

La méthode d'inversion SIR: estimation des paramètres physiques de la calotte polaire sud de Mars à partir d'images hyperspectrales OMEGA

Caroline Bernard-Michel, Sylvain Douté, Laurent Gardes and Stéphane Girard



Plan

I. Introduction

- Images hyperspectrales
- Problème inverse
- Approche LUT
- Approche fonctionnelle: SIR

II. Régression inverse par tranches (SIR)

- Principe
- Régularisation
- Estimation de la relation fonctionnelle

III. Validation

- Critères de validation
- Résultats
- Choix de la base d'apprentissage

IV. Application aux données réelles

V. Conclusions - perspectives

Plan

I. Introduction

- Images hyperspectrales
- Problème inverse
- Approche LUT
- Approche fonctionnelle: SIR

II. Régression inverse par tranches (SIR)

- Principe
- Régularisation
- Estimation de la relation fonctionnelle

III. Validation

- Critères de validation
- Résultats
- Choix de la base d'apprentissage

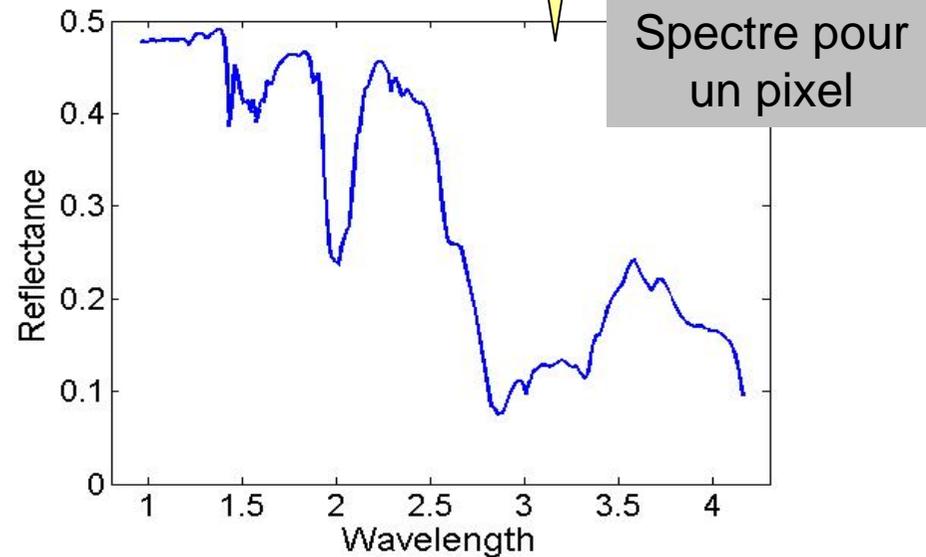
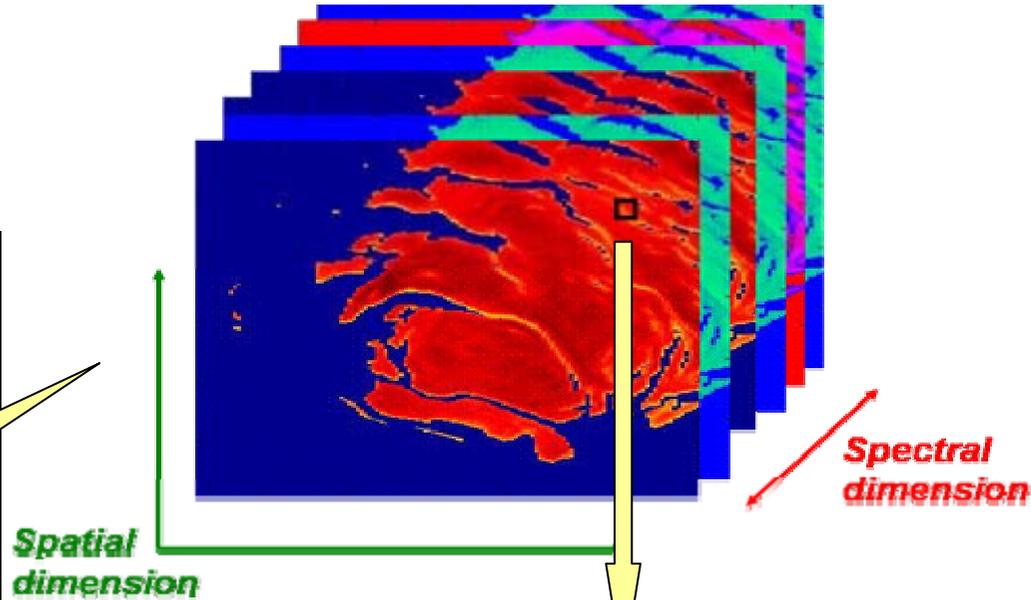
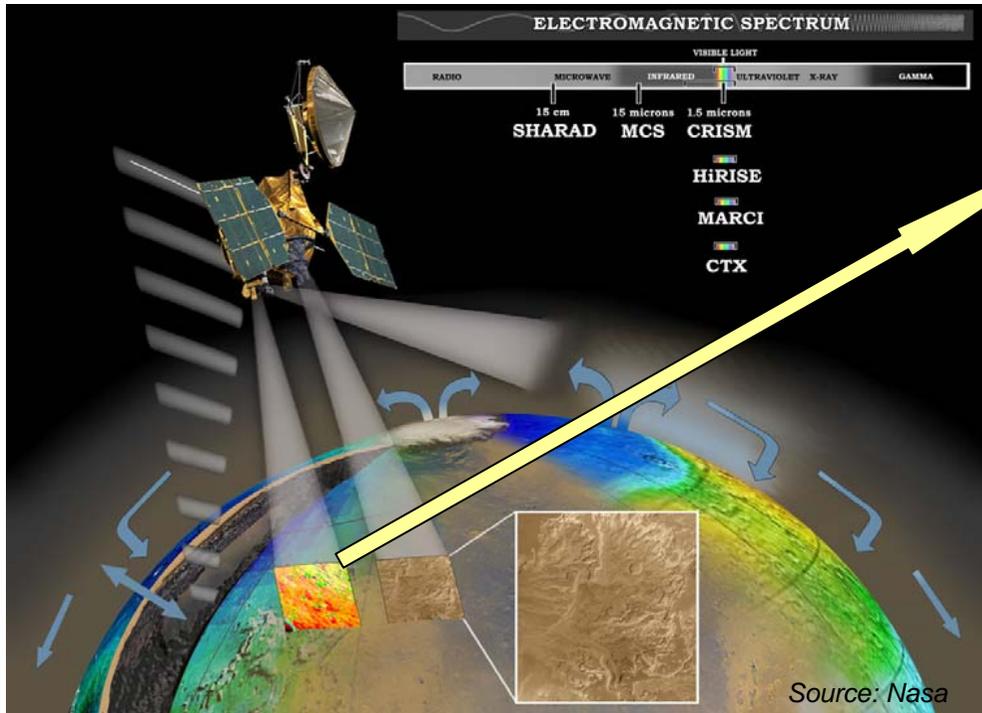
IV. Application aux données réelles

V. Conclusions - perspectives

Images hyperspectrales

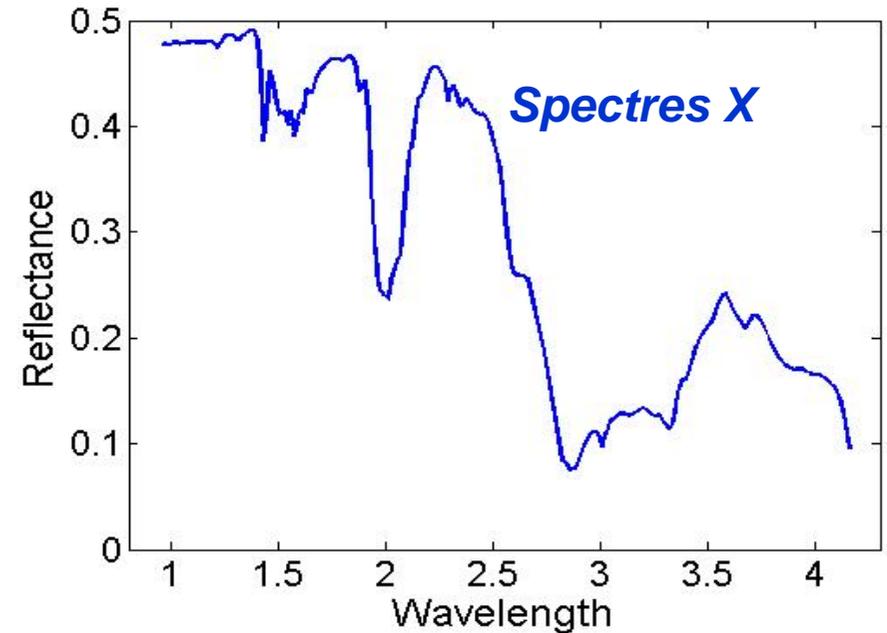
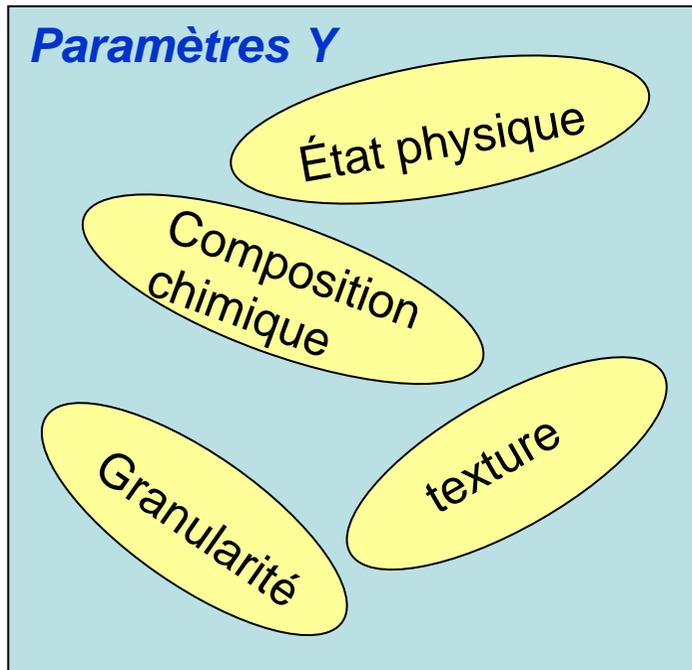
Image hyperspectrale

