Medical MRI, Assignment 1

Anders Olsen (s154043)

April 2020

1 Abstract

Magnetic resonance spectroscopy (MRS) data may give information about the concentration of some metabolites in brain matter, potentially hinting at pathology. General Linear Modeling (GLM) is a well-known statistical technique used to estimate the weightings of individual signals that are linearly mixed using a priori information. This report describes the quantitation of MRS data using GLM, including the work-flow, central equations and potential deviations from actual metabolite concentrations. Code and results were discussed with fellow student Marie Garnæs.

2 Methods

2.1 Data Acquisition

The data were acquired in 2019 using a PRESS sequence on a $(1.5cm)^3$ voxel. The voxel was selected to be far away from white matter and subcutaneous lipids. A water reference scan and a water suppressed scan were run, acquiring an average of 48 free induction decays (FIDs) each. Fairly long repetition time $TR = 3s$ and somewhat short echo time $TE = 0.144s$ were used. An FID is an exponentially decaying (with rate *T2*) mix of sinusoids at different frequencies, depending on the metabolites present. The individual Larmor frequencies of molecules depend on their *chemical shift*, i.e. the electronic cloud surrounding the hydrogen atoms. Demodulation was set to the Larmor frequency of water, so the water reference scan is just an exponentially decaying signal of $0Hz$. In the water-suppressed scan, water is still the dominant signal, but unlike before, metabolites are visible.

Spectroscopy is usually analyzed in the frequency domain since the metabolites are more easily distinguished here. Often the scale used for visualization is distance in parts per million (ppm) from Tetramethyl-Silane. The two acquired spectra can be seen in Figure 1.

2.2 General Linear Modeling

GLM attempts to model a target signal, which here is the water-suppressed FID spectrum, using a number of potential variables called *regressors*. The regressors make up the columns of $\mathbf{X}$ in

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon} \tag{1}$$

and primarily consists of single spectra of metabolites that we know a priori to be present in grey matter. The residual signal, $\boldsymbol{\epsilon}$ is assumed to be Gaussian white noise, i.e. $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and elements of the signal are assumed independent from the model. It is important that the

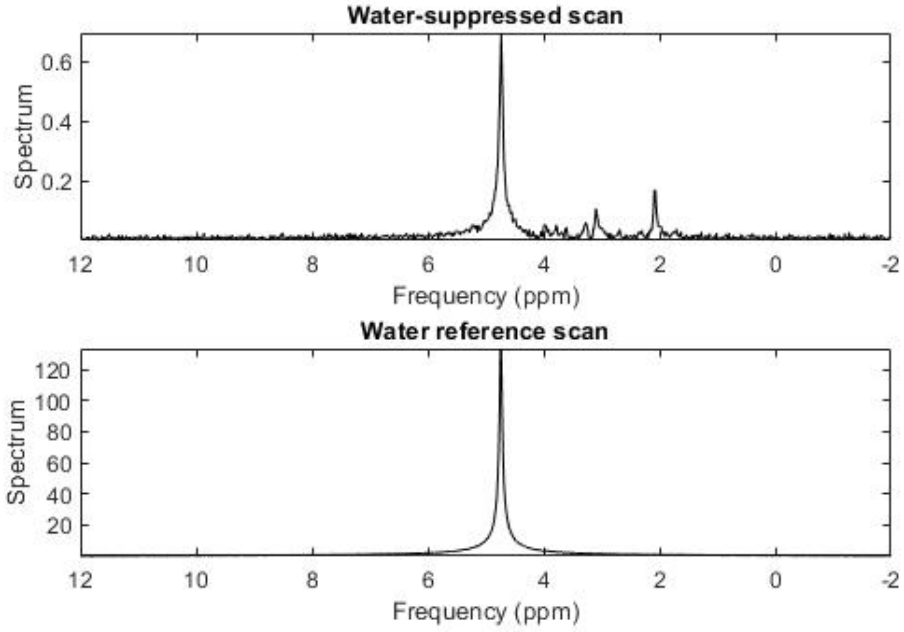

Figure 1: The raw data in ppm-frequency scale

metabolite regressors resemble each other. Therefore, they are constructed as frequency-shifted versions of the water reference spectrum, which is a good line-shape reference. Also, the peaks need to be broadened according to the $T2$-constants of the individual metabolites to correct for line width differences. Each regressor is, in the time domain, built as follows (frequency of water assumed to be 0):

