

HYPER-SPECTRAL IMAGE ANALYSIS WITH PARTIALLY-LATENT REGRESSION

Antoine Deleforge, Florence Forbes and Radu Horaud

INRIA Grenoble Rhône-Alpes, France

ABSTRACT

The analysis of hyper-spectral images is often needed to recover physical properties of planets. To address this inverse problem, the use of learning methods have been considered with the advantage that, once a relationship between physical parameters and spectra has been established through training, the learnt relationship can be used to estimate parameters from new images underpinned by the same physical model. Within this framework, we propose a partially-latent regression method which maps high-dimensional inputs (spectral images) onto low-dimensional responses (physical parameters). We introduce a novel regression method that combines a Gaussian mixture of locally-linear mappings with a partially-latent variable model. While the former makes high-dimensional regression tractable, the latter enables to deal with physical parameters that cannot be observed or, more generally, with data contaminated by experimental artifacts that cannot be explained with noise models. The method is illustrated on images collected from the Mars planet.

Index Terms— Hyper-spectral images; Regression; Dimension reduction; Mixture models; Latent variable model.

1. INTRODUCTION

The analysis of hyper-spectral images often involves to solve an inverse problem to deduce a number of physical parameter values from the observed spectra. This typically requires the estimation of a high-dimensional to low-dimensional mapping, which is challenging. Among others, statistical learning methods have been considered with the advantage that, once a relationship between physical parameters and spectra has been established through training, the learnt relationship can subsequently be used to estimate parameters from new images underpinned by the same physical model [1]. To obtain training data, radiative transfer models have been developed, that link the chemical composition, the granularity, or the physical state, to the observed spectrum. They are generally used to simulate huge collections of spectra in order to perform the inversion of hyper-spectral images [2].

Within this framework, we propose a novel regression method. Regression is a fully supervised method that uses input-response pairs of observed data to infer a mapping, and this mapping is then used to predict a response value from an observed input. It is well known that high-dimensional regression is a difficult problem because of the very high number of parameters to be estimated. To address this problem we propose a probabilistic framework, first to learn the parameters of a low-to-high Gaussian mixture of locally-linear regressions, and second to derive the posterior conditional density of the response given an observed input and the estimated model parameters. We show that, by exchanging the roles of input and response variables during training, high-dimensional regression becomes tractable and the conditional density of the response has a closed-form solution. Moreover, we propose to incorporate a latent component: while the high-dimensional variable remains fully observed, the low-dimensional variable is the concatenation of an observed vector and of an unobserved (latent) vector. The latent component enables us to deal with those physical parameters that cannot be observed during training, or to prevent the observed parameters to be contaminated by experimental artifacts that cannot be explained with noise models.

This model is particularly suited for hyper-spectral imaging, because the computing resources required by radiative transfer models to generate training data increases exponentially with the number of parameters. Thus, they are generally restricted to a small number of parameters, *e.g.*, abundance and grain size of the main chemical components. Other parameters, such as those related to meteorological variability or the incidence angle of the spectrometer are not explicitly modeled and measured. Our method allows to deal with such *non-anotated effects* by explicetely incorporating a latent part in the low dimensional response variable. This allows some form of *slack* in the observations due to unknown effects.

Regression has been extensively studied in the literature. To deal with high-dimensional input data, most methods first reduce the dimensionality based on response variables, before performing the actual regression [3–5]. Such two-step approaches cannot be conveniently expressed in terms of a single optimization problem. To estimate non-linear mappings, mixtures of locally linear models have been investigated [6–8], but not in the case of high dimensional inputs

Support from the European Research Council (ERC) through the Advanced Grant VHIA (#340113) is greatly acknowledged.

and partially-observed responses. An alternative popular approach is to use kernel functions, [4, 9–11], with the drawback that these functions cannot be appropriately chosen automatically, and that the mappings learned cannot be inverted. A more detailed description of the proposed approach and its relation to existing methods is available as a research report [12].

