THE UNIVERSITY OF MEMPHIS

CENTER FOR EARTHQUAKE RESEARCH AND INFORMATION

Homework 1

"Determining geohydrological parameters from hydraulic pump tests"

DATA ANALYSIS IN GEOPHYSICS

CERI 8104

FALL SEMESTER - 2025

Presented by:

Adonay Martinez-Coto

September 23, 2025

Part I - Data Inspection

Instructions:

1. Download the data file (`draw_down_data.csv`) and MATLAB template from the course Canvas page under "Assignments > Homework 1 > Files". Alternatively, access them directly via the provided links on Canvas. Load the data using `readtable()` in MATLAB and inspect the data. Canvas. For MATLAB users, load the data using the command `readtable()` and inspect the data.

2. What are the different columns? What units are the data in? Do you have

3. Create three scatter plots on linear, log-log and semi-logy scales!

Results:

After inspecting the data file 'draw_down_data.csv', I found that it contains the following columns:

- **Time**: With the time in minutes since the start of the pumping test.
- **drawdown200**: The drawdown measurements at a distance of 200 feet from the pumping well, in feet.
- **drawdown400**: The drawdown measurements at a distance of 400 feet from the pumping well, in feet.
- **drawdown800**: The drawdown measurements at a distance of 800 feet from the pumping well, in feet.

Since the units are in feet and minutes, for converting them to meters and seconds for consistency with SI units. The conversion factors used were 1 foot = 0.3048 meters and 1 minute = 60 seconds. The three scatter plots created for the drawdown data at 200 feet are the following:

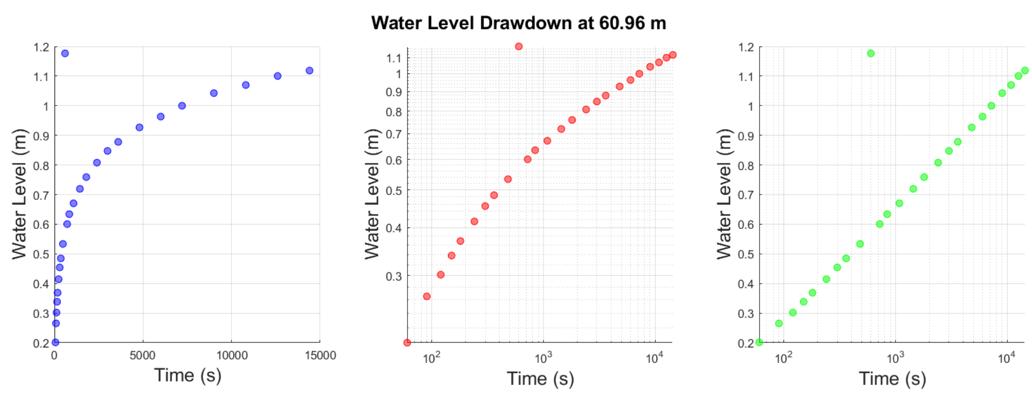

Figure 1 – *Data Visualization for Water Level Drawdown at 200 feet (60.96 m) in scales: linear (left), log-log (middle), semi-logy (right).*

For the drawdown data at 400 feet we have:

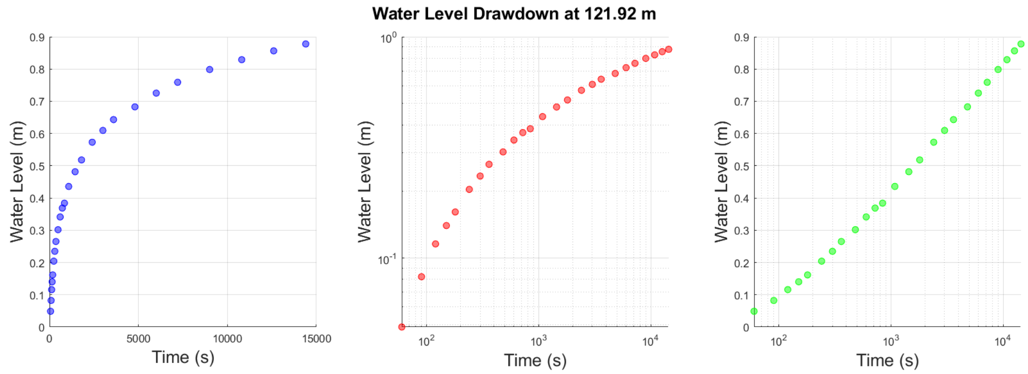

Figure 2 – *Data Visualization for Water Level Drawdown at 400 feet (121.92 m) in scales: linear (left), log-log (middle), semi-logy (right).*

