

Министерство науки и высшего образования Российской Федерации  
Санкт-Петербургский политехнический университет Петра Великого  
Физико-механический институт

Высшая школа теоретической механики и математической физики

Работа допущена к защите

Директор ВШТМиМФ,

д.ф.-м.н., чл.-корр. РАН

\_\_\_\_\_ А. М. Кривцов

«\_\_» \_\_\_\_\_ 2023 г.

## **ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА**

**Определение коэффициентов взаимного влияния однотипных скважин для  
нефтяных месторождений с применением TDP модели**

01.03.03 «Механика и математическое моделирование»

Направленность

01.03.03\_03 Математическое моделирование процессов нефтегазодобычи

Выполнил

Студент гр. 5030103/90301:

С.Д. Боринских

Научный консультант

Руководитель направления программ по

мониторингу и повышению эффективности

активов, ООО Газпромнефть-Ноябрьскнефтегазгеофизика:

Ф.И. Кондратенко

Научный Руководитель

доцент ВШТМиМФ к.ф.-м.н.:

О.С. Лобода

Санкт-Петербург

2023

**САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО**  
**Физико-механический институт**  
**Высшая школа теоретической механики и математической физики**

УТВЕРЖДАЮ

Директор ВШТМиМФ

А.М. Кривцов

«\_\_» \_\_\_\_\_ 20\_\_ г.

**ЗАДАНИЕ**

**на выполнение выпускной квалификационной работы**

студенту Боринских Семену Денисовичу, гр. 5030103/90301

1. Тема работы: Определение коэффициентов взаимного влияния однотипных скважин для нефтяных месторождений с применением TDP-модели
2. Срок сдачи студентом законченной работы: 05.06.2023
3. Исходные данные по работе: актуальные научные публикации, связанные с темой исследования, справочная литература, обезличенные реальные данные с нефтегазового месторождения.
4. Содержание работы: анализ существующих CRM моделей, на основе которых появилась TDP модель, написание программного кода на языке Python для обработки исходных данных, обучения TDP модели и анализа результатов
5. Перечень графического материала (с указанием обязательных чертежей): не предусмотрено
6. Консультанты по работе: Ф. И. Кондратенко, руководитель направления программ по мониторингу и повышению эффективности активов ООО Газпронефть-Ноябрьскнефтегазгеофизика.
7. Дата выдачи задания: 27.02.2023  
Руководитель ВКР \_\_\_\_\_ О.С. Лобода, доцент ВШТМиМФ, к.ф.-м.н.
8. Задание принял к исполнению 27.02.2023

Студент \_\_\_\_\_ С. Д. Боринских

## РЕФЕРАТ

На 48 страниц, 24 рисунка, 2 таблицы.

TDP-МОДЕЛЬ, CRM, ППД, ОПТИМИЗАЦИЯ ППД, ПРОГНОЗ ДЕБИТА, ВРЕМЕННЫЕ РЯДЫ, SSA, МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЕМКОСТНО-РЕЗИСТИВНАЯ МОДЕЛЬ ПЛАСТА, РАЗРАБОТКА МЕСТОРОЖДЕНИЙ, ОПТИМИЗАЦИЯ СИСТЕМЫ РАЗРАБОТКИ МЕСТОРОЖДЕНИЯ

Тема выпускной квалификационной работы: «Определение коэффициентов взаимного влияния однотипных скважин для нефтяных месторождений с применением TDP-модели».

В данной работе рассмотрена модификация известной CRMP модели – TDP модель, основанная на уравнении материального баланса и притока закачки нагнетательных скважин. Предложена модификация TDP модели для учета влияния водоносного горизонта. Так же предложена модификация метода штрафных функций, а именно, придумана собственная штрафная функция для учета ограничения вида  $x > 0$ , где  $x$  – это любой неизвестный параметр, на который наложено ограничение. Использован SSA метод для улучшения сходимости решения и предложен алгоритм поиска коэффициентов взаимовлияния однотипных и разнотипных скважин, работоспособность которого подтверждена приемлемыми результатами. Разработан программный код на языке Python для решения оптимизационной задачи, т.е обучения TDP модели, обработки данных и анализа результатов.

## **ABSTRACT**

On 48 pages, 24 figures, 2 tables.

TDP MODEL, CRM, PPD, OPTIMIZATION PPD, DEBIT PREDICTION, TIME SERIES, SSA, MATHEMATICAL MODELING, CAPACITIVE-RESISTIVE MODEL, FIELD DEVELOPMENT, OPTIMIZATION OF THE FIELD DEVELOPMENT SYSTEM

The topic of the graduate qualification work: "Determination of coefficients of mutual influence of single-type wells for oil fields with the use of TDP-model".

This paper considers a modification of the well-known CRMP model - TDP model based on the equation of material balance and injection flow of injection wells. The modification of TDP model that takes into account the influence of aquifer is proposed. A modification of the penalty function method was also proposed, namely, an own penalty function was created to take into account the restriction of the form  $x > 0$ , where  $x$  is any unknown parameter on which the restriction is imposed. SSA method has been used to improve convergence of the solution and an algorithm has been proposed for searching coefficients of mutual influence of similar and different wells, the efficiency of which has been proved by acceptable results. The program code in Python language was written for solution of optimization problem, i.e. training of TDP model, for data preparation, and for analysis of results.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	5
ГЛАВА 1. CRM – МОДЕЛИ .....	7
1.1 Описание CRM модели .....	7
1.2 Описание TDP модели .....	10
1.3 Учет аквифера в TDP модели .....	12
ГЛАВА 2. ОБУЧЕНИЕ TDP МОДЕЛИ .....	14
2.1 Постановка задачи .....	14
2.2 Метод штрафных функций .....	18
2.3 Метод Оптимизации. ....	20
2.4 Метод сингулярного спектрального анализа (SSA метод) .....	23
2.5 Метод покоординатного спуска .....	28
2.6 Алгоритм обучения TDP модели. ....	29
ГЛАВА 3. ВАЛИДАЦИЯ TDP МОДЕЛИ .....	32
3.1 Постановка тестовой задачи .....	32
3.2 Результаты для тестовой задачи .....	33
ГЛАВА 4. ОБРАБОТКА ИСХОДНЫХ ДАННЫХ ДЛЯ ОБУЧЕНИЯ ....	35
4.1 Проблемы реальных данных .....	35
4.2 Заполнение пропусков .....	36
ГЛАВА 5. РЕЗУЛЬТАТЫ .....	40
5.1 Результаты для основной задачи .....	40
ЗАКЛЮЧЕНИЕ .....	44
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ .....	46

## ВВЕДЕНИЕ

Во время решения задач для управления разработкой месторождений оптимальным образом часто бывает необходима оперативная оценка гидродинамического взаимодействия скважин. Оперативная оценка может использоваться, например, для поддержания пластового давления путем закачки жидкости в пласт через нагнетательные скважины. Так же, подобранные коэффициенты взаимодействия могут указывать на ряд неожиданных случаев, возникших во время разработки месторождения. Например, высокая взаимосвязь удаленных друг от друга скважин может указывать на наличие самопроизвольного гидроразрыва пласта. Низкая взаимосвязь между рассматриваемыми скважинами, может говорить о наличии непроводящего разлома. Таким образом, понимание карты гидродинамической взаимосвязи всех скважин позволит корректно планировать различные геолого-технические мероприятия для достижения необходимых режимов работы добывающих и нагнетательных скважин.

Оценка гидродинамической связи может проводиться с помощью различных инструментов, выбор которых зависит от требований к времени вычисления, качеству исходных данных и к точности результатов.

Один из вариантов – это поиск коэффициентов взаимодействия с помощью гидродинамических симуляторов. В них используются численные модели, которые позволяют учесть сложные физические эффекты, но затрудняют оперативную оценку параметров разработки во время мониторинга. Так же, гидродинамические симуляторы требуют большего количества времени и вычислительных ресурсов для оценки необходимых параметров, что является минусом именно для оперативной оценки гидродинамической связи.

Для решения данной проблемы можно использовать базовые полуаналитические модели Capacitance Resistive Model (CRM), которые

позволяют рассчитать гидродинамическую связь между однотипными скважинами. На данный момент существует огромное количество работ, посвященных CRM моделям. В данной же работе будет рассмотрена модификация CRMP модели – Transmissibility defined productivity (TDP) модель. На текущий момент – это совершенно новая и единственная модель, способная дополнительно учитывать взаимовлияние однотипных скважин типа «добывающая – добывающая» за счет определения коэффициентов взаимной проводимости между добывающими скважинами. Принцип работы указанных моделей подробнее описан в главе 1.

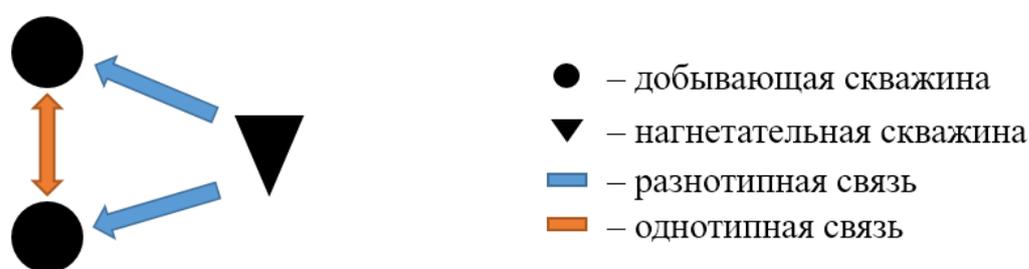


Рис.1. Типы взаимовлияния скважин.

В данной работе представлено исследование и реализация TDP модели и её обучение на данных с реального месторождения. Так же предложена модификация модели, расширяющая количество учитываемых эффектов.

# ГЛАВА 1. CRM – МОДЕЛИ

## 1.1 Описание CRM модели

CRM (сокращение от англ. Capacitance Resistive Model) – емкостно-резистивная модель, в основе которой лежат уравнение материального баланса системы и уравнение Дюпюи. Уравнение материального баланса может быть записано для разных случаев, что и порождает различные CRM модели. Ниже представлено описание лишь одного типа CRM модели, а именно CRMP.

