

# Algorithme d'inversion (durée 1h30)

Sébastien Payan – LPMAA – juin 2009

## Sommaire

1	Introduction .....	3
2	Inversion des spectres .....	5
2.1	Le concept de l'inversion .....	5
2.2	Quelques approches dans le cas multi-spectres.....	6
2.3	Ajustement des spectres par la méthodes des moindres carrés .....	8
2.3.1	Moindre carré linéaire .....	9
2.3.2	Moindre carré non-linéaire.....	13
2.3.3	Linéarisation du modèle direct.....	15
2.3.4	Moindre carré avec amortissement .....	18
2.3.5	Méthode de Levenberg-Marquardt .....	20

2.3.6	Information <i>a priori</i> .....	28
3	Contenu en information.....	31
3.1	Fonctions de poids/ Jacobiens .....	31
3.2	Noyaux moyens.....	35
3.3	Entropie informationnelle / degrés de liberté .....	37
4	Analyse et caractérisation des erreurs .....	39
4.1	Les différentes sources d'erreurs.....	39
4.2	Bruit de mesure .....	41
4.3	Erreur de lissage .....	41
4.4	Paramètres du modèle direct .....	41
4.5	Modèle direct .....	42
4.6	Matrice covariance erreur.....	42
5	Les autres méthodes .....	43

# 1 Introduction

Dans la majorité des cas le signal reçu par un senseur optique n'est pas directement interprétable en terme de paramètre atmosphérique.

Il faut pour chaque instrument mettre au point un **algorithme** permettant, à partir des mesures spectrales (radiances), de remonter aux **paramètres recherchés** (profils verticaux de température ou de concentration par exemple).

On utilise le terme **méthodes d'inversion** pour désigner les grandes classes d'algorithmes utilisés. Dans tous les cas il faut savoir traiter le **problème direct** c'est à dire disposer d'un algorithme permettant de calculer le signal que devrait recevoir l'instrument pour un état donné de l'atmosphère et dans une géométrie donnée.

On comprend que ce calcul sera d'autant plus fiable que la **modélisation de l'instrument** lui même ainsi que celle du **transfert radiatif** sera plus précise (d'où l'importance des paramètres spectroscopiques).

Il faut ensuite construire une méthode inverse précise et performante (souvent itérative, souvent avec amortissement car "problème mal posé") qui tient compte au mieux de l'information contenue dans le signal atmosphérique.

Il faut aussi, pour un dépouillement opérationnel par exemple, automatiser le processus au maximum et **optimiser le temps d'exécution** (et donc quelquefois faire des approximations) ce qui n'est pas toujours

compatible avec la précision ultime qu'on pourrait atteindre avec un algorithme plus rigoureux. Nous ne pouvons bien entendu entrer dans le détail de toutes ces techniques

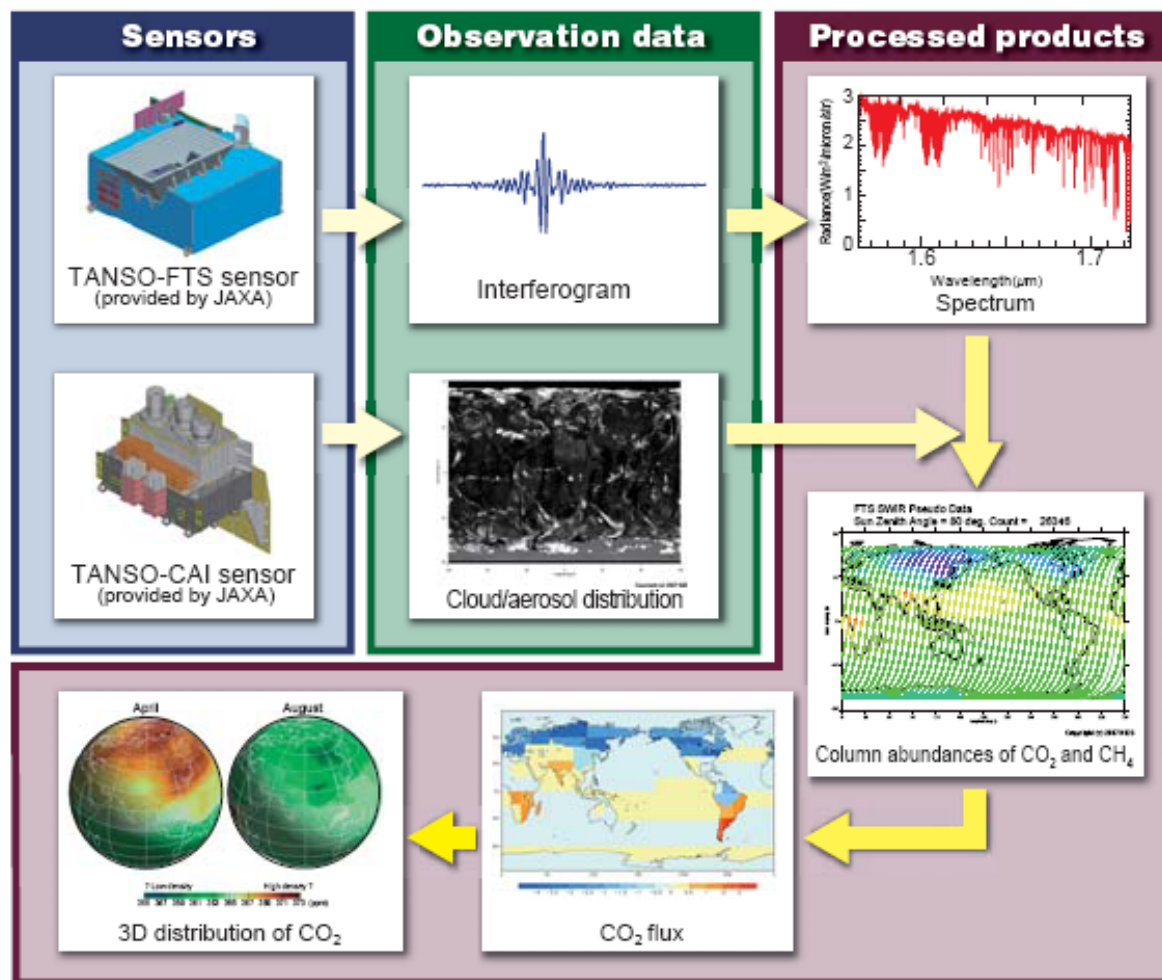


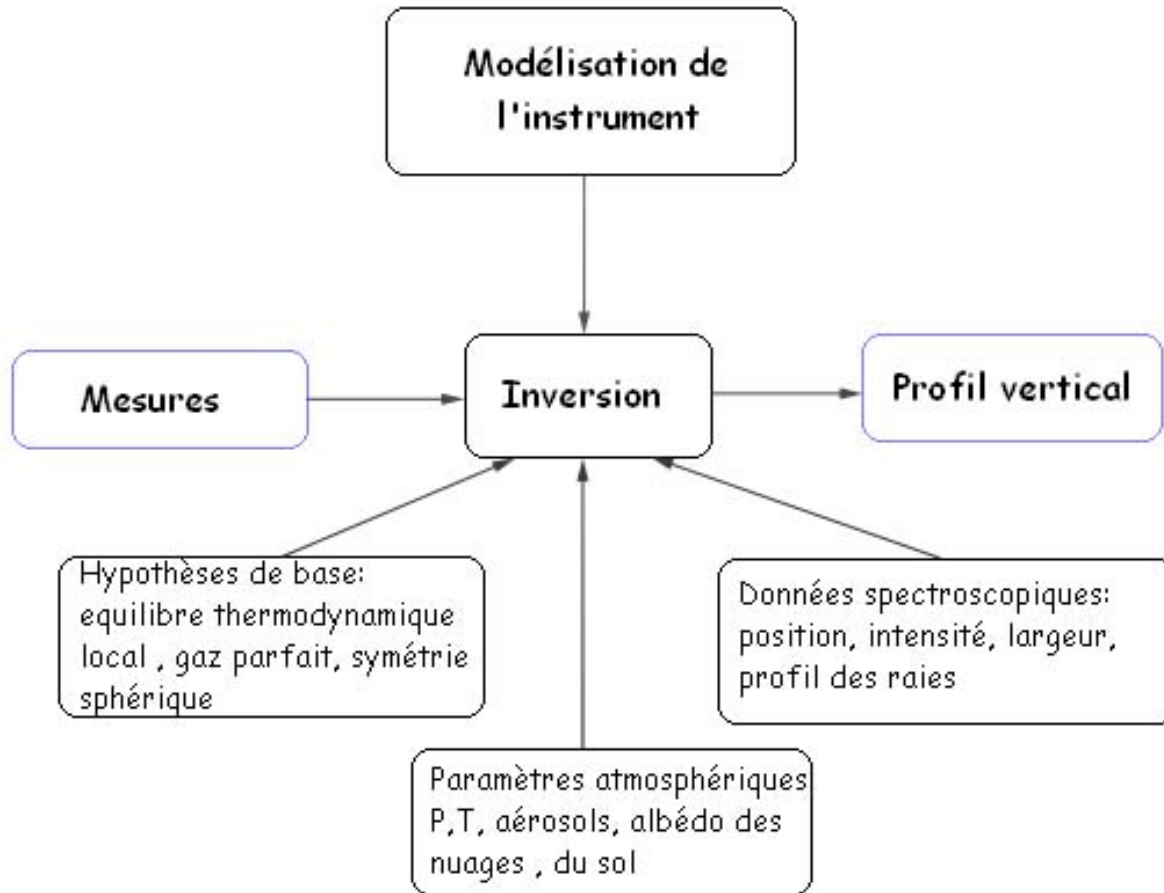
Figure 7. Outline of GOSAT data processing

