A NOVEL COVARIANCE FUNCTION FOR PREDICTING VEGETATION BIOCHEMISTRY FROM HYPERSPECTRAL IMAGERY WITH GAUSSIAN PROCESSES

Utsav B. Gewali [1] and Sildomar T. Monteiro [1,2]

[1]Chester F. Carlson Center for Imaging Science, [2]Dept. of Electrical Engineering
Rochester Institute of Technology, Rochester, NY
ubg9540@rit.edu

ABSTRACT

Remotely extracting information about the biochemical properties of the materials in an environment from airborne- or satellite-based hyperspectral sensor has a variety of applications in forestry, agriculture, mining, environmental monitoring and space exploration. In this paper, we propose a new non-stationary covariance function, called exponential spectral angle mapper (ESAM) for predicting the biochemistry of vegetation from hyperspectral imagery using Gaussian processes. The proposed covariance function is based on the angle between the spectra, which is known to be a better measure of similarity for hyperspectral data due to its robustness to illumination variations. We demonstrate the efficacy of the proposed method with experiments on a real-world hyperspectral dataset.

***Index Terms*—** Hyperspectral images, Biochemistry prediction, Gaussian processes, Covariance/Kernel functions, Bayesian non-parametrics

1. INTRODUCTION

Biochemistry estimation is the problem of retrieving information about the chemical composition of the vegetation from its reflectance spectrum. This is possible because the interaction between a material and the light at different wavelengths depends on the absorption bands of the material, manifested by its atomic or molecular structure. Biochemical distribution maps over a large geographical area can be created using hyperspectral images captured by airborne or space-borne sensors. Hyperspectral imaging is the process of capturing an image of the reflected energy from a scene over hundreds of contiguous, narrow bands in the electromagnetic spectrum, in the visible and infrared spectral regions [1]. Each pixel of a hyperspectral image is a reflectance spectra, unlike a three channel (Red-Green-Blue) color intensity vector as in conventional photography. Biochemistry prediction is a challenging problem because there is usually a non-linear relationship between the biochemical parameters and the spectrum [2]. Traditionally, biochemical predictions were made by developing vegetation indices and relating them to biochemical properties by using linear equations [3]. The vegetation indices are hand-designed ratios involving reflectances at few wavelengths. But recently, there has been a drive to use data-driven regression methods such as kernel ridge regression [4], support vector machines [5] and Gaussian processes [6] for biochemical estimation. A major hurdle in using machine learning models for biochemical estimation is that the hyperspectral datasets are high dimensional with few labeled samples, making the models prone to overfitting. These datasets are created by collecting specimens from an environment, and measuring their reflectance spectra and performing chemical analysis. Due to difficulties in collecting and analyzing specimens, these datasets tend to have very few labeled samples. It is common for datasets to have few hundred dimensional input spectra with less than a hundred labeled examples.

Large scale biochemical distribution maps are essential for various applications such as environmental studies [7]. For instance, relative contents of the photosynthetic pigments, such as chlorophyll-a, chlorophyll-b and carotenoids, can be used to detect stress, assess photosynthetic activity or categorize vegetation. This information can be used to map and isolate the advent of infectious plant disease or to assess the effects of conservation efforts in an ecosystem. Another big application area of biochemical estimation is precision farming [8]. Precision farming is the idea of using site specific crop and soil information to manage agricultural land for optimal utilization of resources, sustainability and maximizing profits. Plant nutrient estimates over an agricultural land may indicate deficiency of certain nutrients at certain sections of the land. Farmers can then apply right fertilizers in appropriate amounts to different sites based on this information. This will result in better yield and also prevent the haphazard over use of fertilizers, which is a big environmental problem.

Gaussian processes (GPs) are attractive for biochemical prediction from hyperspectral images due to their resilience to overfitting and capability to model non-linear relationships. GPs, being Bayesian, tend not to overfit, and, being non-parametric, can adjust their complexity with new data and model complex non-linear relationships, not running into the problem of underfitting. Some studies [9, 4, 6, 10] have ap-

plied GPs to estimate chlorophyll content and vegetation parameters from multispectral and hyperspectral imagery. It has been shown that the GPs outperform commonly used vegetation indices in predicting vegetation properties [4]. Another advantage of the GP is that it provides confidence bound around each prediction indicating the robustness. However, they are much more computationally heavy than the vegetation indices which may produce sufficiently accurate results for some applications. All of the previous studies have used the standard squared exponential covariance function with the GPs.