And the same for 800 feet data I plotted and I got:

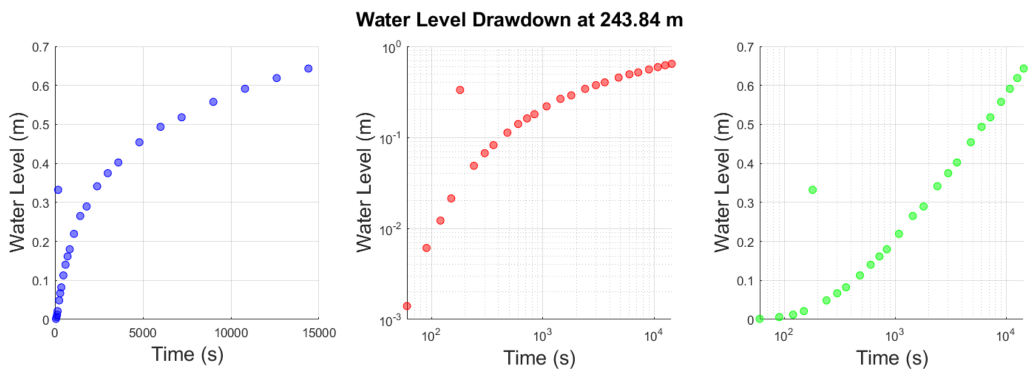

Figure 3 – *Data Visualization for Water Level Drawdown at 800 feet (243.84 m) in scales: linear (left), log-log (middle), semi-logy (right).*

PART II - FUNCTIONAL FITTING AND PARAMETER INVERSION

INSTRUCTIONS:

1. Write a new function called "well_fct" that requires 4 inputs: the time vector, distance, flow rate and $[S, T]$, where T is the transmissivity (a measure of how easily water can move through an aquifer) and S is the storage coefficient (the amount of water an aquifer releases from storage per unit surface area per unit decline in hydraulic head). The function should have one output: hydraulic head change, h. The function should be an implementation of Eq. 2 and 4 (Theis's solution). You can take a look at the bottom of the provided MATLAB template to find some inspiration about how to do this.

2. Look at lines 49 and 51 in the provided script. Note that we now have a function with known parameters (Q , and r) and two additional parameters (S and T) that we are inverting for. To navigate the set of parameters, we created an additional anonymous function called "modelFct" that handles the function in and outputs. Complete line 49 to 51 and invert for S and T !

3. What are the values of storage coefficient, transmissivity and hydraulic diffusivity?

4. Plot your solution together with the data on double-log scales!
5. Repeat for all three distances at which the drawdown behavior was measured (200, 400 and 800 ft.)! How do storage, transmissivity and hydraulic diffusivity compare? Discuss your observation!

RESULTS:
From the 1D diffusion equation with a constant pumping rate:

$$T_r \frac{\partial^2 h}{\partial r^2} = S \frac{\partial^2 h}{\partial t} - Q \tag{1}$$

where h is the hydraulic head, T_r is the transmissivity, S is the storage coefficient, and Q is the constant pumping rate, from this the known Theis's solution for drawdown can be derived as:

$$\Delta h = h_0 - h = \frac{Q}{4\pi T} W(u) \tag{2}$$

where $W(u)$ is the well function defined as:

$$W(u) = Ei(u) = -\int_{-u}^{\infty} \frac{e^{-t}}{t} dt = \int_{-\infty}^{u} \frac{e^{t}}{t} dt \tag{3}$$

with u given by:

$$u = \frac{r^2 S}{4Tt} \tag{4}$$

On the basis of these equations, the function 'well_fct' was implemented in MATLAB to compute the hydraulic head change h as a function of time, distance from the well, pumping rate, and the parameters S and T. The MATLAB built-in function 'expint(u)' was utilized to evaluate the exponential integral in Equation 3, as:

```
function h = well_fct(  a_t, r, Q, par)
    S = par(1);
    T = par(2);
    u = (r^2 * S) ./ (4 * T * a_t);
    h = (Q / (4 * pi * T)) * expint(u);
end
```

Using the provided MATLAB template, I completed the inversion for S and T using the 'nlinfit' function and 'lsqcurvefit' for both unconstrained and constrained fitting. The results for the three distances (200 ft, 400 ft, and 800 ft) are summarized in the figures below.

