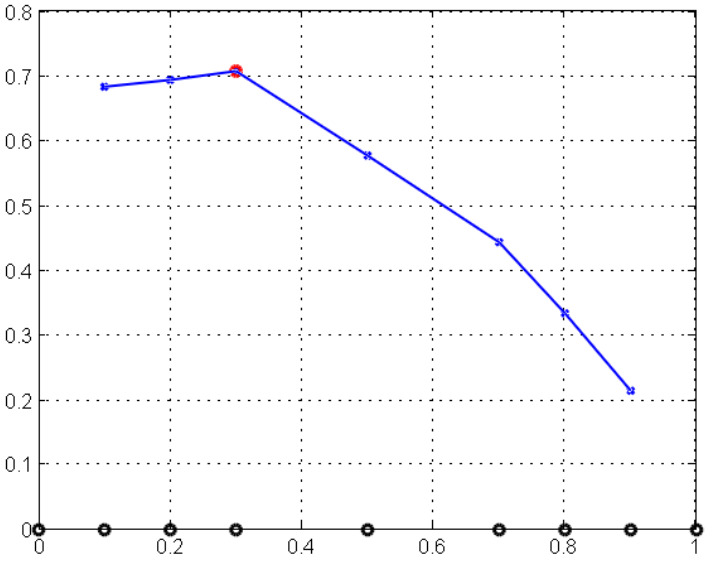
# Заполнение пробелов

- Обзор существующих методов
  - Простые
  - Сложные
- Разработка собственного метода
  - Вариант алгоритма ZET для кубов данных
  - FRiS-функция
  - Количественная оценка компактности подкуба
- Критерий ошибки и реализация модуля для системы

# Заполнение пробелов

1. Алгоритм ищет наиболее компактный подкуб данных

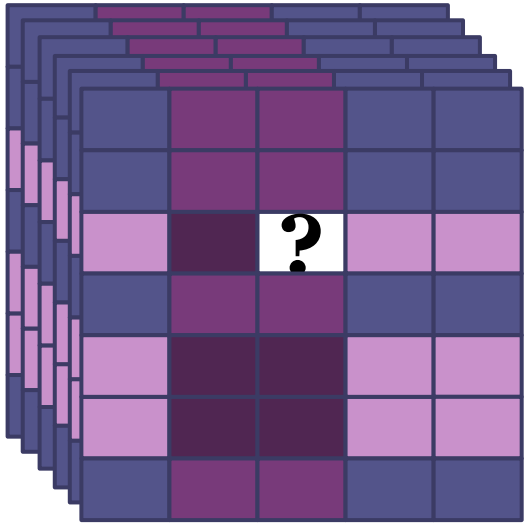
$C$




$\{a_0,\ldots,a_n\}$

# Заполнение пробелов

2. Заполняет пробел на основании модели




- Модели работают лучше в локальной окрестности исследуемого объекта

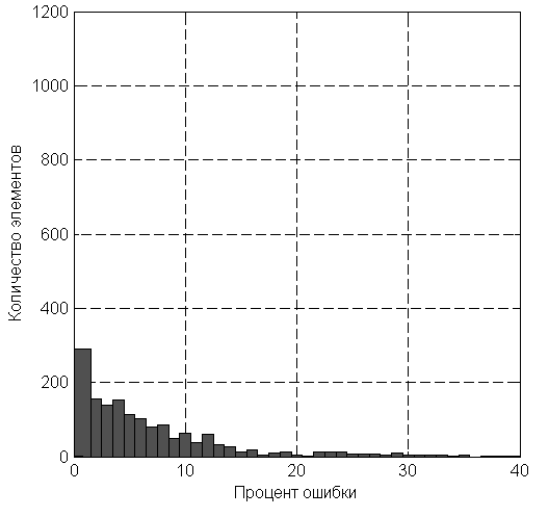
# Заполнение пробелов

- Эксперимент с перекрёстной проверкой при помощи многомерной регрессии
- Отобраны данные по кусту(13) скважин: дебит, давление, температура, …(11 параметров)  за 122 дня
- Каждая клетка дебита полагалась пробелом и заполнялась алгоритмом
- Критерий качества
  - Средняя относительная ошибка