1) Модель CRMT. В этом случае выделяют виртуальную добывающую и нагнетающую скважину, дебит которых равен суммарному дебиту соответствующих реальных скважин

$$c_t V_p \frac{d\tilde{p}}{dt} = I_F(t) - q_F(t) \quad (1)$$

где  $\tilde{p}$  – среднее пластовое давление,  $V_{p_i}$  – поровый объем пласта,  $I_F$  – суммарная закачка в пласт,  $q_i$  – суммарная добыча из пласта.

Данная модель позволяет получить описание пласта в целом, но не предоставляет никакой информации по каждой добывающей скважине. В данной работе не рассматривается данная модель, поскольку она не даёт информации о взаимодействии всех реальных скважин.

2) Модель CRMP. Её суть заключается в том, что каждой  $i$ -й добывающей из  $N_{\text{доб}}$  шт. ставится в соответствие  $N_{\text{наг}}$  шт. нагнетательных. Уравнение материального баланса записывается в виде (2):

$$c_t V_{p_i} \frac{d\tilde{p}_i}{dt} = I_{p_i}(t) - q_i(t) \quad (2)$$

где  $\tilde{p}_i$  – среднее давление в рассматриваемом поровом объеме  $V_{p_i}$ ,  $I_{p_i}$  – суммарная закачка, приходящаяся на данный поровый объем,  $q_i$  – дебит добывающей скважины. Данная модель взята в качестве основной (из CRMP

модели появилась TDP модель). Здесь и далее под CRM-моделью подразумевается именно эта модель.

3) Модель CRMIP. В этом случае рассматривается поровый объем  $V_{p_{ij}}$  между  $i$ -й добывающей и  $j$ -й нагнетательной скважиной. Материальный баланс записывается в виде (3):

$$c_t V_{p_{ij}} \frac{d\bar{p}_{ij}}{dt} = I_{p_{ij}}(t) - q_{ij}(t) \quad (3)$$

где  $\bar{p}_{ij}$  - среднее давление в рассматриваемом поровом объеме  $V_{p_{ij}}$ ,  $I_{p_{ij}}$  - суммарная закачка, приходящаяся на данный поровый объем,  $q_{ij}$  - суммарная добыча, приходящаяся на данный поровый объем. Данная модель взята в качестве основной (из CRMP модели появилась TDP модель). Здесь и далее под CRM-моделью подразумевается именно эта модель.

Далее рассмотрим уравнение материального баланса для CRMP-модели (2):

$$c_t V_{p_i} \frac{d\bar{p}_i}{dt} = I_{p_i}(t) - q_i(t)$$

Для определения доли закачки окружающих скважин, поступающей в поровый объем некоторой добывающей скважины  $i$  для каждой  $j$ -й нагнетательной скважины, вводится параметр связности  $f_{ij}$ :

$$I_{p_i}(t) = \sum_{j=1}^{N_{inj}} f_{ij} I_j(t) \quad (4)$$

Введённый параметр имеет следующие ограничения (5):

$$f_{ij} \leq 1 \quad (5)$$

$$\sum_{j=1}^{N_{prod}} f_{ij} \leq 1, i = \overline{1, \dots, N_{inj}}$$

Следовательно, уравнение материального баланса (3) будет иметь вид (6):

$$c_t V_{pi} \frac{d\bar{p}_i}{dt} = \sum_{j=1}^{N_{inj}} f_{ij} I_j(t) - q_i(t) \quad (6)$$

Предполагается, что дебит скважины пропорционален пластовой депрессии с коэффициентом пропорциональности  $J_i$ :

$$q_i(t) = J_i(\bar{p}_i - p_{wi}) \quad (7)$$

Исключим из уравнения (6) пластовое давление, выраженное из (7):

$$\frac{dq_i(t)}{dt} + \frac{q_i(t)}{\tau_i} = \frac{1}{\tau_i} \sum_{j=1}^{N_{inj}} f_{ij} I_j(t) - J_i \frac{dp_{wi}(t)}{dt}, \text{ где } \tau_i = \frac{c_t V_{pi}}{J_i} \quad (8)$$

Получили неоднородное линейное дифференциальное уравнение первого порядка. Решая его с помощью метода Вариации произвольной постоянной в предположении, что  $J_i$  и  $\tau_i = const$ , можно получить формулу для дебита  $i$ -й добывающей скважины, имеющую вид (9):

$$q_i(t) = q_{0i}(t) e^{-\frac{t-t_0}{\tau_i}} + e^{-\frac{t}{\tau_i}} \int_{t_0}^t \frac{1}{\tau_i} e^{\frac{\psi}{\tau_i}} \left( \sum_{j=1}^{N_{inj}} f_{ij} I_j(\psi) \right) d\psi -$$

$$- J_i \left[ p_{wi}(t) - e^{-\frac{t-t_0}{\tau_i}} p_{wi}(t_0) \right] + \frac{J_i}{\tau_i} e^{-\frac{t}{\tau_i}} \int_{t_0}^t e^{\frac{\psi}{\tau_i}} p_{wi}(\psi) d\psi \quad (9)$$

Предположим, что реальные данные дебитов дискретны. Следовательно, это позволяет вычислить интегралы в (9) и представить производные приемистостей и забойного давления в виде (10):

$$\left. \frac{dp_w(t)}{dt} \right|_k^{k+1} = \beta^k = \frac{p_w(t_{k+1}) - p_w(t_k)}{t_{k+1} - t_k} = const \quad (10)$$

С учетом введенных обозначений, используя суперпозицию по времени (решение на конце одного интервала используется как начальное условие для следующего временного интервала) рекуррентная оценка дебита жидкости в момент времени  $t_n$  запишется в виде:

$$q_i(t_n) = q_i(t_{n-1})e^{\frac{-\Delta t_n}{\tau_i}} + \left(1 - e^{\frac{-\Delta t_n}{\tau_i}}\right) \left[ \sum_{x=1}^{N_{inj}} [f_{xi} I_x^{(n)}] - J_i \tau_i \frac{\Delta p_w^{(n)} i}{\Delta t_n} \right] \quad (11)$$

Данная формула (11) является ключевой для CRMP модели.

## 1.2 Описание TDP модели

Отличие TDP модели от CRMP модели заключается в возможности учета связи однотипных скважин (типа «добывающая – добывающая») за счет добавления к формуле (11) члена  $\sum_{x=1}^{N_{prod}} [T_{ix}(p_{w_x}^n - p_{w_i}^n)]$ . Данное слагаемое описывает поток жидкости между добывающими скважинами через произведение взаимной проводимости (коэффициента  $T_{ij}$ ) на разницу забойного давления. Знак разницы забойного давления определяет направление потоков между добывающими скважинами. Данный подход можно описать формулой (12):

$$q_i(t_n) = q_i(t_{n-1})e^{-\frac{\Delta t_n}{\tau_i}} + \left(1 - e^{-\frac{\Delta t_n}{\tau_i}}\right) * \left[ \sum_{x=1}^{N_{inj}} [f_{xi} I_x^{(n)}] - J_i \tau_i \frac{\Delta p_w^{(n)} i}{\Delta t_n} + \sum_{x=1}^{N_{prod}} [T_{ix} (p_{wx}^n - p_{wi}^n)] \right] \quad (12)$$

где  $q_i$  – дебит жидкости,  $\tau_i$  – постоянная времени,  $f_{ij}$  – коэффициент взаимовлияния разнотипных скважин,  $I_x$  – приемистость,  $J_i$  – коэффициент продуктивности,  $p_w$  – забойное давление,  $T_{ix}$  – коэффициент проводимости между добывающими скважинами.

В общем случае применение данной модели на практике может вызывать некоторые затруднения:

- 1) Если забойные давления меняются мало, а также если их значения на рядом расположенных добывающих скважинах примерно равны друг другу, то значение коэффициента взаимной проводимости между ними не может быть достаточно корректно определено. Это объясняется отсутствием репрезентативной обучающей выборки
- 2) При числе добывающих скважин более 10, задача может стать неустойчивой. Причины неустойчивости требуют дополнительных исследований, однако на простейшем уровне можно понять, что из-за увеличения размерности задачи появляется возможность появления равновероятных, но качественно различных решений. Если переформулировать данную фразу, то одни и те же изменения дебита жидкости можно смоделировать совершенно разными способами
- 3) Модель крайне требовательна к точности данных по забойному давлению. Погрешности в замерах забойного давления на месторождении или полное отсутствие этих данных приводят к неустойчивости задачи и невозможности её использования.

Однако не смотря на ограничения, и на то, что TDP модель является новой и не до конца изученной, эксперименты показали, что в некоторых случаях применение TDP модели способствует существенному улучшению сходимости модели CRMP.

### 1.3 Учет аквифера в TDP модели

Во многих работах, например в O.Izges, C.S. Kabir Quantifying Nonuniform Aquifer Strength at Individual Wells [17], описывается метод моделирования CRM с учётом водоносного горизонта. Адаптировав данную идею для TDP модели, получим формулу (13):

$$q_i(t_n) = q_i(t_{n-1})e^{\frac{-\Delta t_n}{\tau_i}} + \left(1 - e^{\frac{-\Delta t_n}{\tau_i}}\right) * \left[ \sum_{x=1}^{N_{inj}} [f_{xi} I_x^{(n)}] - J_i \tau_i \frac{\Delta p_w^{(n)} i}{\Delta t_n} + \sum_{x=1}^{N_{prod}} [T_{ix} (p_{w_x}^{(n)} - p_w^{(n)} i)] + w_i \right] \quad (13)$$

где  $w_i$  – приток аквифера.

Значение аквифера для каждой скважины рассчитывалось с учётом его равномерного распределения. Представленная формула (13) отличается от (12) дополнительным учётом притоком с водоносного горизонта к приемистости нагнетательной скважины и потоку жидкости между добывающими скважинами.

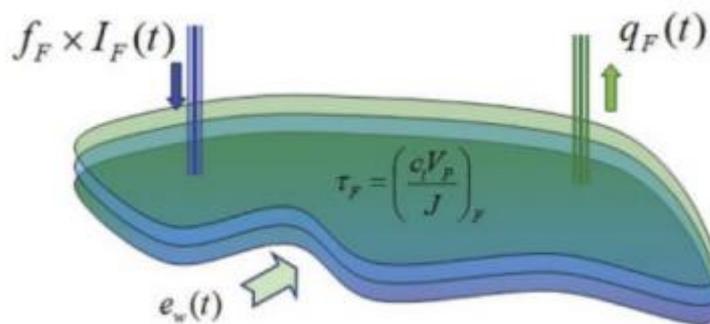


Рис 2 Представление виртуальной скважины, отвечающей за аквифер в модели CRMT  
(заимствовано из [17])

Далее, на рисунке 3 представлено сравнение результатов обучения TDP модели с учетом и без учета аквифера.