1

<sup>1</sup> GOSAT Research announcement, Appendix A, 2008

## 2 Inversion des spectres

### 2.1 Le concept de l'inversion

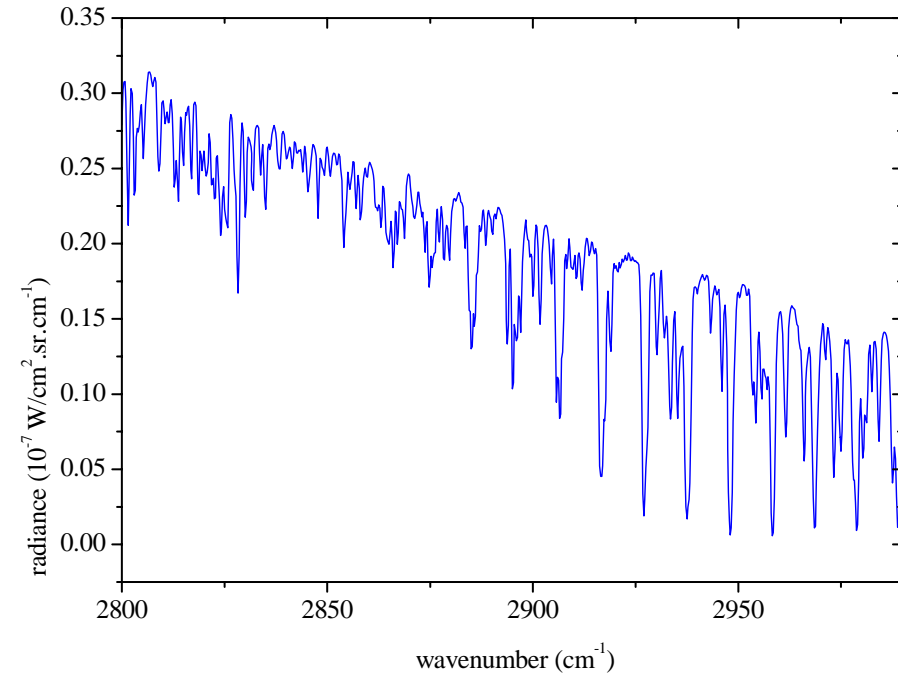
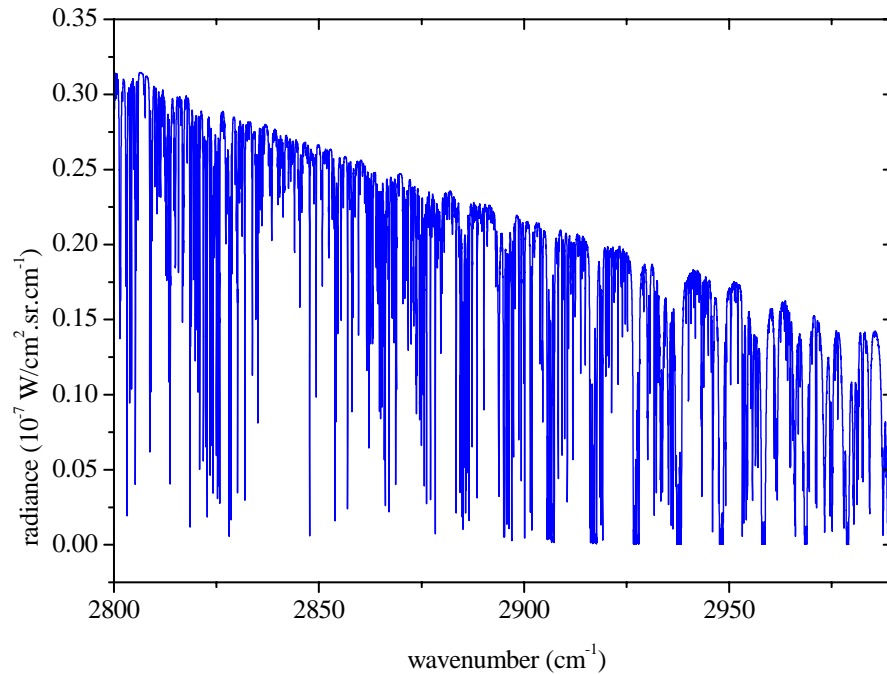


Le problème qui se pose est d'utiliser au mieux l'information dont on dispose : spectre ou ensemble de spectres atmosphériques correspondant à une ou des situations (visées) à partir desquels il est possible de "mettre en parallèle" un ou plusieurs des paramètres que l'on cherche à déterminer.

Mais dans la majorité des cas le signal reçu par un senseur embarqué n'est pas directement interprétable en terme de paramètres atmosphériques. Il faut pour chaque instrument mettre au point un algorithme permettant, à partir des mesures spectrales, de remonter aux paramètres recherchés (des profils de concentration par exemple).

⇒ **Modèle inverse**

Effets instrumentaux : exemple de la fonction d'appareil de IASI :

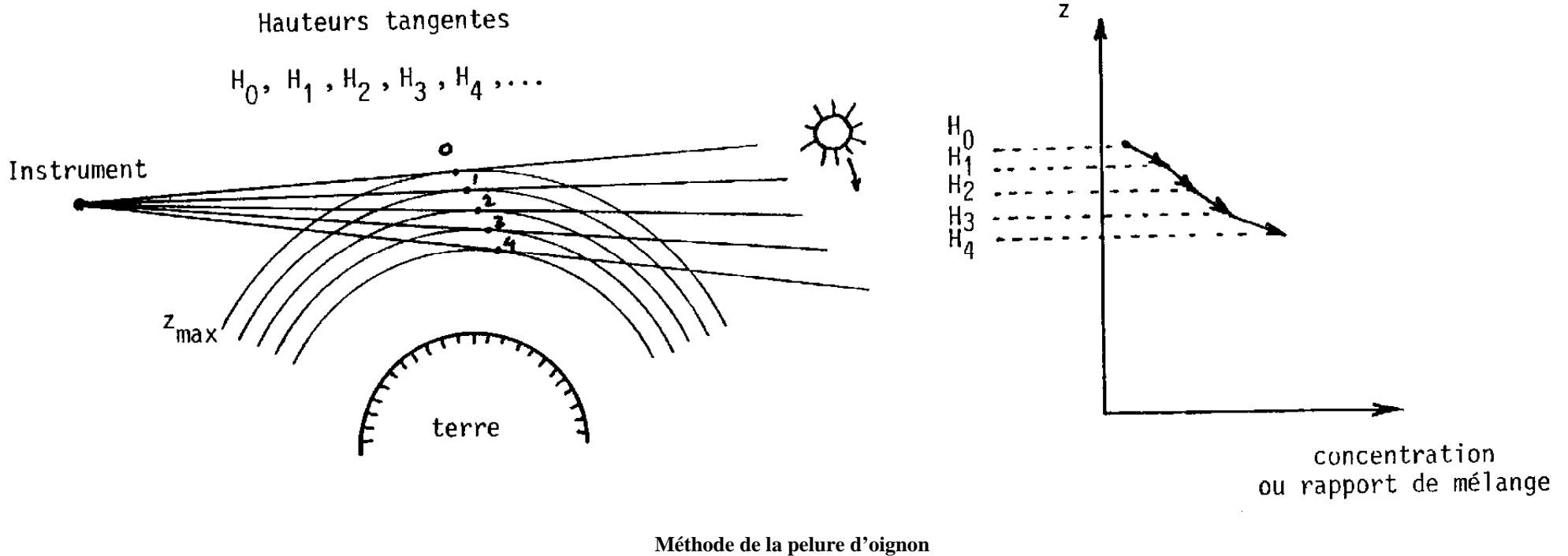


## 2.2 Quelques approches dans le cas multi-spectres

Nous mentionnerons brièvement trois méthodes largement utilisées :

- deux méthodes applicables à basse ou moyenne résolution spectrale
  - la méthode de la pelure d'oignon (onion-peeling technique)
  - la méthode de Mill

- La méthode de l'ajustement global : utilisable pour les spectres à très haute résolution. Elle consiste à ajuster simultanément (au sens des moindres carrés) tous les paramètres (profils verticaux de concentration, mais aussi fonction d'appareil, spectre solaire,...) pour une ou plusieurs espèces, pour toutes les géométries (situations) et pour tous les intervalles (windows, micro-windows) retenus dans l'analyse d'une occultation.



## 2.3 *Ajustement des spectres par la méthodes des moindres carrés*

On dispose donc d'un ensemble de mesures spectrales (spectre) qui sont représentatives d'un ensemble de paramètres dont on cherche à déterminer la valeur.

Nous utiliserons dans la suite les notations suivantes :

- l'indice  $i$  correspond à un élément spectral donné
- on appelle  $n_{pt}$  le nombre total de points spectraux considérés)
- l'indice  $j$  (ou  $k$ ) représentera un des  $np$  paramètres à déterminer

La détermination des paramètres ne sera évidemment possible que si la condition  $n_{pt} \geq np$  est vérifiée.

Nous appellerons :

- $Y$  la grandeur mesurée (ici la radiance spectrale)
- $\hat{Y}$  (ou  $F(X)$ ) la radiance calculée par le modèle
- $X$  les paramètres à déterminer
- $E$  l'écart entre les spectres calculés et mesurés.



La relation générale entre grandeurs mesurées et paramètres du modèle s'écrit sous la forme :