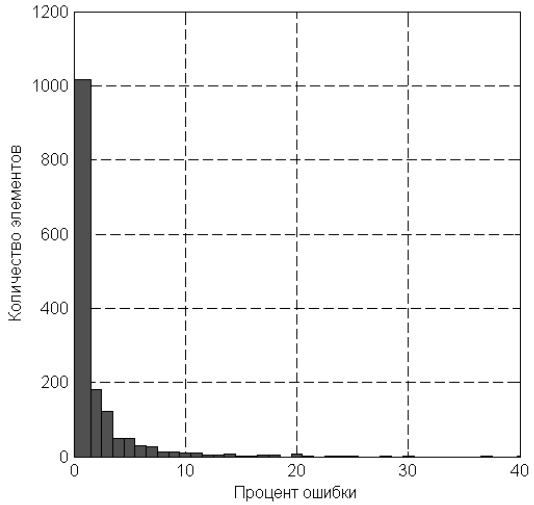
# Заполнение пробелов

- Средняя ошибка:  7%/ 2%
- В интервале [0; 10]: 76% / 95%

Без построения подкуба




С построением подкуба



# Результаты - 1

- Предложен собственный метод для заполнения пробелов и определения ошибок в геолого-промысловых данных.
  - Имеет потенциал для его применения в задачах анализа других данных .

# Результаты - 2

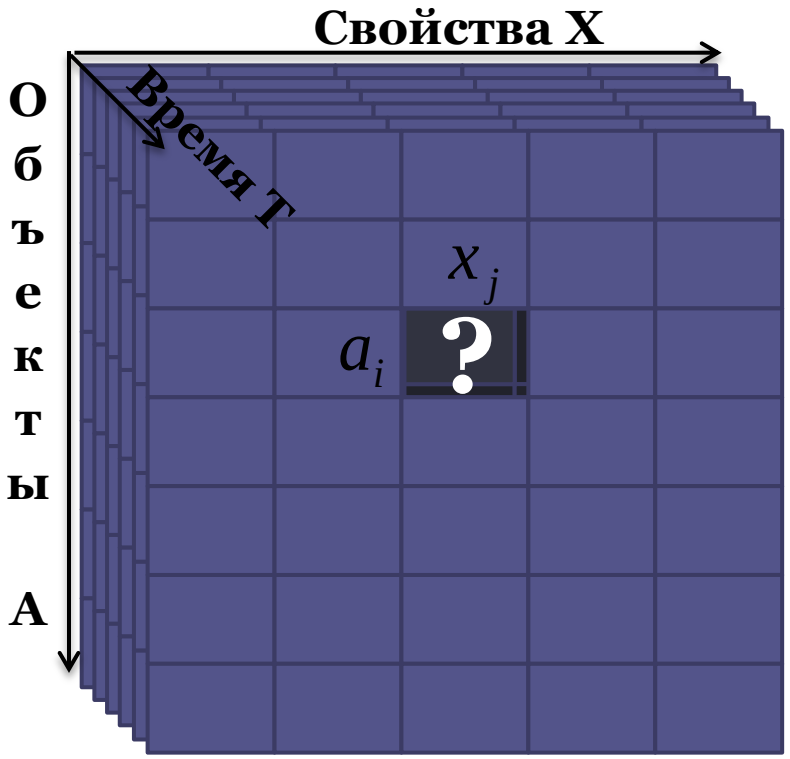
- Разработана интеллектуальная система анализа геолого-промысловых данных с возможностью выделять события для реализации замерной схемы по расписанию.
  - Слабая зависимость компонентов
  - Расширяемость

# Публикации по теме работы

- МНСК-2014 (тезисы)
- Гео-Сибирь-2014 (доклад)
- Сибирский журнал индустриальной математики. Апрель–июнь, 2014 (статья)

Спасибо за внимание!

# Пробелы и ошибки




$\langle A, X, T \rangle$ – *куб данных*

$\langle a_i, x_j, t_k \rangle$ – *элемент куба*

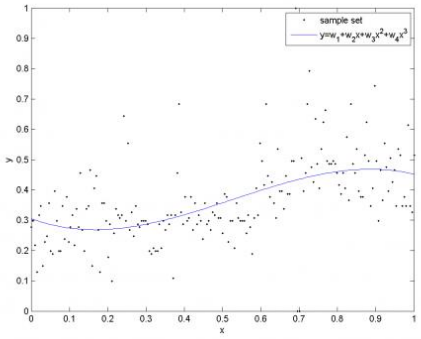
$\langle a_i, X, T \rangle, \langle A, x_j, T \rangle, \langle A, X, t_k \rangle$ – *целевые сечения куба*

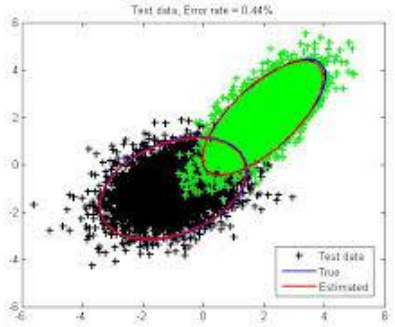
Задача:

- Если элемент не определён, то заполнить.
- Если элемент определён, не содержит ли он ошибку?