Рис.3. Сравнение результатов TDP модели с учетом аквифера и без учета аквифера на обучающей выборке

Из рисунка 3 отчетливо видно, что данная модификация TDP модели позволяет значительно улучшить работоспособность модели и получать гораздо лучшие результаты.

## ГЛАВА 2. ОБУЧЕНИЕ TDP МОДЕЛИ

### 2.1 Постановка задачи

В данной работе в качестве исходных данных используются приемистость нагнетающих скважин, дебит жидкости и забойное давление добывающих скважин. Эти данные используются для оценки дебита жидкости скважин по формуле (13).

Неизвестными параметрами TDP модели являются

- $f_{ij}$  – коэффициент взаимовлияния скважин,  $i = \overline{1, N_{prod}}$ ,  $j = \overline{1, N_{inj}}$

Коэффициент взаимовлияния  $f_{ij}$  между добывающей и нагнетательной скважиной – это безразмерная величина, характеризующая долю полезной закачки  $i$ -ой нагнетательной скважины, приходящейся на  $j$ -ую добывающую скважину.

- $J_i$  – коэффициент продуктивности,  $i = \overline{1, N_{prod}}$

Коэффициент продуктивности добывающей скважины – это коэффициент пропорциональности между дебитом добывающей скважины и пластовой депрессии. Данный коэффициент показывает на сколько может измениться дебит скважины при единичном изменении депрессии на пласт.

- $\tau_i$  – временной шаг (инерционный параметр),  $i = \overline{1, N_{prod}}$

Параметр «тау» связан с запасом упругой энергии в пласте и определяет скорость распространения возмущений в пласте. На рисунке 4 показана динамика дебита жидкости по модели CRM, а также динамика закачки жидкости в пласт.

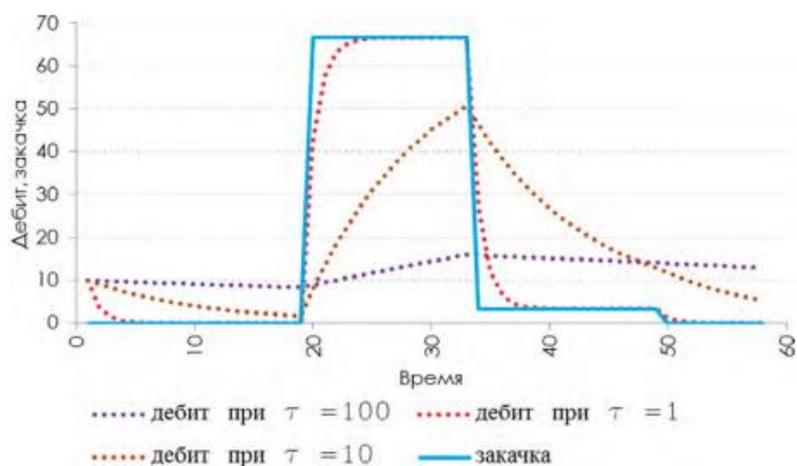


Рис. 4. Влияние параметра «тау» на добычу жидкости (заимствовано из [6])

Видно, что изменение данного параметра «тау» ведет изменению скорости реагирования на закачку. При  $\tau = 100$  возмущения, передаваемые от системы нагнетательных скважин практически незаметны. Уменьшение  $\tau$  в 10 раз приводит к более быстрой реакции добычи на закачку. При значении  $\tau$  равном единице динамика добычи жидкости практически не отличается от динамики закачки жидкости.

- $T_{ij}$  – коэффициент проводимости между добывающими скважинами,  $i, j = \overline{1, N_{prod}}$

Коэффициент проводимости между добывающими скважинами имеет размерность коэффициента продуктивности  $\left(\frac{\text{м}^3}{\text{мес} \cdot \text{Па}}\right)$ . Соответственно, смысл этих коэффициентов очень похож. Коэффициент проводимости показывает насколько может измениться дебит жидкости  $i$ -й добывающей скважины при изменении разницы забойного давления  $i$ -й и  $j$ -й добывающей скважины.

- $w_i$  – приток аквифера,  $i = \overline{1, N_{prod}}$

Приток с аквифера имеет размерность  $\left[\frac{\text{м}^3}{\text{мес}}\right]$  и характеризует количество жидкости, которая перетекает в пласт из водоносного горизонта и оказывает влияние на добывающие скважины.

В сумме имеется  $N_{prod} * (N_{prod} + N_{inj} + 3)$  неизвестных коэффициентов.

Значения данных коэффициентом вычисляются путем решения оптимизационной задачи. В качестве функции потерь для решения оптимизационной задачи выступает функция средней квадратичной ошибки между истинным значением дебита и предсказанным по формуле TDP модели, имеющая вид (14).

$$F = \frac{1}{N_{prod}} \sum_{i=1}^{N_{prod}} \sum_{n=1}^T (\bar{q}_i(t_n) - q_i(t_n))^2 \quad (14)$$

где  $\bar{q}_i(t_n)$  – истинное значение дебита  $i$ -й скважины в  $n$ -й шаг по времени,  $q_i(t_n)$  – вычисленное по формуле (13) значение дебита  $i$ -й скважины в  $n$ -й шаг по времени.

На основе полученной функции потерь (14) возможно решить оптимизационную задачу с целью определения значений параметров  $f_{ij}, T_{ij}, J_i, \tau_i, w_i$ . Оптимизационная задача решается с использованием метода штрафных функций (пункт 2.2), оптимизатора Adam (пункт 2.3), метода сингулярного спектрального анализа (SSA метод) (пункт 2.4) и метода покоординатного спуска (пункт 2.5).

Без учёта метода покоординатного спуска перечисленные выше методы взаимодействуют друг с другом согласно блок-схеме, представленной на рисунке 5.



Рис.5. Блок-схема с алгоритмом решения оптимизационной задачи

Далее, опираясь на физику решаемой задачи, должны быть поставлены следующие ограничения на искомые коэффициенты:

$$\sum_{i=1}^{N_{prod}} f_{ij} \leq 1, \quad j = \overline{1, N_{inj}} \quad (15)$$

$$f_{ij} > 0, \quad i = \overline{1, N_{prod}} \quad (16)$$

$$J_i > 0, \quad i = \overline{1, N_{prod}} \quad (17)$$

Условие (15) означает, что  $j$ -я нагнетательная скважина не может отдавать больше, чем 100% закачиваемой данной скважиной жидкости. Условие (16) накладывает ограничение на тип связи скважин: требование положительность коэффициентов обеспечивает положительную корреляцию между скважинами разного типа (нагнетательная-добывающая). Условие (17)

необходимо для того, чтобы исключить возможность перехода добывающей скважины в режим работы нагнетательной и наоборот.

## 2.2 Метод штрафных функций

Метод штрафных функций (МШФ) используется для сведения условной задачи оптимизации к безусловной. Избавиться от условий необходимо для того, чтобы возможно было воспользоваться градиентным методом оптимизации. Подробно сам метод описан в [20].

Суть метода штрафных функций заключается в следующем:

Пусть требуется найти минимум  $f(x_1, \dots, x_n)$  при ограничениях  $g_i(x_1, \dots, x_n) > 0, i = 1, 2, \dots, m \rightarrow$

Формулируется новая задача для нахождения минимума новой функции  $F(x_1, \dots, x_n) = f(x_1, \dots, x_n) + P(x_1, \dots, x_n)$ , где  $P(x_1, \dots, x_n)$  - штрафная функция вида:

- 1)  $P(\bar{x}) = r \sum_{i=1}^m \frac{1}{g_i(\bar{x})}$ ;
- 2)  $P(\bar{x}) = r \sum_{i=1}^m (g_i(\bar{x}))^2$ ;
- 3)  $P(\bar{x}) = r \sum_{i=1}^m (-\ln(g_i(\bar{x})))$ ;

где  $\bar{x}$  – вектор входных параметров,  $r$ -некий положительный параметр.

Алгоритм МШФ:

- 1) Решается задача оптимизации для  $F(\bar{x})$  с начальным приближением  $\bar{x}_0, r_0$
- 2) Проверяется условие выхода из итерационного цикла:  $|P(\bar{x})| < eps \rightarrow 0$
- 3) А) в случае выполнения условия итерационный цикл заканчивается и на выходе имеется искомый минимум функции

Б) в случае не выполнения условия положительный параметр  $r_k$  уменьшается следующим образом:  $r_k = \frac{r_{k-1}}{C}$ ,  $C > 1$ . После данного действия происходит повторение первого и второго шага МШФ.

В ходе исследования TDP-модели метод штрафных функций был усовершенствован, а именно, была придумана собственная штрафная функция (18) для ограничение  $x > 0$ :

$$P(x) = 1 - e^{-30(x-0.5)^2} \text{ если условие } x > 0 \text{ выполнено;} \quad (18)$$

$$P(x) = e^{70x^2}, \text{ если условие } x > 0 \text{ не выполнено;}$$

Данная функция имеет вид, изображенный на рисунке 6.

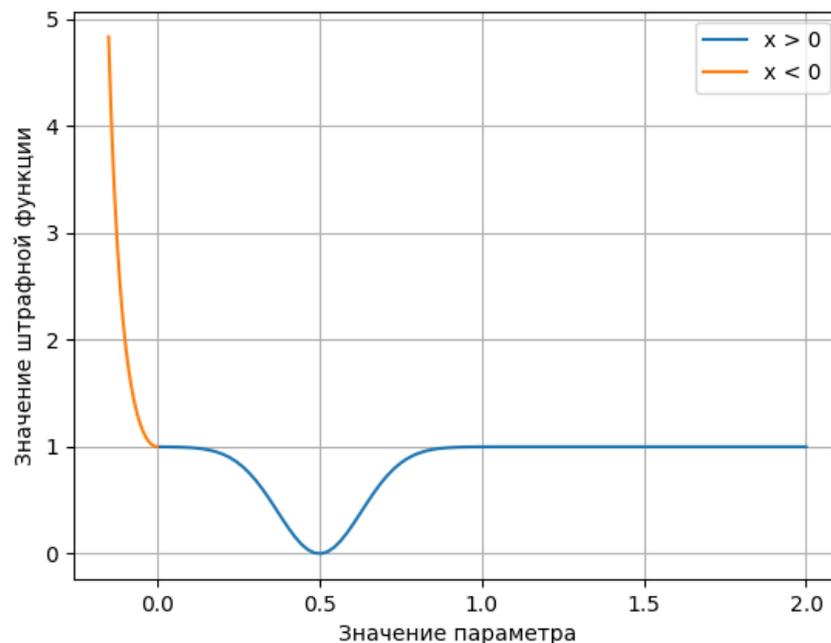


Рис.6. Поведение штрафной функции (18).

