«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Центр непрерывного образования

Факультета компьютерных наук

ИТОГОВЫЙ ПРОЕКТ

МОДЕЛИРОВАНИЕ ДЕБИТА ГАЗОВЫХ СКВАЖИН С ПРИМЕНЕНИЕМ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ ДАННЫХ ДОБЫЧИ В СЕВЕРНОМ МОРЕ

Название темы

Выполнил:

Борисенко Алексей Александрович

Ф.И.О.

Руководитель:

Кантонистова Елена Олеговна

Ф.И.О.

Москва 2021

Оглавление

I. Введение

Нефтегазовая отрасль с самого начала своего существования сталкивается с задачами автоматизации и оптимизации. Новейшие технологии и применение алгоритмов на основе данных дали существенный толчок в развитии отрасли в последние годы. В наши дни методы машинного обучения применяются в бурении (Noshi, 2019, Marana и др., 2010), геологии (Tandon, 2019, Anifowose и др., 2013), петрофизике (Vallabhaneni и др., 2019), добыче (Pennel и др., 2018) и заканчивании скважин (Castiniera и др., 2018, Dan Fu, 2019).

Современные вычислительные мощности также делают более доступными применение нейросетей и методов глубокого обучения для решения проблем отрасли. Многочисленные исследования показывают различные техники и подходы к решению задач по анализу данных добычи нефти и газа (Rastogi и др., 2019), интерпретации сейсмических данных (Lowell и Paton, 2018) и многих других.

В общем случае применение методов машинного обучения в отрасли подразделяется на две категории (Mohaghegh, 2019). Первая категория относится к показателям добычи нефти, дебитам нефтяных и газовых скважин. Вторая категория оперирует большим набором характеристик залегающих пластов, получаемых как при проведении обработок, так и при исследованиях скважин с применением различного инструментария. Данная работа основана на первой категории и представляет моделирование дебита газовых скважин с применением машинного обучения на основе открытых данных о добыче газовых скважин месторождения Viking (Англия), разрабатываемых компанией ConocoPhilips в Северном море.

В области нефтегазового инжиниринга одна из основных задач специалиста по оценке пласта – точный расчет потенциала работы

продуктивного коллектора, его способности отдавать содержащиеся в нем углеводороды. Данный расчет, как правило, выполняется с помощью анализа КПД (кривых падения добычи), которые характеризуют зависимости текущего отбора газа от фактора времени. Иными словами, кривые падения характеризуют изменение добычи газа во времени. В результате анализа строится прогноз добычи на будущие периоды, что в дальнейшем используется при оценке запасов углеводородов в государственных органах по рациональному недропользованию.

С ростом количества внедряемых решений на основе машинного обучения в отрасли возникла необходимость применения имеющихся инструментов и алгоритмов для совершенствования анализа КПД. Одной из причин применения машинного обучения является рост количества данных, получаемых с месторождений и помогающих строить модели продуктивных пластов, а также проводить сопоставление данных моделирования с историческими данными. Традиционные методы анализа КПД энергоемки и ресурсозатратны, что делает новые подходы актуальными.

II. Обзор литературы

Анализ данных по кривой падения добычи является традиционным отраслевым инструментом для определения производительности продуктивного пласта. Основные методы анализа используют эмпирические модели для прогноза добычи (Arps, 1945, Ilk и др., 2008, Duong, 2010). Многочисленные исторические данные по добыче выделяют несколько отраслевых методов, среди которых метод Арпса, применяемый при возможном отсутствии параметров пласта и без непосредственного определения коэффициента извлечения углеводородов, является наиболее универсальным и широко применяемым инструментом.

В современной литературе представлены работы, описывающие как применение некоторых методов машинного обучения (нейронные сети и глубокое обучение) помогают совершенствовать прогнозные расчеты в сравнении с традиционными методиками (Q.Cao и др., 2016, D.Han и др., 2019, Y.Li и Y.Han, 2017). Алгоритм случайного леса, метод опорных векторов (SVM), многомерные адаптивные регрессионные сплайны (MARS) также во многом способствовали совершенствованию качества прогнозных моделей (Vyas, 2017). Часть исследователей показывают лучшие результаты с применением нейронных сетей, в частности для предсказания дебита нефтяных скважин сланцевых месторождений (Suhag, 2017).

В данной работе показано применение алгоритмов SARIMA (Fulton, 2017), Facebook Prophet (Taylor и Letham, 2017) и XGBoost (Chen и др., 2016). Однако, в целом данное исследование не ограничивается вышеуказанными инструментами, и для дальнейшего развития вопроса могут применяться рекуррентные нейронные сети, в том числе LSTM (Sun и др., 2018), зарекомендовавшие себя со стороны высокой точности прогнозирования.

III. Методы обнаружения аномалий и оценка качества алгоритмов

Использование корректных метрик в проектах, касающихся науки о данных, имеет решающее значение. Неправильная метрика влияет как на оптимизацию модели (через функцию потерь), так и на общее представление о модели. В наши дни все чаще публикуются все новые показатели ошибок, однако классические метрики, такие как MAE (средняя квадратичная ошибка), не уходят из поля зрения исследователей данных и используются в большинстве случаев, в том числе при оценке качества моделей прогноза временных рядов.

Однако, стоит отметить, что универсальной метрики не существует, и каждая объединяет большое количество данных в одно значение, поэтому она обеспечивает только одну проекцию ошибок модели, подчеркивая определенный аспект характеристик ошибок производительности модели (T.Chai и R.Draxler, 2014).

Наиболее распространенные показатели ошибок показаны в пяти категориях (R.Hyndman и A.Koehler, 2006):

1) Scale-dependent measures (MAE, MSE, RMSE)
2) Measures based on percentage errors (MAPE, SMAPE)
3) Measures based on relative errors (MRAE, GMRAE)
4) Relative measures (RelMAE)
5) Scaled errors (MASE)

Конечно, будут ситуации, когда некоторые из существующих мер все еще могут быть предпочтительнее. Например, если все серии находятся в одном масштабе, то для оценки качества достаточно метрики MAE. Если все данные положительные и намного больше нуля, то предпочтительно использование MAPE по соображениям простоты. Для оценки качества алгоритмов, описанных в данной работе, воспользуемся обеими метриками – MAE и MAPE.

Ввиду наличия сезонности временного ряда данных дебита газовой скважины с резким снижением значения дебита до 0 на период отключения скважины, а также скачкообразным поведением дебита в ходе эксплуатации скважины, воспользуемся методикой оценки качества работы алгоритмов, описанной в статье о прогнозировании стоимости электричества в Сингапуре (L.Jiang и G.Hu, 2018), с построением диаграмм размаха рассчитанных значений ошибки по 4 сезонам с еженедельной периодичностью. В данной работе диаграммы размаха ошибок MAE и MAPE построены по годам с периодичностью

1 месяц (12 значений в год). Для оценки качества модели как на обучающей, так и на тестовой выборке, взято среднее медианных значений ошибок MAE и MAPE согласно построенных диаграмм размаха за период.

IV. Эксперименты

Для проведения экспериментальной части работы использованы открытые данные по ежедневной добыче нефтегазовых месторождений Великобритании, опубликованные компанией Oil and Gas Authority (**Рисунок 1**).

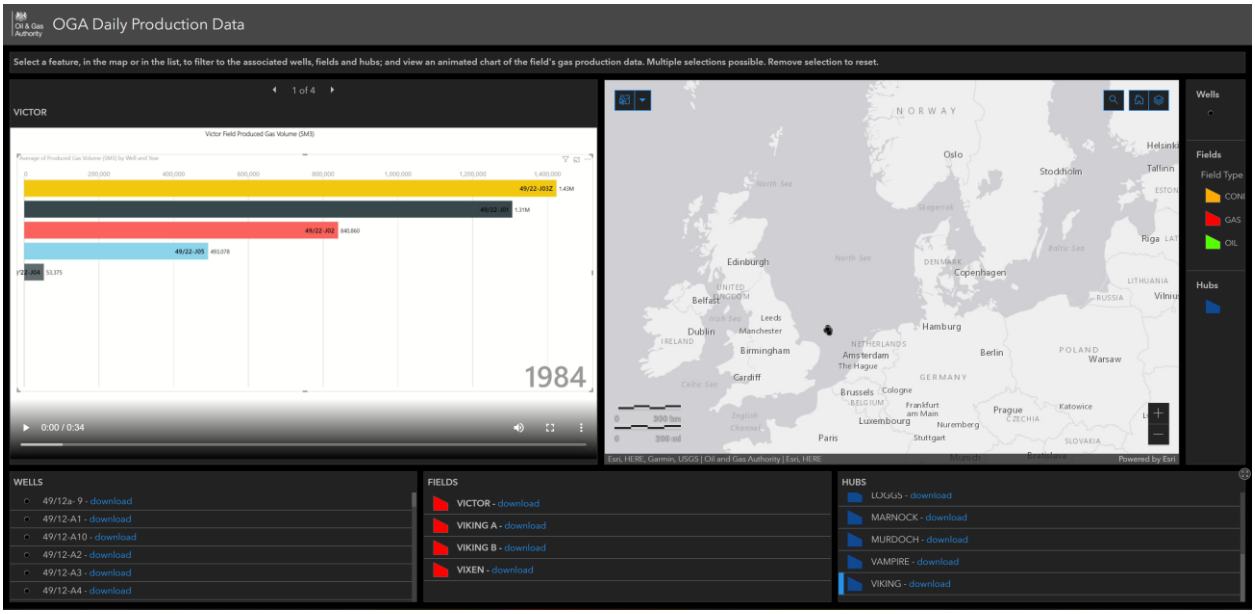

Рисунок 1—Портал Oil and Gas Authority.

Для построения временного ряда и проведения необходимых прогнозных расчетов загружены данные по ежедневной добыче скважины 49/12a-K01 газового месторождения Viking, расположенного в южной части Северного моря в 138 км к востоку от крупного газового терминала Тедлторп (Англия) в квадранте 49 (**Рисунок 2**). Период добычи газа, согласно представленных данных – с 1998 по 2015 год (**Рисунок 3**).