$$S_m(t) = S_w(t) \exp^{-i2\pi f_m t} \exp^{-(R2_m - R2_w)t}$$

$R2_m$ and $R2_w$ are the decay rates of the metabolite and water, respectively. Other factors influencing the line shape of the metabolites like field inhomogeneity within the voxel and eddy currents are seen to be the same for all metabolites and therefore excluded from the equation. The decay rate difference is positive, and the metabolite FID can "blow up" at the end of the signal. This is prevented by multiplying an exponentially decaying function $\exp -ct$ on all FIDs. The constant c is estimated by trial-and-error to $c = 5$.

The richness of using GLM to model the signal spectra is encompassed in the fact that we can include a number of extra columns in the design matrix to account for extra signal deviations. The overall goal is to get the residual signal $\boldsymbol{\epsilon}$ to resemble white noise - if not, we cannot trust concentration estimates and their statistics. These extra columns, along with their parameter estimates $\hat{\beta}_j$ are generally considered nuisance factors, meaning that they are added to account for extra signal variation, but the parameter estimates are not used for anything except sharpen our trust in the computed metabolite concentrations. The two extra features added are derivatives of the regressors and a baseline contribution.

Adding the first and second derivatives of metabolite regressors as separate columns in the model can help account for small inaccuracies in model building. The first derivative represents a small shift of the metabolite frequency, and the second derivative can broaden or narrow the line shape reference, thus giving way for small inaccuracies in $T2$ values.

Baseline contributions also included in the model to account for slow signal variations. There are many ways to do this - using frequency-shifted broad Gaussian or sinusoidal functions or a variety of spline functions. Here we opted for a 2^{nd}-order polynomial representation. That is, we include a constant term, a linear term, and a quadratic term in the design matrix.

Finally, parameters $\hat{\boldsymbol{\beta}}$ are estimated using the pseudo-inverse:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

2.3 Concentration estimation

The concentration of each metabolite is given by the following formula:

$$C_m = \frac{\exp^{-TE/T2_w} N_w |S_m(t=0)| C_w}{\exp^{-TE/T2_m}(1-\exp^{-TR/T1_m}) N_m |S_w(t=0)|} \tag{3}$$

Here, N_m and N_w are the number of hydrogen atoms in the metabolite and water molecules, respectively, and $C_w = 45mM$ is the estimated water concentration in grey matter. The quantity $|S_m(t=0)|$ can be obtained by either summing the absolute metabolite spectra reconstructed from the respective columns of the design matrix $\mathbf{X_m}$ and w_m, or by back-transforming the reconstructed spectra to the time domain and finding the first value. The value $|S_w(t=0)|$ is easily found since this signal is already provided in the time domain as the water reference signal. Equation 3 is weighted according to relaxation time as the signal $S_m(t=0)$ is attenuated with time constant $T2_m$ and for repeated measurements (the signals are averaged over 48 measurements), there is also a *T1*-weighting.

All constants used for estimation of metabolite concentrations are collected in Table 1.

Table 1: Chemical shifts (CS), T1 and T2 values as well as number of hydrogen atoms (N) for each metabolite m: water, N-Acetyl Aspartate (NAA), Choline (Cho), and Creatine (Cr)

	CS_m [ppm]	$T1_m$ [ms]	$T2_m$ [ms]	N_m
Water	4.7		450	2
NAA	2.01	1500	360	3
Cho	3.20	1200	210	9
Cr	3.04	1600	85	5

3 Results

A model validation figure showing the true, estimated, and residual spectra can be seen in Figure 2.

Using the formula in Equation 3, we get the estimated concentration values in Table 2. Also in the Table are the expected concentrations, as found in REF.

