Group 9
Gan Feng
Vivek Gupta
Seth Podhoretz
Arjun Ravikumar
Kou Rui

STAT 626 - Methods of Time Series Analysis
Production Analysis of Oilfield Data
Report 2

Texas A&M University

Abstract

Oilfield data such as production rates carry valuable information that can be used to forecast production volumes and economics of hydrocarbon extraction. This project attempts to extract such information from 60 wells located in the Bakken shale in North Dakota. The Bakken shale is credited with the shale boom in the USA, reversing decades of declining oil production in the country. We will be attempting to characterize the production behavior of the 60 wells, and eventually predict the future performance and ultimate recovery of hydrocarbon.

Our analysis thus far has focused on analyzing oil production from each of the wells as a univariate time series. The conclusions are that after log transformation and differencing of oil production data, white noise and MA(1) seem to be the most appropriate and indeed the simplest model possible for most wells in this dataset. However, operational disruptions and noise can make model identification through ACF and PACF plots ambiguous. In these cases, outlier detection methods and smoothing techniques need to be considered before building an ARIMA model.

Introduction

There are an abundance of oilfield data from approximately two million oil and gas wells in the USA. These include oil, water and gas rates, bottomhole and tubing head pressures, temperatures, etc. These data carry in them crucial information on the physics behind the production of hydrocarbon. An understanding of these physics can help forecast the future production of hydrocarbon, a process that has operational and economic importance.

Our approach to modeling the data follows a streamlined workflow. We plot the oil production rate for each well to determine relevant sections of the data for the analysis. Next, we attempt to bring the data to stationarity using transformations. We determine candidate ARIMA models by looking at the ACF and the PACF. After fitting these candidate models, we validate the fit using residuals diagnostics, including checking for normality of residuals and independence.

Data Description

As part of preprocessing, the 60 wells considered were put into four different bins, based on the nature of the oil production data.

- **Type 1**. Continuous decline trend with low noise.
- **Type 2**. Operational disruptions maintaining decline trend.
- **Type 3**. Operational disruptions changing decline trend.
- **Type 4**. Unclear decline trend with noisy data.

