



**MicroCarb**

**L'estimation optimale pour l'inversion de spectres atmosphériques**  
**COMET-TSI – 26/06/2024**

**Denis Jouglet, Pierre Rieu**

**Contributions de Vincent Quédeville, Pierre Lafrique, Jean-Luc Vergely, Aurélie Bornot**

**CNES DTN/TPI/SA**

# The MicroCarb CO<sub>2</sub> mission at a glance

## Objectives

- Natural CO<sub>2</sub> fluxes at global scale
- Demonstrator for CO<sub>2</sub> anthropogenic emissions

## Main product: CO<sub>2</sub> column integrated concentration

- Rqmt for random error < 1 ppm (0.25%)
- Rqmt for regional bias < 0.1 ppm (0.025%)

## Orbit

- Polar sun-synchronous, alt 649 km, LTAN 22h30
- Cycle 25 days, sub-cycle 7 days,  $\pm 200$ km ACT mirror
- ➔ Any target can be observed once a week

## Observations modes

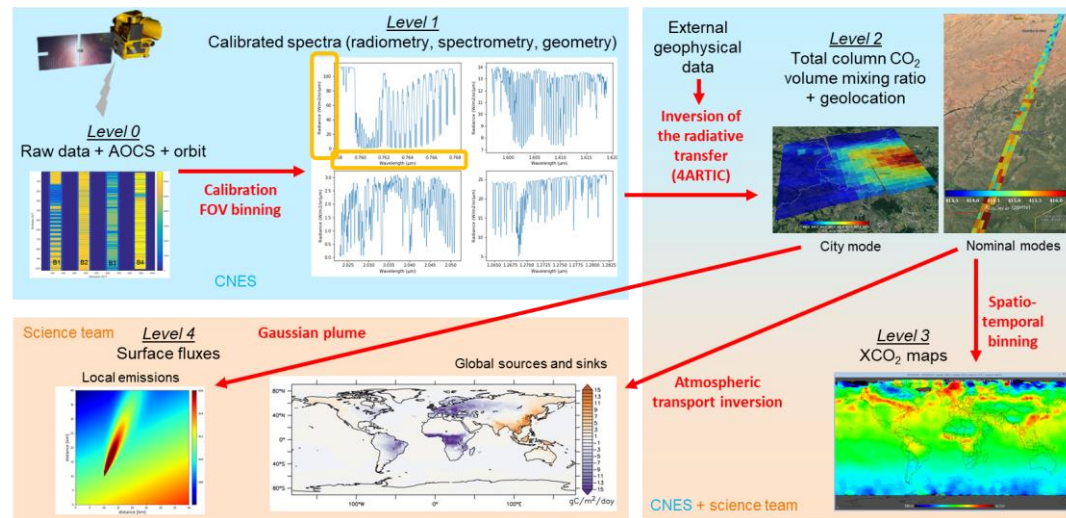
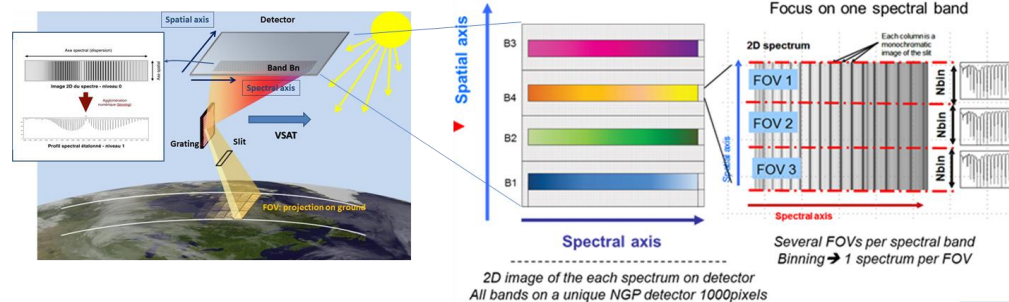
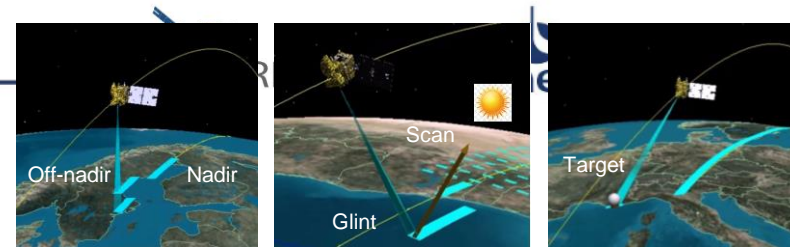
- Science: Nadir, scan, glint, city, region
- Calibration: Shutter, Sun, Lamp, Cold Space, Limb
- Validation: Fixed-Target, Off-nadir target

## Instrument

- Compact instrument (80 kg, 60W) on micro-satellite
- Passive grating spectrometer in 4 NIR / SWIR bands
  - CO<sub>2</sub> 1.61  $\mu$ m & 2.05  $\mu$ m, O<sub>2</sub> 0.76  $\mu$ m & 1.27  $\mu$ m
  - High resolving power 25000, high SNR
  - 3 ACT footprints 4.5x9 km<sup>2</sup>
- Embedded imager (red band resolution 140m)

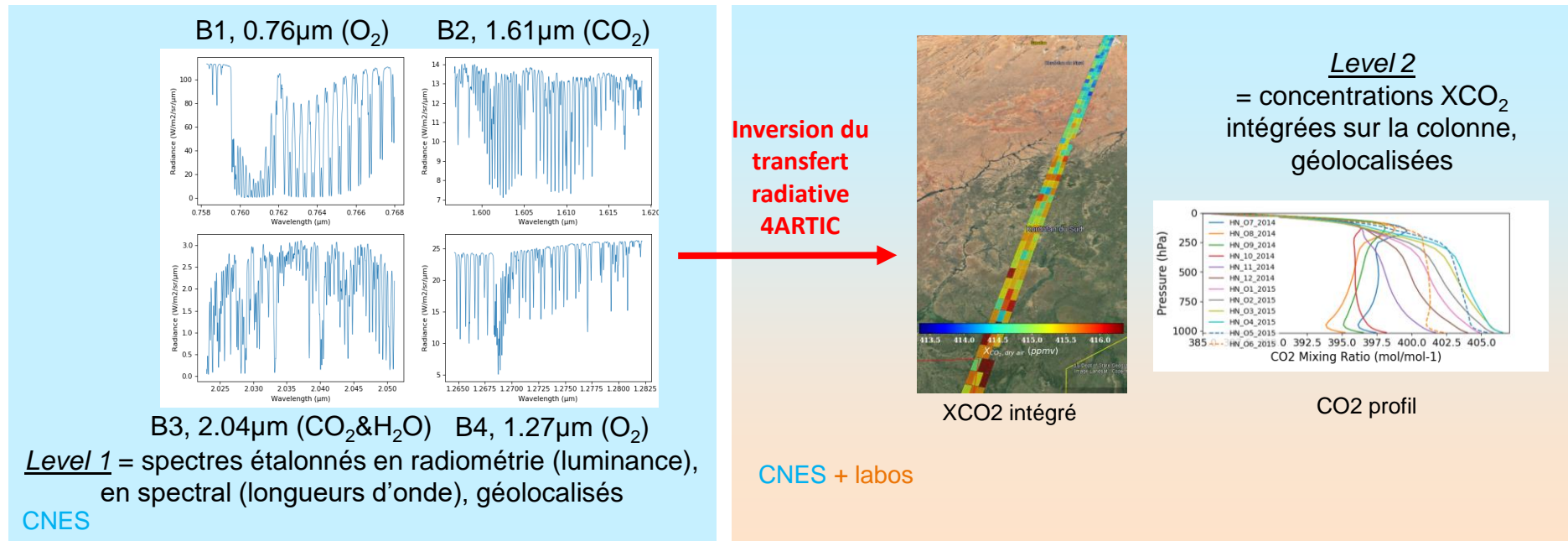
## XCO<sub>2</sub> retrieval : 4ARTIC

- Full physics, optimal estimation, GEISA, LIDORT



# Le problème à résoudre

- Nous souhaitons calculer la concentration de CO<sub>2</sub> (intégrée verticalement) à partir des spectres mesurés

