# Predicting Water Content by Fitting Data to Physical Models

Patric Boardman

September 2022

**Abstract**

This paper outlines a method to extract the water content from terrahertz (THz) based measurements of the refractive index. The code I developed focused on fitting complex data to complex mathematical functions, which stemmed from physical models of the refractive index of biological materials. Here, I outline these models along with their suitability, and how reliably they predicted the water content. The purpose of determining the water content was for cancer detection, with cancerous tissue being known to contain more water. The development of novel cancer detection methods would potentially have major applicability in cancer surgery, since an accurate determination of the precise location of tumours could potentially save the lives of many cancer patients in the future.

## 1 Introduction

Skin tissue is largely comprised of water, protein, fats, and minerals. It is known that tumorous tissue has been shown to have a larger concentration of water than healthy tissue. Therefore, we need to have a way to measure the percentage of water within the tissue. To do this, we measure a property known as the complex refractive index, $\tilde{n}$ using THz waves to probe the tissue sample. Here can be found full details of I extract the index from the measurements. For our analysis, we won't be using actual skin tissue but a material designed to mimic skin tissue, known as a phantom. For simplicity, these phantoms consist of water and gelatin, which was chosen since it is materially similar to skin tissue. The complex index comprises a real component and an imaginary component, $\tilde{n} = n + i\kappa$ The imaginary component $\kappa$ (also known as the *extinction coefficient*) is of particular interest, since it is known that water is highly absorbent at THz frequencies. Therefore, tissue with a low water content to have a much lower value of $\kappa$ than tissue with a high water content. Hence we postulate that this quantity can be used to some extent as a proxy for water content. To precisely correlate the complex refractive index with the water content of these phantoms, we must establish two important things. First, since the index is dependent on the frequency of radiation, we need to know this dependence in the case of pure water. Second, we must determine how we will combine the refractive indices of the individual components of the phantom, which include water and gelatin. With these components known, we can estimate the water content by running multiple parameter fits to the data. In the next section, we outline what the models are, how we combine them to create these models, and how we fit the data.

## 2 Theory and Methodology

### 2.1 Debye Model of Refractive Index for Pure Water

Due to our phantoms primarily comprising water by volume, we expect the dielectric response of the phantom to be similar to, though not quite the same as, that of water. Previous work in the field has shown [1] that in the THz regime, the frequency-dependent refractive index of water fits a two term Debye model with two characteristic relaxation times; $\tau_1$ being for the slow process and $\tau_2$ being that of the fast process. For a material obeying the double Debye model, the frequency dependent complex permittivity is given by

$$\tilde{\varepsilon}(\omega) = \varepsilon_\infty + \frac{\varepsilon_1 - \varepsilon_\infty}{1 + i\omega\tau_1} + \frac{\varepsilon_2 - \varepsilon_\infty}{1 + i\omega\tau_2} \tag{1}$$

where $\omega = 2\pi f$ is the angular frequency, $\varepsilon_\infty$ is the background frequency-independent permittivity, and $\varepsilon_1$ and $\varepsilon_2$ are empirically determined frequency-independent parameters that characterise the slow and fast relaxation processes respectively. Note that the complex refractive index is simply the square root of the complex permittivity, $\tilde{n}(\omega) = \sqrt{\tilde{\epsilon}(\omega)}$. For the code, we use the values of these parameters given by the most reliable data at room temperature [2], given in Table 1. With this, we now have the

| $\varepsilon_\infty$ | $\varepsilon_1$ | $\varepsilon_2$ | $\tau_1$(ps) | $\tau_2$(fs) |
|---|---|---|---|---|
| 3.5 | 78.4 | 4.9 | 8.2 | 180 |

Table 1: Published double Debye parameters we use in our models taken at 25$^{o}$C (298K)

frequency dependence of the index for a significant percentage of the phantom sample. A plot of this is shown in Figure 1.