$$y_i = \hat{y}_i + e_i = f_{ij}(x_j) + e_i \quad (1)$$

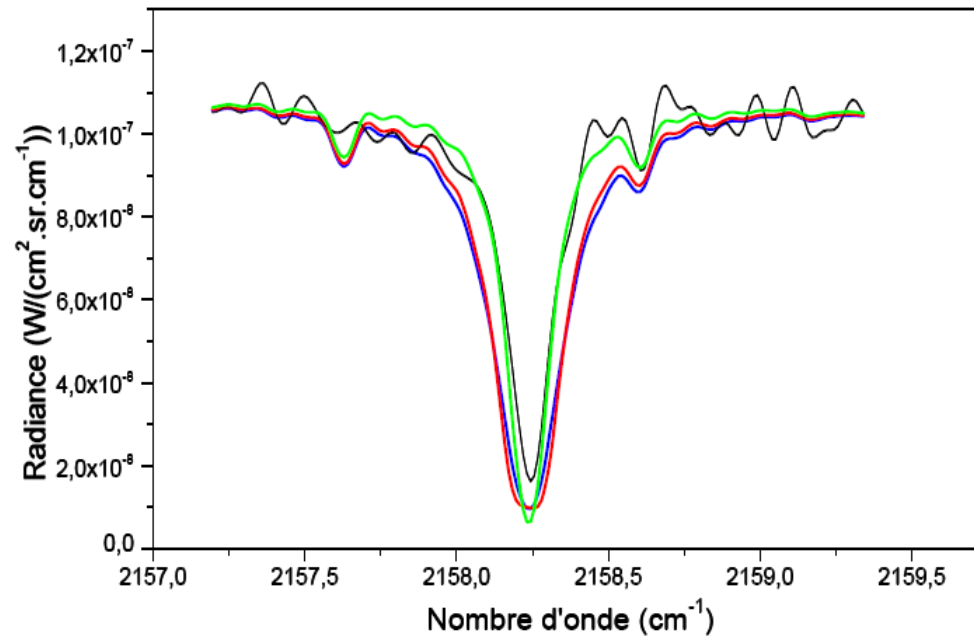
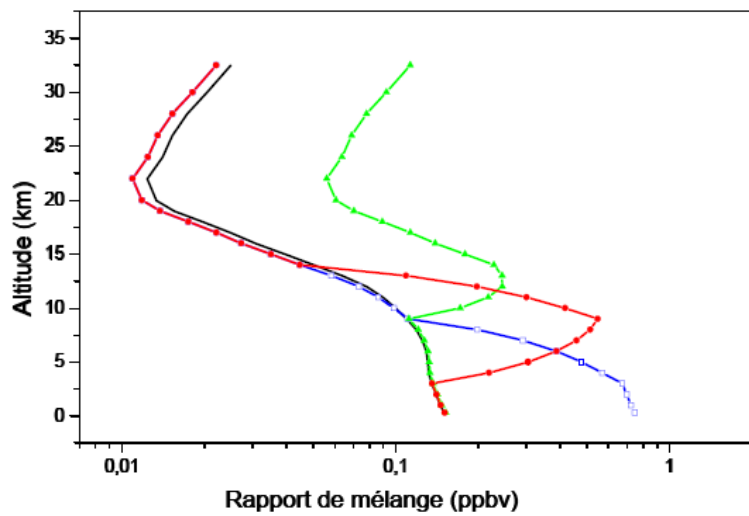
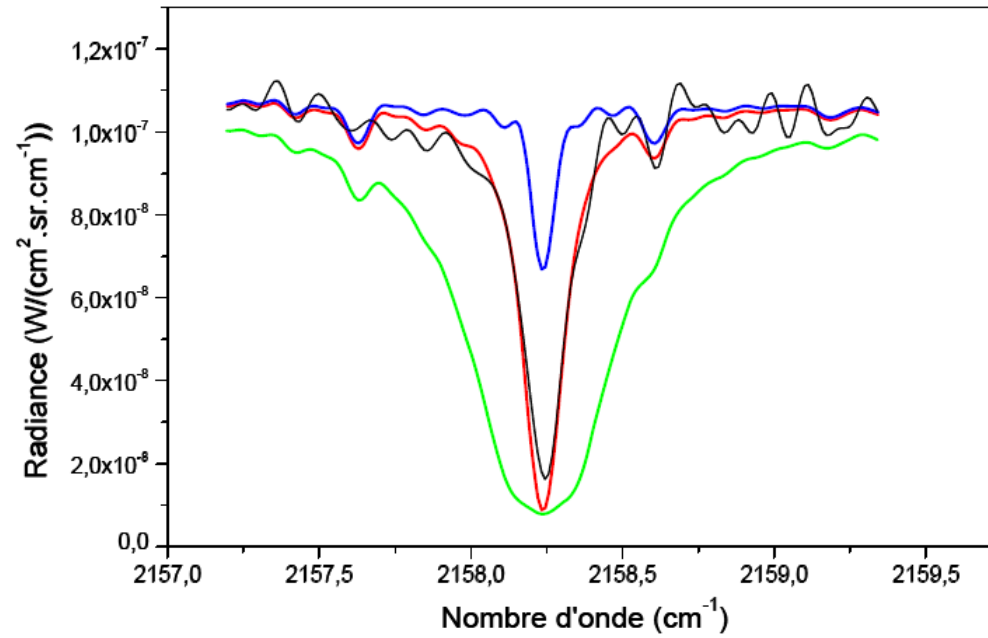
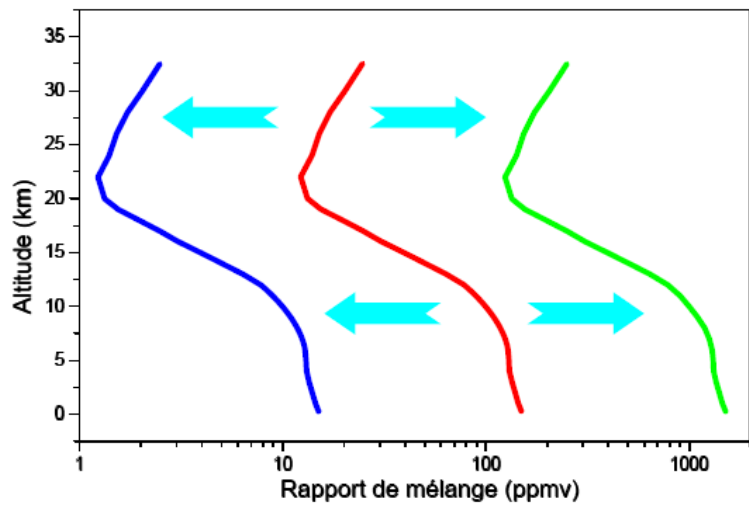
### 2.3.1 Moindre carré linéaire

Plaçons-nous dans un cas simple où la mesure peut-être restituée par un modèle linéaire. L'équation (1) peut alors se réécrire sous la forme matricielle :

$$Y = \hat{Y} + E = FX + E \quad (2)$$

Le spectre  $\hat{Y}$  est aussi appelé spectre synthétique.

Nous cherchons à déterminer les paramètres  $x_j$  qui reproduisent au mieux la mesure, c'est-à-dire ceux pour lesquels les écarts  $e_i$  ( $= y_i - f_{ij}(x_j)$ ) sont les plus faibles possibles (on appelle également les  $e_i$  résidus).



Nous cherchons à déterminer les paramètres  $x_j$  qui reproduisent au mieux la mesure, c'est-à-dire ceux pour lesquels les écarts  $e_i$  sont les plus faibles possibles. Cela revient à minimiser la fonction de  $\chi^2$  défini par :

$$\chi^2 = \sum_{i=1}^{npt} \frac{(y_i - \hat{y}_i)^2}{(\Delta y_i)^2} = \sum_{i=1}^{npt} (y_i - \hat{y}_i)^2 w_i \quad (3)$$

où  $\Delta y_i$  représente l'écart type sur la mesure  $y_i$  et  $w_i = 1/\Delta y_i^2$  le poids correspondant.

On définit également le  $\chi^2$  réduit (noté  $\chi_r^2$ ) par :

$$\chi_r^2 = \frac{\chi^2}{npt - np} \quad (4)$$

On définit ensuite un critère de qualité de l'ajustement qui sera d'autant meilleur au sens statistique qu'il sera proche de 1 (résidus équivalent au bruit de mesure).

Lorsque le  $\chi_r^2$  est  $> 1$ , l'inversion n'est pas satisfaisante.

Lorsque le  $\chi_r^2$  est  $< 1$ , les valeurs des erreurs sur les mesures sont vraisemblablement sur-estimées.

Nous avons vu précédemment que les paramètres  $x_j$  recherchés sont ceux qui minimisent la fonction de  $\chi^2$  que

l'on peut réécrire en notation matricielle :

$$\chi^2 = \mathbf{E}^T \mathbf{W} \mathbf{E} \quad (5)$$

où  $E^T$  représente la transposée de la matrice  $E=Y-\hat{Y}$  et où  $W$  est la matrice diagonale ayant pour éléments diagonaux les poids  $w_i$ , (l'inverse des variances). Il est à noter que  $W$  peut avoir des éléments non diagonaux lorsqu'il y a corrélation entre les points spectraux.

Une condition nécessaire pour que le  $\chi^2$  soit minimum est que sa dérivée par rapport à chaque paramètre soit nulle :

$$\left(\frac{\partial \chi^2}{\partial x}\right)=0 \Leftrightarrow \left(\frac{\partial \chi^2}{\partial x_j}\right)=0 \quad \text{pour } j \in [1, np], j \in \mathbf{N} \quad (6)$$

En utilisant la notation matricielle (5) et en utilisant l'expression de  $E$  déduite de l'équation (2), l'équation ci dessus peut s'écrire :

$$\frac{\partial}{\partial x} \left\{ (Y - Fx)^T W (Y - Fx) \right\} = 0 \quad (7)$$

En utilisant la théorie des matrices on montre que l'équation précédente est équivalente à :

$$(F^T W F) x = F^T W Y \quad (8)$$

L'équation (8), appelée équation normale, a pour solution :

$$x = (F^T W F)^{-1} F^T W Y \quad (9)$$

Dans la mesure où  $npt - np \gg 1$ , une estimation de l'erreur sur les paramètres (intervalle de confiance à 68%) est donnée par :

$$\left[ x_j - \sqrt{\left(\mathbf{F}^T \mathbf{W} \mathbf{F}\right)_{jj}^{-1} \chi^2}, x_j + \sqrt{\left(\mathbf{F}^T \mathbf{W} \mathbf{F}\right)_{jj}^{-1} \chi^2} \right] \quad (10)$$

où  $x_j$  est l'estimation déduite de (9) pour le  $j^{\text{ème}}$  paramètre recherché.

### 2.3.2 Moindre carré non-linéaire

Le modèle utilisé pour reproduire la mesure est dans la plupart des cas non linéaires, c'est-à-dire qu'il n'y a pas de relation matricielle  $Y = F X$ . Cependant dans certains cas particuliers il est possible de linéariser l'équation (1) à l'aide d'un changement de variable approprié :

$$y_i = f_{ij}(x_j) + e_i \xrightarrow{g} y_i' = f_{ij}' x_j + e_i' \quad (11)$$

avec :

$$y_i' = g(y_i) \quad (12)$$

$$f_{ij}' x_j + e_i' = g(f_{ij}(x_j) + e_i) \quad (13)$$

Cette linéarisation nécessite de modifier les poids. En effet, en posant  $y_i = g^{-1}(y_i')$  et  $\hat{y}_i = g^{-1}(\hat{y}_i')$  on peut au voisinage de la solution écrire :

$$y_i - \hat{y}_i = (y_i' - \hat{y}_i') \frac{\partial g^{-1}}{\partial y_i'} \quad (14)$$

On doit donc minimiser :

$$\sum_{i=1}^{npt} (y_i - \hat{y}_i)^2 w_i = \sum_{i=1}^{npt} (y_i' - \hat{y}_i')^2 \left( \frac{\partial g^{-1}}{\partial y_i'} \right)^2 w_i \quad (15)$$

soit :

$$\sum_{i=1}^{npt} (y_i' - \hat{y}_i')^2 w_i' \quad (16)$$

où les nouveaux poids  $w_i'$  ont pour expression :

$$w_i' = \frac{w_i}{\left( \frac{\partial g}{\partial y_i} \right)^2} \quad (17)$$

puisque les dérivées partielles de la fonction de changement de variable et de son inverse sont reliées par :

$$\left( \frac{\partial g^{-1}}{\partial y_i'} \right) = \frac{1}{\left( \frac{\partial g}{\partial y_i} \right)} \quad (18)$$

On ramène alors la minimisation au cas linéaire puisque par hypothèse :

$$y_i' = f_{ij}' x_j + e_i' \quad (19)$$

### 2.3.3 Linéarisation du modèle direct

Lorsque la relation (1) liant les paramètres du modèle et les valeurs mesurées, est intrinsèquement non linéaire, on peut essayer de linéariser le modèle par un développement de Taylor au voisinage de la solution initiale  $X^0$ . On peut alors écrire :

$$y_i = f_i(x^0) + \sum_{j=1}^{np} (x_j - x_j^0) \left( \frac{\partial f_i}{\partial x_j} \right)_{x=x^0} + \sum_{j=1}^{np} \sum_{k=1}^{np} \frac{1}{2} (x_j - x_j^0) (x_k - x_k^0) \left( \frac{\partial^2 f_i}{\partial x_j \partial x_k} \right)_{x=x^0} + \dots \quad (20)$$

En ne retenant dans (20) que les termes linéaires en  $X$ , on obtient à la place de l'équation (2) :

$$Y = F(x^0) + K(x^0)(x - x^0) + E_2 \quad (21)$$

dans laquelle :

- $F(x^0)$  représente la valeur du modèle pour les paramètres initiaux  $X^0$ ,
- $K(x^0)$  représente la matrice dérivée (jacobienne) du modèle par rapport aux paramètres, au voisinage de la solution initiale  $x^0$  avec :

$$k_{i,j} = \frac{\partial f_i(x)}{\partial x_j} \quad (22)$$

- $E_2$  représente le vecteur écart qui peut avoir des propriétés statistiques différentes du vecteur  $E$  introduit en (2), à cause de la linéarisation effectuée.