# Алгоритмы заполнения

- Простые:
  - Интерполяция
  - Скользящее среднее
- Сложные
  - Регрессия
  - EM-алгоритм




- Действуют глобально: зависимость должна быть реализована на всех объектах выборки

# Алгоритм ZET

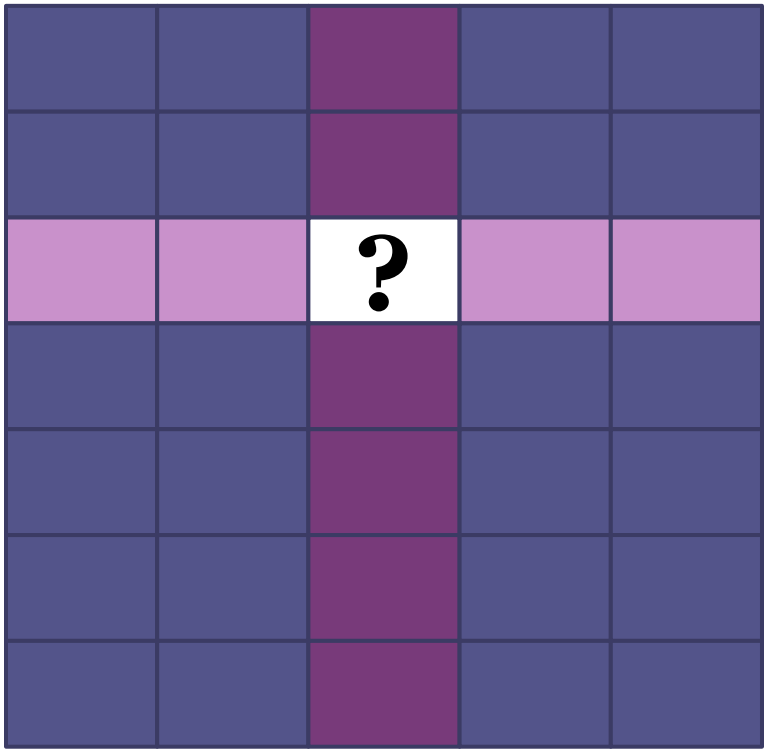
- Используется для двумерных таблиц
- Основные предположения:
  - Избыточность
  - Локальная компактность
  - Линейные зависимости в данных
- Действует локально:

  Выбирает подтаблицу и прогнозирует элемент

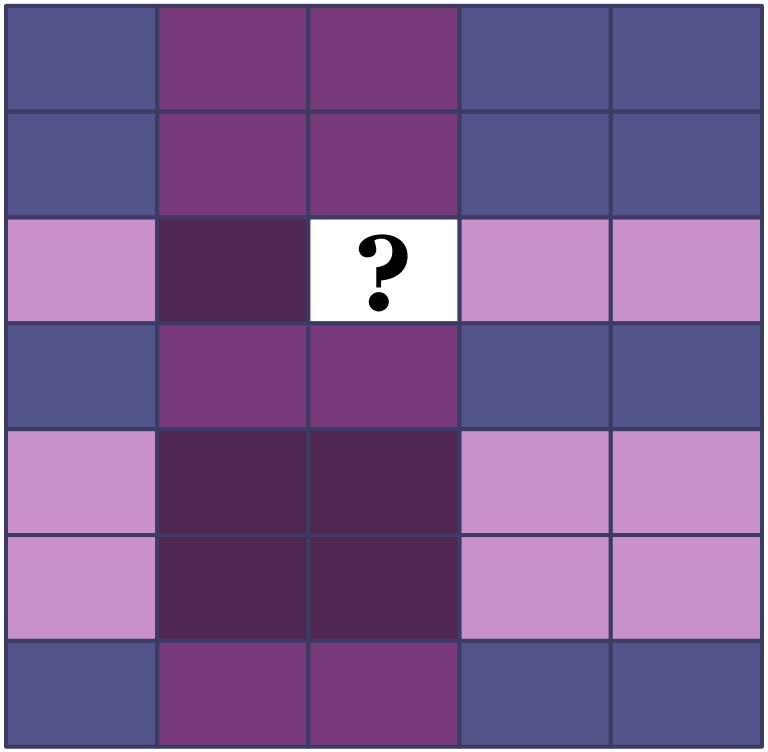
---

Загоруйко Н. Г., Елкина В. Н., Темиркаев В. С. Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм ZET) // Эмпирическое предсказание и распознавание образов . Новосибирск, 1975, Вып.61 , Вычислительные системы. С .3-27.