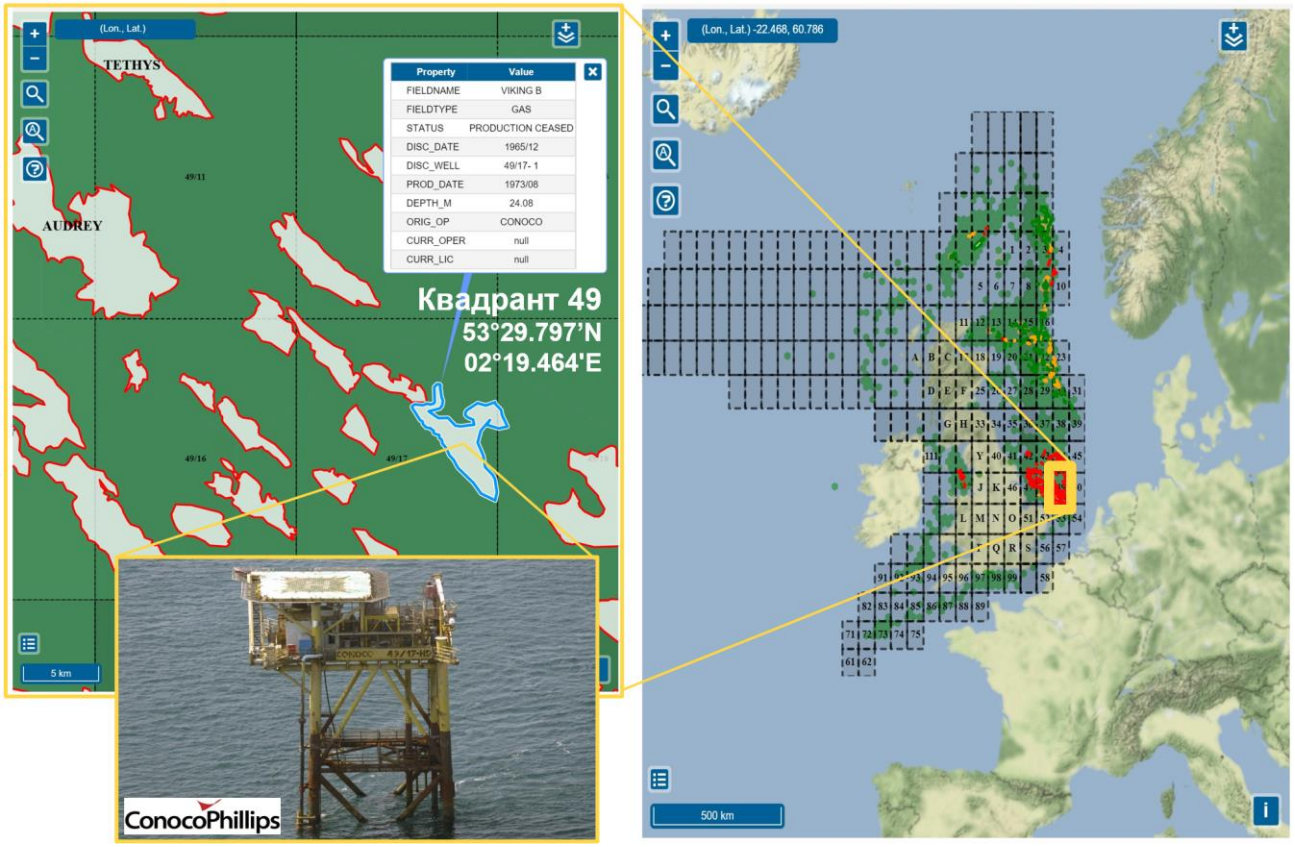

Рисунок 2—Объект исследования: скважина 49/12a-K01 месторождения Viking.

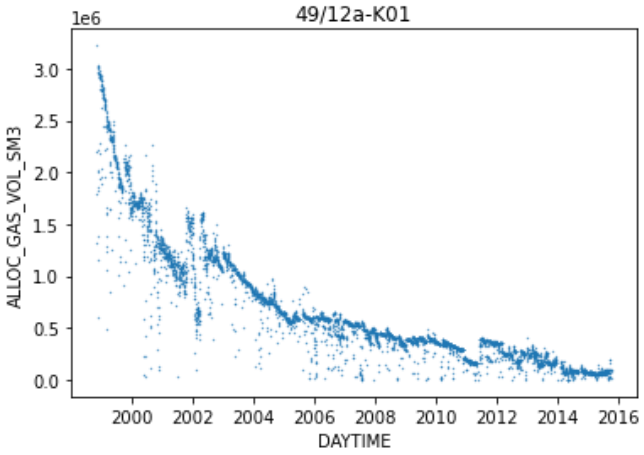

Рисунок 3—Кривая падения добычи на скважине 49/12a-K01 (ежедневная периодичность).

Из графика видно, что наш ряд имеет небольшое кол-во выбросов, которые влияют на разброс. Кроме того, анализировать дебит газа за каждый день не совсем верно, т.к., например, имеют место периодические остановки добычи для проведения различных технологических операций на скважине. Поэтому есть смысл перейти к недельному интервалу (**Рисунок 4**) и среднему значению дебита на нем, это избавит нас от выбросов и уменьшит колебания нашего ряда.

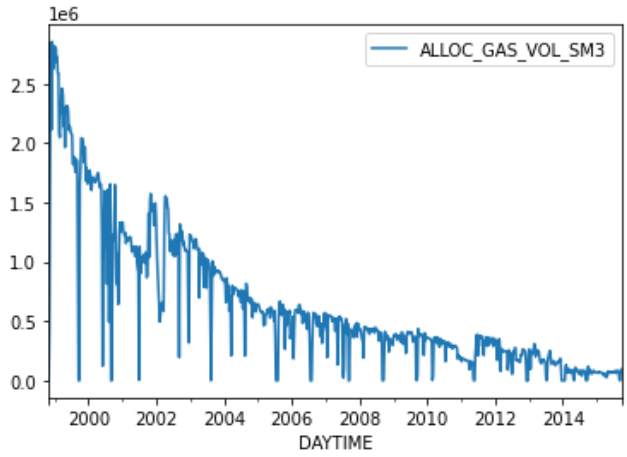

Рисунок 4—Кривая падения добычи на скважине 49/12а-К01 (еженедельная периодичность).

Разделим датасет на обучающую и тестовую выборки. *Для тестовой выборки возьмем период – 2 года.*

1) Модель SARIMA

При близком рассмотрении модели SARIMA (Seasonal Autoregressive Integrated Moving Average) можно выделить три компонента, составляющих данный акроним (Fulton, 2017):

1. AR: Autoregression - характеризует зависимость наблюдаемого значения от отстающих (предшествующих) значений

2. I: Integrated - характеризует скорость изменения наблюдаемых значений путем вычитания текущего наблюдаемого значения из значения на предшествующем шаге, чтобы сделать временной ряд стационарным
3. MA: Moving Average - характеризует зависимость между наблюдением и остаточной ошибкой из модели скользящего среднего, применяемой к предшествующим наблюдениям.

Общий вид модели SARIMA (с учетом сезонности):

$$SARIMA(p,d,q)(P,D,Q)_s$$

где p — число отстающих значений ряда, включенных в модель (порядок авторегрессии)

d — количество раз, когда брали разность наблюдаемых и отстающих значений (степень разности)

q — число отстающих значений, которые прибавляются или вычитаются из наблюдаемого значения (порядок скользящего среднего)

P — порядок сезонной авторегрессии

D — степень сезонной разности

Q — порядок сезонного скользящего среднего

s — сезонность (в нашем случае при ежегодной сезонности с недельным интервалом это значение составит 52)

В статистике временной ряд описывается стохастическим процессом. Это случайная величина, у которой со временем меняется её распределение. У этой величины есть среднее и дисперсия, которые тоже меняются. Стохастический процесс является стационарным (англ. stationary stochastic process), если его

распределение со временем не меняется. Например, к такому процессу относятся периодические колебания значений.

Для построения прогнозной модели SARIMA проверим временной ряд на стационарность. Разложим временной ряд на три составляющие: тренд, сезонность и остаток (англ. residuals). Остаток - это компонента, которая не объясняется трендом и сезонностью, это шум (**Рисунок 5**).

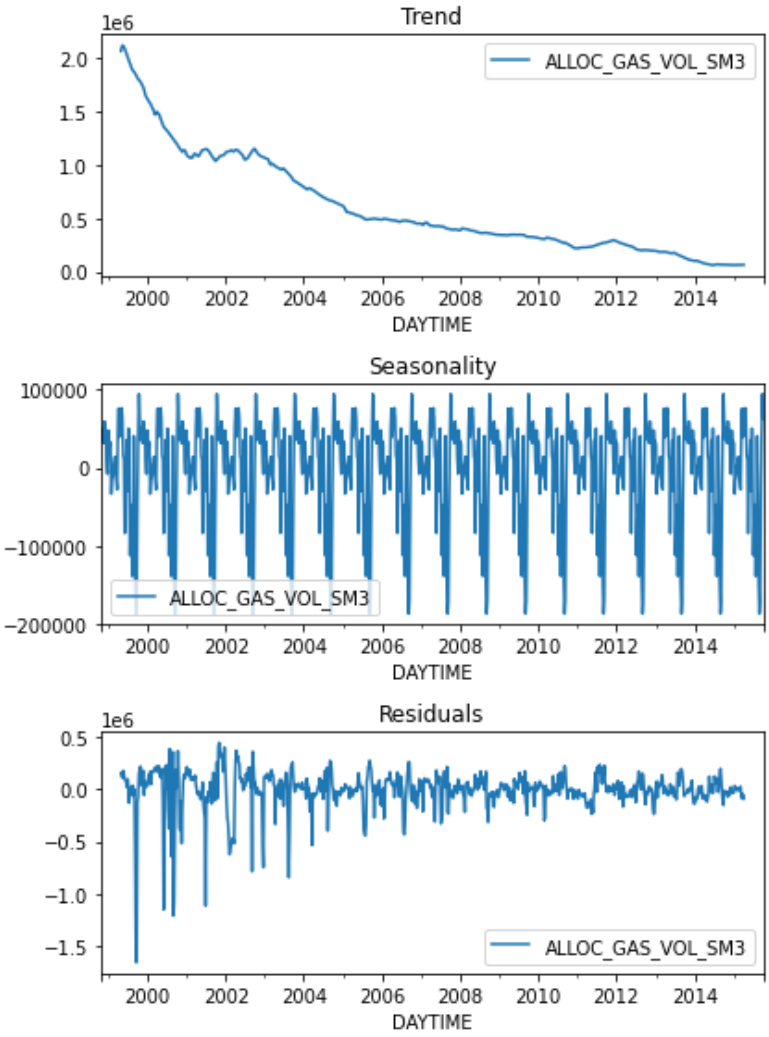

Рисунок 5—Разложение временного ряда.

Временной ряд имеет тренд, характеризующий падение добычи газа со временем вследствие падения пластового давления по мере отборов газа из

продуктивного пласта. Также выявлена сезонность с периодом 1 раз в год, что наиболее вероятно связано с периодическими ежегодными плановыми работами на скважине (ремонт, замена оборудования и пр.). Остаток показывает нерегулярную (не описываемую трендом или сезонностью) составляющую исходного ряда в начальный период работы скважины, что может объясняться подбором оптимальных параметров работы скважины в системе разработки месторождения.

