High dimensional regression with Gaussian mixtures and partially latent response variables: Application to **hyper-spectral image analysis**

- Antoine Deleforge, Florence Forbes, Radu Horaud –

INRIA Grenoble Rhone-Alpes
Mistis & Perception teams

OUTLINE

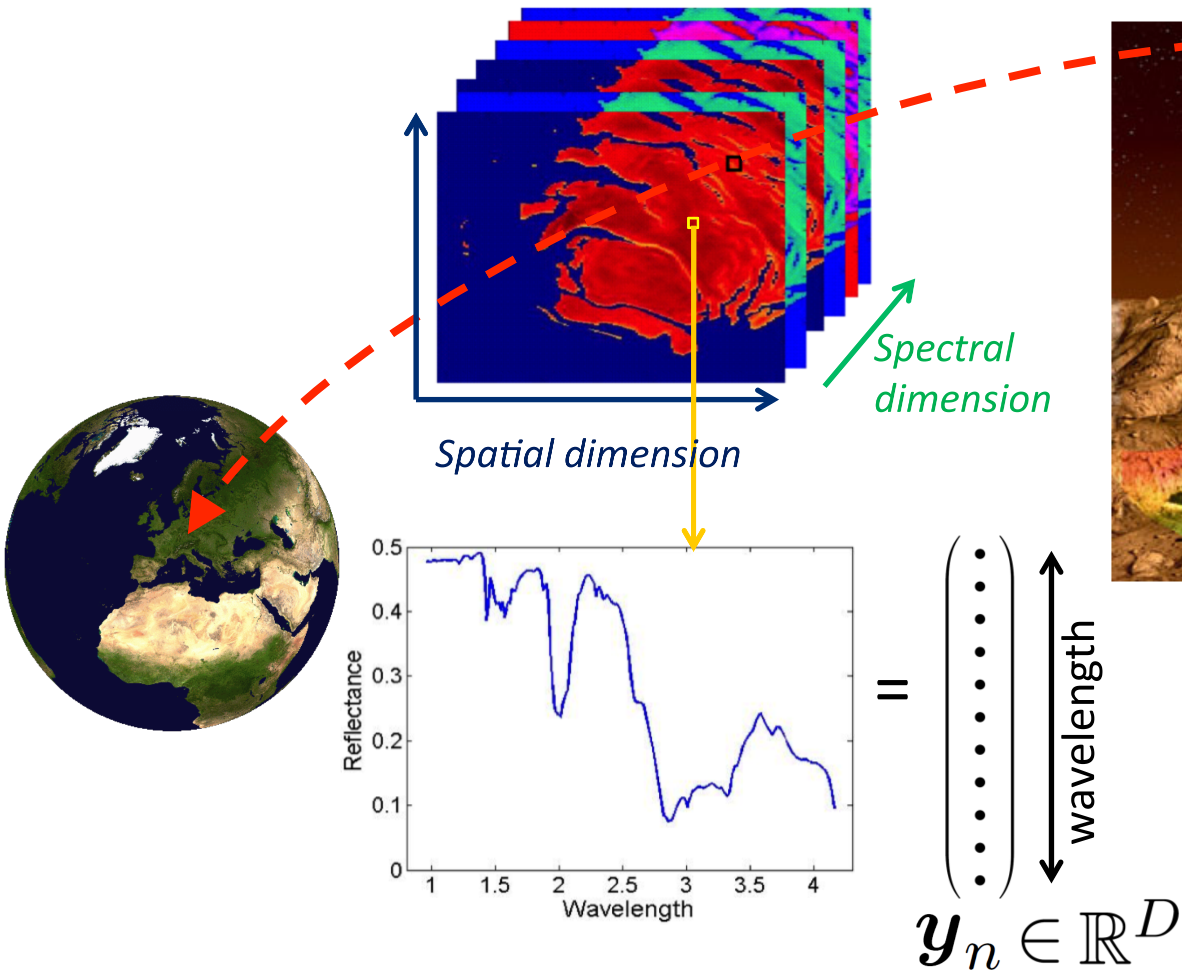
1. Introduction
2. High-Dimensional Regression with GLLiM
3. Extension to Partially-Latent Output
4. Results
5. Conclusion

