# USING GAUSSIAN PROCESS MODELS FOR NEAR-INFRARED SPECTROSCOPY DATA INTERPOLATION

Darren Leamy
Biomedical Engineering Research Group
Department of Electronic Engineering
National University of Ireland Maynooth
Maynooth,
Co. Kildare, Ireland
email: dleamy@eeng.nuim.ie

Tomás Ward
Biomedical Engineering Research Group
Department of Electronic Engineering
National University of Ireland Maynooth
Maynooth,
Co. Kildare, Ireland
email: tomas.ward@eeng.nuim.ie

Juš Kocijan
Jozef Stefan Institute
Ljubljana, Slovenia
and
University of Nova Gorica
Nova Gorica, Slovenia
email: jus.kocijan@ijs.si

## ABSTRACT

Gaussian Process (GP) model interpolation is  used extensively in geostatistics.  We investigated the effectiveness of using GP model interpolation to generate maps of cortical activity as measured by Near Infrared Spectroscopy (NIRS). GP model interpolation also produces a variability map, which indicates the reliability of the interpolated data.  For NIRS, cortical hemodynamic activity is spatially sampled.  When generating cortical activity maps, the data must be interpolated. Popular NIRS imaging software HomER uses Photon Migration Imaging (PMI) and Diffuse Optical Imaging (DOI) techniques based on models of light behaviour to generate activity maps. Very few non-parametric methods of NIRS imaging exist and none of them indicate the reliability of the interpolated data.  Our GP model interpolation algorithm and HomER produced activity maps based on data generated from typical functional NIRS responses.  Image results in HomER were taken as the bench mark as the images produced are commonly considered to be representative of the true underlying hemodynamic spatial response.  The output from the GP approach was then compared to these on a qualitative basis. The GP model interpolation appears to produce less structured image maps of hemodynamic activity compared to those produced by HomER, however a broadly similar spatial response is compelling evidence of the utility of GP models for such applications. The additional generation of a variability map which is produced by the GP method may have some utility for functional NIRS as such information is not explicitly available from standard approaches. GP model interpolation can produce spatial activity maps from coarsely sampled NIRS data sets without any knowledge of the system being modelled. While the images produced do not appear to have the same feature resolution as photonic model-based methods the technique is worthy of further investigation due to its relative simplicity and, most intriguingly, its generation of ancillary information in the form of the variability map. This additional data may have some utility in NIRS optode design or perhaps it may have application as additional input for response classification purposes. This GP technique may also be of use where model information
is inadequate for DOI techniques.

## 1   Introduction

Near-infrared spectroscopy (NIRS) is a non-invasive method for measuring localised hemodynamics in the brain.  This technique uses light at multiple near-infrared wavelengths to measure concentrations of oxygenated haemoglobin (HbO) and deoxygenated haemoglobin (HbR). Since there is a relationship between cerebral vascular responses and neural stimulation, the recorded HbO and HbR signals give an indirect measurement of neuronal activity.

NIRS only interrogates a small area of cortex so interpolation of the data is required to produce a spatial activity map.  The most popular method for producing images from NIRS data is through a linear approximation to the photon diffusion equation [1].  This method is implemented in the NIRS data analysis package HomER. Cruder, more familiar, methods of interpolation have been applied to NIRS signals before [2]. However, the method employed by HomER is widely trusted to be the highest standard of NIRS imaging.

The Gaussian process model is a probabilistic, non-parametric black-box model.  It differs from most of the other black-box modelling approaches as it does not try to approximate the modelled system by fitting the parameters of the selected basis functions but rather searches for the statistical relationship among measured data.

Gaussian process models are closely related to approaches such as Support Vector Machines and especially Relevance Vector Machines. Typically, Gaussian process models have been explored for various applications as a method for classification or regression. Various interesting applications (e.g. [3], [4] in medicine and bioengineering fields) have exploited different properties of Gaussian pro-

cess models. In the field of geostatistics Gaussian Process regression models are used for probabilistic analysis of data and are more commonly known under name Kriging. It is the latter application of GP which we are applying here.