Как можно заметить данные дебита газа имеют тренд. Это означает, что временной ряд - не стационарный. При этом меняется как среднее, так и стандартное отклонение. Прогнозировать при этом сложнее, т.к. свойства ряда меняются слишком быстро. Приведем ряд к стационарному виду (**Рисунок 6**).

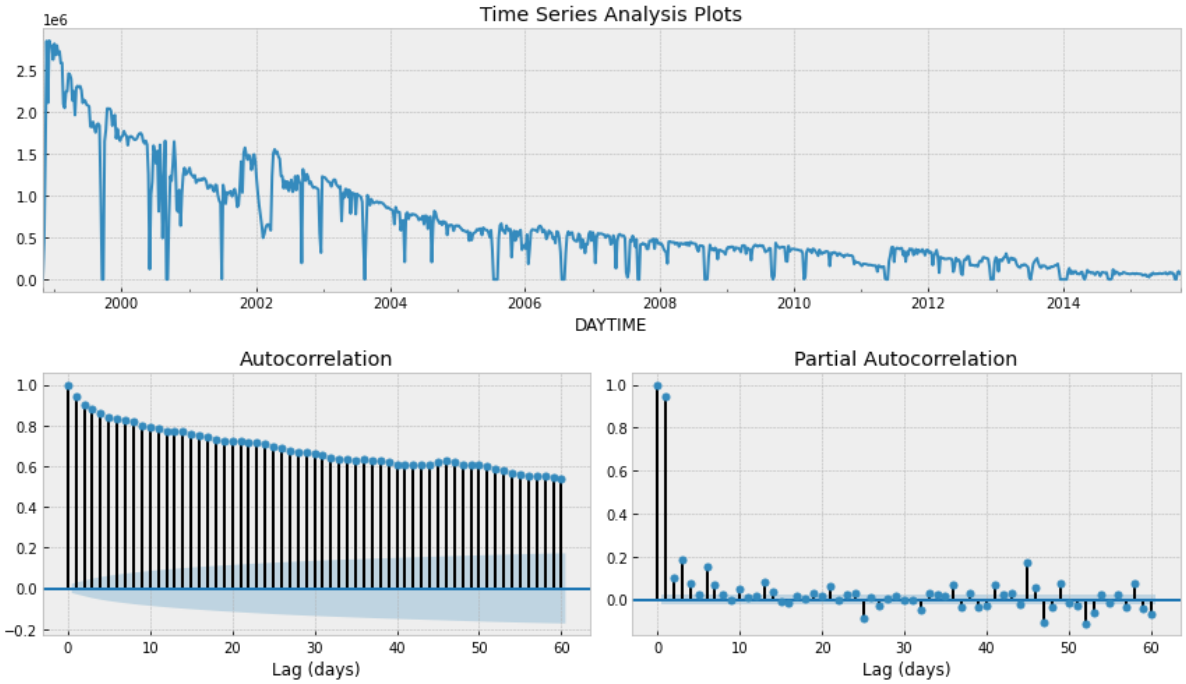

Рисунок 6—Автокорреляционная и частная автокорреляционная функции.

Автокорреляционная функция (коррелограмма ACF) показывает большое число значимых лагов. Так как на графике частной автокорреляционной (PACF)

функции значим лишь один лаг, возьмем первые разности, чтобы привести ряд к стационарному виду (**Рисунок 7**)

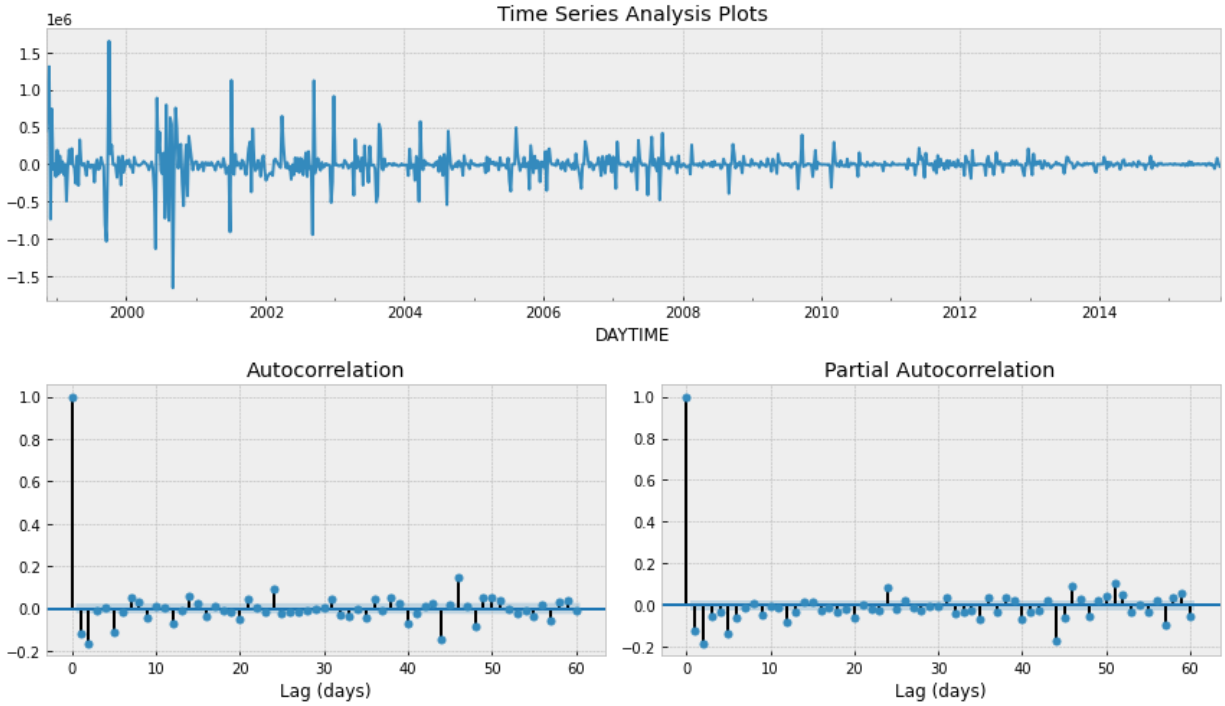

Рисунок 7—Автокорреляционная и частная автокорреляционная функции (ряд первых разностей).

Получили стационарный ряд, по автокорреляционной и частной автокорреляционной функции подберем параметры модели p,d,q, после чего определим порядок сезонных составляющих P,D,Q.

Параметр d есть и он равен 1, осталось определить p и q. Для их определения нам надо изучить автокорреляционную (ACF) и частную автокорреляционную (PACF) функции для ряда первых разностей. ACF поможет нам определить q, т. к. по ее коррелограмме можно определить количество автокорреляционных коэффициентов сильно отличных от 0 в модели MA PACF поможет нам определить p, т. к. по ее коррелограмме можно определить максимальный номер коэффициента сильно отличный от 0 в модели AR.

После изучения коррелограммы PACF можно сделать вывод, что p = 5, т.к. на ней только 5 лаг сильно отличен от нуля. По коррелограмме ACF можно увидеть, что q = 5, т.к. после лага 5 значении функций резко падают.

Теперь перейдем к сезонным составляющим. В нашем случае, при сезонности 52, в лаге 52 параметры P и Q равны 0. В результате наших исследований мы получили модель

$$SARIMA(5,1,5)(0,1,0)_{52}$$

В **Таблице 1** показаны результаты работы модели SARIMA с вышеуказанными параметрами.

Таблица 1—Результаты построения модели $SARIMA(5,1,5)(0,1,0)_{52}$

Statespace Model Results

Dep. Variable:	ALLOC_GAS_VOL_SM3	No. Observations:	780
Model:	SARIMAX(5, 1, 5)x(0, 1, 0, 52)	Log Likelihood	-10097.452
Date:	Tue, 14 Dec 2021	AIC	20216.905
Time:	11:56:06	BIC	20267.383
Sample:	11-01-1998 - 10-06-2013	HQIC	20236.383
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.4925	0.151	-3.263	0.001	-0.788	-0.197
ar.L2	-1.0437	0.155	-6.752	0.000	-1.347	-0.741
ar.L3	0.3205	0.147	2.181	0.029	0.032	0.609
ar.L4	-0.0416	0.092	-0.452	0.651	-0.222	0.139
ar.L5	0.4209	0.057	7.444	0.000	0.310	0.532
ma.L1	0.2104	0.155	1.354	0.176	-0.094	0.515
ma.L2	0.6833	0.135	5.076	0.000	0.419	0.947
ma.L3	-0.7495	0.081	-9.216	0.000	-0.909	-0.590
ma.L4	-0.0917	0.106	-0.869	0.385	-0.299	0.115
ma.L5	-0.4286	0.097	-4.410	0.000	-0.619	-0.238
sigma2	5.819e+10	4.61e-12	1.26e+22	0.000	5.82e+10	5.82e+10

Ljung-Box (Q):	34.22	Jarque-Bera (JB):	3501.44
Prob(Q):	0.73	Prob(JB):	0.00
Heteroskedasticity (H):	0.05	Skew:	-0.15
Prob(H) (two-sided):	0.00	Kurtosis:	13.75

Проверим остаток (residuals) (**Рисунок 8**).

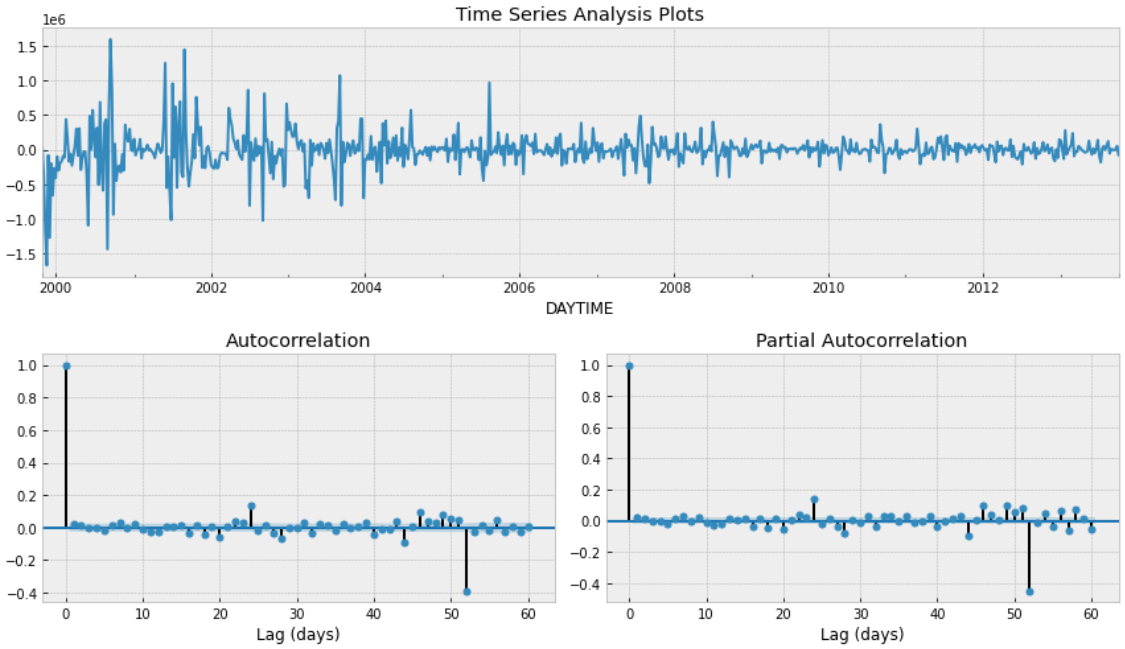

Рисунок 8—Автокорреляционная и частная автокорреляционная функции (остаток).