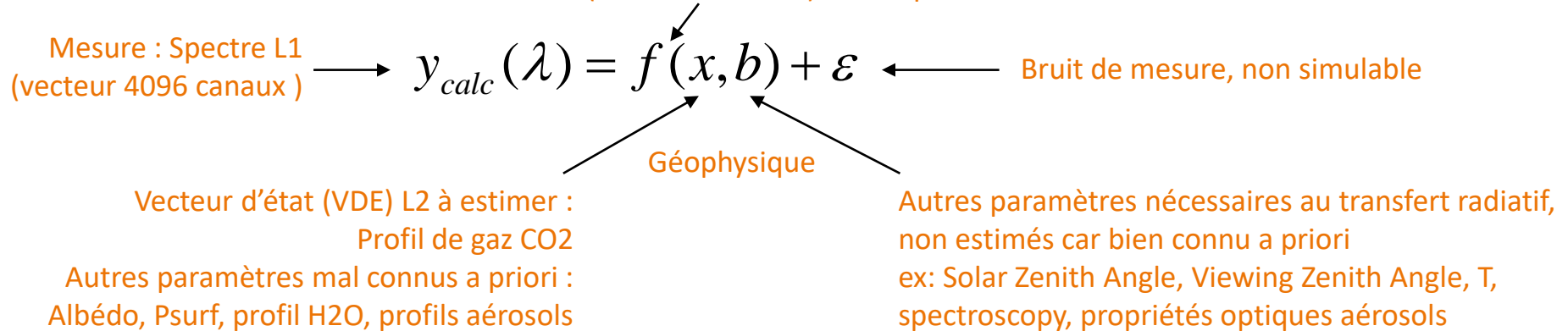


- Au 1<sup>er</sup> ordre, plus la concentration est élevée, plus les raies sont profondes. Mais il y a beaucoup d'autres paramètres influençant le spectre
- Pour modéliser l'effet du CO<sub>2</sub> (et autres) sur les spectres, nous disposons d'un modèle direct (4AOP) : spectre = f(XCO<sub>2</sub> et autres), mais pas directement de modèle réciproque

## Expression mathématique du problème

- Inversion des spectres de luminance  $y$  pour estimer l'état géophysique  $x$

Modèle direct (transfert radiatif) donné par 4AOP



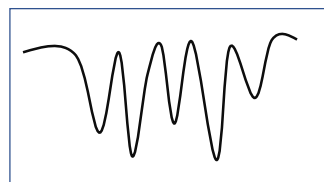
- On cherche  $x$  tel que  $y_{calc}(\lambda)(x) - y_{mesuré}(\lambda) \sim 0 (\sim \varepsilon)$   
 $\rightarrow$  Il s'agit donc d'un problème classique de recherche du 0 d'une fonction
- L'évaluation de (4AOP) est coûteuse en temps de calcul (5 min / iter, max 5 iter. / spectre, 4000 spectres/jour)
- 4AOP fournit aussi les matrices jacobiennes

$$K = \frac{\partial f}{\partial x}$$

# L'inversion avec les mains

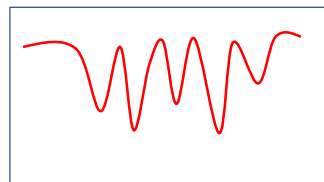
On va chercher à ajuster le spectre 4AOP au spectre mesuré

*Spectre mesuré par MicroCarb*

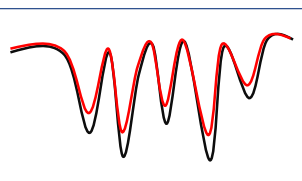


Concentration inconnue

*Spectre calculé par 4AOP*



Concentration a priori (ébauche)  
 $\text{XCO}_2[i=0]$



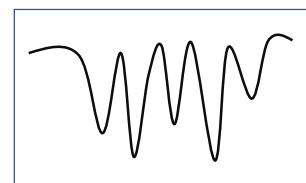
Ecart importants (résidus spectraux) : la mesure contient manifestement plus de gaz que la simulation  
→ 4ARTIC corrige la concentration grâce aux jacobiens et on relance 4AOP pour recalculer le spectre

Processus itératif



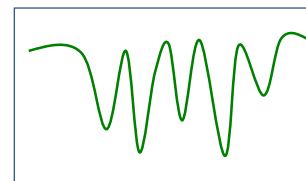
Généralement, le nouveau spectre présente des résidus meilleurs, mais pas encore satisfaisants. En effet on a corrigé la concentration grâce aux jacobiens, mais le problème étant non-linéaire (jacobiens dépendant du  $\text{XCO}_2$ ), on a trop ou pas assez corrigé le  $\text{XCO}_2$ . D'où la nécessité de relancer cette opération plusieurs fois.

*Spectre mesuré par MicroCarb*

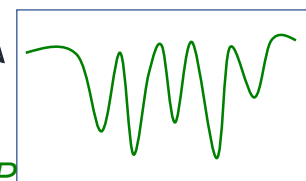


Concentration inconnue

*Spectre calculé par 4AOP*



Concentration  
 $\text{XCO}_2[i=n]$

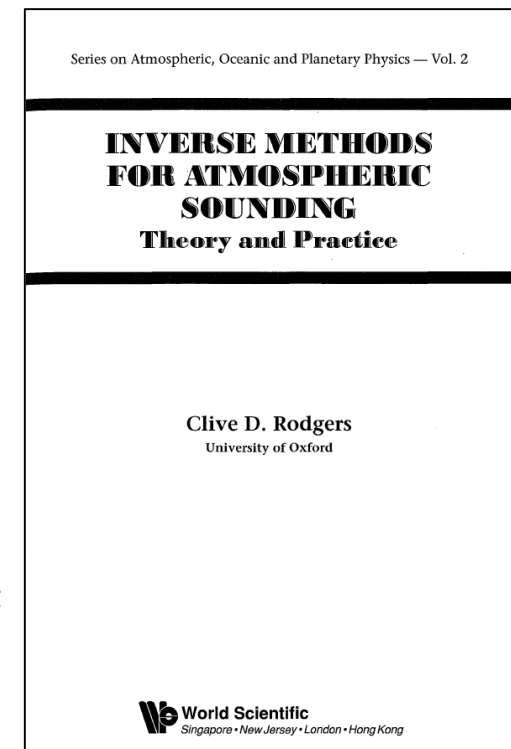


Après plusieurs itérations, les résidus deviennent plus faibles que le bruit de mesure  
→ On arrête les itérations, la concentration retenue est égale à la concentration  $[n]$

## Rodgers 2000

- **Problème mal conditionné**
  - Sous-contraint car trop d'inconnues / information dans les mesures
  - Sur-contraint à cause du bruit ( $K$  est inversible, mais donne n'importe quoi)
- **Problème non linéaire ( $K$  fortement variable avec la scène)**
- **La méthode la plus courante est Rodgers, 2000, notre bible (ancien testament)**
  - Basée sur l'estimation optimale (voir planches suivantes)
  - Régularisation par de l'a priori
- **Au préalable quelques adaptations (nouveau testament)**
  - Réduction de la dimensionnalité du problème
    - Profil de  $\text{CO}_2$  : 19 couches (au lieu de 137)
    - Idem profil de  $\text{H}_2\text{O}$
    - Aérosols : paramétrisé grossièrement par 3 variables (AOD, coeff. Angström kaero, profil vertical gaussien d'altitude Zaero)
    - Albédo idem : offset et pente par bande

→ Vecteur d'état  $x$  à ~50 éléments
  - Recherche des meilleurs a priori
    - Analyses météo ECMWF, CAMS
    - DEM, tapis mondial S2, etc.