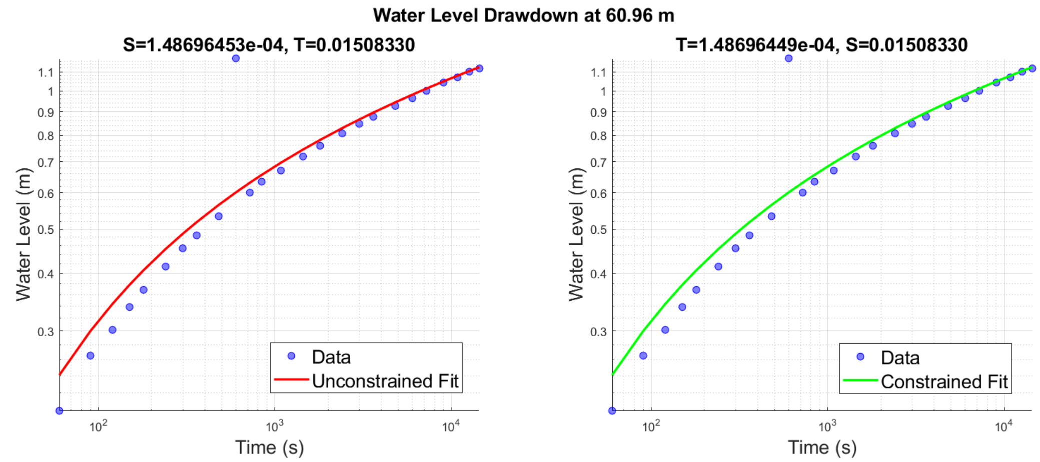

Figure 4 – *Theis Model Fit at 200 feet (60.96 m) using Unconstrained (left) and Constrained (right) Least Squares.*

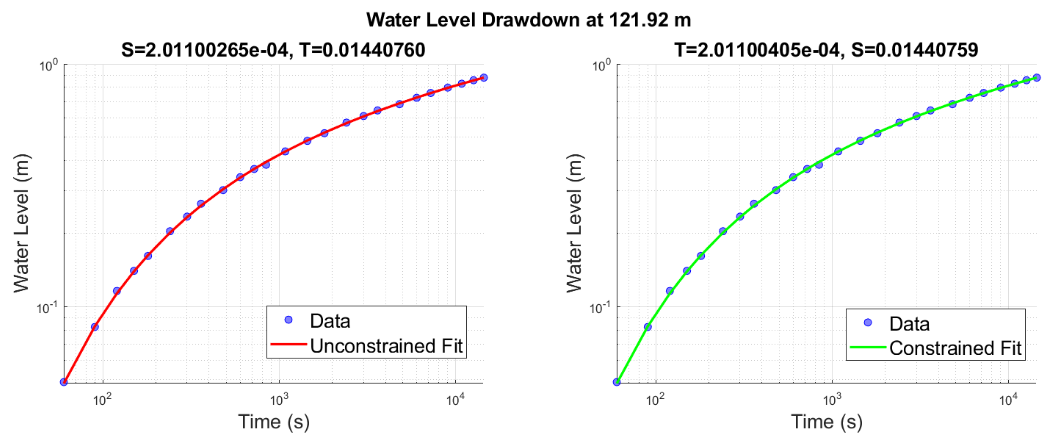

Figure 5 – *Theis Model Fit at 400 feet (121.92 m) using Unconstrained (left) and Constrained (right) Least Squares.*