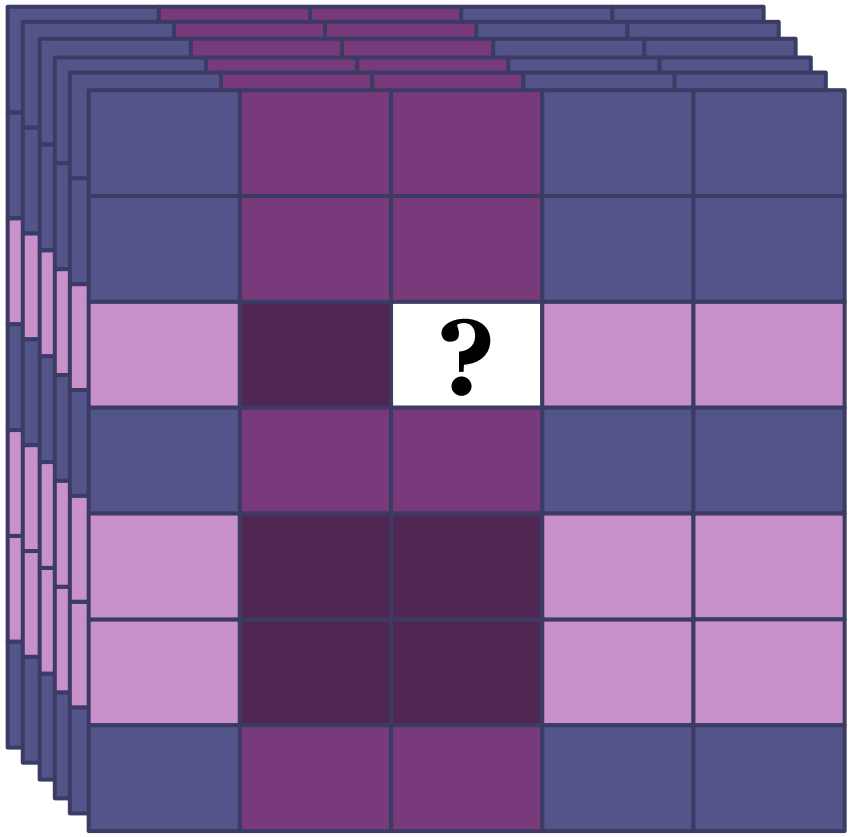
# Алгоритм ZET



# Алгоритм ZET

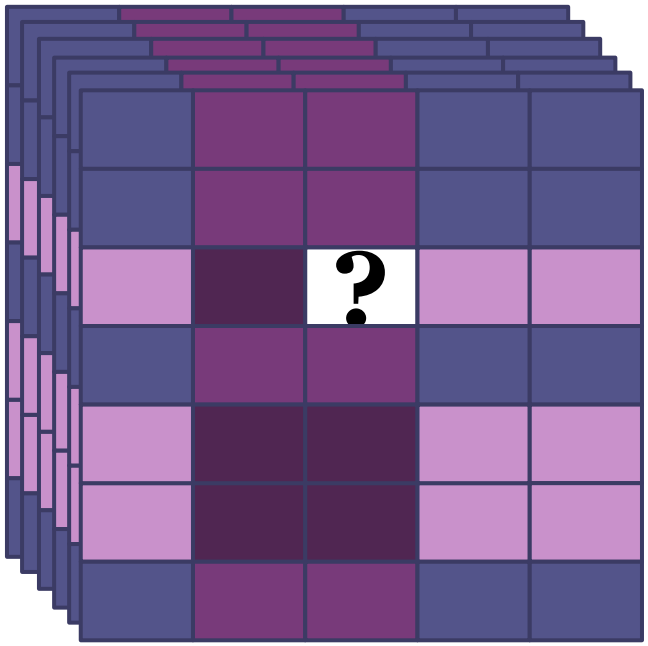


# Переход к кубу данных



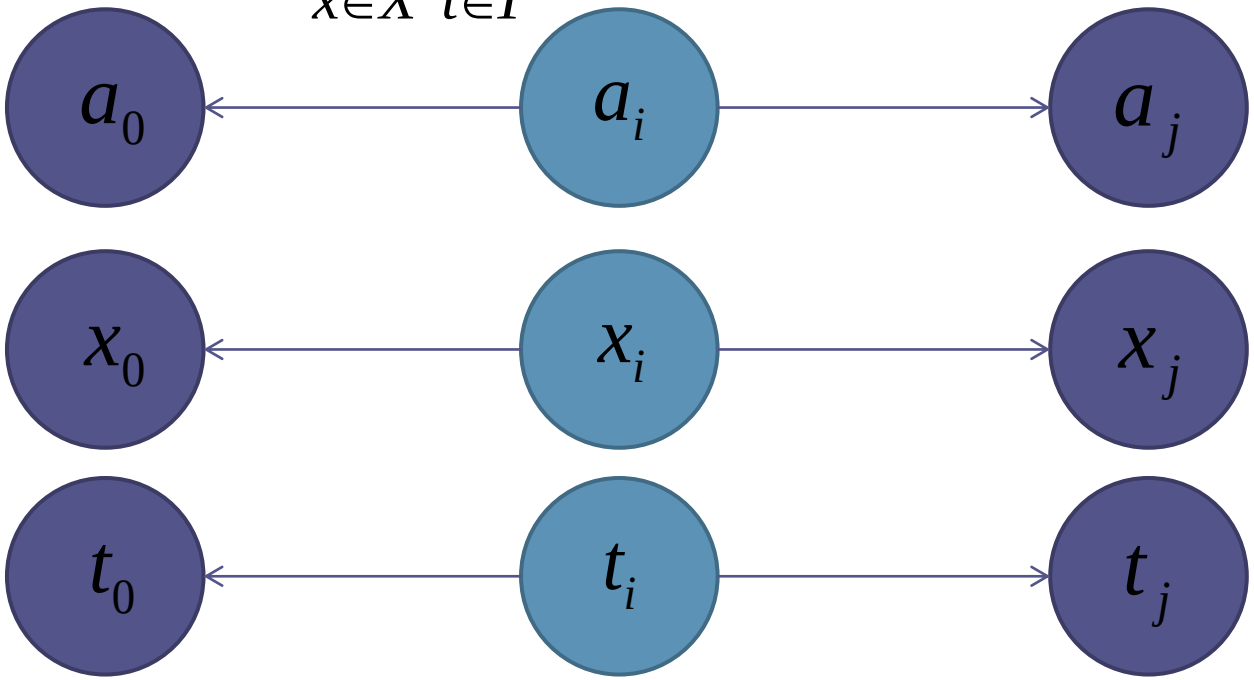
# Переход к кубу данных

- Предложить меру сходства сечений
- Предложить метод оценки компактности



# Сходство сечений

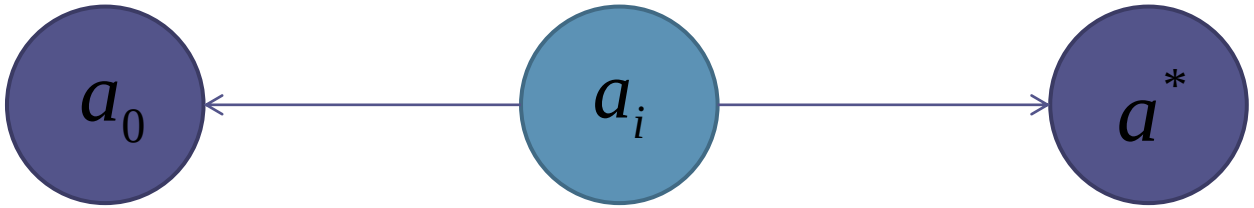
$$r^2(a_i, a_j) = \sum_{x \in X} \sum_{t \in T} \left( \langle a_i, X, T \rangle - \langle a_j, X, T \rangle \right)^2$$



# Конкурентное сходство - 1

$$F(a_i, a_0 \mid a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

$$F(a_i, a_0 \mid a^*) \approx 0$$



# Конкурентное сходство - 2

$$F(a_i, a_0 \mid a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

$$F(a_i, a_0 \mid a^*) \approx 1$$

# Конкурентное сходство – 3

$$F(a_i, a_0 \mid a^*) = \frac{r(a_i, a^*) - r(a_i, a_0)}{r(a_i, a^*) + r(a_i, a_0)}$$