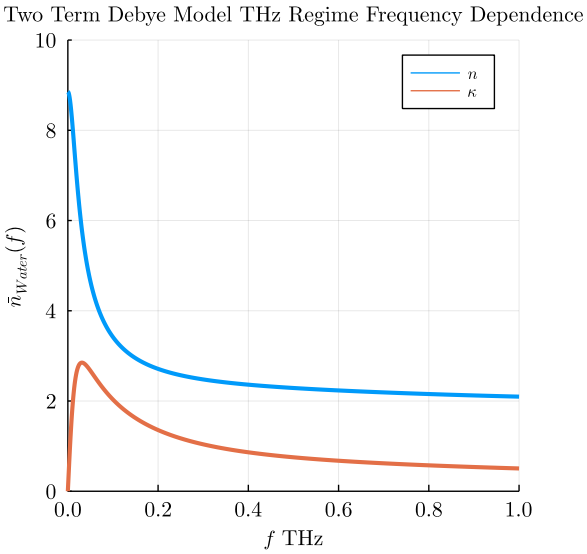


Figure 1: Plot of the frequency dependence for both the real and imaginary parts of the refractive index for pure water, with parameters used in Table 1.

## 2.2 Effective Medium Formulae

The next stage of this analysis is to combine the refractive index (or permittivity) of pure water, with that of the gelatin. In the THz frequency regime, the refractive index of pure gelatin materials has been found experimentally to not vary significantly [3]. We can therefore make the approximation that the frequency dependence of the entire phantom entirely comes from the volumetric fraction of water within the phantom. In this analysis, I compared three different effective medium models, each with varying degrees of complexity. Below we briefly outline these models and subsequently draw comparison between them. Note that most effective medium models usually work with complex permittivity, $\tilde{\varepsilon}$.

### 2.2.1 Weighted Average Model

The simplest way we can combine the permittivities of two different materials is to simply take an average of the two, weighted by the percentage of each material by volume. We define $f_i$, to be the volumetric fill fraction, with the subscript $i$ denoting the material. We can therefore write the effective complex permittivity $\tilde{\varepsilon}_{Eff}$ for water and gelatin as

$$\tilde{\varepsilon}_{Eff} = f_g\tilde{\varepsilon}_g + (1 - f_g)\tilde{\varepsilon}_w \tag{2}$$

with $f_i = f_g$ being the fill fraction of gelatin and the rest comprising water. While we expect this model to be accurate when there is largely one material dominant, the accuracy when there is a

balanced mixture is reduced because it does not take into account any physical effects. Models that do take this into account we outline below.

### 2.2.2 Maxwell-Garnett Model

To take into account more physical effects and obtain a more accurate model, we draw upon mean field theory concepts from electrostatics. For an inclusion of spherical particles within a host medium, a depolarizing field occurs close to the inclusions. This leads to the Clausius-Mosotti relationship between the permittivity $\varepsilon$ and the polarizability $\alpha$. This relation forms the basis for the Maxwell-Garnett model, and a full derivation can be found here [4]. The mixing equation reads

$$\tilde{\varepsilon}_{eff} = \tilde{\varepsilon}_w \frac{\tilde{\varepsilon}_w + \frac{1+2f_g}{3}(\tilde{\varepsilon}_g - \tilde{\varepsilon}_w)}{\tilde{\varepsilon}_w + \frac{1-f_g}{3}(\tilde{\varepsilon}_g - \tilde{\varepsilon}_w)} \tag{3}$$

. While we still expect the Maxwell-Garnett Model to be more accurate than a simple weighted average, it is known to be valid only for small fill fractions. However, since the phantoms we measured in the experiments, mostly consisted of water, this wouldn't be an issue for this model.

### 2.2.3 Bruggeman Model

The most physically accurate model, which is strictly derived from first principles and does not use any approximations, is the Bruggeman mixing Model, and is a generalisation of the Maxwell-Garnett Model. The equation reads:

$$\frac{1}{12}\left(6\tilde{\varepsilon}_w - 3\tilde{\varepsilon}_g + 9f(\tilde{\varepsilon}_g - \tilde{\varepsilon}_w)\sqrt{72\tilde{\varepsilon}_w\tilde{\varepsilon}_g + (6\tilde{\varepsilon}_w - 3\tilde{\varepsilon}_g + 9f(\tilde{\varepsilon}_g - \tilde{\varepsilon}_w))}\right) \tag{4}$$

We expect this equation to give the most accurate results when fitting, however fitting to an equation this complex will compromise performance. This presented a new purpose of comparing these different models; if we could use a model that is less computationally intensive, then the analysis could happen a lot quicker. However, if this happened too much at the cost of accuracy, we would choose to use the more advanced model.