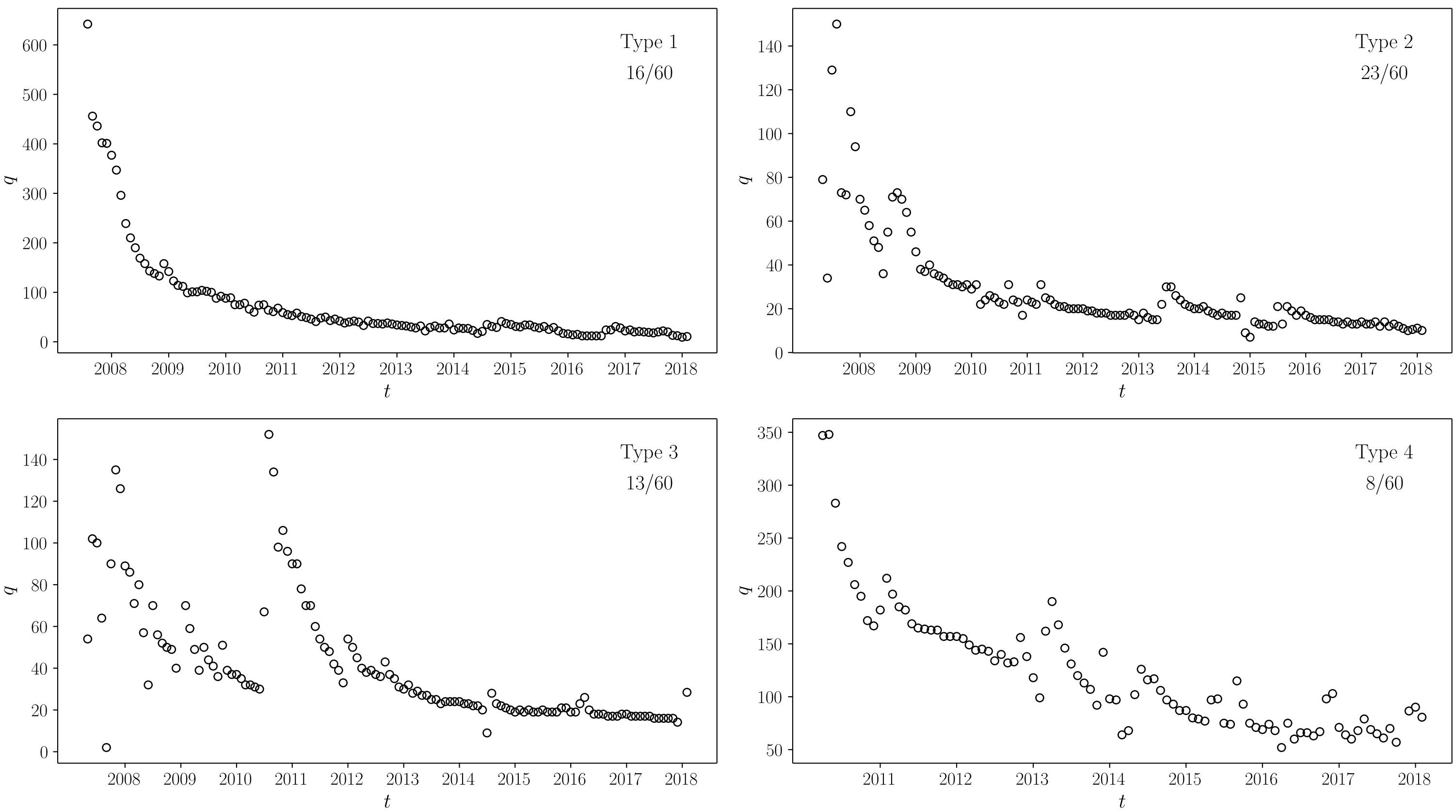

Figure 1: Examples for the Four Broad Types of Wells Analyzed

Figure 1 provides a summary of each of the four types of wells. Types 2 and 3 occur due to wells being shut in for operations. The difference between the two types is that in type 2 behavior, the nature of the well does not change. In type 3 behavior, the performance of the well is significantly improved, due to refracturing or similar intervention.

For forecasting future production, the only relevant periods are when the oil production is declining. All 60 wells in early time have some data in increasing production. This behavior is purely due to operational reasons, and does not reflect the future behavior of the well. We therefore exclude this period from our analysis.

The dataset also contains water and gas production rates. These could be analyzed along with oil production rates as a multivariate time series. From the location data in the dataset, rates from neighboring wells could be analyzed as a multivariate time series for interference between wells.

Data Analysis

Figure 2 shows the well chosen for Type 1, including the differenced, log transformed, and log transformed differenced data. We can see that differencing to remove the trend still has some changing variance in the data. The log transformed differenced data looks more stationary than the differenced data.

Figure 3 shows the ACF and the PACF of the data. From the ACF and PACF, we can reasonably conclude that the log transformed differenced data is quite close to white noise. We therefore attempt an ARIMA(0, 1, 0) model.

Figure 4 shows the residuals, normal Q-Q plot, the ACF of residuals, and the p-values from the Ljung-Box statistic. These diagnostics look quite reasonable for the fitting of an ARIMA(0, 1, 0) model, (i.e.) that the log differenced data could be white noise.