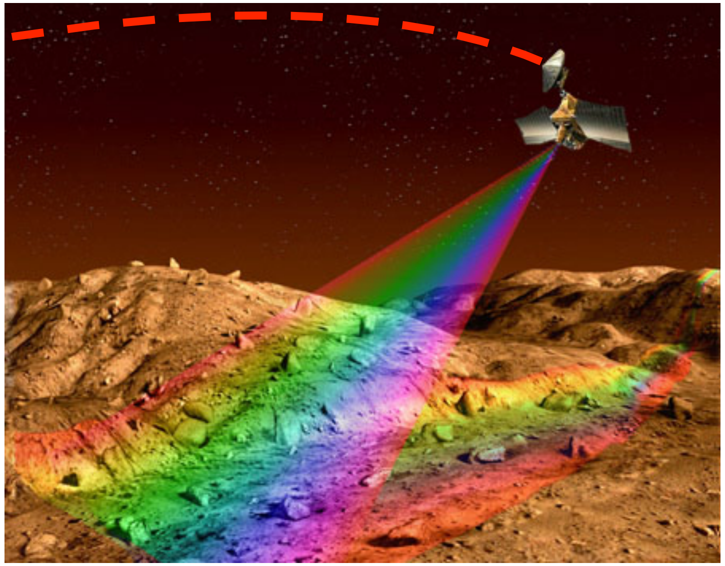
OUTLINE

1. **Introduction**
2. High-Dimensional Regression with GLLiM
3. Extension to Partially-Latent Output
4. Results
5. Conclusion

1 Introduction

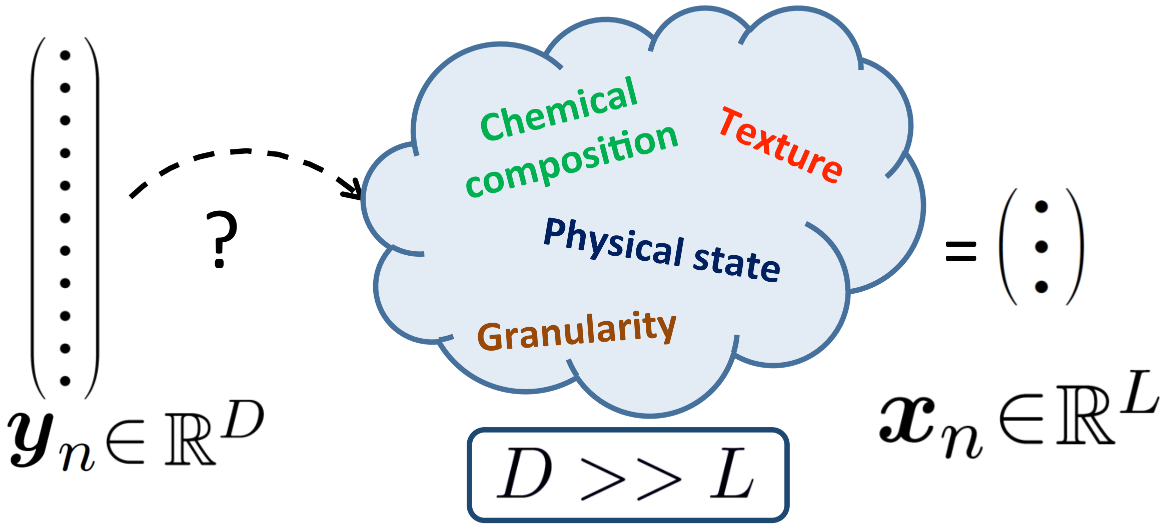
Problem: Retrieving physical properties from hyperspectral images





Mars Express - Omega (2004)
[http://geops.geol.u-psud.fr/]

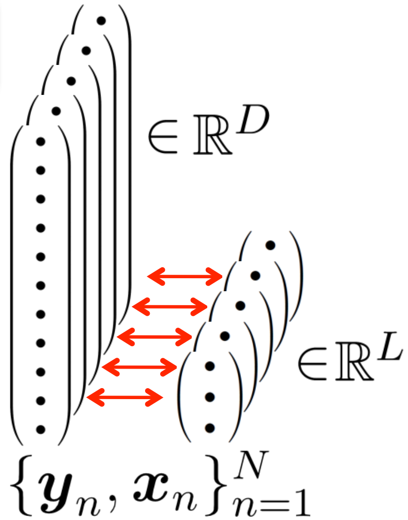
Problem: Retrieving physical properties from hyperspectral images