2. REGRESSION WITH PARTIALLY-LATENT RESPONSE VARIABLES

We describe a regression method that maps a low-dimensional variable \mathbf{X} onto a high-dimensional variable \mathbf{Y} . We start by describing the GLLiM (Gaussian Locally Linear Mapping) model, in which both \mathbf{X} and \mathbf{Y} are fully observed. Then we introduce the *hybrid* GLLiM (hGLLiM) model that treats \mathbf{X} as a partially latent variable. Once the parameters of the hybrid GLLiM model are estimated, we provide a closed-form expression for the high-dimensional to low-dimensional mapping. GLLiM relies on a piecewise linear model in the following way. Let $\{\mathbf{x}_n\}_{n=1}^{n=N} \in \mathbb{R}^L$, $\{\mathbf{y}_n\}_{n=1}^{n=N} \in \mathbb{R}^D$, $L \ll D$, and (\mathbf{y}, \mathbf{x}) is a realization of (\mathbf{Y}, \mathbf{X}) , such that \mathbf{y} is the image of \mathbf{x} by an affine transformation τ_k , among K , plus an error term. This is modeled by a missing variable Z such that $Z = k$ if and only if \mathbf{Y} is the image of \mathbf{X} by τ_k .

The following decomposition of the joint probability distribution will be used: $p(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}; \boldsymbol{\theta}) = \sum_{k=1}^K p(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}, Z = k; \boldsymbol{\theta}) p(\mathbf{X} = \mathbf{x} | Z = k; \boldsymbol{\theta}) p(Z = k; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the vector of model parameters. The locally affine function that maps \mathbf{X} onto \mathbf{Y} is:

$$\mathbf{Y} = \sum_{k=1}^K \mathbb{I}(Z = k) (\mathbf{A}_k \mathbf{X} + \mathbf{b}_k + \mathbf{E}_k) \quad (1)$$

where \mathbb{I} is the indicator function, matrix $\mathbf{A}_k \in \mathbb{R}^{D \times L}$ and vector $\mathbf{b}_k \in \mathbb{R}^D$ define the transformation τ_k and $\mathbf{E}_k \in \mathbb{R}^D$ is an error term capturing both the observation noise in \mathbb{R}^D and the reconstruction error due to the local affine approximation. Under the assumption that \mathbf{E}_k is a zero-mean Gaussian variable with covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{R}^{D \times D}$ that does not depend on \mathbf{X} , \mathbf{Y} , and Z , we obtain:

$$p(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}, Z = k; \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}; \mathbf{A}_k \mathbf{x} + \mathbf{b}_k, \boldsymbol{\Sigma}_k). \quad (2)$$

To complete the above hierarchical definition and enforce the affine transformations to be local, \mathbf{X} is assumed to follow a mixture of K Gaussians defined by

$$p(\mathbf{X} = \mathbf{x} | Z = k; \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}; \mathbf{c}_k, \boldsymbol{\Gamma}_k) \\ p(Z = k; \boldsymbol{\theta}) = \pi_k$$

where $\mathbf{c}_k \in \mathbb{R}^L$, $\boldsymbol{\Gamma}_k \in \mathbb{R}^{L \times L}$ and $\sum_{k=1}^K \pi_k = 1$. The parameters of this model are:

$$\boldsymbol{\theta} = \{\mathbf{c}_k, \boldsymbol{\Gamma}_k, \pi_k, \mathbf{A}_k, \mathbf{b}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K. \quad (3)$$