Spectromètre OMEGA



Problème direct

Modèle
physique
 G



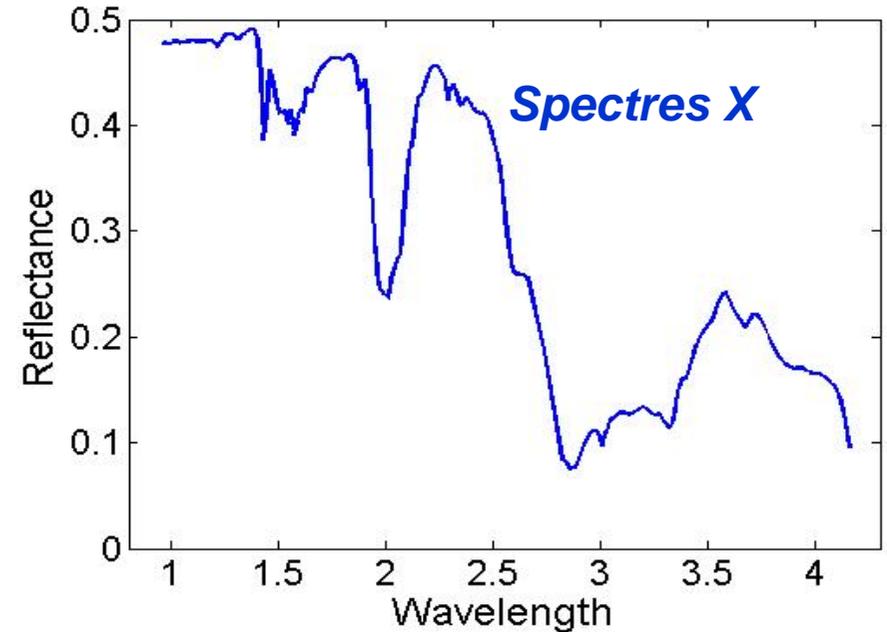
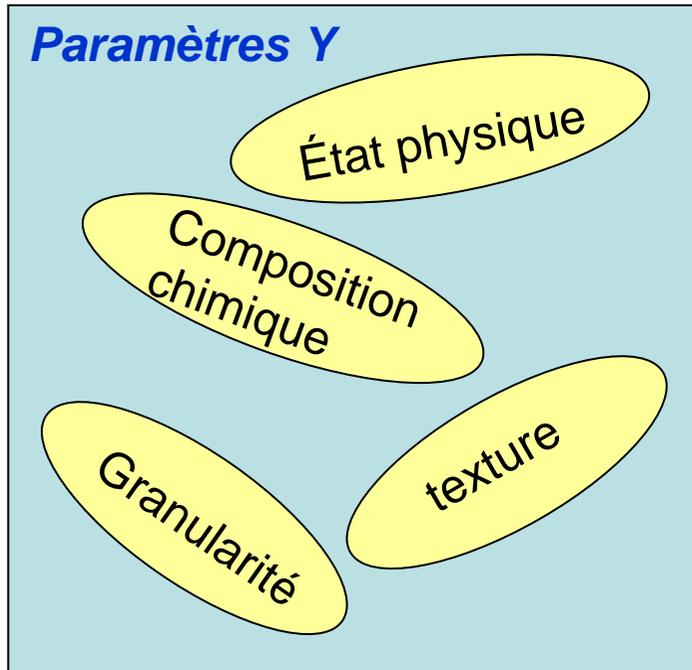
Problème direct:

déduire les spectres X à partir des paramètres Y connaissant G

Ici, par modèles de transfert radiatif.

Problème inverse

Modèle
physique
 G

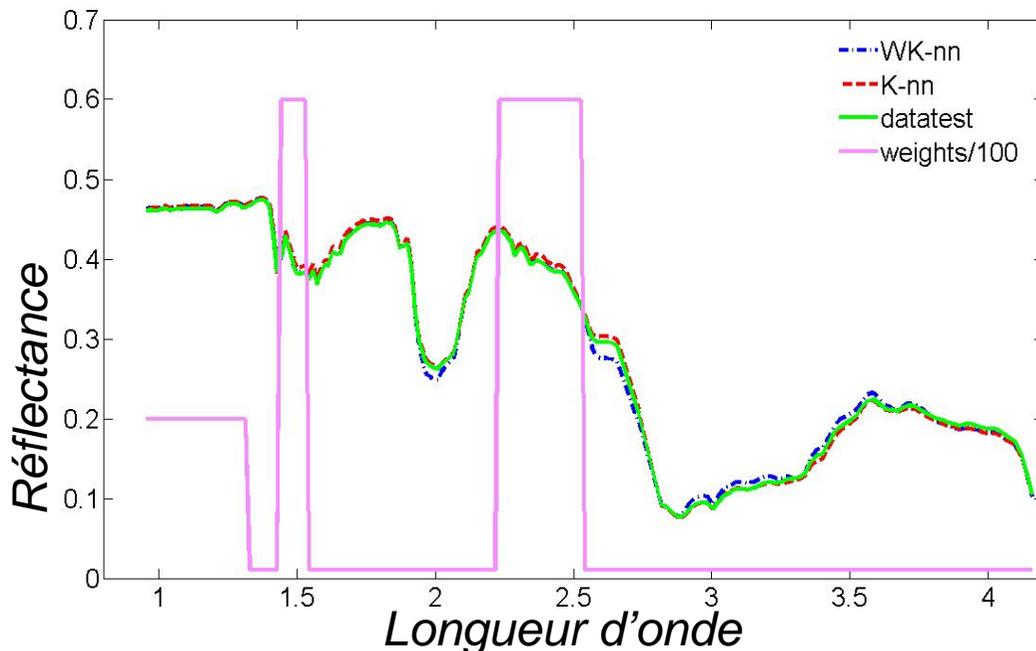


Problème inverse:

déduire les paramètres Y à partir des spectres X connaissant G

Approche actuelle

- Approche “Look-up table / K-NN”
 - Simulation d’une table de spectres pour différentes valeurs de paramètres (LUT)
 - Pour chaque spectre observé, recherche des K plus proches voisins (K-nn) dans la table.
 - Possibilité de pondérer les données (WK-nn).



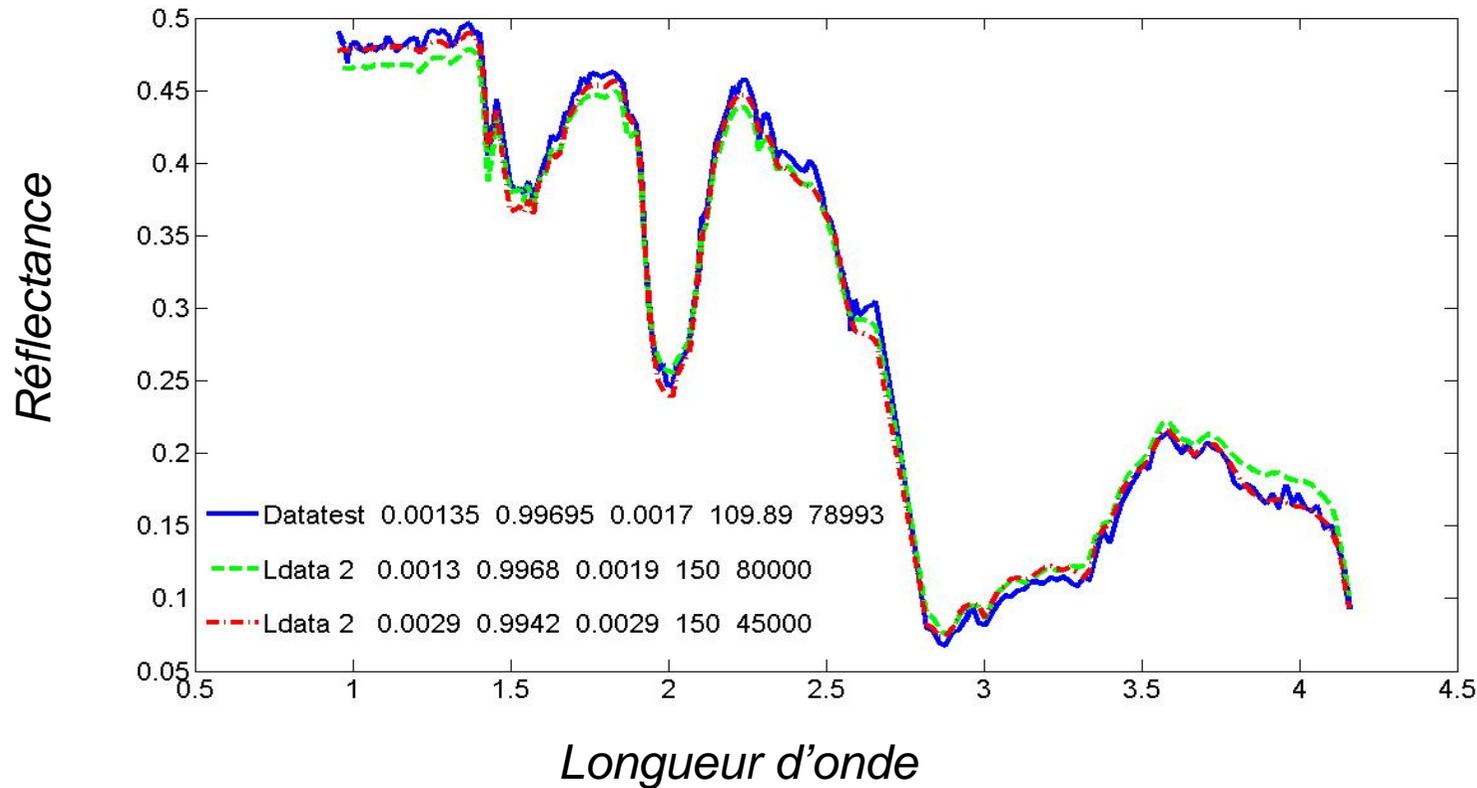
Paramètres	Valeur	K-nn	WK-nn
Proportion d'eau	0.0011	0.0008	0.0014
Proportion de CO2	0.9971	0.9978	0.9966
Proportion de poussière	0.0018	0.0014	0.002
Taille des grains d'eau	157.5	100	200
Taille des grains de CO2	91151	105000	65000