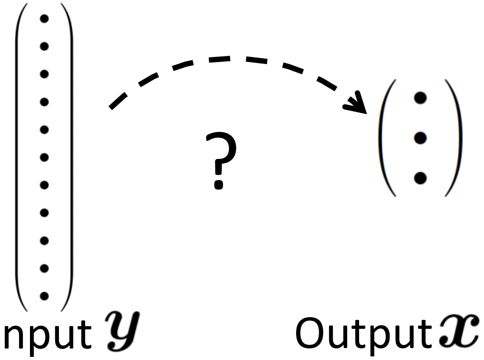
A Regression Problem

Training

Associated input-output pairs



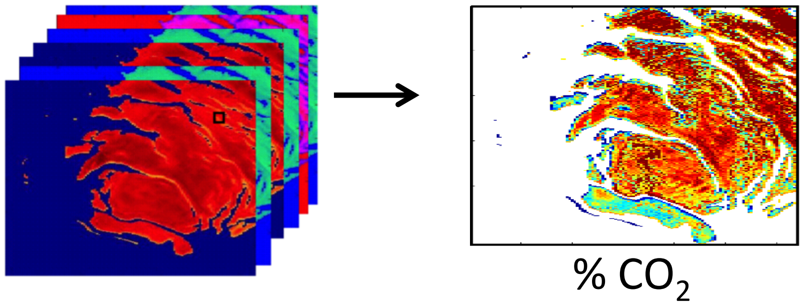
Testing



Training data pairs can be **synthetized** using a *radiative transfer model* (Douté et al. 2007)

Learn ***x=f(y)***

Physical properties can be retrieved from real data using the learned function ***f*** :



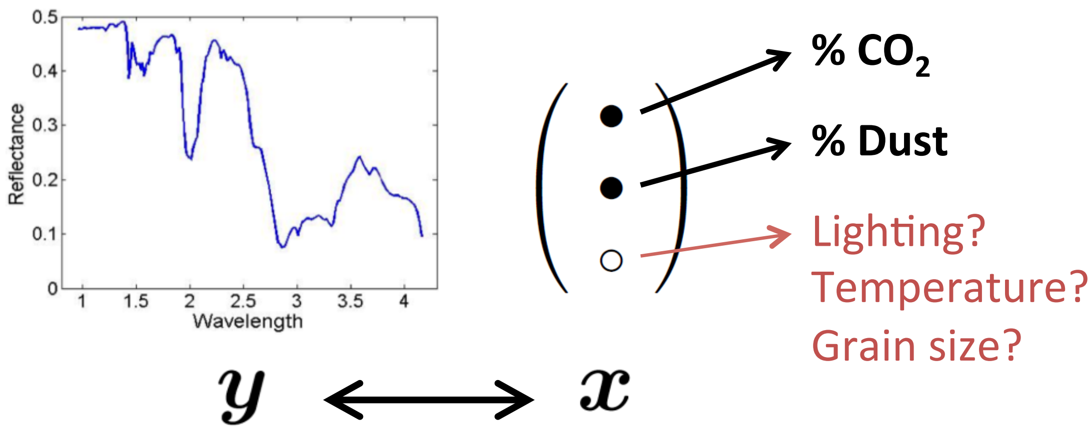
Challenges

1) The **input** is **high-dimensional**

$$\boldsymbol{x} = f(y_1, y_2, \ldots, y_D) \;?$$

Learn the inverse **low-to-high** regression

2) The **output** may be **partially-annotated**



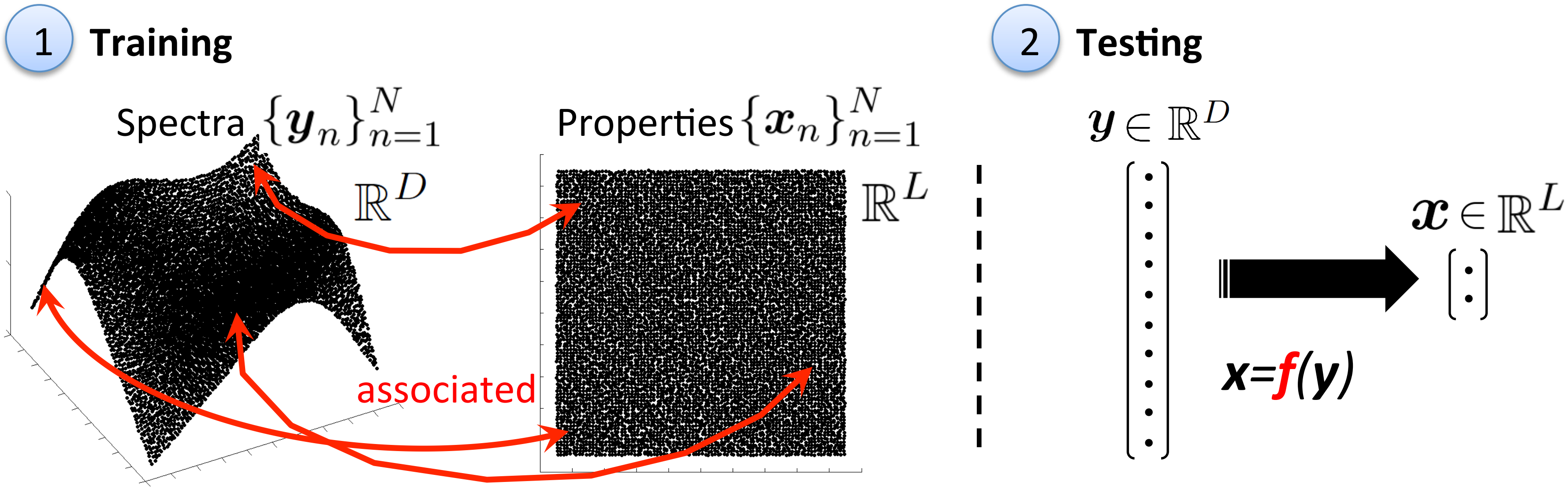
Extension to **partially-latent output**