Остатки стационарны, явных автокорреляций нет. Построим графики истинных и прогнозных значений (**Рисунки 9-10**), а также проведем оценку качества с помощью метрик MAE и MAPE (**Рисунки 11-14**). *В теневой области показан прогнозируемый период на тестовой выборке.*

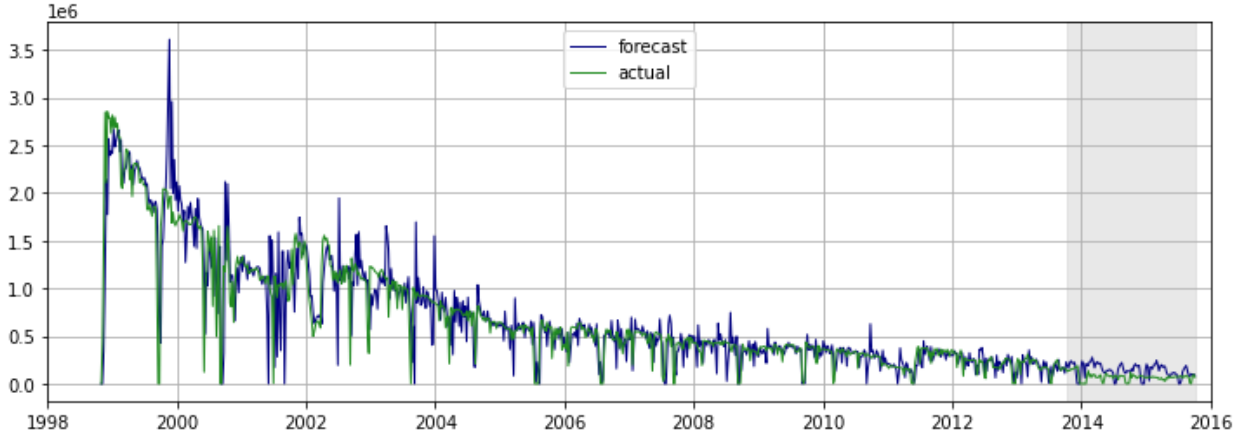

Рисунок 9—Истинные и прогнозные значения. Модель SARIMA (полный датасет).

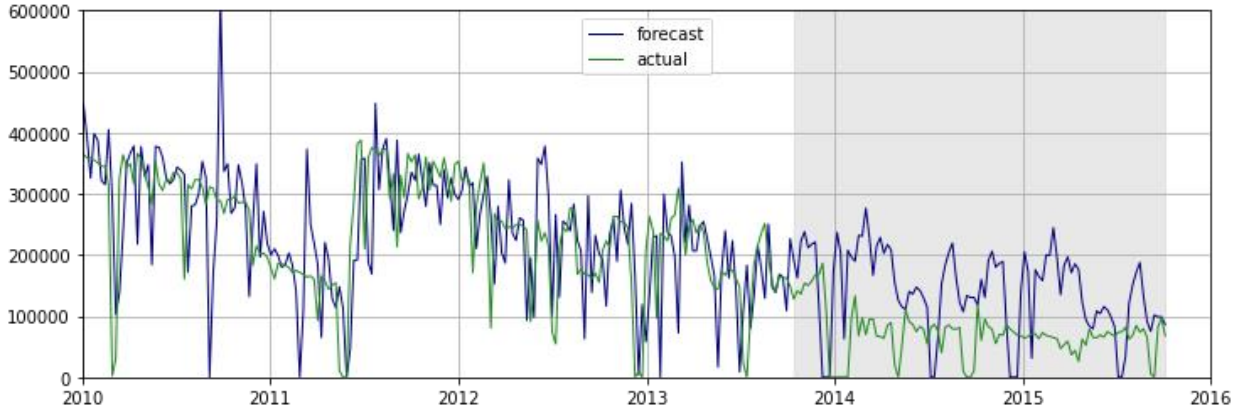

Рисунок 10—Истинные и прогнозные значения. Модель SARIMA.

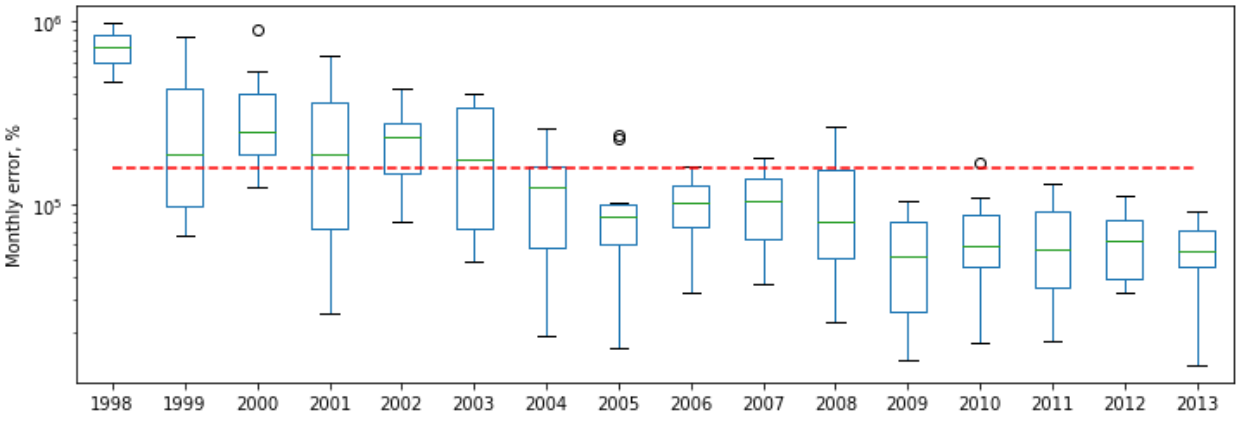

Рисунок 11—Диаграммы размаха значений метрики *MAE* по месяцам для обучающей выборки. Модель SARIMA.

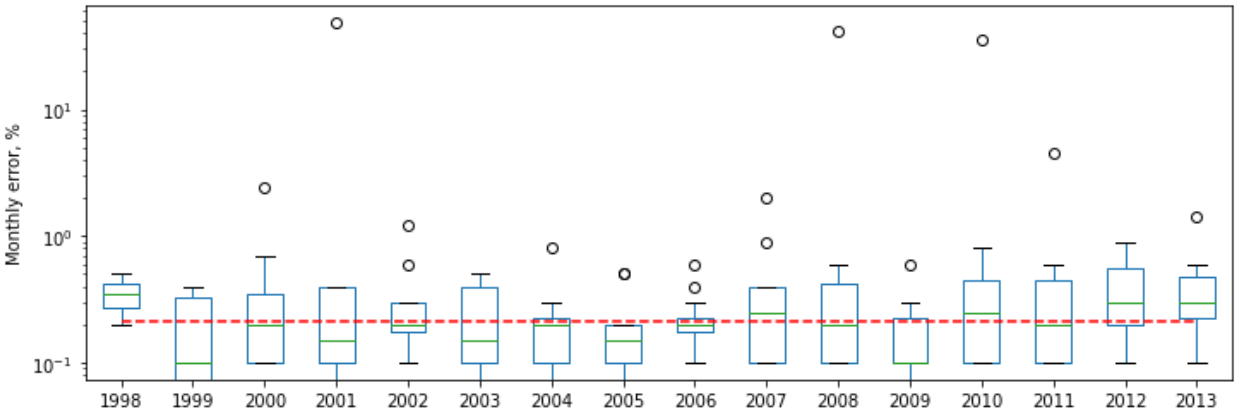

Рисунок 12—Диаграммы размаха значений метрики *MAPE* по месяцам для обучающей выборки. Модель SARIMA.

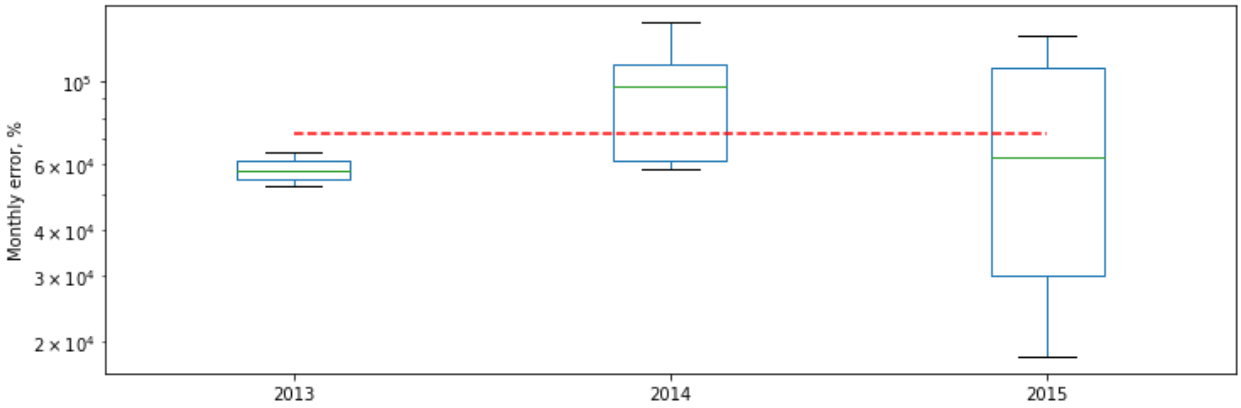

Рисунок 13—Диаграммы размаха значений метрики *MAE* по месяцам для тестовой выборки. Модель SARIMA.