# Bayésien et maximum de vraisemblance (Rodgers 2000)

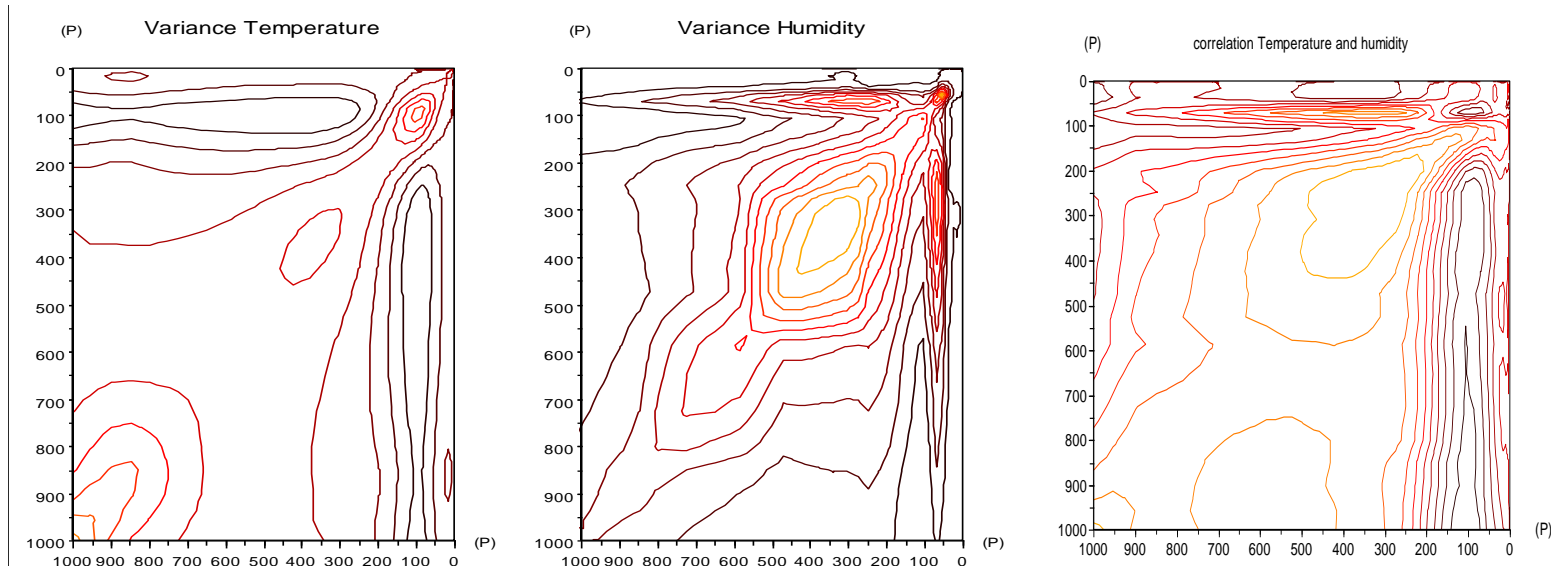
- Formalisme probabiliste bayésien :**

$$P(X / Y).P(Y) = P(Y / X).P(X) \Rightarrow P(X / Y) = \frac{P(Y / X).P(X)}{P(Y)}$$
- Fonctions de probabilités gaussiennes :**
  - Multi-dimensionnel + corrélations
$$\begin{cases} P(X) = c3.\exp\left[-((x - x_a)^T S_A^{-1}(x - x_a))^2\right] & \text{CO2 a priori} \\ P(Y / X) = c2.\exp\left[-((y - f(x))^T S_\varepsilon^{-1}(y - f(x)))^2\right] & \text{Bruit de mesure} \\ P(X / Y) = c1.\exp\left[-((x - \hat{x})^T \hat{S}^{-1}(x - \hat{x}))^2\right] & \text{CO2 estimé} \end{cases}$$
- NB : on passe en gaussien, mais**
  - Les distributions sont en fait rarement gaussiennes!
  - Les distributions ne portent pas forcément sur les mêmes dimensions (temporelle, etc.)
  - On évacue complètement la question du biais
- Maximum de vraisemblance: on cherche  $X = \hat{x}$  qui maximise  $P(X/Y)$ ,  $\frac{\partial(P(X / Y))}{\partial X} = 0$  cad qui minimise la fonction (coût) :**

$$-\ln(P(X/Y)) = \Phi(x) = (y - F(x))^T S_\varepsilon^{-1}(y - F(x)) + (x_a - x)^T S_a^{-1}(x_a - x)$$
  - Dans le cas linéaire ( $f(x) = K.x$ ), par égalisation (analytique) :
 
$$\hat{x} = x_a + (S_a^{-1} + K^T S_\varepsilon^{-1} K)^{-1} K^T S_\varepsilon^{-1} (y - K x_a) \quad \text{et} \quad \hat{S} = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1}$$
  - Dans le cas général, recherche numérique du minimum de la fonction coût

# Matrices de variance - covariance

- Pour  $y$ , variance du bruit de mesure :  $S_\varepsilon$  (matrice diagonale, pas d'erreur modèle)
  - Pour  $x$ , variance de  $x_{\text{a priori}}$  :  $S_a$
  - Pour  $x$ , variance a posteriori estimé :  $\hat{S} = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1}$
- $$S_a : \text{Var}(X) = \begin{pmatrix} \sigma(X_1)^2 & \sigma(X_1, X_N) \\ & \sigma(X_N)^2 \end{pmatrix}$$
- La diagonale représente l'incertitude (écart-type au carré)
  - Les éléments hors diagonale représentent les corrélations, entre niveaux verticaux, ou entre éléments du vecteur d'état



# Méthode Gauss-Newton

- Il s'agit d'une « méthode quadratique »

- Recherche du 0 de la dérivée
- Pour une fonction coût quadratique en X, la dérivée est linéaire, donc le calcul de son minimum est analytique

$$x = x_a - \left[ \frac{\partial^2(P(X/Y))}{\partial X^2}(x_a) \right]^{-1} \frac{\partial(P(X/Y))}{\partial X}(x_a) \Rightarrow \hat{x} = x_a + (S_a^{-1} + K^T S_\epsilon^{-1} K)^{-1} K^T S_\epsilon^{-1} (y - Kx_a)$$

- Ce calcul nécessite le calcul de la dérivée seconde (hessienne)
- La fonction serait quadratique si  $Y=K.X$  (K indépendant de X), ce qui est faux en général

- Algorithme de Newton : on itère, pour chaque  $x_i$  estimé, on recalcule les jacobiens et on recalcule analytiquement la position du minimum**

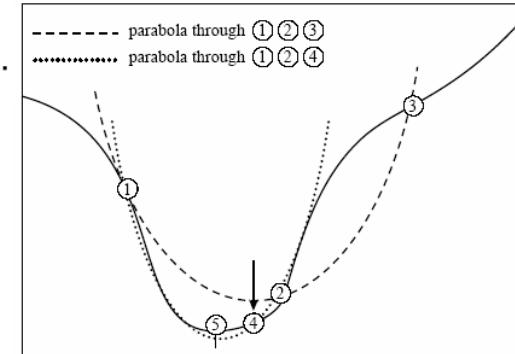
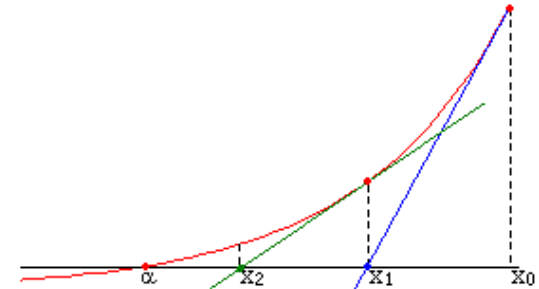
- Cela revient à approximer localement la fonction coût par une nouvelle fonction quadratique

- Méthode de Gauss-Newton : on néglige certains termes d'ordre 2 difficiles à calculer (Hessienne simplifiée)**

$$x_{i+1} = x_i - \left[ \frac{\partial^2(P(X/Y))}{\partial X^2}(x_i) \right]^{-1} \frac{\partial(P(X/Y))}{\partial X}(x_i) \quad \frac{\partial^2(P(X/Y))}{\partial X^2}(x_i) = S_a^{-1} + K^T S_\epsilon^{-1} K - [\nabla_x K^T] S_\epsilon^{-1} [y - F(x)]$$