L'équation (21) se réécrit alors sous la forme :

$$Y = F(x^0) + K(x^0)\Delta x + E_2 \quad \text{avec : } \Delta x = x - x^0 \quad (23)$$

On peut ensuite appliquer la méthode des moindres carrés linéaires sur  $\Delta x$ , ce qui conduit à l'équation normale suivante :

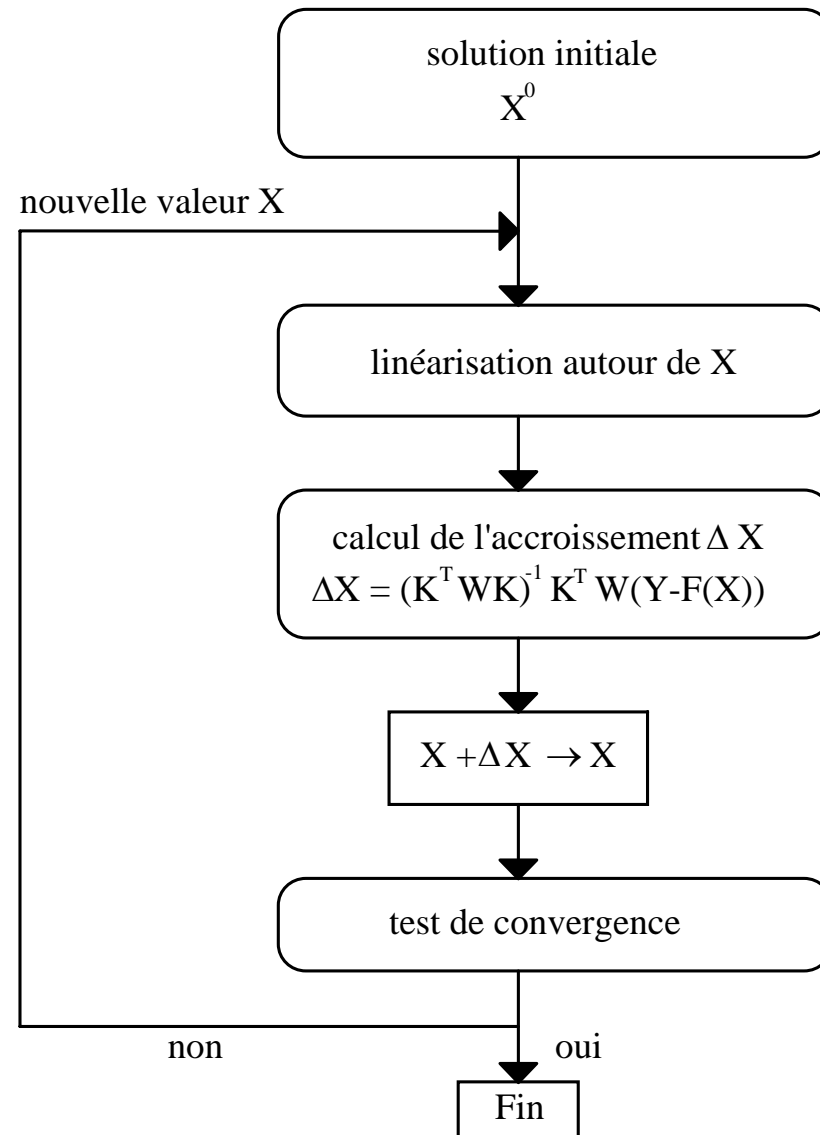
$$(K^T W K)\Delta x = K^T W (Y - F(x)) \quad (24)$$

c'est-à-dire :

$$\Delta x = (K^T W K)^{-1} K^T W (Y - F(x)) \quad (25)$$



Dans la pratique, l'algorithme d'inversion est itératif comme dans l'organigramme représenté sur la figure suivante.



Organigramme de l'algorithme d'inversion de la méthode des moindres carrés avec amortissement.

### 2.3.4 Moindre carré avec amortissement

Dans la mesure où le développement de Taylor limité au premier ordre (20) n'est pas strictement exact, la valeur de l'accroissement des paramètres  $\Delta x$  obtenue par la résolution de l'équation (24) peut s'avérer non réaliste et être en particulier très sensible au bruit de mesure (incertitude sur les observations  $Y$ ). Dans ces conditions, il peut être intéressant au cours du processus itératif de limiter les variations des paramètres. On parle alors de moindres carrés avec amortissement, ce qu'on réalise en remplaçant la fonction du  $\chi^2$  par la nouvelle fonction  $Z$  (fonction coût) définie par :

$$Z = E^T W E + \lambda \Delta x^T S_a^{-1} \Delta x \quad (26)$$

avec :

$$E = Y - \hat{Y} = Y - F(x) \quad (27)$$

où  $\lambda$  est le paramètre d'amortissement et où  $s_a$  est une estimation *a priori* de la matrice variance-covariance des paramètres (supposée diagonale ici). On pénalise ainsi au cours du processus itératif les variations trop importantes des paramètres par rapport à leurs erreurs telles qu'elles peuvent être estimées *a priori*. En effet, les éléments diagonaux de la matrice variance-covariance donnent une estimation statistique de l'erreur sur la détermination des paramètres. L'introduction d'un amortissement implique une contrainte sur la variation de ces paramètres lors de l'ajustement dans un domaine limité approximativement par la valeur de leurs erreurs.

Après linéarisation, la minimisation de la nouvelle fonction coût  $Z$  définie en (26) conduit à un nouvel accroissement calculable par :

$$\Delta x = \left( K^T W K + \lambda S_a^{-1} \right)^{-1} K^T W (Y - F(x)) \quad (28)$$

On retrouve sous cette forme l'expression utilisée par Rodgers.

Deux écueils sont à éviter :

- si  $\lambda$  est trop grand (fort amortissement), la solution aura tendance à rester confinée au voisinage de la solution initiale  $x^0$ , même si ce n'est pas la solution optimale;
- si  $\lambda$  est trop faible, on est quasiment ramené au cas des moindres carrés sans amortissement (c'est le cas si on fait  $\lambda = 0$  dans l'équation (26)).

En fait le paramètre  $\lambda$  peut être modifié au cours du processus d'itération et en principe tend vers zéro lorsque la méthode converge. En pratique, lorsqu'on a obtenu une solution qui semble satisfaisante, on repart de cette solution comme valeur initiale avec un amortissement nul et on vérifie que la nouvelle solution ne diffère pas de la solution initiale (écarts inférieurs aux erreurs sur les paramètres). Cela revient à dire que le choix judicieux du paramètre  $\lambda$  permet de plus ou moins accélérer la convergence, mais n'affecte pas en principe la solution trouvée.

### 2.3.5 Méthode de Levenberg-Marquardt

Nous abordons dans ce paragraphe une méthode qui ne cherche pas à se ramener à une forme linéaire de la relation entre les grandeurs mesurées et les paramètres que l'on cherche à déterminer, mais qui consiste en une procédure permettant d'améliorer à chaque étape d'un processus itératif, les valeurs initiales d'un ensemble de paramètres donnés, et ceci en minimisant la fonction de mérite  $\chi^2$  définie précédemment (équation (3) et (4)). Pour la rédaction de cette partie, nous nous sommes largement inspirés de *Numerical Recipes in Fortran*<sup>2</sup>

Supposons que la fonction  $\chi^2$  (expression (3) et (5)) puisse être approximée par une forme quadratique :

$$\chi^2(\mathbf{x}) = \chi^2(\mathbf{x}^0) + (\mathbf{x} - \mathbf{x}^0)^T \mathbf{B} + \frac{1}{2} (\mathbf{x} - \mathbf{x}^0)^T \mathbf{D} (\mathbf{x} - \mathbf{x}^0) \quad (29)$$

où  $\mathbf{B}$  est le vecteur gradient et  $\mathbf{D}$  la matrice Hessienne dont les éléments sont respectivement définis par :

$$\mathbf{b}_j = \frac{\partial \chi^2(\mathbf{x}^0)}{\partial x_j} = -2 \sum_{i=1}^{\text{npt}} w_i (y_i - \hat{y}_i) \frac{\partial \hat{y}_i}{\partial x_j} \quad j=1, \text{np} \quad (30)$$

$$\mathbf{d}_{j,k} = \frac{\partial^2 \chi^2(\mathbf{x}^0)}{\partial x_j \partial x_k} = 2 \sum_{i=1}^{\text{npt}} w_i \left( \frac{\partial \hat{y}_i}{\partial x_j} \frac{\partial \hat{y}_i}{\partial x_k} - (y_i - \hat{y}_i) \frac{\partial^2 \hat{y}_i}{\partial x_j \partial x_k} \right) \quad j, k=1, \text{np} \quad (31)$$

<sup>2</sup> Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., Numerical Recipes in fortran (second edition), Cambridge University Press, Cambridge, (1992),

Afin de se ramener à une écriture matricielle plus répandue du problème, on définit la matrice  $K$  et la matrice  $L$  de la manière suivante :

$$K_{i,j} = \frac{\partial \hat{y}_i}{\partial x_j} \quad j=1,np \quad i=1,npt \quad (32)$$

$$L_{i,jk} = \frac{\partial^2 \hat{y}_i}{\partial x_j \partial x_k} \quad j,k=1,np \quad i=1,npt \quad (33)$$

On peut alors exprimer le vecteur  $B$  et la matrice  $D$  de la manière suivante :

$$B = -2K^T W E \quad (34)$$

$$D = 2(K^T W K + L^T W E) \quad (35)$$

Si  $\Delta x = x - x^0$ , on peut réécrire l'équation (29) sous la forme :

$$\chi^2(x) = \chi^2(x^0) - 2\Delta x^T K^T W E + \Delta x^T (K^T W K + L^T W E) \Delta x \quad (36)$$

En dérivant l'équation (36) par rapport à  $x$ , on obtient :

$$\nabla \chi^2(x) = -2K^T W E + 2(K^T W K + L^T W E) \Delta x \quad (37)$$

Si  $x$  minimise  $\chi^2(x)$  alors  $\nabla\chi^2(x) = 0$  et d'après ci-dessus on a :

$$\Delta x = \left( \mathbf{K}^T \mathbf{W} \mathbf{K} + \mathbf{L}^T \mathbf{W} \mathbf{E} \right)^{-1} \left( \mathbf{K}^T \mathbf{W} \mathbf{E} \right) \quad (38)$$

Cet ensemble d'équations est résolu pour l'incrément  $\Delta x$  qui nous donne les nouvelles valeurs des paramètres pour l'itération suivante (si celles-ci améliorent effectivement le  $\chi^2$ ) :

$$x \rightarrow x^0 + \Delta x \quad (39)$$

L'utilisation de l'équation (38) implique que l'on soit capable de calculer le gradient et la Hessienne du  $\chi^2$ .

Ceci ne pose pas de problème en principe puisqu'on sait écrire le modèle direct.

