Ionospheric tomography by Over-The-Horizon radar

C. Roy

THÈSE

afin d'obtenir le titre de

Docteur de l'Institut de Physique du Globe de Paris,
Sorbonne Paris Cité; Ecole doctorale des Sciences de la Terre.

Spécialité: Géophysique

soutenue par

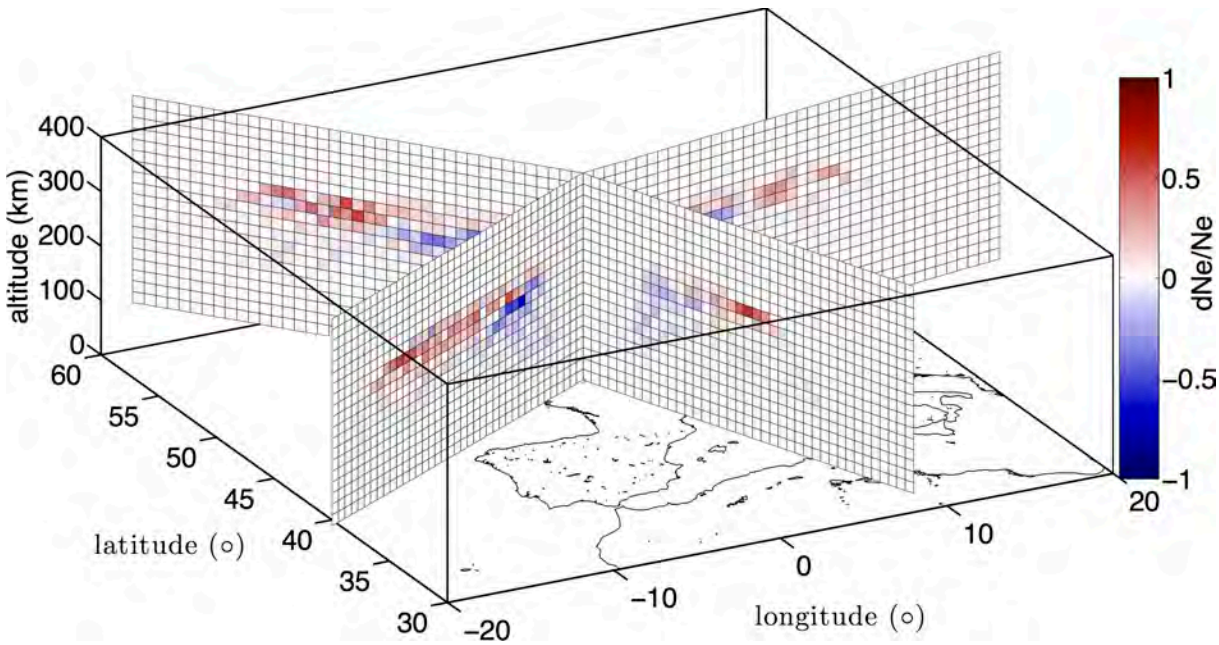
Corinna Roy

et préparée au sein de :

L'équipe de Planétologie et Sciences Spatiales et l'ONERA

sur le sujet :

Ionospheric tomography by Over-The-Horizon radar



le 27 Novembre 2014 devant le jury composé de :

Monsieur	Gauthier Hulot	*Président du jury*
Monsieur	Raphaël F. Garcia	*Rapporteur*
Monsieur	Attila Komjathy	*Rapporteur*
Monsieur	Norbert Jakowski	*Examinateur*
Monsieur	Mark Wieczorek	*Directeur de thèse*
Monsieur	Giovanni Occhipinti	*Co-encadrant de thèse*
Monsieur	Jean-Philippe Molinié	*Invité*

On the front page: Electron density perturbation over Europe obtained by inversion of real data measured by the OTH radar Nostradamus.

Contents

Chapter 1

Résumé étendu en français

1.1 Introduction

Bien que l'ionosphère représente moins de 1% de la masse de l'atmosphère au-dessus de 100 km, son importance résulte de son influence sur les ondes électromagnétiques (EM) à haute fréquence (HF). Tous les signaux des satellites qui traversent l'ionosphère (télécommunications, altimétrie, radar, GNSS, etc.) sont affectés par sa présence. Par conséquent, la connaissance exacte de la structure de l'ionosphère reste un sujet important en Science et pour les applications militaires. La connaissance fine de l'ionosphère joue un rôle clef pour comprendre les processus tels que, par exemple, l'évolution spatiale et temporelle du plasma et pour explorer la propagation d'ondes comme les vagues planétaires, les marées et les perturbations ionosphériques itinérantes (TID), et leur influence dans les mesures des satellites scientifiques, des satellites de télécommunication, ou des satellites de service.

Dans le domaine de la défense où les cibles détectées par le radar doivent être localisées précisément, l'effet de l'ionosphère sur les mesures doit être connu. Pour les applications de grande précision (du centimètre au millimètre) et les applications en temps réel, la correction de ces erreurs est un défi important. Réduire les effets de l'ionosphère permettra d'améliorer les précisions pour différentes applications : navigation par satellites, orbitographie, altimétrie satellitaire, télécommunications…

Néanmoins, les effets de l'ionosphère, s'ils ne sont pas corrigés, peuvent réduire la précision de positionnement, d'estimation de vitesse, et diverses informations estimées au cours du temps, qui sont calculées à partir des signaux reçus. Les erreurs de positionnement peuvent varier de quelques mètres à plusieurs dizaines de mètres.

La précision dans le positionnement par Global Positioning System (GPS) et dans la navigation est profondément affectée par l'ionosphère, car les signaux émis par les satellites sont modifiés par leur chemin dans la densité de son plasma. Le

retard causé par l'ionosphère est devenu la plus grande source d'erreur dans le positionnement par GPS et dans la navigation.

1.2 Ionosphère

L'ionosphère a été découverte par Appleton en 1924: un signal continu émis avec un émetteur de la BBC (British Broadcasting Corporation) à Bournemouth près d'Oxford a été reçu à Cambridge. En mesurant la différence en temps entre l'onde directe, se propageant au sol et l'onde réfléchie dans l'atmosphère, il a pu estimer la hauteur de la couche réfléchissante à 100 km. Appleton l'appelle couche E, et pour cette découverte gagne le prix Nobel de physique en 1947. Dans les années suivantes, avec le développement des radars, l'atmosphère a été explorée de plus en plus. Cette exploration a mené à la découverte d'autres couches et finalement à la définition de l'ionosphère. Le nom ionosphère a été proposé la première fois par le physicien écossais Robert Watson-Watt.

1.2.1 Les caractéristiques de l'ionosphère

L'ionosphère est définie par la partie supérieure de l'atmosphère entre 60 et 1000 km. Elle est constituée de molécules neutres, d'ions et d'électrons libres formant le plasma ionosphérique. Les molécules principales existantes sont l'azote (N2), l'oxygène (O2), et l'argon, qui forment 99% de l'air. Les ions et électrons sont générés en continu soit par photo-ionisation, soit par l'impact d'atomes et de molécules neutres, ou indirectement par des réactions chimiques. Cet effet s'équilibre avec la recombinaison des ions avec des électrons, menant à la production des photons ou la décomposition d'une molécule dans des atomes excités.

Le niveau d'ionisation varie avec l'altitude, en fonction du type de rayonnement, des différents types de recombinaison des ions et des différents procédés de transport. Aux altitudes basses, la densité de l'air est haute, ainsi les effets de collision et de recombinaison prédominent, et par conséquent, l'ionisation est faible. (Figure 1.1). Avec l'attitude croissante, la densité de l'air diminue, le collision sont plus rares, mais d'un autre côté, le rayonnent solaire est plus fort. Dans cet zone une compétition entre ionisation et recombinaison existe. Encore plus haute, l'ionisation par le rayonnement solaire et les rayons X prédominent, et l'ionisation persiste après le coucher du Soleil à cause de faible taux de recombinaison.

L'ionosphère est décrite en général par plusieurs couches. On distingue plusieurs maximums d'ionisation que l'on désigne par des couches différentes appelées régions D, E et F.

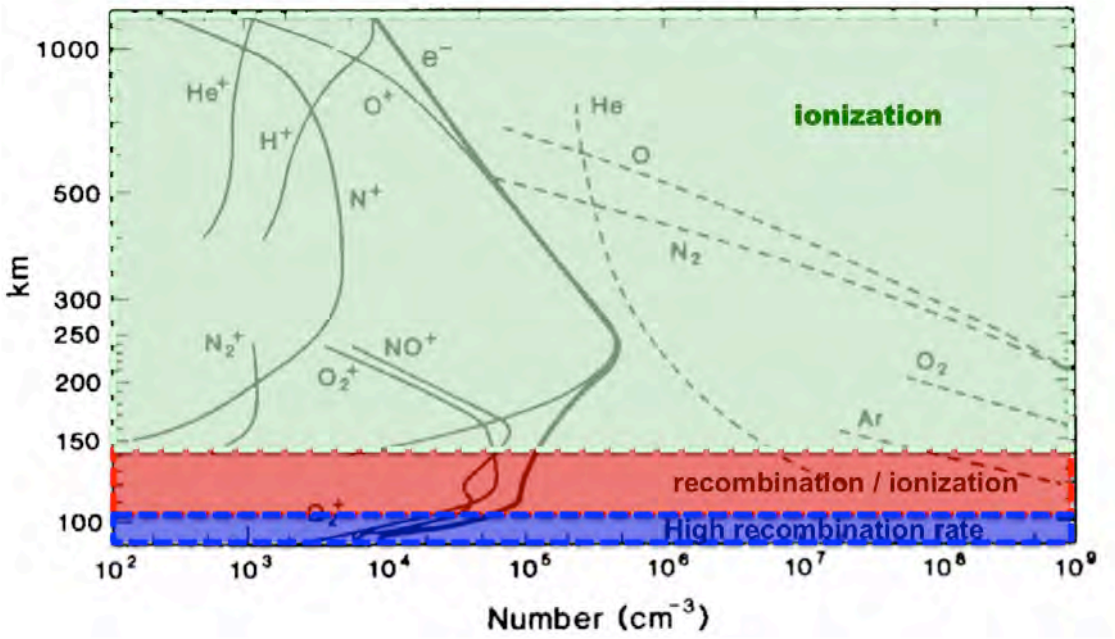


Figure 1.1: Densité des ions dans l'atmosphère et les zones où prédominent l'ionisation et la recombinaison.

La région D

La région D est située approximativement entre 60 et 90 km d'altitude au-dessus de la surface de la Terre. La couche est faiblement ionisée et disparaît en quelques minutes, lorsque le rayonnement solaire diminue et que la recombinaison avec des ions anéantit tous les électrons. A cause de sa faible densité électronique, les ondes électromagnétiques de haute fréquence (3-30 MHz) ne sont pas réfléchies dans la région D, mais subissent une perte d'énergie aboutissant à leur atténuation.

Les régions E et E sporadique

La région E se situe au-dessus de la région D (90 km and 130 km), et son ionisation est générée par les rayonnements ultraviolets et X sur les atomes d'oxygène. Le taux de recombinaison est plus lent que dans la région D, à cause de la densité d'air réduite (Figure 1.1). La région E diminue pendant la nuit, quand le soleil disparaît. Dans certains cas, en particulier pendant les mois d'été, des nuages fins avec une ionisation intense peuvent se former sporadiquement, on les nomme E sporadiques. Ils peuvent durer quelques minutes ou plusieurs heures et réfléchissent des ondes avec des fréquences allant jusqu'à 50 MHz ou plus.

La région F

La couche F (entre 150 km and 800 km) est divisée de jour en deux sous-couches, F1 et F2 qui se rejoignent après le coucher du soleil. Elle contient la densité la plus élevée en électrons libres. En comparaison des autres couches ionosphériques, la hauteur et la densité électronique sont très variables pour cette couche, en raison des variations diurnes, saisonnières et solaires du rayonnement du soleil. Concernant la propagation des ondes HF, elle joue un rôle primordial, parce qu'elle permet les communications radio à l'échelle mondiale.

1.3 Les modèles ionosphériques

Pour décrire l'état de l'ionosphère, des modèles ionosphériques ont été développés qui fournissent des moyennes mensuelles de la densité électronique, ainsi que la température des ions et la composition pour un lieu donné, en fonction de l'heure, la date, l'activité solaire et l'altitude.

Actuellement, environ 170 modèles ionosphériques sont en cours d'utilisation qui diffèrent par leur degré de complexité, le temps de calcul et leur but principal. En général, ils peuvent être classés en modèles empiriques, numériques, paramétrables, et par leur capacité à prendre en compte des données externes.

Les modèles ionosphériques sont constamment améliorés en incluant en outre de nouvelles données ou de meilleures formulations analytiques pour décrire les effets des tempêtes solaires et la variabilité au jour le jour. Les modèles les plus connus sont les modèles empiriques NeQuick et IRI. La structure de l'ionosphère et les modèles sont développés davantage dans le Chapitre 2.

1.4 L'ionosphère et le sondage ionosphérique

Lors de la traversée de l'ionosphère, les ondes électromagnétiques transmises interagissent avec les atomes, molécules et les particules chargées qui composent l'ionosphère. Ces interactions modifient les caractéristiques des signaux: retard temporel, déphasage, déviation angulaire, décalage fréquentiel, atténuation. En HF, les ondes électromagnétiques sont réfractées dans l'ionosphère en raison de la présence d'électrons libres et les signaux transmis peuvent être renvoyés vers le sol. Bien que la réfraction des signaux, le temps de retard et le déphasage introduit par l'ionosphère soient des effets indésirables dans la plupart des cas, ces effets peuvent être utilisés pour obtenir des informations sur la densité d'électrons dans l'ionosphère.

Parmi les principaux instruments utilisés pour étudier l'ionosphère et pour obtenir

des informations sur la densité électronique, on y trouve les sondeurs verticaux et les radars transhorizon. Tous les deux utilisent les propriétés de réfraction de l'ionosphère sur les ondes électromagnétiques à haute fréquence. Les sondeurs verticaux mesurent un ionogramme vertical, c'est à dire la hauteur virtuelle de réflexion d'un signal en fonction de la fréquence émise. Les ionogrammes verticaux peuvent être inversés pour obtenir un profil vertical de densité d'électrons au-dessus du sondeur.

Les radars transhorizon (OTH) sondent l'ionosphère suivant une propagation oblique et fournissent des ionogrammes de rétrodiffusion (Chapitre 3). Le signal émis, après déviation dans l'ionosphère, peut atteindre le sol au-delà de l'horizon radioélectrique, généralement à plusieurs milliers de kilomètres de l'émetteur. Le signal est rétrodiffusé et reçu par le récepteur (localisé prés de l'émetteur) après propagation par le même trajet (dans le cas le plus courant). Aujourd'hui, il existe peu de radars transhorizon dans le monde : Nostradamus en France ; Jindalee en Australie ; ROTHR, ROTHR-TX, et ROTHR-PR aux Etats-Unis ; un radar en Chine dont les informations sont très réservées, et le réseau SuperDARN (localisé près des 2 pôles).

1.4.1 L'inversion des ionogrammes de rétrodiffusion

Les études antérieures sur l'analyse de la densité électronique dans l'ionosphère par radar OTH sont toutes basées sur l'inversion des ionogrammes de rétrodiffusion qui représente l'amplitude du signal rétrodiffusé en fonction de la distance et de l'angle d'élévation ou de la fréquence. Pour estimer les trois paramètres ionosphériques principaux de chaque couche (la fréquence critique, la hauteur du maximum d'ionisation, et la demi-épaisseur de chaque couche), deux approches sont possibles (Chapitre 2) :

- soit la modélisation analytique de la courbe principale de rétrodiffusion par une couche ionosphérique quasi-parabolique,
- soit l'utilisation d'un tracé de rayons pour simuler numériquement cette courbe.

Les deux approches ne fournissent que des profils verticaux de la densité d'électron, valable à une position latitude / longitude donnée (au-dessus du sondeur vertical ou au point milieu du chemin). On suppose généralement qu'ils sont valables dans un rayon autour de la position. Les sondeurs verticaux sont souvent absents des régions d'intérêt ou alors leur répartition est trop lacunaire pour obtenir une bonne représentation de l'ionosphère sur une large zone.

1.5 La tomographie de l'ionosphère

Les méthodes décrites ci-dessus, basées sur les sondeurs verticaux ou les radars trans-horizons, permettent d'obtenir un profil vertical de la densité électronique dans l'ionosphère pour une position locale sur Terre. Le profil est supposé être valable également pour d'autres positions autour du point de mesure.