➡  $x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i)^{-1} [K_i^T S_\epsilon^{-1} (y - F(x_i)) - S_a^{-1} (x_i - x_a)]$

- $x_0$  (first guess) pris égal à l'a priori



## Méthode Levenberg - Marquardt

- Dans GN, le pas  $(x_{i+1} - x_i)$  peut être assez grand  
 → Si la fonction coût est loin d'être quadratique, ces pas peuvent provoquer des oscillations fortes et avoir du mal à trouver le minimum
- La méthode LM propose de modifier GN en « modérant » le pas des oscillations
  - Ajout du terme  $\gamma I$  dans la matrice Hessienne

$$x_{i+1} = x_i - (KK^T + \gamma_i I)^{-1} K^T [y - F(x_i)] \quad \Rightarrow \quad x_{i+1} = x_i + (K_i^t S_\epsilon^{-1} K_i + S_a^{-1} + \gamma D)^{-1} (K_i^t S_\epsilon^{-1} (y - F(x_i)) - S_a^{-1} (x_i - x_a))$$

(exprimé ici sans a priori et sans bruit)

- L'ajout d'un terme en  $\gamma I$  revient à faire un mélange entre GN et l'algorithme du gradient
  - L'algorithme du gradient revient à suivre le gradient avec un pas prédéterminé, cohérent avec l'échelle du problème, mais comment bien le choisir?
  - Rodgers préconise de prendre  $I = S_a^{-1}$
  - Faire évoluer  $\gamma$  selon la linéarité du problème

$$R = \frac{c(x_{i+1}) - c(x_i)}{c_{lin}(x_{i+1}) - c(x_i)}$$

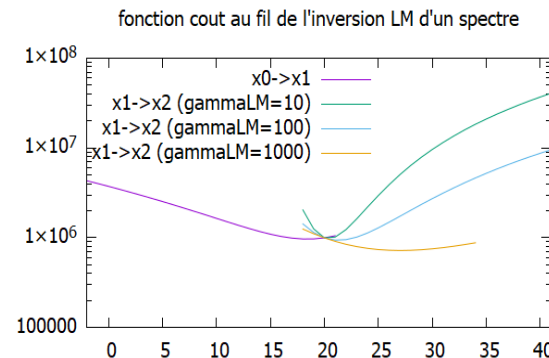
- $c$  est fonction coût  $c(x_i) = (y - F(x_i))^t S_\epsilon^{-1} (y - F(x_i)) + (x_i - x_a)^t S_a^{-1} (x_i - x_a)$
- $c_{lin}(x_{i+1})$  est calculé avec  $F(x_{i+1}) = F(x_i) + K_i(x_{i+1} - x_i)$
- Si  $R < 0$ ,  $x_{i+1}$  est rejeté et  $\gamma = \gamma * R_{mult}$  ( $R_{mult}$  typ.10)
- Si  $0 \leq R < 0.25$ ,  $x_{i+1}$  est accepté et  $\gamma = \gamma * R_{mult}$
- Si  $0.25 \leq R < 0.75$ ,  $x_{i+1}$  est accepté et  $\gamma$  reste inchangé
- Si  $R \geq 0.75$ ,  $x_{i+1}$  est accepté et  $\gamma = \gamma / R_{div}$  ( $R_{div}$  typ.2)

## Autres éléments

- **Les valeurs estimées dans xi peuvent prendre des valeurs non physiques**
  - Ex:  $X_{CO_2} < 0$ , ou même  $X_{CO_2} < 380\text{ppm}$
  - ➔ On s'autorise 3 solutions
    - Arrêt des itérations (on est allé trop loin de la vérité)
    - Seuillage aux valeurs physiques limites (moyen car le VDE n'est plus cohérent)
    - Reset à l'a priori (pas terrible)
- **2 critères d'arrêt (GN et LM)**
  - **Stabilisation du chi ((mesure – simu)/bruit)**
  - **ET Stabilisation du VDE**
  - On évalue aussi le chi pour information

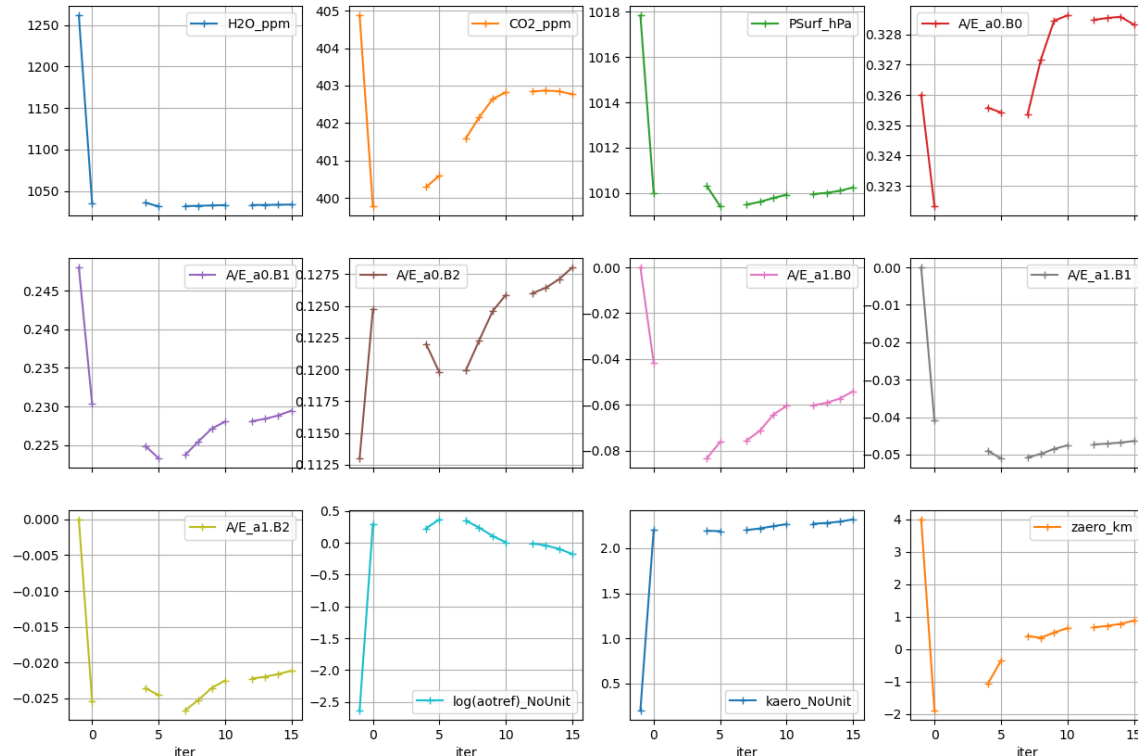
## Quelques réflexions théoriques

- **Gros défaut connu de ces méthodes : incapacité à sortir d'un minimum local**
  - Or des minima locaux sont probables dans MicroCarb !
  - Ex : Augmenter  $P_{surf}$  approfondit les raies sans toucher au continuum, augmenter albédo multiplie le spectre sans changer la profondeur des raies, augmenter les aérosols change les raies ET multiplie le spectre
  - La seule manière de s'en assurer serait de cartographier la fonction coût. Pas évident dans un espace de dimension  $\sim 50$ ...
  - Quelques tentatives dans la direction du gradient:
- **La littérature justifie LM pour les cas où la fonction coût est loin d'être quadratique. Mais cette considération est toujours faite sans utilisation d'a priori. L'utilisation d'a priori dans la fonction coût est une régularisation (accroche à l'a priori) qui limite les variations de la fonction coût**
  - ➔ N'y-a-t-il pas redondance entre utilisation d'a priori et Levenberg-Marquardt?
- **Le modèle direct est loin d'être parfait**
- **Certains défauts instrument utilisent les degrés de liberté de l'inversion**
- **La porte reste ouverte à des simplifications supplémentaires**
  - ➔ Estimer des scaling factors / gradients verticaux de profils a priori (2 paramètres)
  - ➔ Optimiserait par rapport aux DOFS (entre 2 et 3, contre 19 couches estimées)



# Illustration des difficultés : inversion d'un spectre OCO-2

- Inversion d'un spectre OCO-2 acquis à proximité d'une station TCCON (réseau mondial de spectromètres au sol dédiés à la mesure du CO2).
- On dispose de valeurs de référence pour tous les éléments du vecteur d'état, que l'on utilise comme *a priori* et first-guess pour l'inversion.