En effet, le  $\chi^2$  donné par l'expression (3) ne dépend que des valeurs de la grandeur mesurée, de leur estimation par le modèle direct ( $\hat{Y} = F(x)$ ) et de l'incertitude entre valeur mesurée et estimation du modèle. Lorsqu'on dérive par rapport aux paramètres recherchés, il ne reste que le terme dépendant du modèle et le calcul des dérivées est donc possible.

On peut constater que les éléments des matrices  $K$  et  $L$  permettant le calcul de la matrice Hessienne (expression (35)) dépendent des dérivées premières et des dérivées secondes du  $\chi^2$  (expression (32) et (33)) par rapport aux paramètres. Certains traitements classiques ignorent les dérivées secondes sans toujours justifier cette approximation. Nous nous placerons également dans ce cas après quelques explications.

Le terme dérivée seconde peut être écarté lorsqu'il est nul (cas de dépendance linéaire) ou suffisamment petit pour être négligé devant les termes faisant intervenir les dérivées premières. Or le terme multipliant la dérivée seconde est  $(y_i - \hat{y}_i)$ . Pour un modèle bien adapté, ce terme doit être proche de l'erreur aléatoire sur la mesure en chaque point (de signe quelconque) et doit généralement ne pas être corrélé avec le modèle. Par conséquent, les termes dérivées secondes tendent à s'annuler lorsqu'on somme sur les points de mesures. En outre, il faut remarquer que la présence des termes dérivées secondes peut être déstabilisante si le modèle reproduit mal les points de mesure ou s'il est contaminé par des points aberrants (« outliers ») qui seront difficilement « compensés » par des points avec des écarts de signes opposés.

A la lumière des arguments nous utiliserons donc comme définition des éléments de la matrice Hessienne :

$$D = 2K^T W K \quad (40)$$

et nous ré-écrivons l'expression (38) sous la forme :

$$\Delta x = (K^T W K)^{-1} (K^T W E) \quad (41)$$

Si maintenant l'expression (36) est une mauvaise approximation locale de la fonction que l'on cherche à minimiser, la *méthode du gradient de plus forte pente à pas prédéterminé* (*steepest descent method*) consiste à estimer une nouvelle valeur des paramètres recherchés à partir du vecteur gradient multiplié par un coefficient d'amortissement  $\mu$  positif et

suffisamment petit pour que cette estimation ne donne pas une valeur de  $x$  trop éloignée de la solution de départ  $x^0$  soit :

$$\Delta x = -\mu B = -\mu 2K^T W E \quad (42)$$

et on détermine de cette façon une nouvelle valeur des paramètres :  $x \rightarrow x^0 + \Delta x$

On considère deux méthodes correspondant aux cas où le modèle représente bien et mal les mesures lorsqu'on utilise un développement à l'ordre 2 de la fonction à minimiser. La méthode de Levenberg-Marquardt permet de réunir ces deux méthodes en une méthode unique.

La méthode du gradient de plus forte pente à pas prédéterminé (42) est utilisée loin du minimum, laissant graduellement la place à la méthode de la Hessienne inverse (41) lorsqu'on s'approche du minimum. Cette méthode est bien sûr itérative et permet de déterminer une solution de proche en proche en modifiant à chaque étape l'importance relative des deux processus permettant la diminution du  $\chi^2$ .

Le  $\chi^2$  est sans dimension alors que les éléments de  $B$  ont la dimension de  $1/x_j$  (chaque élément de  $B$  peut être de dimension différente). La constante multiplicative  $\mu$  dans l'équation (42) doit par conséquent avoir la dimension de  $x_j^2$ .

Ainsi l'inverse des éléments diagonaux de la matrice Hessienne  $1/(K^T W K)_{j,j}$  (ayant la dimension de  $x_j^2$ ) nous permettent d'obtenir une estimation des « bornes » de la constante  $\mu$ , mais cet intervalle pouvant être non réaliste, on le



divise par un facteur  $\lambda$  sans dimension avec la possibilité de prendre ce facteur très inférieur à 1 ce qui permet d'augmenter l'intervalle dans lequel se situe la nouvelle solution. On remplace donc l'équation (42) en intégrant le facteur 2 dans le coefficient  $\lambda$  par :

$$\Delta x_j = \frac{1}{\lambda \left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{j,j}} \left( \mathbf{K}^T \mathbf{W} \mathbf{E} \right)_j \Rightarrow \lambda \left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{j,j} \Delta x_j = \left( \mathbf{K}^T \mathbf{W} \mathbf{E} \right)_j \quad (43)$$

et en notation matricielle, on peut écrire :

$$\gamma \mathbf{I} \Delta \mathbf{x} = \mathbf{K}^T \mathbf{W} \mathbf{E} \quad (44)$$

où  $\mathbf{I}$  est la matrice unité et  $\gamma$  la matrice dont les éléments sont donnés par :

$$\left( \gamma \right)_{j,j} = \lambda \left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{j,j} \quad j=1, np \quad (45)$$

$$\left( \gamma \right)_{j,k} = 0 \quad j,k=1, np ; j \neq k$$

La relation (43) ou (44) n'est vraie que si  $1/\left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{j,j}$  est positif pour tout  $k$ , ce qui est vrai vu sa définition (équation (31)) pour  $j = k$  et en négligeant les dérivées secondes).

La méthode de Levenberg-Marquardt réunit les équations (44) et (41) permettant de déterminer un nouvel incrément  $\Delta \mathbf{x}$  :

$$\Delta \mathbf{x} = \left( \mathbf{K}^T \mathbf{W} \mathbf{K} + \gamma \mathbf{I} \right)^{-1} \left( \mathbf{K}^T \mathbf{W} \mathbf{E} \right) \quad (46)$$

ce qui revient à définir une nouvelle matrice Hessienne  $D' = \mathbf{K}^T \mathbf{W} \mathbf{K} + \gamma \mathbf{I}$

Quand  $\lambda$  est très grand, la matrice  $D'$  est « forcée » par ses éléments diagonaux et l'équation (36) tend à être identique à l'équation (44).

Au contraire, lorsque  $\lambda$  est proche de zéro, l'équation (36) tend à être identique à l'équation (41).

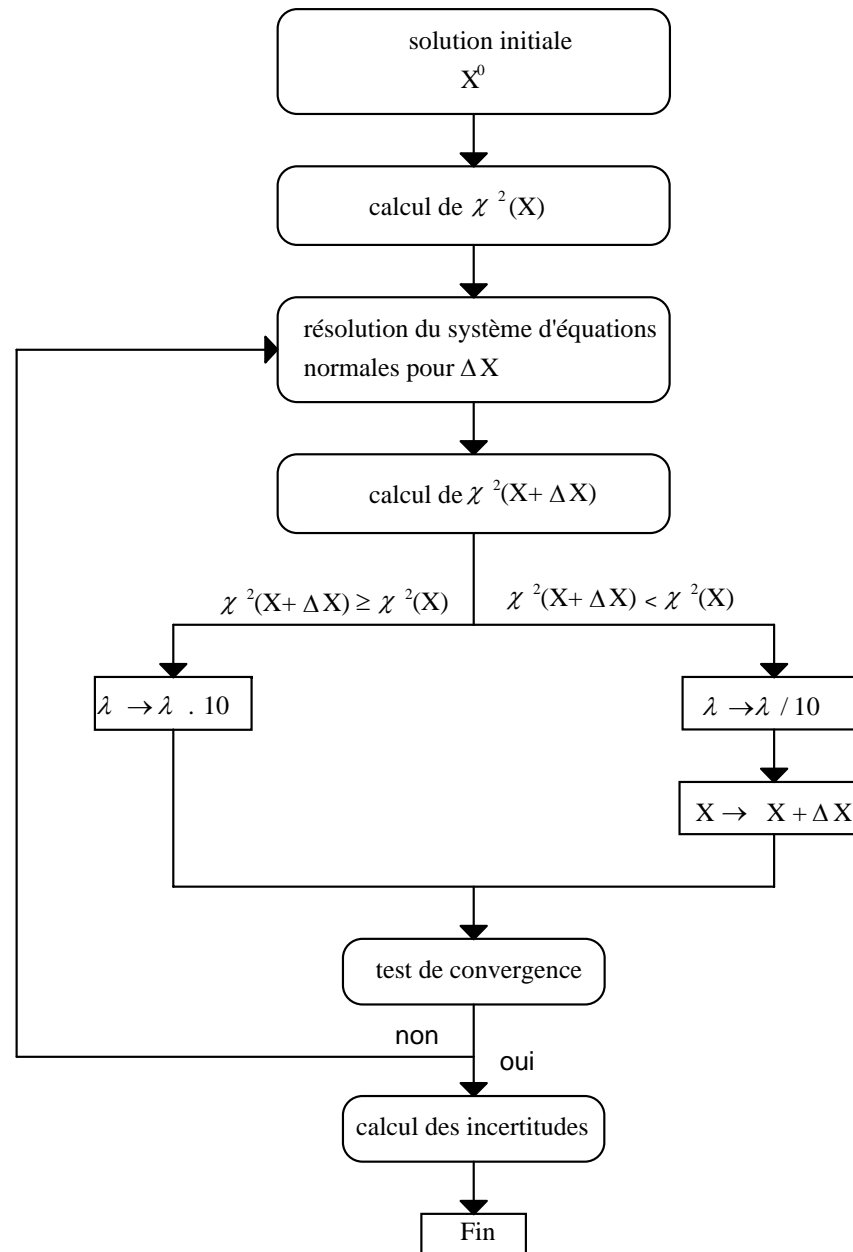
Un ensemble de données de départ étant fixé et pour un choix de paramètres  $\mathbf{X}$  à inverser, le processus d'inversion est itératif.

Il est nécessaire de fixer un critère pour stopper l'itération car en pratique on n'atteint que rarement l'égalité  $\chi^2=1$ . En fait, le minimum recherché ne correspond qu'à la meilleure estimation statistique des paramètres  $x$  et un changement des paramètres qui change le  $\chi^2$  d'une quantité très inférieure à 1 est rarement significatif.

Lorsqu'un minimum acceptable a été trouvé, on recalcule la matrice Hessienne  $D'$  pour  $\lambda = 0$ , matrice qui, une fois inversée, donne une estimation de la matrice variance-covariance des erreurs sur la détermination des paramètres inversés (les éléments diagonaux de cette matrice correspondant aux variances). On peut alors obtenir un intervalle de confiance (seuil de confiance de 68 %) pour chaque paramètre inversé:

$$\left[ x_j - \sqrt{\chi_r^2 \left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{jj}^{-1}}, x_j + \sqrt{\chi_r^2 \left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{jj}^{-1}} \right] \quad (47)$$

où  $\chi_r^2$  est le  $\chi^2$  réduit (expression (4)) et où  $\left( \mathbf{K}^T \mathbf{W} \mathbf{K} \right)_{jj}^{-1}$  est le  $j^{\text{ème}}$  élément diagonal de la matrice Hessienne inverse.



Organigramme de l'algorithme d'inversion de Levenberg-Marquardt.