Il est souvent nécessaire de connaître l'état *a priori* de l'ionosphère non seulement localement, mais aussi sur une échelle mondiale ou régionale. La distribution du plasma peut être calculée par imagerie ionosphérique. Pour y parvenir, des méthodes de tomographie ionosphérique ont été développées.

La tomographie est une technique qui permet de reconstruire le volume d'un objet (d'un corps humain dans le cas de l'imagerie médicale ou d'une structure géologique en géophysique) à partir d'une série de mesures effectuées à l'extérieur de l'objet. Il s'agit d'une technique non invasive qui permet la visualisation des structures internes de l'objet sans modification de l'objet (Chapitre 4).

Ces mesures permettent une reconstruction de certaines propriétés de l'objet en fonction du type d'informations fournies par les capteurs (la pression acoustique, l'atténuation d'un faisceau lumineux, la variation de la vitesse ou de la polarisation des ondes sismiques)

1.5.1 La tomographie de l'ionosphère par GNSS

GNSS (Global Navigation Satellite System) est un terme général pour décrire des constellations de satellites de géopositionnement qui sont opérationnels et ont une couverture mondiale. En Avril 2013, le Système de positionnement global américain (GPS) et le système russe GLONASS sont les seuls GNSS opérationnels. Ils permettent de déterminer une position (longitude, latitude, altitude) avec une grande précision à l'aide des signaux radio transmis par des satellites et reçus par des stations au sol.

Les satellites transmettent des signaux bi-fréquences qui se propagent différemment dans l'ionosphère car l'indice de réfraction est à la fois dépendant de la densité du plasma et de la fréquence du signal propageant. La mesure de déphasage entre les deux signaux permet de calculer l'intégrale de la densité sur le chemin suivi par les ondes entre le satellite et le récepteur. On s'appelle cette intégrale le TEC (Contenu Electronique Total). Le TEC est généralement utilisé pour un représentation 2D de l'ionosphère (Figure 1.2) en supposant que toute la densité électronique est concentrée au maximum de l'ionisation (hypothèse de couche fine).

La mesure du TEC, conjointement avec certaines informations *a priori*, permet en outre d'estimer la densité électronique locale de l'ionosphère en utilisant la théorie

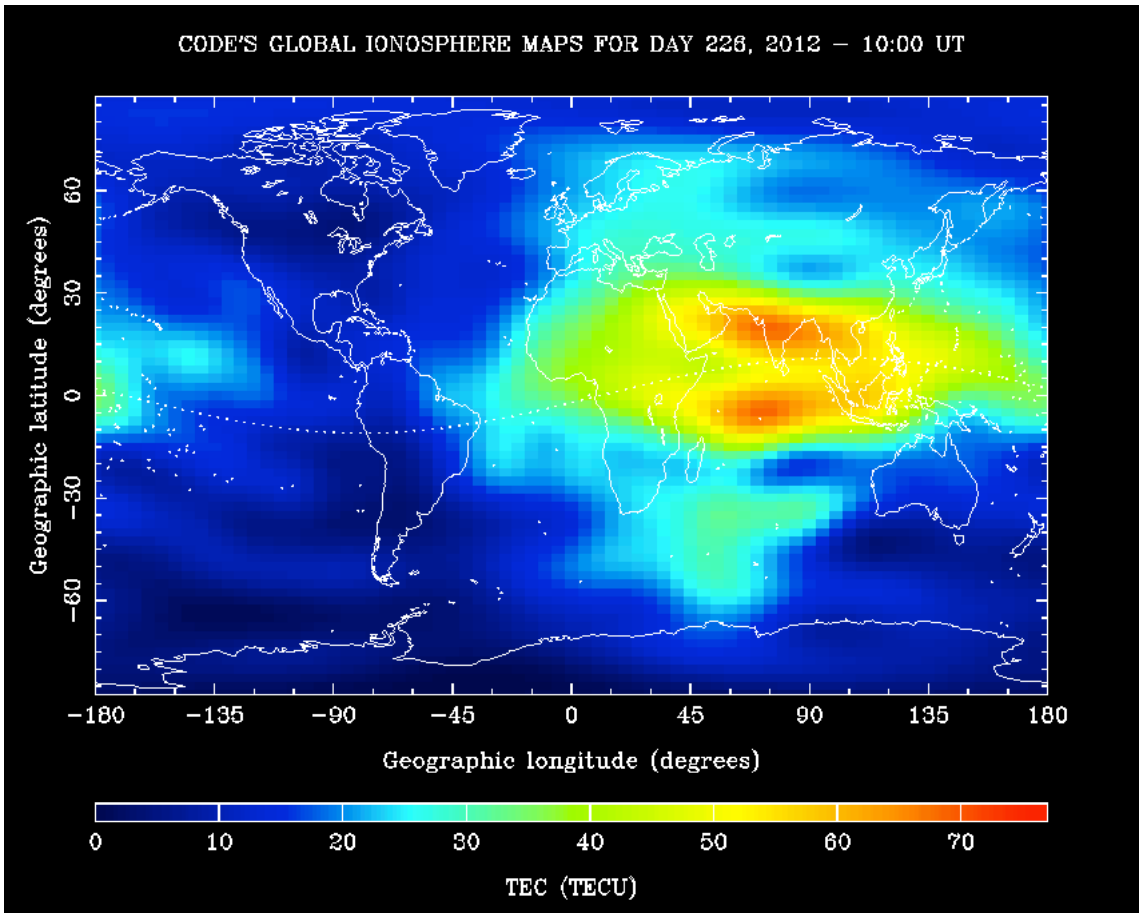


Figure 1.2: Carte global du TEC pour le jour 112 en 2012 (22 avril) à 10 TU, générée sur la base de 200 données GPS/GLONASS. Source: University of Bern

des problèmes inverses (Figure 1.3).

Les premières études de faisabilité de la tomographie ionosphérique par GPS ont été faites par *Austen et al. (1988)*. Grâce à la quantité croissante de données disponibles avec l'installation de réseaux mondiaux et régionaux de récepteurs GPS au sol, cette méthode a subi un développement continu et s'est fortement améliorée au cours des dernières années. Aujourd'hui, c'est la méthode la plus utilisée pour étudier les variations spatiales et temporelles de l'ionosphère en 3D avec affichage de l'évolution dans le temps (3D + 1).

Bien que le GPS soit un outil puissant pour étudier l'ionosphère, il a aussi ses limites. En dehors des limitations spatiales et temporelles associées aux conditions expérimentales, les ensembles de données sont souvent incomplets en raison de l'intervalle d'échantillonnage fini dans le temps et l'angle de vue limité de chaque récepteur. A cela s'ajoute le problème majeur de la tomographie ionosphérique utilisant les signaux GPS : l'absence de rayons à faible incidence, qui conduit à une faible résolution verticale. Les signaux GPS utilisent des fréquences de l'ordre du gigahertz ce qui limite leur sensibilité à la couche ionosphérique la plus dense (couche

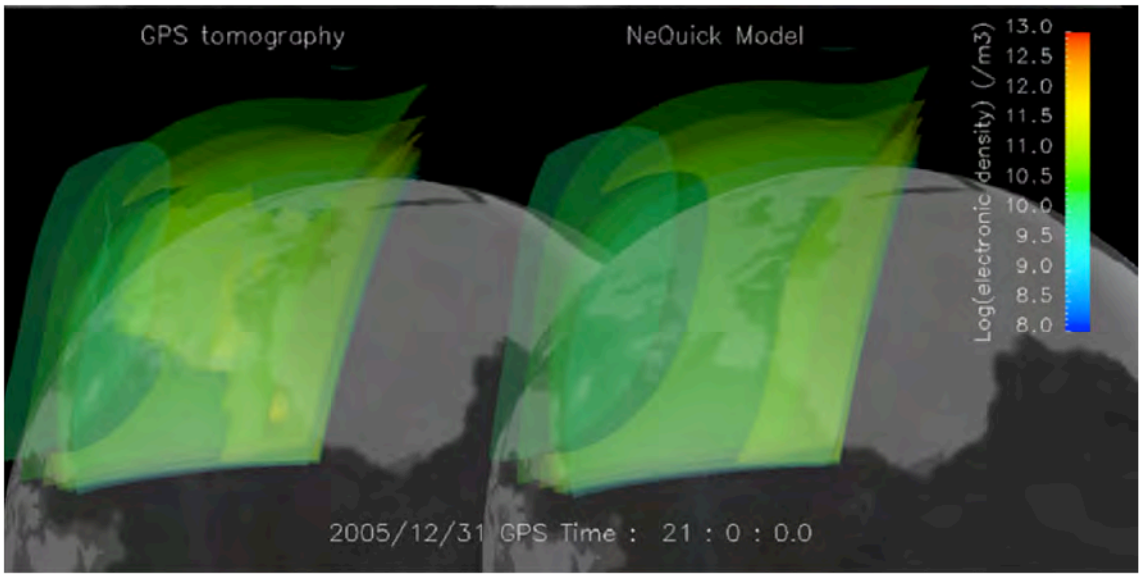


Figure 1.3: La tomographie de l'ionosphère par GPS. La densité électronique de l'inversion (gauche) et du modèle (à droite) à 21 TU présentée comme des isosurfaces. Source: *Garcia and Crespon (2008)*

F). La tomographie de l'ionosphère à partir de mesures de radar transhorizon en HF permet de contourner ces 2 problèmes et d'étendre la sensibilité tomographique aux couches inférieures.

1.6 La tomographie de l'ionosphère par radar transhorizon

1.6.1 Le tracé de rayons

La première étape pour développer une méthode de tomographie de l'ionosphère nécessite une modélisation du temps de propagation des ondes électromagnétiques (EM) dans le plasma ionosphérique. Ceci est réalisé par une modélisation de la propagation des ondes EM dans l'ionosphère en utilisant la théorie des rayons. Le code TDR (Tracé De Rayon) utilisé ici, modélise la propagation des ondes EM dans l'ionosphère sur la base de l'optique géométrique. Cela signifie qu'il modélise les ondes EM comme des rayons se propageant dans un plasma de densité locale Ne, qui affecte l'indice de réfraction. La propagation est faite dans un système tridimensionnel cartésien centré sur une terre ellipsoïdale entourée par une ionosphère tridimensionnelle obtenue par le modèle empirique NeQuick. Le modèle est donc totalement tridimensionnelle et prend un compte les hétérogénités latérales de l'ionosphère.

Nous utilisons l'équation Eikonal, qui décrit la propagation des rayons dans un milieu et nous la résolvons par la méthode de Runge-Kutta *(Press et al., 1992)* au

quatrième ordre. Le résultat fournit le chemin suivi par les rayons dans l'ionosphère et donc, le temps de propagation.

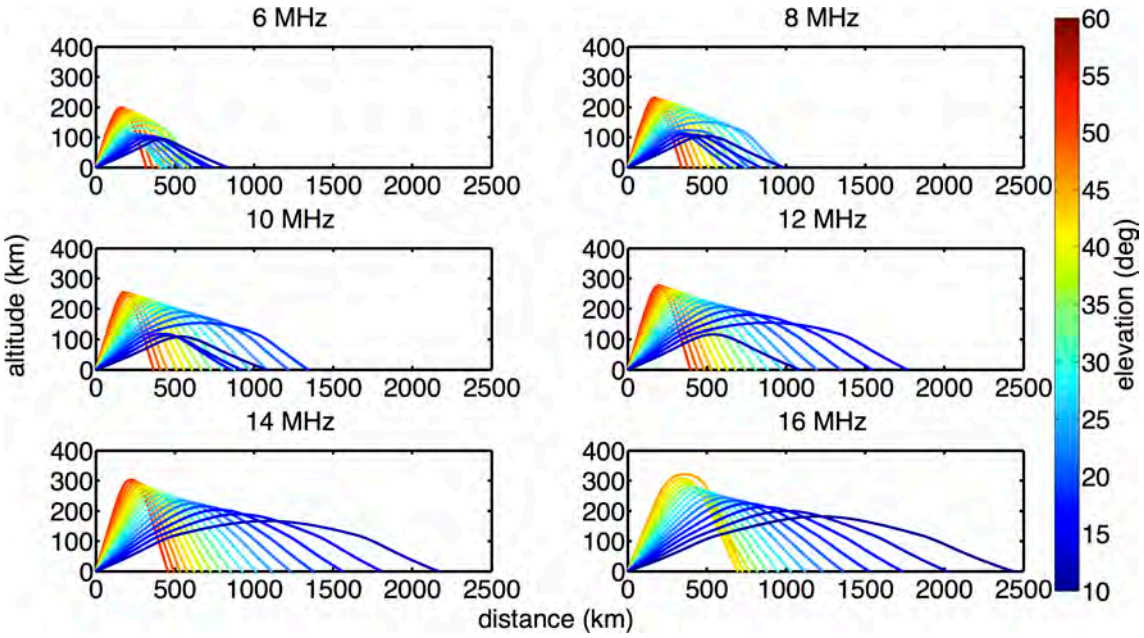


Figure 1.4: Exemple de Tracé de rayon dans l'ionosphère générée pour Octobre à midi en utilisant le modèle NeQuick. Pour chaque fréquence (6 MHz, 8 MHz, 10 MHz, 12 MHz, 14 MHz and 16 MHz), l'angle d'élévation varie entre 10° and 60°.

1.6.2 La théorie de la tomographie de l'ionosphère

Comme brièvement évoqué auparavant, le temps de propagation des ondes EM est influencé par la densité d'électrons Ne, qui affecte l'indice de réfraction $n(\vec{r})$ selon la formule suivante :

$$n(\vec{r}) = \sqrt{1 - \frac{e^2 N_e(\vec{r})}{4\pi^2 \epsilon_0 m_e f_e^2}} = \sqrt{1 - \frac{80.6 N_e(\vec{r})}{f_e^2}} = \sqrt{1 - \frac{f_p^2}{f_e^2}} \tag{1.1}$$

Toute variation du modèle ionosphérique introduit une variation de la vitesse de propagation des ondes et donc une modification du temps de propagation. La différence entre le temps de propagation dans un modèle ionosphérique T_{phase}^{synth} et dans l'ionosphère réelle T_{phase}^{real} est liée à une perturbation dans la densité électronique $\delta N_e(\vec{r})$

$$\delta T_{phase} = T_{phase}^{real} - T_{phase}^{synth} = -\frac{40.3}{c f_e^2} \int_{s_0} \delta N_e(\vec{r})\, ds, \tag{1.2}$$

avec la vitesse de la lumière c, la fréquence du signal f_e^2 et le chemin du rayon ds. Le problème inverse qu'il convient de résoudre est donc de calculer la perturbation dans

la densité électronique $\delta N_e(\vec{r})$ à partir d'une différence dans le temps de propagation mesuré, δT_{phase}.

Afin de résoudre ce problème inverse, trois méthodes d'inversion ont été développées et appliquées : deux méthodes linéaires (inversion par les moindres carrés (Section 4.5.2) avec et sans contraintes (Section 4.5.3)) et une méthode non-linéaire (méthode du gradient conjugué, Section 4.6.5). Les méthodes sont explicitement développées avec l'ensemble des détails analytiques dans le Chapitre 4 et également dans *Roy et al. (2014)* contenu dans l'appendice de ce manuscrit. Par la suite ne sont détaillées que les méthodes linéaires inversées par les moindres carrés.

1.6.3 L'inversion par les moindres carrés

Dans l'inversion par les moindres carrés, l'ionosphère a été paramétrée par N blocs homogènes, ne se recouvrant pas en distance et en altitude. Les problèmes inverses sont souvent des problèmes mal conditionnés parce que le nombre de paramètres à estimer est supérieur au nombre d'observations. Par conséquent, la solution n'est pas unique et de petites variations dans les données peuvent introduire de grands changements dans la solution. Donc, la solution doit être régularisée, en prenant en compte l'écart dans les données et l'écart au modèle initial $\mathbf{m^0}$. Pour trouver une solution stable, satisfaisant les conditions ci-dessus, *Menke (1989)*, propose une inversion par les moindres carrés en utilisant un paramètre de régularisation λ.

$$\frac{\delta \mathbf{m}}{\mathbf{m^0}} = \left(\mathbf{A}^T \cdot \mathbf{A} + \lambda \cdot \mathbf{I}\right)^{-1} \cdot \mathbf{A}^T \cdot \frac{\delta \mathbf{T}}{\mathbf{T^0}}, \tag{1.3}$$

où $\frac{\delta \mathbf{m}}{\mathbf{m^0}}$ est un vecteur avec N perturbations dans la densité électronique, le vecteur $\frac{\delta \mathbf{T}}{\mathbf{T^0}}$ contient M perturbations dans le temps de propagation et $\mathbf{A}$ représente une matrice géométrique avec $M \times N$ segments du trajet des rayons ds_{ij} du rayon j dans le bloc i. $\mathbf{I}$ est la matrice d'identité.