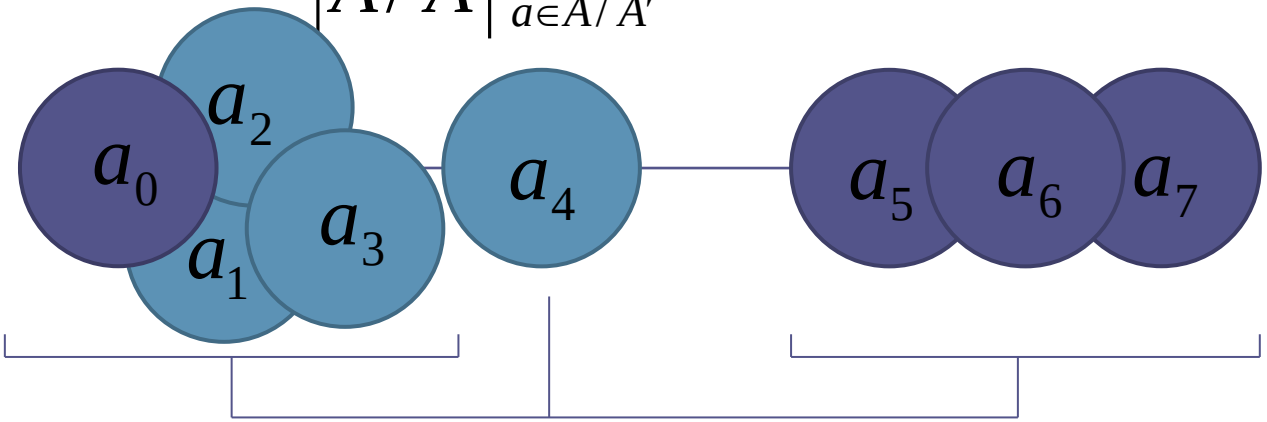
$$F(a_i, a_0 \mid a^*) \approx -1$$

# Компактность сечений

$$C(A') = \sum_{a \in A'} F(\bar{a}_0, a \mid \bar{\bar{a}}^*)$$

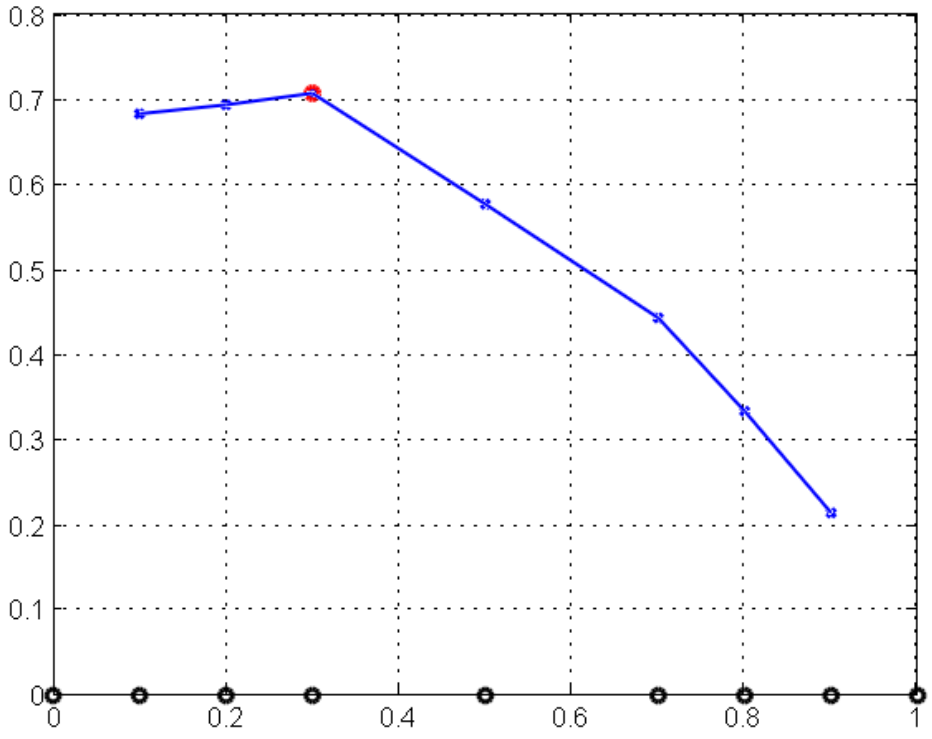
$$\bar{a}_0 : r(a_0, \bar{a}_0) = \frac{1}{|A'|} \sum_{a \in A'} r(a_0, a)$$

$$\bar{\bar{a}}^* : r(a_0, \bar{\bar{a}}^*) = \frac{1}{|A / A'|} \sum_{a \in A / A'} r(a_0, a)$$