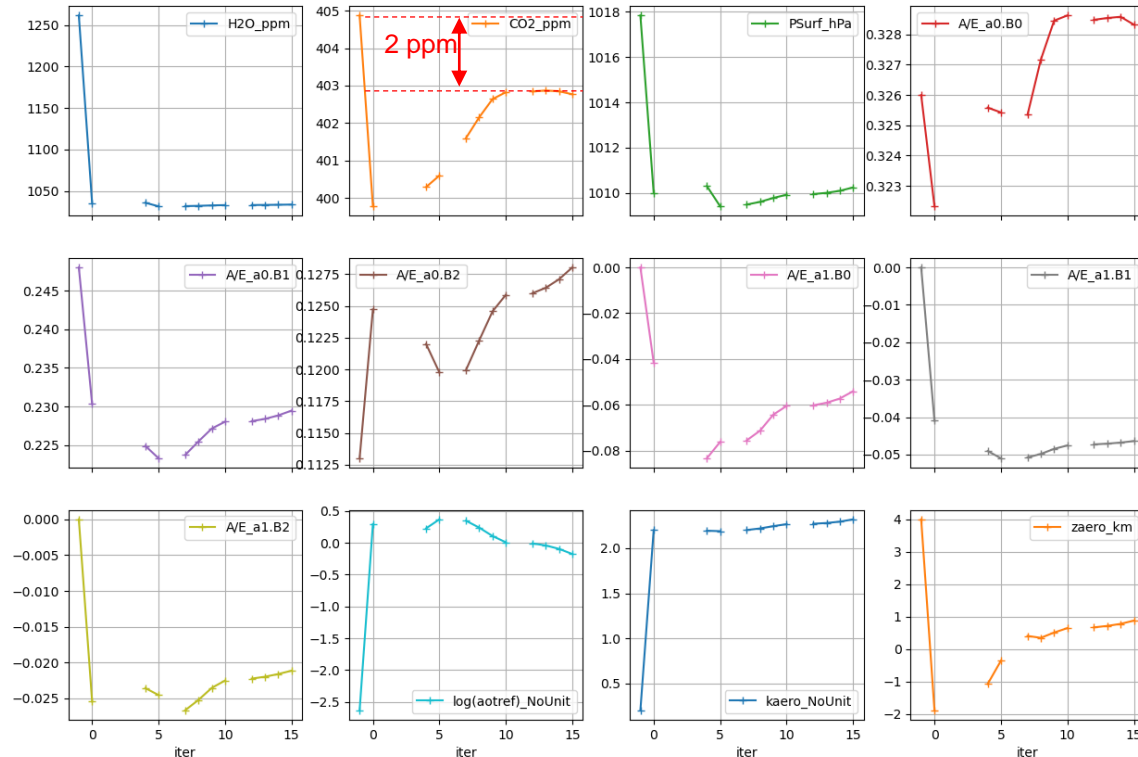


Inversion OCO-2 : paramètres estimés au cours des itérations. Point manquant = VDE rejeté car ne permet pas de réduire la fonction de coût

# Illustration des difficultés : inversion d'un spectre OCO-2

- **Pas de convergence** : arrêt après 15 itérations (augmentation successive du paramètre  $\gamma$  jusqu'à  $10^4$  sans parvenir à réduire la fonction de coût)
- **Environ 2 ppm de différence sur le CO2 estimé** par rapport à TCCON (très supérieur à l'erreur aléatoire)
- Résidu important en fin d'inversion :

$$\chi = \sqrt{\frac{(y - F(x_i))^t S_\epsilon^{-1} (y - F(x_i))}{m}} = 2.22$$



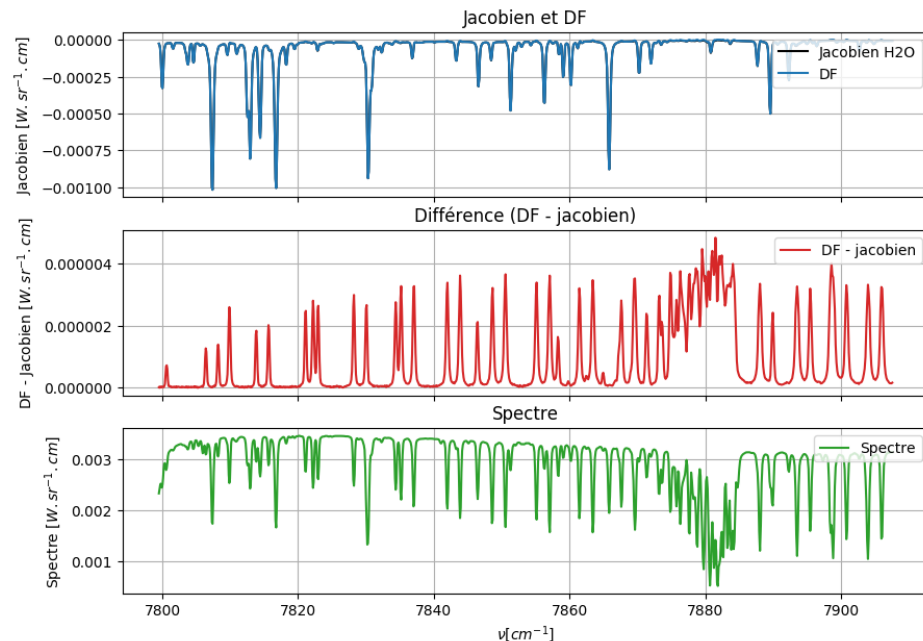
Inversion OCO-2 : paramètres estimés au cours des itérations. Point manquant = VDE rejeté car ne permet pas de réduire la fonction de coût

## **Comment expliquer les problèmes de convergence ainsi que les écarts obtenus, et comment les améliorer ?**

- **Plusieurs causes probables ont été identifiées et des pistes d'améliorations ont été proposées :**
  - Améliorer la précision des dérivées partielles du modèle direct
  - Améliorer le paramétrage de l'algorithme de Levenberg-Marquardt
  - Améliorer la modélisation du spectre
- **Dans la suite de la présentation, on détaille ces pistes d'améliorations et les résultats obtenus.**

## Amélioration de la précision des jacobiens

- Le modèle direct permet le calcul des dérivées partielles par rapport aux paramètres estimés (les estimations par différences finies sont trop coûteuses en temps de calcul).
- Le modèle direct est complexe (appel au code de transfert radiatif 4A) donc certains termes qui contribuent aux dérivées partielles sont négligés.
- Une matrice jacobienne imprécise peut perturber l'inversion (le gradient de la fonction de coût est un peu faux)
- Il n'est pas évident d'estimer l'impact d'une erreur de matrice jacobienne sur les inversions, et donc l'erreur acceptable sur la jacobienne.