Here, we apply the GP interpolation process to NIRS data and compare results to those obtained in HomER. GP model interpolation is unique in that it produces an interpolated magnitude map and also a variance map. The levels of variance at each interpolation point give an indication of how reliable the estimated magnitude value is. This is the first time as far as the authors are aware that a Kriging approach has been applied to data generated by near infrared spectroscopy methods and it may open up new vistas of research at least for non-parametric approaches to image construction in this domain.

## 2 Background

### 2.1 NIRS Theory

In NIRS, near-infrared wavelengths of light are shone on to the scalp. Some of the emitted photons are absorbed and some are scattered. Some of the scattered photons are 'back-scattered' so that they exit the scalp again. A photodetector can be placed on the scalp to record the intensity of the transmitted light exiting at that point. A pairing of light source to photodetector is called an 'optode'. The light recorded by a photodetector is predicted to have travelled in a roughly banana-shaped path [5]. By controlling the separation of the light source and photodetector, the expected depth of the photon path can be controlled. This source-detector separation is set so that a small area of cortex is interrogated by the photons.

The intensity of the back-scattered light is related to the amount of light that has been absorbed and scattered. The absorption spectrum of HbO and HbR has been recorded before [6]. Using the absorption coefficients of HbO and HbR at the two wavelengths used, the modified Beer-Lambert law can be used to calculate concentrations of HbO and HbR along the path [7]. This indicates the changes in HbO and HbR in the interrogated area of cortex. A more detailed introduction to NIRS can be found elsewhere [8].

### 2.2 The Gaussian process model

The Gaussian process model is an example of the use of a flexible, probabilistic, non-parametric model with uncertainty predictions. It fits naturally in the Bayesian modelling framework in which instead of parameterizing mapping function $f(\mathbf{x})$, a prior is placed directly on the space of possible functions $f(\mathbf{x})$ which could represent the nonlinear mapping from input $\mathbf{x}$ to output $y$. This prior represents the modeller's beliefs about the mapping, usually involving smoothness assumptions. This prior is combined with the likelihood of the identification (training) set of $N$ observed input-output data pairs, $\{\mathbf{x}^i, y^i\}_{i=1}^N$, to provide us with the
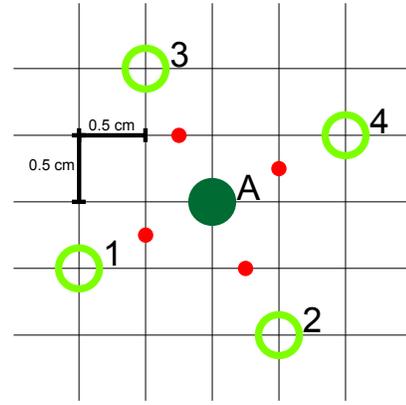


Figure 1. Optode configuration. The solid dark green circle labelled A indicates the location of the light source. The open circles, labelled 1-4 indicate the locations of the photodetectors. The solid red circles indicate the midpoints of the paths, which are used as the data source locations for GP model interpolation. Points on the grid are 0.5 cm apart.

posterior distribution for model predictions. $\mathbf{x}^i \in \Re^D$ (so that $\mathbf{X}$ is the $N \times D$ matrix of inputs) and $y^i \in R$. For an introduction to the background on this work, see [9].

The simplest type of priors over functions is the Gaussian one.