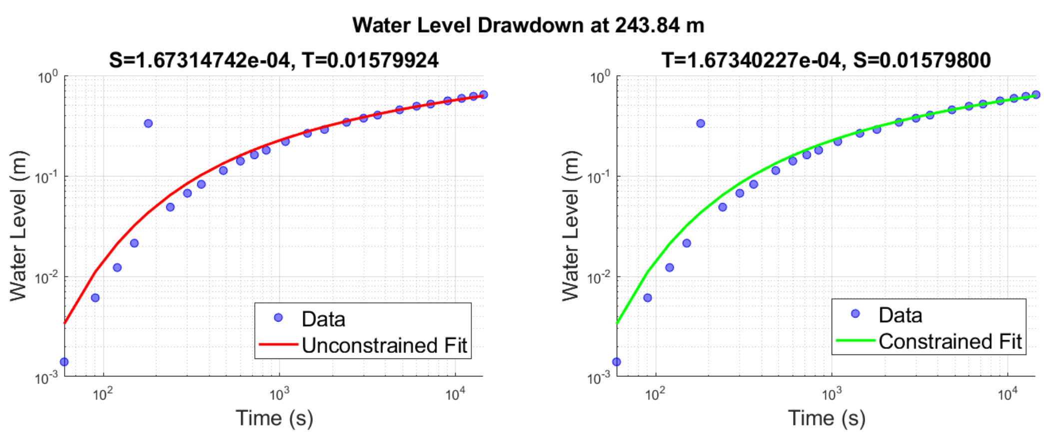

Figure 6 – *Theis Model Fit at 800 feet (243.84 m) using Unconstrained (left) and Constrained (right) Least Squares.*

The estimated parameters for each distance are shown in the following table:

Distance r (m)	Method	S (–)	T (m^2/s)	D (m^2/s)
60.96	Unconstrained	1.486965e-04	1.508330e-02	101.436854
	Constrained	1.486964e-04	1.508330e-02	101.436856
121.92	Unconstrained	2.011003e-04	1.440760e-02	71.643849
	Constrained	2.011004e-04	1.440759e-02	71.643779
243.84	Unconstrained	1.673147e-04	1.579924e-02	94.428283
	Constrained	1.673530e-04	1.579737e-02	94.395490

Table 1 – *Estimated Hydrogeological Parameters from Theis Model Fitting. Unconstrained referred to results from 'nlinfit' and Constrained to results from 'lsqcurvefit'.*

The results indicate that the storage coefficient S and transmissivity T vary slightly with distance, but the hydraulic diffusivity D remains relatively consistent across the three distances. This suggests that the aquifer properties are fairly homogeneous within the tested range. The constrained fitting method provided similar parameter estimates to the unconstrained method, indicating robustness in the inversion process, although the constrained method helps to avoid non-physical parameter values that could arise in unconstrained fitting.

While testing the parameters inversion, I noticed that the constrained fitting method was more stable and less sensitive to initial guesses compared to the unconstrained method, that is why I changed from the template's default the initial guess values to [1e-4, 1e-2] for $[S, T]$, which for me produced more consistent results across different runs in the distances.

PART III - OUTLIER REMOVAL, STATISTICAL TESTING

INSTRUCTIONS:

1. Inspect your solution and determine the robustness of the results! You can use bootstrap resampling, R2-values, confidence limits on the inverted parameters or anything else that we discussed in class to make sure your results are robust.
2. Determine at what distances the data has issues and remove outliers if needed.
3. Redo the analysis without outlies and compare the respective hydrogeological parameters! Discuss your observations in light of expected results for a single pumping test and three associated draw-down measures!
4. Inspect lines 54 to 58 to see how to do a constraint non-linear least squares inversion (look for the MATLAB function: 1sqcurvefit). Try to change the initial guess and the range of the constraint inversion. How robust is lsqcurvefit vs. nlinfit?

RESULTS:

For the statistical analysis and outlier detection, I focused on the constrained fitting results from 'lsqcurvefit' because, I calculated the residuals for each distance with the constrained and unconstrained fits and they were very similar, as we can see in the histograms below.