Jacobien H2O en bande B4 : mise en évidence d'une erreur liée au jacobien O2 de l'ordre de 1%

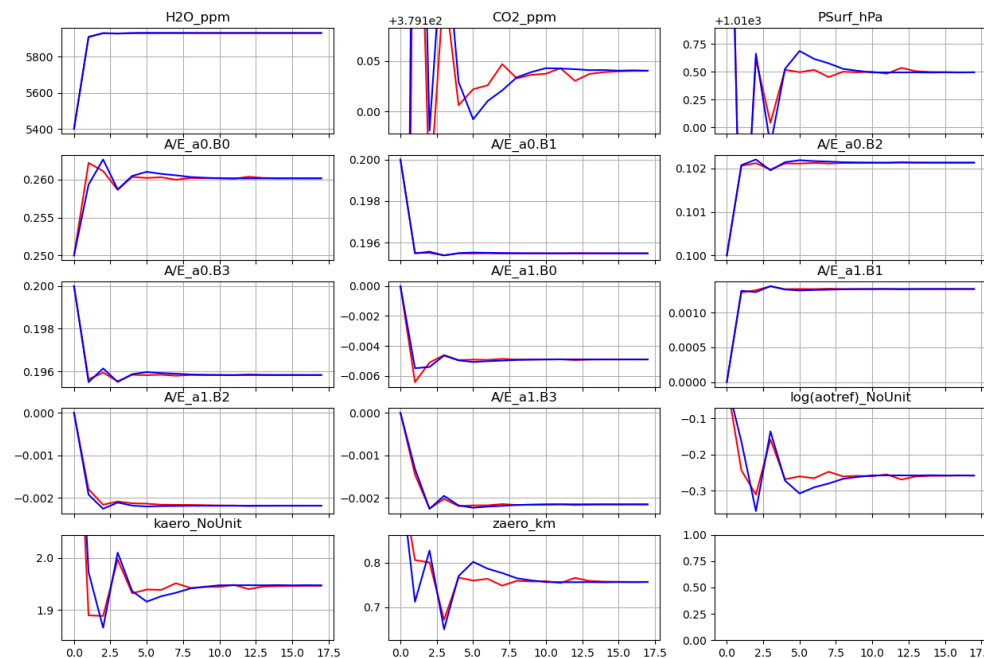
# Amélioration du paramétrage de Levenberg-Marquardt

- Itération Levenberg-Marquardt :

$$x_{i+1} = x_i + (K_i^t S_\epsilon^{-1} K_i + S_a^{-1} + \gamma D)^{-1} (K_i^t S_\epsilon^{-1} (y - F(x_i)) - S_a^{-1} (x_i - x_a))$$

- La matrice D doit être définie positive (pour que  $x_{i+1} - x_i$  se rapproche d'une direction de descente quand  $\gamma$  est grand), mais le choix diffère selon les référence / implémentation :
  - Rodgers propose  $D = S_a^{-1}$  (mais  $S_a$  n'est pas forcément adaptée à cet usage, elle est complètement indépendante du modèle
  - Le choix de l'implémentation MINPACK revient dans notre cas à choisir est  $D = \text{diag}(K_i^t S_\epsilon^{-1} K_i + S_a^{-1})$  : diagonale de la Hessienne
- Sur des inversions de spectres synthétiques, le second choix donne une convergence plus rapide et régulière.

Impact du choix de D sur une inversion de spectre synthétique



$$D = \text{diag}(K_i^t S_\epsilon^{-1} K_i + S_a^{-1}) \quad D = S_a^{-1}$$

# Amélioration du paramétrage de Levenberg-Marquardt : cas OCO-2

- Mise à jour du paramétrage LM, et inversion du même spectre OCO-2 que précédemment (sans rien changer au reste de l'inversion)

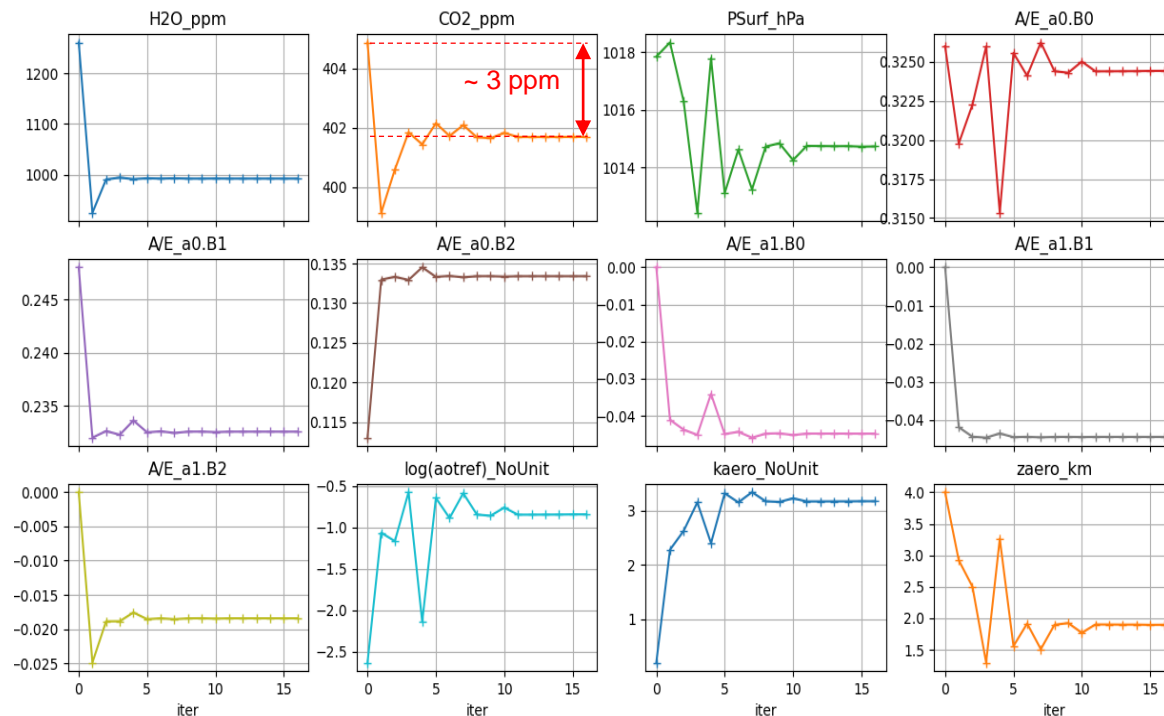
→ **Convergence en 10 itérations environ**

Le paramètre  $\gamma$  est d'environ 0.01 en fin d'inversion. **C'est mieux, mais encore trop ! Les oscillations sont faiblement amorties**

→ **Environ 3 ppm de différence sur le CO2 estimé : c'est pire !**

→ Résidu un peu plus faible en fin d'inversion (donc solution a priori meilleure), mais qui reste important :

