Extracting histological parameters from multi-spectral retinal images: a Bayesian inverse problem approach

Yuan Shen[1]
y.shen.2@cs.bham.ac.uk
Antonio Calcagni[12]
a.calcagni@aston.ac.uk
Ela Claridge[1]
e.claridge@cs.bham.ac.uk
Frank Eperjesi[2]
f.eperjesi@aston.ac.uk
Hannah Bartlett[2]
h.e.bartlett@aston.ac.uk
Jonathan Gibson[2]
j.m.gibson@aston.ac.uk
Andrew Palmer[13]
adp551@cs.bham.ac.uk
Iain Styles[1]
I.B.Styles@cs.bham.ac.uk

[1] School of Computer Science
University of Birmingham
Birmingham, UK

[2] School of Life and Health Sciences
Aston University
Birmingham, UK

[3] PSIBS Doctoral Training Centre
University of Birmingham
Birmingham, UK

Abstract

Extracting histological parameters, especially macular pigment, from multispectral images of the ocular fundus is a potential technique for the assessment of age-related macular degeneration. Such approaches make use of a Monte Carlo radiation transport model relating spectral reflectance of the tissue to tissue histology. We develop a probabilistic surrogate for this computationally expensive physicalmodel using Gaussian processes (GP). Further, we present a Bayesian inversion algorithm that uses the surrogate model to recover model input parameters. This methodology is tested both on synthetic data generated from the Monte Carlo model and on real image data. It is shown that our inversion methods can recover macular pigment concentrations in human retina with good accuracy and the spatial distribution is consistent with known physiology.

1 Introduction

A reduction in the quantity of macular pigment (MP) in the retina is thought to be positively correlated to the onset of Age-related Macular Degeneration (AMD), the most common cause of blindness in the Western world. There are no established objective assessment

methods. Research techniques compare a spectrum measured at the fovea (where the macular pigments are present) with a background measurement from nearby and deduce the macular pigment optical density (MPOD) from the difference between the two measurements. An important unresolved problem is to compensate for the effect of scattering by ocular tissues and of absorption by other pigments (melanins, haemoglobins) present in the fovea region.

In [4], we proposed a solution which involves the use of a computational forward model that predicts spectral reflectance for a histologically plausible set of retinal composition parameters. The model is parameterised by the quantities of pigments (MP, melanin in retinal pigment epithelium and choroid, haemoglobins in retina and choroid) and assumes constant tissue thickness and scattering properties derived from values found in the literature [2]. The ability to isolate the effect of MP on tissue reflectance depends on being able to spectrally separate MP from other pigments present in the sample. It has been shown [4] that this separation can be achieved optimally through measurements at six wavelengths, namely 507nm, 525nm, 552nm, 585nm, 596nm and 610nm. Image data is acquired at these wavelengths and the images are divided pixel-wise by the image at 610nm to normalise for uneven illumination to form so-called image quotients.MP concentration can then be estimated from this set of spectral measurements by inverting the forward model.

The forward model is constructed by means of a Monte Carlo-based simulation of light transport in the retinal tissue. Previous attempts to inverting such a mdoel have adopted a lookup-table approach to compute MP values by interpolation [4]. This approach is sensitive to image quality and underlying approximations in the model, and we are driven towards a Bayesian inverse problem approach in order to subvert these difficulties. However, the Monte Carlo simulation is costly and unsuitable for iterative inversion, which makes a robust solution scheme for the inverse problem very challenging. A solution that could circumvent this difficulty is to approximate the forward model by a computationally cheap surrogate model.

We have employed Gaussian Process modelling (GP) [3] to approximate the functional relationship between the histological parameters and the spectral reflectance. This surrogate model is constructed using a set of parameter-reflectance pairs generated from the physical forward model to learn the model's parametrised mean and covariance functions. Within the GP framework, the approximate forward model can be formulated analytically and its computation is much more efficient. Moreover, the probabilistic nature of GPs allows us to naturally adopt a Bayesian approach to our inverse problem providing a maximum a posteriori (MAP) estimate of the model parameters from reflectance measurements.

We first provide a description of our methodology in Sect. 2. In Sect. 3, we validate our approach with tests on synthetic and real data. We conclude with a discussion in Sect. 4.

2 Methodology

Our approach to estimate macular pigment has three major ingredients: 1) a surrogate approximation of the forward model using GPss; 2) a Bayesian inverse model for estimating the model input parameters; and 3) a spatial regularization scheme for the estimation of macular pigment across pixels.