Эта функция используется для того, чтобы давать небольшой штраф в случае положительного значения  $x$ , тем самым слегка приближая его к значению 0.5, и чтобы накладывать большой, причём быстро увеличивающийся, штраф в случае отрицательного значения  $x$ .

В рамках поставленной задачи метод штрафных функций применяется к формуле дебита TDP модели (13) для добывающей скважины.

### 2.3 Метод Оптимизации

В данной работе используется оптимизатор Adam, реализованный в библиотеке Tensorflow для языка программирования Python. Adam относится к градиентным методам оптимизации. В его основе лежат такие методы, как стохастический градиентный спуск, метод Нестерова и метод Adagrad.

Разберём подробнее эволюцию градиентных методов. Начало лежит в методе градиентного спуска. Вспомним, что градиент – это направление наискорейшего локального возрастания функции. В таком случае отрицательный градиент – это направление наискорейшего локального убывания.

Пусть  $x_0$  – начальная точка для метода градиентного спуска. Тогда каждое следующее приближение выбирается следующим образом:

$$x_{k+1} = x_k - \alpha \nabla loss \quad (19)$$

где  $\alpha$  – это размер шага (learning rate).

Недостатком данного метода является застраивание в локальных минимумах.

Если разделить всю выборку на подмножества (мини-батчи) и считать градиент по ним, то получим метод стохастического градиентного спуска. В данном случае следующее приближение считается так:

$$x_{k+1} = x_k - \alpha \sum_n \nabla loss_n \quad (20)$$

Различие в действии двух перечисленных методов можно рассмотреть на рисунке 7.



Рис.7. Принцип действия метода градиентного спуска и стохастического градиентного спуска (заимствовано из [18]).

Далее, чтобы приближение не застревало в локальных минимумах, было предложено ввести некую инерцию в предыдущую формулу (20). Математически идея метода Моментов (21) реализуется следующим образом: прежнее значение градиента умножается на дисконтирующий множитель  $0 < \gamma < 1$ . К нему и прибавляется текущее значение градиента для параметра  $\theta$ :

$$u_{k+1} = \gamma u_k - \alpha \nabla \text{loss}(\theta_k) \quad (21)$$

$$\theta_{k+1} = \theta_k - u_k$$

Развитие этой идеи заключалось в том, чтобы вычислять градиент не в самой точке  $\theta$ , а в следующей точке  $\theta_t - \gamma u_k$ , тем самым заглядывая немного вперед при вычислении градиента. Эта идея метода Нестерова, которая математически записывается следующим образом (22):

$$u_{k+1} = \gamma u_k - \alpha \nabla \text{loss}(\theta_t - \gamma u_k) \quad (22)$$

$$\theta_{k+1} = \theta_k - u_k$$

Метод моментов и метод Нестерова учитывали лишь историю изменения градиента, но никак не были связаны с самими оптимизируемыми параметрами. Следующая идея заключалась в том, что некоторые параметры могут достигать своего оптимального значения быстрее, чем другие. Учёт

это в градиентном спуске, изменяя learning rate у каждого параметра индивидуально. Одним из первых алгоритмов, в котором была реализована эта идея, был метод Adagrad. В данном методе шаг для параметров зависит от величины их флуктуаций: чем они больше, тем меньше шаг. Математически его можно записать так:

Пусть градиент для  $i$ -го параметра, равен:

$$g_{k,i} = \nabla_{\theta_k} \text{loss}(\theta) \quad (23)$$

Тогда обновление  $i$ -го параметра будет иметь вид:

$$\theta_{k+1} = \theta_k - \frac{\alpha}{\sqrt{G_{k,ii} + \epsilon}} g_{k,i} \quad (24)$$

где  $G_{k,ii} = G_{k-1,ii} + g_{k,i}^2$  – сумма квадратов градиентов для  $i$ -го параметра,  $\epsilon > 0$  – положительная бесконечно малая величина, введённая для отсутствия деления на ноль.

Главный недостаток данного алгоритма – это постоянное уменьшение шага обучения. Кроме того, иногда бывает сложно выбрать глобальную скорость обучения  $\eta$ .

Для преодоления недостатка алгоритма Adagrad был придуман метод оптимизации под названием Adadelta. Идея метода в сохранении не всей истории по квадратам градиентов, а лишь некоторую небольшую её часть. Реализуем это путём добавления дисконтирующего множителя:

$$G_{k,ii} = \rho G_{k-1,ii} + (1 - \rho) g_{k,i}^2 \quad (25)$$

$$0 < \rho < 1$$

Во всем остальном логика работы совпадает с методом Adagrad.

Наконец рассмотрим метод оптимизации, используемый в данной работе, т.е метод Adam. Он часто применяется для обучения нейронных сетей

и хорошо себя показал в поиске минимума функции потерь. Этот алгоритм является очередной модификацией алгоритма Adagrad, использующей сглаженные версии среднего и среднеквадратического градиентов. Математически это выглядит следующим образом:

$$\begin{aligned}
 v_{k+1} &= \beta_1 v_k + (1 - \beta_1) \nabla \text{loss}(u_k), \\
 G_{k+1} &= \beta_2 G_k + (1 - \beta_2) (\nabla \text{loss}(u_k))^2, \\
 u_{k+1} &= u_k - \frac{\alpha}{\sqrt{G_{k+1} + \epsilon}} v_{k+1};
 \end{aligned}
 \tag{26}$$

где  $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$ ,  $\epsilon = 10^{-8}$  – заданные заранее гиперпараметры.  $u_k$  – оптимизируемая величина.

## 2.4 Метод сингулярного спектрального анализа (SSA метод)

SSA (Singular spectrum analysis) – это метод, используемый для анализа временных рядов. Метод заключается в том, что происходит преобразование одномерного временного ряда в многомерный ряд, к которому в последующем применяется метод главных компонент [2]. Основная цель SSA – это выделить из входного сигнала тренд, периодические компоненты и шум.

Приведём математическое описание SSA метода, следуя [2]. Рассмотрим вещественный временной ряд  $F = (f_0, \dots, f_{N-1})$  длины  $N$ . Предположим, что длина временной ряда  $> 2$  и что он ненулевой, то есть существует хотя бы одно значение  $i$  такое, что  $f_i \neq 0$ . Алгоритм состоит из двух этапов, дополняющих друг друга. Рассмотрим их по порядку.

## Этап разложения

### 1) Процедура вложения

Данная процедура переводит временной ряд в последовательность многомерных векторов. Введём параметр  $L$  – длину окна ( $1 < L < N$ ). Вложение образует  $K = N - L + 1$  новых векторов вложения

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, 1 \leq i \leq K,$$

имеющих размерность  $L$ .

Далее, введём понятия траекторной матрицы временного ряда –  $X$ . Данная матрица состоит из векторов вложения

$$X = [X_1: \dots : X_K]$$

Иначе говоря, траекторная матрицы – это матрица вида:

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix} \quad (27)$$

### 2) Сингулярное разложение

На втором шаге происходит сингулярное разложение траекторной матрицы временного ряда (27).

Обозначим  $S = XX^T$ . Пусть  $\lambda_1, \dots, \lambda_L$  – неотрицательные собственные числа матрицы  $S$ , взятые в порядке возрастания, и  $U_1, \dots, U_L$  – ортонормированная система собственных векторов матрицы  $S$ , соответствующая собственным числам.

Далее, пусть  $d$  – максимальное неотрицательное собственное число. Если обозначить  $V_i = X^T U_i \frac{1}{\sqrt{\lambda_i}}, i = 1 \dots d$ , то сингулярное разложение матрицы  $X$  может быть записано в виде (21):

$$X = X_1 + \dots + X_d \quad (28)$$

где  $X_i = \sqrt{\lambda_i} U_i V_i^T$ . Каждая матрица  $X_i$  имеет ранг, равные единице, следовательно, эти матрицы можно называть элементарными.

Набор матриц  $(\sqrt{\lambda_i} U_i V_i^T)$  называют  $i$ -й сингулярной тройкой сингулярного разложения  $X$  (28).

Этап восстановления

### 3) Процедура группировки

На основе сингулярного разложения  $X$  (28) процедура группировки делит множество индексов  $\{1, \dots, d\}$  на  $m$  штук непересекающихся подмножеств  $I_1, \dots, I_m$ .

Пусть  $I = \{i_1, \dots, i_p\}$ . В таком случае результирующая матрица сгруппированного сингулярного разложения  $X_I$ , которая соответствует группе  $I$ , будет определена как (29):

$$X_I = X_{i_1} + \dots + X_{i_p} \quad (29)$$

Такого вида матрицы вычисляются для  $I = I_1, \dots, I_m$ . Процедура выбора множества  $I$  называется группировкой сингулярных троек.

### 4) Диагональное усреднение

Данный шаг является последним в базовом алгоритме SSA. На этом шаге результирующая матрица сгруппированного сингулярного разложения  $X_I$  (29) преобразуется в новые временной ряд, имеющий начальную длину  $N$ .

Пусть  $Y$  – некоторая матрица размера  $L \times K$ , имеющая элементы  $y_{ij}$ , где  $1 \leq i \leq L$ ,

$1 \leq j \leq K$ . Введём значение  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  и  $N = L + K - 1$ . Далее, предположим, что  $y_{ij}^* = y_{ij}$  при  $L < K$  и  $y_{ij}^* = y_{ji}$  в обратном

случае. Диагональное усреднение преобразует матрицу  $Y$  в временной ряд  $g_0, \dots, g_{N-1}$ , используя следующую формулу (30):

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^*, & \text{для } 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^*, & \text{для } L^* - 1 \leq k < K^* \\ \frac{1}{N-k} \sum_{m=1}^{N-K^*+1} y_{m,k-m+2}^*, & \text{для } K^* \leq k < N \end{cases} \quad (30)$$

Данное выражение (30) имеет смысл усредненных элементов матрицы вдоль «диагоналей»  $i + j = k + 2$ . Применяя диагонально усреднение (30) к результирующим матрицам (29), получаем новый временной ряд  $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ , из чего следует, что исходный временной ряд раскладывается в сумму временных рядов (31):

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)} \quad (31)$$

В качестве основного параметра базового алгоритма метода SSA выступает длина окна  $L$ . Выбор длины окна всегда зависит от условий решаемой задачи и предварительной информации об временном ряде, который необходимо проанализировать. В общем случае нет универсальных правил для выбора длины окна.