The model just described can be learned with standard EM inference methods if \mathbf{X} and \mathbf{Y} are both observed. The key idea in this paper is to treat \mathbf{X} as a *partially-latent* variable, namely

$$\mathbf{X} = \begin{bmatrix} \mathbf{T} \\ \mathbf{W} \end{bmatrix},$$

where $\mathbf{T} \in \mathbb{R}^{L_t}$ is observed and $\mathbf{W} \in \mathbb{R}^{L_w}$ is latent ($L = L_t + L_w$). In hybrid GLLiM, parameter estimation uses observed pairs $\{\mathbf{y}_n, \mathbf{t}_n\}_{n=1}^N$ while it must also be constrained by the presence of the latent variable \mathbf{W} . This can be seen as a *latent-variable augmentation* of classical regression, where the observed realizations of \mathbf{Y} are affected by the unobserved variable \mathbf{W} . It can also be viewed as a variant of dimensionality reduction since the unobserved low-dimensional variable \mathbf{W} must be recovered from $\{(\mathbf{y}_n, \mathbf{t}_n)\}_{n=1}^N$. The decomposition of \mathbf{X} into observed and latent parts implies that some of the model parameters must be decomposed as well, namely \mathbf{c}_k , $\boldsymbol{\Gamma}_k$ and \mathbf{A}_k . Assuming the independence of \mathbf{T} and \mathbf{W} given Z we write:

$$\mathbf{c}_k = \begin{bmatrix} \mathbf{c}_k^t \\ \mathbf{c}_k^w \end{bmatrix}, \boldsymbol{\Gamma}_k = \begin{bmatrix} \boldsymbol{\Gamma}_k^t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_k^w \end{bmatrix}, \mathbf{A}_k = \begin{bmatrix} \mathbf{A}_k^t & \mathbf{A}_k^w \end{bmatrix}.$$

We devised two EM algorithms to estimate the parameters of the proposed model. The principle of the suggested algorithms is based on a data augmentation strategy that consists of augmenting the observed variables with the unobserved ones, in order to facilitate the subsequent maximum-likelihood search over the parameters. There are two sets of missing variables, $Z_{1:N} = \{Z_n\}_{n=1}^N$ and $\mathbf{W}_{1:N} = \{\mathbf{W}_n\}_{n=1}^N$, associated with the training data set $(\mathbf{y}, \mathbf{t})_{1:N} = \{\mathbf{y}_n, \mathbf{t}_n\}_{n=1}^N$, given the number K of linear components and the latent dimension L_w . Two augmentation schemes arise naturally. The first scheme is referred to as general hybrid GLLiM-EM, or *general-hGLLiM*, and consists of augmenting the observed data with both variables $(Z, \mathbf{W})_{1:N}$ while the second scheme, referred to as *marginal-hGLLiM*, consists of integrating out the continuous variables $\mathbf{W}_{1:N}$ previous to data augmentation with the discrete variables $Z_{1:N}$. A particularly interesting feature of general-hGLLiM is that both the E-step and M-step can be computed in closed-form for various constraints on the noise covariances $\{\boldsymbol{\Sigma}_k\}_{k=1}^K$, including “equal for all k ”, isotropic, and diagonal constraints. On the other hand, marginal-hGLLiM only has closed-form steps in the isotropic case, but is easier to initialize. Therefore, the latter is used to initialize the former in practice. Full details on the analytical steps of the resulting *hGLLiM* algorithm are given in [12], and a Matlab toolbox implementing it is available online¹.

Once the parameter vector $\boldsymbol{\theta}$ has been estimated by the

¹https://team.inria.fr/perception/gllim_toolbox/

algorithm, one can derive the following conditional density:

$$p(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}^*) = \sum_{k=1}^K \frac{\pi_k^* \mathcal{N}(\mathbf{y}; \mathbf{c}_k^*, \boldsymbol{\Gamma}_k^*)}{\sum_{j=1}^K \pi_j^* \mathcal{N}(\mathbf{y}; \mathbf{c}_j^*, \boldsymbol{\Gamma}_j^*)} \mathcal{N}(\mathbf{x}; \mathbf{A}_k^* \mathbf{y} + \mathbf{b}_k^*, \boldsymbol{\Sigma}_k^*) \quad (4)$$

which is a Gaussian mixture with K components. Notice that the parameters associated with this mixture:

$$\boldsymbol{\theta}^* = \{\mathbf{c}_k^*, \boldsymbol{\Gamma}_k^*, \pi_k^*, \mathbf{A}_k^*, \mathbf{b}_k^*, \boldsymbol{\Sigma}_k^*\}_{k=1}^K \quad (5)$$

are obtained analytically from the learned parameters (3) using the following formulae:

$$\begin{aligned} \mathbf{c}_k^* &= \mathbf{A}_k \mathbf{c}_k + \mathbf{b}_k, \quad \boldsymbol{\Gamma}_k^* = \boldsymbol{\Sigma}_k + \mathbf{A}_k \boldsymbol{\Gamma}_k \mathbf{A}_k^\top, \quad \pi_k^* = \pi_k, \\ \mathbf{A}_k^* &= \boldsymbol{\Sigma}_k^* \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1}, \quad \mathbf{b}_k^* = \boldsymbol{\Sigma}_k^* (\boldsymbol{\Gamma}_k^{-1} \mathbf{c}_k - \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{b}_k), \\ \boldsymbol{\Sigma}_k^* &= (\boldsymbol{\Gamma}_k^{-1} + \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{A}_k)^{-1}. \end{aligned} \quad (6)$$

To predict an output one can use the expectation of (4):

$$\mathbb{E}[\mathbf{X} | \mathbf{y}; \boldsymbol{\theta}^*] = \sum_{k=1}^K \frac{\pi_k \mathcal{N}(\mathbf{y}; \mathbf{c}_k^*, \boldsymbol{\Gamma}_k^*)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{y}; \mathbf{c}_j^*, \boldsymbol{\Gamma}_j^*)} (\mathbf{A}_k^* \mathbf{y} + \mathbf{b}_k^*). \quad (7)$$

3. RETRIEVING MARS PHYSICAL PROPERTIES FROM HYPER-SPECTRAL IMAGES

Visible and near infrared imaging spectroscopy is a key remote sensing technique used to study and monitor planets. It records the visible and infrared light reflected from the planet in a given wavelength range and produces cubes of data where each observed surface location is associated with a spectrum. Physical properties of a planet surface, such as chemical composition, granularity, texture, etc., are some of the most important parameters that characterize the morphology of the spectra. In case of Mars, radiative transfer models have been developed to numerically evaluate the connection between these parameters and observable spectra. Such models allow to simulate spectra from a given set of parameter values [2]. In practice, the goal is to scan Mars's ground from an orbit in order to observe gas and dust in the atmosphere and seek for signs of specific materials such as silicates, carbonates, or ice at the surface. We are therefore interested in solving the associated *inverse problem*, namely to infer physical parameter values from the observed spectra. Since this inverse problem cannot generally be solved analytically, the use of optimization or statistical methods has been investigated, e.g., [1]. In particular, supervised statistical learning has been considered with the advantage that, once a relationship between physical parameters and spectra has been established through training, the learned relationship can then be used for very large datasets and for all new images underpinned by the same physical model.

Within this category of methods, we investigate the potential of the proposed hybrid GLLiM model using a dataset of hyper-spectral images collected from the imaging spectrometer OMEGA instrument [13] onboard of the Mars express spacecraft. To this end a database of synthetic spectra with their associated physical parameter values were generated using a radiative transfer model. This database is composed of 15,407 spectra associated with five real parameter values, namely, (i) *proportion of water ice*, (ii) *proportion of CO₂ ice*, (iii) *proportion of dust*, (iv) *grain size of water ice*, and (v) *grain size of CO₂ ice*. Each spectrum is made of 184 wavelengths. Hybrid GLLiM is used, first to learn a *low-to-high dimensional regression function* between physical parameters and spectra from the database, and second to estimate the unknown physical parameters corresponding to a new spectrum using the learned function. Since no ground truth is available for Mars, the synthetic database will also serve as a test dataset to evaluate the accuracy of the predicted parameter values. In order to fully illustrate the potential of hybrid GLLiM, we deliberately ignore two of the five parameters in the database and consider them as *latent variables*. During training, we excluded the observed values of the proportion of water ice and the grain size of CO₂ ice. A previous study [1] revealed that these two parameters were sensitive to the same wavelengths, in comparison with the proportion of dust and are suspected to *mix* with the other two parameters in the synthetic transfer model. Therefore, not only that they are harder to estimate but they seem to affect the estimation of two other parameters. Moreover, we observed that if these two parameters are treated as observed variables during the learning stage, they tend to degrade the estimation of the other three parameters, which are of particular interest.