Mathematically, a GP is formulated as a probability distribution of functions $\mathbf{f} : \mathbf{x} \longmapsto \mathbf{y}$ with input $\mathbf{x} \in \Re^{d_x}$ and output $\mathbf{y} \in \Re^{d_y}$. This distribution is characterized by a mean function $\mathbf{m}(\mathbf{x}) \in \Re^{d_y}$ and a covariance function $\mathbf{cov}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2)$. The mean function is modelled by $\mathbf{m}(\mathbf{x}) = \mathbf{B}^\intercal \mathbf{h}(\mathbf{x})$ where $\mathbf{h} \in \Re^K$ is a multivariate regression function and $\mathbf{B} \in \Re^{K \times d_y}$ is its

regression coefficient matrix. The covariance function $\mathbf{cov}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)$ is approximated by a product of output covariance $\Sigma(\mathbf{y}_1,\mathbf{y}_2)\in\Re^{d_y\times d_y}$ and input covariance $c(\mathbf{x}_1,\mathbf{x}_2)\in\Re^{d_x\times d_x}$ specified by $c(\mathbf{x}_1,\mathbf{x}_2)=\exp\{-(\mathbf{x}_1-\mathbf{x}_2)^\intercal \mathrm{diag}(\mathbf{r})(\mathbf{x}_1-\mathbf{x}_2)\}$ with $\mathbf{r}\in\Re^{d_x}$ as a positive roughness parameter vector. Note that $\mathbf{B}$, $\mathbf{r}$, and Σ are the hyper-parameters that specify a GP.

To train a Gaussian process model, a set of inputs $\mathbf{S}=\{\mathbf{x}_1,...,\mathbf{x}_n\}$ is first selected, and the corresponding output of each input in the set is computed by running the forward model. The set of outputs $\mathbf{D}=(\mathbf{f}(\mathbf{x}_1)\cdots\mathbf{f}(\mathbf{x}_n))^\intercal\in\Re^{n\times d_y}$ can be seen as the observations from which we infer the underlying GP. The hyperparameters $\mathbf{B}$, $\mathbf{r}$ and Σ are estimated by maximizing the likelihood $p(\mathbf{D}|\mathbf{B},\mathbf{r},\Sigma)\propto\frac{\exp\left(-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(\mathbf{D}-\mathbf{HB})^\intercal\mathbf{A}^{-1}(\mathbf{D}-\mathbf{HB})]\right)}{(2\pi)^{nd_y/2}|\mathbf{A}|^{d_y/2}|\Sigma|^{n/2}}$ where $\mathbf{H}^\intercal=(\mathbf{h}(\mathbf{x}_1)\cdots\mathbf{h}(\mathbf{x}_n))\in\Re^{K\times n}$ and $\mathbf{A}\in\Re^{n\times n}$ with $\mathbf{A}_{ij}=c(\mathbf{x}_i,\mathbf{x}_j)$. To reduce the number of hyperparameters to be estimated, $\mathbf{B}$ and Σ in the likelihood are both marginalised out: $p(\mathbf{r}|\mathbf{D})=\int d\mathbf{B}d\Sigma\cdot p(\mathbf{D}|\mathbf{B},\mathbf{r},\Sigma)\cdot p(\mathbf{r})\cdot p(\Sigma)$ with a log-logistic prior on $\mathbf{r}=\prod_{i=1}^{d_x}(1+r_i^2)^{-1}$ and a Jeffreys-type prior on $\Sigma\propto|\Sigma|^{-\frac{d_y+1}{2}}$. Let $\hat{\mathbf{r}}$ denote the estimate of $\mathbf{r}$ that maximize $p(\mathbf{r}|\mathbf{D})$ and note that $\hat{\mathbf{r}}$ is the only hyperparameter that specifies the GP.

For any input $\mathbf{x}\notin\mathbf{S}$, the corresponding output $\mathbf{f}(\mathbf{x})$ is given in terms of a predictive distribution as $p(\mathbf{f}(\mathbf{x}))=\mathscr{T}_{d_y}(\hat{\mathbf{m}},\hat{\Gamma},n-K)$, which is a multivariate student distribution with its location vector $\hat{\mathbf{m}}$, scale matrix $\hat{\Gamma}$ and degrees of freedom equal to $n-K$. For the expression of $\hat{\mathbf{m}}(\mathbf{x})$ and $\hat{\Gamma}(\mathbf{x})$, we refer to [1] but note that the evaluation of these expressions does use the training sets $(\mathbf{S},\mathbf{D})$, the multivariate regression function $\mathbf{h}(\cdot)$ and the input covariance function $c(\cdot,\cdot;\hat{\mathbf{r}})$ which is specified by the hyperparameter $\hat{\mathbf{r}}$.