OUTLINE

1. **Introduction**
2. High-Dimensional Regression with GLLiM
3. Extension to Partially-Latent Output
4. Results
5. Conclusion

OUTLINE

2 High-Dimensional Regression with GLLiM



Problem: The mapping is non-linear

Solution: ***Gaussian Locally Linear Mapping (GLLiM)***

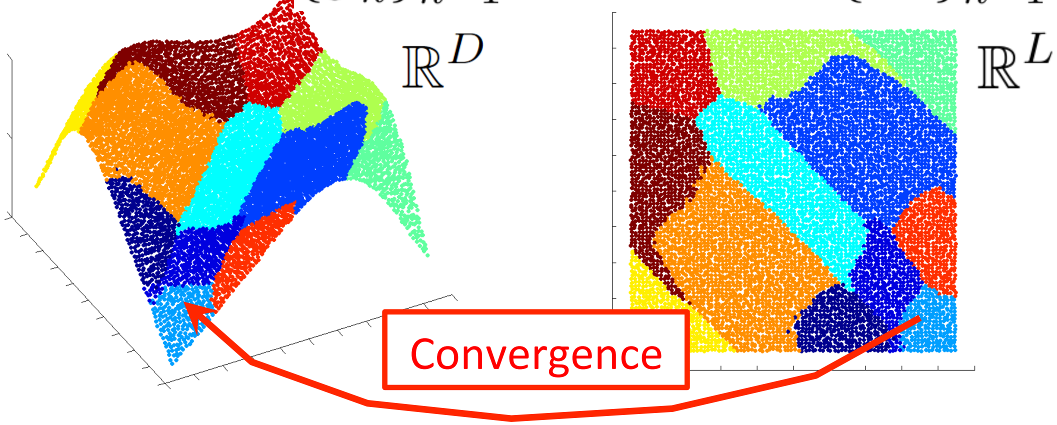
Problem: High- to low-dimensional regression is hard because f has a high-dimensional support: $f(y_1,y_2,\dots,y_D)$

Solution: - Learn the regression **the *other way around***: $y=g(x)$

- Use the inverse function $f=g^{-1}$ to map y to x.

Gaussian Locally-Linear Mapping (GLLiM)

Spectra $\{\boldsymbol{y}_n\}_{n=1}^N$ Properties $\{\boldsymbol{x}_n\}_{n=1}^N$



$$\boldsymbol{y}_n = \mathbf{A}_k \boldsymbol{x}_n + \boldsymbol{b}_k + \boldsymbol{e}_n$$

Closed-form EM algorithm

E-Step Posterior update

Assign points to ***regions***

M-Step Parameters update

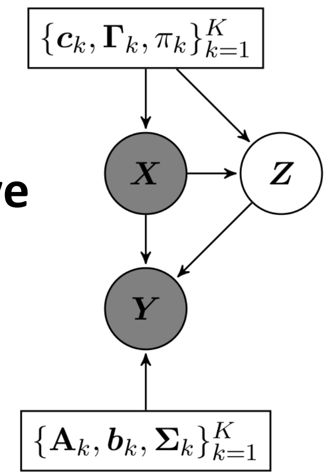
Calculate ***transformations***

$$p(\boldsymbol{x}_n | z_{kn} = 1; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{c}_k, \boldsymbol{\Gamma}_k)$$

$$p(z_{kn} = 1; \boldsymbol{\theta}) = \pi_k$$

$$p(\boldsymbol{y}_n | z_{kn} = 1, \boldsymbol{x}_n; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}_n; \mathbf{A}_k \boldsymbol{x}_n + \boldsymbol{b}_k, \boldsymbol{\Sigma})$$