# Компактность сечений

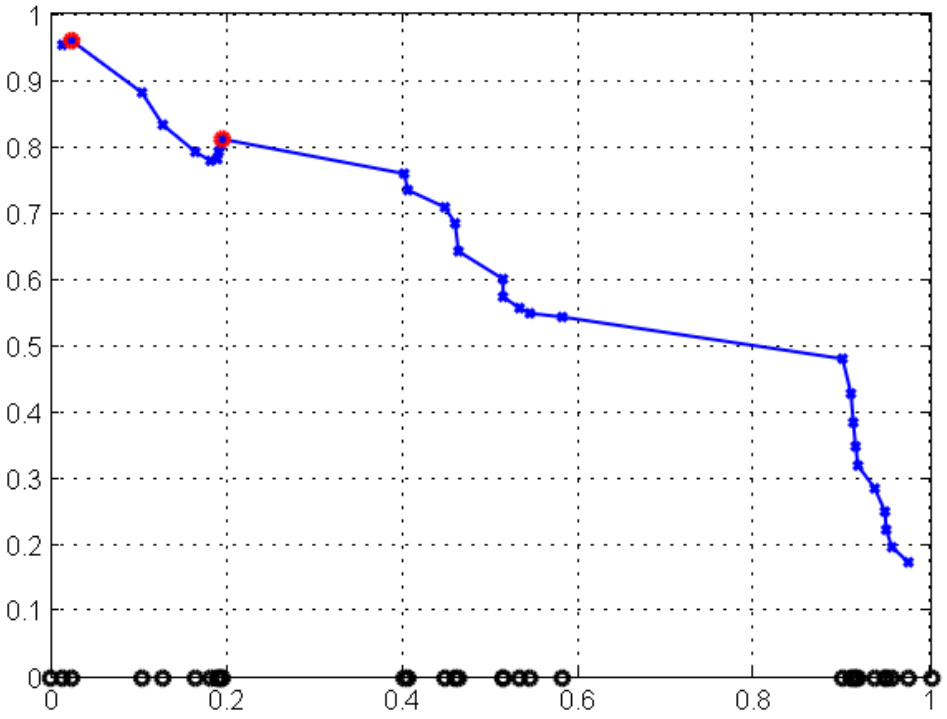
$C(\{a_0,\ldots,a_i\})$




$\{a_0,\ldots,a_n\}$

# Компактность сечений

$C(\{a_0,\ldots,a_i\})$




$\{a_0,\ldots,a_n\}$

# Построение подкуба

- Дано:

$$\langle A, X, T \rangle - \textit{куб данных}.$$

$$\textit{Значение} \langle a_0, x_0, t_0 \rangle - ?$$

- Найти:

$$\langle A', X', T' \rangle - \textit{компактный подкуб}.$$

$$\langle a_0, x_0, t_0 \rangle = M_{\langle A', X', T' \rangle}(a_0, x_0, t_0)$$

$$C\langle A', X', T' \rangle \xrightarrow[\substack{A' \subset A \\ X' \subset X \\ T' \subset T}]{} \max$$

# Приближённый алгоритм - 1

$$\langle A', X', T' \rangle = \langle A, X, T \rangle$$

- Положим искомый подкуб равным исходному кубу:

$$C\langle A', X', T' \rangle$$

- Повторять:
  - Найдём наиболее удалённые от целевых сечения $a^*, x^*, t^*$.
  - Вычислим компактность
  - Если не выполнен критерий остановки:
    - Исключаем сечения с отрицательным сходством
  - Иначе: конец.

# Приближённый алгоритм - 2

- Критерии остановки:
  1. Локальный максимум
  2. Достижение подкубом минимального размера
  3. Если по одному из направлений количество сечений с отрицательным сходством слишком велико и их исключение приведёт к значительному уменьшению размеров подкуба, целесообразно не исключать сечения на этом шаге, а подождать следующего. Выполнение условия 3 по всем направлениям останавливает алгоритм.

# Приближённый алгоритм - 3

- После итерации удаления плоскостей целесообразно пополнить подкуб небольшим числом(<50%)  случайно выбранных сечений отброшенных на предыдущих шагах

