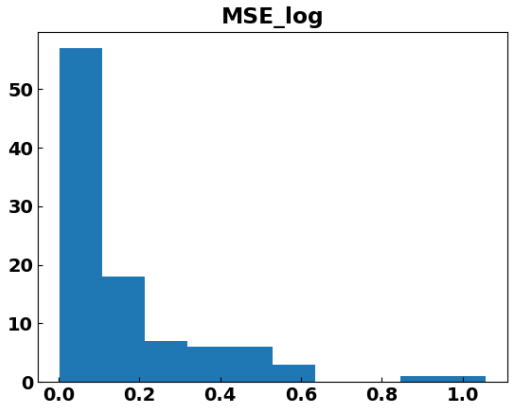
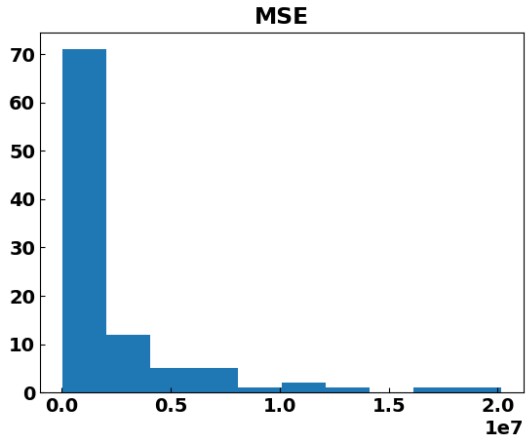
Decline Curve Analysis