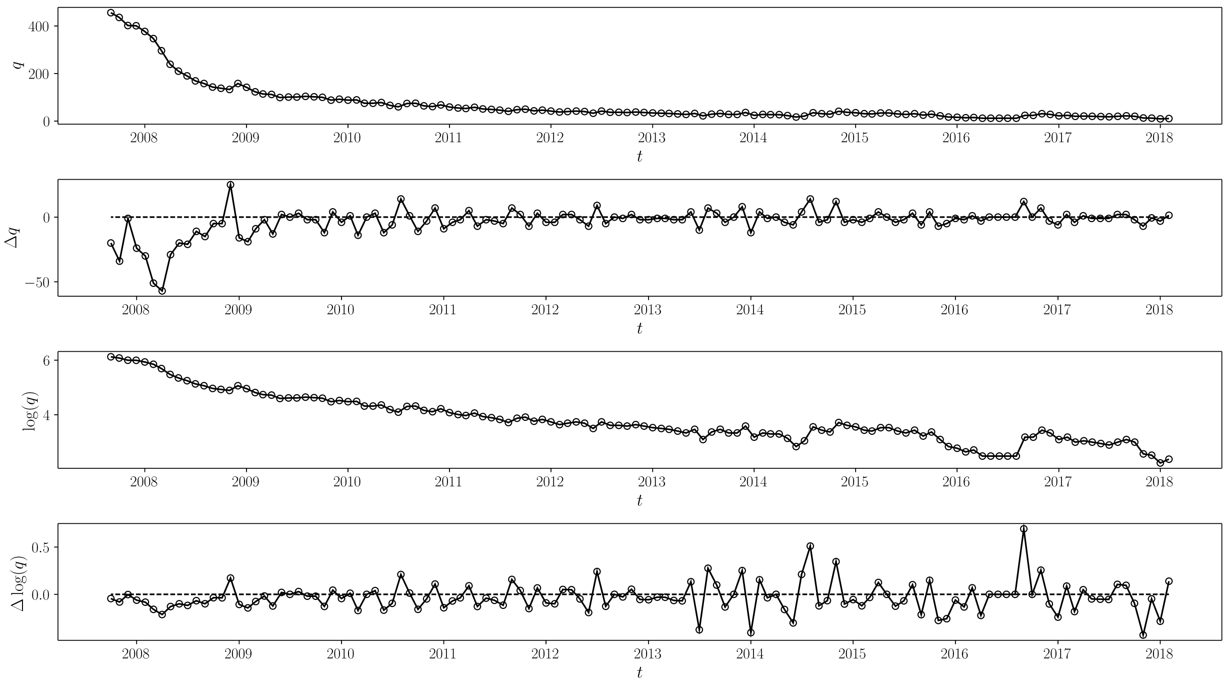

Figure 2: Differenced, Log Transformed, and Log Differenced Data, Type 1

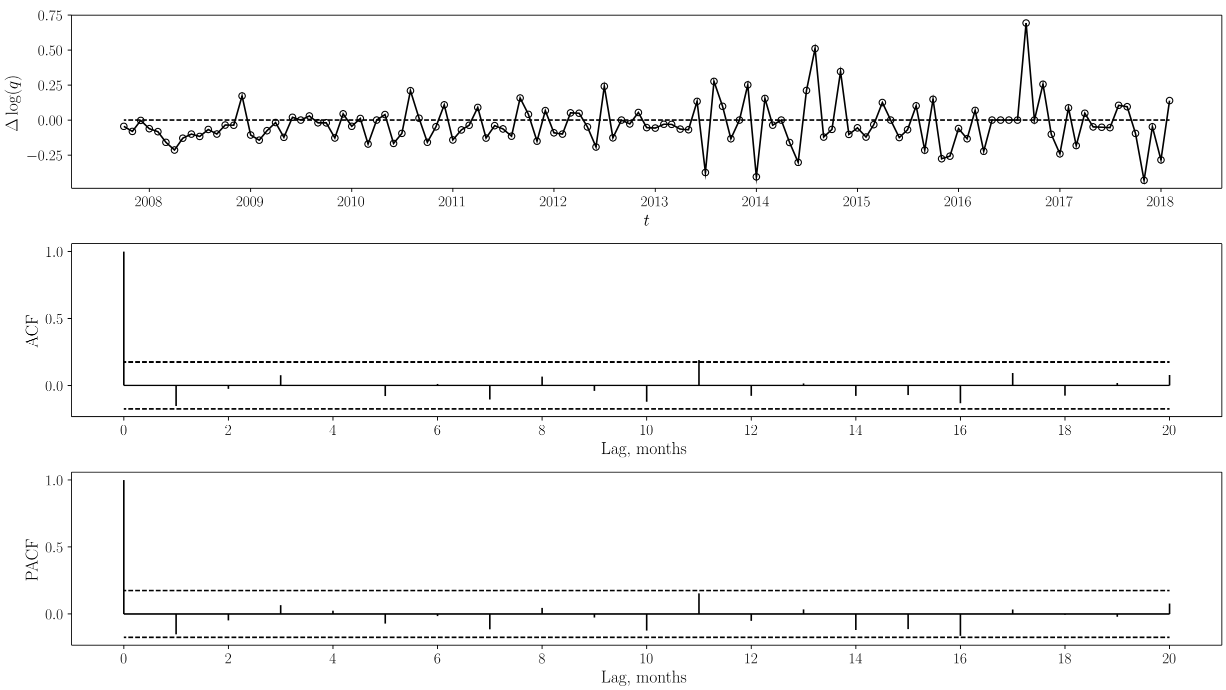

Figure 3: ACF, PACF for Log Transformed Data, Type 1

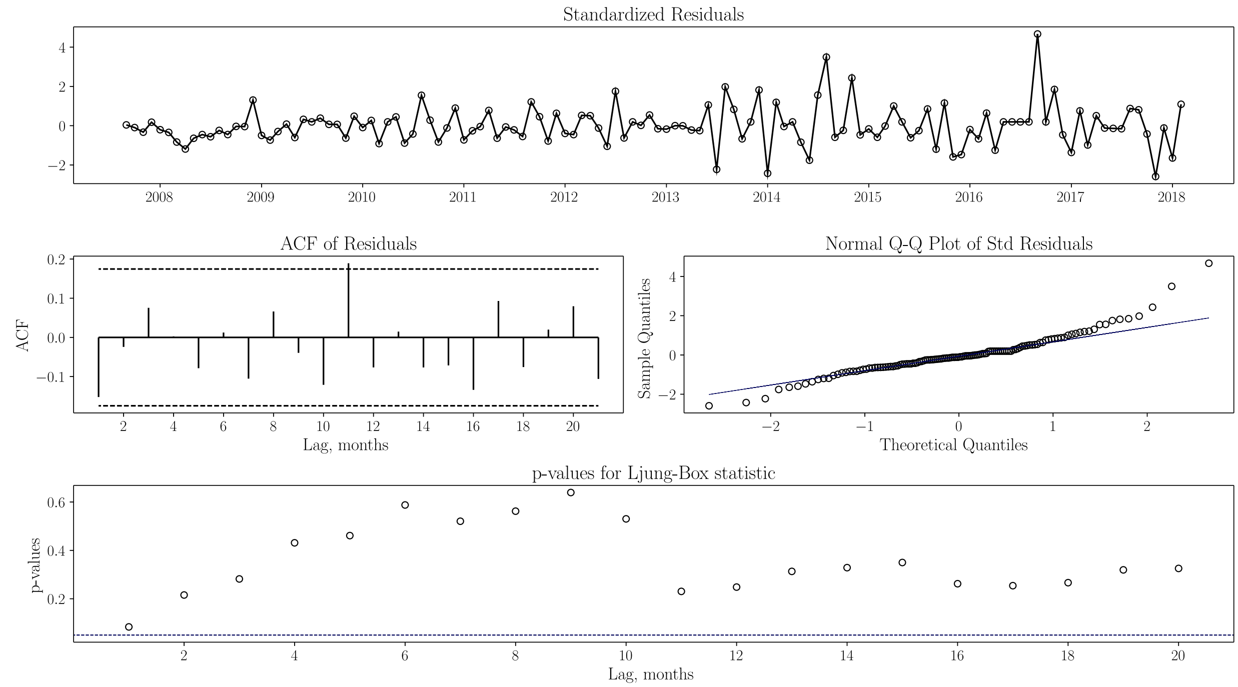

Figure 4: Residuals Diagnostics for ARIMA(0, 1, 0) for Log Transformed Data, Type 1

This procedure is repeated for all 60 wells. The various types of wells only differ in proprocessing; all the other procedures remain the same. Table 1 summarizes the models chosen for representative wells of the four different types.

Table 1: Model Summary

Type	**API**	**Order**
Type 1	33061005340000	ARIMA(0, 1, 0)
Type 2	33025006070000	ARIMA(0, 1, 1)
Type 3	33025006180000	ARIMA(0, 1, 1)
Type 4	33053028000000	ARIMA(0, 1, 0)

Discussions

The prevailing theory on decline curve analysis holds that the production behavior of oil wells is either exponential or hyperbolic (Arps, 1945). From the form of equations for both these behaviors, it can be seen that we expect the log differenced data to be white noise. However, the ACF and PACF of the wells reveal that this is not always the case.

This is due to the fact that the presence of noise and outliers in the data impact ACF and PACF estimation from data. Since model identification was done by looking at the ACF and PACF, a few outliers due to operational reasons can lead to the selection of a completely different model. This calls for an outlier detection and removal algorithm to preprocess the data, or a different approach to model identification.

Another issue with model identification is the presence of zero rate points. These points, when they occur in small groups in the data, prevent the log transform from being applied. We dropped these zero rate points before data analysis, since adjusting the rates by adding a constant to them would create more outliers, and possibly impact the relative magnitude of rate.