### 2.2.4 Comparison Between Models

To see how these models differ, we initially compare how they change for different fill fractions. A plot of this comparison is shown in Figure 2.

We observe that for low and large fill fractions, the models agree, as expected. Throughout the entire $p_w$ domain, the values of the real component predicted by the three models are very similar, with the Maxwell-Garnett and Bruggeman models being almost indistinguishable. In the case of the imaginary part, the real part differs slightly more, by predicting higher values than the other two, though still not differing significantly. This tells us that the refractive index is indeed mostly determined simply by the water percentage, and the way in which they combine is less important.

# 3 Fitting

For each set of refractive index data (which corresponded to a THz measurement), we performed a fitting procedure to the predicted refractive index given by each of the models. The first task was to decide what our fitting parameters were. Since the task of this overall analysis is to extract the water content, we of course chose $f_g$, the volume fill fraction of gelatin, from which the water fill fraction $f_w$ can simply be obtained by subtracting it from 1. In order to reduce the complexity of the procedure, we aimed to minimise the number of free parameters, as is common with fitting. Therefore, we chose only one other free parameter, which was the complex gelatin permittivity $\tilde{n}_g$ which we assumed constant over this frequency range. Because the index is complex, we have to guess both the real and imaginary component, which would technically make this a three parameter fit, with three input guesses. The fitting algorithm used was a function called `curve fit`, which was part of the Julia package `LsqFit.jl`. For more information on the precise fitting procedure, click here. In essence, the fit function utilised the well-known Levenberg-Marquardt algorithm [5], and could fit given appropriate guess choices.

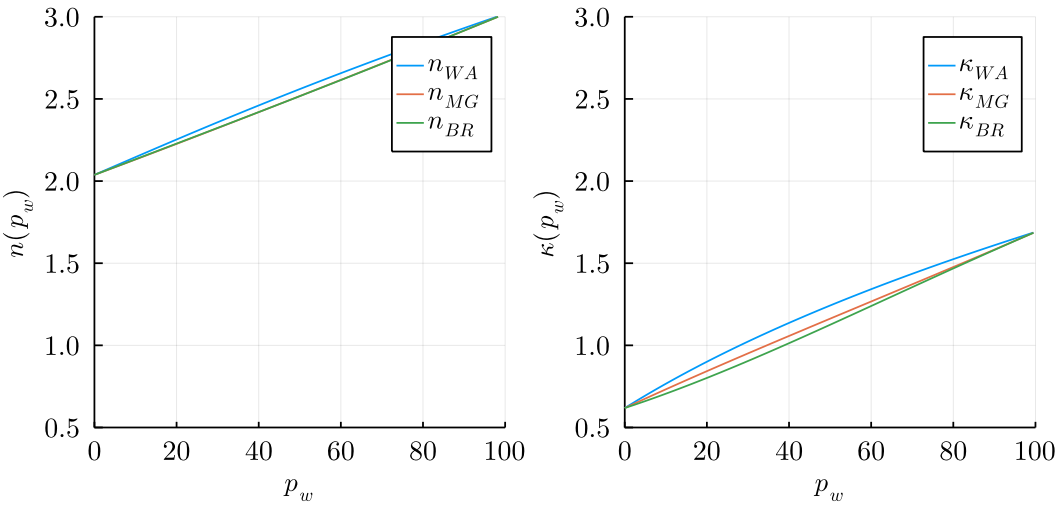


Figure 2: Comparison between effective medium mixing formulae, showing both the real and imaginary parts of the refractive index, at a frequency of 140GHz. We assume a frequency independent gelatin index at this regime $\tilde{n}_{gel} = 2.03 + 0.65i$ Note that rather than the fill fraction of the inclusion (in our case gelatin, $f_g$), we plot the water percentage $p_w = (1 - f_g) \times 100\%$