### 2.3.6 Information *a priori*

Jusqu'alors, dans tout le processus de calcul, **aucune contrainte sur la réalité physique** des solutions n'est prise en compte durant l'inversion. Il se peut donc que la solution déterminée par l'algorithme de Levenberg-Marquardt corresponde effectivement à un **minimum de la fonction de coût mais n'ait aucune réalité physique**. On introduit donc une contrainte supplémentaire pour prendre en compte la cohérence des solutions trouvées, via une information *a priori* sur les paramètres recherchés.

**L'information *a priori*** peut, par exemple, être le profil vertical de concentration d'une espèce chimique qu'on souhaite mesurer. En général, elle correspond à **l'état de l'atmosphère, connu, avant la mesure**. Le profil *a priori* peut être utilisé pour plus de simplicité comme profil de départ (*first guess*) mais ce n'est pas une obligation.

Cette contrainte supplémentaire est rajoutée à la fonction de coût définie précédemment ( $\chi^2$  ou  $Z$ ) mais  $\Delta X = x - x_0$  devient  $x - x_a$  où  $x_a$  est le vecteur des **valeurs *a priori*** des paramètres recherchés (avant la mesure), c'est aussi une solution de départ qui a un sens physique.

On peut aussi introduire l'inverse de la matrice erreur,  $S_e^{-1}$  dont les valeurs correspondent aux éléments diagonaux  $w_i$ . On a ainsi :

$$Z = E^T S_e^{-1} E + \lambda \left( x - x_a \right)^T S_a^{-1} \left( x - x_a \right) \quad (48)$$

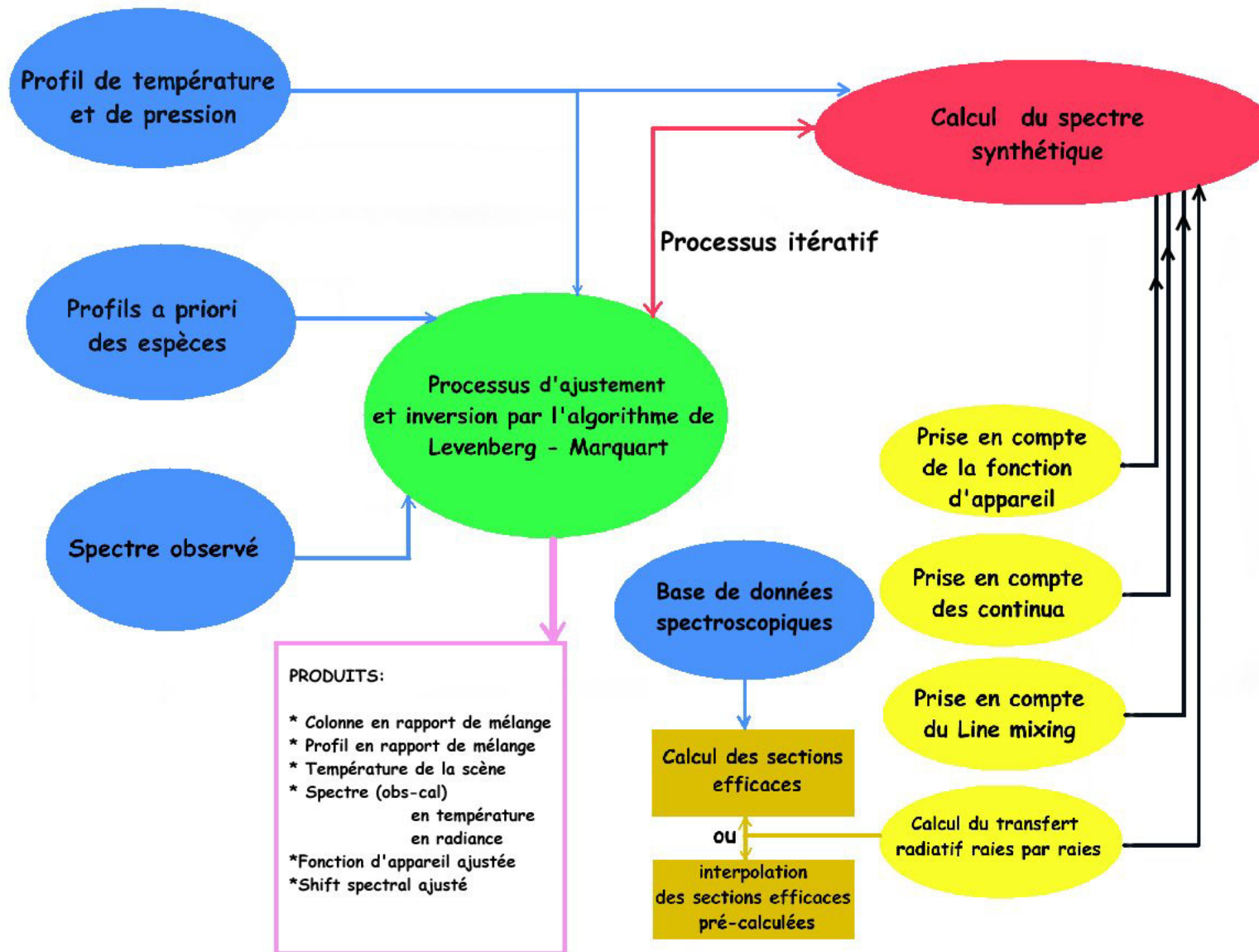
$S_a$  est une estimation *a priori* de la **matrice variance-covariance des paramètres**. On pénalise ainsi au cours du processus itératif les variations trop importantes des paramètres par rapport à la solution *a priori* en tenant compte des « fluctuations » acceptables autour de cette solution en fonction des erreurs également estimées *a priori*, les éléments diagonaux de la matrice variance-covariance donnant une estimation statistique de l'erreur sur la détermination *a priori* des paramètres.

Combiné avec la méthode du gradient de plus forte pente à pas déterminé (comme dans la méthode de Levenberg-Marquardt décrite avant) l'incrément d'itération a pour nouvelle expression:

$$\Delta x = \left( \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1} + \gamma \mathbf{I} \right)^{-1} \left( \left( \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{E} \right) - \mathbf{S}_a^{-1} (x - x_a) \right) \quad (49)$$

On obtient ainsi une nouvelle expression de intervalle de confiance (seuil de confiance de 68 %) pour chaque paramètre inversé qui tient compte de l'information *a priori* et qui contraint la solution dans un domaine de validité physique.

$$\left[ x_j - \sqrt{\chi_r^2 \left( \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1} \right)_{jj}^{-1}}, x_j + \sqrt{\chi_r^2 \left( \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1} \right)_{jj}^{-1}} \right] \quad (50)$$



3

<sup>3</sup> Victor Ferreira, «Inversion de spectres infrarouges en émission thermique de la basse atmosphère terrestre enregistrés sous ballon par transformée de Fourier en visée au nadir : préparation de la mission spatiale IASI», thèse de l'Université Pierre et Marie Curie, 7 février 2005

### 3 Contenu en information

Prenons le cas de la mesure à distance d'un profil vertical de concentration. L'efficacité et la résolution (verticale) d'un sondage à distance peuvent être exprimées de deux manières différentes suivant qu'on considère uniquement la qualité du modèle direct (**fonction de poids**) ou qu'on tient compte de la « puissance du modèle d'inversion à résoudre les structures fines verticales de l'atmosphère » (**noyaux moyens**). Ces deux grandeurs permettent de représenter pour chaque spectre, la contribution à l'absorption des molécules de l'espèce étudiée suivant leur répartition en altitude.

#### 3.1 Fonctions de poids/Jacobiens

Pour faire apparaître les fonctions de poids il faut reprendre l'expression reliant le spectre observé au vecteur d'état et au modèle de calcul direct :

$$y = F(x, b) + \varepsilon \quad (51)$$

Dans cette expression  $x$  est le vecteur d'état de  $np$  paramètres,  $b$  représente les paramètres du modèle qui ne sont pas parfaitement connus de l'utilisateur comme par exemple les paramètres spectroscopiques, la fonction d'appareil, les données météorologiques de type profils de température et de pression. Le terme  $\varepsilon$  est l'erreur expérimentale.

Lors de l'inversion on cherche à déterminer le vecteur d'état qui permet au mieux de simuler le spectre observé au travers l'algorithme du transfert radiatif, ce qui revient à écrire le vecteur sous la forme :

$$\hat{x} = R[y, b, x_a] = R[F(x, b) + \varepsilon, b, x_a] \quad (52)$$

où

- $\hat{x}$  est le vecteur d'état restitué
- $R$  la fonction du modèle inverse (fonction de transfert)
- $x_a$  le vecteur d'état avec les valeurs *a priori* des paramètres recherchés.

En linéarisant le modèle direct au voisinage de  $x_a$  on a :

$$\hat{x} = R[F(x_a, b) + K_x(x - x_a) + \varepsilon, b, x_a] \quad (53)$$

On fait ainsi apparaître la matrice de fonction de poids ou jacobien  $K_x$  :

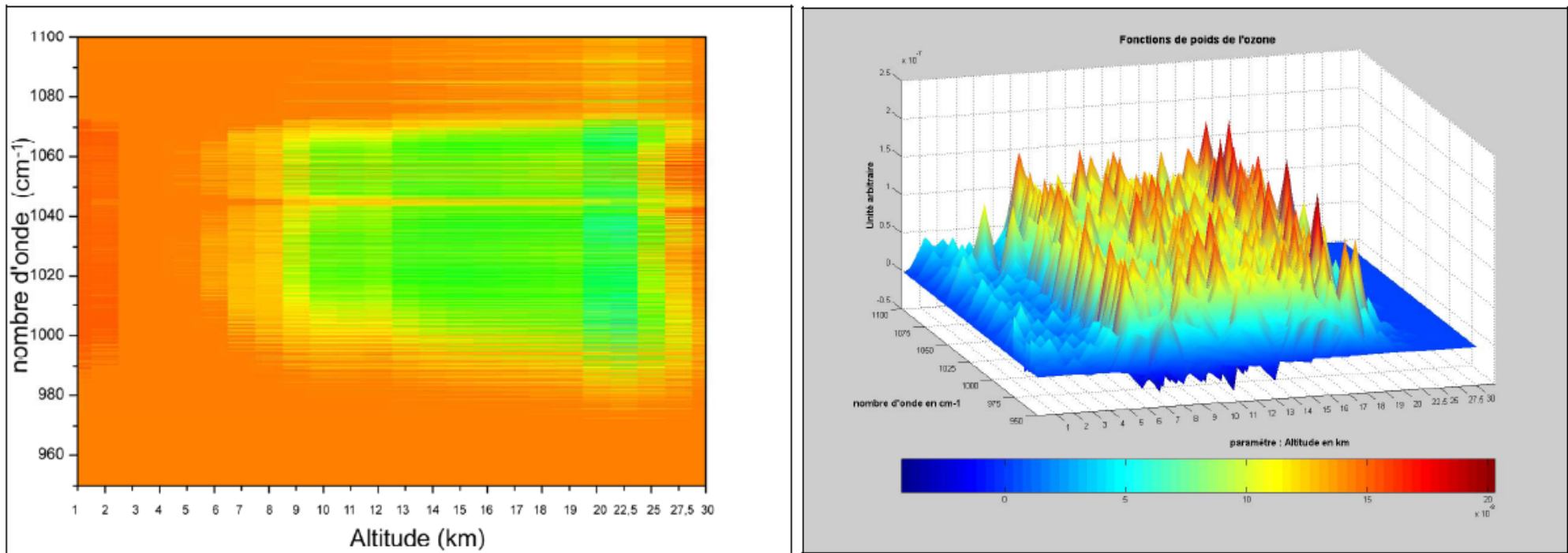
$$k_{i,j} = \frac{\partial f_i(x)}{\partial x_j} \quad (54)$$