1. Time for 100 wells regression:38.5s
2. MSE of log transformed production data



MSE of original production data

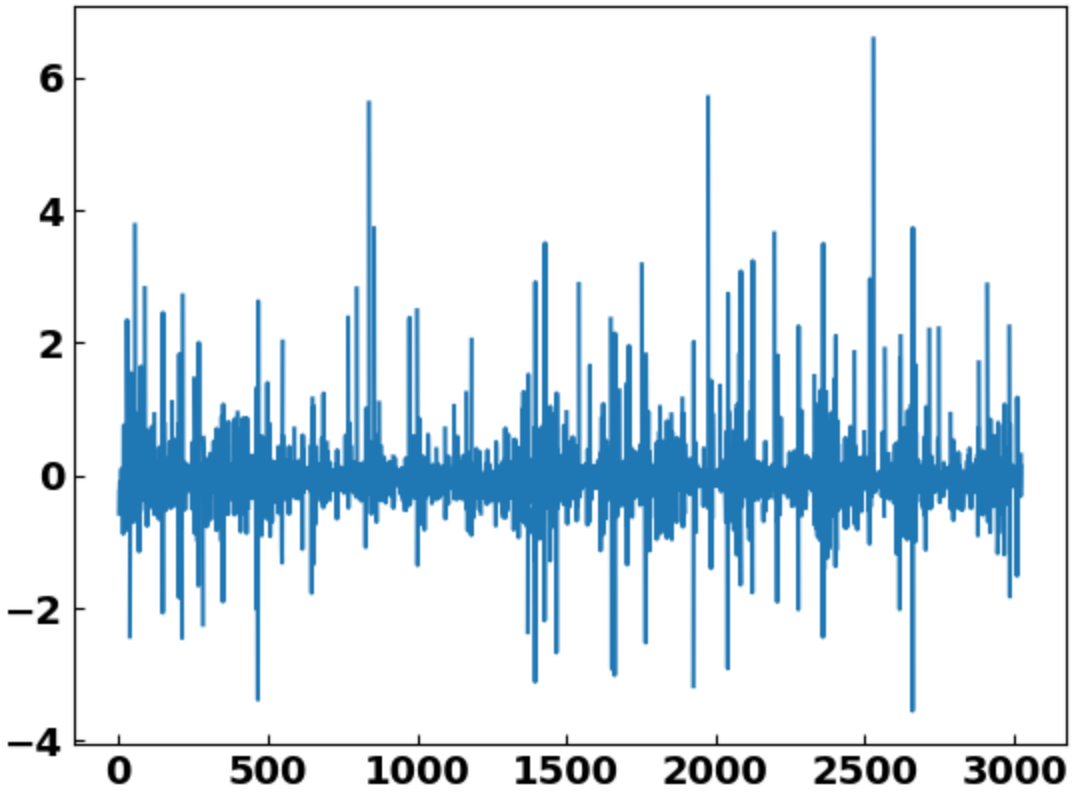


3. See attached image files in "./img/" repo
4. Loss function: mse of log transformed production data.

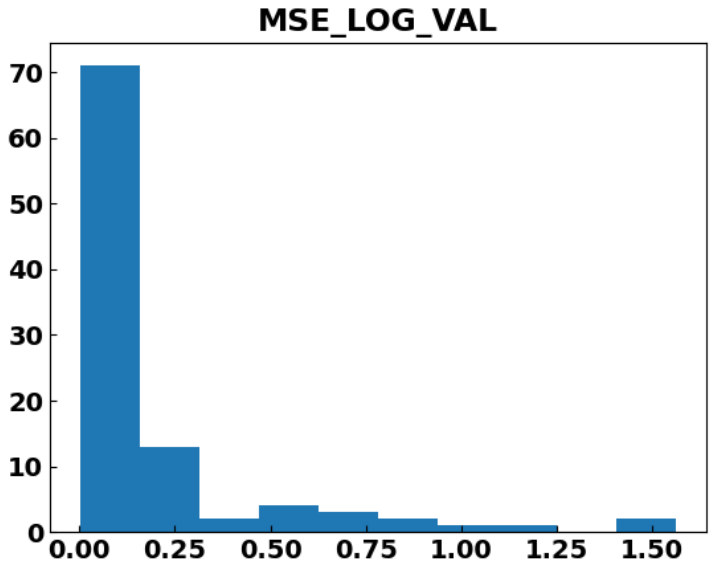
   Reason: if I don't use the original production data, the production forecast tends to bias to high

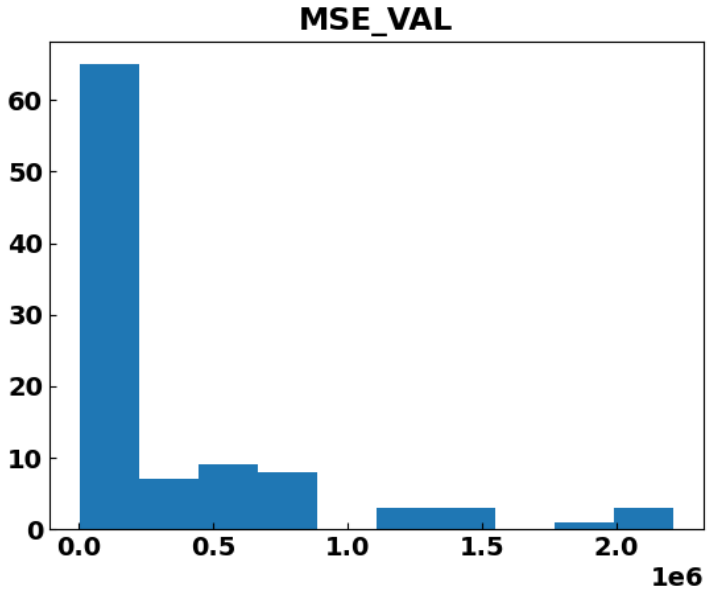
production. Log transformation also makes the fitting more robust to outliers, given that I removed the extremely low production data. We are interested in forecasting, which has a stronger correlation with more recent data, which are generally small.

5. Outlier Detection: I removed the extremely low values for the 1st iteration because low production usually indicates a well workover that are not considered in the decline curve analysis. For the 2nd iteration I am planning to check the 2nd and 3rd order difference of production, which are more stationary and remove outliers that have very high absolute values. The residual of first pass will also be examined.



6. Evaluation Scheme: As an MVP, I trained on the first 70% of the data and predict the rest 30%. As shown in the figure below, the error does not increase much, indicating that it is not overfitted.


MSE_LOG_VAL
70
60
50
40
30
20
10
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50


MSE_VAL
60
50
40
30
20
10
0
0.0
0.5
1.0
1.5
2.0
1e6