**Средство защиты от локальных экстремумов**

# Заполнение пробела

$$M_{\langle A',X',T'\rangle}(a_0,x_0,t_0) \approx \langle a_0,x_0,t_0\rangle$$

- Выбор модели прогнозирования

$$\langle a_0,x_0,t_0\rangle = \sum_{t\in T'} \langle a_0,x_0,t\rangle$$

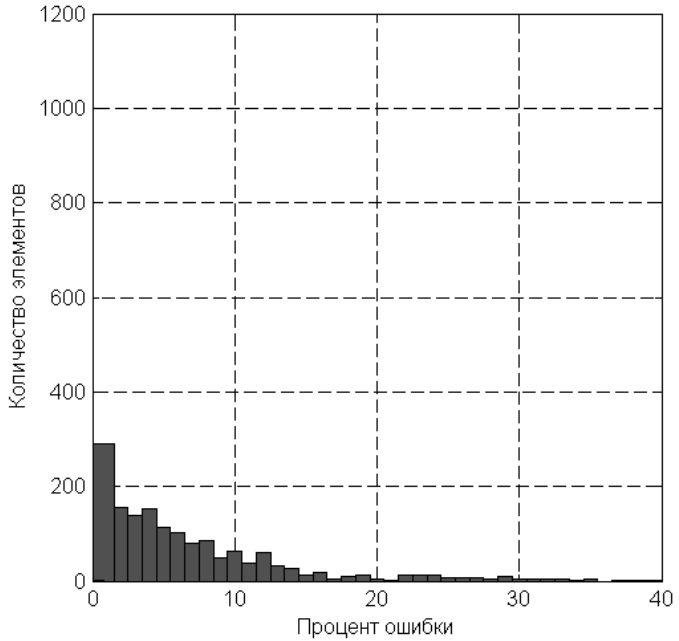
- Зависит от природы данных и характера задачи:
  - Усреднение по времени:
  - Усреднение по объектам (алгоритм kNN)
  - Регрессия по строкам/столбцам
  - …

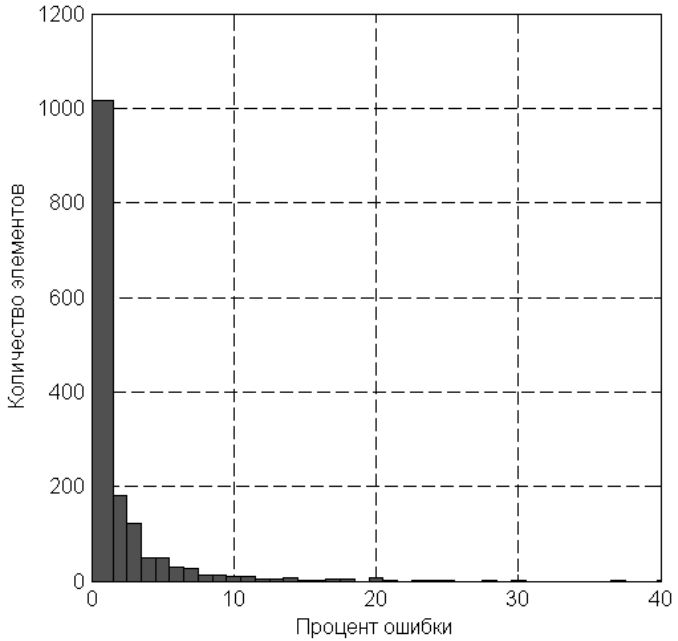
# Данные

- Данные о работе нефтяных скважин
  - 13 скважин
  - 11 параметров
  - 122 дня

| Объект | Параметр | 14.01.2012 | 15.01.2012 | 16.01.2012 | 17.01.2012 | 18.01.2012 | 19.01.2012 | 20.01.2012 | 21.01.2012 | 22.01.2012 | 23.01.2012 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Qж | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| | Qж | 28 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| | Qж | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| Скважина 4 | Qж | 28 | 28 | 27 | 27 | 27 | 27 | 28 | 28 | 28 | 28 |
| | Qн, т/с | 22,6 | 22,6 | 21,8 | 21,8 | 21,8 | 21,8 | 22,6 | 22,6 | 22,6 | 22,6 |
| | Нд | 1963 | 1963 | 2162 | 2162 | 2162 | 2162 | 2162 | 2146 | 2146 | 2146 |
| | Рб/Ру | 6,5 | 6,5 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| | Рз (Нд) | 7,8 | 7,8 | 8,4 | 8,4 | 8,4 | 8,4 | 8,4 | 7,4 | 7,4 | 7,4 |
| | Рл | 6 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| | Рзаб кров. | 43,1 | 43,1 | 42,1 | 42,1 | 42,1 | 42,1 | 41,3 | 41,1 | 41,1 | 41,1 |
| | Время работы, час | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| | Рпр | 35 | 35 | 34 | 34 | 34 | 34 | 33,2 | 33 | 33 | 33 |
| | F (Гц) | 54 | 54 | 54 | 54 | 54 | 54 | 54 | 54 | 54 | 54 |
| | Вода(%) | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |

# Данные




- 7%/2%  средняя относительная ошибка
- 76%/95% клеток в интервале [0;10]