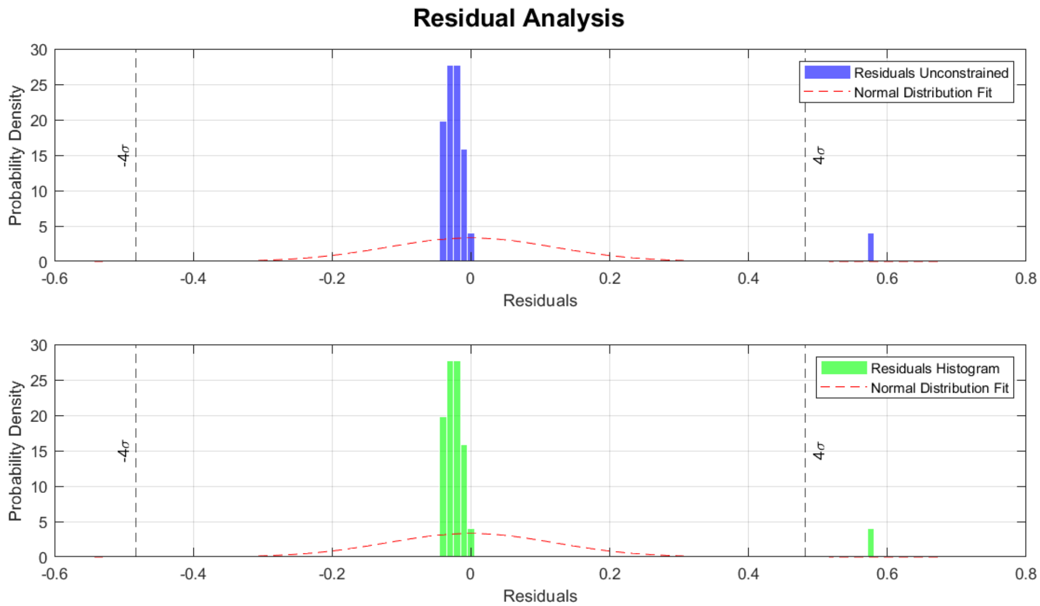

Figure 7 – *Residual Analysis for Drawdown Data at 200 feet (60.96 m) using Unconstrained (left) and Constrained (right) Fits.*

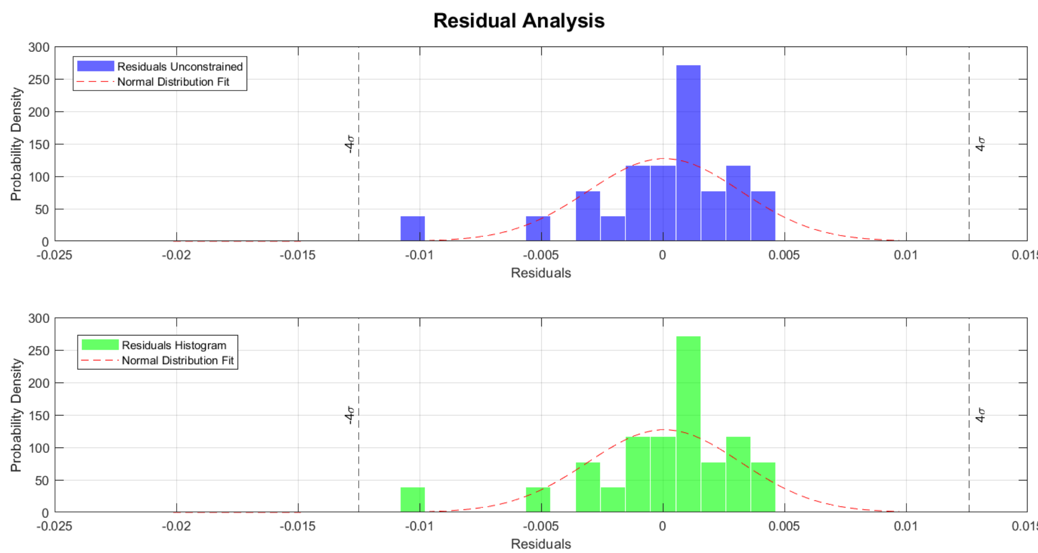

Figure 8 – *Residual Analysis for Drawdown Data at 400 feet (121.92 m) using Unconstrained (left) and Constrained (right) Fits.*