In this paper, we introduce a new spectral angle based covariance function, called exponential spectral angle mapper (ESAM), for biochemical prediction, and compare its performance with commonly used stationary and non-stationary covariance functions. ESAM makes an assumption that the angle between the spectra is a better metric for similarity compared to the Euclidean distance between the spectra used by squared exponential covariance function and all of the stationary covariance functions. We further compare the performance of ESAM with the non-stationary neural network covariance function [11] and the non-stationary observation angle dependent (OAD) covariance function [12].

2. GAUSSIAN PROCESS FOR REGRESSION

Gaussian process (GP) regression is a probabilistic model where the output variable values of all the training and the testing data points are considered to be samples of a joint multivariate normal distribution [11]. The mean of this joint distribution is generally taken to be a zero vector while the covariance matrix is estimated using a covariance function defined over the input values of a pair of data points. Under these assumptions, the probability distribution of the output values of the data points is

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right), \quad (1)$$

where y is the vector of the output values at the training points, f_* is the vector of the output values at the testing points, X is the training input matrix with the input vector at each training point as its rows, X_* is the testing input matrix with the input vectors at each testing points as its rows, and $K(X_1, X_2)$ is a covariance matrix whose *i*-th row and *j*-th column element is the $k(\mathbf{x_1}, \mathbf{x_2})$ of *i*-th row of X_1 and *j*-th row of X_2, where $k(\mathbf{x_1}, \mathbf{x_2})$ is the covariance function. It should be noted that an independent and identically distributed Gaussian noise variance, σ_n, is added to the main diagonal of the training data covariance matrix to model noise in the observations. During inference, the posterior predictive distribution at the testing points is given by

$$f_*|X, y, X_* \sim \mathcal{N}(\bar{f}_*, cov(f_*)), \quad (2)$$

where,

$$\bar{f}_* = K(X_*, X)\left[K(X,X)+\sigma_n^2 I\right]^{-1} y, \quad (3)$$

$$cov(f_*) = K(X_*, X_*) - K(X_*, X)\left[K(X,X)+\sigma_n^2 I\right]^{-1} K(X, X_*). \quad (4)$$

Mostly, the mean value of the predictive distribution is used as the estimate and the variance is used to measure the uncertainty. The covariance function, $k(\mathbf{x_1}, \mathbf{x_2})$, defines the relationship between a pair of data points, $\mathbf{x_1}$ and $\mathbf{x_2}$. Depending on the characteristics of the covariance function, different behavior is exhibited by the GP. Hence, the choice of an appropriate covariance function becomes essential while modeling data with GPs. Only positive semi-definite functions (or Mercer kernels) can be a valid covariance function. A covariance function can be stationary if it is only a function of the Euclidean distance between the points, or otherwise be non-stationary. The covariance functions are usually parameterized by few hyperparameters, which can be learned from the data by maximizing the log marginal likelihood function of the GP,

$$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^T\left(K(X,X)+\sigma_n^2 I\right)^{-1}\mathbf{y} - \frac{1}{2}\log|K(X,X)+\sigma_n^2 I| - \frac{n}{2}\log 2\pi. \quad (5)$$

n in the equation is the number of training points.

3. EXPONENTIAL SPECTRAL ANGLE MAPPER COVARIANCE FUNCTION

Our proposed covariance function, the exponential spectral angle mapper (ESAM), is a non-stationary function that uses the angle between two spectra as the similarity metric for covariance calculation. The motivation to use spectral angle comes from a widely used spectra similarity measure in hyperspectral remote sensing, called spectral angle mapper (SAM), $\alpha(\mathbf{x_1}, \mathbf{x_2}) = \cos^{-1}\left(\frac{\mathbf{x_1}\cdot\mathbf{x_2}}{\|\mathbf{x_1}\|\,\|\mathbf{x_2}\|}\right)$.