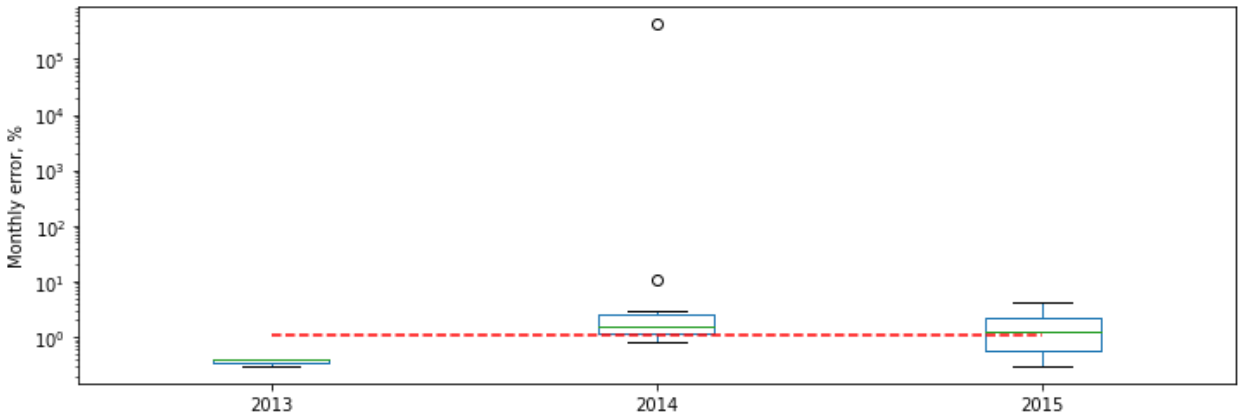

Рисунок 14—Диаграммы размаха значений метрики *MAPE* по месяцам для тестовой выборки. Модель SARIMA.

Прогнозная кривая падения добычи качественно совпадает с истинной. Значения метрик качества модели $SARIMA(5,1,5)(0,1,0)_{52}$ на обучающей выборке – MAE = 159461 м^3, MAPE = 0.21, на тестовой – MAE = 72585 м^3, MAPE = 1.05.

2) Модель Facebook Prophet

Модель состоит из следующих компонент:

$$y(t)=g(t)+s(t)+h(t)+\epsilon_t$$

Сезонные компоненты s(t) отвечают за моделирование периодических изменений, связанных с недельной и годовой сезонностью.

Тренд g(t) — это кусочно-линейная или логистическая функция. Логистическая функция вида:

$$g(t) = \frac{C}{1 + \exp\left(-k(t-b)\right)}$$

позволяет моделировать рост с насыщением, когда при увеличении показателя снижается темп его роста. Библиотека приспособлена выбирать оптимальные точки изменения тренда.

Компонента h(t) отвечает за заданные аномальные и нерегулярные дни.

Ошибка ϵ_t содержит информацию, которая не учтена моделью.

Построим графики истинных и прогнозных значений (**Рисунки 15-17**), а также проведем оценку качества с помощью метрик MAE и MAPE (**Рисунки 18-21**). *В теневой области показан прогнозируемый период на тестовой выборке.*

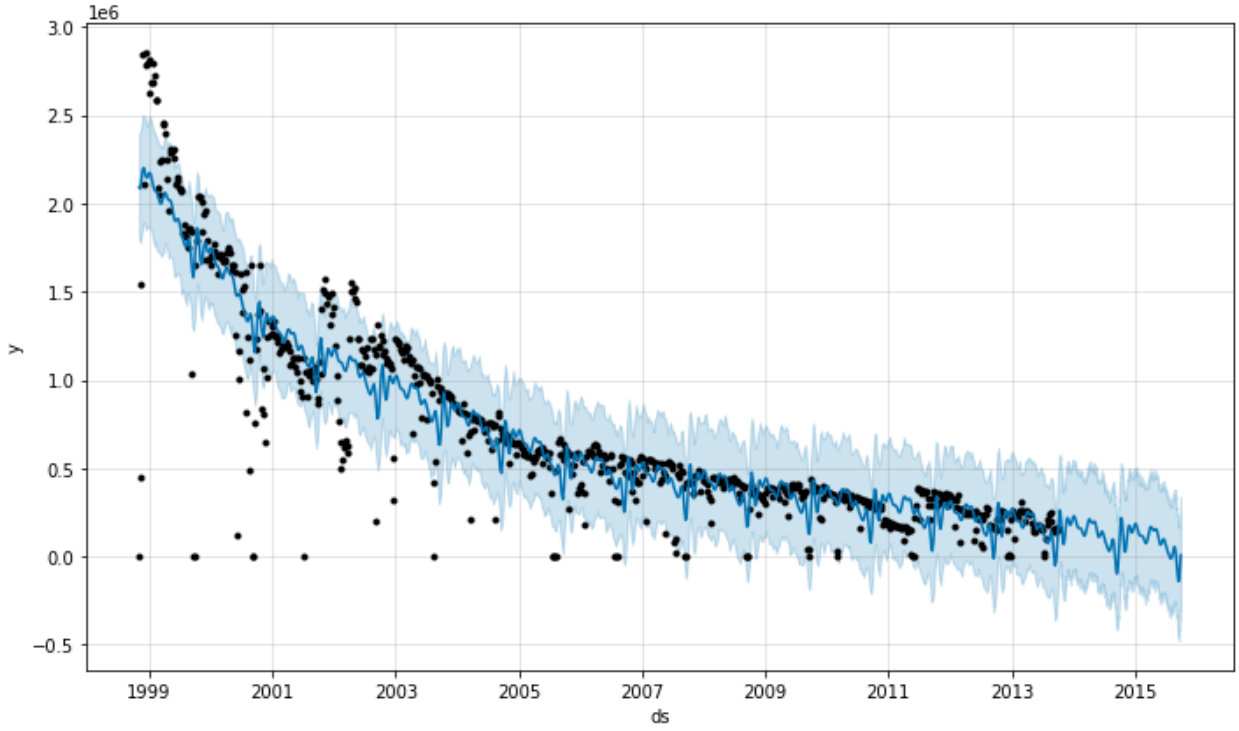

Рисунок 15—Истинные и прогнозные значения. Модель Facebook Prophet (график по умолчанию, полный датасет).

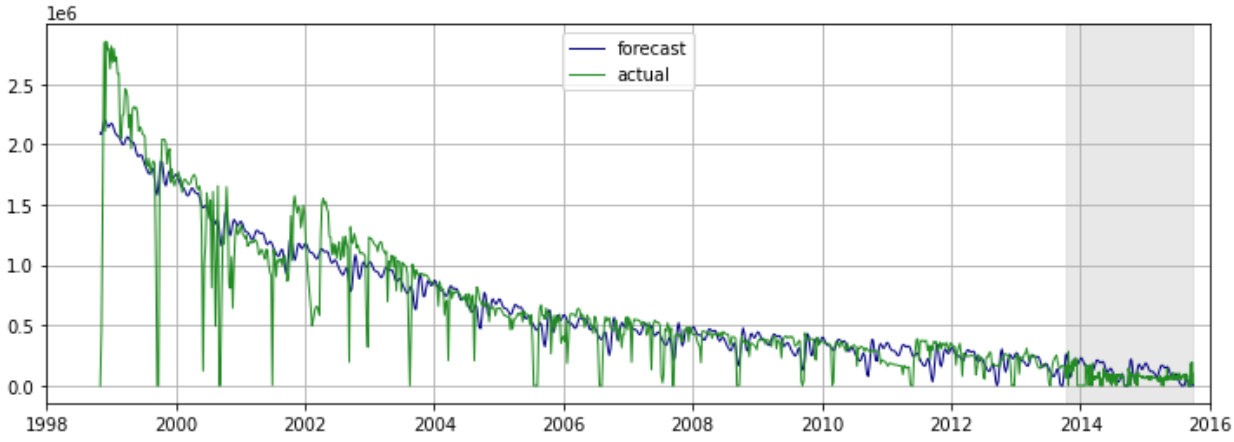

Рисунок 16—Истинные и прогнозные значения. Модель Facebook Prophet (полный датасет).

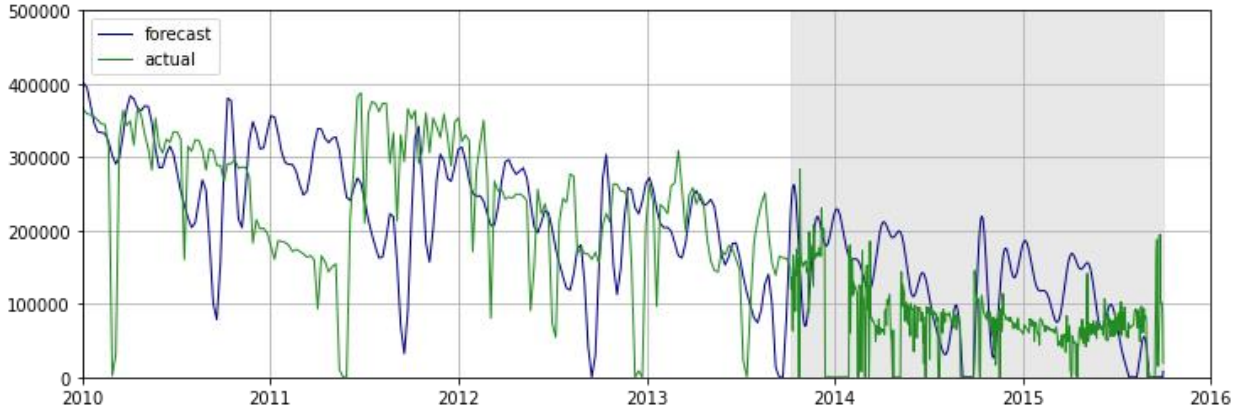

Рисунок 17—Истинные и прогнозные значения. Модель Facebook Prophet.

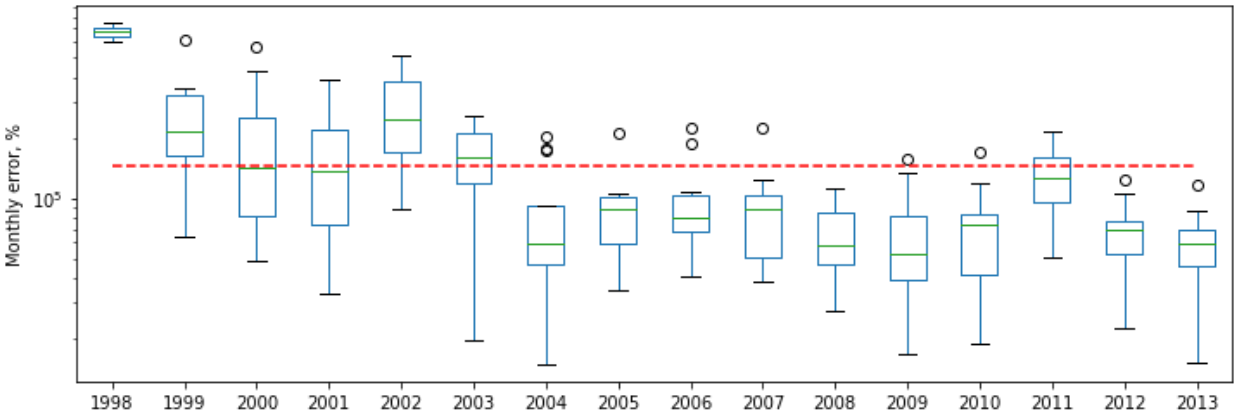

Рисунок 18—Диаграммы размаха значений метрики *MAE* по месяцам для обучающей выборки. Модель Facebook Prophet.