1.6.4 Les méthodes "Velocity" et "Velocity&Ray"

Deux méthodes linéaires ont été développées : la méthode v et la méthode $v\&r$. La première prend un compte seulement l'effet de la densité électronique sur la vitesse des ondes électromagnétique, c'est à dire qu'on suppose que le chemin des rayons dans l'ionosphère réelle reste le même que dans le modèle initial. Cependant, cette hypothèse est fausse car le défi majeur dans le développement de la tomographie de l'ionosphère par radar trans-horizon est liée à l'indétermination sur la position des points de rétrodiffusion au sol. Ceux-ci peuvent changer pour un angle d'élévation constant car ils dépendent de la variation de la densité d'électrons dans l'ionosphère.

La déviation du trajet des rayons, introduit par la variation du point de rétrod-

iffusion, peut ajouter une variation supplémentaire dans le temps de propagation. Pour considérer cet effet, la méthode $v\&r$ a été mise au point. Cette méthode tient compte de la déviation du trajet des rayons en imposant une perturbation d'électrons connu, aléatoire dm^* dans les cases traversées par les rayons et en calculant l'effet sur le temps de propagation δT^*. La méthode se base donc sur l'hypothèse démontrée que le rapport $\frac{\delta T^*}{dm^*}$ est constant et indépendant de la perturbation imposée ; cela permet donc le calcul du kernel $K_{ji} = \frac{\delta T_j^*}{dm^*}$. L'inversion par les moindres carrés décrite dans la section 1.6.3 (l'équation 1.3) est appliqué à la matrice $\mathbf{A}$ pour la méthode v et la matrice $\mathbf{A} + \mathbf{K} = \mathbf{M}$ pour la méthode $v\&r$. Le meilleur paramètre de régularisation λ est choisi par le critère de la L-curve, après avoir déterminé les bornes de calcul de la L-curve à l'aide d'une courbe d'erreur. Les résultats sont également résumés par *Roy et al. (2014)*.

1.7 Résultats obtenus

Pour valider les méthodes mises au point (v et $v\&r$), elles ont été appliquées à un ensemble de tests sur signaux synthétiques (perturbations localisées ou en damier). Un cas idéalisé et irréel ou la variation de la densité électronique n'affecte pas du tout la propagation des rayons (appelée ici *frozen ray*) a aussi été calculé pour quantifier les limites de résolution envisageables par la géométrie et la paramétrisation de notre problème. Les résultats de l'inversion de *frozen ray* représentent les meilleurs résultats que les méthodes v et $v\&r$ peuvent atteindre. En général, ces tests montrent les meilleures performances de la méthode $v\&r$ et justifient la prise en compte de la déviation du trajet des rayons dans l'inversion. Seule la méthode $v\&r$ est capable de trouver la position correcte des anomalies. Pour une itération, la méthode v n'est pas suffisante, car elle ne reconstruit pas bien les modèles cibles (Figure 1.5). La quantification de performance montre également que la méthode $v\&r$ reproduit à 60% le modèle cible, là où la méthode v est limitée à 41%. *Roy et al. (2014)* démontrent aussi que le méthode itérative améliore les résultats à 70% et 60% respectivement.

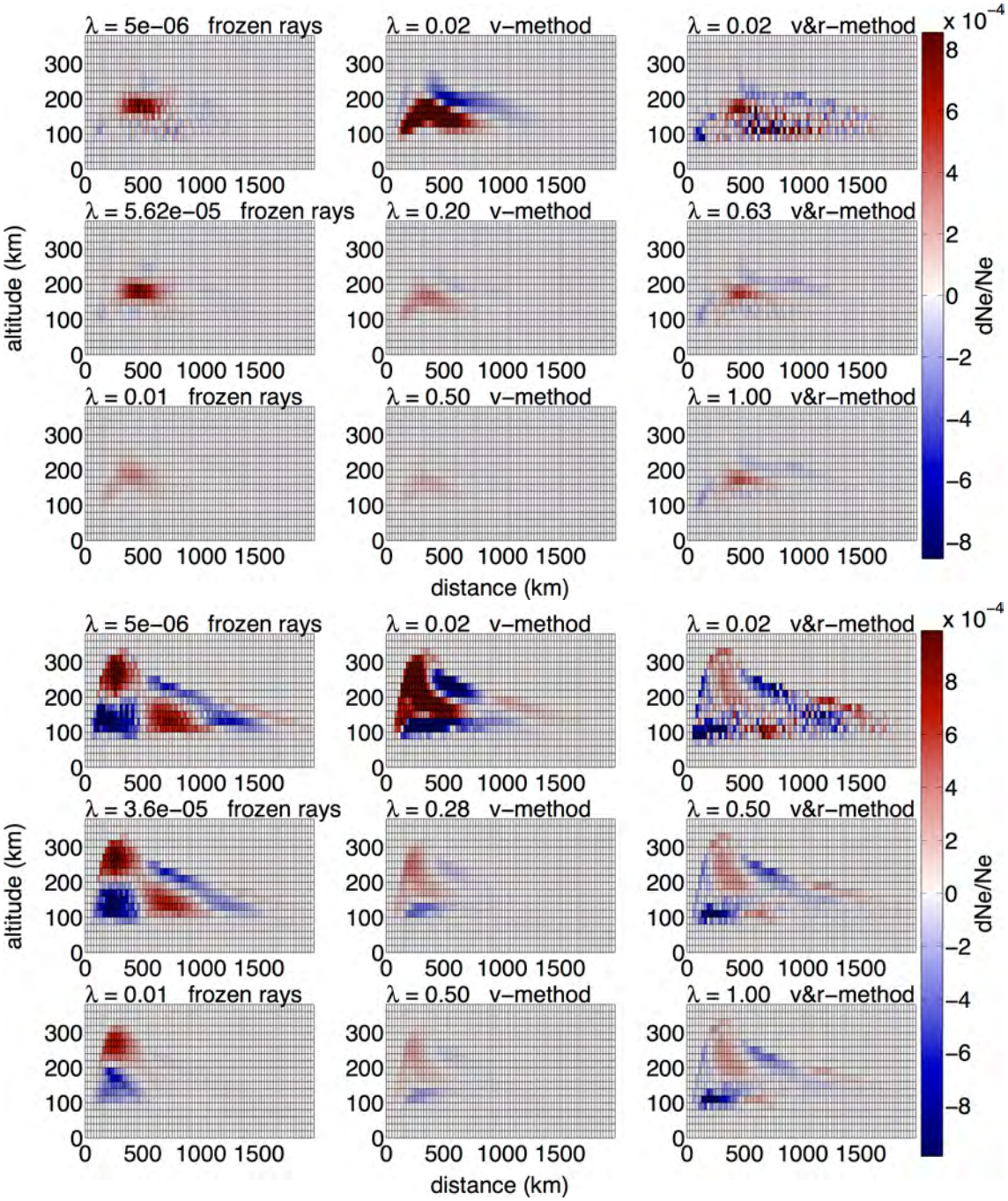


Figure 1.5: Résultats d'inversion après la première itération pour un test de damier et une perturbation localisée de 0.1 % pour différents paramètres de régularisation en utilisant l'approche des *rayons figés*, la v- et la $v\&r$-méthode. Le résultat de l'inversion pour le meilleur paramètre de régularisation est présenté dans chaque colonne dans la seconde figure à partir du haut.

1.7.1 Tests de résolutions

Les méthodes v et $v\&r$ ont été appliquées à un ensemble de tests sur signaux synthétiques de tailles et d'amplitudes différentes pour analyser la résolution accessible à

la tomographie de l'ionosphère (Chapitre 5). Dans ces essais, l'influence de la grille, les dimensions de la cellule, ainsi que la plus petite amplitude de la perturbation ont été analysées (Chapitre 5).

Les principales conclusions de ces tests de résolution sont que les deux méthodes d'inversion (v et $v\&r$) sont indépendantes de la grille. Ceci a été validé en utilisant le test du damier ou des perturbations localisées. Pour ces deux types de perturbations, les deux méthodes peuvent identifier des perturbations de la densité électronique, même pour des petites grilles de 25 km × 20 km en distance et en altitude. En plus, deux petites perturbations localisées de 165 km × 20 km peuvent être identifiées et séparées par les deux méthodes (Figure 1.6).

Pour la première itération, la méthode $v\&r$ donne de meilleurs résultats pour des perturbations en damier et pour une perturbation localisée. Elle permet d'identifier la position exacte de l'anomalie maximale et peut reconstruire plus de détails concernant la perturbation en damier à grande distance. La méthode v localise mal la perturbation, ne peut pas reproduire la forme et dans le cas de l'échiquier, le résultat est fortement amorti.

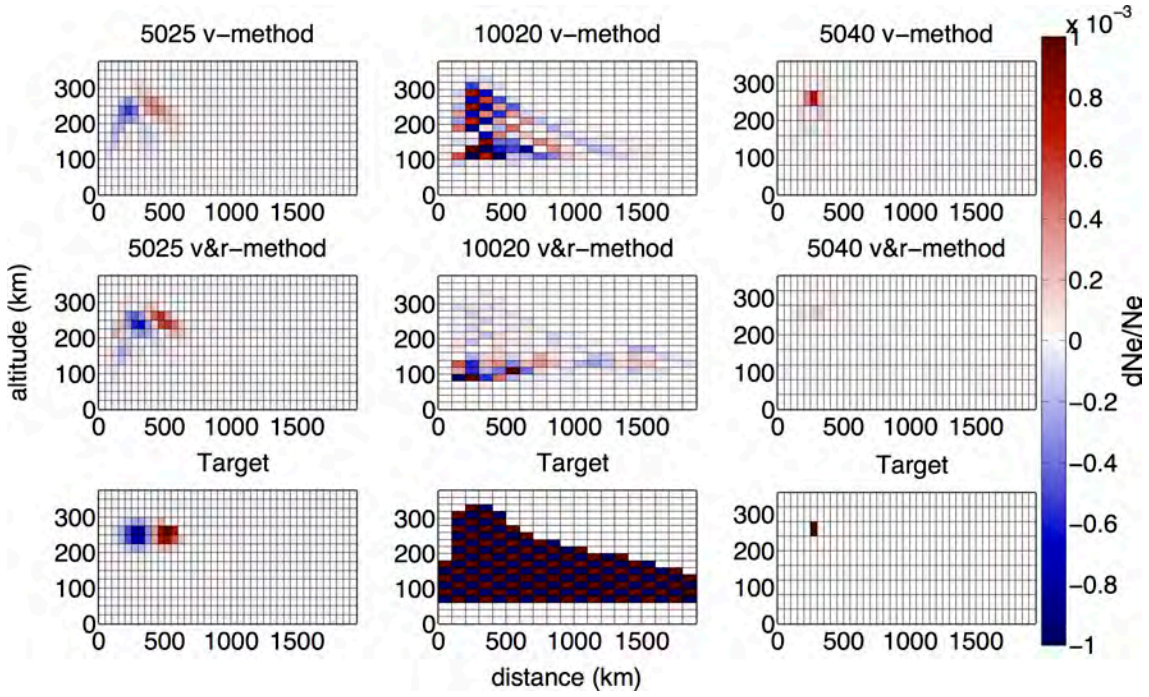


Figure 1.6: Résultats d'inversion pour trois modèles cibles et différentes grilles en utilisant la méthode v et la méthode $v\&r$.

Cependant, dans les deux résultats de l'inversion (v et $v\&r$), des anomalies négatives à grande échelle sont visibles pour une perturbation cible localisée. Ces anomalies ne sont ni présentes dans le modèle cible ni dans l'inversion *frozen ray* (Figure 1.5). Ces anomalies peuvent être expliquées par des effets non linéaires dans la dévi-

ation des chemins des rayons. A temps de propagation constant, une surévaluation positive de la perturbation entraîne l'apparition d'une perturbation négative à une autre position.

Tous les essais illustrent la dépendance de la solution à la couverture des rayons et à la zone de sensibilité maximale dans le milieu. Cette zone correspond aux endroits où la fréquence du plasma (liée à la densité électronique locale) est proche de la fréquence du signal. Le test en damier démontre clairement ces dépendances : les perturbations ne sont pas bien résolues ni à grandes distances, ni à basse altitude, ni dans les zones peu couvertes par les rayons (Figure 1.5).

1.7.2 L'approche itérative

Les résultats d'inversion après la première itération pour une perturbation localisée ont montré que, pour la méthode $v\&r$, l'amplitude de la solution est affaiblie, et pour la méthode v, la perturbation est mal localisée. Afin d'améliorer les résultats de l'inversion, une technique itérative a été développée pour les 2 méthodes. Les résultats préliminaires pour une approche itérative publiés par *Roy et al. (2014)* ont montré une décroissance des écarts en temps de propagation et des écarts par rapport au modèle cible avec les itérations, qui permet de reproduire le modèle cible à 70% avec la méthode $v\&r$ et 60% avec la méthode v. Cependant, *Roy et al. (2014)* ont aussi mis en évidence une perte des performances de la solution après une nombre critique d'itérations. Cela a conduit à la conclusion que l'approche itérative améliore généralement les résultats de l'inversion, car les écarts sont réduits par rapport à la première itération, mais le système devient instable après un nombre critique d'itérations.

L'instabilité de l'inversion est plus importante dans la méthode $v\&r$ que dans la méthode v. Ceci peut s'expliquer par le bruit numérique induit par la discrétisation des cellules, et par la non-linéarité de la déviation du trajet des rayons (Chapitre 6).

Pour explorer davantage ce problème et réduire l'instabilité, on a introduit un gain de boucle appelé coefficient de feedback (Chapitre 6). Les effets du feedback et de la régularisation ont été testés pour les deux méthodes. Le premier paramètre règle l'amplitude, la seconde amortit la solution ré-injectée.

Ces deux paramètres ralentissent la convergence dans les deux méthodes d'inversion, mais ils ne réduisent pas l'instabilité. Cependant, la méthode $v\&r$ atteint l'optimum généralement plus rapidement que la méthode v. Ceci peut s'expliquer par le fait que la méthode $v\&r$ tient compte de la déviation du trajet des rayons dès la première itération, alors que la méthode v prend en compte cette déviation avec le tracé de rayon dans le modèle mis à jour après la première itération. L'analyse des temps de calcul a montré que pour une itération de la méthode $v\&r$, 15 itérations avec la

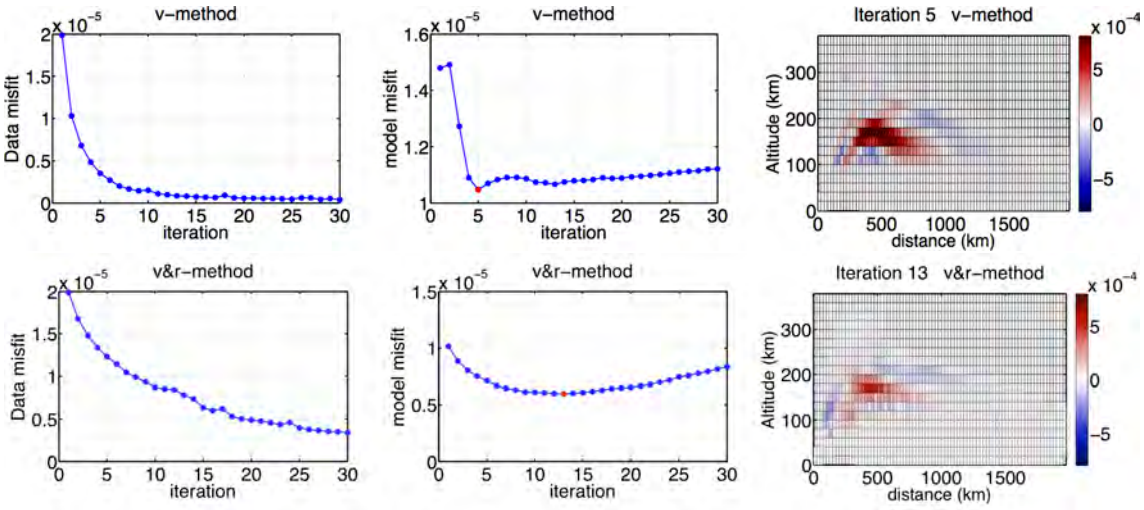


Figure 1.7: Ecarts en temps de propagation et des écarts par rapport au modèle cible avec les itérations pour la méthode v (haut) et la méthode $v\&r$ (bas). *Roy et al. (2014)*

méthode v peuvent être calculées dans le même temps après parallélisation du code. Cependant, bien souvent la méthode v n'a pas encore atteint un optimum après 30 itérations.