When there is a perfect match between the spectra, the SAM value between them becomes zero, and increases in magnitude as the two spectra being compared start getting more dissimilar. Since, SAM is only dependent on the direction of the spectra, which corresponds to their shape not magnitudes, it is known to produce better results compared to other metrics (e.g., Euclidean distance between the spectra). The equation for SAM by itself cannot be used as a covariance function with GPs as it is not a positive semi-definite function. We have encapsulated the SAM metric inside an exponential function to produce ESAM covariance function which is a positive semi-definite function and also captures the spectral similarity property of SAM. The ESAM covariance function is defined as

$$k_{\text{ESAM}}(\mathbf{x_1}, \mathbf{x_2}) = \sigma_0^2 \exp(-\alpha(\mathbf{x_1}, \mathbf{x_2})/\sigma_1^2), \quad (6)$$

where σ_0 and σ_1 are the hyperparameters of the covariance function.

3.1. Analysis

We prove the validity of ESAM covariance function with the help of the observation angle dependent (OAD) covariance function [12]. We define a new function, $k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2})$, as an exponential of the OAD covariance function, $k_{\text{OAD}}(\mathbf{x_1}, \mathbf{x_2})$, and show that under certain hyperparameter settings, this function converges to the ESAM function. The hyperparameters of $k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2})$ are the multiplier, a^2, and the hyperparameters of the OAD function, c^2 and ϕ. The $k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2})$ is a valid positive semi-definite function as exponentiating a positive semi-definite function, i.e., OAD function, produces a positive semi-definite function and so does multiplying a positive semi-definite function with a positive number, a^2. The proof that OAD is a positive semi-definite function can be found in [12].

$$k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2}) = a^2 \exp\left(k_{\text{OAD}}(\mathbf{x_1}, \mathbf{x_2})\right) \tag{7}$$

where,

$$k_{\text{OAD}}(\mathbf{x_1}, \mathbf{x_2}) = c^2 \left(1 - \frac{1 - \sin\phi}{\pi}\alpha(\mathbf{x_1}, \mathbf{x_2})\right) \tag{8}$$

$$k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2}) = a^2 \exp\left[c^2 \left(1 - \frac{1 - \sin\phi}{\pi}\alpha(\mathbf{x_1}, \mathbf{x_2})\right)\right]$$

Taking $\phi = 0$.

$$= a^2 \exp\left[c^2 \left(1 - \frac{1}{\pi}\alpha(\mathbf{x_1}, \mathbf{x_2})\right)\right]$$

$$= a^2 \exp(c^2) \exp\left(-\frac{c^2}{\pi}\alpha(\mathbf{x_1}, \mathbf{x_2})\right)$$

Both $a^2 \exp(c^2)$ and $\frac{c^2}{\pi}$ are independent positive quantities. Taking $a^2 \exp(c^2) \to \sigma_0^2$ and $\frac{c^2}{\pi} \to \frac{1}{\sigma_1^2}$.

$$k_{\text{new}}(\mathbf{x_1}, \mathbf{x_2}) = k_{\text{ESAM}}(\mathbf{x_1}, \mathbf{x_2}) \tag{9}$$

Hence, ESAM is a valid covariance function.

4. EXPERIMENTS

4.1. Dataset

The dataset used in our experiments was obtained from the National Ecological Observatory Network (NEON) [13]. It contains Analytical Spectral Devices (ASD) Fieldspec-3 portable spectrometer measured leaf reflectances and the corresponding nitrogen and carbon contents of the 54 samples collected as part of 2013 field campaign at San Joaquin, Soaproot Saddle and Teakettle in California [1]. The spectra ranges from $350\,\text{nm}$ to $2050\,\text{nm}$ in wavelengths with $1\,\text{nm}$ interval, and the contents of nitrogen and carbon is measured in terms of percentage dry foliage weight. We have used the sample hyperspectral image provided with the dataset for testing. The 250×250 image was collected by NEON imaging spectrometer (NIS) flown over an area in San Joaquin. Each pixels of the image has a spectral range of $382\,\text{nm}$ to $2511\,\text{nm}$ and covers an area of $1\,\text{m} \times 1\,\text{m}$ on the ground.

[1]National Ecological Observatory Network. 2015. Available online http://data.neoninc.org/ from National Ecological Observatory Network, Boulder, CO, USA.

