NONLINEAR STATISTICAL RETRIEVAL OF SURFACE EMISSIVITY FROM IASI DATA

Valero Laparra[1], Jordi Muñoz-Marí[1], Luis Gómez-Chova[1], Xavier Calbet[2] and Gustau Camps-Valls[1]

(a) Image Processing Laboratory (IPL). Universitat de València, Spain. (b) AEMET, Spain.

ABSTRACT

Emissivity is one of the most important parameters to improve the determination of the troposphere properties (thermodynamic properties, aerosols and trace gases concentration) and it is essential to estimate the radiative budget. With the second generation of infrared sounders, we can estimate emissivity spectra at high spectral resolution, which gives us a global view and long-term monitoring of continental surfaces. Statistically, this is an ill-posed retrieval problem, with as many output variables as inputs. We here propose nonlinear multi-output statistical regression based on kernel methods to estimate spectral emissivity given the radiances. Kernel methods can cope with high-dimensional input-output spaces efficiently. We give empirical evidence of models performance on Infrared Atmospheric Sounding Interferometer (IASI) simulated data. Kernel regression model largely improves previous least squares linear regression model quantitatively, with an average reduction of 25% in mean-square error.

1. INTRODUCTION

Ultraspectral resolution infrared (IR) radiances obtained from nadir observations provide information about the atmosphere, surface, aerosols, and clouds. Surface spectral emissivity (SSE) from current and future operational satellites can and will reveal critical information about the Earth's ecosystem and land-surface-type properties, with implications on the long-term monitoring of the Earth's environment and global climate change [1]. In this study, we deal with the statistical retrieval of SSE from IR data. The inversion scheme has been applied to the Infrared Atmospheric Sounding Interferometer (IASI) synthetic data. The IASI instrument poses a major dimensionality challenge to statistical retrieval algorithms due to its dense spectrum sampling that induces a high dimensional feature problem. IASI spectra consist of 8461 spectral channels, between 3.62 and 15.5 μm, with a spectral resolution of 0.5 cm^{-1} after apodization [2, 3]. Its spatial resolution is 25 km at nadir with an Instantaneous Field of View (IFOV) size of 12 km at an altitude of 819 km. On top of this, we aim to retrieve an also high dimensional state vector; the surface spectral emissivity [1]. This huge data dimensionality typically requires simple and computationally efficient processing techniques.

The remainder of this paper is organized as follows. Section 2 reviews the main characteristics of the employed nonlinear regression method, describes the training strategy, and discuss on the selection of free parameters. Section 3 shows the experimental results, and section 4 concludes.

This paper has been partially supported by the European Research Council under Consolidator Grant SEDAL ERC-2014-CoG 647423, and the Spanish Ministry of Economy and Competitiveness (MINECO) through the projects TIN2015-64210-R and TEC2016-77741-R.

2. NONLINEAR STATISTICAL RETRIEVAL

In this work, we aim to compare the power of nonlinear regression versus linear regression (LR) for emissivity retrievals. We concentrate on the field of kernel methods, which deal with the concept of regularization and capacity control naturally. An introduction to kernel methods in remote sensing data analysis is available in [4]. Specifically, we will focus on the kernel implementation of the standard (regularized) least squares linear regression. The method is known as kernel ridge regression (KRR) or least squares support vector machine [5]. Kernel regression gives rise to non-parametric and nonlinear decision functions: the nonlinearity is obtained implicitly through the use kernels, which are controlled by the so-called kernel hyperparameters, while the regularization is tunned via a conditioning constant (which is related to the noise variance). In this section, we first review the standard formulation of the KRR method departing from the LR model. Then, we discuss on the alternatives to deal with (very large) multi-output regression problems as in our case of emissivity estimation. Finally, we study the sensitivity of the hyperparameters and discuss on for tuning them.

2.1. Kernel Ridge Regression

We are given n pairs of $\mathbf{x}_i \in \mathbb{R}^D$ (spectra) and $\mathbf{y}_i \in \mathbb{R}^M$ (emissivity), where $i = 1, ..., n$ indicates the index of the n training samples (or field of views in our context, FOVs). In matrix notation, the input data (radiances) are collectively expressed as $\mathbf{X} \in \mathbb{R}^{n \times D}$, and the outputs (emissivitites) are given by $\mathbf{Y} \in \mathbb{R}^{n \times M}$. The canonical linear model assumes $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{W}$, and typically one estimates the weights/coefficients $\mathbf{W}$ by least squares minimization under the assumption of an additive $i.i.d.$ noise model, $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{N}$

with Gaussian noise $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ of zero mean and standard deviation σ_n.