1.7.3 L'inversion de données réelles

Les méthodes v et $v\&r$ sont appliquées avec succès à des données réelles obtenues à l'aide du radar transhorizon NOuveau Système TRAnshorizon Décamétrique Appliquant les Méthodes Utilisées en Studio (Nostradamus) mis en œuvre par l'ONERA. L'inversion produit une carte de la densité électronique dans l'azimut choisi qui permet de calculer un profil vertical de la densité d'électrons à toute distance dans la zone atteinte par le radar (Figure 1.8).

Il s'agit d'une grande amélioration par rapport aux techniques classiques d'inversion des ionogrammes de rétrodiffusion. En effet, ces techniques ne donnent qu'un profil vertical au point milieu du trajet des rayons. Pour la première fois, la distribution réelle de la densité électronique dans l'ionosphère a été obtenue par l'inversion de données d'un radar transhorizon.

Nous avons montré le fort potentiel de cette méthode de tomographie pour reconstruire la distribution du plasma sur le continent européen. Dans un test préliminaire, les profils verticaux de la densité électronique obtenus par inversion étaient comparés aux profils fournis par le sondeur vertical de Chilton. La possibilité de combiner la tomographie par radar transhorizon avec la tomographie par GPS a été démontrée dans un deuxième test préliminaire sur des données synthétiques (Figure 1.9, et Chapitre 7)

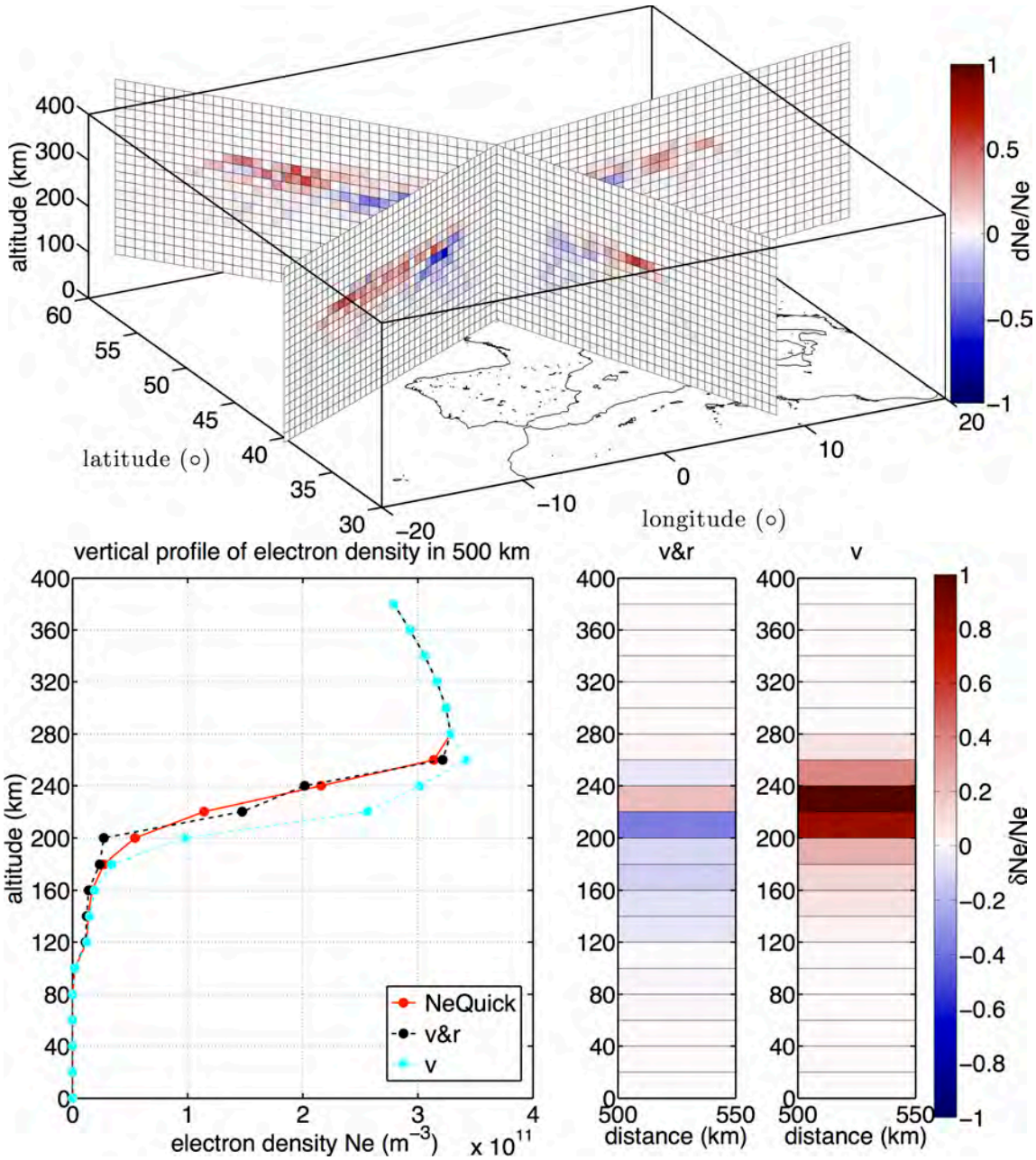


Figure 1.8: Perturbation dans la densité électronique sur l'Europe dans 4 azimut obtenue par l'inversion des données réelles (haut). Profil vertical de la densité d'électrons à 500 km en distance du radar obtenue par l'inversion des données réelles mesurées par Nostradamus en mars 2006 à 247° azimut (bas). La perturbation de la densité d'électrons $\delta Ne/Ne$ obtenue par l'inversion est montrée dans les images à droit. La perturbation de la densité électronique d'un point dans le tracé de gauche correspond au bloc à partir de l'altitude du point dans l'image à droite.

1.8 Conclusion et Perspectives

Dans cette thèse, une méthode de tomographie pour le radar transhorizon monostatique a été développée pour la première fois. Cela a eu pour but de contourner

les limitations de la tomographie par GPS, liées à la haute fréquence utilisée et l'absence des chemins horizontaux des rayons. Deux méthodes ont été développées, la méthode v et la méthode $v\&r$. La première prend en compte seulement l'effet de la densité électronique sur la vitesse des ondes électromagnétique. La méthode $v\&r$ considère aussi la déviation du trajet des rayons, introduit par la variation de la densité du plasma. Cela impose l'indétermination du point de rétrodiffusion, ajoutant ainsi une variation supplémentaire dans le temps de propagation.

Les méthodes ont été validées par des données simulées. Les tests montrent que les deux méthodes peuvent trouver les modèles cibles, mais que la méthode $v\&r$ peut mieux reconstruire leur position et la forme. Les tests de résolutions effectués ont révélés les possibilités de cette méthode de la tomographie par radar transhorizon et ses limites. Plus important, ils ont montré l'indépendance des résultats à la grille utilisée.

Une approche itérative améliore les résultats d'inversion, parce que des écarts en temps de propagation et des écarts par rapport au modèle décroissent avec les itérations (Chapitre 6).

La ré-augmentation d'écart par rapport au modèle cible après un nombre critique d'itérations peut être expliqué par le bruit introduit par la discrétisation de modèle ou par les effets non linéaires du déviation du trajet des rayons.

Après validation sur les données simulées, la méthode v et la méthode $v\&r$ ont été appliquées avec succès aux données réelles obtenues à l'aide du radar transhorizon Nostradamus mis en œuvre par l'ONERA. L'inversion produit une carte de la densité électronique dans l'azimut choisi qui permet de calculer un profil vertical de la densité d'électrons à toute distance dans la zone atteinte par le radar, couvrant ainsi l'ensemble de l'Europe.

Les profils doivent être validés par comparaison avec les profils des sondeurs verticaux. Cela était fait dans un test préliminaire avec le sondeur à Chilton, qui se trouve dans un azimut et une distance sondée par Nostradamus. La comparaison montre que l'écart entre le profil du modèle NeQuick et celui de Chilton peut se réduire avec les perturbations d'électron reconstruites par la méthode $v\&r$. Ce test préliminaire extrêmement encourageant peut être davantage amélioré en explorant plus en détail la méthode d'inversion utilisée par l'équipe du sondeur vertical de Chilton. L'utilisation des données brutes des sondeurs pourrait nous permettre de les inverser suivant les mêmes hypothèses que nos méthodes, voir les intégrer dans une inversion conjointe sondeur vertical - radar transhorizon.

La méthode de la tomographie de l'ionosphère par radar trans-horizon peut être effectivement combinée avec d'autres techniques, non seulement les sondeurs verticaux, mais surtout les mesures de TEC obtenues par les stations GPS au sol, ou embarquées sur satellites (occultation). La possibilité d'une inversion conjoint avec

le GPS a été démontre dans une test préliminaire (Chapitre 7). L'avantage d'une inversion conjointe est que les deux méthodes peuvent se compléter : La tomographie avec le GPS permet uniquement une bonne reconstruction de la densité d'électrons dans la région F2 ($\sim$ 300 km), et avec le radar transhorizon, une excellent sondage des altitude inférieure à 300 km est possible.

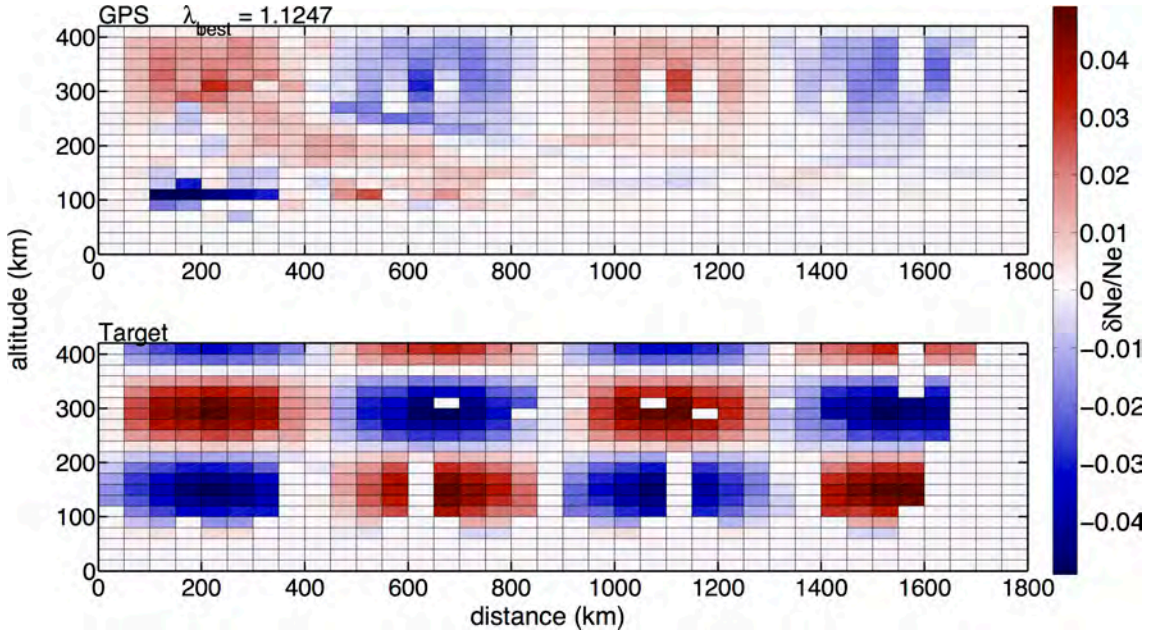


Figure 1.9: L'inversion conjoint des données simulées du radar transhorizon et du GPS (haut) pour un test de damier.

Chapter 2

Earth's ionosphere and ionospheric sounding

2.1 Introduction

The ionosphere is defined as the upper part of the atmosphere extending above the mesosphere from 60 km altitude up to about 1000 km, with its maximum of ionization around 300 km (Figure 2.1). This part is electrically neutral, but solar radiation produces electrically charged atoms and molecules. The ionosphere is a dynamic and highly variable environment that depends on the activity of the Sun and the effect of geomagnetic storms occurring there.

Although the ionosphere makes up less than 1% of the mass of the atmosphere above 100 km, its great importance results from its influence on electromagnetic (EM) waves. All satellite signals that cross the ionosphere (telecommunication, altimetry, radar, GPS, etc.) are affected by its presence. Hence, the exact knowledge of the structure of the ionosphere remains an important scientific and technological topic. The ionosphere has to be known precisely in Science to understand processes such as, for instance, spatial and temporal evolution of the plasma and to investigate all scales of wave activity ranging from planetary waves and tides down to small-scale Traveling Ionospheric Disturbances (TID)s generated at the ground. Apart from that, in the field of military and defense, where targets detected by radar need to be localized exactly as well as in satellite systems, the effect of the ionosphere on the measurements has to be known precisely. Consequently, the fine knowledge of the plasma density of the ionosphere is extremely important to detect perturbations, TIDs as well as to better understand the ionospheric behavior. In this chapter, the characteristics of the ionosphere as well as the methods of ionospheric sounding will be presented in details.

2.2 Discovery of the ionosphere: A brief history

The discovery of the ionosphere goes back to the first suggestion made by the physicist Carl F. Gauss in 1839 *(Schunk and Nagy, 2009)*, who speculated that the upper atmosphere might contain ionized regions to explain observed variations of the magnetic field at the surface of the Earth. Important steps in the discovery of the ionosphere were the proof of the existence of radio waves by Hertz in 1887 and the first transatlantic communication established in 1901 by Guglielmo Marconi with a radio transmission at a frequency of 300 kHz.

In 1902, Kennelly and Heaviside independently explained this achievement by suggesting the presence of a permanent electrically conducting layer high in the atmosphere *(Gillmor, 1982)*. In the following years, theories were developed on how the conducting properties of such a layer can be produced, but it was not until 1924 that Appleton was able to prove the existence of such a layer in the atmosphere. This was achived using a BBC transmitter at Bournemouth to broadcast a continuous signal to Oxford. By measuring the difference between the direct wave traveling along the ground and the signal reflected in the atmosphere, Appleton and his student Barnett were able to determine the height of the reflecting layer at an altitude of 100 km *(Gillmor, 1982)*. On December 12th 1924, Appleton called this layer the E layer. The reason for his choice is not clear, but one possible explanation is that in describing the reflected radio waves, he used the letter E to represent the electric field of the waves.

In 1925, Breit and Tuve in the USA were the first who measured the reflecting layer directly using the first radar-like device. It transmitted a very short radio pulse and the time it took to be reflected determined *(Schunk and Nagy, 2009)*. This method quickly became a standard technique for measuring the height of the reflecting layers and led to extensive investigations of the atmosphere. Finally, this technique resulted in the discovery of two other ionospheric layers, which were named in alphabetical order D and F layer, based on their location below and above the E layer, respectively. The F region is usually subdivided into F1 and F2 layers.

In 1926, the Scottish physicist Robert Watson-Watt first proposed the term ionosphere *(Hagfors and Schlegel, 2001)*. In the following years, there was a rapid progress in understanding the ionospheric behavior. Appleton developed the equation that describes the complex index of refraction and the polarization for a plane wave propagating in a plasma, taking into account the magnetic field, absorbing effects and collisions of the electrons,known as the Appleton-Hartree equation. A few years later, Sydney Chapman presented his theory of the formation of an ionized layer due to the action of solar Ultraviolet (UV) radiation *(Chapman, 1931)*.

2.3 Characteristics of the ionosphere

The Earth's atmosphere is composed primarily of the gases molecular nitrogen (N_2) (78 %), dioxygen (O_2) (21 %) and Argon (1 %) that make up more than 99 % of the air. All charged atoms in the ionosphere are produced either by photoionization, impact ionization of neutral atoms and molecules or indirectly by ionic-chemical reactions. The rate of ionization depends on neutral densities that decrease with height and incoming solar radiation that increases with height.

The most important quantities needed to understand the ionosphere are electron and ion concentrations as a function of altitude. They are solutions of the *continuity* equation describing the conservation of mass. Since ions and electrons are continuously created by ionization and also disappear in recombination producing neutral molecules, the *continuity* equation for these particles is *(Rees, 1989)*

$$\frac{\partial N_s}{\partial t} + \nabla \left(n_s \mathbf{u_s} \right) = P_s + L_s, \tag{2.1}$$

where n_s is the number density, P_s the production rate per unit volume, and L_s the loss rate per unit volume for ion species s. The second term on the left side, $\nabla \left(n_s \mathbf{u_s} \right)$, is a transport term and represents the flux divergence that depends on the velocity $\mathbf{u_s}$. In other words, it describes the difference between the flux of ions entering and leaving a given altitude.

