

Branch: master stat_comp / MCMC_project /

Create new fileUpload filesFind fileHistory

deckboy centerLatest commit f32730a a minute ago

..

.ipynb_checkpoints	center	a minute ago
images	final report	44 minutes ago
input	upload MCMC project	20 days ago
presentation	final report	44 minutes ago
src	final report	44 minutes ago
.DS_Store	final report	44 minutes ago
Arps_Testing.ipynb	python code	42 minutes ago
MCMC_Production_Analysis_v6.ipynb	python code	42 minutes ago
README.md	center	a minute ago
theta_testing.ipynb	python code	42 minutes ago
trans_testing.ipynb	python code	42 minutes ago

README.md

MCMC implementation for oil and gas production models

Rui Kou, Wajahat Ali

04/30/2019

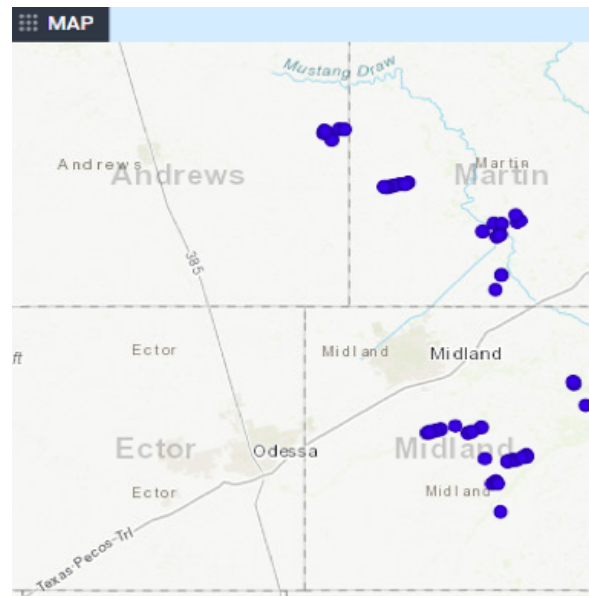
1. Introduction

Production Data Analysis has been particularly useful in forecasting the oil and gas production rate and estimating the ultimate recovery (EUR). The commonly used deterministic method relies heavily on the experience of the evaluator and provides little information on the uncertainty of the estimation. In this project, we apply the Bayesian probabilistic method to estimate the parameters in both empirical and physics-based models. We aim to achieve the following objectives:

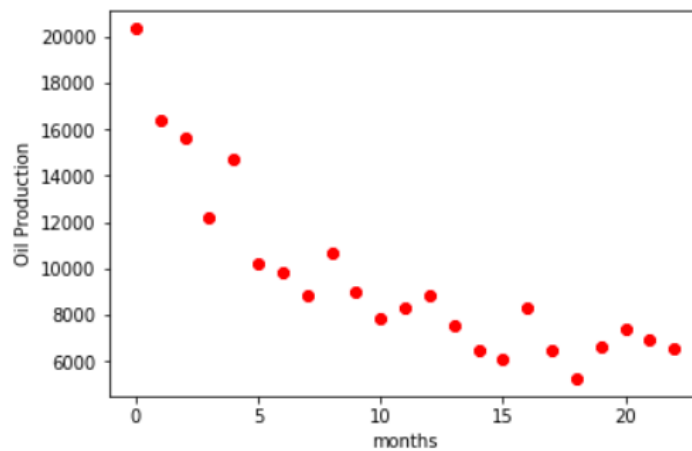
- To develop a Python program that generates probabilistic decline curves using Bayesian statistics and the Markov Chain Monte Carlo sampling method
- To investigate the performance of both empirical (Arps model) and physics-based decline curve models (e.g. Transient Hyperbolic Model, θ Function Model) using production history data from unconventional oil producers

2. Oil and Gas Production Data

In this project, we use real production data from 150 wells located in Midland Texas. The map below shows the location of wells. Note that each of the blue dot may represent around 10 wells close to each other.