The Kernel Ridge Regression (KRR) is a nonlinear version of the previous linear regression model based on the use of kernels [6, 5]. Now, we want to fit a linear model in an alternative Hilbert space, $\mathcal{H}$, of very high dimensionality (possibly infinite) $D_{\mathcal{H}}$, where samples have been mapped to through a mapping $\boldsymbol{\phi} : \mathbf{x}_i \in \mathbb{R}^D \rightarrow \boldsymbol{\phi}(\mathbf{x}_i) \in \mathbb{R}^{D_{\mathcal{H}}}$. The prediction model is thus given by $\hat{\mathbf{Y}} = \boldsymbol{\Phi}\mathbf{W}_{\mathcal{H}}$, where the bias term was intentionally dropped as it can be easily fitted.

The problem is solved by minimizing the regularized squared loss function

$$\mathcal{L} = |\mathbf{Y} - \boldsymbol{\Phi}\mathbf{W}_{\mathcal{H}}\|^2 + \lambda\|\mathbf{W}_{\mathcal{H}}\|^2.$$

One proceeds as in standard least squares regression: first takes derivatives w.r.t. $\mathbf{W}_{\mathcal{H}}$ and equates it to zero, which gives the (primal) solution $\mathbf{W}_{\mathcal{H}} = (\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda\mathbf{I})^{-1}\boldsymbol{\Phi}^\top\mathbf{Y}$, where $\boldsymbol{\Phi}$ is the matrix of mapped samples whose size is now $n \times D_{\mathcal{H}}$. Then, by applying the Representer's theorem[1], the dual solution

$$\boldsymbol{\alpha} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \lambda\mathbf{I})^{-1}\mathbf{Y} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{Y},$$

where we replaced the inner product matrix with a similarity matrix between samples, known as the *kernel matrix*, $\mathbf{K}$. This training matrix is squared, symmetric and positive definite, and it contains all the similarities between training samples, $[\mathbf{K}]_{ij} = K(\mathbf{x}_i,\mathbf{x}_j)$. The problem is now solvable and closed-form, since we only need to compute the inverse of the (regularized) matrix kernel $\mathbf{K}$ of size $n \times n$. By applying the same trick to new incoming test samples, $\mathbf{X}_*$, it is easy to show that one can do retrievals in testing without the explicit mapping to Hilbert spaces, just resorting to similarities between training and test samples, i.e.: $\hat{\mathbf{Y}}_* = \boldsymbol{\Phi}_*\mathbf{W} = \boldsymbol{\Phi}_*\boldsymbol{\Phi}^\top\boldsymbol{\alpha} = \mathbf{K}_*\boldsymbol{\alpha}$, where the matrix $\mathbf{K}_*$ contains the similarities between all test and training samples.

2.2. Large input and output regression problems

In the emissivity estimation problem we have huge both input and output dimensional data, which makes the direct application of the previous methods unaffordable. Remember that the IASI instrument provides radiances (and hence emissivities) in $D = M = 8461$ spectral channels [3]. In this case, we follow the strategy in [1] by which the models are applied to the top *principal component scores* of both $\mathbf{X}$ and $\mathbf{Y}$, that is $\mathbf{X}_p = \mathbf{X}\mathbf{V}_x$ and $\mathbf{Y}_q = \mathbf{Y}\mathbf{V}_y$ [1], where p and q are the number of component scores selected for $\mathbf{X}$ and $\mathbf{Y}$ respectively. This strategy is very effective, and it is motivated by the high spectral correlation of both radiances and emissivities. It is also computationally very efficient because only one model is used for the estimation of the emissivities, instead of using M models or alternatively the top q scores.

[1]The representer's theorem states that one can express the solution matrix $\mathbf{W}_{\mathcal{H}}$ defined in $\mathcal{H}$ as a linear combination of mapped samples in that space, $\mathbf{W}_{\mathcal{H}} = \boldsymbol{\Phi}^\top\boldsymbol{\alpha}$.