A Gaussian process is a Gaussian random function, fully characterized by its mean and covariance function. It can be viewed as a collection of random variables which have a joint multivariate Gaussian distribution, where for simplicity, we assume a zero-mean process: $f(\mathbf{x}^1), \dots, f(\mathbf{x}^n) \sim \mathcal{N}(0, \mathbf{\Sigma})$, where $\Sigma_{pq}$ gives the covariance between $f(\mathbf{x}^p)$ and $f(\mathbf{x}^q)$ and is a function of the corresponding $\mathbf{x}^p$ and $\mathbf{x}^q$: $\Sigma_{pq} = C(\mathbf{x}^p, \mathbf{x}^q)$. The covariance function $C(.,.)$ can be of any kind, provided that it generates a positive definite covariance matrix $\mathbf{\Sigma}$. Assuming a stationary process, which means that the covariance between two points depends only on the distance between them and is invariant to translation in the input space, a common choice of covariance function is

$$C(\mathbf{x}^p, \mathbf{x}^q) = v_1 \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d(x_d^p - x_d^q)^2\right] \quad (1)$$

where $D$ is the input dimension and $v_1, w_1, \dots, w_D$ are free parameters. Typically, covariance functions are chosen such as (1) so that points close together in the input space are more correlated than points far apart (a smoothness assumption). The parameter $v_1$ controls the vertical scale of variation and the $w_i$'s are inversely proportional to the horizontal length-scale in dimension $i$ ($\lambda_i = 1/\sqrt{w_i}$). Other forms of covariance functions are discussed in [9].

Let the input/target relationship be $y = f(\mathbf{x}) + \epsilon$. We assume an additive white noise with variance $v_0$, $\epsilon \sim$

$\mathcal{N}(0, v_0)$, and put a GP prior on $f(.)$ with covariance function as (1) with unknown parameters. Within this probabilistic framework, we have $y^1, \ldots, y^n \sim \mathcal{N}(0, \mathbf{K}_{N+1})$ with $K_{N+1_{pq}} = \Sigma_{pq} + v_0 \delta_{pq}$, where $\delta_{pq} = 1$ if $p = q$, 0 otherwise.

Based on a set of $N$ training data pairs, $\{\mathbf{x}^i, y^i\}_{i=1}^N$, we wish to find the predictive distribution of $y^*$ corresponding to a new given input $\mathbf{x}^*$.

We need to estimate the unknown parameters of the covariance function, as well as the noise variance $v_0$. This is done via maximization of the log-likelihood

$$\log(p(\mathbf{y}|\mathbf{X})) = -\frac{1}{2}\log(|\mathbf{K}|) - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{N}{2}\log(2\pi) \tag{2}$$

where $\Theta$ is the vector of parameters, $\Theta = [w_1 \ldots w_D \ v_0 \ v_1]^T$ and $\mathbf{K}$ is the $N \times N$ training covariance matrix.

The predictive distribution of $y^*$ is $p(y^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y}|\mathbf{X})}$. It can be shown that this distribution is Gaussian with mean and variance

$$\mu(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{y} \tag{3}$$
$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*) \tag{4}$$

where $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}^1, \mathbf{x}^*), \ldots, C(\mathbf{x}^N, \mathbf{x}^*)]^T$ is the $N \times 1$ vector of covariances between the test and training cases and $k(\mathbf{x}^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance between the test input and itself.

The vector $\mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1}$ can be interpreted as a vector of smoothing terms which weight training outputs $\mathbf{y}$, to make a prediction at the test point $\mathbf{x}^*$. This is the reason why predictions of GP model can be used for interpolation of missing data of the function of interest. GP models can be effectively used as a smoothing mechanism for noisy and multiple data.

If the new input is far away from the data points, the term $\mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*)$ will be small hence $\sigma^2(\mathbf{x}^*)$ large. This indirectly also means that GP models are more suitable for interpolation of data than for its extrapolation. Areas of the input space where there is little data, where the data has high complexity or the data is noisy are highlighted through a high variance.