Premières constatations:

- Erreurs parfois importantes
- Choix des poids difficile

Problèmes de l'approche K-NN (1)

Deux spectres peuvent être proches avec des paramètres très différents!



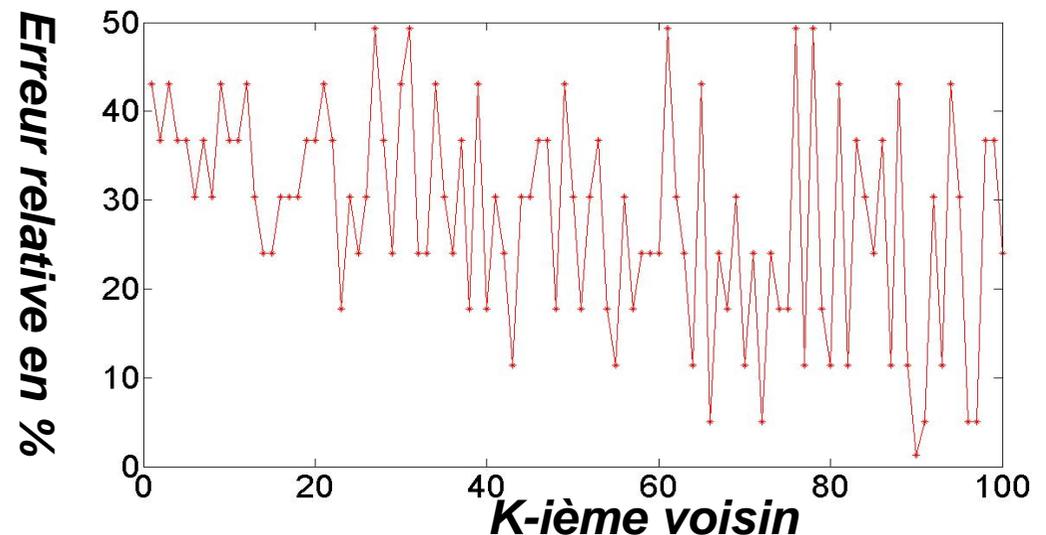
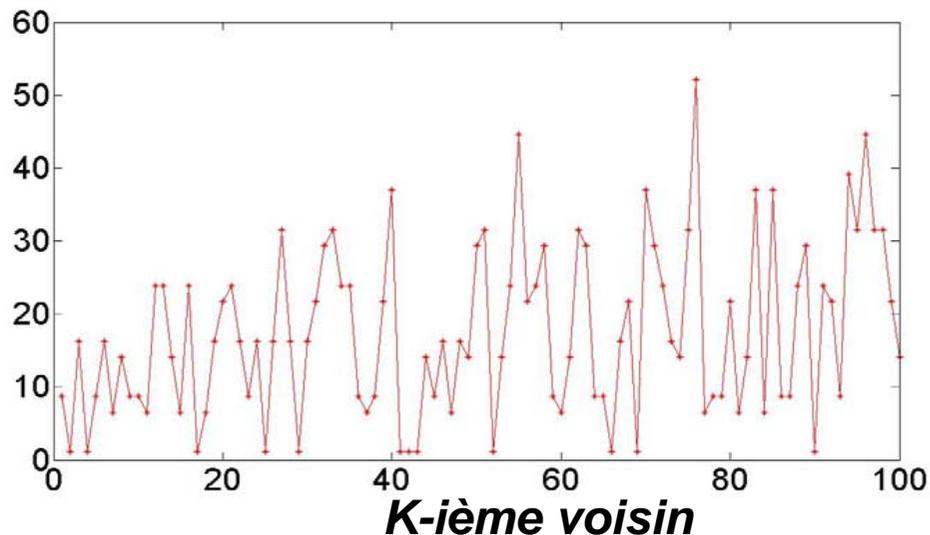
Problèmes de l'approche K-NN (2)

- Instabilité des estimations
- Choix du nombre de voisins ?

	Spectre 1		Spectre 2	
	Valeur	Estimation	Valeur	Estimation
Proportion d'eau	0.0019	0.0021	0.0013	0.0029
Proportion de CO2	0.9969	0.9966	0.9969	0.9942
Proportion de poussière	0.0012	0.0013	0.0017	0.0029
Taille des grains d'eau	156	150	109	150
Taille des grains de CO2	65721	60000	78993	45000

Spectre 1

Spectre 2



Approche proposée

- Base d'apprentissage
- Etablir une relation fonctionnelle entre:
 - Les spectres $x \in \mathbb{R}^p$ ($p=184$)
 - Les paramètres physiques $y \in \mathbb{R}$: proportion d'eau, de poussière, taille des grains de CO₂...
 - Construire f dans le but d'estimer les paramètres pour n'importe quel nouveau spectre:
 $f : \mathbb{R}^p \rightarrow \mathbb{R}$
 $x \rightarrow y$

Difficulté

- Fléau de la dimension (184 longueurs d'onde): la dimension de x doit être réduite
- Trouver un axe de projection $a \in \mathbb{R}^p$ (ici, seul le premier axe est nécessaire)
- Au lieu d'estimer f telle que $y = f(x)$, on supposera qu'il existe $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que:

$$y = g(\langle a, x \rangle)$$



REGRESSION INVERSE PAR TRANCHES (SIR)

Plan

I. Introduction

- *Images hyperspectrales*
- *Problème inverse*
- *Approche LUT*
- *Approche fonctionnelle: SIR*