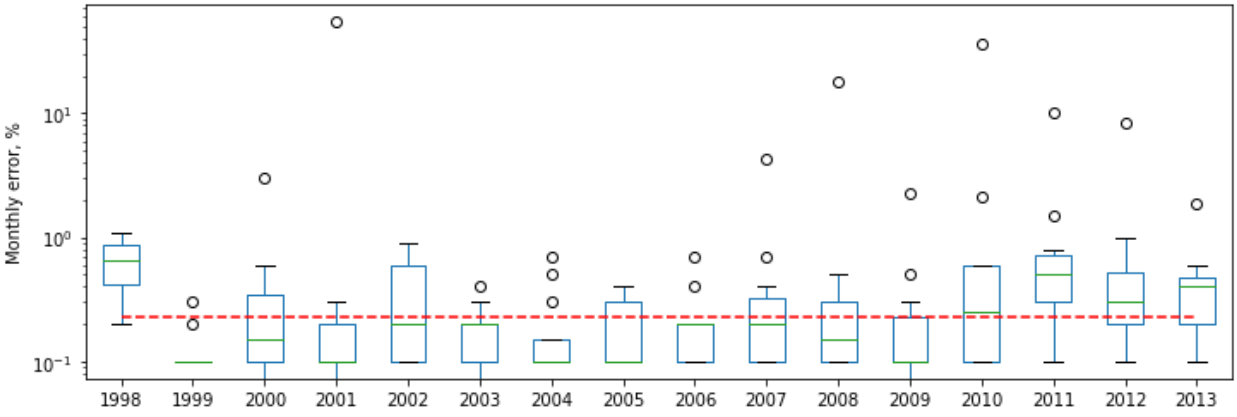

Рисунок 19—Диаграммы размаха значений метрики *MAPE* по месяцам для обучающей выборки. Модель Facebook Prophet.

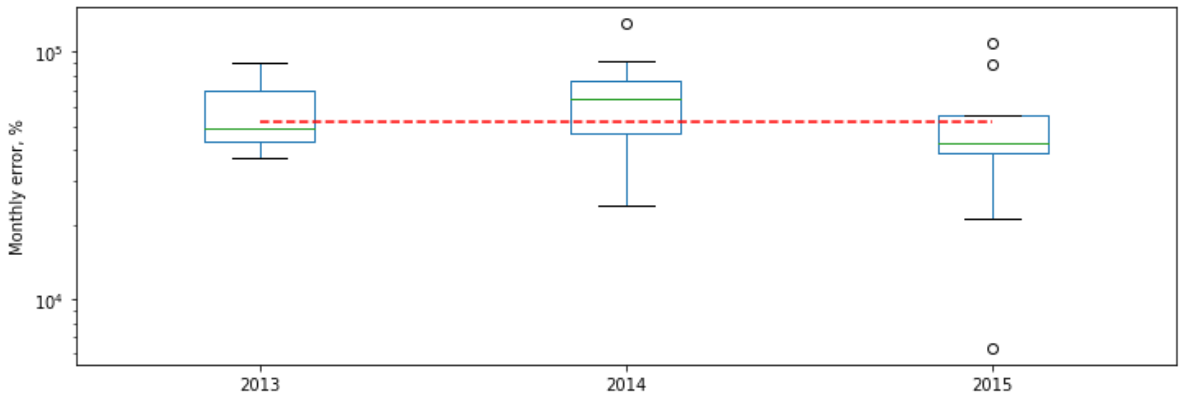

Рисунок 20—Диаграммы размаха значений метрики *MAE* по месяцам для тестовой выборки. Модель Facebook Prophet.

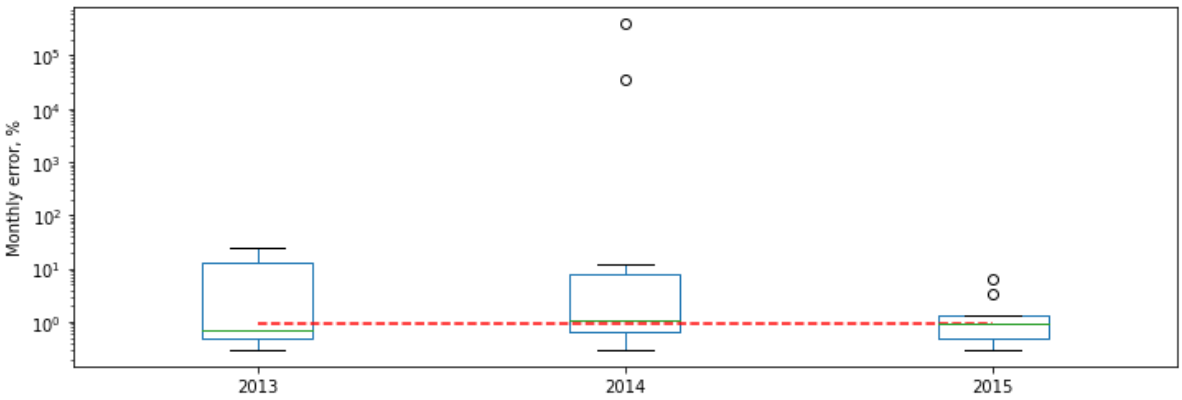

Рисунок 21—Диаграммы размаха значений метрики *MAPE* по месяцам для тестовой выборки. Модель Facebook Prophet.

Прогнозная кривая падения добычи качественно совпадает с истинной. Значения метрик качества модели Facebook Prophet на обучающей выборке – MAE = 145772 м^3, MAPE = 0.23, на тестовой – MAE = 52159 м^3, MAPE = 0.90.

3) Модель XGBoost

XGBoost (Extreme Gradient Boosting) представляет собой эффективную реализацию алгоритма машинного обучения стохастического повышения градиента для задач классификации и регрессии.

XGBoost также можно использовать для прогнозирования временных рядов, хотя для этого требуется, чтобы набор данных временных рядов сначала был преобразован в задачу контролируемого обучения. Это также требует использования специальной техники для оценки модели, называемой пошаговой проверкой. Для обучения модели возьмем лаг в количестве 6 последних значений.

Построим графики истинных и прогнозных значений (**Рисунки 22-23**), а также проведем оценку качества с помощью метрик MAE и MAPE (**Рисунки 24-27**). *В теневой области показан прогнозируемый период на тестовой выборке.*

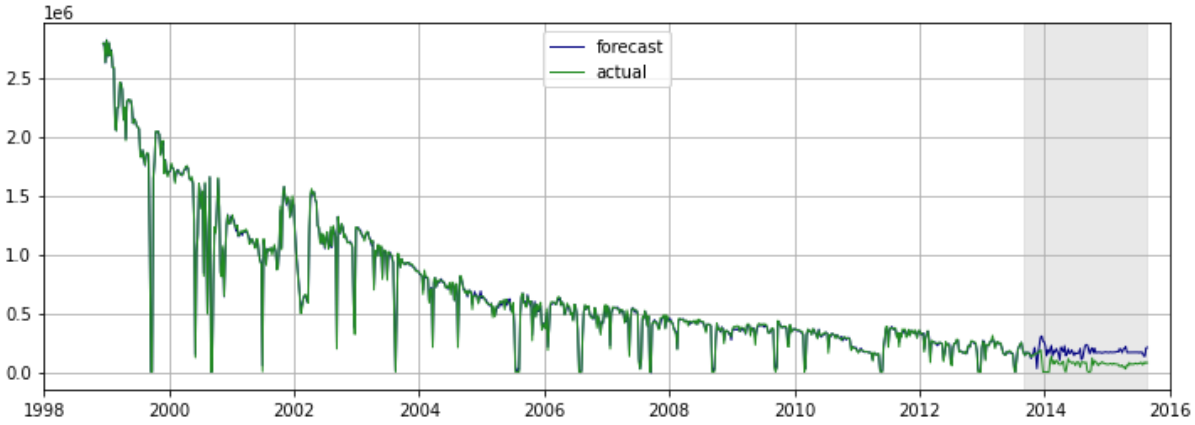

Рисунок 22—Истинные и прогнозные значения. Модель XGBoost (полный датасет).

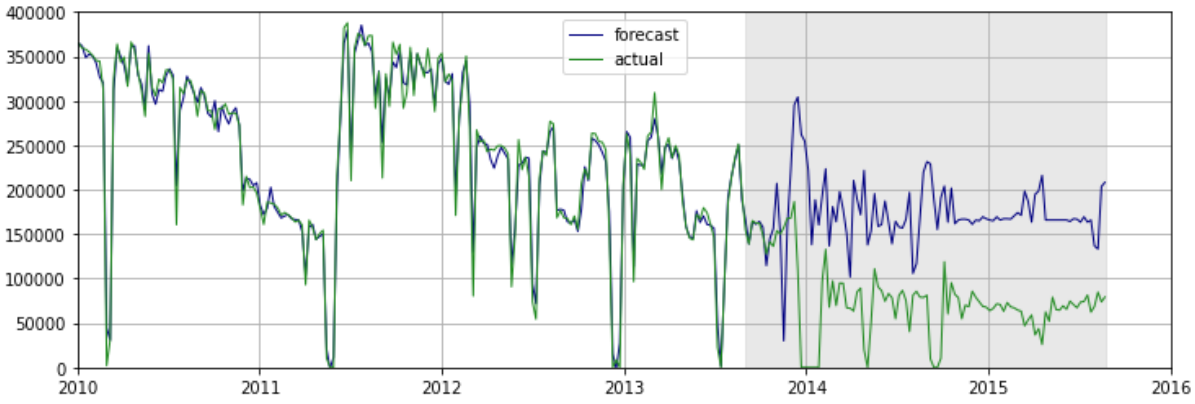

Рисунок 23—Истинные и прогнозные значения. Модель XGBoost.