4.2. Methodology

In the first experiment, the predictive performances of the ESAM covariance function and the other covariance functions was compared using leave-one-out cross-validation. The performance metric used was the Pearson correlation coefficient, r^2 value, calculated between the true value and the predicted value of the biochemical parameter. Since the dataset has very few labeled examples, leave-one-out cross-validation was chosen to get accurate estimates for the performance. All the covariance functions used were isometric, and the Gaussian likelihood was used. The hyperparameters of the GP were optimized by minimizing the negative log likelihood function using quasi-Newton method. To prevent local minima, five optimization trials with random initial guesses, sampled uniformly from [-10,10], were performed. The covariance functions compared were the exponential spectral angle mapper (ESAM), the observation angle dependent (OAD), the squared exponential (SE), the matern 3/2 (Mat3) and the neural network function (NN). The results from random forest regression (RF) with 100 trees was compared as baseline. All the results were averaged over ten trials. Since, the standard deviations of the r^2 over these trials were insignificant, they have been omitted in the results.

The second experiment evaluated the best performing nitrogen and carbon predicting models from the first experiment on the hyperspectral image. Since, the models were learned only on vegetation spectra, the non-vegetation regions in the test image were masked. To segment the vegetation pixels from non-vegetation pixels, normalized difference vegetation index (NDVI) at each pixel was calculated, and any pixel having a value less that 0.3 was masked. To match the sampling wavelengths of the ground spectra and the hyperspectral pixel, linear interpolation was used. Water absorption bands were removed from both the training spectra and the hyperspectral image. Unfortunately, the ground truth information regarding nitrogen and the carbon contents about the hyperspectral image are unavailable. We were unable to quantitatively validate the performance of using ground spectra trained GP model on airborne hyperspectral imagery.

4.3. Experimental Results

Table 1 shows the r^2 values for prediction of nitrogen and carbon using different covariance functions with the GP. The ni-

Table 1: Predictive performance measured by r^2.

Method	Nitrogen	Carbon
GP-ESAM	0.5316	**0.5592**
GP-OAD	0.5337	**0.5592**
GP-NN	**0.5580**	0.5481
GP-Mat3	0.4424	0.4335
GP-SE	0.4611	0.4145
RF	0.4712	0.4134

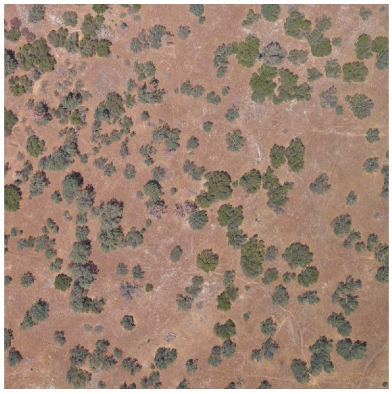
trogen distribution map and the carbon distribution map over the test area in shown in Fig 1.

5. RELATED WORK

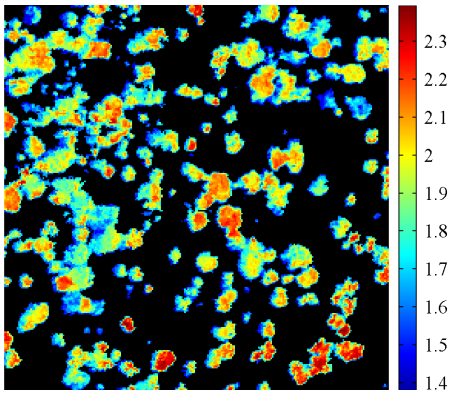
Few covariance functions have been proposed to model hyperspectral data. The observation angle dependent (OAD) covariance function [12] was proposed for Gaussian process classification of hyperspectral data [14]. It uses the angle between the spectra as the similarity metric for covariance computation. Our proposed covariance function, exponential spectral angle mapper (ESAM), is similar to the OAD covariance function. However, the OAD function is linearly related to the spectral angle, while the ESAM function is exponentially related to the spectral angle. Both of these covariance functions are closely related to the neural network covariance function [15] and the arc-cosine kernel function [16]. Functions similar to ESAM have been used as kernel function for hyperspectral classification using support vector machines [17], but there have been no studies to date that uses spectral covariance functions, such as OAD and ESAM, for biochemical estimation from hyperspectral data.