## 3 Method

### 3.1 HomER NIRS data analysis

HomER (Hemodynamic Evoked Response) is a piece of software for the analysis and imaging of NIRS signals. It is developed by The Center for Functional Neuroimaging Technologies [10] and the Martinos Center's Photon Migration Imaging Lab and is freely available online [11]. HomER is based on Matlab and utilises the Photon Migration Imaging (PMI) toolbox [12] (also developed by The Center for Functional Neuroimaging Technologies)
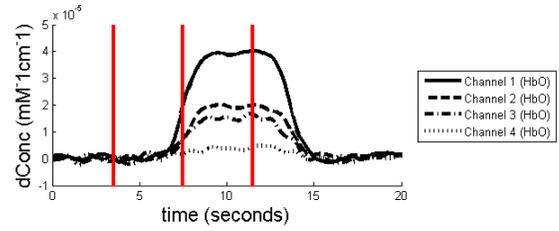


Figure 2. Input signals to both GP model and HomER. Red lines indicate image sample times at 3.5, 7.5 and 11.5 seconds.

to solve Diffuse Optical Imaging (DOI) forward and inverse problems. HomER can be used to generate image sequences of spatially resolved hemodynamic responses [13]. For the purposes of this study, specific images from the sequence at various points of time during a response were used for comparison with those produced by the GP method. The resolution was set at 20 x 20 interpolation points.

### 3.2 GP model interpolation

The GP model as described above was implemented in Matlab. For each time sample, the GP model calculates a field of Gaussians, each with a mean estimated magnitude and estimation variance. Maps of estimated magnitudes and their corresponding variances are plotted. These maps are the two results of the GP model interpolation technique. The process was then repeated for each time sample. Each time sample calculated a field of 20 x 20 Gaussians. Because GP models require that each signal has a point of activity, we chose that the location of signal activity be at the midpoint of the path since this is where the photon path interrogates cortical tissue according to models [5].