В данной работе было использовано лишь выделение тренда и шума из входного сигнала. При реализации программного кода метод SSA был взят из библиотеки Pyts для языка программирования Python [22]. Пример выделения шума и тренда их входного сигнала представлен на рисунке 8.

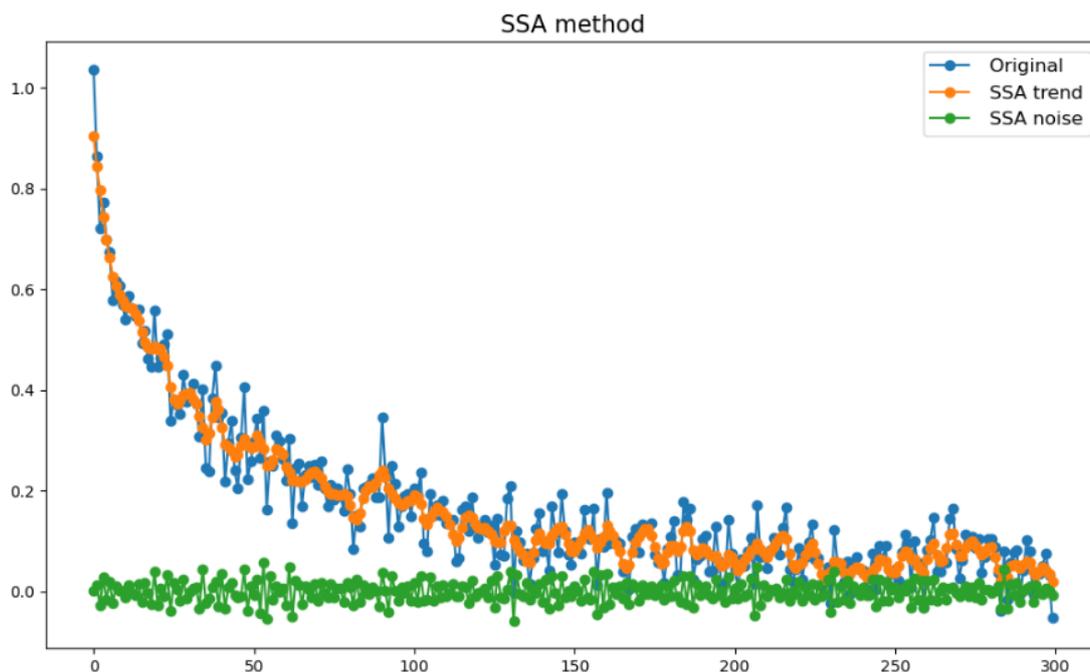


Рис.8. Выделения шума и тренда их входного сигнала.

На рисунке 8 «оранжевые» точки, соединенные тонкими линиями, в совокупности являются трендом исходного, «синего», сигнала. «Зеленые» же точки – это выделенный из исходного сигнала шум. Если по компонентам сложить «оранжевые» и «зеленые» точки, то получится исходный сигнал. Эксперименты показали, что при обучении TDP модели на тренде дебита скважин, а не на реальном значении дебита, в большинстве случаев улучшают результаты. Пример выделения тренда из дебита для одной скважины показан на рисунке 9.

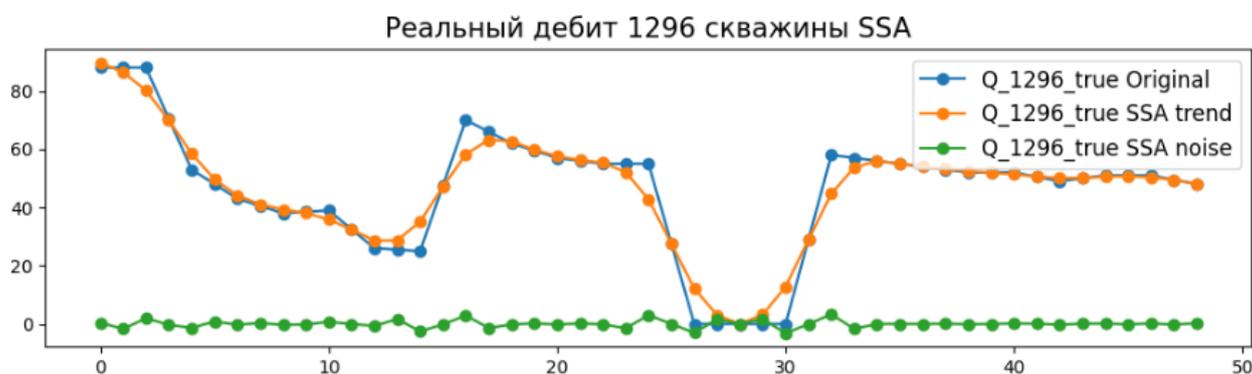


Рис.9. Выделение тренда дебита 1296й скважины с помощью SSA метода.

Процедура, описанная выше, проводилась для каждой, рассматриваемой в рамках текущей задачи, добывающей скважины.

## 2.5 Метод покоординатного спуска

Параметр  $\tau$ , который является одним из неизвестных параметров в решаемой задаче, в ходе действия градиентного метода Adam (пункт 2.3) изменяется на очень малую величину от начального приближения. Это связано с тем, что для корректного подбора всех остальных неизвестных, параметр learning rate метода Adam принимается за число порядка  $10^{-4}$  или  $10^{-5}$ . Таким образом, параметр  $\tau$  не достигает своего оптимального значения, соответствующего минимуму функции потерь (14). В связи с этой проблемой, в работе [9] было предложено использовать метод покоординатного спуска.

Его суть заключается в том, что после решения основной оптимизационной задачи, в результате которой были определены параметры  $f_{ij_{opt}}, T_{ij_{opt}}, J_{i_{opt}}, w_{i_{opt}}$ , формулируется еще одна задача оптимизации, цель которой определить значения параметра  $\tau_{i_{opt}}$ . Делается это следующим образом:

- 1) Предполагается, что формула (14) теперь зависит лишь от значений  $\tau_1, \dots, \tau_{N_{prod}}$ . Остальные же параметры принимаются за константы.
- 2) Задаётся шаг  $\Delta\tau = 1$
- 3) Вычисляется начальное значение функции потерь (14) с предположением о зависимости только от «тау».
- 4) Поочередно выбирается параметр  $\tau_i, i = 1, \dots, N_{prod}$ , для него вводится возмущение  $\tau_i = \tau_i + \Delta\tau$ , после чего вновь вычисляется значение функции потерь (14).

А) если в направлении увеличения  $\tau_i$  функция уменьшается, то делаются шаги по  $\Delta\tau$ , пока она не начнет увеличиваться.

Б) если в направлении увеличения  $\tau_i$  функция увеличивается, то делаются шаги по  $\Delta\tau$  в другую сторону.

Шаг повторяется, пока не останется направлений ни по одной координате, при которых шаг в  $\Delta\tau$  приводит к уменьшению целевой функции.

5) Задаётся шаг  $\Delta\tau = 0.1$  и повторяется пункт 4, после чего производится выход из алгоритма.

По итогам алгоритма все  $\tau$  будут определены с точностью до 0,1, что является вполне приемлемой точностью.

Немного забегаая вперед, стоит сказать, что использование данного метода снизило среднюю ошибку на 3 процента.

## 2.6 Алгоритм обучения TDP модели

В данном пункте представлен алгоритм, по которому были получены результаты из главы 5. Обучение TDP модели проводилось в несколько этапов:

- 1) Для каждой добывающей скважины ищется тренд реального дебита с использованием SSA метода. Он используется для обучения моделей.
- 2) Вторым шагом решается задача с использованием CRMP модели. Модель обучается на полученных в первом шаге входных данных. Для обучения в модель поочередно подается система входных данных, состоящая из значений приемистости, дебита и забойного давления для: первой нагнетательной скважины и всех добывающих; второй нагнетательной скважины и всех добывающих; итд... Для пояснения приведен рисунок 10. В качестве оптимизируемых параметров выступают  $f_{ij}, J_i, w_i$ .



Рис.10. Иллюстрация второго пункта алгоритма обучения TDP модели.

После того, как оптимизационная задача была решена, полученные коэффициенты используются в качестве начального приближения к третьему шагу.

- 3) Далее решается задача с использованием для TDP модели, причем модель обучается сразу же на всём пласте со всеми скважинами. Для пояснения приведен рисунок 11. В качестве оптимизируемых параметров выступают  $f_{ij}, T_{ij}, J_i, w_i$ .

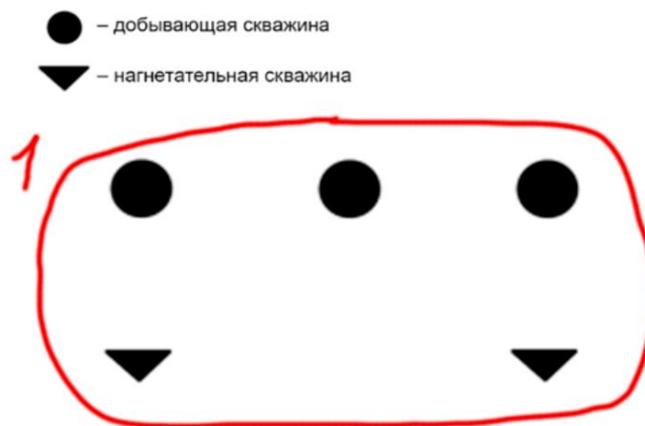


Рис.11. Иллюстрация третьего пункта алгоритма обучения TDP модели.

- 4) Четвертым шагом подбирается значение параметра  $\tau$ . Алгоритм подбора подробно описан в пункте 2.5.

В результате завершения данного алгоритма вычисляются оптимальные значения всех необходимых параметров, описанных в постановке задачи пункта 2.1.