Oil production rate from a single well declines over the time. Depends of the formation property, operation condition and well completions quality, the decline curve could be smooth or noisy. We show the production history data of a representative well for 24 months. Our goal is to use the first 12 month data to build a probabilistic model, and compare second year data with the P10-P90 interval.



3. Model Formulation

We first introduce the 3 decline curve models that we are going to implement.

The first model is Arps model. Arps model is an empirical model, where the D parameter can be interpreted as the first order derivative line, and b parameter (constant) is the second derivative.

$$q_t = q_i (1 + b D_i t)^{-\frac{1}{b}} \quad \text{Arps Model, Empirical}$$

$$D = -\frac{1}{q} \frac{dq}{dt} \quad \text{Decline rate, first derivative}$$

$$b = \frac{d}{dt} \left[\frac{1}{D} \right] \quad \text{b-factor (constant), second derivative}$$

To implement the probabilistic Arps model, we need to estimate 3 parameters: q_i, D_i, b

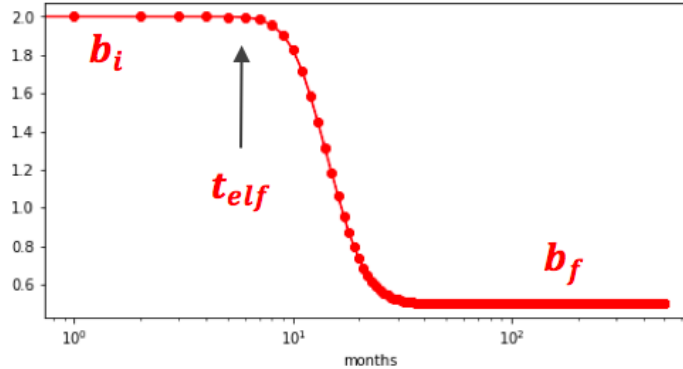
The second model is Transient Hyperbolic Model (THM). This model was introduced by Fulford and Blasingame in 2013. Different from traditional Arps model, they bring the physics meaning into the " b " parameter.

$$q_t = q_i(1 + b D_i t)^{-\frac{1}{b(t)}}$$

$$b_t = b_i - (b_i - b_f) \exp(-\exp(-c(t - t_{elf})) + e^\gamma)$$

$$c = \frac{e^\gamma}{1.175 t_{elf}}$$

Instead of using a constant b parameter, the THM model assumes the flow was started in the "linear" flow regime, with a constant $b_i = 2.0$. At the end of linear flow (at time t_{elf}), the b parameter smoothly transit to a constant value of $b_f < 1.0$. One example of the changing b parameter is shown below.



To implement the probabilistic Arps model, we need to estimate 4 parameters: q_i, D_i, b_f, t_{elf} .

The third model that we are implementing is the Jacobi θ Function model. This model was proposed by Gilding and Valko in 2018. This model is fully analytical based on solving a linear flow partial differential equation (PDE). The solution to the PDE is in the form of second Jacobi θ function.

$$q(t) = \frac{kA}{\mu L} (p_i - p_{wf}) \sum_{n=0}^{\infty} 2e^{-\frac{k}{\phi \mu c_i} \left[\frac{\pi}{2L} (1+2n) \right]^2 t} \cos \left[\frac{\pi x_i}{2L} (1+2n) \right]$$

$$\chi = \frac{\pi x_i}{2L} \quad \eta = \frac{\pi^2}{L^2} \frac{k}{\phi \mu c_i}$$

$$q(t) = q_i \theta_2(\chi, e^{-\eta t}) \quad \text{2nd Jacobi } \theta \text{ Function}$$

After grouping the variable, we only need to estimate three parameters for the Jacobi θ function model, which are: q_i, χ, η .