Let $\mathbf{y}^i$ denote the multispectral image data. It is assumed that $\mathbf{y}^i$ is the observed model output $\mathbf{y}$ with noise contamination. Thus $\mathbf{y}^i=\mathbf{f}(\mathbf{x})+\varepsilon$ where ε is assumed to be i.i.d. multivariate Gaussian noise with zero mean and spatially homogeneous error covariance $\mathbf{R}$. This gives rise to our definition of the likelihood as $p(\mathbf{y}^i|\mathbf{f},\mathbf{x})\propto\exp\left\{-\frac{1}{2}\varepsilon^\intercal\mathbf{R}^{-1}\varepsilon)\right\}$. However, the forward model $\mathbf{f}(\mathbf{x})$ is now approximated by a GP which is represented by a probability distribution of $\mathbf{f}$, i.e. $p(\mathbf{f})$. Therefore, $\mathbf{f}$ in the likelihood is just a realization of the GP $p(\mathbf{f})$ and needs to be integrated out. The resulting likelihood is obtained by $p(\mathbf{x}|\mathbf{y}^i)\propto\int d\mathbf{f}\cdot p(\mathbf{y}^i|\mathbf{f},\mathbf{x})\cdot p(\mathbf{f})$. The GP $p(\mathbf{f})$ can be seen as a prior on $\mathbf{f}$. A GP based inversion maximizes the posterior distribution $p(\mathbf{x}|\mathbf{y}^i)$. Given $\mathbf{x}$, $p(\mathbf{f}(\mathbf{x}))$ is a Student distribution over $\mathbf{f}(\mathbf{x})$. Thus, we apply the fact that the student distribution can be represented as an infinite mixture of scaled Gaussian distribution. The resulting posterior is now given by $p(\mathbf{x}|\mathbf{y}^i)\propto p(\mathbf{x})\int d\lambda\mathscr{G}(\lambda,\nu/2,\nu/2)\cdot\mathscr{N}(\mathbf{y}^i,\hat{\mathbf{m}}(\mathbf{x}),\mathbf{R}+\lambda\hat{\Gamma}(\mathbf{x}))$ where $\nu=n-K$. The gamma function here is strongly localized due to large ν value. Note that the number of training data n and the number of regression functions K usually differ by two order of magnitudes. Thus, the computational cost for numerical integration over λ is negligible.

As it is assumed that the parameters vary smoothly across all pixels in the spatial domain $\mathbf{s}\in\mathscr{I}$, the parameter field $\mathbf{x}(\mathscr{I})$ needs to be estimated jointly from the image $\mathbf{y}^i(\mathscr{I})$. The joint posterior is given by $p(\mathbf{x}(\mathscr{I})|\mathbf{y}^i(\mathscr{I}))=\prod_{\mathbf{s}}p(\mathbf{x}(\mathbf{s})|\mathbf{y}^i(\mathbf{s}))\cdot p(\mathbf{x}(\mathscr{I}))$. The prior $p(\mathbf{x}(\mathscr{I}))$ accounts for the smooth variation of each $\mathbf{x}$-component in $\mathscr{I}$ and is specified by a Gaussian Markov random field for each $\mathbf{x}$-component with its regularization parameter ξ:

$$p(\mathbf{x}(\mathscr{I}))\propto\prod_{i=1}^{d_x}\xi_i^{-|\mathscr{I}|}\cdot\exp\left(-\frac{\sum_{\mathbf{s}\in\mathscr{I}}\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{s}}(x_i(\mathbf{m})-x_i(\mathbf{s}))^2}{2\xi_i^2}\right),$$

where $\mathscr{N}_\mathbf{s}$ defines a neighbourhood of $\mathbf{s}$ in $\mathscr{I}$. Here, such a neighbourhood represents the 8 adjacent pixels to every individual pixel.