The formation of the ionosphere mostly depends on the Sun radiation and is generated by photoionization. The Sun emits enormous quantities of radiation of all wavelengths that impacts atoms (X) or molecules (XY) in the atmosphere leading to free electrons and positively charged ions, *e.g.*,

$$X + h\nu \rightarrow X^+ + e \tag{2.2}$$

$$XY + h\nu \rightarrow XY^+ + e. \tag{2.3}$$

By contrast, the amount of ions decreases by recombination. The most important recombination processes are direct and dissociative recombination described by

$$X^+ + e \rightarrow X + h\nu \tag{2.4}$$

$$XY^+ + e \rightarrow X^* + Y^*. \tag{2.5}$$

The first process leads to the production of a photon, the second one to decomposition of a molecule forming two atoms in an excited state. The recombination rate depends on ion densities, which are decreasing with height. Therefore, recombination is more important in lower ionospheric layers and dissociative recombination of

electrons with molecular ions is the most important loss mechanism for electrons in the E-region.

In addition, collisions become frequent at low altitude and negative ions may form by attachment of an electron:

$$XY + e \rightarrow XY^- \tag{2.6}$$

Negative ions are generated in the D region (60 to 90 km altitude), where ion chemistry is complicated. They may undergo backward reaction under detachment of the electron:

$$XY^- + h\nu \rightarrow XY + e \tag{2.7}$$

At high latitudes, electric fields, particle precipitation, and field-aligned currents play an important role in ionization. Figure 2.1 illustrates ion densities in the atmosphere. O^+ predominates at altitudes around 200 km and H^+ starts to increase above 300 km (F-region). In the E-region ($\sim$ 150 km), the NO^+ and O_2^+ ions prevail.

In general, the ionosphere is described by layers. This is an incorrect description and could be misleading, since ionization occurs over the whole ionosphere, but other physical phenomena, including transport and diffusion, play an important role. The level of ionization varies with altitude, depending on the type of radiation, different types of ion recombination and various transport processes. The maximum ionization may be considered as different layers or, more correctly, regions called D, E and F. At lower altitude, a C region can also be present, but its level of ionization is too low for having any effect on radio signals, therefore it is rarely mentioned. A general overview of ionospheric 'layers' during day and night is shown in Figure 2.2.

D region

The D region reaching from 60 to 90 km is the lowest region within the ionosphere that affects radio communication signals to any degree. It is generated mainly by H-Lyman$-\alpha$ radiation ($\lambda = 121$ nm). This radiation corresponds to a UV emission line of the hydrogen atom and is emitted when the electron returns from the second to the lowest energy level. The energy of this radiation is sufficiently high to ionize nitric oxide (NO), which is found in small amounts.

Due to the air density and the resulting high recombination rate of ions, the layer is weakly ionized and disappears within a few minutes at night, when the incident radiation for ionization declines. On the other hand, the collision frequency of electrons and other particles during the day is very high with ca. 10^6 collisions per second, so High Frequency (HF) radio waves are not reflected in the D layer, but suffer loss of energy leading to their attenuation and decrease of intensity.

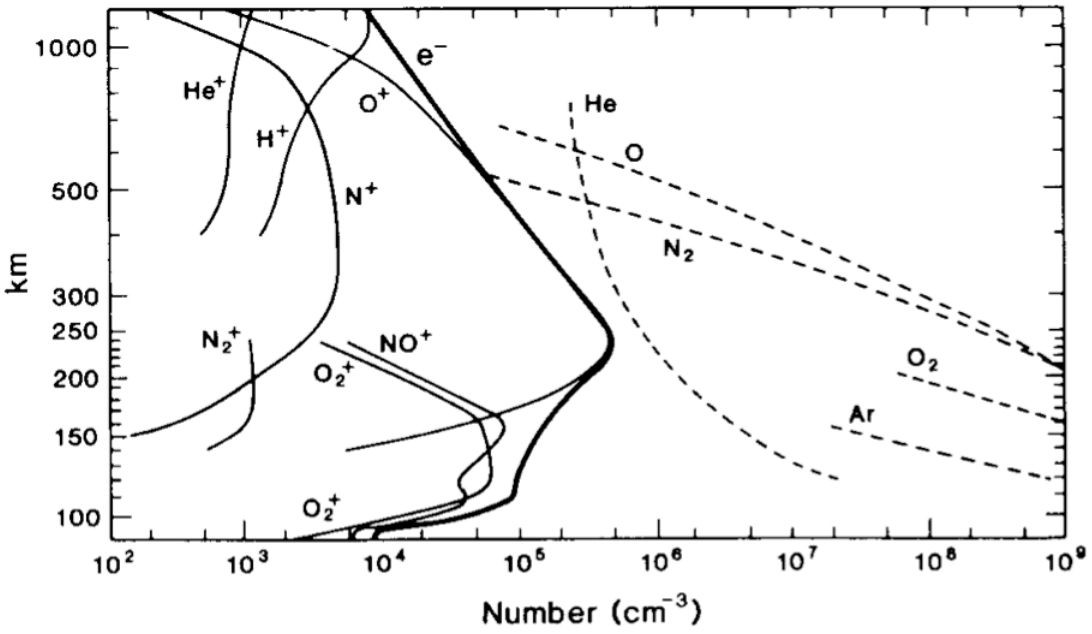


Figure 2.1: Daytime ion densities in the atmosphere based on mass spectrometer measurements. Source: *Luhmann (1995).*

E region

The E layer located above the D layer mainly consists of ionized nitric oxide (NO^+) and ionized oxygen (O^+) atoms produced by the action of X-ray and extreme ultraviolet (EUV) solar radiation on oxygen atoms. The ionization reaches a maximum at noon. It diminishes during night as electrons recombine with molecular ions such as ionized molecular oxygen (O_2^+) and NO^+ and the ionization source is no longer present. The recombination occurs slower than in the D-region and recombination with metallic ions such as Na^+ is very inefficient. Regarding the propagation of radio waves, this layer can only reflect radio waves with frequencies lower than about 10 MHz, higher frequencies are slightly attenuated.

Sporadic E layer

In certain cases, especially during the summer months, thin clouds of intense ionization, the so called sporadic E layer, may sporadically form between 90 km and 130 km altitude. They may last for just a few minutes or up to several hours and reflect signals with frequencies up to 50 MHz and higher *(Hagfors and Schlegel, 2001)*. The layer densities reach up to one order of magnitude greater than the background densities, and this layers contain primarily metallic ions (*e.g.,* Fe^+, Mg^+), originating from meteoric sources *(Schunk and Nagy, 2009)*.

In general, sporadic E layers are very narrow ($0.6 - 2$ km), but they can be found at all latitudes and multiple layers can occur simultaneously, separated by

$6-10\,\mathrm{km}$. The mechanisms leading to the formation of sporadic E layers are not well understood. Several phenomena are assumed to give rise to their formation, *e.g.,* electrical storms, auroral activity, and upper atmosphere winds, but the involvement of gravity waves is also discussed *(Budden, 1985, Chapter 1).*

F region

The F layer, also known as Appleton layer, contains the highest concentration of free electrons and ions in the atmosphere. Therefore, it is greatly affected by the position of the Sun and varies in the course of a day, falling at night as the radiation from the Sun disappears. During the day, the F layer splits into two sub-layers F1 and F2. After sunset these two layers merge again into one single F layer. Unlike other ionospheric layers, the height of the F layer as well as its electron density is highly variable due to large daily, seasonal, and sunspot-cycle variations leading in combination to a generally highly variable behaviour. During day, the electron density can vary between $10^{11}\,\mathrm{m}^{-3}$ to $10^{12}\,\mathrm{m}^{-3}$ (Figure 2.4).

The F layer acts as a "reflector" of signals in the HF component of the radio spectrum enabling ground based worldwide radio communications to be established. It is the most important region associated with HF signal propagation. As the density of gases at this altitude is much lower, recombination of ions and electrons takes place more rarely, with nearly about a quarter of the rate occuring in the E region.

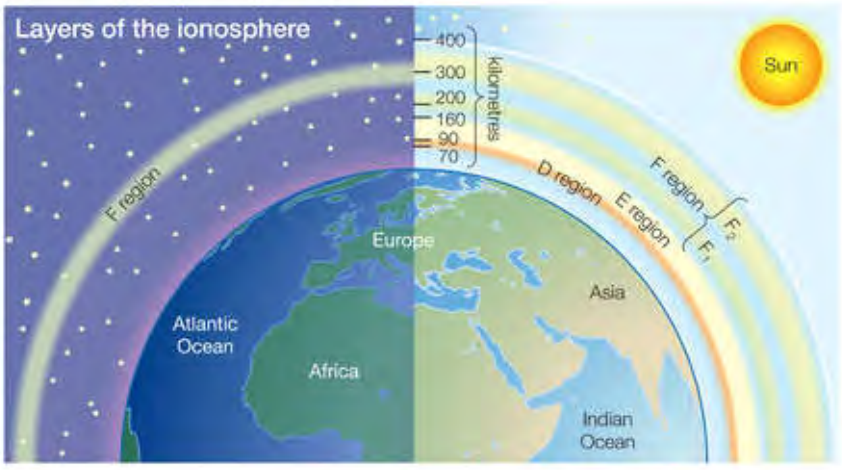


Figure 2.2: Ionospheric layers during day and night. Source: Encyclopaedia Britannica

Influence of the Sun

Since the Sun is the most important source of ionization and the Earth axis of rotation is inclined with respect to its orbital plane, the ionosphere undergoes diurnal,

seasonal and solar cycle variations. There are two factors influencing the ionosphere, namely the solar zenith angle and the solar radiation. Whenever they change, the ionosphere will change, too. Seasonal and diurnal variations of the ionosphere are related to a solar zenith angle change, while its solar cycle variation corresponds to a change in solar EUV and X-ray radiation fluxes. At solar maximum, solar EUV fluxes are greater than those at solar minimum and, as a consequence, these conditions lead to higher electron densities.

The ionosphere undergoes a diurnal variation due to the Earth rotation. At sunrise, the electron density begins to increase rapidly due to photoionization. It rises slowly further during the day and starts decaying afterwards when the photoionization source disappears. Electron density of the ionosphere can be visualized

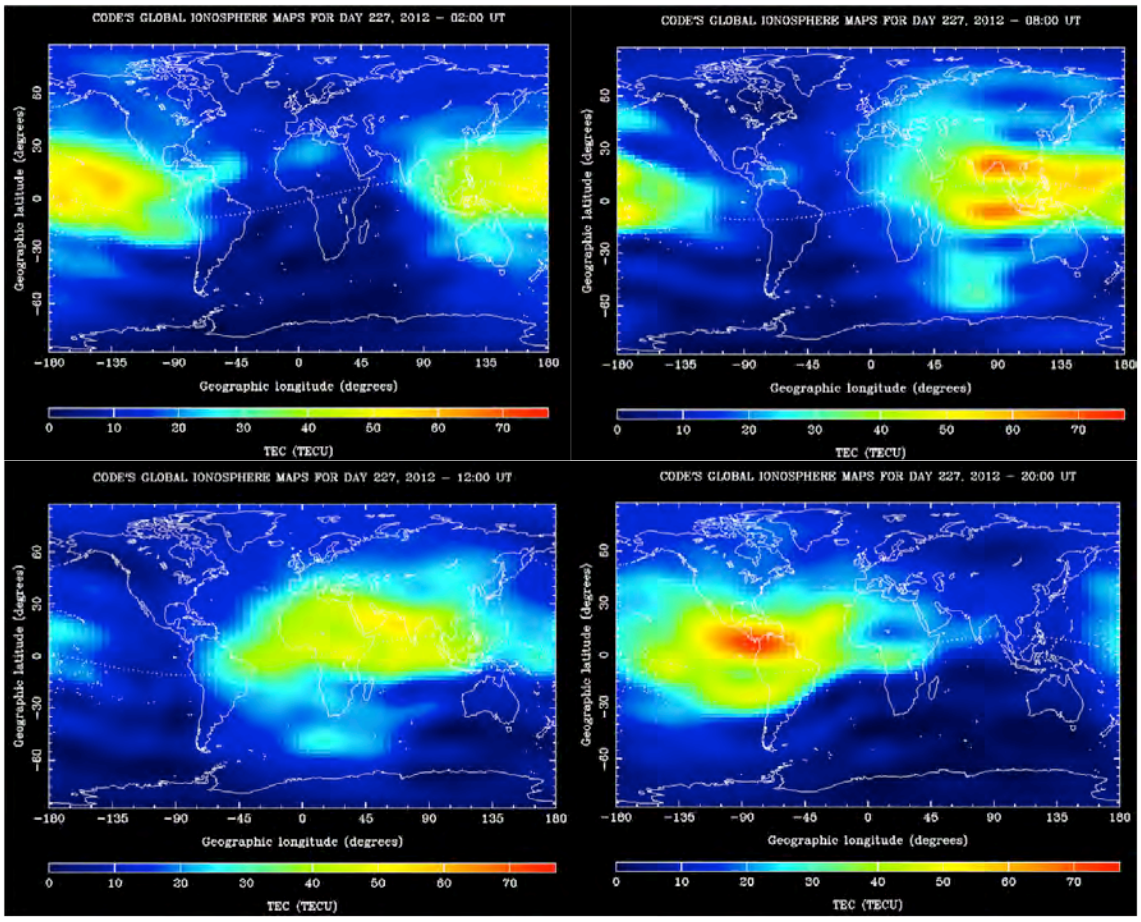


Figure 2.3: Global Total Electron Content (TEC) maps for day 112 in 2012 (April 22) at 2 UT, 8 UT, 12 UT and 20 UT. This map was generated on a daily basis with data from about 200 GPS/GLONASS sites. Source: University of Bern

by various sensing techniques that will be explained in detail in Section 2.6. All of them take advantage on the refractive properties of the ionosphere on HF radio waves. Ionosondes, for instance, are special radars that transmit vertically short EM pulses, which are reflected in the ionosphere. The electron density above the

sounder can be estimated from the time it takes the EM pulse to return. Ionospheric sounding techniques based on monostatic Over-the-horizon (OTH) radars use the signal backscattered from the ground to estimate a vertical profile of electron density at the midpoint of the oblique ray path. More recently, techniques based on Global Positioning System (GPS) were developed that enable imaging of the ionosphere.

Contrary to the previous mentioned techniques, they are based on the inversion of the TEC defined as the integral of electron density along the ray-path between a satellite and a receiver at the ground, which is computed from the phase difference between two signals emitted by the GPS satellite and measured at the receiver at the ground (Mannucci et al., 1998). Because of the very dense coverage of satellites and ground stations, with this technique, it is possible to visualize the ionosphere globally, and consequently the global variation linked to the Sun, too.

Figure 2.3 shows the daily variations of the global TEC. TEC is usually visualized as global or regional two-dimensional maps showing the state of the ionosphere *(Mannucci et al., 1998)*. In addition, the measured TEC in combination with some *a priori* information about the background ionospheric model allows estimation of the local electron density of the ionosphere using the inverse problem theory (see section 4.2). The electron density in the ionosphere can also be visualized as vertical profile as shown in Figure 2.4. It illustrates that the ionization below the F region peak reaches its maximum at noon with the solar zenith angle being smallest and decreases subsequently (Figure 2.4, right).

It is straightforward to conclude that a smaller zenith angle results in a higher electron density. Nevertheless, the electron density in the F2 layer during daytime is greater in winter than in summer, although the solar zenith angle is smaller in summer. This phenomenon, clearly evident in Figure 2.4, left, is called *seasonal anomaly* and is explained by seasonal changes in the ion concentration, where loss of ions by recombination is higher than their generation, so total ionization diminishes *(Schunk and Nagy, 2009, Chapter 11)*. The reason for this phenomena is not clearly understood.