Generative Model



2 High-Dimensional Regression with GLLiM

Learned $\boldsymbol{\theta}$

- The *inverse conditional density:*

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

↑ spectrum ↑ properties

→ Inverse mapping: ***y*** = *g*(***x***)= $E[\mathbf{Y}|\mathbf{X}=\mathbf{x}\ ;\ \theta]$

- The ***forward conditional density*** (Bayes' inversion)

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

↑ properties ↑ spectrum

→ Forward mapping: ***x*** = *f*(***y***)= $E[\mathbf{X}|\mathbf{Y}=\mathbf{y}\ ;\ \theta]$

$$\left(\begin{array}{lll} \text{where } \boldsymbol{c}_k^* = \mathbf{A}_k\boldsymbol{c}_k + \boldsymbol{b}_k, & & \boldsymbol{\Sigma}_k^* = (\boldsymbol{\Gamma}_k^{-1}+\mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1}\mathbf{A}_k)^{-1}, \\ \boldsymbol{\Gamma}_k^* = \boldsymbol{\Sigma}_k + \mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^\top, & & \boldsymbol{b}_k^* = \boldsymbol{\Sigma}_k^*(\boldsymbol{\Gamma}_k^{-1}\boldsymbol{c}_k - \mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1}\boldsymbol{b}_k) \\ \mathbf{A}_k^* = \boldsymbol{\Sigma}_k^*\mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1}, & & \end{array}\right)$$

2 High-Dimensional Regression with GLLiM

Regression: *low-to-high* vs. *high-to-low*?

- Example: $D = 1000,\ L = 2,\ K = 10$,
  Isotropic and equal noise covariances $\{\boldsymbol{\Sigma}_k\}_{k=1}^{K}$

- Low-to-high regression ($\boldsymbol{X} \rightarrow \boldsymbol{Y}$) model size:

$$\mathcal{D}(\boldsymbol{\theta}) = K(1+L+DL+L^2+D+1) = 30,080$$

- High-to-low regression ($\boldsymbol{Y} \rightarrow \boldsymbol{X}$) model size:

$$\mathcal{D}(\boldsymbol{\theta}^*) = K(1 + D + LD + D^2 + L + 1) = 10,030,040$$

  \+ requires the inversion of 1000 x 1000 covariance matrices

OUTLINE

OUTLINE

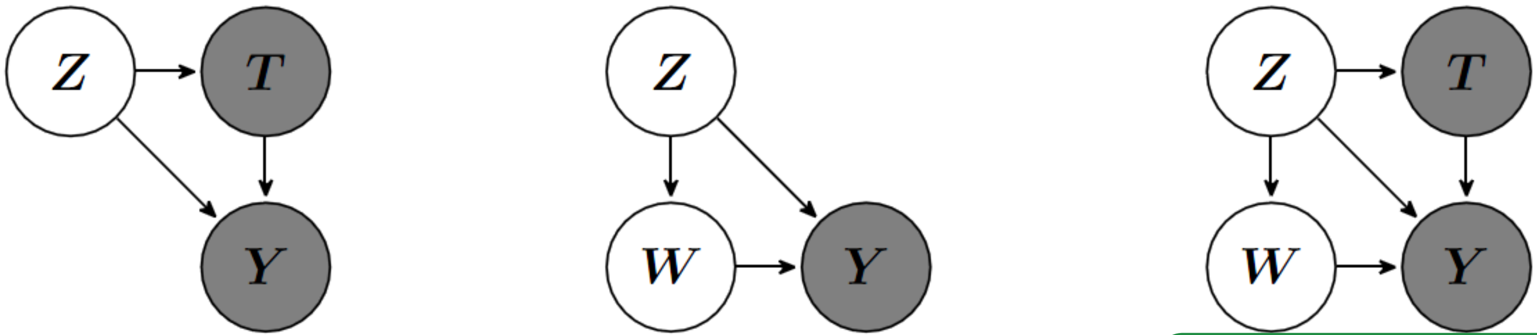
1. Introduction
2. High-Dimensional Regression with GLLiM
3. **Extension to Partially-Latent Output**
4. Results
5. Conclusion

The Hybrid GLLiM Model

The low-dimensional output variable is split into a fully-observed and a fully-latent component:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{T} \\ \boldsymbol{W} \end{bmatrix}$$