### 3.1 Fitting Inputs

Similar to the refractive index extraction process, it was necessary to specify an initial guess at the parameters. In our case, the two fit parameters are the fill fraction of gelatin, $f_g$, and the complex gelatin permittivity $\varepsilon_g + i\varepsilon'_g$. The Levenberg-Marquardt algorithm however does have the capability to fit complex data non-independently so it could be treated as if it was a two parameter fit. As well as guesses, we also needed to specify upper and lower bounds for all three parameters. These values were chosen in accordance with what was physically reasonable for the materials in which we were considering. If the procedure had extracted parameters that had the values of these bounds, then we can assume that the fit was not successful and the parameters were not physical. By restricting the bounds, and ensuring that the fitting procedure produced values that were strictly within the limits, we can reasonably assume that these were physically significant and corresponded to the actual values. Table 2 shows the initial guesses, upper and lower bounds for all three fit parameters.

| Parameter | Initial Guess | Lower Bound | Upper Bound |
|---|---|---|---|
| $f_g$ | 0.5 | 0 | 1 |
| $\varepsilon_g$ | 3.5 | 1 | 5 |
| $\varepsilon'_g$ | 0.5 | 0 | 4 |

Table 2: Initial guess, upper and lower bound values for each of the three fit parameters we chose to fit the data.

### 3.2 Fitting Range

It was also important to consider the range of frequency values for which we will fit the data. Since the probing THz signal does not extend across the entire 0.1-2THz domain that we are analysing, we must ensure that only those frequencies for which the refractive index value is reliable are included within the fit. This range will not be the same for any two measurements, however, since in these experiments

we were largely measuring the same thing, so it was reasonable to define a constant fitting range for all measurements. For this particular fit, we chose the range 0.3-1.5THz, which was the frequency range for which the probing pulse was the strongest.

# 4 Fitting Results and Discussion

The results of this fitting procedure are shown both in Figure 3 and Table 3.

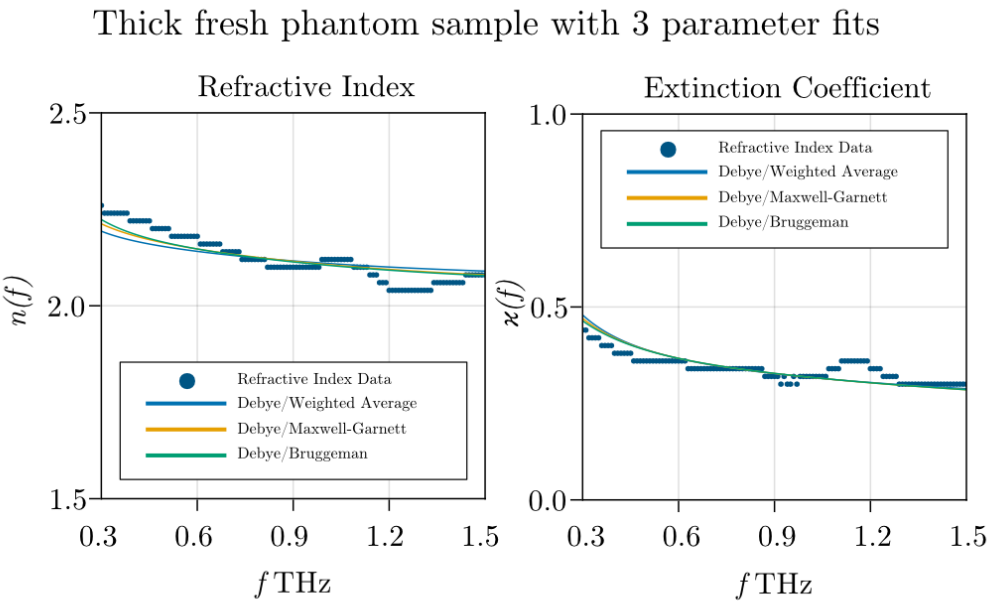


Figure 3: Plots showing the data, with fits to each of the three effective medium models, for both the real and imaginary parts of the refractive index. The results shown use the data from the THz measurement of a phantom taken on April 29th, and is the same exemplary piece of data used in the THz analysis paper.