II. Régression inverse par tranches (SIR)

- *Principe*
- *Régularisation*
- *Estimation de la relation fonctionnelle*

III. Validation

- *Critères de validation*
- *Résultats*
- *Choix de la base d'apprentissage*

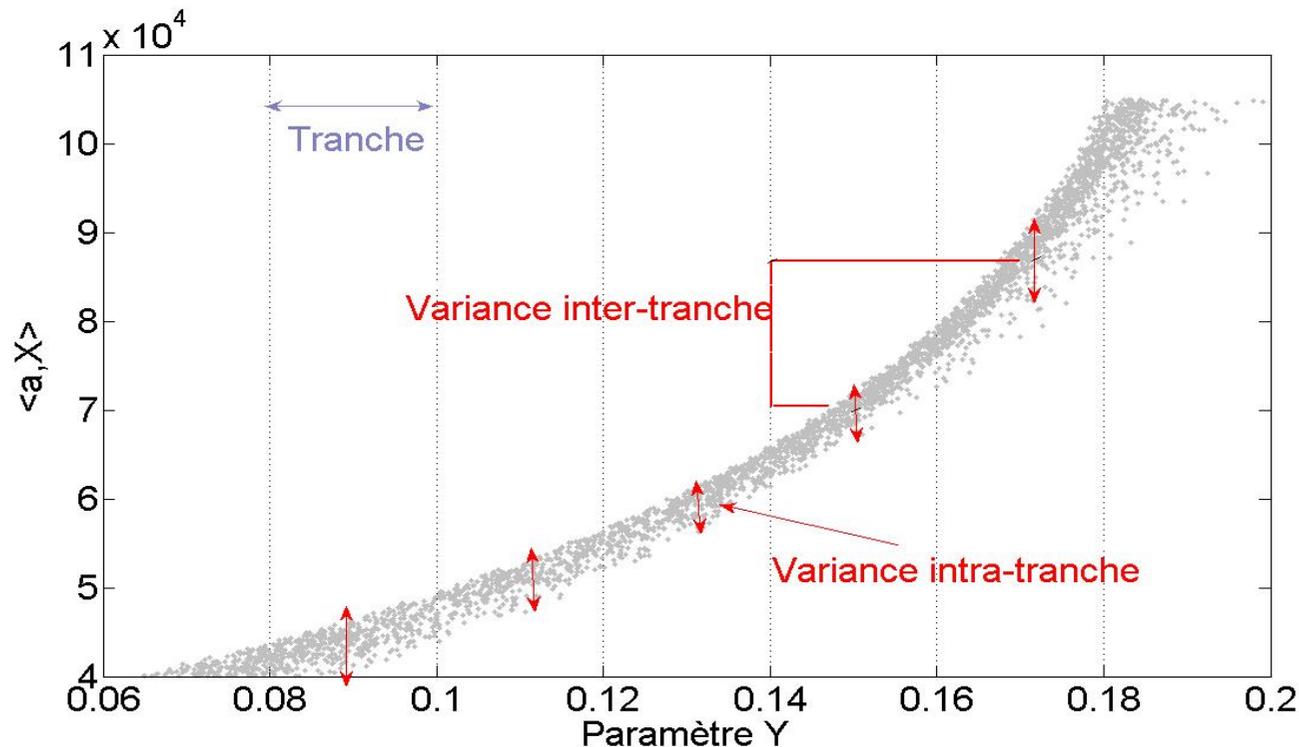
IV. Application aux données réelles

V. Conclusions - perspectives

Régression inverse par tranches

- Proposée par Li (1991)
- Maximise la variance inter-tranche
- ACP de $E(Z/T(Y))$ avec $Z = \Sigma^{-1/2} X$, T tranchage.
- Valeurs propres de $\Sigma^{-1} \Gamma$ avec $\Sigma = \text{var}(X)$

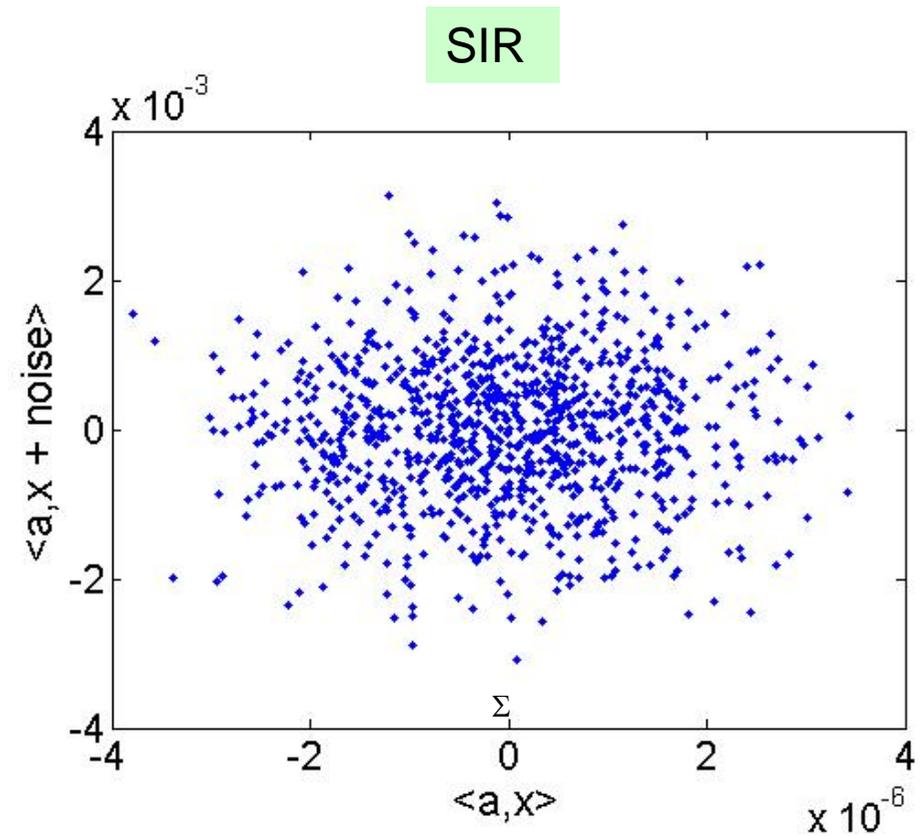
$$\Gamma = \text{var}(E(X / T(Y)))$$



Régularisation (1)

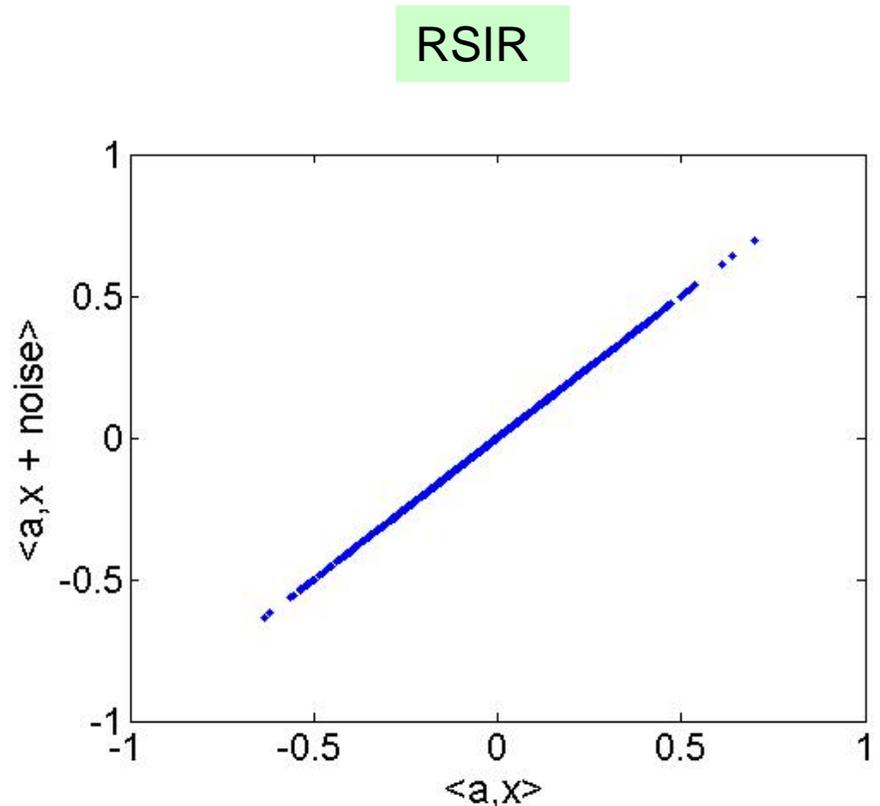
SIR usuel

- valeurs propres de $\Sigma^{-1}\Gamma$
- **Problème mal posé! : Σ est mal conditionnée**



Régularisation (2)

- SIR régularisée (RSIR)
 - s'appuie sur la régularisation de Tikhonov
 - Calcul des valeurs propres de $(\Sigma^2 + \delta Id)^{-1} \Sigma \Gamma$

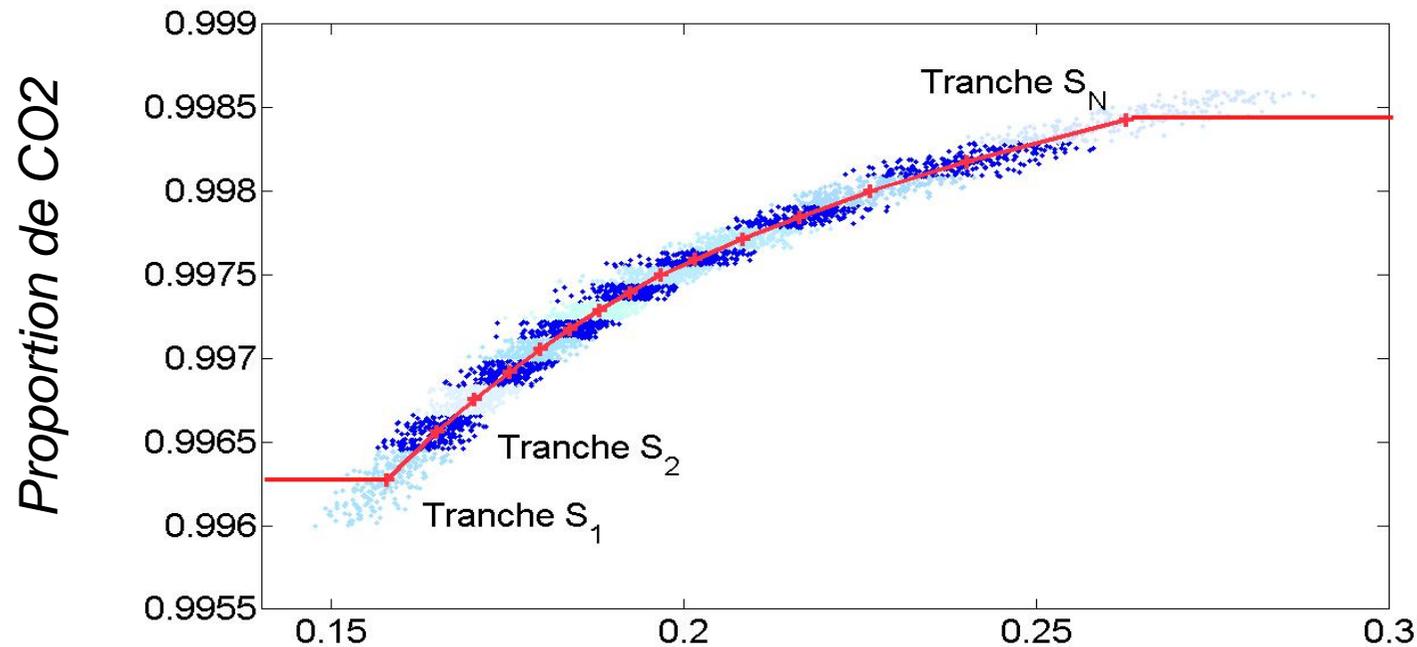


Référence: C. Bernard-Michel, L. Gardes & S. Girard (2007). Gaussian Regularized Sliced Inverse Regression. Soumis à *Statistics and computing*.