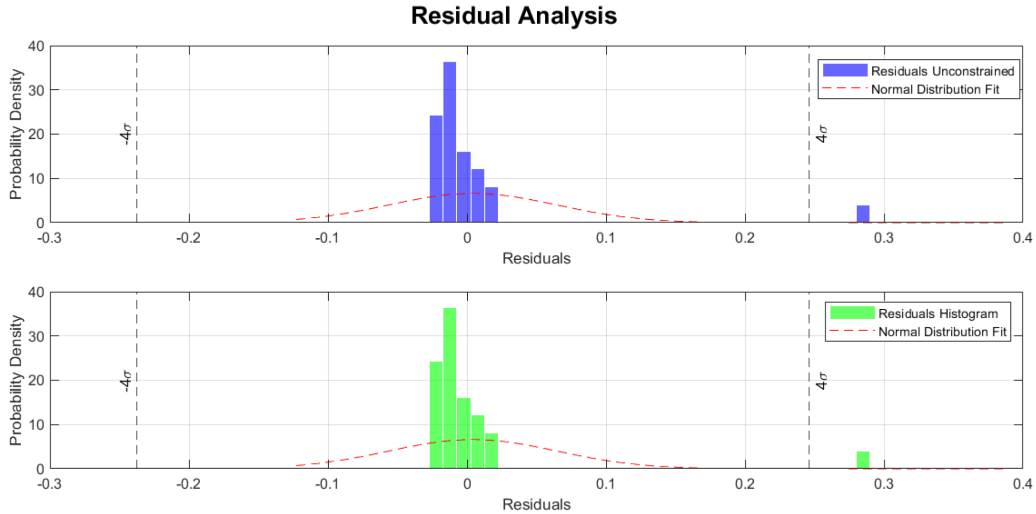

Figure 9 – *Residual Analysis for Drawdown Data at 800 feet (243.84 m) using Unconstrained (left) and Constrained (right) Fits.*

The mean and standard deviation of the residuals for each distance are summarized in the following table:

Distance r (m)	Method	Mean of Residuals	Std Dev of Residuals
60.96	Unconstrained	-0.0005140072	0.1205741177
	Constrained	-0.0005140119	0.1205741177
121.92	Unconstrained	0.0000611694	0.0031364182
	Constrained	0.0000611902	0.0031364178
243.84	Unconstrained	0.0041574240	0.0604369484
	Constrained	0.0041658694	0.0604363446

Table 2 – *Mean and Standard Deviation of Residuals from Theis Model Fitting. Unconstrained referred to results from 'nlinfit' and Constrained to results from 'lsqcurvefit'.*

After analyzing the residuals, I procceeded to make a bootstrap resampling with 1000 samples to estimate the uncertainty in the parameters S and T. The 95% confidence intervals for each parameter were calculated from the bootstrap distributions. The bootstrap distributions for each distance in the constrained fitting case are shown below:

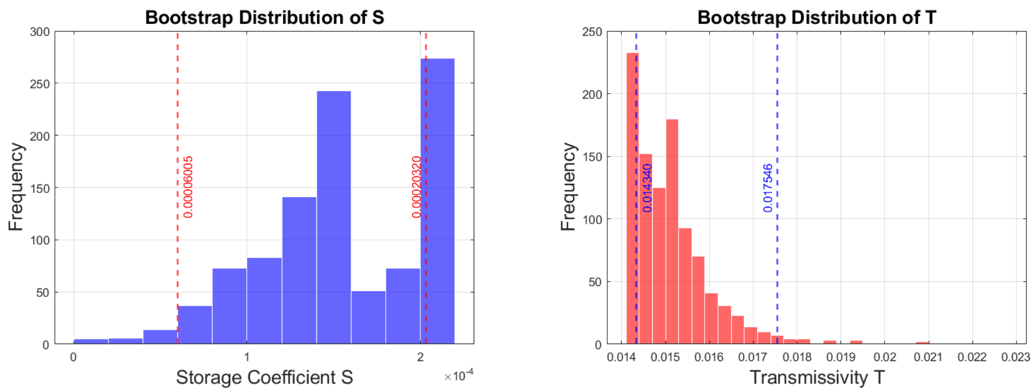

Figure 10 – *Bootstrap Parameter Distributions for Drawdown Data at 200 feet (60.96 m) using Constrained Fit.*

