BIOPHYSICAL PARAMETER ESTIMATION WITH ADAPTIVE GAUSSIAN PROCESSES

G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, and J. Calpe-Maravilla

Imaging & Processing Laboratory (IPL). Universitat de València. València, Spain.
gustavo.camps@uv.es, http://www.uv.es/gcamps

1. INTRODUCTION

Biophysical parameter estimation and model inversion are active research areas in remote sensing. Lately, *non-parametric* models, such as neural networks and support vector machines (SVM), have demonstrated successful performance [1–3]. Alternative kernel-based prediction models are the Relevance Vector Machines (RVM) and Gaussian Processes (GP) [4, 5], which are formalized under the Bayesian framework. Despite their nice properties for biophysical parameter estimation –accurate predictions, sparse models, confidence intervals– limited attention has been paid in the remote sensing community [6, 7].

Being a kernel-based method, the key for obtaining good models is the definition of a suitable kernel function. The adoption of traditional kernels, such as RBF, does not constitute a good choice since there is no adaptation to the different intrinsic relevance of the spectral channels. In [6] we proposed an alternative RVM to deal with this problem. In this paper we propose adaptive GPs to *learn* the relative relevance of each spectral band by optimizing a different kernel width per band. Besides, GPs can provide an estimate of the noise variance in the data and confidence intervals for the predictions. The proposed adaptive GP is tested in ocean chlorophyll concentration, LAI and sea/land temperature estimation problems.

2. PROPOSED METHOD: ADAPTIVE GAUSSIAN PROCESSES

Gaussian processes provide a probabilistic approach for learning regression and classification problems with kernels [5]. A GP defines a distribution over functions $f : \mathcal{X} \rightarrow \mathbb{R}$ fully described by a mean $m : \mathcal{X} \rightarrow \mathbb{R}$ and a covariance function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that $m(x) = \mathbb{E}[f(x)]$ and $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$. Hereafter we set m to be the zero function for the sake of simplicity. Now, given a finite labeled samples dataset $\{x_1, \dots x_n\}$ we first compute its covariance matrix K in the same way as done for the Gram matrix in SVM. The covariance matrix defines a distribution over the vector of output values $f_x = (f(x_1), \dots, f(x_n))^\top$, such that $f_x \sim \mathcal{N}(\mathbf{0}; K)$, which is a multivariate Gaussian distribution.

For training purposes, we assume that the observed values of the biophysical parameter are only noisy observations of the true underlying function $y = f(x) + \epsilon$. Moreover we assume the noise to be additive independently identically Gaussian distributed with zero mean and variance σ. Let us define the stacked output values $y = (y_1, \dots, y_n)^\top$, the covariance terms of the test point $k_* = (k(x_*; x_1), \dots, k(x_*, x_n))^\top$, and $k_{**} = k(x_*, x_*)$. From the previous model assumption, the output values are distributed according to:

$$\begin{pmatrix} \mathbf{y} \\ f(x_*) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} K + \sigma^2 I & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} \end{pmatrix}\right) \tag{1}$$

For prediction purposes, the GP is obtained by computing the conditional distribution

$$f(x_*)|\mathbf{y}\{x_1, \dots, x_n\}; x_* \sim \mathcal{N}(k_*^\top (K + \sigma^2 I)^{-1}\mathbf{y}, k_{**} - k_*^\top (K + \sigma^2 I)^{-1}\mathbf{k}_* \tag{2}$$

Note that the GP mean predictor is exactly the same solution that the obtained for the kernel ridge regression where now the noise variance term σ^2 appears as the regularization constant. Even more important is the fact that not only a mean prediction is obtained for each sample but a full distribution over the output values including an uncertainty of the prediction.

Traditional choices of the GP covariance (kernel) form involve the radial basis function (RBF) kernel, or the polynomial kernel. A good characteristic of the GP framework is the possibility to optimize all involved hyperparameters, such as different kernel width *per* spectral channel. This is done by maximizing the negative log marginal likelihood and its partial derivatives wrt the hyperparameters. Details on the formulation will be given in the full paper.

This paper has been partially supported by the Spanish Ministry for Education and Science under project CONSOLIDER/CSD2007-00018.

3. EXPERIMENTAL RESULTS

In this preliminary experiment we used *simulated* data of the medium resolution imaging spectrometer (MERIS) to predict chlorophyll concentration, CC, in subsurface waters. We selected the eight channels in the visible range (412-681 nm), following [1]. The range of variation of the chlorophyll concentration is from 0.02 to 25 mg/m^3. The data were generated according to a fixed (noise-free) model and thus uncertainty is not encountered. This fact allows us to test the capabilities of GPs. The total 5000 pairs of *in situ* concentrations and received radiances was randomly divided in a training (500) and validation (500) sets to build the model, and a test set (4000) to assess performance. Before applying the GP, concentration was transformed logarithmically, $\mathbf{y} = \log(CC)$.