Where $\boldsymbol{T} \in \mathbb{R}^{L_{\mathrm{t}}}$ is observed and $\boldsymbol{W} \in \mathbb{R}^{L_{\mathrm{w}}}$ is latent ($L = L_{\mathrm{t}} + L_{\mathrm{w}}$)



Supervised GLLiM ($\boldsymbol{X} \equiv \boldsymbol{T}$)

Unsupervised GLLiM ($\boldsymbol{X} \equiv \boldsymbol{W}$)

Hybrid GLLiM ($\boldsymbol{X} \equiv [\boldsymbol{T}; \boldsymbol{W}]$)

3 Extension to Partially-Latent Output

- **Good news: A general closed-form EM algorithm**

- Observed: $\{(\boldsymbol{Y}_n, \boldsymbol{T}_n), n = 1:N\}$, missing: $\{(Z_n, \boldsymbol{W}_n), n = 1:N\}$

E-Step Update posterior probabilities:

- $p(Z_n = k|\boldsymbol{t}_n, \boldsymbol{y}_n; \boldsymbol{\theta}^{(i)}) = \dfrac{\pi_k^{(i)} p(\boldsymbol{y}_n, \boldsymbol{t}_n | Z_n = k; \boldsymbol{\theta}^{(i)})}{\sum_{j=1}^K \pi_j^{(i)} p(\boldsymbol{y}_n, \boldsymbol{t}_n | Z_n = j; \boldsymbol{\theta}^{(i)})}$ } **« GMM like »**

- $p(\boldsymbol{w}_n | Z_n = k, \boldsymbol{t}_n, \boldsymbol{y}_n; \boldsymbol{\theta}^{(i)}) \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}_{nk}^{\mathrm{w}}, \widetilde{\mathbf{S}}_k^{\mathrm{w}})$ } **« Probabilistic PCA or Factor Analyse like »**

3 Extension to Partially-Latent Output

- **Good news: A general closed-form EM algorithm**

- Observed: $\{(\boldsymbol{Y}_n, \boldsymbol{T}_n), n = 1:N\}$, missing: $\{(Z_n, \boldsymbol{W}_n), n = 1:N\}$

M-Step Update parameters

- $\pi_k, \boldsymbol{c}_k^{\mathrm{t}}, \boldsymbol{\Gamma}_k^{\mathrm{t}}$ } **« GMM like »**

- $\mathbf{A}_k, \boldsymbol{b}_k, \boldsymbol{\Sigma}_k$ } A hybrid between **linear regression** and **PPCA / FA** :

$$\widetilde{\mathbf{A}}_k = \widetilde{\mathbf{Y}}_k \widetilde{\mathbf{X}}_k^\top \left(\begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\mathbf{S}}_k^{\mathrm{w}} \end{bmatrix} + \widetilde{\mathbf{X}}_k \widetilde{\mathbf{X}}_k^\top \right)^{-1}$$

*A **hybrid** method between **regression** and **dimensionality reduction***

The Hybrid GLLiM algorithm - Particular instances

Model	c_k	Γ_k	π_k	$\mathbf{A}_k$	$\boldsymbol{b}_k$	Σ_k	L_t	L_w	K
MLE [Xu et al 95]	-	-	-	-	-	diag	-	0	-
MLR [Jedidi et al 96]	$\mathbf{0}_L$	$\infty\mathbf{I}_L$	-	-	-	iso+eq	-	0	-
JGMM [Qiao et al 09]	-	-	-	-	-	-	-	0	-
PPAM [Deleforge et al 12]	-	|eq|	eq	-	-	diag+eq	-	0	-
GTM [Bishop et al 98]	fixed	$\mathbf{0}_L$	eq.	eq.	$\mathbf{0}_D$	iso+eq	0	-	-
PPCA [Tipping et al 99a]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	iso	0	-	1
MPPCA [Tipping et al 99b]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	iso	0	-	-
MFA [Ghahramani et al 96]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	diag	0	-	-
PCCA [Bach et al 05]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	block	0	-	1
RCA [Kalaitzis et al 11]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	fixed	0	-	1

- Blue = **regression** methods
- Red = **dimensionality reduction** methods

OUTLINE

1 Introduction

2 High-Dimensional Regression with GLLiM

3 Extension to Partially-Latent Output

4 Results

5 Conclusion

OUTLINE

1 Introduction

2 High-Dimensional Regression with GLLiM

3 Extension to Partially-Latent Output

4 **Results**

5 Conclusion

Synthetic Data

- 15,407 spectra (D = 187 wavelengths) using a *radiative transfer model* [Douté et al. 2007]
- 5 physical parameters: % CO_2 ice, % Water ice, % Dust, Grain size CO_2, Grain size water