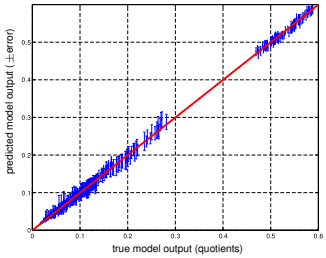

Figure 1: Plot of simulated vs emulated spectral reflectance quotients for the Monte Carlo model. Error bars indicate twice standard deviation of predictive error

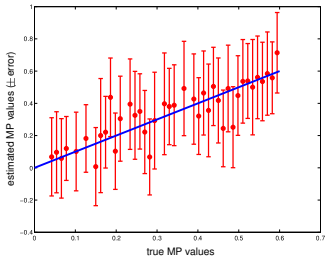

Figure 2: Plot of estimated vs true macular pigment concentration in the retina. Dots and error bars denote the mean twice standard deviations of the estimates from independently repeated observations.

3 Numerical Validation

In this work, we consider only three input parameters of the model: macular pigment concentration in retina, $\mathscr{C}_{MP}$, melanin concentration in retina, $\mathscr{C}_{RH}$, haemoglobin concentration in RPE, $\mathscr{C}_{RM}$. Other parameters such as haemoglobin and melanin in choroid, are set to typical values. The output is a five dimensional vector of image quotients.

To train a GP emulator for the Monte Carlo model, a set of 100 input-output pairs are generated by running the simulation for 100 input parameter vectors which are sampled from the three-dimensional input space using a Latin hypercube algorithm that maximizes the Euclidean distance between these input parameters. Further, the pivoted Cholesky decomposition is applied to detect those vectors that are not sufficiently far apart. After training, 50 input vectors are sampled randomly and their corresponding output vectors are simulated. The predictive mean and standard deviation of their emulator output are computed and compared in Fig. 1. It is seen that most of the mean predictions are very close to the true values. Those that clearly deviate are within of two standard deviations of the prediction error. However, the level of prediction error is somewhat high (about 5%).

Next, we compare the estimated macular pigment concentration with its true value. The other two variable input parameters are not further investigated. For each of 7 $\mathscr{C}_{MP}$-values evenly sampled from its normal range, we have generated 10 repeated observations by adding 10 i.i.d Gaussian noise with $\sigma^2 = 0.01$ to the true value. The mean and its twice standard deviation of those $\mathscr{C}_{MP}$-estimates are displayed in Fig. 2. The large error bars indicate that the estimate of $\mathscr{C}_{MP}$ from a single observation is prone to random fluctuation. In practice, however, one can use the observation on neighbouring pixels of the same image if we assume $\mathscr{C}_{MP}$ is constant. Equivalently, we estimate $\mathscr{C}_{MP}$ jointly for all pixels on a retinal image while imposing some smoothness prior on the parametric map of $\mathscr{C}_{MP}$. In this work, a Markov random field prior is imposed on parameter maps.

The algorithm to jointly estimate the parameters across all pixels is tested on a set of artificial images of size 100×100, with an uniform distribution of all histological parameters except for that of macular pigment. The spatial distribution of $\mathscr{C}_{MP}$ is modeled by a superposition of an uniform background concentration with $\mathscr{C}_{MP} = 0.14$ and a two-dimensional isotropic Gaussian with its centre at (50,50) and its full width at half maximum around 10 pixels. The Gaussian is scaled so that its peak value is about $\mathscr{C}_{MP} = 0.48$. Also, we add i.i.d

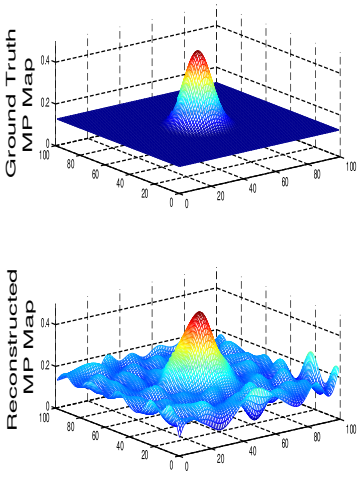

Figure 3: Numerical experiment with synthetic data: the true macular pigment field (upper panel) and the estimated macular pigment field (lower panel)

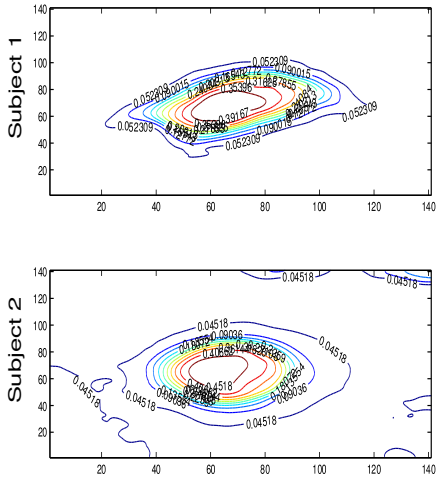

Figure 4: Contour plots of the estimated macular pigment field for two real images from two subjects.