C'est une matrice  $K$  ( $m$  points spectraux  $\times n$  paramètres inversés)

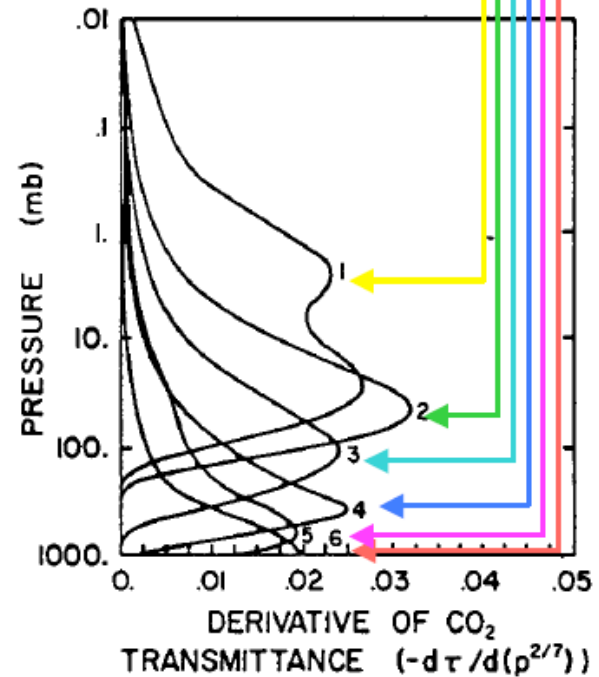
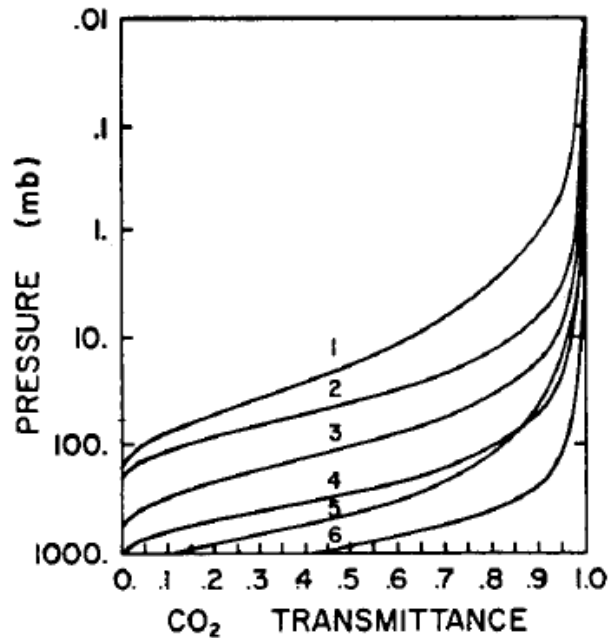
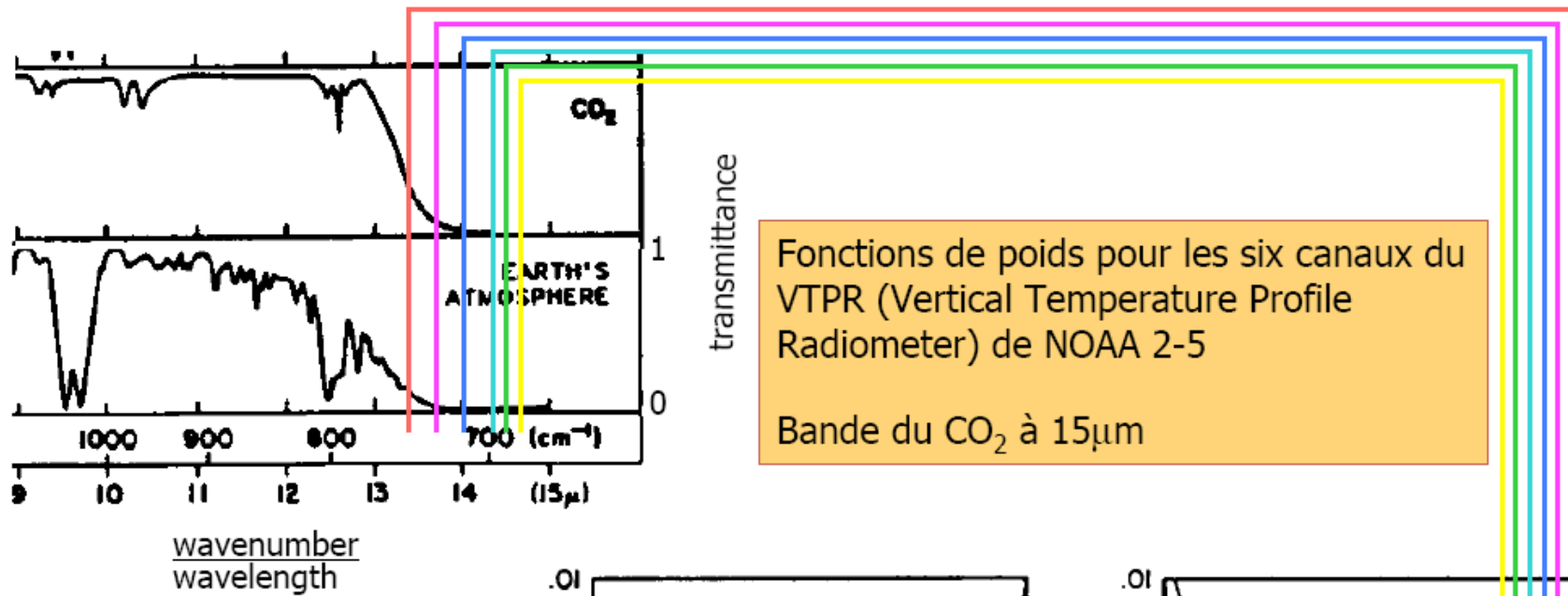


Les fonctions de poids représentent la sensibilité du modèle aux paramètres que l'on cherche à déterminer. Ces fonctions ont généralement la forme d'un pic en visée au limbe et sont plus élargies en visée à partir du sol ou au nadir.

La largeur à mi-hauteur des pics des fonctions de poids donne une information sur la résolution verticale de la mesure, ainsi que sur la sensibilité du spectre par rapport à chaque paramètre inversé (sensibilité en fonction de l'altitude par exemple).



Visualisation des fonctions de poids en 2 et 3 dimensions pour chaque altitude du profil d'ozone (expérience IASI ballon)



<sup>4</sup> [www.lmd.ens.fr/jpduvel/Cours](http://www.lmd.ens.fr/jpduvel/Cours)

### 3.2 Noyaux moyens

En linéarisant le modèle inverse R par rapport à y on obtient :

$$\hat{x} = R[F(x_a, b), b, x_a] + G_y [K_x (x - x_a) + \varepsilon, b, x_a] \quad (55)$$

où  $G_y = \frac{\partial R}{\partial y}$  est la sensibilité de l'inversion par rapport à la mesure. On peut réécrire l'expression précédente de la manière suivante :

$$\begin{aligned} \hat{x} - x_a &= R[F(x_a, b)] - x_a && \dots \text{biais} \\ &+ A(x - x_a) && \dots \text{lissage} \\ &+ G_y \varepsilon && \dots \text{erreur inversée} \end{aligned} \quad (56)$$

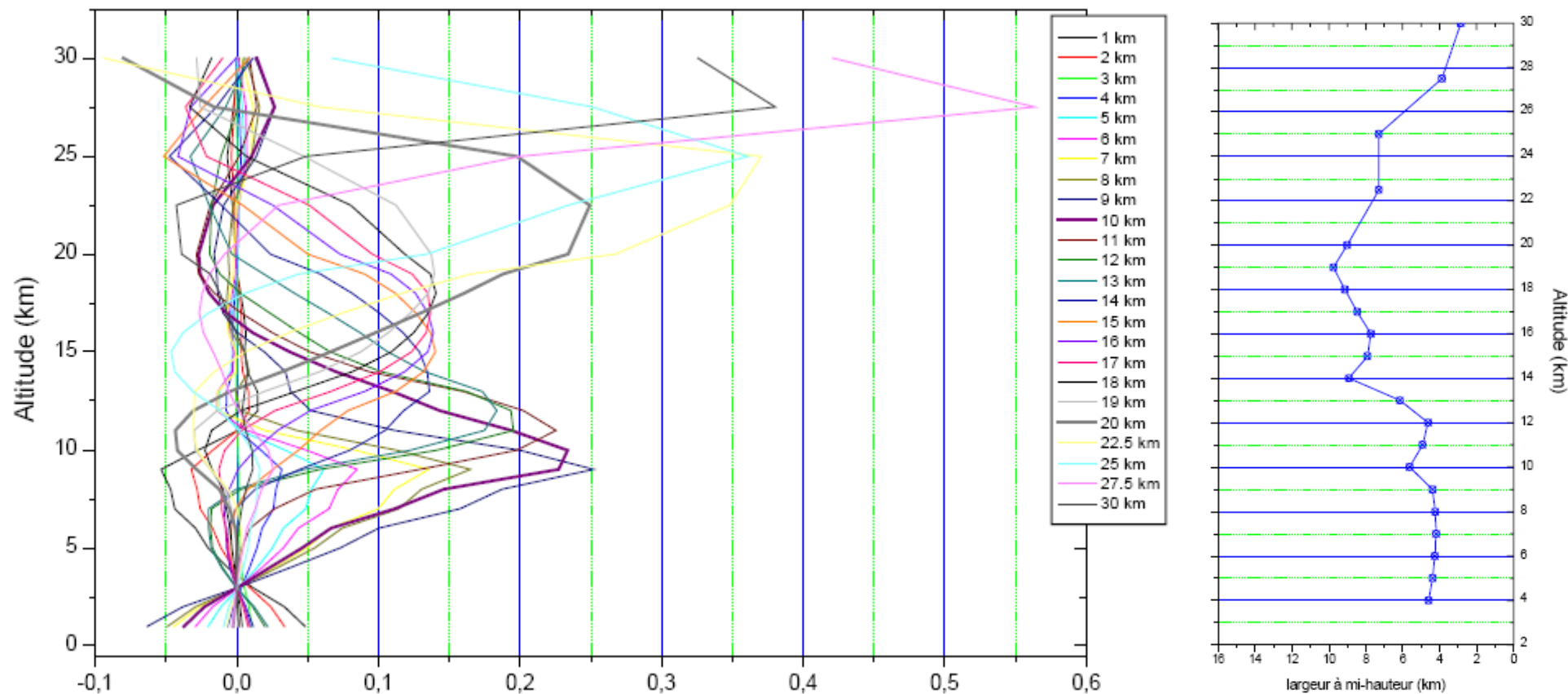
où

$$A = G_y K_x = \frac{\partial \hat{x}}{\partial X} \quad (57)$$

Les lignes de la matrice A ( $n \times n$ ) sont les noyaux moyens (*averaging kernel*) qui peuvent être considérés un par un comme les fonctions responsables du lissage (fonction de moyennage) des paramètres inversés correspondant.