### 3.3 NIRS data

A sample NIRS data set included in the HomER package was modified slightly and used to compare the two processes. We decided to use this data because it featured a very obvious change in activity and would make comparison clearer. The source-detector layout for the data is shown in Figure 1. The layout consisted of a single light source and four photodetectors. The light transmitted was at wavelengths of 690 nm and 830 nm. There are 4 paths of light and 2 wavelengths so the data set had 8 channels of data. Samples of the results were taken at three sample times (3.5, 7.5 and 11.5 seconds) for comparison. Variance maps from GP model interpolation at these sample times were also recorded.
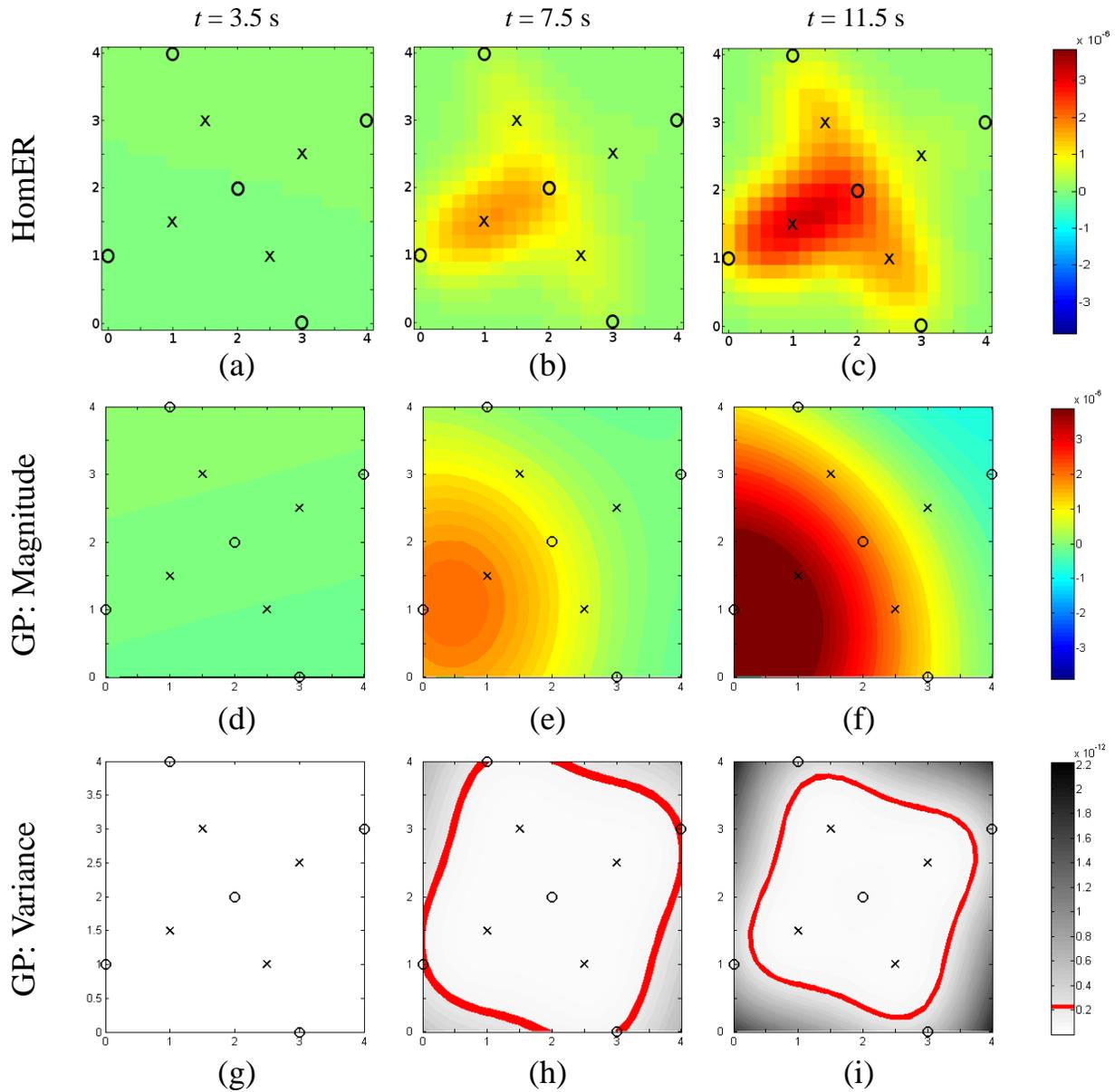
Figure 3. Comparisons of image maps of all methods at three different sample times. (a), (b) and (c) show image maps generated with HomER at times $t = 3.5$, $t = 7.5$ and $t = 11.5$ seconds, respectively. (d), (e) and (f) show image maps for same data for Gaussian Process model interpolation. (g), (h) and (i) are variance maps from Gaussian process model interpolation.

### 3.4 Preprocessing

Previous NIRS studies have used various cut-off frequencies for band-pass filtering of the raw NIRS signals. Our investigation used $0.01 - 0.8$ Hz as the band-pass range. The raw data was first low-pass filtered with a 3rd-order Butterworth filter at 0.8 Hz. The signals were then normalised by dividing each signal by its mean value. The data was then high-pass filtered using a 3rd-order Butterworth filter at 0.01 Hz. The signal was then offset to unity mean. The negative logarithm of these filtered intensity signals were taken to convert them to delta Optical Density measurements. By implementing the modified Beer-Lambert law as it is standard for NIRS [14], signals indicating changes in HbO and HbR (delta concentration signals) were calculated. The HbO delta concentration signals were then used as input to the GP model and to HomER since their activation was greater compared to the HbR signals, although both sets of signals were similar. The HbO delta concentration signals are shown in Figure 2. "Channel 1 (HbO)"

is the HbO delta concentration signal for the photon path from the light source to detector 1. Similar naming applies to the other three channels.