Ignored during training

- Harder to estimate
- Deteriorate estimation of other parameters

⇒ $L_t = 3$, $L_w = 2$

- Test hGLLiM with different values of L_w and **BIC**

$$BIC(\widetilde{\boldsymbol{\theta}}, N) = -2\mathcal{L}(\widetilde{\boldsymbol{\theta}}) + \mathcal{D}(\widetilde{\boldsymbol{\theta}}) \log N$$

- Training on 10,000 randomly selected spectra, testing on the remaining ones (repeated 20 times)

4 Results

Synthetic Data

Normalized root mean square errors for different methods

Method	Proportion of CO2 ice	Proportion of dust	Grain size of water ice
JGMM [1]	0.83 ± 1.61	0.62 ± 1.00	0.79 ± 1.09
SIR-1 [2]	1.27 ± 2.09	1.03 ± 1.71	0.70 ± 0.94
SIR-2 [2]	0.96 ± 1.72	0.87 ± 1.45	0.63 ± 0.88
RVM [3]	0.52 ± 0.99	0.40 ± 0.64	0.48 ± 0.64
MLE [4]	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*†	$\mathbf{0.34 \pm 0.63}$	$\mathbf{0.25 \pm 0.44}$	$\mathbf{0.39 \pm 0.71}$
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	$\mathbf{0.34 \pm 0.63}$	$\mathbf{0.25 \pm 0.44}$	$\mathbf{0.39 \pm 0.71}$

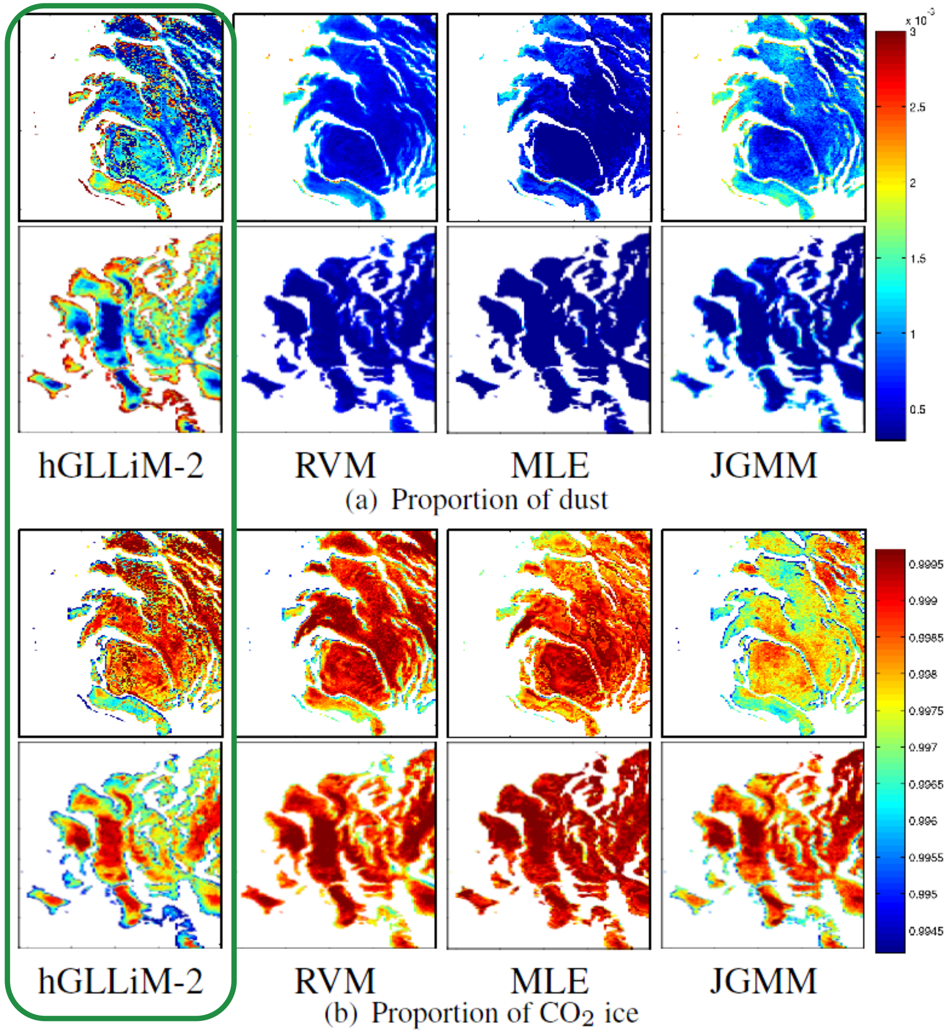
[1] Y. Qiao et al. « Mixture of Probablistic Linear Regressions (…) » (2009) ⟷ **hGLLiM-D**