4. Probabilistic Model Formulation

We use Arps model as an example to show our implementation of the probabilistic model. The same formulation was applied to both THM and Jacobi θ models. We assume a non-informative prior distribution for each of the parameter. We pick normal distribution as the proposal distribution. The variance of the proposal normal distribution is selected in an trial and error manner.

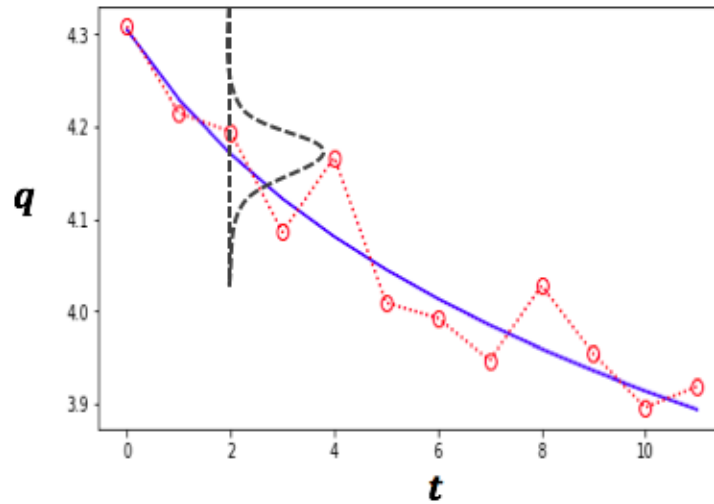
Parameters

q_i	Uniform Prior [0.1, 1,000,000]
D_i	Uniform Prior [0.1, 50]
b	Uniform Prior [0, 2]

Proposal Distribution

$q_{i,propose} \sim N(q_{i,n-1}, 1)$
$D_{i,propose} \sim N(D_{i,n-1}, 0.1)$
$b_{propose} \sim N(b_{n-1}, 0.1)$

We assumed the likelihood of observing a data point given the parameters follows normal distribution. This is illustrated in the figure below



The likelihood function formulation is shown below:

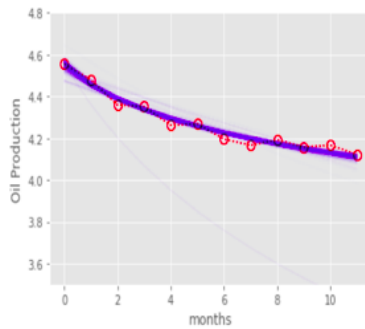
$$f(q|\theta, \sigma, t) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{[q_i - q_t(\theta, t_i)]^2}{\sigma^2}\right)$$

Finally, we show the acceptance ratio formulation. Since we are using a uniform distribution as prior distribution, the ratio of posterior distribution became the ratio of likelihood because constant prior.

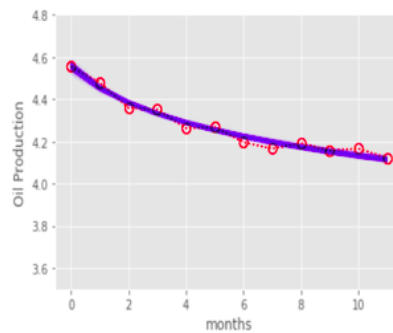
$$\alpha = \min \left[1, \frac{\pi(\theta_{proposal} | q) Q(\theta_{n-1} | \theta_{proposal})}{\pi(\theta_{n-1} | q) Q(\theta_{proposal} | \theta_{n-1})} \right] = \min \left[1, \frac{f(q | \theta_{proposal})}{f(q | \theta_{n-1})} \right]$$

5. Results

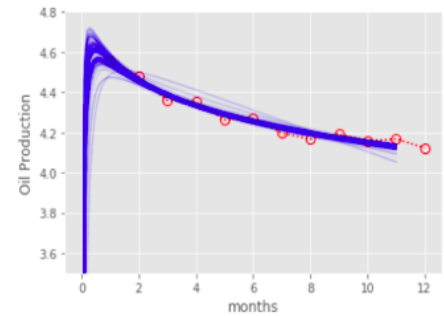
We implemented all three probabilistic models in Python. Detailed code is provided in the current director. We take the last 1000 samples of the parameters and showed the fitted line below.