Важным замечанием является то, что этот алгоритм применяется лишь для решения задачи с данными из реального месторождения. Результаты работы программы, использующей данный алгоритм, представлены в главе 5. Для главы 3 используется лишь третий пункт алгоритма, описанного выше.

## ГЛАВА 3. ВАЛИДАЦИЯ TDP МОДЕЛИ

### 3.1 Постановка тестовой задачи

Для проверки работоспособности TDP модели была сформулирована следующая задача:

Рассмотрим месторождение, которое состоит из 2 нагнетательных и 2 добывающих скважин. Приемистость первой нагнетательной скважины в один момент резко увеличивается, приемистость второй же скважины остается без изменений. Из-за этого дебит первой добывающей скважины резко увеличивается, а дебит второй скважины в этот же момент уменьшается. Смоделированные условия можно увидеть на рисунке 12 и рисунке 13.

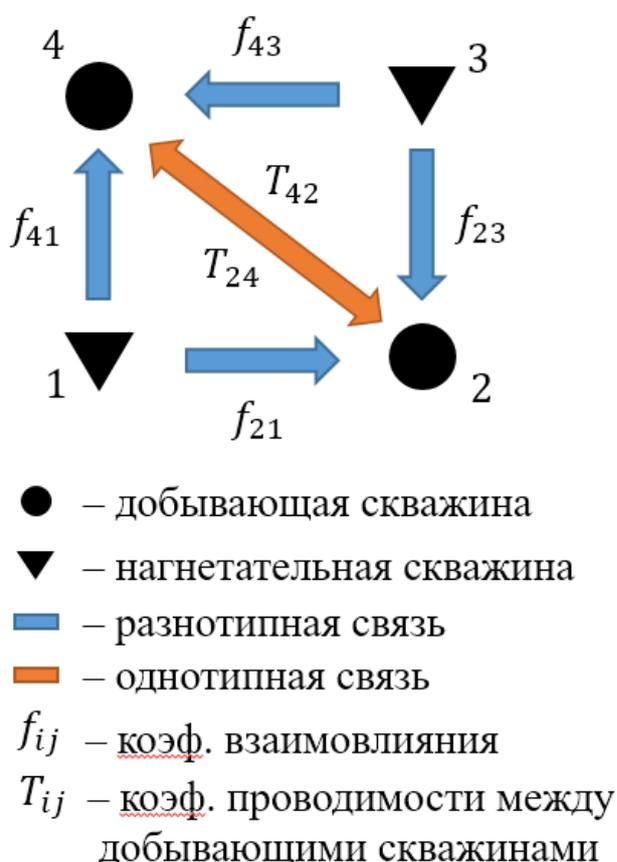


Рис.12. Примерная схема искусственного месторождения.

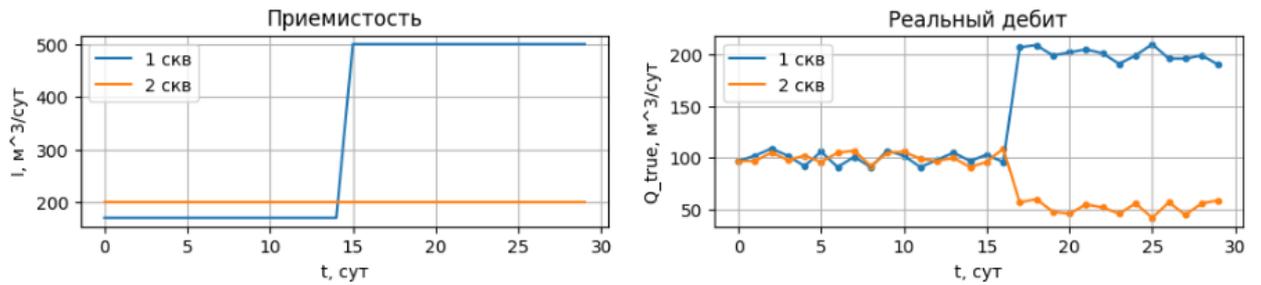


Рис.13. Изменение приемистости и дебита рассматриваемых скважин.

Для упрощения задачи было принято, что упругая энергия пласта равно нулю. Вследствие этого появилась возможность сделать ряд допущений в формуле (12):

$$\tau \rightarrow 0; \quad e^{-\frac{\Delta t_k}{\tau_i}} \underset{\tau_i \rightarrow 0}{\rightarrow} 1; \quad \left(1 - e^{-\frac{\Delta t_k}{\tau_i}}\right) \underset{\tau_i \rightarrow 0}{\rightarrow} 0;$$

Следовательно, формула (12) примет вид (32):

$$q_i(t_n) = \sum_{x=1}^{N_{inj}} [f_{xi} I_x^{(n)}] + \sum_{x=1}^{N_{prod}} [T_{ix} (p_{wx}^{(n)} - p_{wi}^{(n)})] \quad (32)$$

Формула (32) является упрощенной версией приближения дебита добывающей скважины для TDP модели, используемой для её валидации.

### 3.2 Результаты для тестовой задачи

На рисунке 14 приведены результаты, на которых отображена полученная динамика дебита жидкости, вычисленная согласно TDP модели, для тестовой, упрощенной, задачи с нулевой энергией пласта.

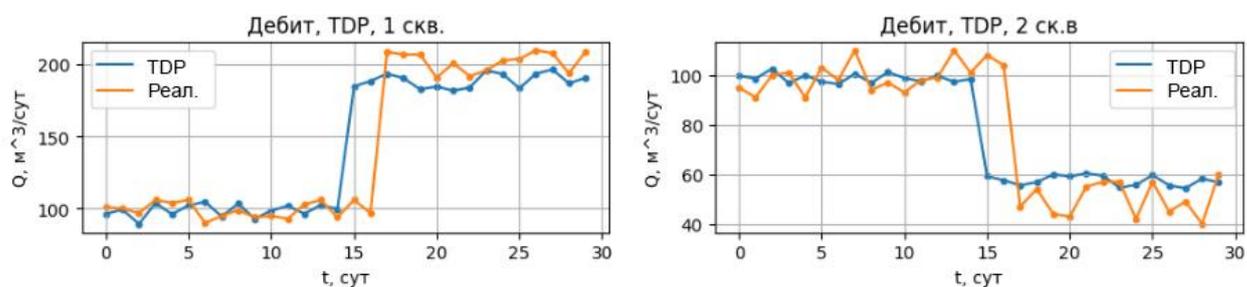


Рис.14. Дебит добывающих скважин в тестовой задаче.

Анализируя рисунок 14, можно увидеть, что TDP модель корректно обрабатывает и вычисляет коэффициенты взаимовлияния скважин. Модель абсолютно верно предсказала поведение добывающих скважин в зависимости от приемистости нагнетательных скважин с учётом приведенных в пункте 3.1 предположений. Несмотря на то, что результаты далеки от идеальных, суть решения тестовой задачи заключалась в том, чтобы доказать способность TDP модели корректно обрабатывать и вычислять коэффициенты гидродинамической связи не только разнотипных, но и однотипных скважин. Коэффициенты вычислены и соответствуют заданному поведению скважин. Это значит, тестовую задачу можно считать успешно решенной, а её результаты можно считать корректными.

## ГЛАВА 4. ОБРАБОТКА ИСХОДНЫХ ДАННЫХ ДЛЯ ОБУЧЕНИЯ

### 4.1 Проблемы реальных данных.

Как было сказано в пункте 2.1, в качестве входных данных для обучения выступали: приемистость нагнетательных скважин, дебит и забойное давление добывающих скважин. Однако, для решения некоторых проблем, рассмотренных в текущем и следующем пункте данной работы, были использованы дополнительные данные, о которых пойдет речь в пункте 4.2. Работа происходила с обезличенными данными по реальному нефтяному месторождению. Как многим известно, такого рода информация имеет достаточное количество проблем, в частности связанных с пропусками значений в таблицах.

Промысловые данные о дебитах и забойных давлениях скважин могут быть очень неточными. В таких случаях отклонения полученного по модели дебита жидкости от фактического дебита будут принимать большие значения. Так же, в ходе разработки месторождения скважины могут останавливаться по различным причинам. Классические CRM модели и, рассмотренная в данной работе, TDP модель не учитывают остановки скважин. Существует OmegaCRM модель, описанная в работе [23], в которой учтены остановки скважин, но для TDP модели это всё является проблемой, мешающей адекватно применять модель на реальных данных.

Для демонстрации изначального качества имеющихся данных приведена матрица пропусков данных на рисунке 15.

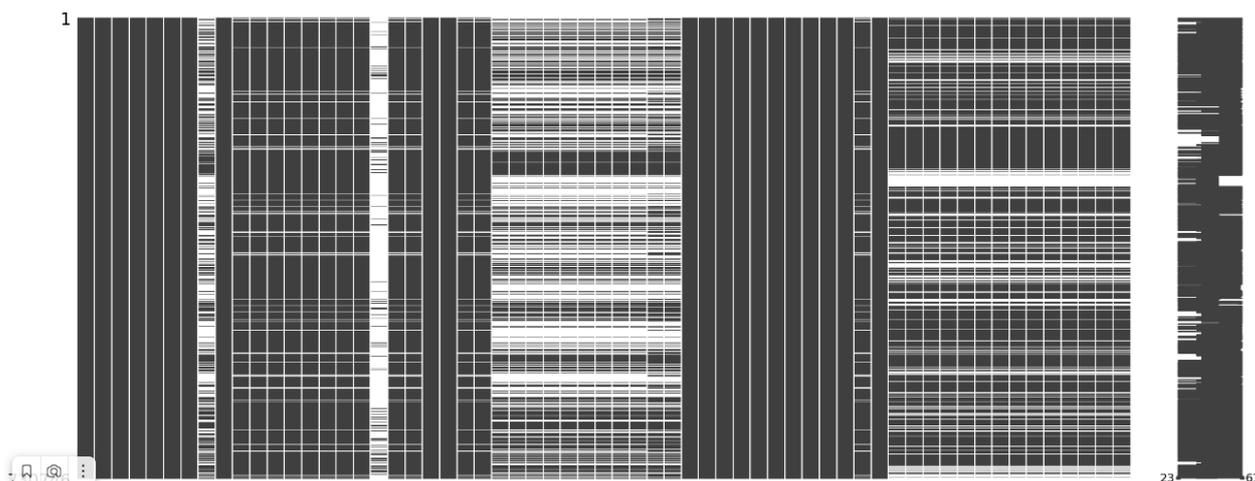


Рис.15. Матрица пропусков начальных данных.