[2] K. C. Li. « Sliced-Inverse Regression for Dimension Reduction » (1991)

[3] A. Thayananthan et al. « Multivariate Relevance Vector Machines for Tracking » (2006)

[4] L. Xu et al. « An alternative Model for Mixture of Experts » (1995) ⟷ **hGLLiM-0**

Real Data



- Hyperspectrometer OMEGA (Mars Express 2004)

- Two points of view of Mars' south polar cap (Orbits 41 and 61)

- Training with synthetic data

- *No ground-truth available*

- Our method:
  - Consistency between the two orbits
  - Complementarity of proportions
  - Higher concentration of dust on the edge of the glacier
  - smoothness?

OUTLINE

OUTLINE

5 Conclusion

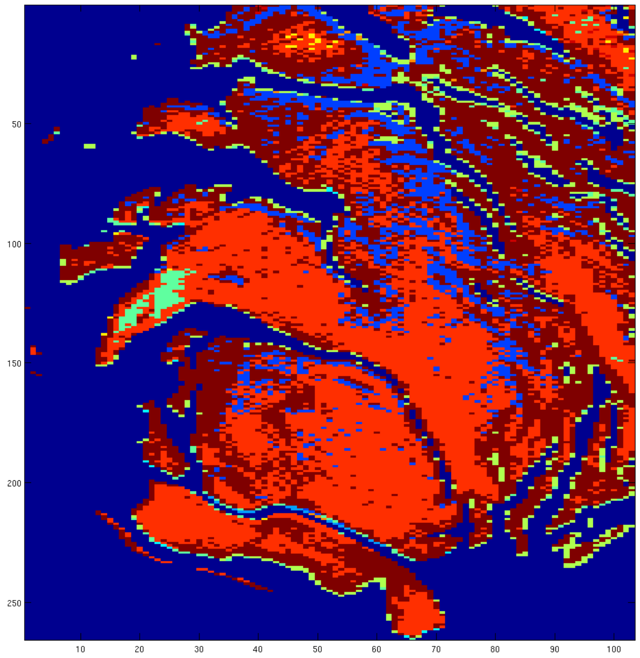
Summary:

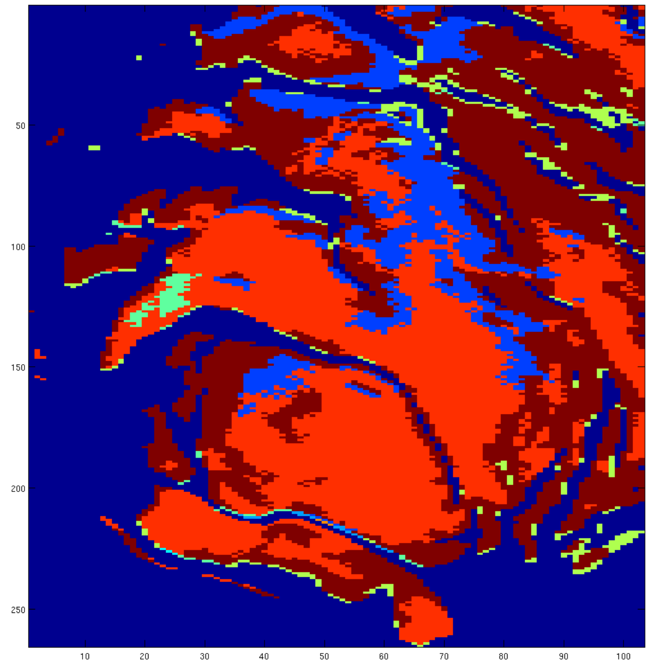
- A general framework for high-to-low dimensional locally-linear regression
- Extension to partially latent output
- A hybrid method between regression and dimensionality reduction
- Promising results for the hyperspectral analysis of Mars

Future work:

- Control the smoothness of the maps using Markov Random Fields
- Other applications of GLLiM:
  - Face Pose Estimation
  - Sound Source Localization
  - **Matlab code available online**: team.inria.fr/perception/gllim_toolbox/

THANK YOU !


50
100
150
200
250
10 20 30 40 50 60 70 80 90 100


50
100
150
200
250
10 20 30 40 50 60 70 80 90 100