$$\chi = 2.20$$

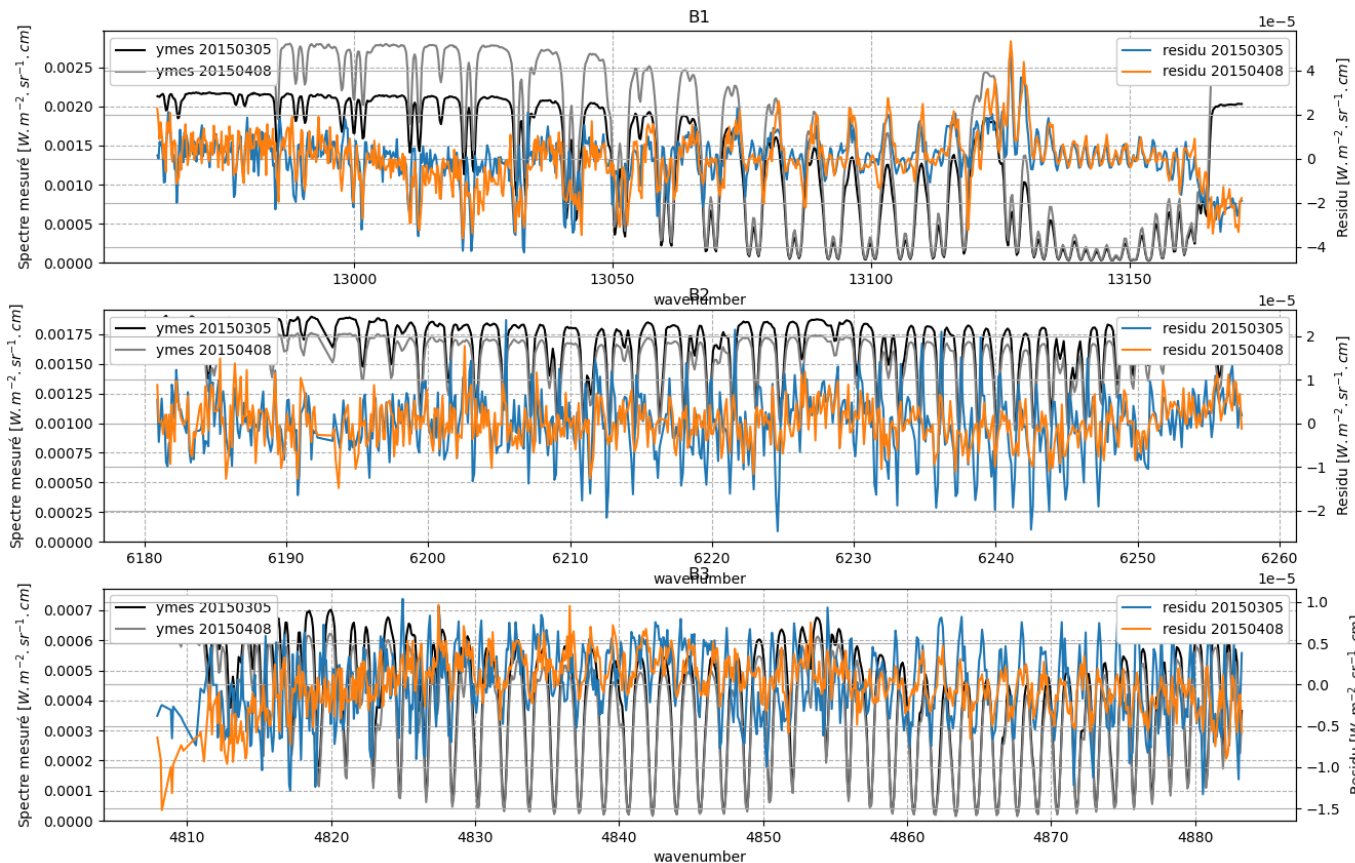


Inversion OCO-2 après maj LM : paramètres estimés au cours des itérations.

# Amélioration du paramétrage de Levenberg-Marquardt : cas OCO-2

- L'analyse des résidus d'inversion mets en évidence le fait que le spectre n'est pas bien modélisé
- Certains paramètres peuvent absorber partiellement les effets non modélisés, ce qui peut conduire à des biais sur les estimations en particulier sur le CO<sub>2</sub>

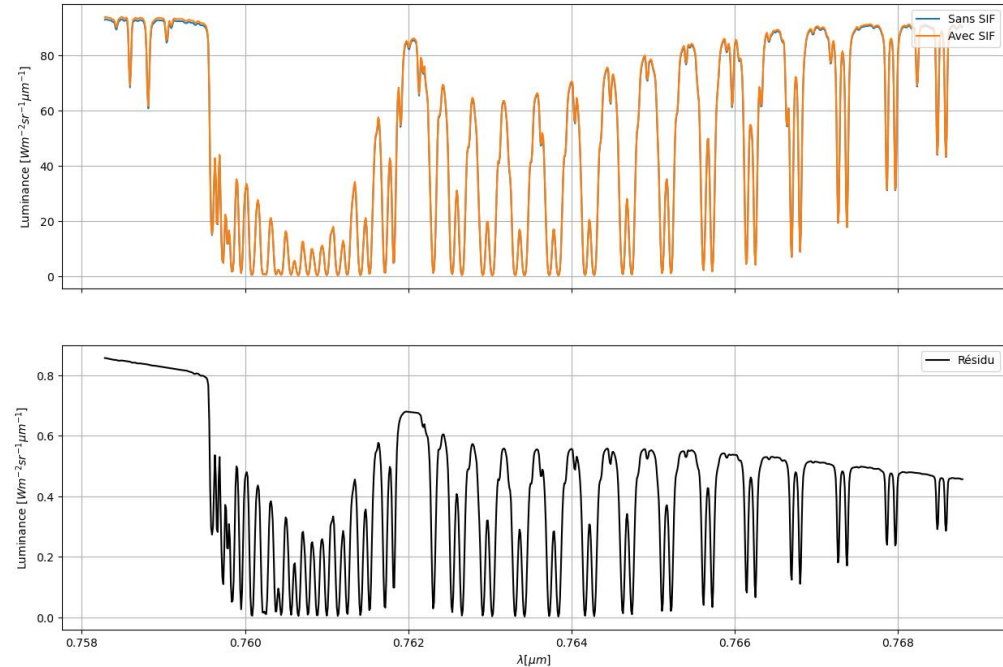
→ Nécessité d'améliorer la précision du modèle direct



Résidu inversion spectres OCO-2 sur deux stations TCCON différentes

# Amélioration de la modélisation : exemple de la SIF (Solar Induced Fluorescence)

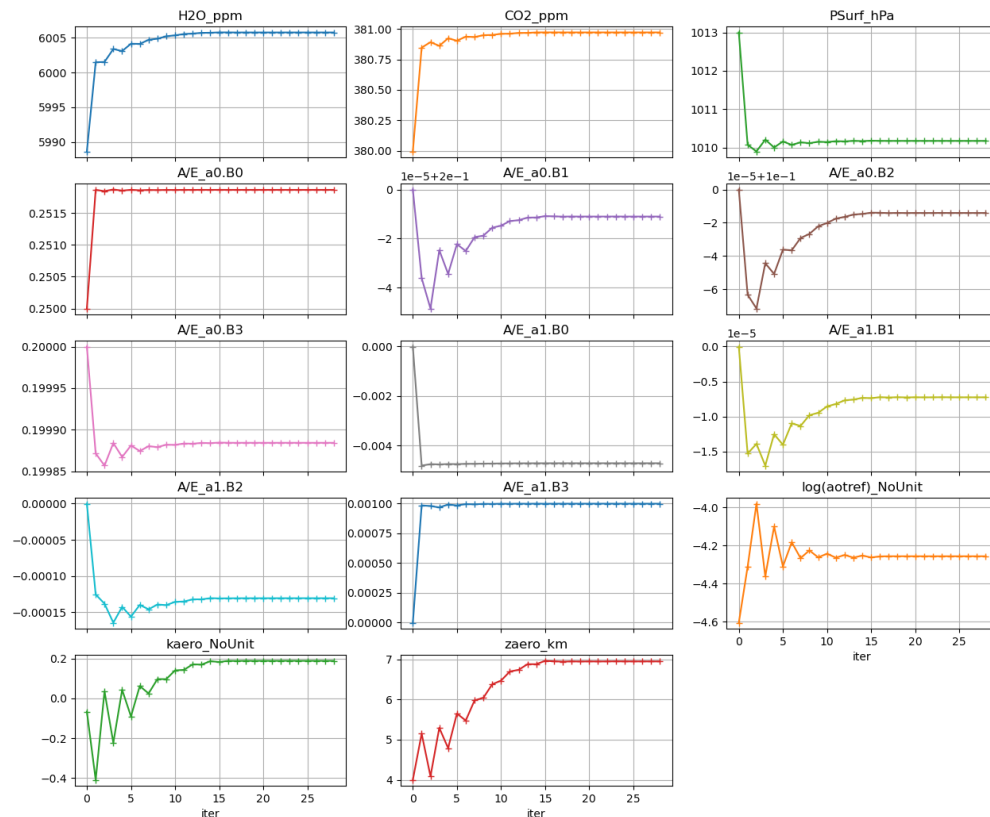
- La végétation émet au niveau du sol un flux lumineux qui subit une absorption atmosphérique sur le trajet montant. Le signal atteint typiquement 1% du signal total dans le continuum (zone sans raies d'absorption) de la bande 1 de MicroCarb
- On regarde en simulation l'effet qu'a une non prise en compte de cet effet dans le modèle utilisé pour l'inversion.