4 Discussion

Model validation in this report was carried out purely graphically by figures as seen in Figure 2. Tests do exist that may confirm whether the residual spectrum is in fact Gaussian white noise, however, these are not in general used for simple models. Figure 2 shows that the residual

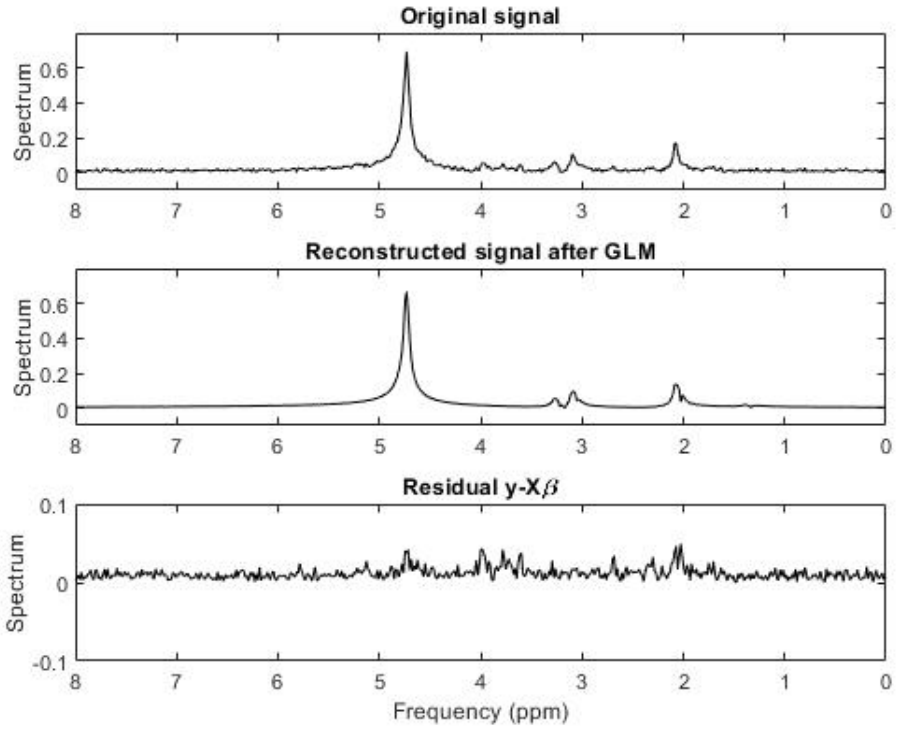

Figure 2: The original, reconstructed and residual spectra for model validation and interpretation.

Table 2: Estimated and expected concentration values for the metabolites N-Acetyl Aspartate (NAA), Choline (Cho), and Creatine (Cr)

	Estimated [mM]	Expected [mM]
NAA	12.2	12.0
Cho	1.4	2.0
Cr	6.1	7.9

signal (bottom subfigure) to a high degree resembles white noise. However, there are some small residual peaks residing at *ppms* 4.7, 4.0 and 2.05, as well as some minor peaks in between. This indicates that some water and some NAA has not been modeled perfectly.

This imperfection may be explained by several things. When using a model, such as GLM, that relies heavily on a priori information, it is necessary to question this. The residual peak at the ppm-location of NAA may be explained by the table-value of ppm being wrong.

The voxel chosen was supposed to be purely gray matter. However, as seen in Figure 3, which shows the selected voxel in the horizontal slice, the voxel contains some CSF-filled space between the two hemispheres as well as some white matter. Thus, the assumption of $C_w = 45mM$ may be wrong and directly alter the estimated metabolite concentrations. Also, the field inhomogeneity within the voxel may differ between metabolites, since water seems to be collected in the middle, and gray/white matter to the sides.

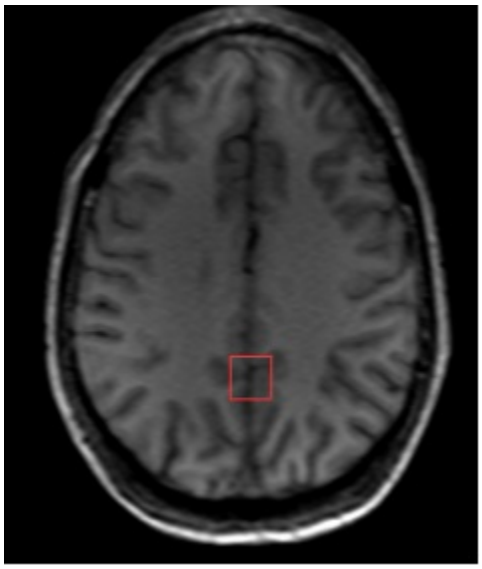

Figure 3: A poor quality image showing the horizontal slice of the subject's brain, as well as the chosen voxel location for spectroscopic analysis.

aaa