Как видно из рисунка 15, действительно, исходные данные имеют огромное количество пропусков. Однако, проблемы эти можно частично решить и довести качество данных до уровня, который будет являться приемлемым для обучения модели и получения адекватных результатов. Все действия, совершенные для исправления вышеперечисленных проблем, описаны в пункте 4.2.

#### 4.2 Заполнение пропусков.

В данном пункте подробно описаны все действия, которые были совершены для улучшения качества исходных данных.

Изначально данные хранились в двух excel таблицах, имеющих огромное количество ненужных столбцов. Файл с данным для добывающих скважин будем называть «oil-файл», а для нагнетательных «ppd-файл». Так же, следует обратить внимание на то, что работа происходила с данными, представленными на период с 01.08.2014 по 01.08.2022.

Первым шагом из общего «oil-файла» были выделены только необходимые столбцы: дата, название месторождения, название пласта, номер скважины, забойное давление на кровлю, дебит жидкости в м<sup>3</sup>/сут, причина простоя, ZAB\_X, ZAB\_Y (ZAB\_X – координата забоя скважина по оси X, ZAB\_Y – координата забоя скважина по оси Y). Полученный набор данных



давления не рассматривался в данной работе. Таким образом, набор данных был сужен до временного интервала с 01.08.2019 по 01.08.2022. Так же, скважины, которые на протяжении всего временного интервала являлись остановленными, были удалены из датасета. Полученный новый набор данных назовем как «третий датасет». Имеющаяся выборка данных в последствии делится на две части: первые 24 месяца – это обучающая выборка, последние 12 месяцев – это выборка для попытки прогнозирования дебита на основе полученных в ходе решения оптимизационной задачи коэффициентов.

Для заполнения пропусков в столбце забойного давления «третьего датасета» была использована Регрессия ближайших соседей (KNeighborsRegressor) из библиотеки scikit-learn для языка программирования Python [23]. В качестве данных для обучения регрессора ближайших соседей использовались значения координат забоя добывающих скважин по осям X и Y. После обучения регрессора ближайших соседей был получен предикт забойного давления на все недостающие даты.

После всех вышеперечисленных действий по заполнению пропусков были найдены месторождения и пласты, имеющиеся в обоих исходных файлах. Произвольным образом было выбрано одно месторождение и один пласт, на котором производилось обучение моделей. Матрицы пропусков итоговых датасетов, готовых для обучения, представлены на рисунках 17 и 18.

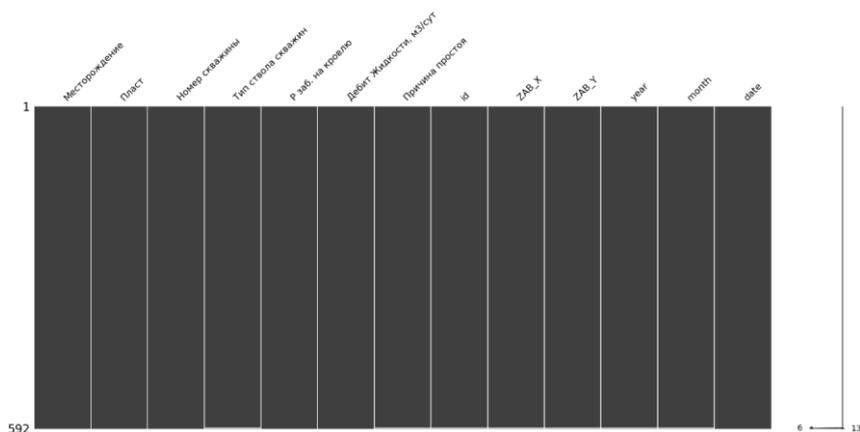


Рис 17. Матрица пропусков итогового датасета с добывающими скважинами.

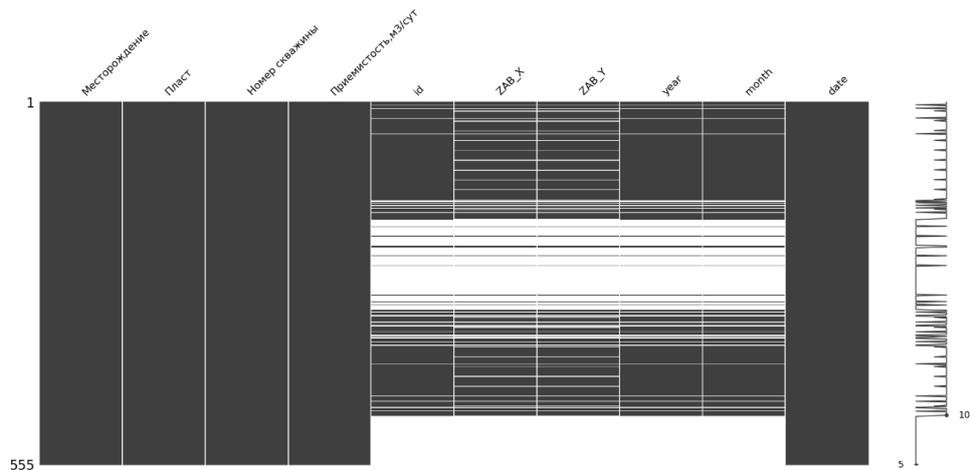


Рис 18 Матрица пропусков итогового датасета с нагнетательными скважинами

Последним шагом полученные два датасета были интерполированы так, чтобы в каждый месяц имелось не одно значение, а два. Т.е, условно говоря, в 8й месяц 2019го года имеется значение приемистости, дебита и забойного давления не только в 1е число месяца, а еще и в 15е.

# ГЛАВА 5. РЕЗУЛЬТАТЫ

## 5.1 Результаты для основной задачи

Для получения результатов был выбран пласт, в котором присутствует 10 добывающих и 15 нагнетательных скважин. В этом пласте был проведён расчет коэффициентов взаимного влияния и всех остальных неизвестных параметров согласно пункту 2.5.

На рисунках 19, 20, 21, 22, 23, 24 приведены результаты для основной задачи. На графиках представлено приближение дебита во временном промежутке размером в 2 года, на котором обучалась модель. Данный промежуток отмечен кривой маленькими точками. Кривая с большими точками – это прогноз дебита жидкости на основе полученных в ходе решения задачи коэффициентах. В данном документе приведена лишь часть результатов.

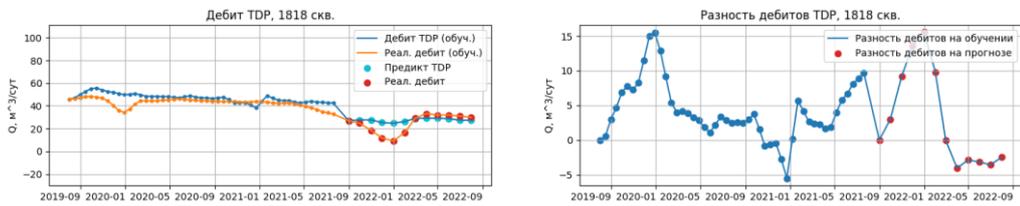


Рис.19. Дебит 1818й добывающей скважины.

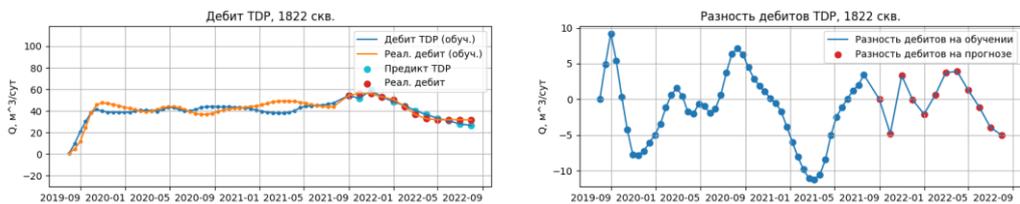


Рис.20. Дебит 1822й добывающей скважины.

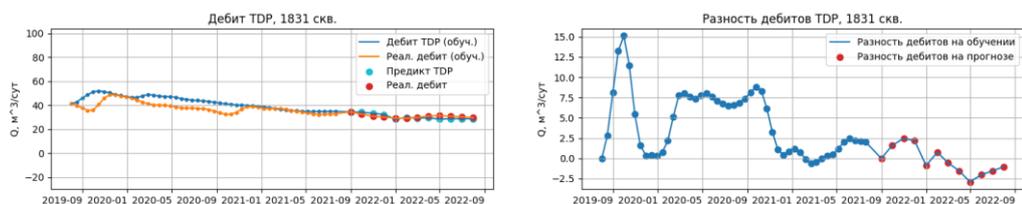


Рис.21. Дебит 1831й добывающей скважины.

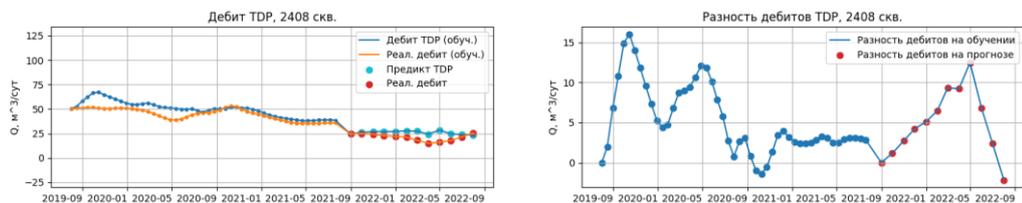


Рис.22. Дебит 2408й добывающей скважины.

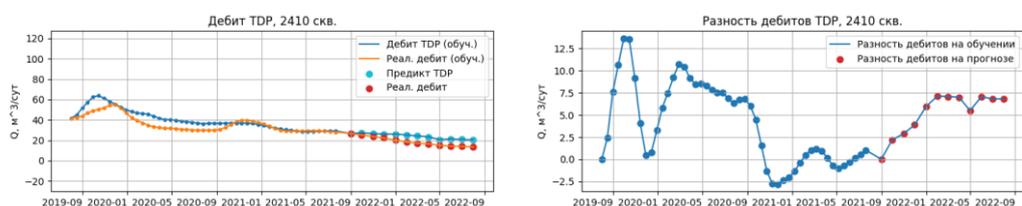


Рис.23. Дебит 2410й добывающей скважины.