The hGLLiM algorithm was compared to JGMM (joint Gaussian mixture model) [7], SIR (sliced inverse regression) [3], RVM (multivariate relevance vector machine) [11] and MLE (mixture of linear experts) [6]. SIR is used with one (SIR-1) or two (SIR-2) principal axes for dimensionality reduction, 20 slices (the number of slices is known to have very little influence on the results), and polynomial regression of order three (higher orders did not show significant improvements in our experiments). SIR quantizes the low-dimensional data \mathbf{X} into *slices* or clusters which in turn induces a quantization of the \mathbf{Y} -space. Each \mathbf{Y} -slice (all points \mathbf{y}_n that map to the same \mathbf{X} -slice) is then replaced with its mean and PCA is carried out on these means. The resulting dimensionality reduction is then informed by \mathbf{X} values through the preliminary slicing. RVM may be viewed as a multivariate probabilistic formulation of *support vector regression* [14]. As all kernel methods, it critically depends on the choice of a kernel function. Using the authors' freely available code², we ran preliminary tests to determine an optimal kernel choice for each dataset considered. We tested

²http://www.mvrvm.com/Multivariate_Relevance_Vector

Table 1. Normalized root mean squared error (NRMSE) for Mars surface physical properties recovered from hyper-spectral images, using synthetic data and different methods.

Method	Proportion of CO2 ice	Proportion of dust	Grain size of water ice
JGMM	0.83 ± 1.61	0.62 ± 1.00	0.79 ± 1.09
SIR-1	1.27 ± 2.09	1.03 ± 1.71	0.70 ± 0.94
SIR-2	0.96 ± 1.72	0.87 ± 1.45	0.63 ± 0.88
RVM	0.52 ± 0.99	0.40 ± 0.64	0.48 ± 0.64
MLE	0.54 ± 1.00	0.42 ± 0.70	0.61 ± 0.92
hGLLiM-1	0.36 ± 0.70	0.28 ± 0.49	0.45 ± 0.75
hGLLiM-2*†	0.34 ± 0.63	0.25 ± 0.44	0.39 ± 0.71
hGLLiM-3	0.35 ± 0.66	0.25 ± 0.44	0.39 ± 0.66
hGLLiM-4	0.38 ± 0.71	0.28 ± 0.49	0.38 ± 0.65
hGLLiM-5	0.43 ± 0.81	0.32 ± 0.56	0.41 ± 0.67
hGLLiM-20	0.51 ± 0.94	0.38 ± 0.65	0.47 ± 0.71
hGLLiM-BIC	0.34 ± 0.63	0.25 ± 0.44	0.39 ± 0.71

14 kernel types with 10 different scales ranging from 1 to 30, hence, 140 kernels for each dataset in total.