Estimation de la relation entre projections et paramètres

Interpolation linéaire entre les points ayant pour:

- Abscisses: les projections moyennes par tranche
- Ordonnées: valeurs moyennes des paramètres par tranche



Projection des spectres sur le premier axe SIR

Plan

I. Introduction

- *Images hyperspectrales*
- *Problème inverse*
- *Approche LUT*
- *Approche fonctionnelle: SIR*

II. Régression inverse par tranches (SIR)

- *Principe*
- *Régularisation*
- *Estimation de la relation fonctionnelle*

III. Validation

- *Critères de validation*
- *Résultats*
- *Choix de la base d'apprentissage*

IV. Application aux données réelles

V. Conclusions - perspectives

Validation

- Simulation d'une base test : paramètres choisis aléatoirement, ajout d'un bruit gaussien sur les spectres.
- Comparaison des estimations avec les valeurs réelles
- Deux bases d'apprentissage (une adaptée et une trop grande)
- Deux critères de validations:
 - **NRMSE**: doit être proche de **0**

$$\frac{\|\hat{y} - y\|}{\|y - \bar{y}\|} = \frac{\text{Variance résiduelle}}{\text{Variance totale}}$$

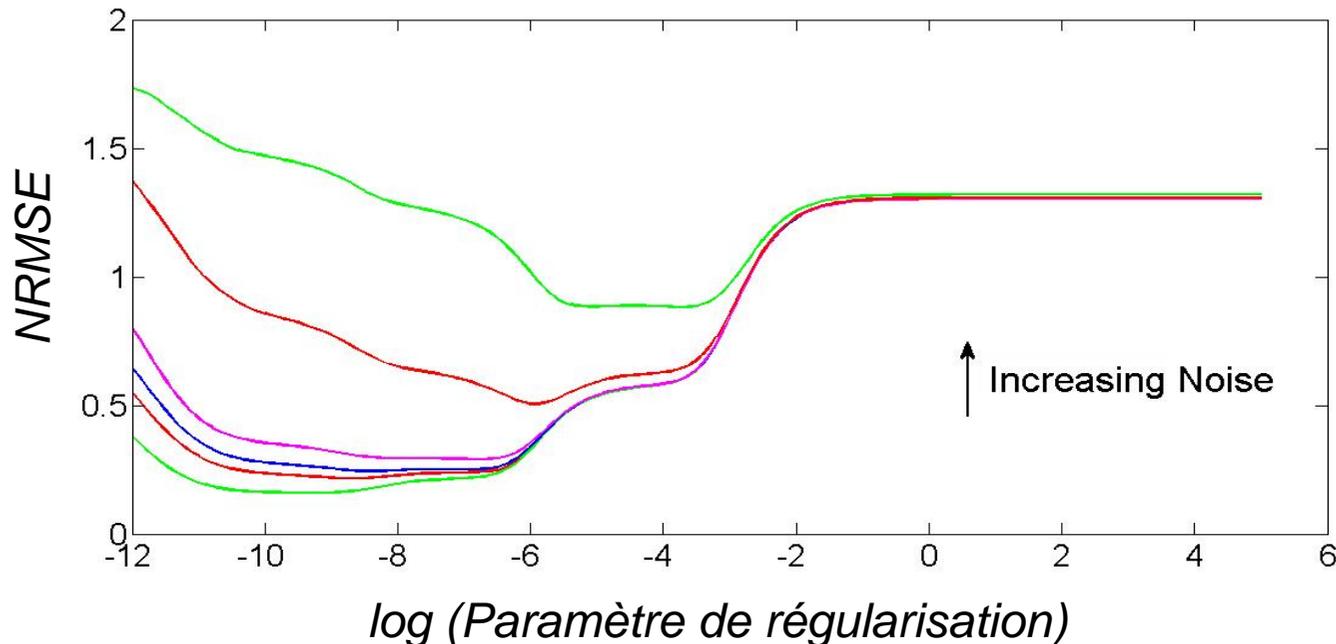
- **SIRC**: doit être proche de **1**

$$\frac{{}^t a \Gamma a}{{}^t a \Sigma a} = \frac{\text{Variance inter-tranche}}{\text{Variance totale}}$$

Choix du paramètre de régularisation

- Augmenter le paramètre de régularisation dégrade la relation fonctionnelle
- MAIS, dans le cas de données bruitées, un faible paramètre de régularisation conduit à de mauvaises estimations

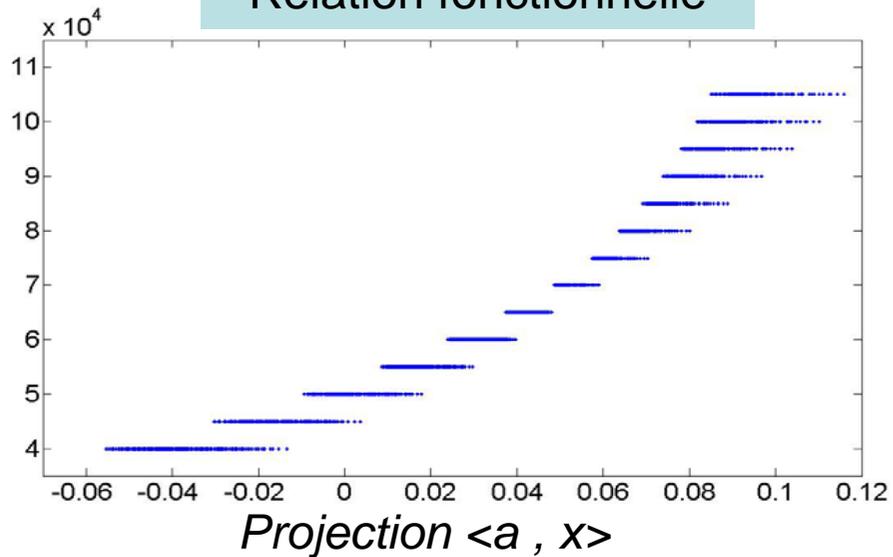
CHOIX FINAL: Minimisation du NRMSE dépend du bruit.



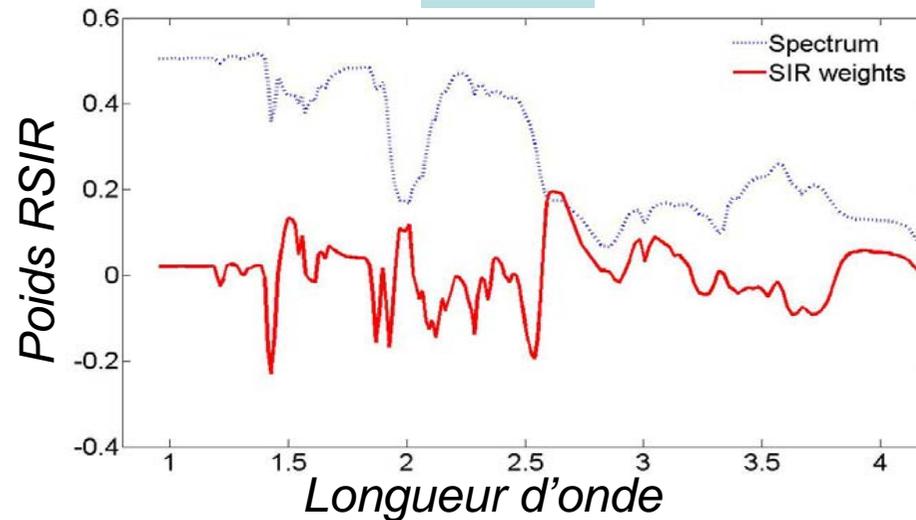
Résultats (1)