6. CONCLUSION

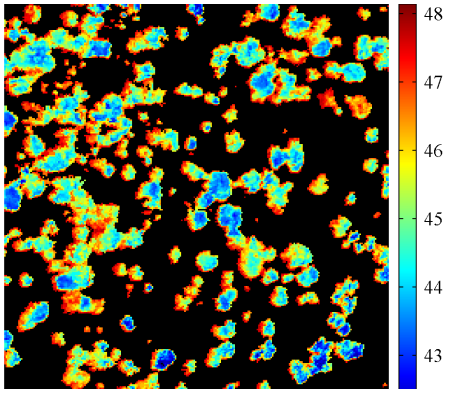
Our experiments demonstrate that Gaussian processes are well suited for biochemical prediction from hyperspectral imagery. They further indicate that the squared exponential covariance function used by most of the previous studies is not optimal for biochemical prediction, and better results could be obtained using non-stationary covariance functions, such as, the neural network covariance function, the OAD covariance function and the ESAM covariance function. The proposed ESAM covariance function demonstrated superior performance, competitive with the state-of-the-art functions, such as the neural network function. It would be interesting to evaluate the performance of the ESAM function for hyperspectral classification using Gaussian processes in future studies.



(a) Color Image



(b) Nitrogen Contents (% of leaf weight)



(c) Carbon Contents (% of leaf weight)

Fig. 1: (a) A colored image of the 250 m $\times$ 250 m test area. (b) Nitrogen distribution map generated by GP-NN. (c) Carbon distribution map generated by GP-ESAM.

7. REFERENCES

[1] M.T. Eismann, *Hyperspectral remote sensing*, SPIE Press, 2012.

[2] A. Ramoelo, A.K. Skidmore, M.A. Cho, R. Mathieu, I.M.A. Heitkönig, N. Dudeni-Tlhone, M. Schlerf, and H.H.T. Prins, "Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 82, pp. 27–40, 2013.

[3] D. Haboudane, J.R. Miller, N. Tremblay, P.J. Zarco-Tejada, and L. Dextraze, "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture," *Remote Sensing of Environment*, vol. 81, no. 23, pp. 416 – 426, 2002.

[4] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1832–1843, 2012.

[5] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 248–252, April 2009.

[6] J. Verrelst, L. Alonso, J.P.R. Caicedo, J. Moreno, and G. Camps-Valls, "Gaussian process retrieval of chlorophyll content from imaging spectroscopy data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 867–874, 2013.

[7] S.L. Ustin, D.A. Roberts, J.A. Gamon, G.P. Asner, and R.O. Green, "Using imaging spectroscopy to study ecosystem processes and properties," *BioScience*, vol. 54, no. 6, pp. 523–534, 2004.

[8] Y. Ge, J.A. Thomasson, and R. Sui, "Remote sensing of soil properties in precision agriculture: A review," *Frontiers of Earth Science*, vol. 5, no. 3, pp. 229–238, 2011.

[9] G. Camps-Valls, L. Gomez-Chova, J. Muoz-Mari, J. Vila-Frances, J. Amoros, S. del Valle-Tascon, and J. Calpe-Maravilla, "Biophysical parameter estimation with adaptive Gaussian processes," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 4, pp. 69–72, 2009.

[10] M. Lazaro-Gredilla, M.K. Titsias, J. Verrelst, and G. Camps-Valls, "Retrieval of biophysical parameters with heteroscedastic Gaussian processes," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 838–842, 2014.

[11] C.E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2005.

[12] A. Melkumyan and E. Nettleton, "An observation angle dependent nonstationary covariance function for Gaussian process regression," in *Neural Information Processing*, Lecture Notes in Computer Science, pp. 331–339. Springer, 2009.

[13] T. Kampe, N. Leisso, J. Musinsky, S. Petroy, B. Karpowicz, K. Krause, R.I. Crocker, M. DeVoe, E. Penniman, T. Guadagno, W. Gallery, T. Ramond, L. Wasser, D. Barnett, J. van Aardt, K. Cawse-Nicholson, and S. Serbin, "The NEON 2013 airborne campaign at domain 17 terrestrial and aquatic sites in California," *NEON Technical Memorandum Series, TM-005*, 2013.

[14] S. Schneider, A. Melkumyan, R.J. Murphy, and E. Nettleton, "Gaussian processes with OAD covariance function for hyperspectral data classification," *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, vol. 1, pp. 393–400, 2010.

[15] C.K.I. Williams, "Computing with infinite networks," *Advances in neural information processing systems (NIPS)*, pp. 295–301, 1997.

[16] Y. Cho and L.K. Saul, "Kernel methods for deep learning," in *Advances in neural information processing systems (NIPS)*, 2009, pp. 342–350.

[17] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2003, vol. 1, pp. 288–290.