Solar cycles are periodic changes in activity and appearance of the Sun with an average cycle duration of about 11 years. The solar cycle modulates the solar flux as well as the solar wind that is defined as the stream of particles released by the Sun. Perturbations of the solar wind can be caused by specific solar events, particularly Coronal Mass Ejections (CME), were plasma consisting primarily of electrons and protons is ejected by the Sun. When this plasma reaches the Earth, the Interplanetary Magnetic Field (IMF) is compressed by the shock of the traveling mass of solar energetic particles and pressure is exerted on the magnetic field. This causes strong disturbance of the Earth's magnetic field resulting in disruption of communications and navigation systems, intense auroras, damage to satellites as

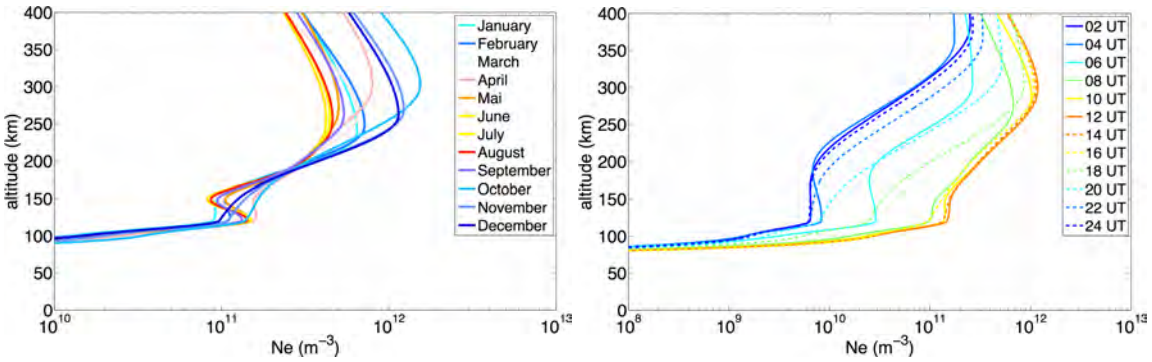


Figure 2.4: Vertical electron density profiles calculated using the ionospheric model NeQuick for the year 2013 (left) at noon and every two hours at March 28, 2014 (right).

well as induced currents in power lines and pipelines that lead to power outages and corrosion in the case of the most extreme storms.

Another typical perturbation of the ionosphere is introduced by Solar Flares. Solar Flares are intense emissions of UV and X rays by the Sun leading to strong ionization in the D region of the ionosphere which is referred to as sudden ionospheric disturbance (SID). They can occur very suddenly developing within a few seconds and cause strong ionization when the increased high-energy solar radiation reaches the upper atmosphere. The strong ionization can absorb radio waves and interrupt radio communication may be interrupted. Flares are frequent around peaks of sunspot cycle.

2.4 Ionospheric models

A large number of different models of the ionosphere have been developed and are used for various purposes, including scientific and practical applications such as correction of ionospheric effects on Global Navigation Satellite System (GNSS) and telecommunication channels. These models differ by their degree of complexity, calculation time and their primary purpose. At present, approximately 170 ionospheric models are in use *(Schunk, 2013)* and discussed in an overview *(American Institute of Aeronautics and Astronautics, 1999)*. Basically, they can be classified into empirical, physics-based numerical, parameterized, and data assimilation models.

Empirical models are based on measurements collected over an extended period of time using in situ and remote methods. The data consist of easily measured parameters of each ionospheric layer such as critical frequencies (f_oE, f_oF_1, f_oF_2), peak heights (h_mE, h_mF_1, h_mF_2) and half-thicknesses (y_mE, y_mF_1, y_mF_2). Subsequently, the collected data are averaged and fitted to simple analytical expressions or orthogonal polynomials in order to construct an electron density profile.

Physics-based numerical models (*e.g.,* Sami2 is Another Model of the Ionosphere (SAMI2)) *(Huba et al., 2000)* describe the distribution of electron density together with ion temperatures and drifts. Hence, they allow to study ionospheric behavior over time, including variation of electron and ion densities and temperatures with altitude, latitude, longitude, and solar cycle as well as the behavior under geomagnetic conditions and at different seasons. These models are principally calculated by solving numerically conservation equations (continuity, momentum, energy, etc.) for the ions and electrons taking into account chemical and transport processes in the ionosphere.

Parameterized models such as the Parametrized Ionosphere Model (PIM) *(Daniell et al., 1995)* are empirical models fitting orthogonal functions to the output obtained from a large number of numerical simulations. It returns ionospheric parameters as well as ion composition on a global scale.

Data assimilation models, *e.g.*, Global Assimilation of Ionospheric Measurements (GAIM) *(Schunk et al., 2004)*, include different types of data from various measurements resulting in a real time ionospheric model. Data sources are, for example, (1) electron density profiles of the bottom side ionosphere from a network of ionosondes, (2) TEC between a large network of ground stations and GPS satellites, or (3) TEC from occultations between satellites. Details on ionospheric monitoring techniques will be presented in Section 2.6.

In the following, two widely used empirical ionospheric models, namely International Reference Ionosphere (IRI) and NeQuick, are discussed in detail. A visual comparison of the TEC obtained using this two models is presented in Figure 2.5. In both parts of the figure, the equatorial anomaly (the areas with higher TEC about 20° north and south of the equator) becomes evident. Empirical models benefit from their independence of the evolving theoretical understanding of processes in the ionospheric plasma, but a disadvantage is their dependence on the underlying database.

The International Reference Ionosphere (IRI)

The IRI model *(Bilitza et al., 2014)* is the standard reference model for the ionosphere. It is a global empirical model based on worldwide collected ground and space data. Data sources are the worldwide network of ionosondes and storm time ionosondes, the Incoherent Scatter Radar (ISR), incoherent scatter data obtained,

for instance, from the European Incoherent Scatter (EISCAT) radar, and satellite data such as EUV data from the Aeronomy Satellite (AEROS), measurements by the International Satellites for Ionospheric Studies (ISIS) and Alouette satellite provide the basis of the IRI model.

Established in the late 1960s by the Commitee on Space Research (COSPAR) and the Union Radio-Scientifique Internationale (URSI), IRI has been steadily improved over the years using updated data and better modeling techniques. Nowadays, a working group of 58 experts is in charge of developing and improving the model. A detailed list of improvements provided since 1968 can be found in *(Bilitza et al., 2014)*.

For a given location, time, date and solar activity, IRI provides monthly averages of electron density and temperature, ion temperature and ion composition in an altitude range from about 60 km to about 2000 km. Thus, IRI can describe monthly varying electron densities, but no day-to-day variability that requires real-time data and an update or assimilation technique combining IRI with these data. This is briefly discussed in *Bilitza et al. (2014)*.

To obtain the electron density with IRI, the vertical profile is described by 7 subsections: the D region, the E-bottomside, the E-valley, the intermediate region between E and F1 layer, the F1 layer, the F2-bottomside, and F2-topside *(Bilitza, 1990)*. The topside and the bottom side of the electron density profiles are normalized to the F2 peak density and heights. The latest version of the model, IRI-2012, is also able to describe storm effects in the auroral E-region and includes auroral boundaries that allow a better representation of density and temperature features at these boundaries *(Bilitza et al., 2014)*.

Ionospheric models such as IRI rely on solar indices involving daily and seasonal variations as well as the impact of solar activity on ionospheric conditions. In general, these solar indices are the sunspot number R (number of dark spots on the solar disc) and the solar radio flux at 10.7 cm wavelength (F10.7). Both can be observed from the ground and long data records exist. IRI uses the 12-months running mean of the sunspot number observed at the Zurich observatory (R_{12}) and the IG12 index *(Bilitza, 1990)*. R_{12} is a smoothed value over the values obtained for six months before and six months after the month for which R_{12} is computed. IG12 is based on F peak plasma frequencies measured by 13 ionosonde stations and on the linear regression with solar activity taken from the International Telecommunication Union-Radiocommunication (ITU-R) model.

NeQuick model

The "quick calculation model" NeQuick *(Radicella and Leitinger, 2001)* is an empirical model because it is based on a model introduced by *Di Giovanni and Radicella (1990)* taking into account the physical properties of the ionospheric layers. It has been developed by the International Centre for Theoretical Physics (ICTP) in Trieste, Italy, in collaboration with the University of Graz in Austria. NeQuick is a three-dimensional and time-dependent ionospheric electron density model, which provides electron densities in the ionosphere as a function of position and time. The input parameters of this model are the position (longitude, latitude, and height), the period (month and UT), and the solar activity (given by monthly-mean sunspot number R_{12} or 10.7 cm solar radio flux).

The electron density distribution is reproduced analytically up to the F2 layer peak using five semi-Epstein layers. The model uses the peaks of the E, F1, and F2 layers as anchor points modeled from the ionosonde parameters (f_oE,f_oF_1,f_oF_2). However, the critical frequency for the F2 layer, f_oF_2, is modeled by the Comité Consultatif International pour la Radio (CCIR) maps, the critical frequency for the E layer, f_oE, is a formulation modified by John Titheridge, and the critical frequency for the F1 layer, f_o, F_1, is assumed to be proportional to f_oE in daytime and 0 during night *(Leitinger et al., 2005)*. The topside is represented by another semi-Epstein layer, with a height-dependent, empirical thickness parameter.

As for the IRI model, several efforts have been made to improve the analytical formulation of the NeQuick model. *Leitinger et al. (2005)* improved the bottom side description of the model, and *Coïsson et al. (2006)* made major changes in the topside formulation leading to NeQuick 2 *(Nava et al., 2008)* that is now adopted by ITU-R recommendation. The NeQuick model is also used to correct the ionospheric delay for the Galileo satellite system *(Arbesser-Rastburg, 2006)*.

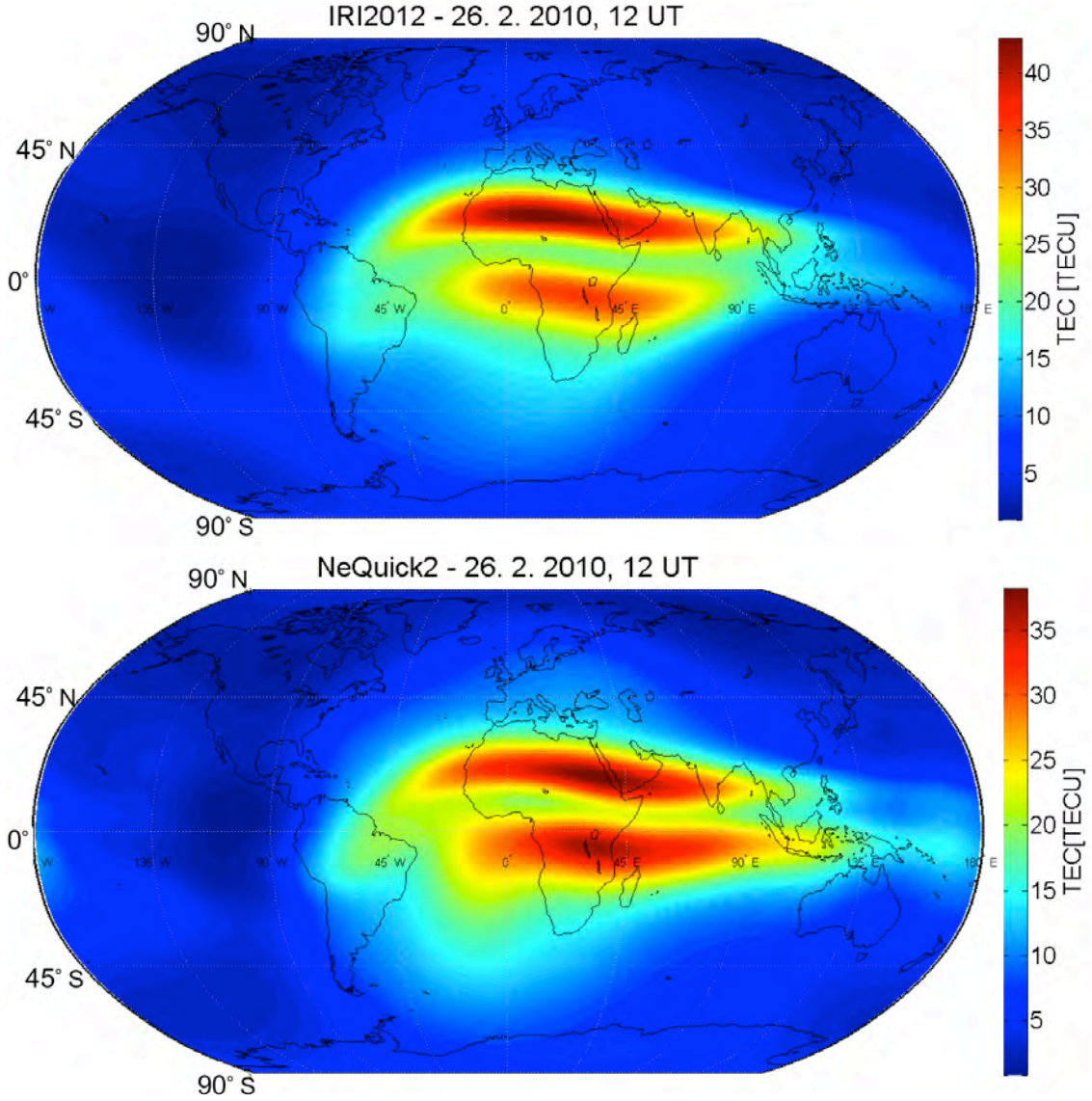


Figure 2.5: Global ionospheric TEC maps for 26th February 2010 at 12 UT modeled by the empirical models NeQuick and IRI. Although they are different in the way they model the ionosphere, both distinguish the equatorial anomaly. Source: *Najman and Kos (2014)*

2.5 Propagation of electromagnetic waves in the ionosphere

The main topic of this thesis is related to radio signals of high frequency radars, therefore details of the physics regarding the propagation of radio waves in a plasma, known as magneto-ionic theory, are discussed here. The cornerstone of this theory is the formula for the refractive index of an ionized medium in a magnetic field, known as Appleton-Hartree formula *(Davies, 1965, Chapter 3)*. The refractive index determines how much radio waves are bent when propagating through the ionosphere.

The complex refractive index is given by

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1-X-iZ)} \pm \left(\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right)^{\frac{1}{2}}}, \tag{2.8}$$

where the dimensionless quantities X, Y_T, Y_L and Z are defined as

$$X = \frac{\omega_0^2}{\omega^2}, \qquad Y_T = \frac{\omega_B \sin\theta}{\omega}, \qquad Y_L = \frac{\omega_B \cos\theta}{\omega}, \qquad Z = \frac{\nu}{\omega} \tag{2.9}$$

with the electron plasma frequency ω_0 and the electron gyro-frequency ω_B

$$\omega_0 = \sqrt{\frac{Ne^2}{\epsilon_0 m}}, \qquad \omega_B = \frac{Be}{m}. \tag{2.10}$$

In these equations ϵ_0 denotes the vacuum permittivity, e and m_e are the charge and the mass of the electron, respectively, ω is the frequency of the radio signal, ν the electron collision frequency, θ the angle between the direction of propagation and the magnetic field, and B is the ambient magnetic field strength. The plasma frequency is the characteristic frequency of the medium. In equilibrium, the electric fields of electrons and ions cancel each other out in a plasma. If an electron is displaced from an uniform background of ions due to thermal motion of the particles, an electric field is set up pulling the electron back to its original position. Because of their inertia, the electrons will overshoot and oscillate around the equilibrium position with a characteristic frequency, the plasma frequency.

In equation 2.8 Z is the collision term describing energy loss due to collision of particles. Y are the two propagation modes depending on the orientation of the magnetic field, where the subscripts L and T refer to the longitudinal and transverse components of the magnetic field. The $\pm$ sign in the Appleton-Hartree equation gives two separate solutions for the refractive index, since radio signals are split up due to the magnetic field of the Earth as soon as they penetrate into the ionosphere. The EM wave propagating perpendicular to the magnetic field lines takes the name *ordinary mode* and is usually indicated with a + sign; following the same approach, the *extraordinary mode* is the EM wave propagating parallel to the magnetic field lines, and it is usually indicated with a − sign. At higher frequencies, the *ordinary* and the *extraordinary* waves often follow very similar paths, but at lower frequencies they diverge and travel along completely different paths through the ionosphere. The existence of two split waves is clearly detectable in ionograms. At high frequencies, $f > 8\,\mathrm{MHz}$, the collisional term Z can be neglected and the refractive index is purely real or imaginary. Neglecting the magnetic field ($Y_L = Y_T = 0$) too and considering

that the angular frequency is linked to the frequency f by $\omega = 2\pi f$, equation 2.8 reduces to

$$n^2 = 1 - \frac{f_p^2}{f_e^2}. \tag{2.11}$$

At ground level, where the electron density (and consequently the plasma frequency) is zero, the refractive index is 1. It decreases with altitude until it is zero, then the plasma frequency f_p equals the frequency of the signal f_e. At this condition $n = 0$ the signal is reflected. If the frequency of the signal is too high, the signal is not reflected and escapes into space.