Taille des grains de CO2

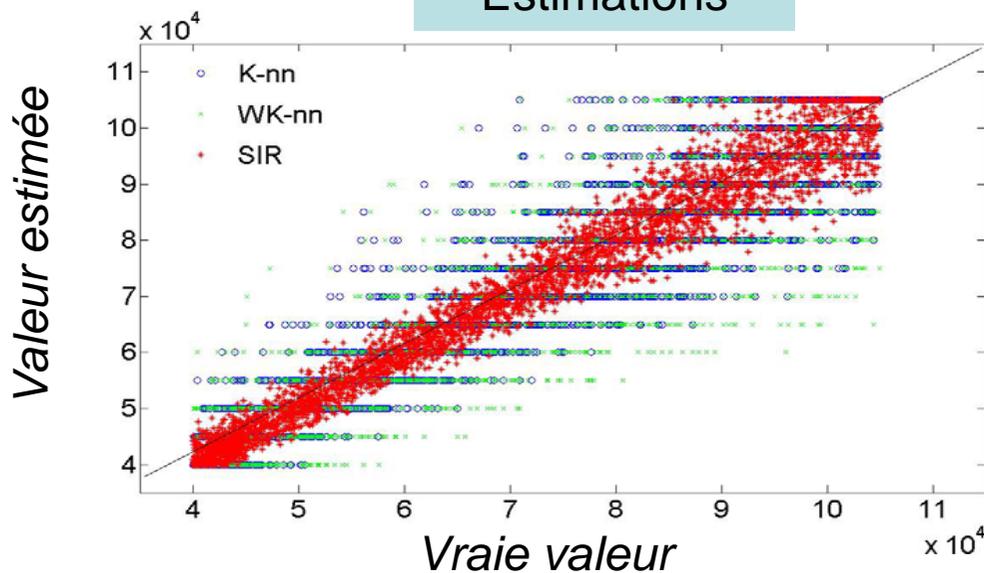
Relation fonctionnelle



Poids



Estimations



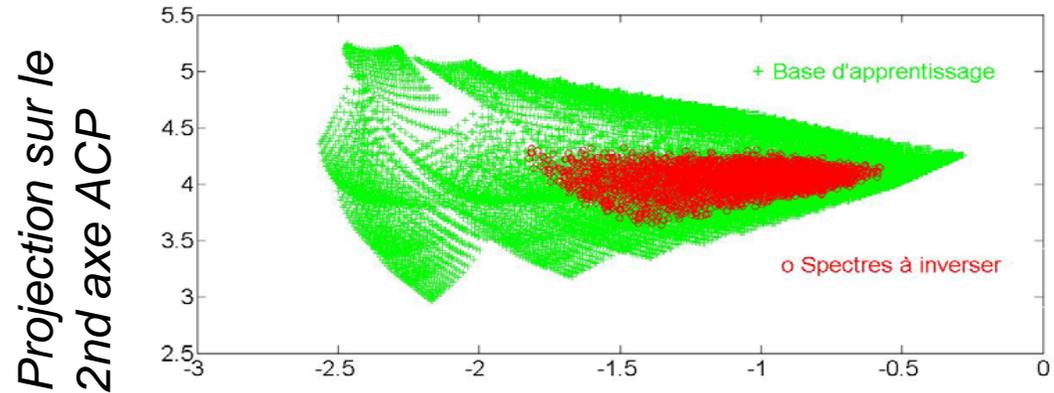
Résultats (2)

- Estimations toujours meilleures par SIR
- WK-nn améliore estimations des proportions, dégrade estimations des tailles de grain
- Paramètre le mieux estimé: proportion de poussière
- Dégradation des résultats en utilisant une base trop grande

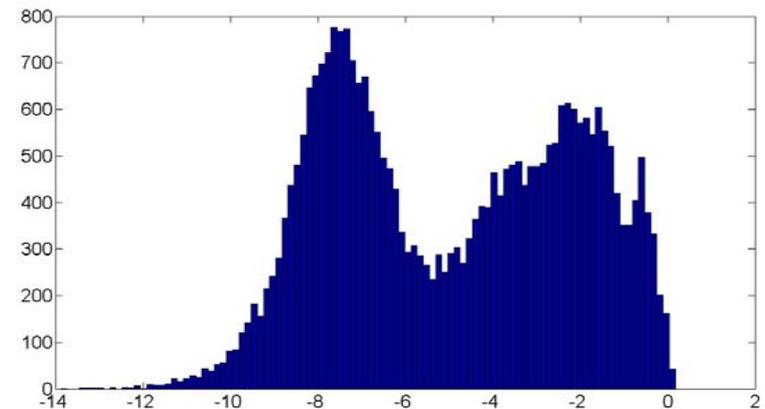
	Base adaptée				Grande base			
	RSIR		K-nn	WK-nn	RSIR		K-nn	WK-nn
	NRMSE	SIRC	NRMSE	NRMSE	NRMSE	SIRC	NRMSE	NRMSE
Proportion d'eau	0.29	0.92	0.5	0.38	0.63	0.88	0.86	0.6
Proportion de CO2	0.22	0.99	0.54	0.46	0.4	0.97	0.88	0.68
Proportion de poussière	0.13	0.99	0.34	0.35	0.31	0.99	0.44	0.41
Taille des grains d'eau	0.37	0.92	0.39	0.45	0.41	0.8	0.43	0.48
Taille des grains de CO2	0.19	0.98	0.35	0.46	0.27	0.93	0.53	0.67

Réflexion sur le choix de la base

- K-nn: si base trop grande, non unicité des solutions
- SIR: si base trop grande, ajout d'information inutile
- Sélection de la base pour SIR par ACP et classification des distances « base-observations »



Projections des spectres sur le 1er axe ACP de la base d'apprentissage



Distances des points de la base d'apprentissage à leur plus proche voisin dans les spectres observés

Plan

I. Introduction

- *Images hyperspectrales*
- *Problème inverse*
- *Approche LUT*
- *Approche fonctionnelle: SIR*

II. Régression inverse par tranches (SIR)

- *Principe*
- *Régularisation*
- *Estimation de la relation fonctionnelle*

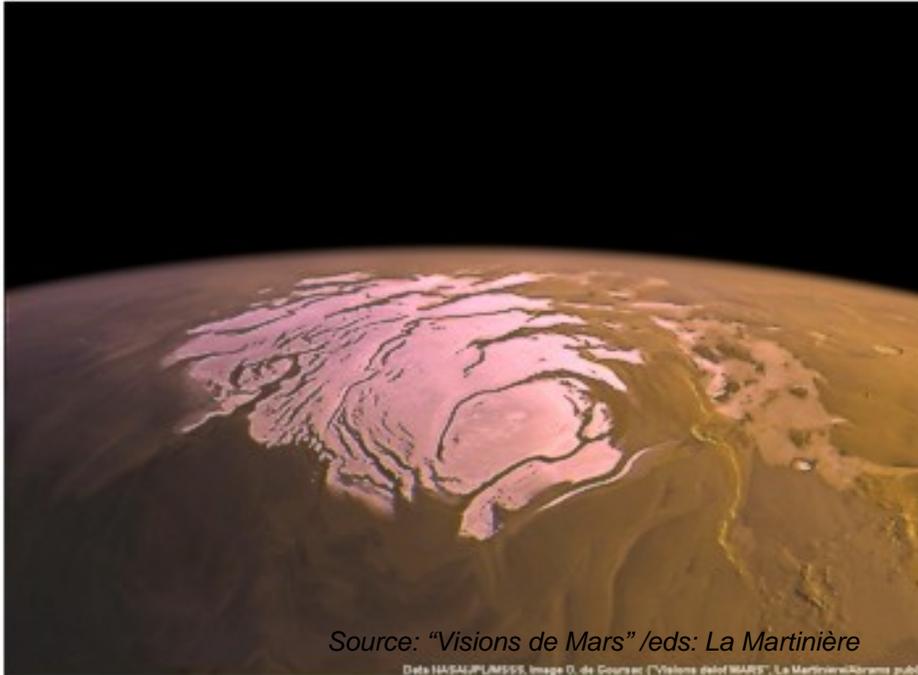
III. Validation

- *Critères de validation*
- *Résultats*
- *Choix de la base d'apprentissage*

IV. Application aux données réelles

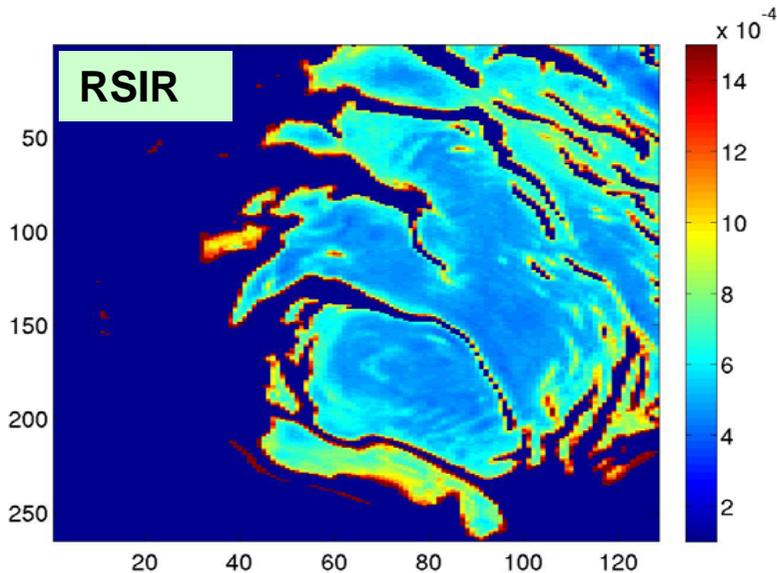
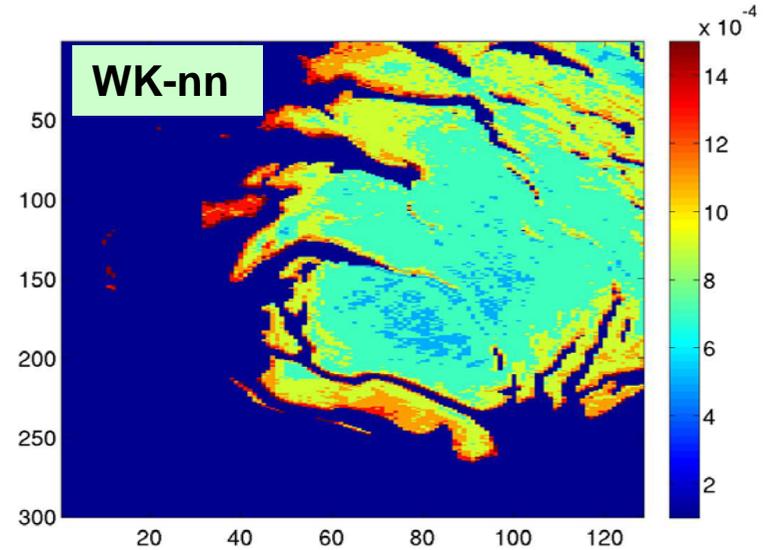
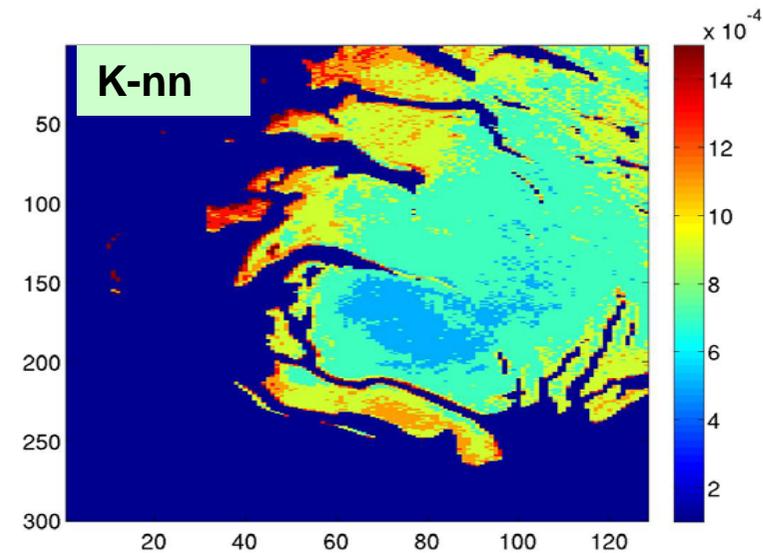
V. Conclusions - perspectives