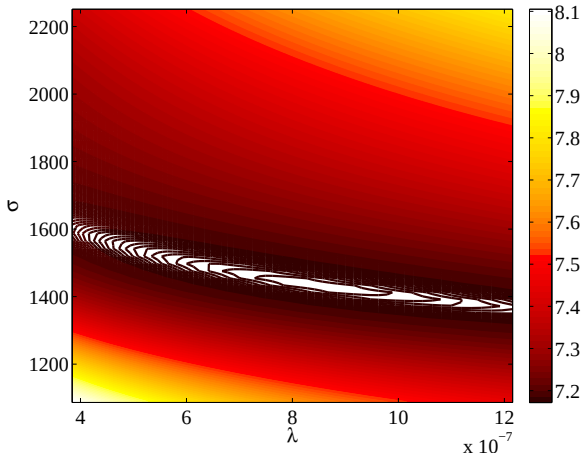

Fig. 1. Error surface (RMSE) for different combinations of the KRR hyperparameters (σ, λ).

2.3. KRR selection parameters

In our experiments we use the squared exponential kernel, also known as Radial Basis Function (RBF) kernel: $K(\mathbf{x}_i,\mathbf{x}_j)$ = exp(-$\|\mathbf{x}_i - \mathbf{x}_j\|^2/(2\sigma^2)$). With this kernel, only two free parameters have to be tuned in KRR, the regularization parameter λ, and the kernel parameter σ, which were selected via standard cross-validation [2]. Essentially, a training set was split into training and validation input-output pairs; the former was used to adjust the dual weights $\boldsymbol{\alpha}$, and the latter to evaluate its performance for each particular choice of (λ, σ). After the best pair (λ, σ) was found, a final set of coefficients $\boldsymbol{\alpha}$ were used to generate predictions on an independent (unseen) test set. Figure 1 shows the error surface as a function of the KRR parameters. We analyzed a grid of several combinations of the parameters. We observed a clear minimum for a particular combination of the parameters λ and σ.

3. EXPERIMENTAL RESULTS

This section shows the experimental results when using LR and KRR to predict emissivities using IASI data.

3.1. Experimental setup

In this work, we follow the same scheme presented in [1] to evaluate the use of non-linear retrieval models for retrieving emissivities using radiances. For this task, we employ IASI simulated data together with the corresponding emissivities. We reduce the dimensionality of the input data (radiances) and the output data (emissivities) by using Principal Components Analysis (PCA) [7]. Following [1] we reduce the

[2]We have used the KRR implementation included in the simpleR toolbox: `http://www.uv.es/gcamps/software.html`

radiances dimensionality to $p = 200$ components, and the emissivity dimensionality to $q = 11$ components.

3.2. Model comparison

We evaluate two different models: Tikhonov regularized linear regression (LR), and Kernel Ridge Regression (KRR), which has been previously used in atmospheric parameter retrieval [4, 8, 9, 10, 11], see sec. 2.1 for details. The two methods are trained using 20000 samples and a cross-validation scheme where half of the data are used to obtain different sets of regression coefficients for each combination of parameters. The other half of the training data is used to test all sets of obtained coefficients, and to select the one that obtains the smallest error. The data to test the models in the 3.3 section is different from the data employed for training. Figure 1 shows the error surface when using different parameters with KRR, see sec. 2.3 for details.

3.3. Emmissivity retrieval

Here we test the trained model on the test data, 47475 samples. Figure 2(a) shows the error of each regression model for estimating the value of the dimensions corresponding to the principal components of the emissivities. KRR clearly obtains less error than LR. This difference is more noticeable in the first principal components. Since KRR is better for predicting these values, the error has to be also smaller when reconstructing the emissivities using the KRR predictions.

Figures 2(b) and 2(c) show the bias error and the root mean square error (RMSE) for the reconstructed (inverted) emissivities[3]. As in Fig. 2(a), KRR errors are smaller than for LR, for both bias and RMSE.

Although differences in the performance between LR and KRR are clear, we also show in figure 2(d) the relative results between KRR method and LR in percentage. This figure is helpful to analyze the difference in performance for each particular wave number.