2.5.1 From waves to rays: Mathematical approximation

Typically, the propagation of waves in a medium is described by wave equations. A detailed description of the propagation of radio waves in the ionosphere can be found in *(Budden, 1985, Chapter 4)*. However, here ray theory will be used to represent the wave field as rays. The well known homogeneous wave equation for EM waves is given by

$$\left(\frac{\partial^2}{\partial x} + \frac{\partial^2}{\partial y} + \frac{\partial^2}{\partial z}\right)\phi = \frac{1}{v\left(\vec{x}\right)^2}\frac{\partial^2 \phi}{\partial t^2} \tag{2.12}$$

where v is the propagation velocity of the EM waves in the medium and ϕ the electric field. For a constant velocity, the solution of equation 2.12 could be written as a plane wave. Since the velocity varies slowly with local position, a solution where the amplitude A and the velocity are functions of position is assumed.

$$\phi\left(\vec{x}, t\right) = A\left(\vec{x}\right) \cdot \exp\ i\omega\left(\frac{W\left(\vec{x}\right)}{v_0} - t\right), \tag{2.13}$$

with $W\left(\vec{x}\right)$ as the *Eikonal* and v_0 a reference velocity. In an isotropic medium, the wavefronts are given by the surfaces, where W is constant and the rays are orthogonal trajectories to these surfaces. Substituting this modified solution in equation 2.12 gives two sets of equations. For the real part

$$\nabla^2 A\left(\vec{x}\right) - \frac{\omega}{v_0}^2 \nabla^2 W\left(\vec{x}\right) A\left(\vec{x}\right) = -\frac{\omega^2}{v\left(\vec{x}\right)^2} A\left(\vec{x}\right) \tag{2.14}$$

and for the imaginary part

$$i\left(\frac{2\omega}{v_0}\nabla A \nabla W\left(\vec{x}\right) + A\left(\vec{x}\right)\frac{\omega}{v_0}\nabla^2 W\left(\vec{x}\right)\right) = 0 \tag{2.15}$$

Equation 2.15 is the *transport equation* and can be used to compute the amplitude of propagating waves. In order to obtain information on propagation of the EM waves, only the real part is considered, for which follows

$$\nabla^2 W(\vec{x}) - \frac{v_0^2}{v(\vec{x})^2} = \frac{v_0^2}{\omega^2}\frac{\nabla^2 A(\vec{x})}{A(\vec{x})}. \tag{2.16}$$

The right hand side of this equation is a ratio of the spatial Laplacian of the amplitude to the product of amplitude and ω^2. For high frequencies (like in the case of OTH radar), this term is small. In that case it follows

$$\nabla^2 W(\vec{x}) = \frac{v_0^2}{v(\vec{x})^2} = n^2. \tag{2.17}$$

Equation 2.17 is called *Eikonal equation*. It means that the gradient of a wavefront at a position $\vec{x}$ is equal to the refractive index. The direction of maximum change of the wavefront defines the direction of the wave. Basically, this equation is one of several equivalent theorems in geometrical optics like Snell's law and Fermat's principle.

Further assumptions are required to satisfy the *Eikonal equation*: from equation 2.16 follows that the variation of the gradient of amplitude $A(\vec{x})$ for a given wavelength is smaller than the amplitude $A(\vec{x})$. This implies that

$$\lambda_0^2 \frac{\nabla^2 A(\vec{x})}{A(\vec{x})} \ll \nabla^2 W, \tag{2.18}$$

with $\lambda_0 = c_0 \cdot \frac{2\pi}{\omega}$ as a reference wavelength. Consequently, an implicit requirement for the *Eikonal equation* is that the ratio of amplitude variation to amplitude has to be much smaller than the refractive index

$$\lambda_0^2 \frac{\nabla^2 A(\vec{x})}{A(\vec{x})} \ll n^2. \tag{2.19}$$

Furthermore, the *Eikonal equation* (equation 2.17) implies that knowing the refractive index n allows to reconstruct the direction of the ray by ray tracing. Since the *Eikonal equation* is nonlinear and therefore difficult to solve, the ray path is calculated instead of the wavefront. Based on 2.17, the equation for the ray-path is derived *(Lay and Wallace, 1995)* and given by

$$\frac{d}{ds}\left(n\frac{d\vec{x}}{ds}\right) = \nabla n. \tag{2.20}$$

This is a second order differential equation for the ray path $\vec{x}$ and means that the change of the refractive index is related to the change of ray geometry, or more

precisely, the change in ray geometry is proportional to the spatial change in the refractive index n. The solution of this equation system is calculated numerically using the fourth order Runge-Kutta method *(Press et al., 1992)*.

2.6 Ionospheric sounding

The basic instruments used to study the ionosphere and to obtain information about the electron density in the ionosphere are ionosondes, GPS and radars. All of them take advantage of the refractive properties of the ionospheric plasma on EM waves. However, the information they provide about the electron density in the ionosphere is limited to a certain region, because they use different frequencies and incident angles. In this Section, methods and instruments used to investigate the ionosphere and to obtain information about the electron density are described.

2.6.1 Ionosondes

Breit and Tuve developed the first prototype of an ionosonde to measure the height of the ionospheric layers (see Section 2.2), and this device has become a basic instrument in studying the ionosphere. Ionosondes are special radars that transmit vertically short EM pulses with a frequency range of 0.1 - 30 MHz. They consist of an antenna with suitable radiation pattern as well as digital control and data analysis circuits.

The EM signals emitted vertically by the ionosonde propagate in the ionosphere as long as the local plasma frequency is smaller than the emission frequency and are reflected in a certain height depending on the signal frequency. From equation (2.11) clearly follows that signals with higher frequencies penetrate deeper into the ionosphere until they are reflected. Their echos are received by the receiver and analyzed by the control system. Modern Digital Ionospheric Goniometric Ionosonde (Digisonde) measure in addition to the travel time of the signal, angles of arrival, polarization and Doppler frequency shift.

The reflection height that can be calculated from the propagation time is called virtual height of reflection, because the group velocity of the EM waves (velocity with which the envelope of a pulse propagates in a medium) is smaller than the speed of light in the vacuum due to ionospheric dispersion. The virtual reflection height is always larger than the real reflection height in the ionosphere. The propagation time of the signal is related to the virtual height of reflection, h_v, by

$$h_v = \frac{1}{2} \cdot c \cdot \Delta t, \tag{2.21}$$

with c being the speed of light in the vacuum. Usually, the reflected pulse consists of two components, the ordinary and the extraordinary wave with different values of h_v resulting from the influence of the Earth magnetic field on HF propagation (Section 2.5).

Figure 2.6 shows an example of a measured ionogram being the virtual height of reflection as a function of frequency at the Digisonde in Juliusruh, Germany. In an ionogram, ionospheric parameters such as the critical frequency and the maximum electron density of each ionospheric layer can be identified, but it contains no information about the vertical profile of electron density. The relation between the virtual height of reflection and the electron density given by *(Reinisch and Xueqin, 1983)*

$$h_v = \int_0^{h_r} \frac{dh}{\sqrt{1 - \frac{N(h)e^2}{\epsilon_0 m w^2}}} \tag{2.22}$$

allows inversion of the ionogram to obtain a vertical profile of electron density. A number of methods for ionogram inversion have been proposed and developed since invention of the ionosonde.

One of the most common technique is a model-fitting method or POLynomial ANalysis (POLAN), a FORTRAN program developed by *Titheridge (1985)*, where the graph true height versus plasma frequency is approximated by polynomials. This program solves the inversion problem by breaking up the profile into simpler sections, for which physically acceptable solutions can be found, and using extrapolation and interpolation for the rest. It calculates only a profile for the bottomside (250 km) of the ionosphere. The exponential decrease of electron density in the topside ionosphere is normally described by an $\alpha-$Chapman or Epstein layer.

Automatic Real-Time Ionogram Scaler with True height (ARTIST) is the second commonly used ionogram inversion program that automatically scales digital ionograms by combining image recognition and analytical function fitting techniques. It provides a vertical profile of electron density from data of each Digisonde that measures polarization and incidence angles *(Reinisch and Xueqin, 1983)*. This technique has been used at the Digisonde in Germany to produce the ionogram in Figure 2.6.

A major drawback of this ionospheric sounding method is the fact that analysis of ionosonde data allows the reconstruction of vertical electron density profiles only for the bottom side ionosphere up to 250 km altitude for a local position on Earth. Typically, an ionosonde station obtains one ionogram recorded every 15 minutes. Figure 2.7 shows a map of the worldwide distribution of ionosondes constructed from the list provided by the National Oceanic and Atmospheric Administration

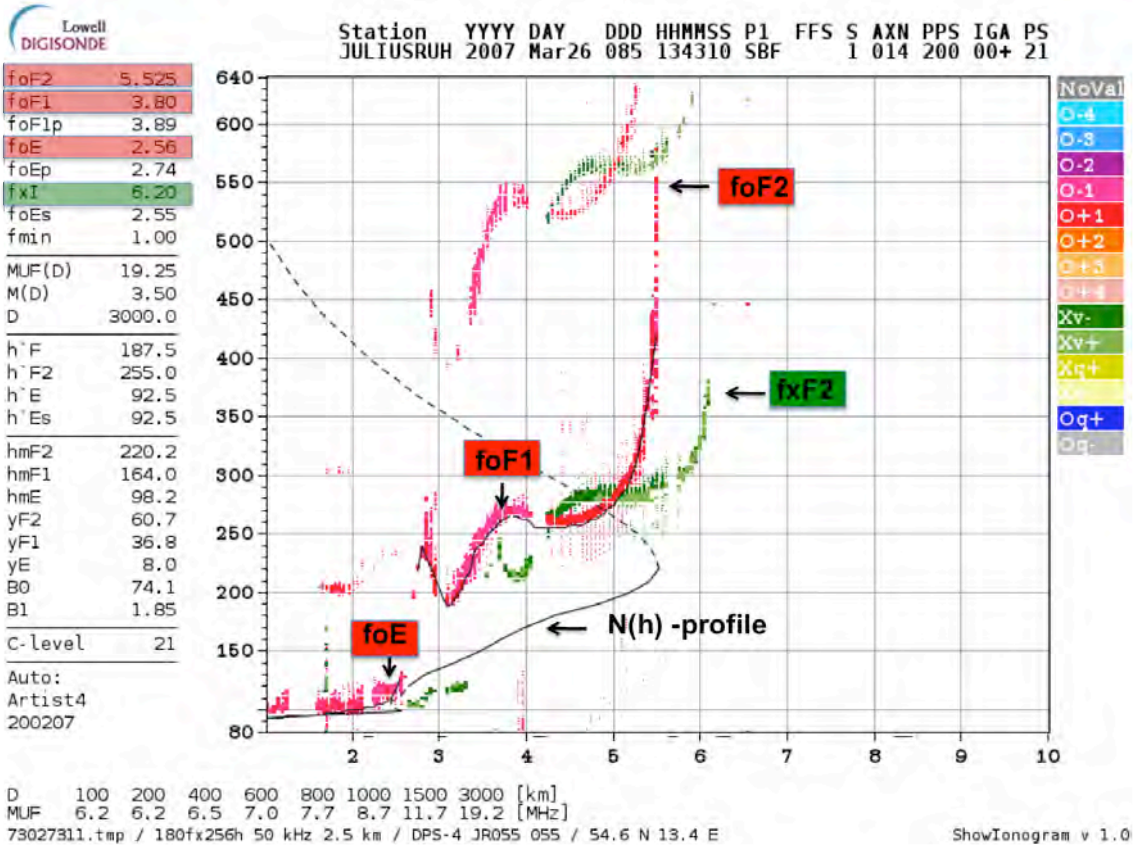


Figure 2.6: Example of an ionogram (virtual height of reflection in km as a function of frequency in MHz) measured at the Digisonde in Juliusruh, Germany. The black solid line shows the vertical electron density profile obtained by inversion up to 250 km. The topside profile (dotted line) is extrapolated using a Chapman layer. The E, F1 and F2 layers can be identified, as well as the ordinary (red) and extraordinary (green) modes.

(NOAA). Only 53 stations out of 420 ionosondes provide ionograms in real time to the scientific community, *e.g.*, the one at Chilton, U.K.

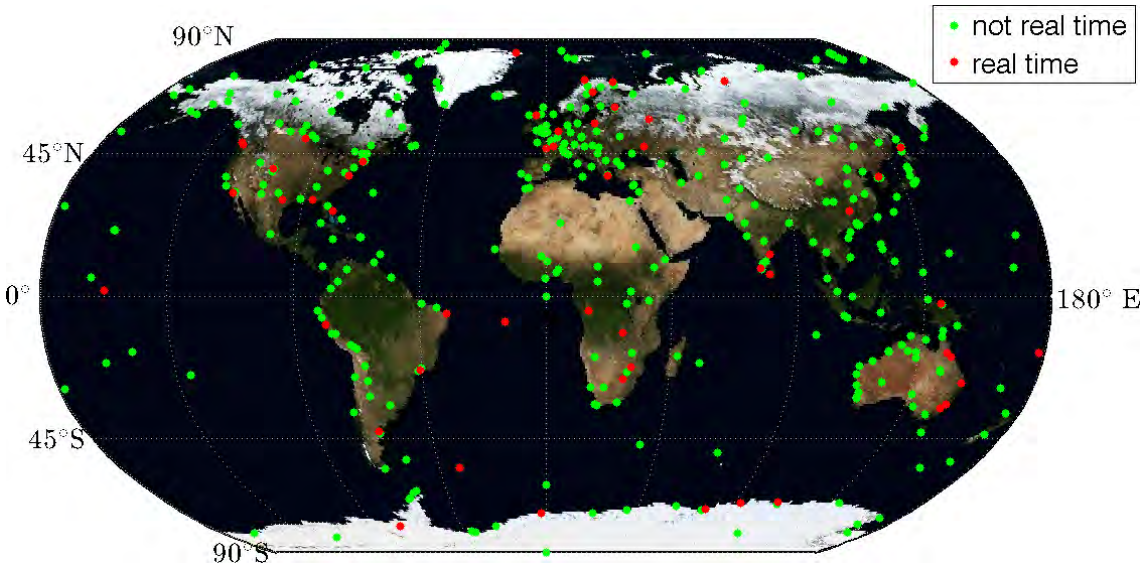


Figure 2.7: Worldwide distribution of ionosondes created on the basis of the station map from NOAA as at June 2014.

2.6.2 Inversion of backscatter ionograms

Another well known method is the inversion of backscatter ionograms obtained with OTH radar. This radar transmits a signal in an azimuthal direction and sweeps in frequency or in elevation. The signal is refracted by the ionosphere and backscattered from the ground and then analyzed to determine its energy when returning from different ranges. The results are called backscatter ionograms, which represent the amplitude of the backscattered signal as a function of range and elevation angle or frequency.

An example of a backscatter ionogram is shown in Figure 2.8. The strongest signal in the backscattered ionogram represents the leading edge, that is, for each frequency the ray with minimal group delay from the radar to a point of first contact with the ground. In other words, these are the fastest rays for each frequency. The leading edge is a function of two variables operating frequency and minimal group delay and contains information about ionospheric regions which are located thousands of kilometers away from the transmitter and are in some cases inaccessible, for instance the Arctic or the Antarctic. This turns the backscatter sounding into a powerful tool for investigation of the ionosphere. The backscattered ionograms can be inverted to obtain ionospheric parameters, but in general only the leading edge is used, since a complete inversion is extremely difficult and requires information regarding complex phenomena such as sea backscatter and absorption. This approach imposes a real limit in data exploration, because it neglects the information contained in the total backscattered signal.

The idea of exploiting data of OTH radar in order to obtain information on the

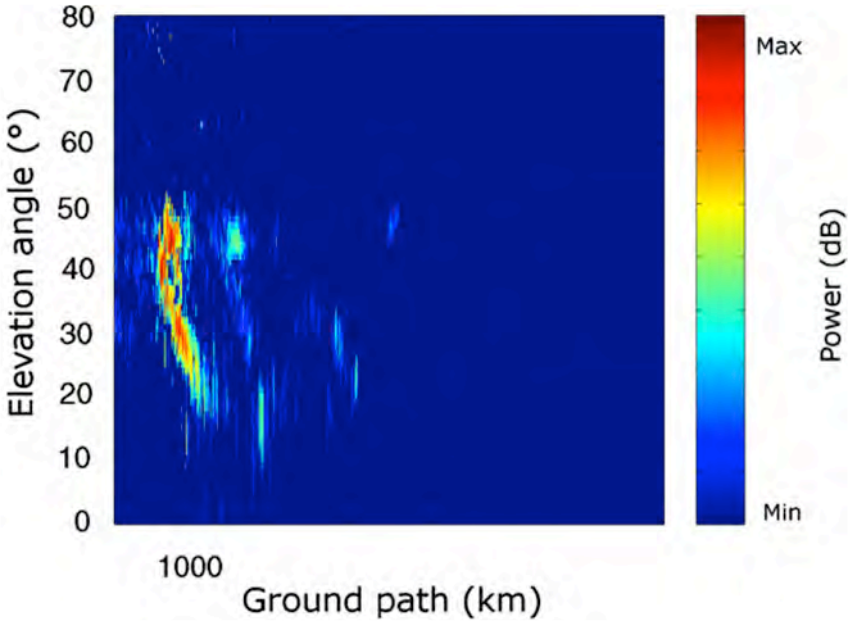


Figure 2.8: An elevation scan backscatter ionogram from OTH radar Nostradamus. Axis labels are not shown for confidential reasons. Source: *Benito et al. (2008)*

electron density in the ionosphere came up in the 1970s *(Hatfield, 1970; Rao, 1974)* and was further developed later on (*e.g.*, *Ruelle and Landeau (1994); Landeau et al. (1997)*). However, these methods use 2D ray tracing in a 1D ionosphere described by quasi-parabolic (QP) layers and invert for the three major ionospheric parameters (critical frequency f_c, peak height and semi-thickness for each layer) for a local position on Earth. More recently, *Benito et al. (2008)* developed an inversion method of backscatter ionograms optimized by simulated annealing, which was validated for real data of OTH radar Nostradamus.