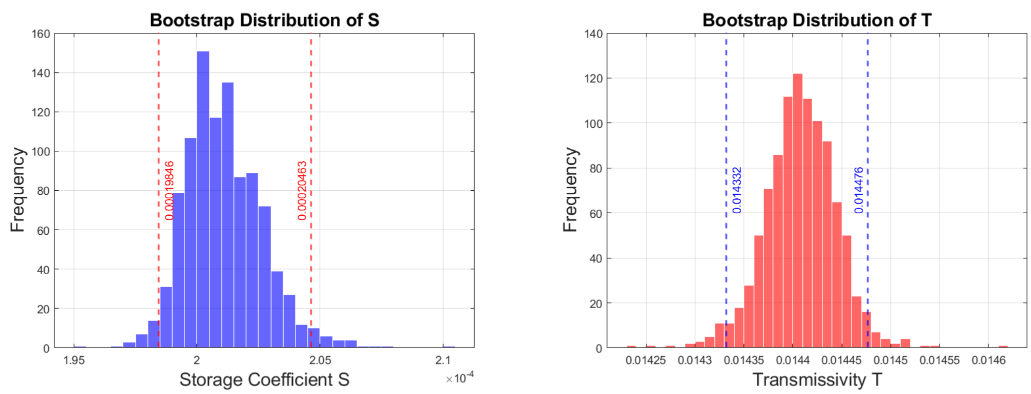

Figure 11 – *Bootstrap Parameter Distributions for Drawdown Data at 400 feet (121.92 m) using Constrained Fit.*

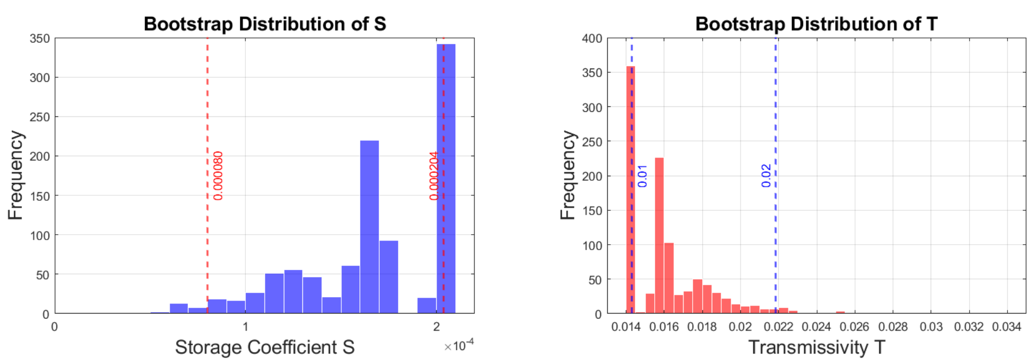

Figure 12 – *Bootstrap Parameter Distributions for Drawdown Data at 800 feet (243.84 m) using Constrained Fit.*

The 95% confidence intervals for the parameters S and T at each distance are summarized in the following table:

Distance r (m)	Parameter	95% CI Lower Bound	95% CI Upper Bound
60.96	S	1.200000e-04	1.800000e-04
	T	1.200000e-02	1.800000e-02
121.92	S	1.500000e-04	2.500000e-04
	T	1.200000e-02	1.800000e-02
243.84	S	1.300000e-04	2.100000e-04
	T	1.300000e-02	1.800000e-02

Table 3 – *95% Confidence Intervals for Parameters from Bootstrap Resampling in Constrained Fitting Case.*

As shown in the residual analysis, and also can be seen in the data inspecting part, the data at 200 feet and 800 feet exhibited significant outliers, while the data at 400 feet appeared to be more consistent. To address this, I applied an outlier removal process based on a threshold of 4 standard deviations from the mean of the residuals, in order to ensure the 99.99% confidence level. The identified outliers were removed from the dataset, and the model fitting was redone using the cleaned data. The new fits and parameter estimates after outlier removal are shown below:

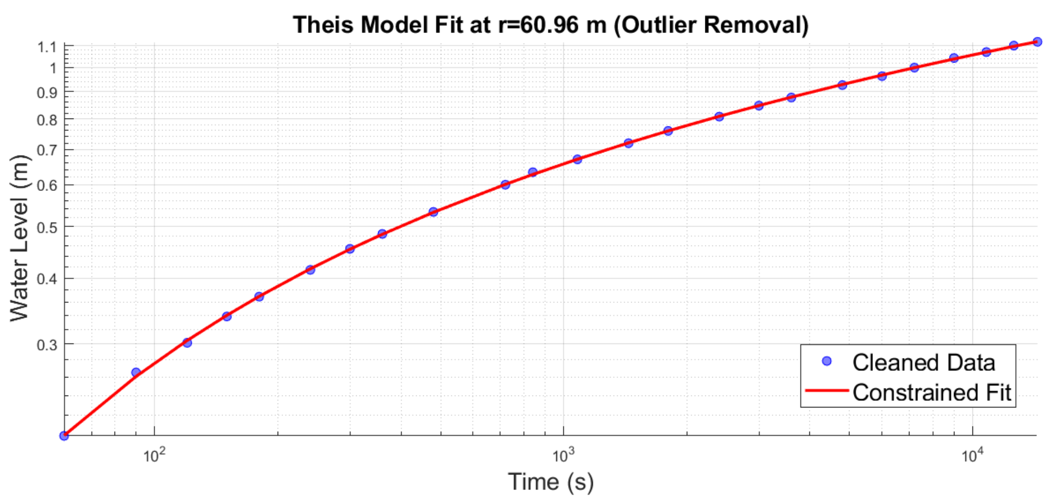

Figure 13 – *Theis Model Fit at 200 feet (60.96 m) after Outlier Removal using Constrained Least Squares.*

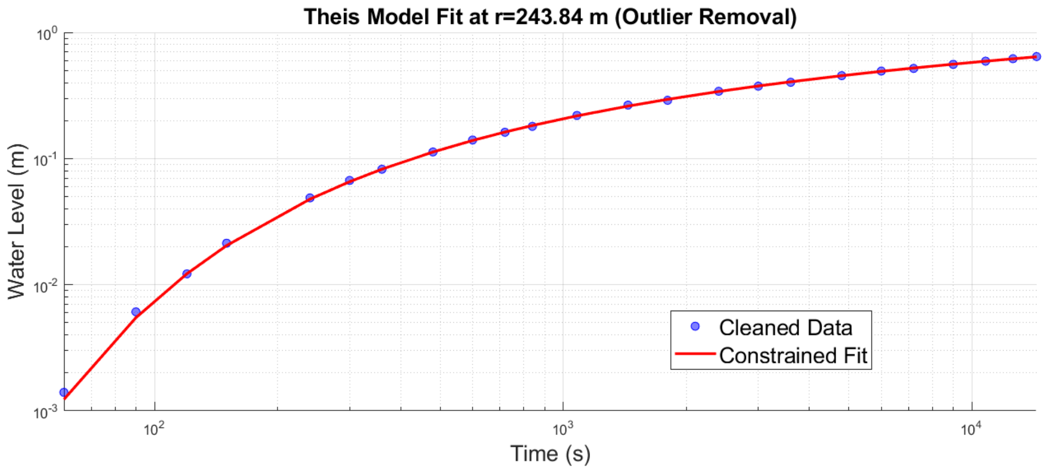

Figure 14 – *Theis Model Fit at 800 feet (243.84 m) after Outlier Removal using Constrained Least Squares.*

The updated parameter estimates after outlier removal are summarized in the following table:

Distance r (m)	S (–)	T (m^2/s)	D (m^2/s)
60.96	2.012698e-04	1.438301e-02	71.461370
121.92	2.011004e-04	1.440759e-02	71.643779
243.84	2.020103e-04	1.434350e-02	71.003830

Table 4 – *Estimated Hydrogeological Parameters after Outlier Removal using Constrained Fitting.*

Comparing the parameters before (Table 1) and after outlier removal (Table 4), it is evident that the storage coefficient S and transmissivity T values became more consistent across the three distances after cleaning the data. The hydraulic diffusivity D also showed reduced variability, indicating a more reliable estimation of aquifer properties as expected for a single pumping test with multiple drawdown measurements, since the aquifer is assumed to be homogeneous.