Application à la calotte polaire sud de Mars



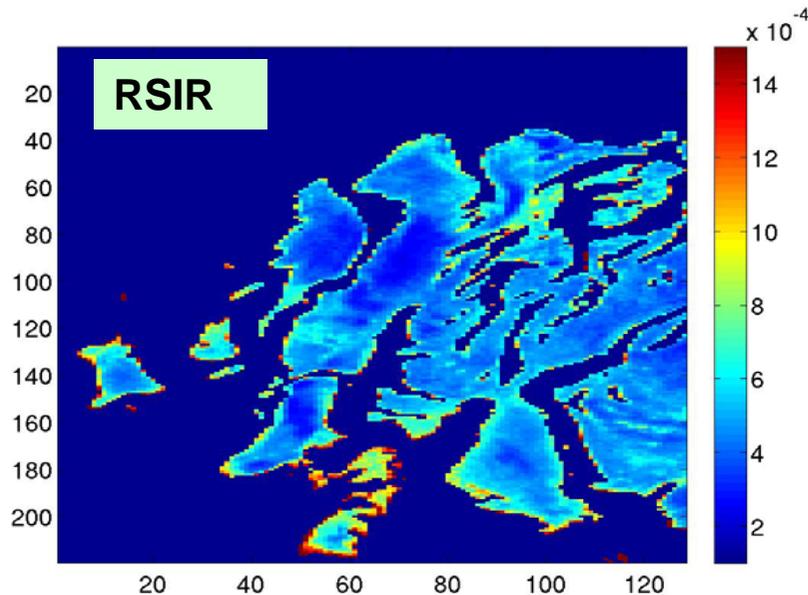
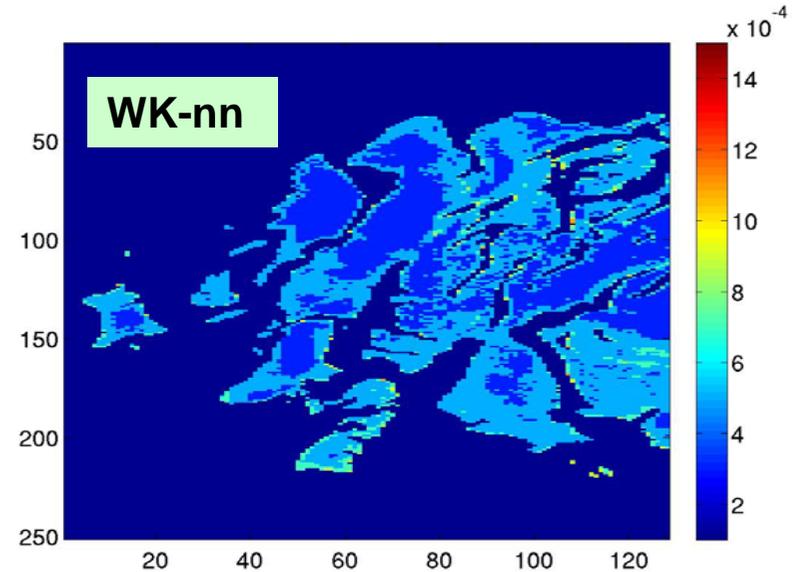
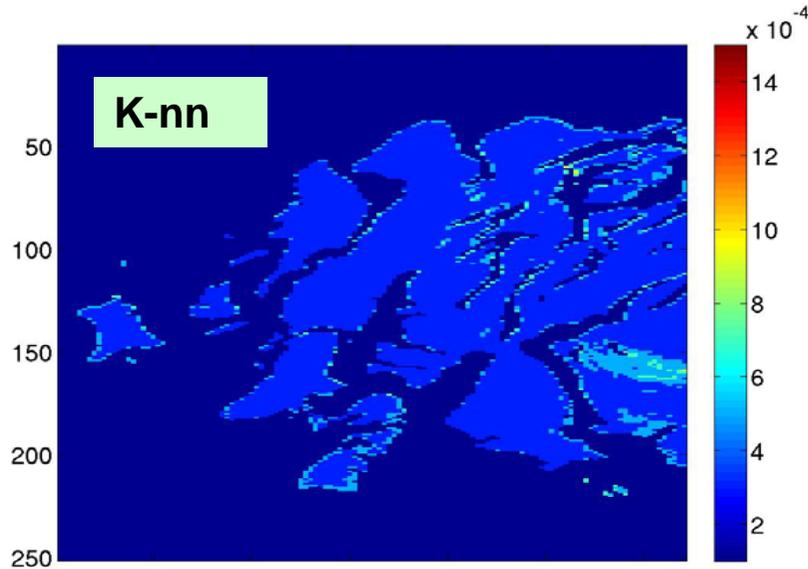
- Modèle déterminé par les physiciens: poussière + eau + CO₂
- Orbites 30, 41, 61 et 103
- ~ 10000-20000 spectres à inverser par image
- 184 longueurs d'onde
- Base d'apprentissage par modèle de transfert radiatif
- 5 paramètres: proportions de poussière, CO₂ et eau, taille des grains de CO₂ et d'eau.

Orbite 41: proportion de poussière



- Estimations par RSIR spatialement cohérentes
- RSIR fournit des cartes plus lisses que par K-nn et WK-nn
- Les cartes d'estimations ne semblent pas en contradiction avec celles données par K-nn et WK-nn

Orbite 103: proportion de poussière

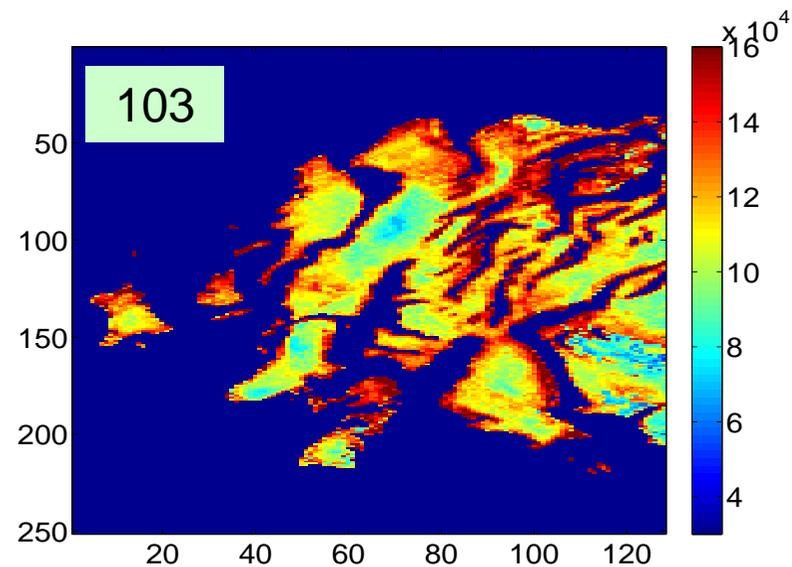
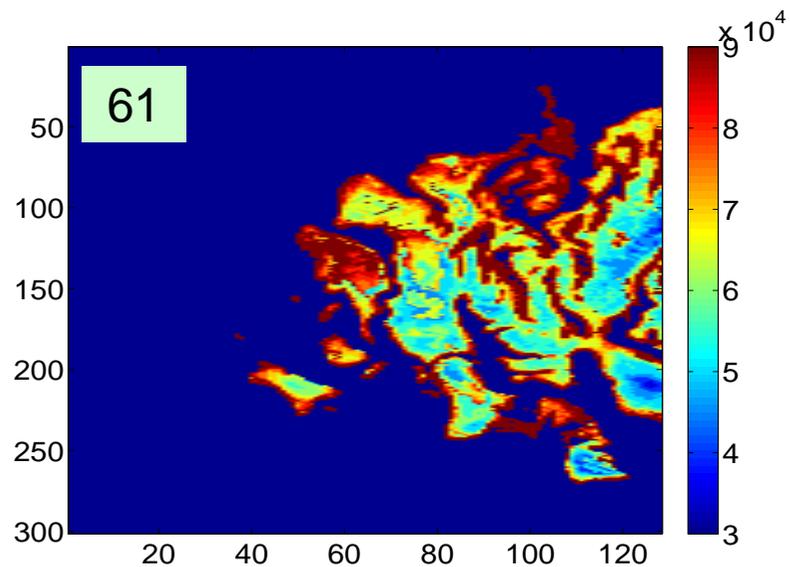


Par K-nn, même estimation pour presque tous les pixels!