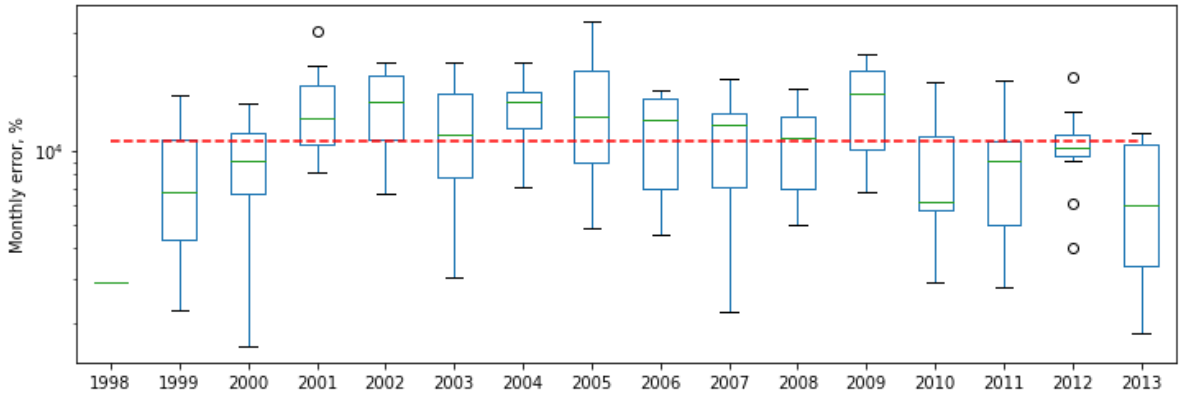

Рисунок 24—Диаграммы размаха значений метрики *MAE* по месяцам для обучающей выборки. Модель XGBoost.

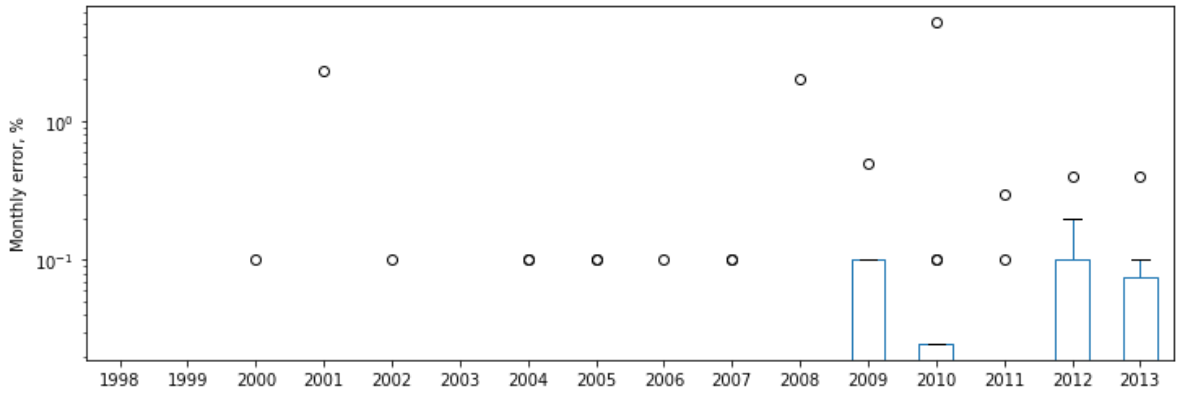

Рисунок 25—Диаграммы размаха значений метрики *MAPE* по месяцам для обучающей выборки. Модель XGBoost.

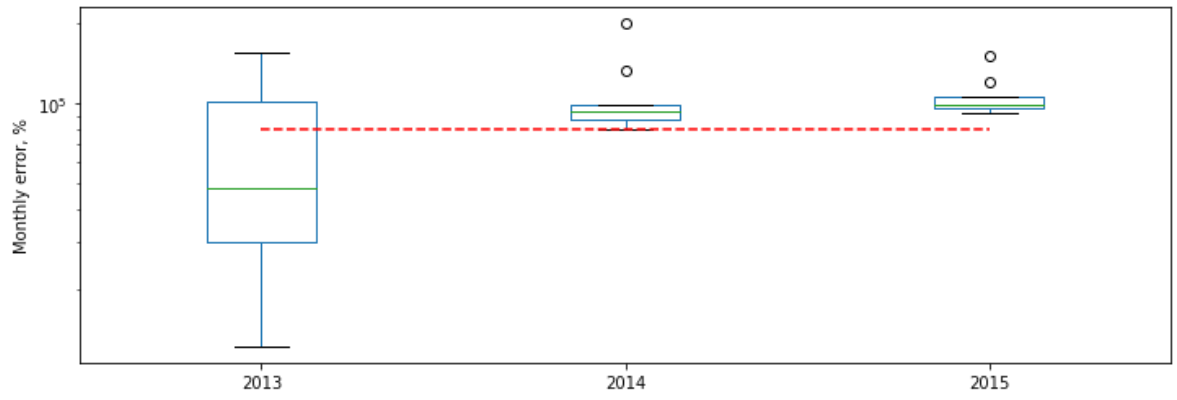

Рисунок 26—Диаграммы размаха значений метрики *MAE* по месяцам для тестовой выборки. Модель XGBoost.

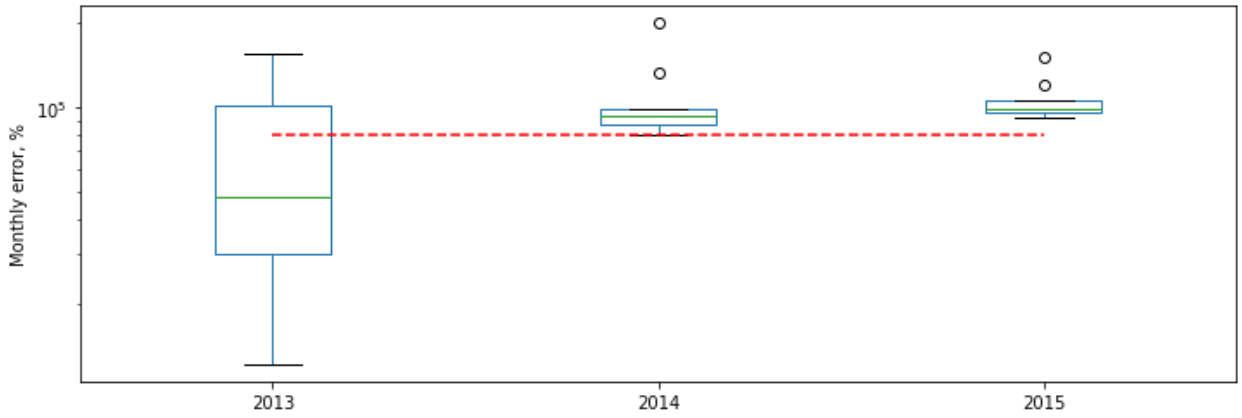

Рисунок 27—Диаграммы размаха значений метрики *MAPE* по месяцам для тестовой выборки. Модель XGBoost.

Прогнозная кривая падения добычи качественно совпадает с истинной. Значения метрик качества модели XGBoost на обучающей выборке – MAE = 10914 м^3, MAPE = 0.01, на тестовой – MAE = 80262 м^3, MAPE = 1.02.

V. Заключение

Графики прогнозных значений, построенных с помощью трех различных алгоритмов – SARIMA, Facebook Prophet и XGBoost, качественно совпадают с истинной и представлены на **Рисунках 28-29**. *В теневой области показан прогнозируемый период на тестовой выборке.*

Соответствующие прогнозным моделям метрики MAE и MAPE представлены в **Таблице 2.**

Лучшее качество в решении задачи прогнозирования дебита газовой скважины 49/12a-K01 месторождения Viking в рамках данного исследования показала модель Facebook Prophet – MAE (тест) = 52159 м^3, MAPE (тест) = 0.90.

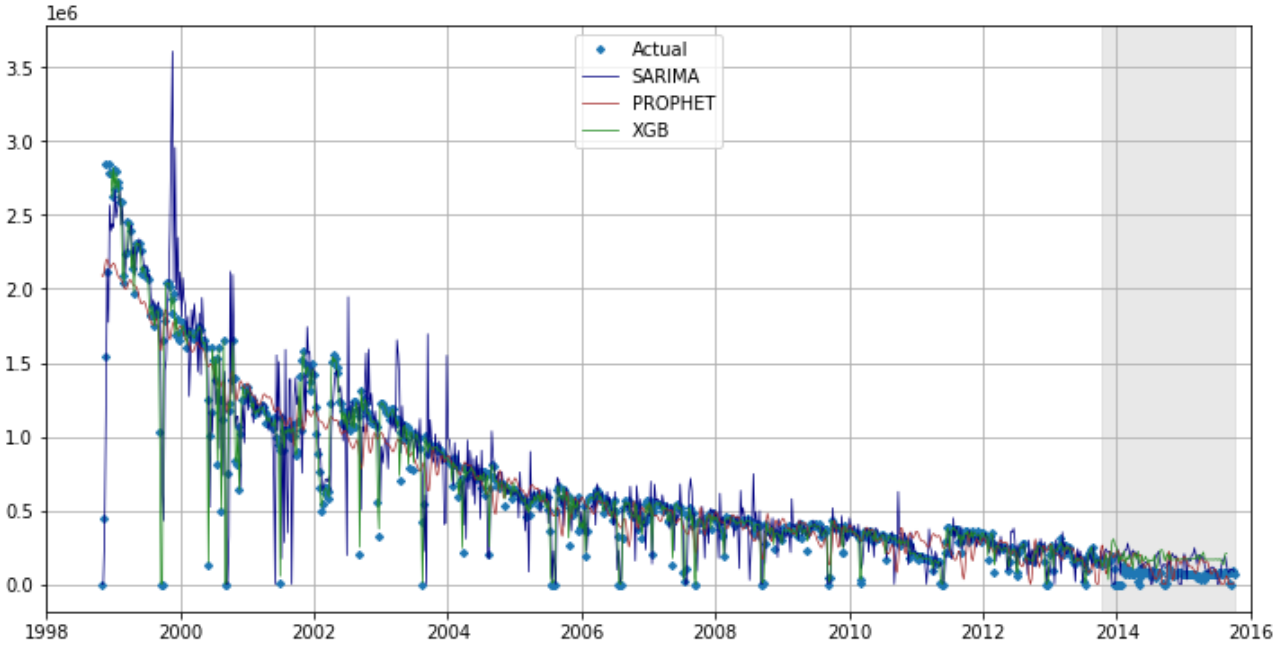

Рисунок 28—Истинные и прогнозные значения моделей SARIMA, Facebook Prophet и XGBoost (полный датасет).