An objective evaluation was done by cross validation. We selected 10,000 training couples at random from the training set, tested on the 5,407 remaining spectra, and repeated this 20 times. For all algorithms, training data were normalized to have 0 mean and unit variance using scaling and translating factors. These factors were then used on test data and estimated output to obtain final estimates. This technique showed to noticeably improve results of all methods. We used $K = 50$ for MLE, hGLLiM and JGMM. MLE and JGMM were constrained with equal, diagonal covariance matrices as they yield the best results. For RVM, the best out of 140 kernels was used. A third degree polynomial kernel with scale 6 showed the best results using cross-validation on a subset of the database. As a quality measure of the estimated parameters, we computed the normalized root mean squared error, or NRMSE³, which quantifies the difference between the estimated and real parameter values. NRMSE is normalized enabling direct comparison between the parameters which are of very different range: the closer to zero, the more accurate the predicted values. Table 1 shows NRMSE values obtained for the three parameters considered here. The ground-truth latent variable dimension is $L_w^* = 2$, and accordingly, the empirically best dimension for hGLLiM was $L_w^\dagger = 2$. hGLLiM-2 outperformed all the other methods on that task, more precisely the error is 36% lower than the second best-performing method, RVM, closely followed by MLE. No significant difference was observed between hGLLiM-2 and hGLLiM-3. Note that due to the computation of the $D \times D$ kernel matrix, the computational and memory costs of RVM for training were about 10 times higher than those of hGLLiM, using MATLAB implementations. We also tested the selection of the latent dimension (L_w) based on BIC [12]. For each train-

ing set, the hGLLiM-BIC method minimized BIC with the latent dimension in the range $0 \leq L_w \leq 20$, and used the corresponding model to perform the regression. Interestingly, hGLLiM-BIC performed very well on these large training sets ($N = 10,000$) as it correctly selected $L_w = 2$ for the 20 training sets used by BIC (the BIC selection could differ with the training set), yielding the same results as those obtained with hGLLiM-2.

Finally, we used an adequately selected subset of the synthetic database, *e.g.*, [1] to train the algorithms, and test them on real data made of observed spectra. In particular, we focus on a dataset of Mars South polar cap. Since no ground truth is currently available for the physical properties of Mars’s South pole regions, we propose a qualitative evaluation using hGLLiM-2 and the three best performing methods, among the tested ones, namely RVM, MLE and JGMM.

We used these four methods in order to extract physical parameters from two hyper-spectral images that correspond, approximately, to the same area but from different view points (orbit 41 and orbit 61). Since we are looking for proportions between 0 and 1, values smaller than 0 or higher than 1 are not acceptable and hence they were set to either one of the bounds. As it can be seen in Fig. 1, hGLLiM-2 estimates proportion maps with similar characteristics for the two view points, which suggests a good consistency of the method. Such a consistency is not observed using the other three tested methods. In addition, RVM and MLE estimate a much higher number of values falling outside the interval $[0, 1]$. Moreover, hGLLiM-2 is the only method featuring less dust at the South pole cap center and higher concentrations of dust at the boundaries of the CO2 ice, which matches expected results from planetology [15]. Finally, note that the proportions of CO2 ice and dust clearly seem to be complementary using hGLLiM-2, while this complementarity is less obvious using other methods.

³NRMSE = $\sqrt{\frac{\sum_{m=1}^M (\bar{t}_m - t_m)^2}{\sum_{m=1}^M (t_m - \bar{t})^2}}$ with $\bar{t} = M^{-1} \sum_{m=1}^M t_m$.