## 4 Results

### 4.1 Imaging results

Samples of the interpolation results were taken at 3.5, 7.5 and 11.5 seconds (as shown on Figure 2) to demonstrate typical results at three different states of activity. Activity magnitude results produced by HomER are shown in Figure 3 (a)–(c). Activity magnitude results produced by our Gaussian Process model interpolation method are shown in Figure 3 (d)–(f) and the corresponding variance results are shown in Figure 3 (g)–(i).

HomER is expected to produce the most accurate results, so the GP model interpolation results are compared directly to them. At $t = 3.5$ seconds, there is low activity for all 4 HbO signal channels (Figure 2). The HomER result (Figure 3 (a)) displays low activity as does the GP result (Figure 3 (d)). The two image maps for this time sample are very similar. The GP variance map (Figure 3 (g)) shows that the variance was very low for each of the 400 (20 x 20) Gaussians plotted. From the low levels of variance for this time sample, we believe that we can trust the GP interpolated magnitude result. By comparing the GP result to the HomER result, we know this to be true.

At $t = 7.5$ seconds, HbO signal activity is increasing. From Figure 2 we expect that channel 1 displays the highest level of activity, channels 2 and 3 display medium levels of activity and that channel 4 display very low levels of activity. The HomER result for this time sample (Figure 3 (b)) indeed displays the highest activity along the path from source A to detector 1 (channel 1). Lower activity is displayed along the paths from source A to detectors 2 and 3 (channel 2 and channel 3). Activity along channel 4 is only just noticeable. The GP result (Figure 3 (e)) displays a high centre of activity on the channel 1 path, but is not 'stretched' along the direction of the path. Looking at the corresponding variance map (Figure 3 (h)), we see that the variance has increased along with activity. The variance increases with distance from the nearest data source. While we trust the brighter areas of the map where variance is low, where the variance has increased and the map gets darker we mustn't place much trust in the corresponding interpolated magnitude result. A variance threshold of $0.2 \times 10^{-12}$ was arbitrarily selected and highlighted on the variance maps during image processing. We decided to trust the data within the threshold and disregard data values outside of the threshold marker. Many of the data points are actually extrapolated, not interpolated, since the points do not lie between data sources. As mentioned earlier, the GP model is more suited to interpolation than extrapolation, so extrapolated magnitudes typically have much higher variance than interpolated magnitudes.

Comparing the GP magnitude result to the HomER result, we can clearly see differences. The most obvious differences are near the edges of the magnitude map, e.g. in the bottom-left corner. The GP result shows high activity, but the HomER result does not. By taking the variance map into consideration, we can discard those extrapolated values because their variance is relatively high. Looking only at the mainly interpolated values within the threshold, we can still see differences, but they're not as large. By choosing a lower threshold, the results might look more similar.

At $t = 11.5$ seconds, the signal activity is at its highest. The HomER result (Figure 3 (h)) displays high activity along channel 1 and medium activity along channels 2 and 3. There is still very little activity along channel 4. These are the results we expect given the signals in Figure 2. The corresponding GP result (Figure 3 (f)) looks very different compared to the HomER result. The GP model has extrapolated Gaussians with very high mean values to the bottom-left of the image map. Such high activity is not seen in the HomER result. The GP variance map, however, tells us that those estimated values are of a relatively high variance, are outside of our threshold, and are therefore not to be trusted. The increase in activity has given rise to higher variance around the edge of the image map.

By again only concerning ourselves with the interpolated magnitude data within the variance threshold limit, the GP magnitude map is improved and is qualitatively similar to that produced by HomER. The GP magnitude map displays a large area of high activity around the midpoint of channel 1, where the HomER result only displays a small area of high activity. This is most likely a data extrapolation effect.

### 4.2 Variance map

The variance map is shown to be very important for GP model interpolation. When there is little difference between the channel data, the variances of the interpolated Gaussians stay low and the variance map indicates we can trust the interpolated magnitude values. When the difference in magnitude between data sources increases, the variance of the mean magnitudes also increases. In the case of extrapolation, the variances can be very large. The variance map tells us which data on the magnitude map we can trust.