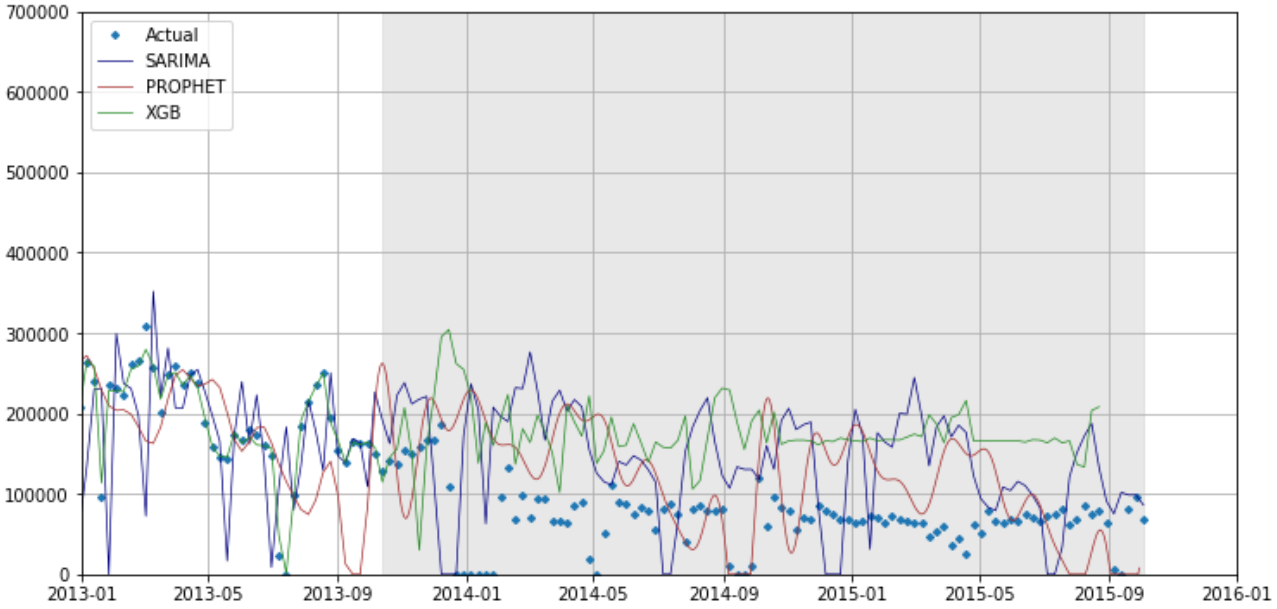

Рисунок 29—Истинные и прогнозные значения моделей SARIMA, Facebook Prophet и XGBoost.

Таблица 2—Метрики качества моделей SARIMA, Facebook Prophet и XGBoost для тестовой выборки.

Model	**MAE**	**MAPE**
SARIMA	72585	1.05
Facebook Prophet	52159	0.90
XGBoost	80262	1.02

Приложение

Библиотеки, классы и функции, использованные в работе.

Язык программирования - Python

```
import pandas as pd
import numpy as np
import seaborn as sns

from urllib.request import urlopen
from io import BytesIO
from zipfile import ZipFile
from datetime import timedelta
from datetime import datetime
import warnings
warnings.filterwarnings('ignore')

import matplotlib.pyplot as plt
import statsmodels.tsa.api as smt
import statsmodels.api as sm
from statsmodels.tsa.seasonal import seasonal_decompose
from sklearn.metrics import mean_absolute_error as mae
from sklearn.metrics import mean_absolute_percentage_error as mape
from sklearn.metrics import r2_score

from statsmodels.tsa.statespace.sarimax import SARIMAX
from fbprophet import Prophet
from xgboost import XGBRegressor



class Plots():
  """Класс для отрисовки графиков, используемых в проекте.
  """

  def wells(data):
    """Функция для построения графиков добычи из скважин.
    """

    for well in wells:
      data_well = data[data['WELLNAME'] == well].reset_index()
      data_well.plot('DAYTIME', 'ALLOC_GAS_VOL_SM3', kind = 'scatter', s=0.1
, title = well)
      plt.show()
```

```
  def forecast_plot(actual, forecast, predict_period):
    """Функция для построения графика истинных и прогнозных значений.
    """

    predict_start = forecast.index[-predict_period]
    predict_end = forecast.index[-1]

    plt.figure(figsize=(12, 4))
    plt.plot(forecast, color='navy', label='forecast', linewidth=1)
    plt.plot(actual, color='forestgreen', label='actual', linewidth=1)
    plt.legend()
    plt.axvspan(predict_start, predict_end, alpha=0.5, color='lightgrey')
    plt.grid(True)
    plt.xlim(pd.Timestamp(forecast.index[0].year, 1, 1), pd.Timestamp(foreca
st.index[-1].year+1, 1, 1))

  def tsplot(y, lags=None, figsize=(12, 7), style='bmh'):
    """Функция для построения графиков анализа временных рядов.
    """

    if not isinstance(y, pd.Series):
      y = pd.Series(y)
    with plt.style.context(style):
      fig = plt.figure(figsize=figsize)
      layout = (2, 2)
      ts_ax = plt.subplot2grid(layout, (0, 0), colspan=2)
      acf_ax = plt.subplot2grid(layout, (1, 0))
      pacf_ax = plt.subplot2grid(layout, (1, 1))
      acf_ax.set_xlabel('Lag (days)')
      pacf_ax.set_xlabel('Lag (days)')

      y.plot(ax=ts_ax)
      ts_ax.set_title('Time Series Analysis Plots')
      smt.graphics.plot_acf(y, lags=lags, ax=acf_ax, alpha=0.5)
      smt.graphics.plot_pacf(y, lags=lags, ax=pacf_ax, alpha=0.5)

      print("Критерий Дики-Фуллера: p=%f" % sm.tsa.stattools.adfuller(y)[1])
      plt.tight_layout()
    return

  def boxplot(data, error):
    """Функция для оценки ошибок MAE и MAPE применительно к временным рядам
    воспользуемся методикой, изложенной в статье (Jiang и Hu, 2018),
    выделив распределение ошибки по месяцам года с помощью диаграммы
    разброса.
```

```
    """

    years = data.index.year.unique().tolist()
    months = range(1, 13)

    error_per_year = pd.DataFrame(index=months)
    for year in years:
      error_per_month = {}
      for month in months:
        data_error = data[
                          (data.index.year == year) &
                          (data.index.month == month) &
                          (data['actual'] != 0) &
                          (data['forecast'] != 0)
                          ]
        if data_error.shape[0] != 0:
          error_per_month[month] = round(error(data_error['actual'], data_er
ror['forecast']), 1)
        else:
          error_per_month[month] = np.nan
          pass

      error_per_year[year] = pd.DataFrame.from_dict(error_per_month, orient=
'index', columns=[year])

    error_per_year.plot(kind='box', figsize=(12, 4), logy=True, ylabel='Mont
hly error, %')
    error_mean = round(error_per_year.median().mean(), 2)
    print('Среднее значение ежемесячной ошибки по годам: {}'.format(error_me
an))
    pd.Series({1: error_mean, len(error_per_year.columns): error_mean}).plot
.line(linestyle='--', color='r')

  def seasonal_decomposition(data):
    """Функция для построения графиков декомпозиции.
    """

    decomposed = seasonal_decompose(data)
    plt.figure(figsize=(6, 8))
    plt.subplot(311)

    # Чтобы график корректно отобразился, указываем его
    # оси ax, равными plt.gca() (англ. get current axis,
    # получить текущие оси)
    decomposed.trend.plot(ax=plt.gca())
    plt.title('Trend')
```

```
    plt.subplot(312)
    decomposed.seasonal.plot(ax=plt.gca())
    plt.title('Seasonality')
    plt.subplot(313)
    decomposed.resid.plot(ax=plt.gca())
    plt.title('Residuals')
    plt.tight_layout()

  def results():
    """Функция для построения графиков результатов анализа временных рядов.
    """

    plt.figure(figsize=(12, 6))
    plt.plot(df_fc_arima['actual'], 'D', label='Actual', markersize=3)
    plt.plot(df_fc_arima['forecast'], color='navy', label='SARIMA', linewidt
h=0.7)
    plt.plot(df_fc_prophet['forecast'], color='brown', label='PROPHET', line
width=0.7)
    plt.plot(df_fc_xgb['forecast'], color='forestgreen', label='XGB', linewi
dth=0.7)
    plt.axvspan(df_fc_arima['forecast'].index[-
predict_period], df_fc_arima['forecast'].index[-
1], alpha=0.5, color='lightgrey')
    plt.legend()
    plt.grid(True)
```

Список литературы:

F. A. Anifowose, “Ensemble Machine Learning: The Latest Development in Computational Intelligence for Petroleum Reservoir Characterization,” SPE Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, 2013.

J.J. Arps, “Analysis of Decline Curves. Published in Petroleum Transactions”, AIME, 160 (1945): 228 –247.

D. Castiniera, R. Toronyi, and N. Saleri, “Machine Learning and Natural Language Processing for Automated Analysis of Drilling and Completion Data,” SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 2018.

T. Chai, and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?”, 2014

T. Chen, and C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, 2016

A. Duong, “An Unconventional Rate Decline Approach for Tight and Fracture-Dominated Gas Wells”. Paper CSUG/SPE 137748, 2010.

D. Fu, “Unlocking Unconventional Reservoirs With Data Analytics, Machine Learning, and Artificial Intelligence,” Journal of Petroleum Technology, pp. 14–15, January 2019.

C. Fulton, “Estimating time series models by state space methods in Python: Statsmodels”, 2017

R. J. Hyndman, and A. B. Koehler, “Another look at measures of forecast accuracy. International Journal of Forecasting”, 22, 679–688, 2006

D. Ilk, J. A. Rushing, and T.A. Blasingame, “Exponential vs. Hyperbolic Decline in Tight Gas Sands Understanding the Origin and Implications for Reserve Estimates Using Arps’ Decline Curves”. Paper SPE 116731, 2008.

L. Jiang, and G. Hu, “Day-Ahead Price Forecasting for Electricity Market using Long-Short Term Memory Recurrent Neural Network”, 2018

A. Marana, J. Papa, M. V. Ferreira, K. Miura, and F. A. Torres, “An Intelligent System To Detect Drilling Problems Through Drilled-CuttingsReturn Analysis” , 2010.

S. Mohaghegh, “How Does the Use of Artificial Intelligence and Machine Learning Differ for Conventional vs. Unconventional Plays?”, 2019

C. I. Noshi, “Application of Data Science and Machine Learning Algorithms for ROP Optimization in West Texas: Turning Data into Knowledge”, 2019.

M. Pennel, J. Hsiung and V. Putcha, “Detecting Failures and Optimizing Performance in Artificial Lift Using Machine Learning Models,” SPE Western Regional Meeting, Garden Grove, CA, 2018.

A. Rastogi and A. Sharma, “Quantifying the Impact of Fracturing Chemicals on Production Performance Using Machine Learning”, 2019.

J. Sun, X. Ma and M. Kazi, “Comparison of Decline Curve Analysis DCA with Recursive Neural Networks RNN for Production Forecast of Multiple Wells” SPE Western Regional Meeting, Garden Grove, CA, 2018.

S. Tandon, “Integrating Machine Learning in Identifying Sweet Spots in Unconventional Formations,” SPE Western Regional Meeting, San Jose, CA, 2019

S. J. Taylor, and B.Letham, “Facebook Prophet – forecasting at scale”, 2017

S. Vallabhaneni, R. Saraf, and S. Priyadarshy, “Machine-LearningBased Petrophysical Property Modeling”, 2019.