Gaussian noise with noise level $\sigma^2 = 0.01$ to the signal at each pixel. The results are shown in Fig. 3. It can be seen that the estimated $\mathscr{C}_{MP}$s show a spatial distribution very close to the true one. Moreover, both the baseline and peak $\mathscr{C}_{MP}$ values are estimated with good accuracy, not to mention that all parameter fields are initialized as a constant field with its value randomly chosen from the corresponding permissible range. On average, the optimization procedure is terminated after 500 iterations.

Finally, we apply the algorithm to 2 real retinal images of 2 healthy subjects. For each image, a ROI of size 140×140 is selected so that the foveal region is located in the middle of the image. The estimated $\mathscr{C}_{MP}$ maps are displayed in Fig. 4. It can be seen that there is a distinct peak of $\mathscr{C}_{MP}$ in the foveal area for each subject, with $\mathscr{C}_{MP} = 0.39$ (upper panel) and $\mathscr{C}_{MP} = 0.45$ (lower panel). It is believed that the estimated peak values are reliable. This is because we also observe that the baseline $\mathscr{C}_{MP}$ is very small (about 0.05) in both cases. In addition, it is clear to see a rapid decrease of $\mathscr{C}_{MP}$ from its peak to baseline. Both observations are consistent with known physiology.

4 Discussion

We have proposed and tested a surrogate model based inversion method for analysing medical images. The numerical experiments in Sect. 3 demonstrate that our methodology is feasible and the reconstructed macular pigment map is consistent with known physiology. In the following, we discuss issues to be addressed in follow-up work.

Our surrogate model for the Monte Carlo model needs to be improved with regards to the large uncertainty observed (high prediction error). There are two possible explanations:

1) The output of a Monte Carlo model is, in principle, stochastic whereas the GP emulator method is developed for approximating a deterministic model; 2) due to the ill-conditioned nature of the problem, only a limited amount of training data can be used. In this work, those training data are generated by the Latin hypercube algorithm which does not take into account the model itself. Clearly, the optimal choice of training data is model dependent. Therefore, a more sophisticated algorithm needs to be developed.

In this work, we predict the model output vector of length d_y using the optimized GP for each pixel individually. In fact, we can also predict the multivariate output field jointly. This can further reduce the prediction error as the former approach can be considered to be an approximation of the latter by setting all off-diagonal blocks of size $d_y \times d_y$ in its predictive covariance matrix to zero. However, the resulting covariance matrix is too large to be dealt with in respect of the computing power. On the other hand, this matrix could be a low-rank matrix as the parameter field is smooth. Therefore, low-rank approximation techniques could be used here, in conjunction with the surrogate approximation method.

Applications of GPs as surrogate model in medical image analysis can go beyond the model inversion problem which we have explored in this paper. A surrogate approximation approach could also be useful in optimal design of experiments. For example, we can use the surrogate model to determine the optimal set of wavelengths for multispectral imaging. For Gaussian processes, some theoretical results have been developed in the area of experimental design. More importantly, Gaussian processes can be used for simultaneously approximating a forward model and accounting for model error. This is potentially useful because the forward model used in a inverse problem may not adequately describe the full complexity and variability of the problem. GPs can include such information statistically.

5 Acknowledgements

We acknowledge financial support from the Dunhill Medical Trust under grant R116/0509, and from the EPSRC under grant EP/F50053X/1 (studentship to AP in the PSIBS Doctoral Training Centre).

References

[1] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams. Multi-task gaussian process prediction. *IEEE Transactions on Neural Networks*, 2:409–435, 2008.

[2] S. J. Preece and E. Claridge. Monte Carlo modelling of the spectral reflectance of the human eye. *Phys. Med. Biol.*, 47:2863–2877, 2002.

[3] Carl Edward Rasmussen and Chris Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge MA, 2006.

[4] I. B. Styles, A. Calcagni, E. Claridge, F. Orihuela-Espina, and J. M. Gibson. Quantitative analysis of multi-spectral fundus images. *Medical Image Analysis*, 10:578–597, 2006.