## 5 Conclusion

What we have presented here is a unique method of producing image maps of cortical activity as measured by NIRS. The interpolation method is based on generating a large number of Gaussians and estimating their mean interpolated magnitude and the variance of those magnitudes. This method was compared to the current standard for generating cortical activity maps and was shown to be comparable, but clearly not exactly the same. The results show that GP models can interpolate data points to produce hemo-

dynamic images which may have utility in applied neuroscience applications. This method is unique because the GP model is non-parametric; it has no knowledge about the medium being interrogated, the measurement modality or even the direction of the channel paths. By comparison, DOI does know these parameters and utilises them effectively. Although the GP interpolation method lacks in this respect, it is still interesting to see how the results are not too dissimilar and have compared quite well. This GP interpolation technique may be of some use when NIRS is applied in other biological contexts where model information, as it would be used by DOI, is sparse or less adequate.

The GP model interpolation technique is also unique in that it indicates the reliability of its results. Using the interpolated magnitude map alone would produce inaccurate and misleading results. The variance map is therefore very important for producing useful results, indicating the reliability of and the amount of trust to be placed in the magnitude estimations. No other interpolation technique has this feature and it is perhaps something that can be exploited during subsequent response processing.

Future work on this subject would be a general improvement to the results presentation by combining the magnitude and variance maps into a single result. Investigations into the ideal variability threshold level should also be carried out since the variance map is so important for this method. There is also a possibility of utilising the variability map for guiding optode placement in NIRS experiments.

## Acknowledgements

## References

[1] Arridge, S. R., "Optical tomography in medical imaging.", *Inverse Problems*, 15(2), 1999, R41–R93.

[2] Franceschini, M. A., Toronov, V., Filiaci, M., Gratton, E. & Fantini, S., "On-line optical imaging of the human brain with 160-ms temporal resolution.", *Opt Express*, 6, 2000, 49–57.

[3] S. Faul, G. Gregorčič, G. Boylan, W. Marnane, S. Lightbody, G. Connolly, "Gaussian process modelling of EEG for the detection of neonatal seizures.", *IEEE Transactions on Biomedical Engineering*, 54(12), 2007, 2151–2162.

[4] K. Ažman and J. Kocijan, "Application of Gaussian processes for black-box modelling of biosystems.", *ISA Transactions*, 46(4), 2007, 443–457.

[5] Okada, E., Firbank, M., Schweiger, M., Arridge, S. R., Cope, M. & Delpy, D. T., "Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head.", *Appl Opt*, 36(1), 1997, 21–31.

[6] Cope, M., "The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant.", PhD thesis, University College London, 1991.

[7] Sassaroli, A. & Fantini, S., "Comment on the modified Beer-Lambert law for scattering media", *Phys. Med. Biol.*, 49(14), 2004, N255–N257.

[8] Yoko Hoshi, "Functional near-infrared spectroscopy: current status and future prospects.", *J. Biomed. Opt.*, 12(6), 2007, 062106.

[9] C. E. Rasmussen & C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2006.

[10] The Center for Functional Neuroimaging Technologies, `http://www.nmr.mgh.harvard.edu/martinos/ncrr/index`

[11] HomER - Hemodynamic Evoked Response NIRS Data Analysis GUI, `http://www.nmr.mgh.harvard.edu/PMI/resources/homer/download.htm`

[12] Photon Migration Imaging Toolbox for Matlab, `http://www.nmr.mgh.harvard.edu/~jstott/PMI/`

[13] Huppert, T., Diamond, S. G., Franceschini, M. A. & Boas, D. A., "HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain.", *Appl Opt*, 48(10), 2009, D280–D298.

[14] Matthews, F., Pearlmutter, B. A., Ward, T. E., Soraghan, C. & Markham, C., "Hemodynamics for Brain-Computer Interfaces" *IEEE Signal Processing Magazine*, 25(1), 2008, 87–94.