Arps Model

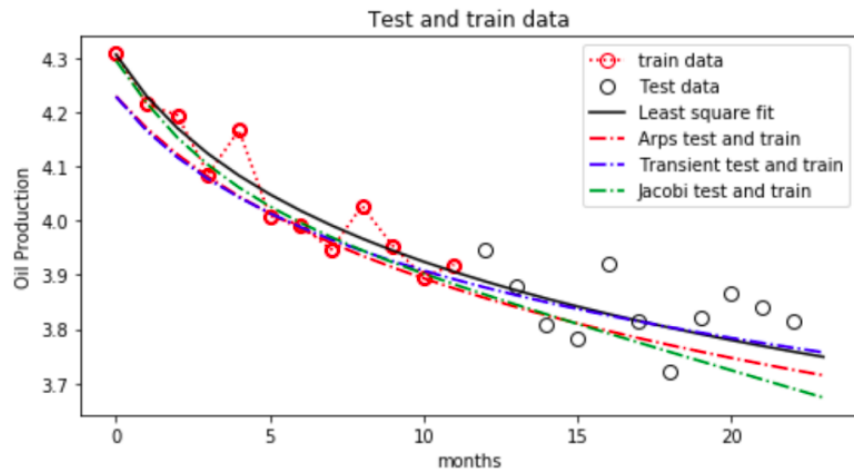


THM

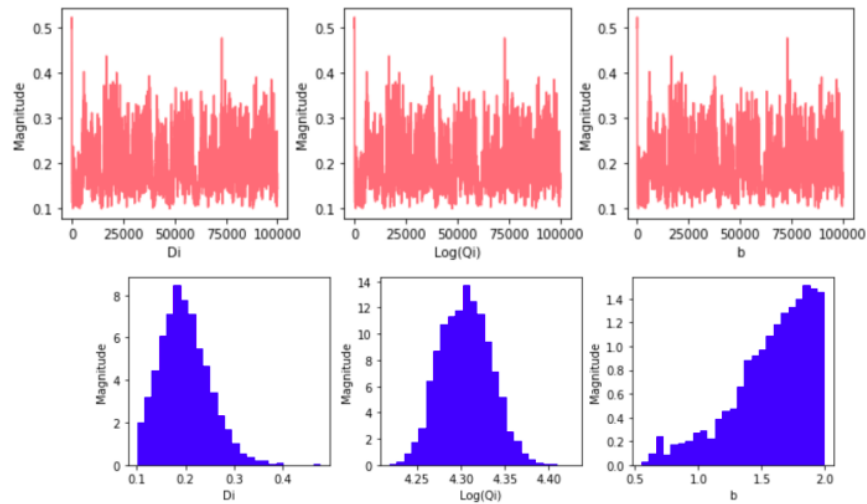


Jacobi θ

Also, we take the posterior mean of each parameter and plotted the fitted line together with original data. We also plotted the deterministic Arps model fitted by least square regression.

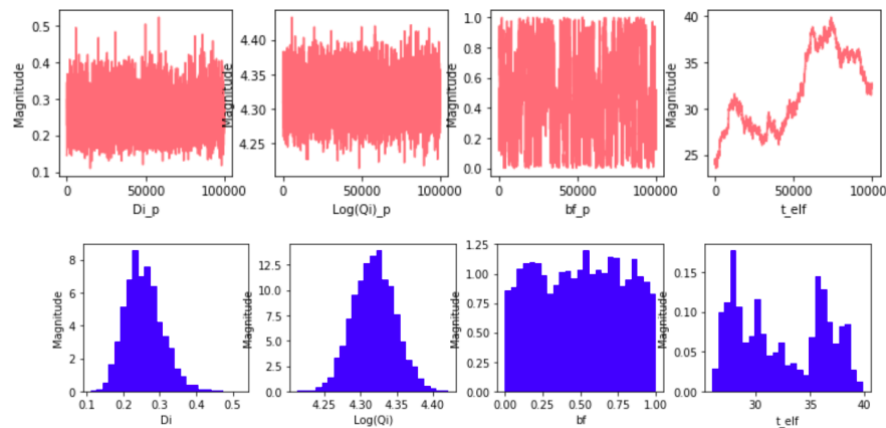


We first plot the mixing and posterior distribution of parameters in the Arps model.