**Chaque élément du résultat qui provient de l'inversion apparaît comme le produit des valeurs vraies ou recherchées par la fonction de moyennage correspondante.**

Dans des conditions favorables, les noyaux moyens sont des fonctions en « pics » dont la largeur à mi-hauteur fournit une estimation de la résolution verticale du système d'observation. Dans le cas idéal ces noyaux moyens sont des pics de Dirac. L'intégrale de la surface des noyaux moyens, généralement proche de l'unité, détermine la contribution aux paramètres restitués de la mesure.



Noyaux moyens pour l'inversion du profil d'ozone à partir de l'expérience IASI-ballon.

A l'aide des noyaux moyens et des fonctions de poids il est possible d'**établir la sensibilité de l'inversion** pour chaque paramètre inversé et donc **d'effectuer une sélection optimale des paramètres** en retenant en priorité ceux qui ont le plus de sensibilité sur le spectre et donc qui sont les plus exploitables par le modèle de transfert radiatif. On gagne ainsi en temps de calcul, en limitant le nombre de paramètres que l'on cherche à déterminer, mais aussi en qualité d'ajustement des spectres (et d'inversion). En ayant moins de paramètres à inverser tout en gardant une bonne souplesse dans l'ajustement on évite une trop grande corrélation entre paramètres, qui en général introduit des effets de compensation pouvant conduire à des solutions non physiques.

### **3.3 Entropie informationnelle / degrés de liberté**

Le nombre de degrés de liberté d'un signal peut par exemple être considéré comme une mesure de l'information. Rodgers reprend le concept de l'information selon la formulation de Shannon et l'applique aux problèmes d'inversion de sondages atmosphériques. A partir de l'entropie informationnelle et de la notion de nombre de degrés de liberté il est possible d'estimer l'information contenue dans un spectre pour une espèce chimique donnée.

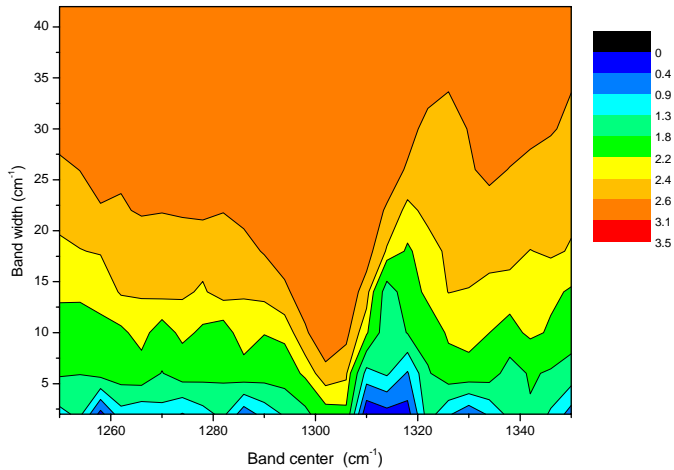
Pour plus de détail, vous pouvez consulter le livre de Clive Rodgers. Nous donnerons simplement ici l'expression du nombre de degrés de liberté :

$$d_s = \sum_i d_{s_i} = \sum_i \frac{\lambda_i^2}{1 + \lambda_i^2} = \text{tr}(A) \quad (58)$$

Où  $\lambda_i$  sont les valeurs singulières de la matrice  $\tilde{\mathbf{K}}$  définie par :

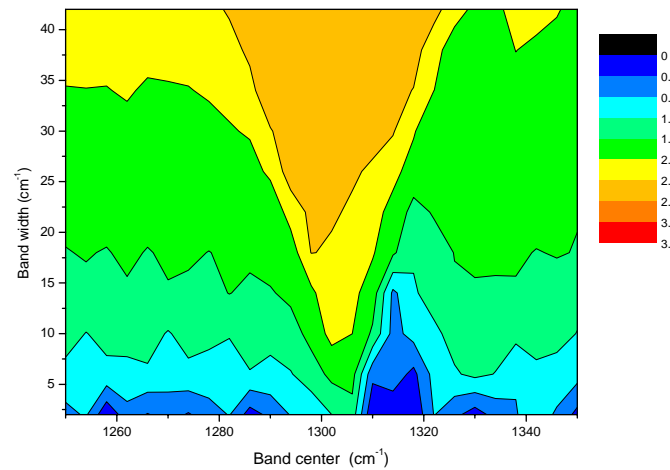
$$\tilde{\mathbf{K}} = \mathbf{S}_\varepsilon^{-1/2} \mathbf{K} \mathbf{S}_a^{1/2} \quad (59)$$

Exemple : DOFS pour l'inversion de  $\text{CH}_4$  par visée au nadir



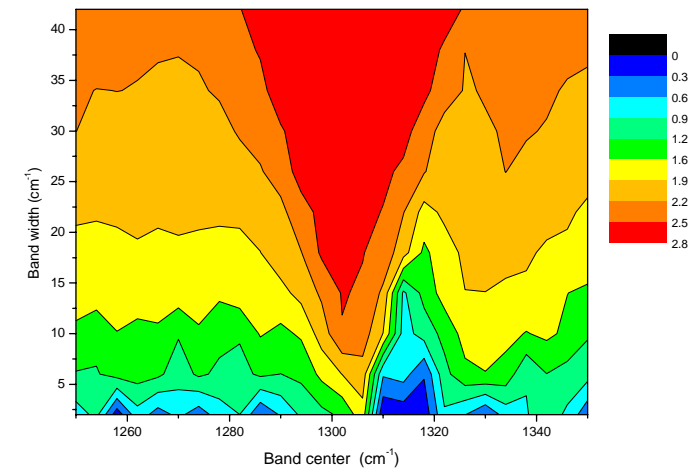
FWHM =  $0.125 \text{ cm}^{-1}$

NE $\Delta$ T = 0.1 K



FWHM =  $0.125 \text{ cm}^{-1}$

NE $\Delta$ T = 0.20 K



FWHM =  $0.250 \text{ cm}^{-1}$

NE $\Delta$ T = 0.10 K]

## 4 Analyse et caractérisation des erreurs

Dans l'approximation linéaire de la méthode inverse présentée précédemment, on peut écrire l'erreur totale comme la somme de quatre termes

### 4.1 *Les différentes sources d'erreurs*

Reprenons la fonction de transfert. Le vecteur d'état inversé est une fonction de la forme :

$$\hat{\mathbf{x}} = \mathbf{R}(\mathbf{y}, \hat{\mathbf{b}}, \mathbf{c})$$

Le lien avec l'état réel  $\mathbf{X}$  est le suivant :

$$\hat{\mathbf{x}} = \mathbf{R}(\mathbf{f}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\epsilon}, \hat{\mathbf{b}}, \mathbf{c})$$

Que nous pouvons représenter de la manière suivante :

$$\hat{\mathbf{x}} = \mathbf{T}(\mathbf{x}, \mathbf{b}, \boldsymbol{\epsilon}, \hat{\mathbf{b}}, \mathbf{c})$$

En linéarisant la fonction de transfert, Clive Rodgers montre qu'on peut écrire l'erreur de l'inversion sous la forme suivante :

$$\begin{aligned} \hat{\mathbf{x}} - \mathbf{x} = & (\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}_a) && \textit{smoothing} \\ & + \mathbf{G}_y \mathbf{K}_b (\mathbf{b} - \hat{\mathbf{b}}) && \textit{model parameters} \\ & + \mathbf{G}_y \Delta \mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{b}') && \textit{modelling error} \\ & + \mathbf{G}_y \epsilon && \textit{measurement noise} \end{aligned}$$

$$\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} = \mathbf{A} \quad \text{Sensibilité à l'état réel : noyaux moyens.}$$

$$\frac{\partial \hat{\mathbf{x}}}{\partial \epsilon} = \mathbf{G}_y \quad \text{Sensibilité au bruit de mesure}$$

$$\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{b}} = \mathbf{G}_b \quad \text{Sensibilité aux paramètres non inversés}$$

$$\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{c}} = \mathbf{G}_c \quad \text{Sensibilité aux paramètres de la méthode inverse}$$

The sum of the first two terms (smoothing and measurement errors) will be referred to as the *internal* error, which is inherent to the observing geometry and instrumental capabilities, whereas the sum of the last two terms, which are related to the knowledge of the atmosphere, constitute the *external* error.

Certains de ces termes sont faciles à estimer, et d'autre pas.



## 4.2 *Bruit de mesure*

L'erreur de mesure est due au bruit instrumental. Sa matrice covariance est donnée par:

$$\mathbf{S}_n = \mathbf{G}_y \mathbf{S}_\epsilon \mathbf{G}_y^T$$

## 4.3 *Erreur de lissage*

Cette erreur considère le lissage du profil vrai (ou du vecteur d'état vrai) par les noyaux moyens. La matrice covariance de l'erreur de lissage est donnée par :

$$\mathbf{S}_s = (\mathbf{A} - \mathbf{I}) \mathbf{S}_a (\mathbf{A} - \mathbf{I})^T$$

Pour bien déterminer cette erreur il est donc nécessaire de bien connaître la covariance de la climatologie des paramètres recherchés.

## 4.4 *Paramètres du modèle direct*

C'est l'erreur sur les paramètres non inversés du modèle direct. Sa covariance s'écrit :

$$\mathbf{S}_p = \mathbf{GK}_b \mathbf{S}_b (\mathbf{GK}_b)^T$$

où  $\mathbf{S}_b$  est la matrice covariance représentant les incertitudes sur les paramètres fixés du modèle direct.

## 4.5 *Modèle direct*

$$\text{modelling error} = \mathbf{G}_y \Delta \mathbf{f} = \mathbf{G}_y (\mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{b}') - \mathbf{F}(\mathbf{x}, \mathbf{b}))$$

Souvent très difficile à déterminer car il faut un modèle  $f$  permettant de quantifier les écarts entre le modèle direct et la réalité physique ...

## 4.6 *Matrice covariance erreur*

The total error covariance matrix is then given by:

$$\mathbf{S}_T = (\mathbf{S}_s + \mathbf{S}_m) + (\mathbf{S}_p + \mathbf{S}_{cs})$$

## 5 Les autres méthodes

Réseau de neurone

Méthodes statistiques

## Pour aller plus loin :

C.D. Rodgers, Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore (2000).

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., Numerical Recipes in fortran (second edition)., Cambridge University Press, Cambridge, (1992) .

Physique moléculaire Physique de l'atmosphère - Molecular Physics, Atmospheric Physics, Editions du CNRS 1982, Montfoulon (près d'Alençon), 1-10 décembre 1980, C. Camy-Peyret editor, isbn 1-222-03104-4