Handling the outliers could be done with a smoothing algorithm, or with an algorithm that detects and removes them, or by interpolating the point based on the surrounding points. The issue of whether these zero rate points and outliers should have been removed or interpolated is a subject that is ultimately undecided, and calls for more analysis.

Conclusions

The analysis and model identification for the 60 wells reveals that most of the wells match an ARIMA(0, 1, 0) or ARIMA(0, 1, 1) model, once the oil production data has been log differenced. Wells with significant noise and outliers are much more difficult to match with this simplistic approach, and require more sophisticated preprocessing before attempting to identify or fit a model.

This conclusion comes from a decision to drop zero points from the data, to enable us to take the log transform of the data. Outlier analysis was not performed, and in particularly noisy wells, they had an impact on model determination through the ACF and PACF plots.

References

Arps, J. J. (1945). Analysis of Decline Curves. Society of Petroleum Engineers. doi:10.2118/945228-G

Society of Petroleum Evaluation Engineers (2014). Estimating Developed Reserves in Unconventional Reservoirs. Monograph 4.

Roles and Responsibilities

Table 2: Contributions of Group Members

Member	**Contribution**
Gan Feng	Analyzed 12 wells, prepared analysis for Type 3 wells
Vivek Gupta	Analyzed 12 wells, prepared analysis for Type 1 wells
Seth Podhoretz	Analyzed 12 wells, prepared analysis for Type 2 wells
Arjun Ravikumar	Analyzed 12 wells, prepared analysis for Type 4 wells
Kou Rui	Selected representative wells, and compiled presentation