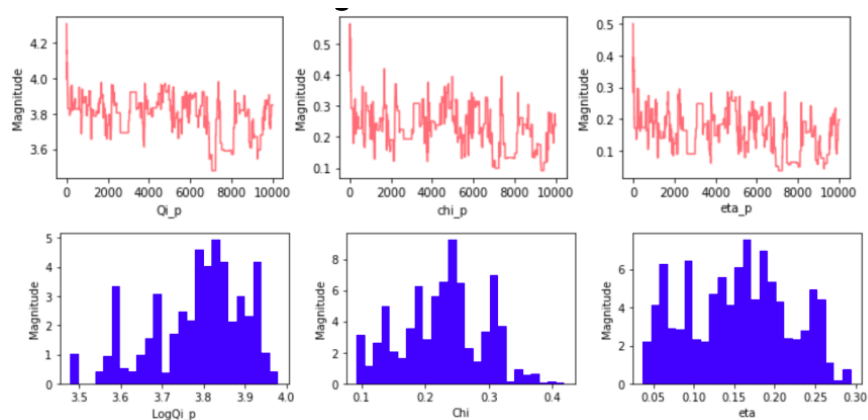
One can observe that the mixing of the chain for each parameter is good. Also, The posterior distribution of D_i , Q_i and b is in the shape of normal distribution. Here the distribution of b is truncated for b values larger than 2.0. This is because the largest possible b value is 2, which is the case of transient linear flow.

Next, we show the posterior distribution and mixing of THM model in the plot below.



One can observe the mixing for $Q_{i,p}$ and $q_{i,p}$ is good. And the distribution of $Q_{i,p}$ and $q_{i,p}$ is also in normal distribution shape. However, the distribution and mixing of $b_{f,p}$ and $t_{\text{elf},p}$ is very different. We argue this is because we train the model using only 12 months of production data. The "switching point", which is $t_{\text{elf},p}$ usually happens during the second year. Thus, our current training data set has little or no information about $t_{\text{elf},p}$ and $b_{f,p}$. This explains why $b_{f,p}$ is evenly possible throughout the range of $[0,1]$.

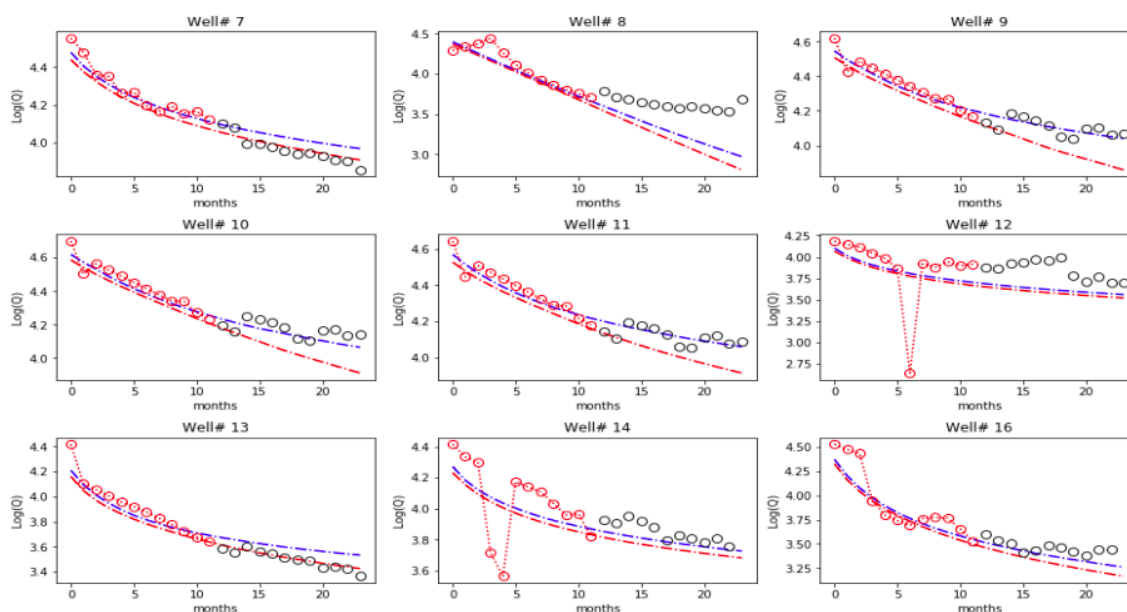
Lastly, we show the posterior distribution of parameters for the Jacobi θ model.