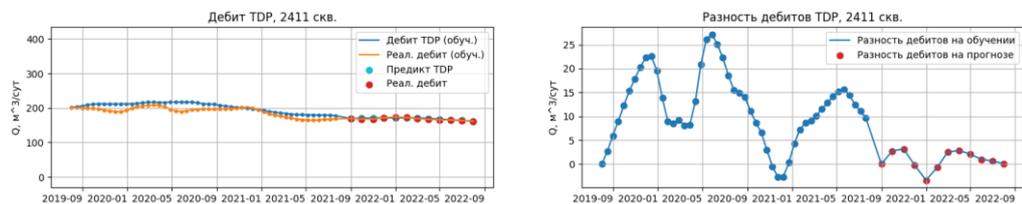


Рис.24. Дебит 2411й добывающей скважины.

Как можно заметить, модель отрабатывает в разы лучше, в сравнении с тестовой задачей. «Синяя» кривая, соответствующая приближению дебита, посчитанному согласно TDP модели, во многих местах совпадает с «оранжевой» кривой, которая описывает реальную ситуацию динамики изменения дебита добывающих скважин. Так же в правой части рисунков представлен график разности реального дебита и его вычисленного приближения. По этому графику можно оценить величину отклонения вычисленного дебита от реального в каждой точке временного ряда.

Далее, для сравнения работы TDP и CRMP модели, приведена таблица, в которой посчитана средняя абсолютная процентная ошибка (MAPE) для каждой скважины. Зеленым цветом выделен лучший результат для каждой скважины.

Таблица 1. Сравнение CRMP и TDP модели на обучающей выборке

Обучающая выборка, MAPE		
скважина	CRMP	TDP
1296	45.75 %	38.92 %
1817	8.09 %	8.47 %
1818	9.76 %	9.67 %
1819	16.56 %	14.12 %
1822	21.36 %	11.15 %
1831	10.72 %	9.81 %
2408	9.49 %	10.20 %
2410	12.27 %	10.65 %
2411	6.08%	5.94 %
2412	21.25 %	15.56 %

Из таблицы 1 отчетливо видно, что в период обучения TDP модель ведёт себя гораздо лучше. Среднее значение ошибки для CRMP модели равняется 16.13%, а для TDP модели 13.42%.

Далее приведём таблицу 2, в которой будет представлено сравнение моделей по MAPE во время прогнозирования.

Таблица 2. Сравнение CRMP и TDP модели на прогнозе

Предикт, MAPE		
скважина	CRMP	TDP
1296	16.86 %	15.21 %
1817	19.91 %	20.92 %
1818	22.59 %	21.20 %
1819	46.24 %	44.22 %
1822	14.30 %	6.77 %
1831	5.18 %	4.80 %
2408	20.32 %	19.63 %
2410	25.33 %	22.43 %
2411	0.93 %	0.94 %
2412	30.39 %	24.83 %

Как и в предыдущей таблице, CRMP модель оказалась лучше всего лишь в двух из десяти случаев. Среднее значение ошибки для CRMP модели равняется 20.21%, а для TDP модели 18.09%.

Таким образом, из полученных таблиц и графиков видно, что для данной задачи при одинаковых условиях TDP модель позволяет гораздо точнее и корректнее вычислять коэффициенты взаимного влияния скважин.

## ЗАКЛЮЧЕНИЕ

В данной работе была рассмотрена и использована CRMP модель и ее модификация – TDP модель для оценки гидродинамической связи однотипных скважин, а также предложена модификация TDP модели для учёта влияния на динамику дебита жидкости притока с водоносного горизонта. Далее, был написан программный код на языке Python для обработки исходных данных, обучения TDP и CRMP моделей и анализа результатов. Так же была предложена модификация метода штрафных функций (придумана собственная штрафная функция), был использован метод сингулярного спектрального анализа и метод покоординатного спуска для улучшения сходимости результатов.

На основании проделанной работы можно сделать вывод о том, что TDP модель позволяет оценивать гидродинамическую связь как однотипных, так и разнотипных скважин. Поскольку приемлемое отклонение предсказанного дебита от реального не должно превышать 10%, можно сделать вывод о том, что полученные результаты в большинстве своем устраивают нас лишь на обучающей выборке. Однако, в пункте 1.3 уже было сравнение работы TDP модели с учетом и без учета аквифера. Сравнение показало, что модификация в разы улучшила сходимость и уменьшила среднюю абсолютную ошибку в процентах. Это значит, что имеется возможность улучшать TDP модель, тем самым приводя её к более корректному и стабильному режиму работы в будущем.

Поскольку TDP модель является новой и не до конца изученной, необходимо еще провести большое число исследований и модификаций для улучшения результатов. Так же стоит заметить, что для решенной задачи, в которой имеется 10 добывающих и 15 нагнетательных скважин, количество неизвестных параметров равняется числу 280. Из-за этого функция, минимум которой ищется в ходе решения, имеет огромное число локальных минимумов. Помимо этой проблемы, результаты очень сильно зависят от

заданного начального приближения, так что есть шанс просто «не угадать» точку, с которой будет производиться расчет. Таким образом, для достижения устойчивых и интерпретируемых результатов задачи обучения модели на данных реальных месторождений необходимо модифицировать как саму модель, так и метод обучения. Так же еще одна модификация, над которой в обязательном порядке необходимо работать – это учет остановок скважин.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Вержбицкий В.М., Основы численных методов: учеб. для вузов по спец. "Прикладная математика" / В.М. Вержбицкий Москва: Высшая школа, 2002. – 306 с
2. Данилова Д.Л., Главные компоненты временных рядов. [Текст] / Под редакцией Д.Л. Данилова и А.А. Жиглявского // СПб: Пресском, 2007. – 308 с.
3. Крец В.Г., Основы нефтегазового дела / В.Г. Крец, А.В. Шадрина – Издательство Томского политехнического университета 2011
4. Попова Т.М., Методы безусловной оптимизации / Т.М. Попова – Хабаровск, Издательство ТОГУ, 2013
5. Поршнева С.В., Исследование особенностей применения метода сингулярного спектрального анализа в задаче анализа и прогнозирования временных рядов / С.В. Поршнева, Ф. Рабайя – Ульяновск 2016. – 21-24 с
6. Степанов С.В., Сопровождение разработки нефтяных месторождений с использованием моделей CRM / С.В. Степанов [и др.]. – Тюмень 2021.
7. Sayarpour Morteza, Development and Application of Capacitance-Resistive Models to Water/CO<sub>2</sub> Floods / Morteza Sayarpour [и др.]. – The University of Texas at Austin August 2008.
8. Афанаскин И.В., Использование CRM-моделей интерференции скважин для оценки фильтрационно-емкостных свойств пласта по данным разработки / П. В. Крыганов, А. А. Глушаков, П. В. Ялов // Федеральное государственное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», 2020. – 1-5 с
9. Бекман А.Д., Новый алгоритм нахождения оптимального решения задачи определения коэффициентов взаимовлияния скважин в рамках модели CRM / Сергей Викторович Степанов, Александр Альфредович

- Ручкин, Дмитрий Валерьевич Зеленин // Вестник Тюменского государственного университета, 2019. – 166-178 с.
10. Оленчиков Дмитрий, Применение CRM-подобных моделей для экспресс прогнозирования и оптимизации разработки месторождений / Дмитрий Посвянский, Рохар // Авторское право 2019 г. Общество инженеров нефтегазовой промышленности 2019. – 1-7 с.
  11. Сидельников К.А., Моделирование на базе методов трубок и линий тока / В.Е. Лялин, И.М. Григорьев // Вестник Удмуртского университета 2012 – 109 с.
  12. Хатмуллин И.Ф., Полуаналитические модели расчета интерференции скважин на базе класса моделей CRM/ И.Ф. Хатмуллин, А.П. Цанда и др // Нефтяной хозяйство 2018. – 1-4 с.
  13. Цанда Артём, «Оценка гидродинамической связности скважин добывающего и нагнетательного фондов зрелых месторождений на базе класса моделей CRM», 2019 г. – 1-54 с
  14. Шпетный Д.Н., «Применение машинного обучения для повышения точности прогноза добычи жидкости и нефти моделью "CRM"», 2022 г. – 1-43 с
  15. Emre Artun, Characterizing Reservoir Connectivity and Forecasting Waterflood Performance Using Data-Driven and Reduced-Physics Models / Middle East Technical University - Northern Cyprus Campus // Авторское право 2019 г. Общество инженеров нефтегазовой промышленности , Alaska, USA, 23–26 May 2016. – 26-27 с.
  16. Izgec O., Kabir C.S.. Quantifying Nonuniform Aquifer Strength at Individual Wells // SPE Reservoir Evaluation & Engineering -2010
  17. Kingma P. Diederik, Adam: a method for stochastic optimization / University of Amsterdam, OpenAI // Published as a conference paper at ICLR 2015
  18. Академия Яндекс, учебник по машинному обучения – <https://academy.yandex.ru/handbook/ml/article/optimizaciya-v-ml> (дата обращения: 18.02.2023)

19. Документация библиотеки Tensorflow. -  
[https://www.tensorflow.org/api\\_docs](https://www.tensorflow.org/api_docs). (дата обращения: 10.02.2023)
20. Профессиональный информационно-аналитический ресурс,  
посвященный машинному обучению. -  
[http://www.machinelearning.ru/wiki/index.php?title=Метод\\_штрафных\\_функций](http://www.machinelearning.ru/wiki/index.php?title=Метод_штрафных_функций). (дата обращения: 11.02.2023)
21. Профессиональный информационно-аналитический ресурс,  
посвященный машинному обучению. -  
[http://www.machinelearning.ru/wiki/index.php?title=Прогнозирование\\_временных\\_рядов\\_методом\\_SSA\\_%28пример%29](http://www.machinelearning.ru/wiki/index.php?title=Прогнозирование_временных_рядов_методом_SSA_%28пример%29) (дата обращения: 07.04.2023)
22. Документация библиотеки Pyts. -  
[https://pyts.readthedocs.io/en/stable/auto\\_examples/decomposition/plot\\_ssa.html](https://pyts.readthedocs.io/en/stable/auto_examples/decomposition/plot_ssa.html) (дата обращения: 08.04.2023)
23. Документация библиотеки scikit-learn. – <https://scikit-learn.ru/1-6-nearest-neighbors> (дата обращения: 27.05.2023)