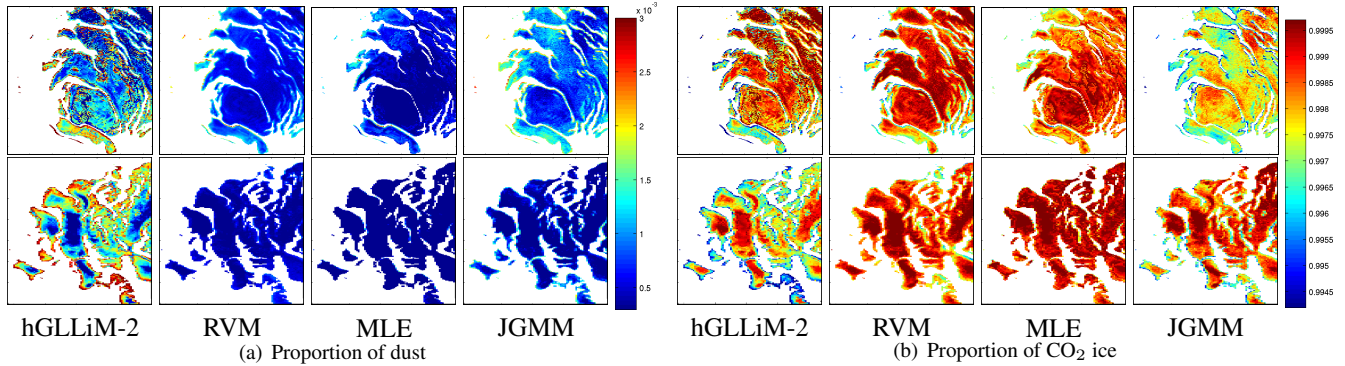


Fig. 1. Proportions obtained with our method (hGLLiM-2) and three other methods. The data correspond to hyper-spectral images grabbed from two different viewpoints of the South polar cap of Mars. Left: orbit 41, Right: orbit 61. White areas correspond to unexamined regions, where the synthetic model does not apply.

4. CONCLUSIONS

We proposed a new high-dimensional regression method allowing for partially latent responses. This novel approach outperformed 4 state-of-the-art regression methods in hyper-spectral image analysis. A promising extension would be to take into account dependencies between adjacent pixels in images to enforce smoothness. This could be done using Markov Random field. More complex noise models could also be investigated via Student distributions (*e.g.* [8]) to allow for outliers accommodation and more robust estimation.

REFERENCES

- [1] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard, “Retrieval of Mars surface physical properties from OMEGA hyperspectral images using regularized sliced inverse regression,” *Journal of Geophysical Research: Planets*, vol. 114, no. E6, 2009.
- [2] S. Douté, E. Deforas, F. Schmidt, R. Oliva, and B. Schmitt, “A comprehensive numerical package for the modeling of Mars hyperspectral images,” in *Lunar and Planetary Science XXXVIII*, March 2007.
- [3] K. C. Li, “Sliced inverse regression for dimension reduction,” *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 316–327, 1991.
- [4] H.M. Wu, “Kernel sliced inverse regression with applications to classification,” *Journal of Computational and Graphical Statistics*, vol. 17, no. 3, pp. 590–610, 2008.
- [5] K. P. Adragni and R. D. Cook, “Sufficient dimension reduction and prediction in regression,” *Philosophical Transactions of the Royal Society A*, vol. 367, no. 1906, pp. 4385–4405, 2009.
- [6] L. Xu, M. I. Jordan, and G. E. Hinton, “An alternative model for mixtures of experts,” in *Advances in Neural Information Processing Systems*. December 1995, pp. 633–640, MIT Press, Cambridge, MA, USA.
- [7] Y. Qiao and N. Minematsu, “Mixture of probabilistic linear regressions: A unified view of GMM-based mapping techniques,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2009, pp. 3913–3916.
- [8] S. Ingrassia, S. C. Minotti, and G. Vittadini, “Local statistical modeling via a cluster-weighted approach with elliptical distributions,” *Journal of classification*, vol. 29, no. 3, pp. 363–401, 2012.
- [9] M.E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [10] N. Lawrence, “Probabilistic non-linear principal component analysis with gaussian process latent variable models,” *The Journal of Machine Learning Research*, vol. 6, pp. 1783–1816, 2005.
- [11] A. Thayananthan, R. Navaratnam, B. Stenger, P. Torr, and R. Cipolla, “Multivariate relevance vector machines for tracking,” in *European Conference on Computer Vision*. 2006, pp. 124–138, Springer.
- [12] A. Deleforge, F. Forbes, and R. Horaud, “High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables,” *Statistics and Computing*, 2014.
- [13] J-P Bibring et al., “OMEGA: Observatoire pour la Minéralogie, l’Eau, les Glaces et l’Activité,” in *Mars Express: The Scientific Payload*, 2004, vol. 1240, pp. 37–49.
- [14] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [15] S. Douté et al., “Nature and composition of the icy terrains of the south pole of Mars from MEX OMEGA observations,” in *Lunar and Planetary Science XXXVI*, March 2005.