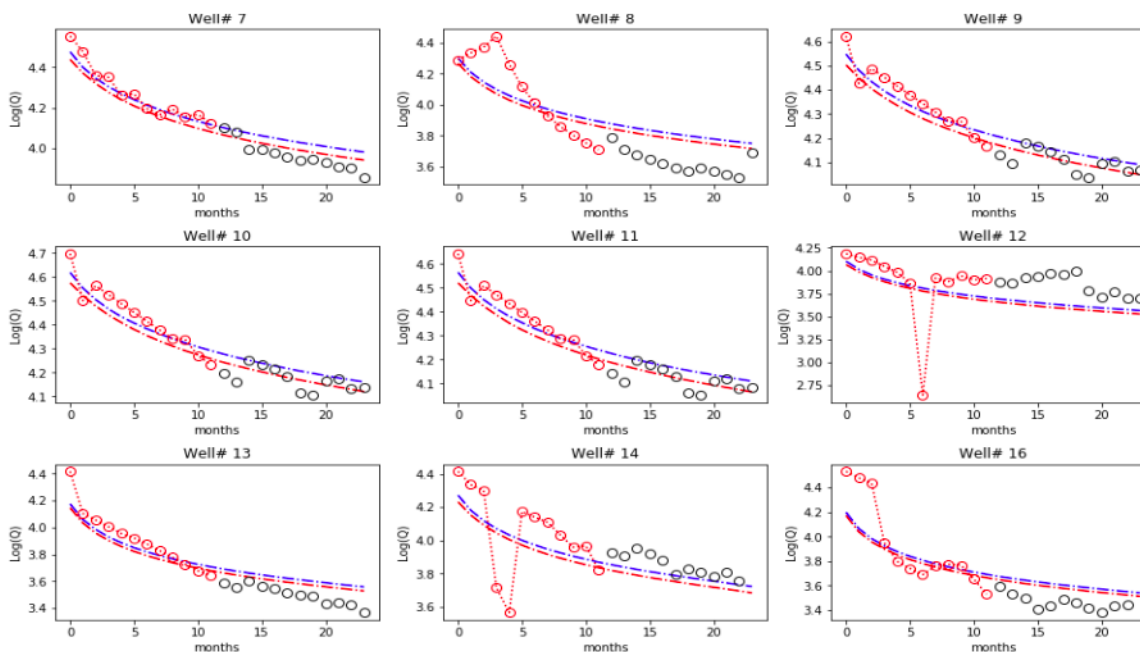
Different from previous two models, the Jacobi θ model is very computationally expensive, because it requires a special function which has infinite summation term. We can observe the Markov chain of parameters are converging. However, the mixing parameters are not as good as previous models, because we haven't found the best "step size" (variance) for the proposal distribution. Also, we only run the Markov chain until 10,000 samples, whereas previous models run up to 100,000 samples.