| Model | $f_g$ | $\varepsilon_g$ | $\varepsilon'_g$ |
|---|---|---|---|
| Debye/Weighted Average | 0.751 | 4.424 | 1.085 |
| Debye/Maxwell-Garnett | 0.706 | 4.397 | 1.018 |
| Debye/Bruggeman | 0.708 | 4.384 | 1.018 |

Table 3: Extracted parameter values for each of the three models. The results shown use the data from the THz measurement of a phantom taken on April 29th, and is the same exemplary piece of data used in the THz analysis paper.

## 4.1 Analysis

It is clear from this analysis that each of the effective medium formulae were in close agreement, with the weighted average formula deviating slightly more so than the other two. This would make sense from a physical perspective as this is the only formula that does not take into account any electrostatic effects of each of the two media that are being mixed, such as molecule shape and size, which cause depolarizing effects. However, it is clear that the majority of the frequency dependence comes from the Debye model for water.
If we compare the predicted values in Table 3, we see that most models predict that the phantom

consists from between 70-75 % gel content, with the rest being water. Values of approximately this would intuitively make sense considering that this particular observation was taken some time after the phantoms had been exposed to the air. However, it appears that not all of the water has left the sample, which suggests that within the gel matrix, there exists water that is bound. This is consistent with existing research into this area [3], which introduces the concept of bound water. Because of the nature of this water, it is possible that we would have to increase the complexity of the models towards a three component mixture as opposed to simply a two component mixture. This would require modifying the three formulaes to account for this, but it would be possible. However, in terms of accuracy, it is likely that there are other sources of uncertainty within this analysis which would have to be resolved before increasing the complexity. Several evaluations of several data sets would need to produce equally compelling results before going to this stage.

### 4.2 Next Stages

The next phase of this analysis is to compare these results with what we can observe experimentally using a different spectroscopic technique which uses Raman spectroscopy. In essence, this inspects a portion of the wave-number spectrum and observes changes, which correlate to the frequency of the vibrations of water molecules. If the values are in agreement, then this would be one step further into determining the water content within skin tissue, leading on to exciting novel applications of imaging and potential early cancer diagnosis.

## 5 Acknowledgements

I would like to acknowledge my supervisor, Professor Euan Hendry from the University of Exeter physics department, for the exceptional amount of assistance given to me on this project and the co-development of this code. I am very grateful for the assistance I was given whilst working on this project. The experience I gained during that year I have no doubt will provide invaluable for future endeavors. I would also like to acknowledge Dr. Sonal Saxenna, a post-doctorate who was working closely alongside me on this project. Many of the measurements in this experiment were conducted by her, without which this analysis would not be possible.

## References

[1] E. Pickwell, B. Cole, A. Fitzgerald, V. Wallace, and M. Pepper, "Simulation of terahertz pulse propagation in biological systems," *Applied Physics Letters*, vol. 84, no. 12, pp. 2190–2192, 2004.

[2] O. Cherkasova, M. Nazarov, M. Konnikova, and A. Shkurinov, "Thz spectroscopy of bound water in glucose: direct measurements from crystalline to dissolved state," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 41, no. 9, pp. 1057–1068, 2020.

[3] Z. Tianmiao, Z. Maria, V. Anna, P. Aleksandr, F. Maria, N. Ravshanjon, K. Anna, D. Petr, U. Mayya, and K. Mikhail, "Terahertz optical and mechanical properties of the gelatin-starch-glycerol-bentonite biopolymers," *Journal of Biomedical Photonics & Engineering*, vol. 6, no. 2, p. 20304, 2020.

[4] V. A. Markel, "Introduction to the maxwell garnett approximation: tutorial," *JOSA A*, vol. 33, no. 7, pp. 1244–1256, 2016.

[5] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*, pp. 105–116, Springer, 1978.