Spectre sans SIF (bleu) et avec SIF à 1% de continuum (orange) et la différence (noir)

# Amélioration de la modélisation : exemple de la SIF (Solar Induced Fluorescence)

- Sans prise en compte de la SIF dans le modèle direct :
  - **biais d'1 ppm sur le CO<sub>2</sub>**, comme prévu par le transfert d'erreur linéaire
  - **convergence lente** : oscillations faiblement amorties, 12 itérations nécessaires et 14 appels au modèle direct
- Avec prise en compte de la SIF:
  - **pas de biais**
  - **convergence quadratique, 3-4 itérations max** (perturbation du first-guess)

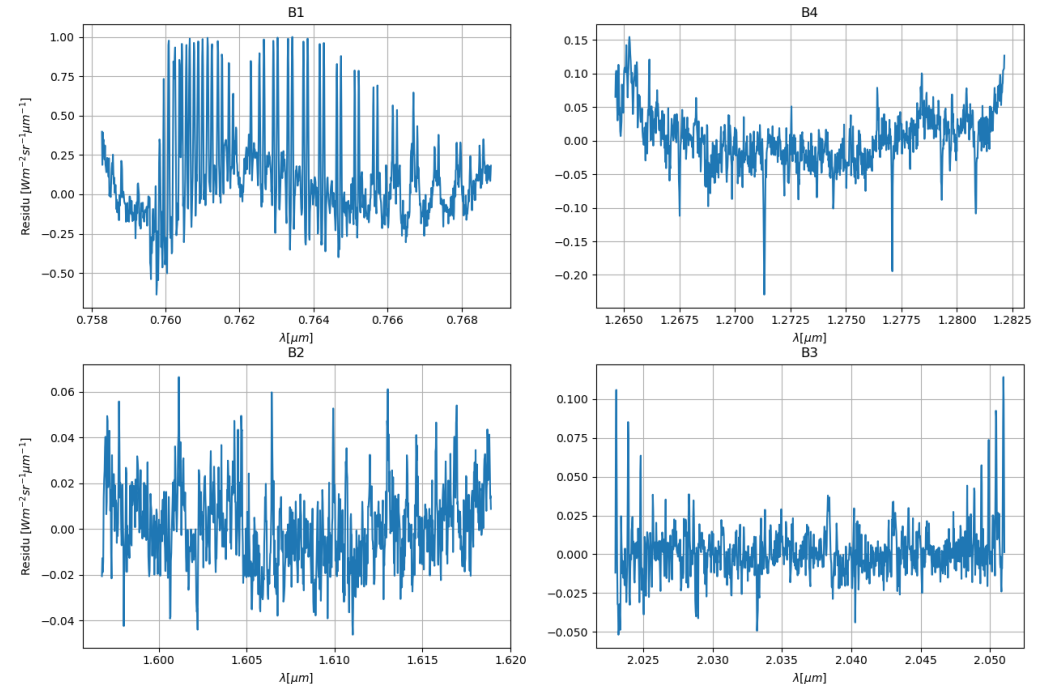


Paramètres estimés au cours de l'inversion d'un spectre avec modèle sans prise en compte de la SIF

# Résidus : mise en évidence de l'impact sur la convergence avec un résidu synthétique orthogonal

- Comme aperçu sur l'exemple SIF, la convergence semble gênée par la présence d'un résidu dans le spectre inversé.
- Pour le vérifier, on simule un spectre et on calcule la matrice de gain  

$$G = (K^t S_{\epsilon}^{-1} K + S_a^{-1})^{-1} K^t S_{\epsilon}^{-1}$$
 (transport de biais linéaire)
- On génère un vecteur résidu synthétique  $r$  tel que  $G r = 0$ , donc qui n'engendre pas de biais au premier ordre.
- On injecte ce résidu dans le spectre que l'on inverse.

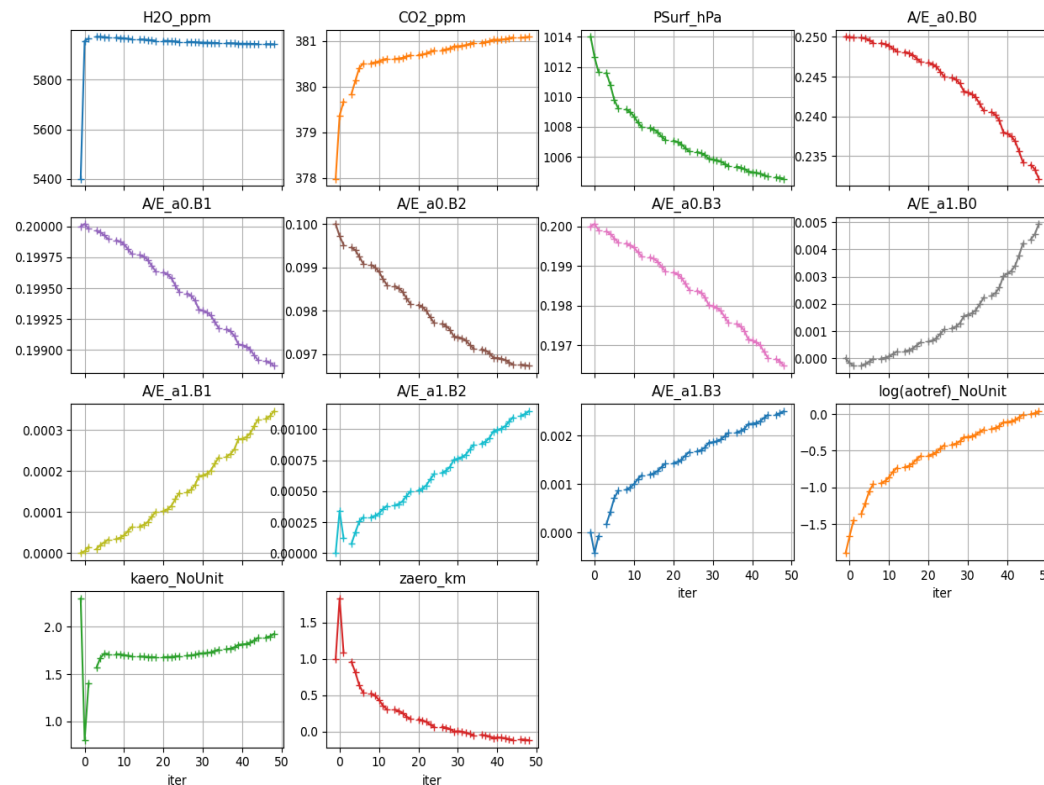


Exemple de résidu orthogonal pour un spectre simulé MicroCarb (environ 1% du spectre)

# Résidus : mise en évidence de l'impact sur la convergence avec un résidu synthétique orthogonal

- Si le résidu est petit, le biais est nul et la convergence n'est pas ralentie
- A mesure que l'on augmente le résidu, la convergence est de plus en plus lente avec un biais assez faible.
- Ci-contre le cas où le résidu atteint 1% du spectre en valeur max : **pas de convergence**

→ Une mauvaise modélisation engendre des résidus lors de l'inversion qui peuvent ralentir fortement la convergence.



Exemple de résidu orthogonal pour un spectre simulé MicroCarb (environ 1% du spectre)

## Amélioration du modèle : autres sources d'erreur

- Plusieurs autres sources d'erreurs de modélisation
  - Erreurs de spectroscopie
  - Effets instrumentaux non corrigés : shift spectral, résidus de non-linéarité
  - Paramétrisation aérosol très simpliste
  - Autres ...
- Une solution possible est de déterminer des résidus empiriques à partir des résultats d'inversion et de les décomposer en composantes principales (Empirical Orthogonal Functions). On ajoute un paramètre dans le vecteur d'état pour quelques composantes principales du résidu.

## Conclusions

- **De la théorie à la pratique, il y a un monde!**
- **Avec des impacts potentiellement importants sur la perfo mission et le dimensionnement du segment sol**
- **D'autres pistes testées (dans le cadre d'un code de transfert radiatif)**
  - Estimation amplitudes de composantes principales
  - Régularisation Tikhonov (norme de la dérivée du profil vertical)
  - Ajout terme de régularisation par proximité aux limites physiques des valeurs
  - Etc.
- **On est loin d'avoir résolu tous nos problèmes**
- **De l'aide est la bienvenue ☺**
  - Expertise
  - Thèses, CDD, post-doc, assistance technique...