To test the prediction performance of each of the model, we randomly selected 9 wells and plotted the true second year production data and the P10, P90 prediction interval.

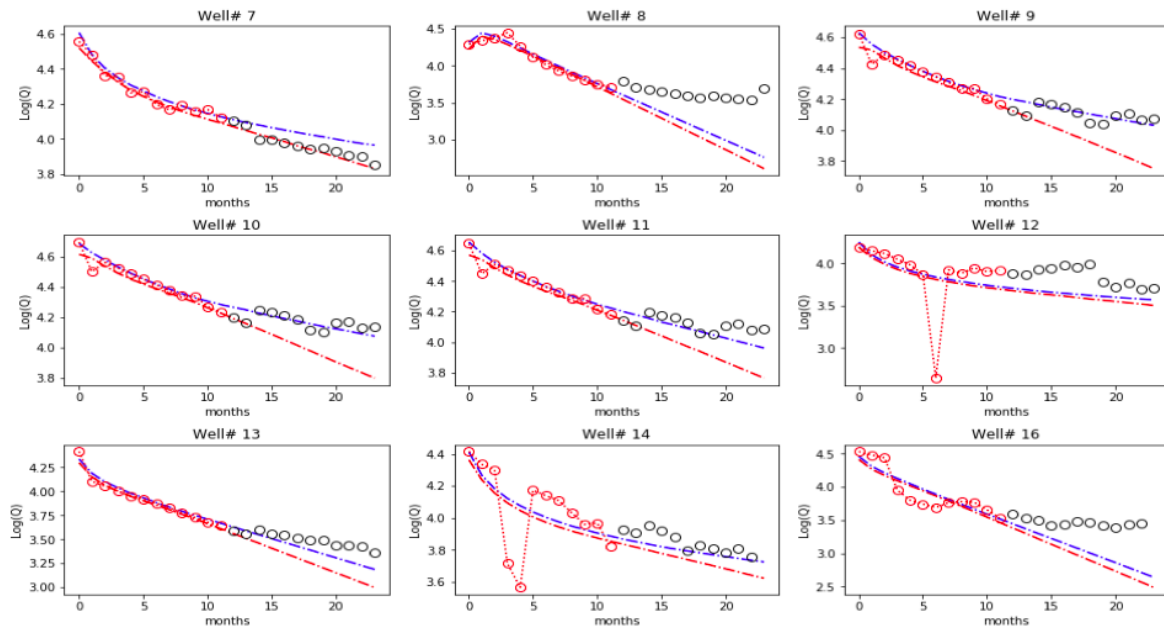
2nd Year Prediction: Arps



2nd Year Prediction: THM



2nd Year Prediction: Jacobi θ function



Comparing the prediction of the three models, we have the following observation:

- For majority of the wells (6 out of 9), the uncertainty of 2nd year production is "roughly" captured by the p10-p90 interval.
- Outlier/anomaly points exists in 2 out of 9 wells. The the prediction is largely influenced by the outlier.
- For both Arps and THM Jacobi θ models, the uncertainty (p10-p90 interval) increases for points further in the future.
- For THM model, the uncertainty interval does not increase. This is because the initial b-parameter value are fixed at 2, until it reached the time of t_{elf} .

6. Conclusion and Future Work

We implemented the probabilistic decline curve using three models. As shown in the results section, the MCMC sampling method generate a reasonable uncertainty interval. However, there are many aspects that we need to improve on the current model:

- A outlier detection and replacement method is much in need. As shown in the 9 well comparison plots, one single outlier will drag the prediction below the true data.
- Jacobi θ function model is very time consuming. It would be better if we can parallel the code and speed up the sampling process.
- It would also be useful to build a hierachical model. Because the parameters of neighbouring wells in the same oild field may be consider as sampled from a hyper-parameter distribution. Stan package should be incorporated into the hierachical model.