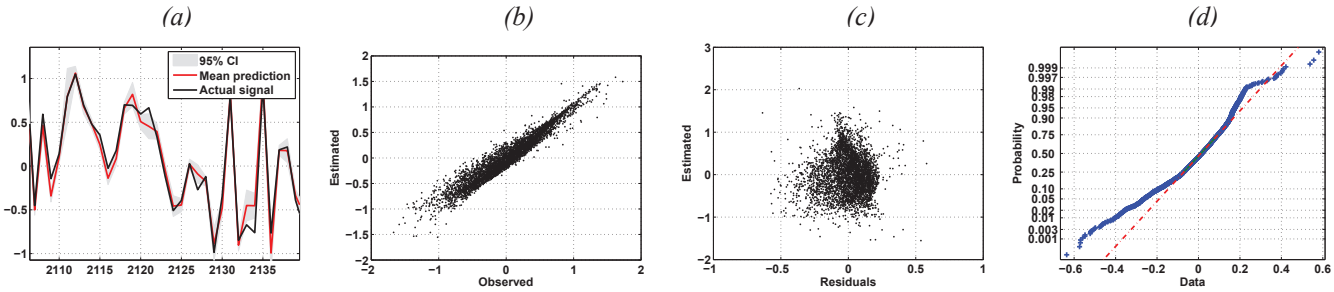

Fig. 1. (a) Actual concentration (black), mean prediction (red) and 95% confidence intervals (gray); (b) predicted versus observed plot; (c) residuals versus predicted plot; and (d) normality plot. All plots are for the test set.

A summary of the results is shown in Fig. 1. GPs provide better results in the test set (RMSE=0.1371, r=0.9610) than in the validation set (RMSE=0.1497, r=0.9535), suggesting good generalization capabilities. It is also noticeable the good alignment in the normality plot. In fact, the estimated noise variance is 0.024 which is a much lower than the signal variance and confirms that the model captured the noise-free nature of the synthetic data. The trained GP also returns the fitted lengthscale per spectral band: $\{227.79, 139.20, 9.34, 15.70, 15.35, 21.79, 11.99, 22.03\}$. The lower the width the higher the relevance of the band, that is the most relevant channels are $\lambda_3 = 490$ nm and $\lambda_7 = 665$ nm, which partially agrees with previous studies [8]. Results on other regression problems, such as LAI estimation and sea/land temperature estimation, will be shown in the conference paper.

4. CONCLUSIONS

An adaptive GP is proposed for model inversion. Preliminary results in the estimation of ocean chlorophyll concentration show good prediction performance. Also GPs provide 1) confidence intervals for the predictions, 2) a noise variance estimation of the data, and 3) allow us to optimize the parameters of the kernel for each spectral band thus providing a feature ranking directly from the model. All these advantages are achieved without any additional computational burden.

5. REFERENCES

[1] P. Cipollini, G. Corsini, M. Diani, and R. Grass, "Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks," *IEEE Trans. On Geosc. Rem. Sens.*, vol. 39, pp. 1508–1524, 2001.

[2] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, "Robust support vector regression for biophysical variable estimation from remotely sensed images," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 3, pp. 339–343, 2006.

[3] G. Camps-Valls, J. Mu noz Marí, L. Gómez-Chova, and J. Calpe-Maravilla, "Semi-supervised support vector biophysical parameter estimation," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'08*, Boston, USA, July 2008.

[4] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. of Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, New York, 2006.

[6] G. Camps-Valls, L. Gomez-Chova, J. Vila-Francés, J. Amorós-López, J. Muñoz-Marí, and J. Calpe-Maravilla, "Retrieval of oceanic chlorophyll concentration with relevance vector machines," *Remote Sensing of Environment*, vol. 105, no. 1, pp. 23–33, Nov 2006.

[7] R. Furfaro, R. D. Morris, A. Kottas, M. Taddy, and B. D. Ganapol, "A Gaussian Process Approach to Quantifying the Uncertainty of Vegetation Parameters from Remote Sensing Observations," *AGU Fall Meeting Abstracts*, pp. A261+, Dec. 2006.

[8] S. Dransfeld, A.R. Tatnall, I.S. Robinson, and C.D. Mobley, "Prioritizing ocean colour channels by neural network input reflectance perturbation," *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 1043–1048, 2005.