3.4. Final remarks

Two main conclusions can be derived from the experiments. First, the cross-validation strategy yielded a clear optimum point in the error surface to select the best hyperparameters of the KRR algorithm (see Fig. 1). And second, KRR clearly outperforms LR both in bias and RMSE. Actually, KRR achieves for some wavenumbers 50% less error than LR, and an average reduction of 25% in mean-square error. It is noticeable that most of the gain is obtained for the first PCs (the ones with higher eigenvalues).

[3]Note that PCA learns a projection matrix $\mathbf{V} \in \mathbb{R}^{d \times d}$ which is orthogonal, $\mathbf{V}^\top \mathbf{V} = \mathbf{I}$, so the inverse transformation is just its transpose, $\mathbf{V}^{-1} = \mathbf{V}^\top$. Hence, since we predicted PCA-projected emissivities, $\hat{\mathbf{Y}} = \mathbf{E}\mathbf{V}$, we can derive the estimated emissivities simply as $\hat{\mathbf{E}} = \hat{\mathbf{Y}}\mathbf{V}^\top$.

4. CONCLUSIONS

In this paper we presented a kernel-based regression algorithm for the statistical retrieval of emissivities from IASI data. Given the high input and output data dimensionality, we applied KRR in dual PCA/EOF subspaces. The proposed methodology largely improved the accuracy results over previous methodologies based on LR.

Future work will be addressed in two directions. On the one hand, we aim to improve the regression scheme with alternatives to PCA [12, 13] for dimensionality reduction, and to KRR for computational efficiency [14] and uncertainty characterization [15]. On the other hand, the generality of the scheme allows its application to data coming from other sensors, such as the upcoming Meteosat Third Generation infrared sounder (MTG-IRS).

5. REFERENCES

[1] Daniel K. Zhou, Allen M. Larar, Xu Liu, William L. Smith, L. Larrabee Strow, Ping Yang, Peter Schlussel, and Xavier Calbet, "Global Land Surface Emissivity Retrieved From Satellite Ultraspectral IR Measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 1277–1290, 2011.

[2] G. Chalon, F. Cayla, and D. Diebel, "IASI: an advanced sounder for operational meteorology," in *Proceedings of the 52nd Congress of IAF*, Toulouse, France, 2001.

[3] Chalon G. Siméoni D., Singer C., "Infrared atmospheric sounding interferometer," *Acta Astronautica*, vol. 40, pp. 113–118, 1997.

[4] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for Remote Sensing Data Analysis*, Wiley & Sons, UK, Dec 2009.

[5] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[6] B. Schölkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press Series, 2002.

[7] I.T. Jolliffe, *Principal component analysis*, Springer, 2002.

[8] G. Camps-Valls, V. Laparra, J. Muñoz-Marí, L. Gómez-Chova, and X. Calbet, "Kernel-based retrieval of atmospheric profiles from IASI data," in *IEEE Proc. IGARSS 11*, Jul 2011, pp. 2813–2816.

[9] G. Camps-Valls, J. Muñoz and, L. Gómez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Trans. Geosci. Rem. Sens.*, vol. 50, no. 5, pp. 1759–1769, 2012.

[10] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1759–1769, 2012.

