Analysis of MRI Data: Diffusion Signal Curve Fitting

We have measured the following points in individual pixels:

b-factor [s/mm^2]	S Normal	S Tumor	S Noise
5	13167.9	6383.1	101.3
238	8159.6	5143.9	72.7
471	5031.2	4337.7	67.1
704	3138.5	3685.0	66.3
937	2211.0	3322.1	62.9
1170	1592.5	2915.5	57.4
1403	1234.9	2635.2	58.7
1636	1001.6	2437.6	60.2
1869	817.4	2206.3	56.3
2102	691.3	2025.8	52.1
2335	603.7	1868.9	60.3
2568	527.3	1695.0	50.9
2801	552.2	1620.2	61.4
3034	99.2	1508.2	55.2
3267	463.5	1410.0	58.1
3500	455.5	1310.9	57.8

Initial Tasks:

1. Plot the signal decay curves with regular y-scale and logarithmic y-scale.
2. Does the noise change with changing signal level?
3. What type of signal data is typically measured, i.e., magnitude, real, imaginary or phase?
4. Which noise distribution and average do you expect for real data?
5. Which noise distribution and average do you expect for magnitude data? For more detailed reading see: *The Rician Distribution of Noisy MRI Data*, Magn Reson Med, 1995; 34(6): 910–914). One important conclusion from this article is that the magnitude signal follows a Rician distribution and that in the absence of signal the mean noise $\bar{M}$ within a region is governed by the following relations

   $$\bar{M} = \sigma\sqrt{\pi/2} = \sigma_M/\sqrt{(4/\pi - 1)} \approx 1.91\sigma_M.$$

   This allows us to estimate the signal-to-noise ratio (SNR= S/σ) and to verify if images have been filtered or not. Verify if the repeated noise measurements follow the rule and estimate the true SNR for $b = 3,500 s/mm^2$ in normal tissue.
6. Plot the diffusion related MR signal decay of water at 20 C (D = 2.0 μm^2/ms, $S(b) = S_0 \exp -bD$) and compare with the given data.
7. What is the SNR of the data?
8. How would the SNR improve, if you could use the sum of N pixels rather than a single data point.
9. What is the least SNR you need for a meaningful analysis.

Curve Fitting:

1. How do you deal with outliers?

2. Perform a least-square fit with a monoexponential function $S(b) = S_0 \exp(-bD)$ for b-factors below 1,000 s/mm^2 (Hint: use the logarithm of the signals to fit a linear function).

3. Use a non-linear fitting method to fit with a monoexponential function $S(b) = S_0 \exp(-bD)$ (Hint: use the Levenberg-Marquardt Method).

4. Compare the fits on plots with regular y-scale and logarithmic y-scale.

5. Perform a non-linar fit over the entire range of data with a monoexponential function $S = S_0 \exp(-bD)$. Note the degrees of freedom and sum of square residuals.

6. Perform a non-linar fit over the entire range of data with a stretched exponential function $S(b) = S_0 \exp(-(bD)^\alpha)$. Note the degrees of freedom and sum of square residuals.

7. Perform a non-linar fit over the entire range of data with a Taylor expansion of a Kurtosis fit $S(b) = S_0 \exp(-(bD) + \frac{1}{6}b^2D^2K)$. Note the degrees of freedom and sum of square residuals.

8. Perform a non-linar fit over the entire range of data with a biexponential function $S(b) = S_0(f_1 \exp(-bD_1) + (1 - f_1)\exp(-bD_2))$. How could you interpret D_1, D_2, f_1 and $f_2 = 1 - f_1$? Note the degrees of freedom and sum of square residuals.

9. (Difficult!) Show that the signal decay of a diffusion coefficient distribution according the Gamma distribution:

$$Gamma(k, \Theta) = \frac{1}{\Gamma(k)\Theta^k} D^{k-1} \exp(-D/\Theta)$$

   equals

$$S(b) = S_0 \frac{1}{(1 + b\Theta)^k}$$

10. Perform a non-linar fit over the entire range of data with a the signal decay for a gamma-distributed diffusion coefficient (see previous question). Calculate the mode, i.e., maximum $(k - 1)\Theta$ of the distribution. Note the degrees of freedom and sum of square residuals.

11. Examine visually how the curves fit the data.

12. Perform an analysis of variance (Fisher-test) to determine the best curve fit.

$$F = \frac{(RSS_1 - RSS_2)/RSS_2}{(DF_1 - DF_2)/DF_2}$$

    Determine the p-values associated with the F-values you find. There are useful F-test calculators on the internet, such as *http://www.statdistributions.com/f/* or *http://www.danielsoper.com/statcalc3/calc.aspx?id=7*

The null hypothesis is that the simpler model (the one with fewer parameters) is correct. The improvement of the more complicated model is quantified as the difference in sum-of-squares. You expect some improvement just by chance, and the amount you expect by chance is determined by the number of data points and the number of parameters in each model. The F test compares the difference in sum-of-squares with the difference you would expect by chance. The result is expressed as the F ratio, from which a p value is calculated.

The p value answers this question:

If the null hypothesis is really correct, in what fraction of experiments (the size of yours) will the difference in sum-of-squares be as large as you observed, or even larger?

If the p value is small, conclude that the simple model (the null hypothesis) is wrong, and accept the more complicated model. Usually the threshold P value is set at its traditional value of 0.05. If the p value is less than 0.05, then you reject the simpler (null) model and conclude that the more complicated model fits significantly better.