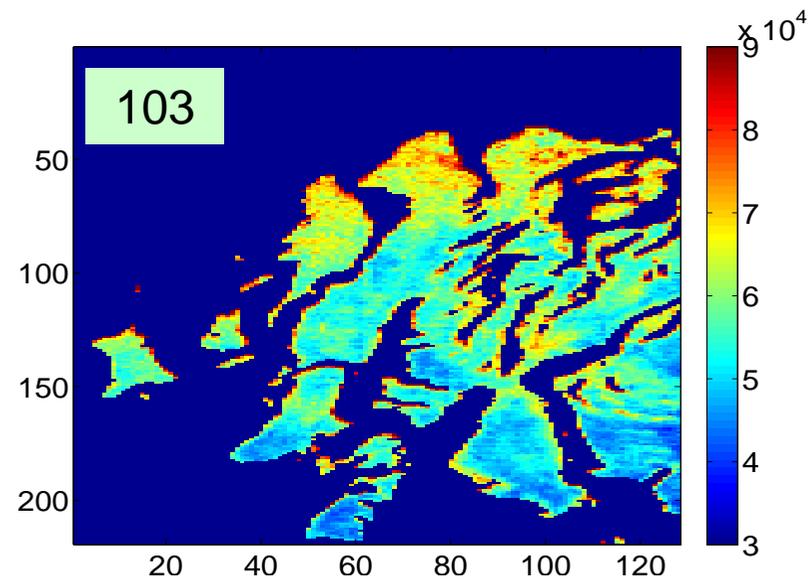
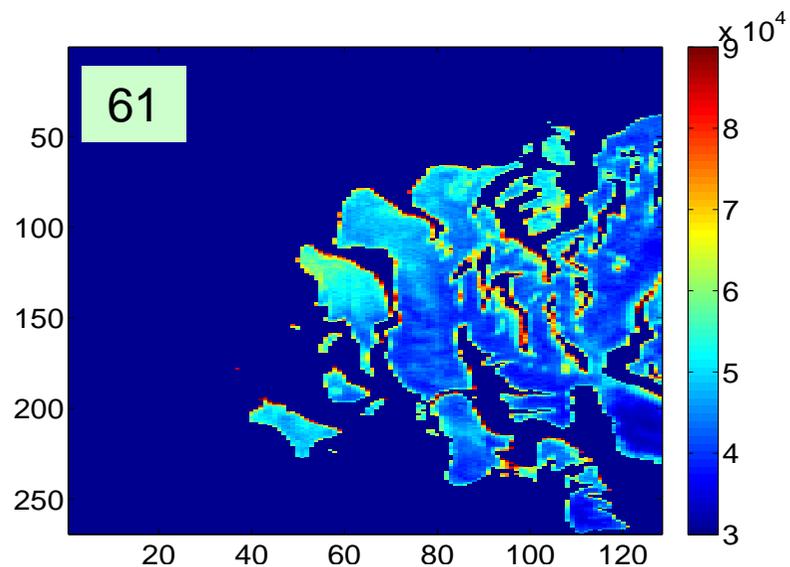
Taille des grains de CO2

Orbites 103 et 61

K-nn



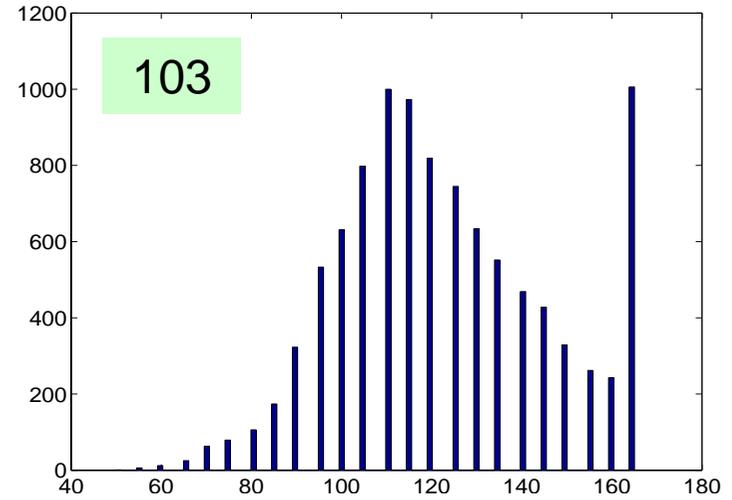
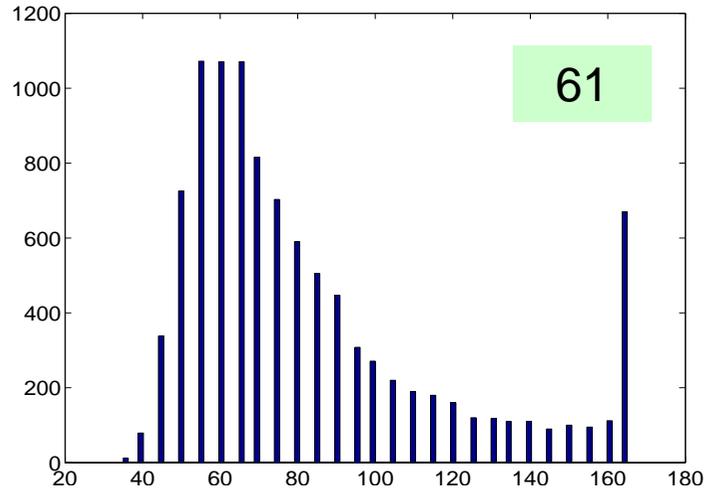
RSIR



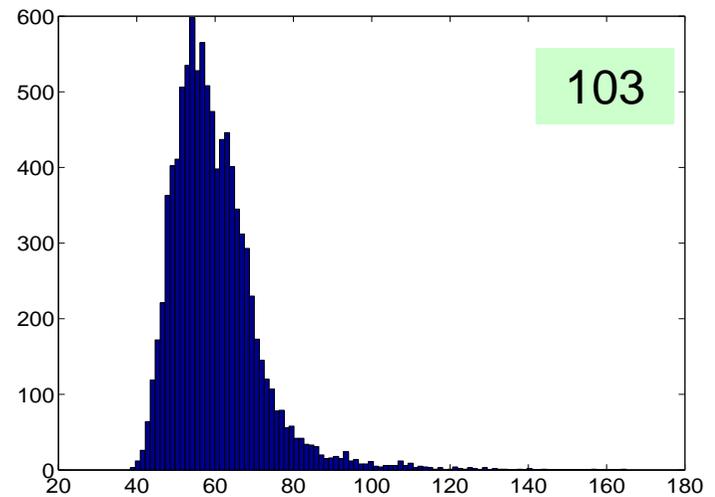
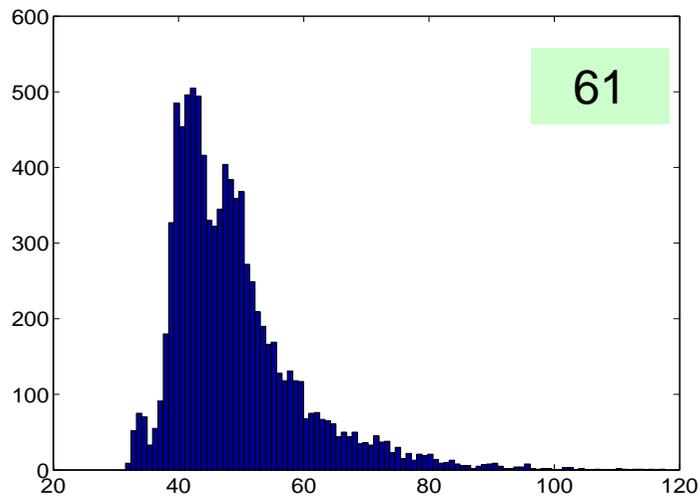
Taille des grains de CO₂

Orbites 61 et 103

K-nn



RSIR



Plan

I. Introduction

- *Images hyperspectrales*
- *Problème inverse*
- *Approche LUT*
- *Approche fonctionnelle: SIR*

II. Régression inverse par tranches (SIR)

- *Principe*
- *Régularisation*
- *Estimation de la relation fonctionnelle*

III. Validation

- *Critères de validation*
- *Résultats*
- *Choix de la base d'apprentissage*

IV. Application aux données réelles

V. Conclusions - perspectives

Conclusions

- SIR: fournit une pondération des longueurs d'onde pour chaque paramètre
- Bons résultats sur simulations
- Résultats cohérents sur données réelles
- La validation est difficile car aucune mesure terrain n'est disponible
- Critères de validation donnent une idée de l'incertitude

Perspectives

- Calcul d'incertitude
- Choix du paramètre de régularisation
- Estimation et influence du bruit
- SIR multivarié avec contrainte
- Validation avec des données terrain