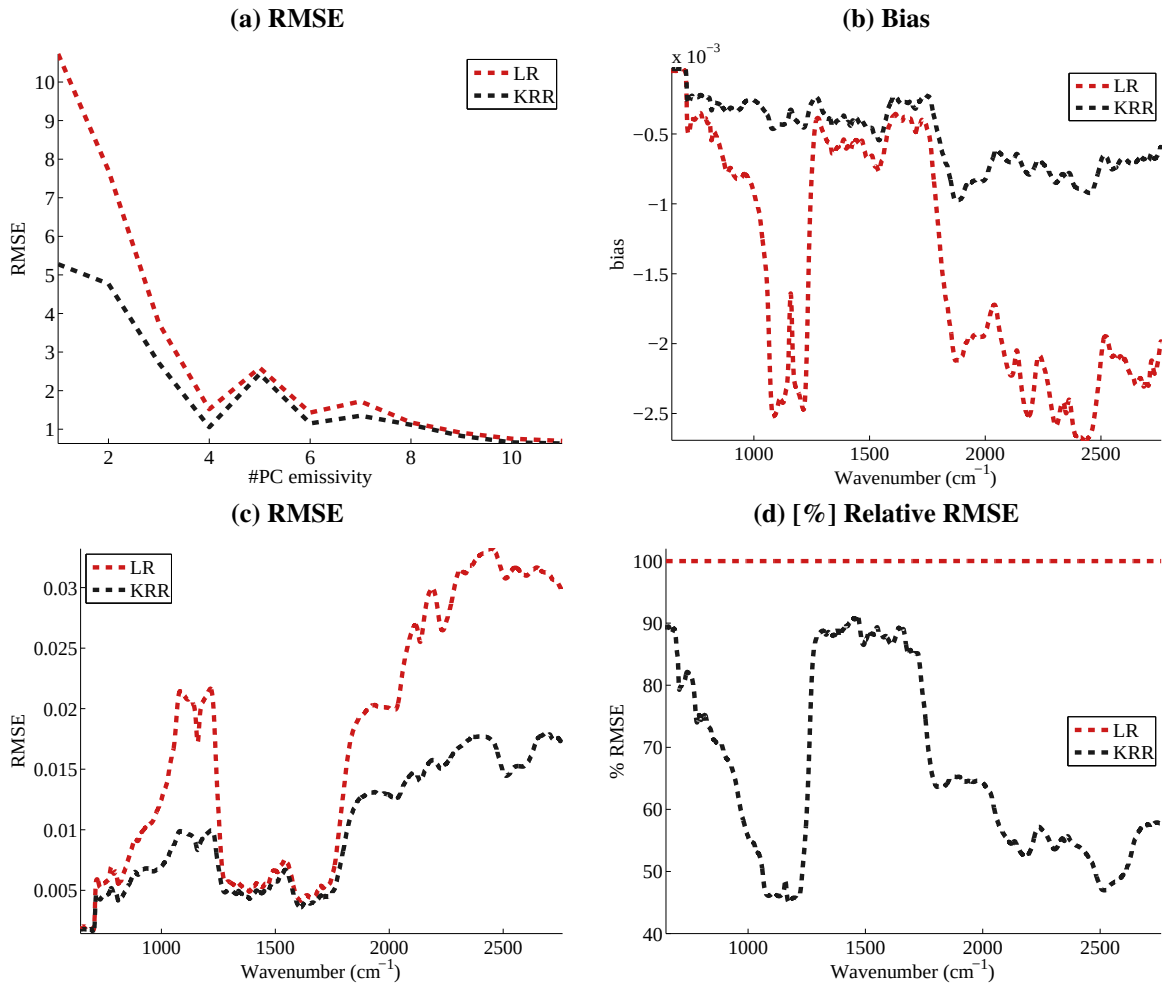

Fig. 2. **Regression results.** The regression models tested are regularized linear regression (LR) and Kernel Ridge Regression (KRR). a) RMSE of each method to predict the emissivities principal components. b) Bias of each method predicting the emissivities for each. c) RMSE of each method predicting the emissivities for each wave number. d) % RMSE of each method divided by the RMSE of the LR method.

[11] J. García Sobrino, Serra-Sagristà, V. Laparra, X. Calbet, and G. Camps-Valls, "Statistical atmospheric parameter retrieval largely benefits from spatial-spectral image compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 1, no. 1, pp. 1–1, 2016.

[12] J. Arenas-García, K.B. Petersen, G. Camps-Valls, and L.K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 16–29, 2013.

[13] V. Laparra, J. Malo, and G. Camps-Valls, "Dimensionality reduction via regression in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1026–1036, Sept 2015.

[14] V. Laparra, D.M. Gonzalez, D. Tuia, and G. Camps-Valls, "Large-scale random features kernel regression," in *IEEE International Geoscience and Remote Sensing Symposium 2015 (IGARSS 2015)*, 26-31 July, 2015, Milan, Italy, 2015, pp. 1–4.

[15] G. Camps-Valls, J. Verrelst, J. Muñoz-Marí, V. Laparra, F. Mateo-Jiménez, and J. Gómez-Dans, "A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 58–78, June 2016.