Since previous methods were only able to provide ionospheric parameters for one local position, *Fridman and Fridman (1994)* extended the inversion of the leading edge from backscatter ionograms to reconstruct the two-dimensional electron density distribution by using additionally the vertical electron density profile measured over the sounding station. Finally, incorporation of data obtained from eight azimuthal beams and the vertical profile measured by quasi-vertical-incident (QVI) sounders *(Fridman, 1998)* allowed a 3 Dimensions (3D) reconstruction of the ionosphere.

Two results are shown in Figure 2.9 for one day in summer and winter at night-time. Their iterative method is able to produce two-dimensional snapshots of the horizontal structure of the lower F2 region, but it seems to be less effective for other ionospheric layers than the F layer. In addition, four data sets are needed as input for the inversion, that is, the leading edge data (group delay versus frequency and azimuth), the vertical profile of electron density from QVI sounders, and the root-mean-square error of the leading edge and vertical profile. Furthermore, the method

do not account for ray-path deflection, needed to be taken in account for monostatic OTH radar. In monostatic OTH radars, the endpoints of the rays (where the signal is backscattered at the ground) are unknown, so the location of the scattering point at the ground can change for a constant elevation angle and depends on the electron density variation $\delta N_e(\vec{r})$ in the ionosphere. The shift of the scattering point introduces an additional in the propagation time. This thesis focalizes on this effect and takes it into account in the inversion of OTH radar. The complete description of the method is presented in Chapter 4.

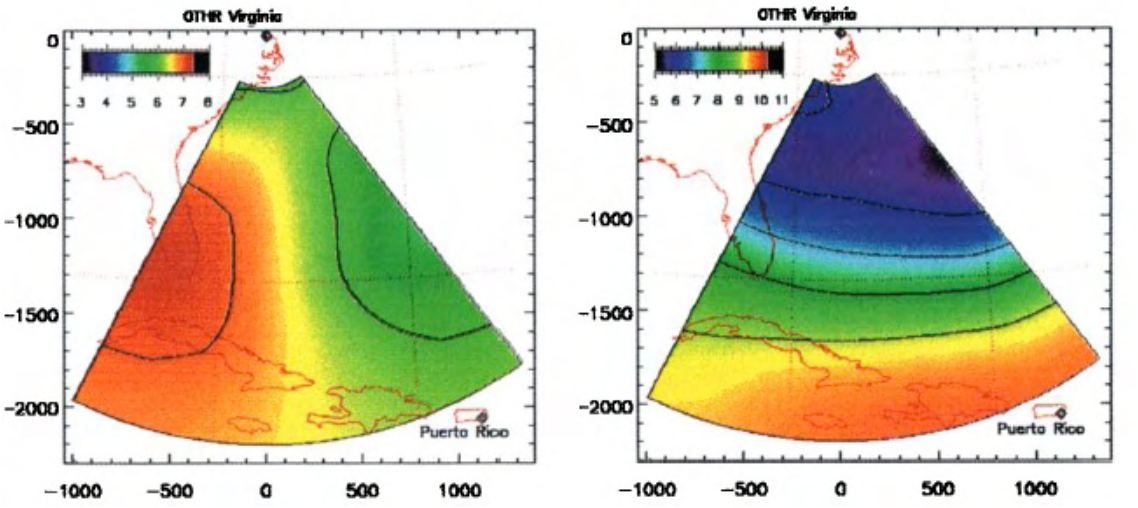


Figure 2.9: Inversion result for the plasma frequency in MHz at 235 km altitude on December 8, 1994, 2244 UT (left) and on August 5, 1994, 2257 UT at 260 km (right). Source: *Fridman (1998)*.

TEC measurements by satellites

Beginning in 1960, rockets and satellites came into use and enabled to study the ionosphere from above, instead of using instruments on Earth. Shortly after the launch of the first man-made satellite Sputnik 1 by the Soviet Union in 1957, the development of the world's first navigation satellite system TRANSIT for the U.S. Navy started. The first navigation satellite of this system, which consisted of 15 navigation satellites and eight related research satellites, was launched in 1959. After more than 32 years of continuous successful service to the U.S. Navy, the TRANSIT system was replaced by the American Global Positioning System (GPS) and is now used as the Navy Ionospheric Monitoring System (NIMS).

GPS developed quickly for military purposes thereafter and was a great improvement with higher accuracy and stable atomic clocks on board to achieve precise time transfer. However, it wasn't until a civilian Korean Air airplane carrying 269 passengers, was shot down after mistakenly entering Soviet airspace that the Reagan

Administration in the US had opened up GPS for civilian applications. So aircraft, shipping, and transport could fix their positions and avoid straying into restricted foreign territory.

Since December 2012, the GPS system consists of 32 satellites arranged in orbital planes at approximately 20 200 km height. All of these satellites broadcast at the same two frequencies in the L-band, $f_1 = 1575.42\,\mathrm{MHz}$ and $f_2 = 1227.60\,\mathrm{MHz}$. By receiving the transmitted signals from several of these satellites, a GPS receiver on Earth can accurately determine its position. Although initially intended for military applications, GPS is considered a dual-use technology, having significant military and civilian applications, *e.g.* cartography, tectonics, navigation, and ionospheric science.

An alternative to the American GPS is the Russian system Global'naya Navigatsionnaya Sputnikovaya Sistema (GLONASS). Both constellations, which are fully operational and have global coverage, are generally described as Global Navigation Satellite System (GNSS). Other GNSS are under development, for instance the European Union Galileo or the Chinese COMPASS positioning system, which is an expansion of the existing Chinese BeiDou-1 satellite system. Apart from global positioning systems, also regional networks like the the Indian Regional Navigation Satellite System (IRNSS) and the Japanese Quasi-Zenith Satellite System (QZSS) are under development.

Measurements with GNSS using two or more frequencies are employed to obtain the TEC on the Line Of Sight (LOS) between transmitter and receiver, regardless of the receiver being a stationary ground station or on board of a satellite. The TEC is proportional to the phase difference introduced by the ionosphere between the two signals and is defined as the integral of electron density N_e along the ray path between satellite and receiver. Usually it is visualized as global or regional two-dimensional map to show the state of ionosphere *(Mannucci et al., 1998)*. The TEC can be inverted for the local electron density and then used in ionospheric tomography (Section 4.2).

The vertical resolution in ionospheric tomography with GNSS is usually poor due to the absence of horizontal ray paths. This problem can be addressed by including TEC occultation data. In GPS occultation measurements, the signal emitted by the GPS satellites is received by a receiver on a Low Earth Orbit (LEO) satellite, orbiting around the Earth at approximately 160 km altitude (Figure 2.10). Along its way through the ionosphere, the signal has been refracted due to the presence of free electrons with the amplitude of refraction depending on the electron density. Since both satellites move over time, this technique allows a vertical scanning of successive layers of the atmosphere. The LEO satellites Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC) and CHAllenging Minisatellite

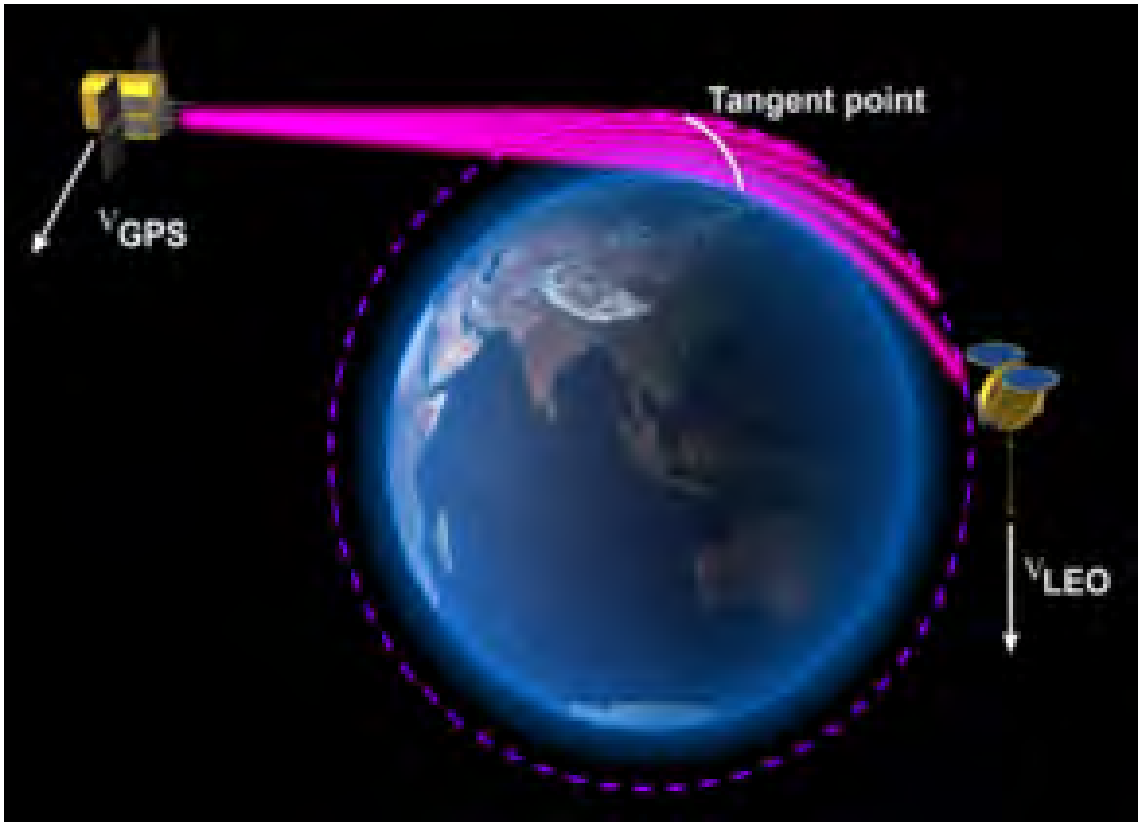


Figure 2.10: Principle of GPS occultation measurement. The LEO satellite receives a signal from the GPS satellite that has been refracted along its way through the ionosphere. Source: `https://www.image.ucar.edu/DAReS/DART/Research/GPS_Liu/`

Payload (CHAMP) are used in GPS radio occultation measurements.

The installation of global (*e.g.*, EUREF Permanent Network (EPN), International GNSS Service (IGS)) and regional (*e.g.*, GPS Earth Observation Network (GEONET) in Japan) ground-based GPS networks greatly increased the amount of available data and ray-path coverage, allowing a 3D reconstruction of the ionosphere with time evolution (3D+1). The IGS collects, archives, and distributes GPS observation data sets from an international network of more than 350 continuously operating dual-frequency GPS stations. The generated orbit and tracking raw data are sent to Operational Data Centers, were they are formatted and forwarded to regional or Global Data Centers, which make the data online available. Apart from tracking data and satellite ephemerides, the IGS provides products such as global ionospheric maps.

GEONET is a Japanese GPS station network operated by the Geospatial Information Authority of Japan and consists of approximately 1240 GNSS sites with an average spatial resolution of 25 km *(Seemala et al., 2014)*. Its equivalent in Europe is the EPN with 247 permanently operating GNSS reference stations, which also provide data in real-time.

Chapter 3

Over-the-horizon radars

3.1 Introduction

All kinds of radar transmit EM signals that are scattered at a "target", with a part of the signal energy returning to the emitter, where it is detected by a receiver. Comparing the energies returned and transmitted and measuring the propagation time allows determination of the target range, as well as its direction, corresponding to the radar emission direction. Hence, the acronym RADAR stands for RAdio Detection And Ranging. The velocity of an object can be computed based on the doppler shift of a signal that is scattered at an object.

The distance over which a radar works efficiently is limited by obstacles and, especially, by the curvature of the Earth. The problem of detecting a target at very long range, *i.e.*, more than 1000 km away from the transmitter, can be solved by using the reflective properties of the ionospheric plasma located more than 100 km above the ground.

For this purpose powerful Over-the-horizon (OTH) radars using the ionospheric refraction properties on EM waves in the HF band (3−30 MHz) have been developed. There are two types of OTH radars, namely OTH sky wave and OTH surface wave radar. As the names suggests, the EM waves propagate in the ionosphere or along the ground and the sea surface, respectively. In this thesis, only data obtained from the OTH sky wave radar are used, so this kind of radar will be described more detailed.

As described in section 2.5, EM waves are refracted in the ionosphere due to the presence of free electrons. Since ionization increases with altitude up to 300 km, HF EM waves are bent towards the ground to locations beyond the geometric horizon, typically up to thousands of kilometers away from the transmitter. Therefore, this type of radar is called *Over-the-horizon radar*. Figure 3.1 illustrates the difference between a classical radar and an OTH sky wave radar.

The distance reached with an OTH radar depends on the signal frequency and the elevation angle of the radar, but for a single-path the signal may be backscattered at about 2000 km, and for multi-path at about 4000 km. A detailed description of the working principle of OTH radar is given in *Headrick and Anderson (2008)*.

OTH radars can be monostatic or bistatic. For monostatic OTH radar, transmitter and receiver are located at the same place, and signals emitted by the radar are refracted by the ionosphere and then reach the ground, where they are backscattered in all directions, especially in the incident direction. Presumably they follow the same path back to the radar, where they can be detected. In addition to information about the target, the received signal contains information about the medium it propagated through, so it can be used to study phenomena in its interior. For bistatic OTH radars emitter and receiver are located at different places, and the signal transmitted at one site is detected by a receiver several hundred kilometer away.

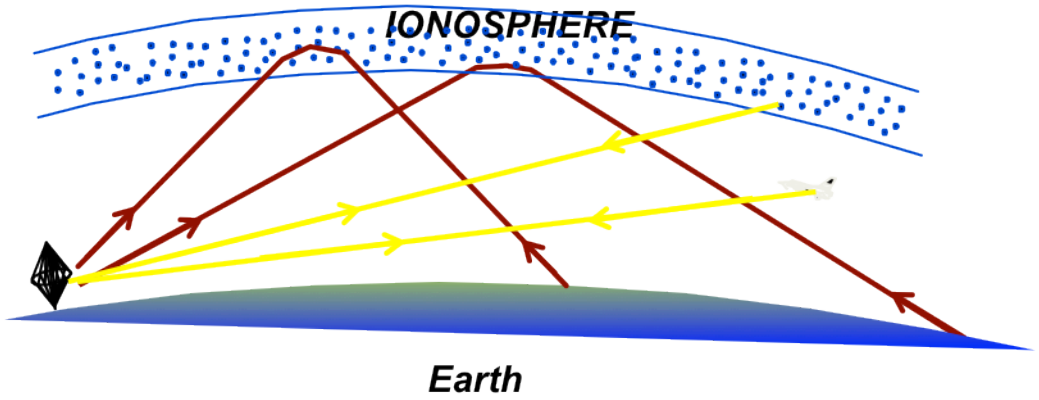


Figure 3.1: Schematic representation of OTH (red) and classical radar (yellow). Source: *Molinié (2011)*

3.2 OTH radar worldwide

Since World War II, the request for identifying a target without being limited by the curvature of the Earth lead to the development of OTH systems all over the world. Some of these systems are described in the following sections, and an exhaustive description can be found in *Liu (2007)*.

OTH systems in the United States of America (USA) are described in detail by *Thomason (2003)*. The first experimental monostatic radar, Magnetic-Drum Radar Equipment (MADRE), was built in the late 1950s by the Naval Research Laboratory (NRL) in Chesapeake, Virginia. Data were recorded on magnetic drum devices (hence the name). Another experimental OTH radar built in the United

States is the Wide-Aperture Research Facility (WARF) in central California. This bistatic radar represents a major improvement as compared to MADRE and its large aperture results in improved azimuthal resolutions *(Thomason, 2003)*. The first attempt of establishing an operational OTH radar was the Anglo-American project AN/FPS-95, also known as Cobra Mist. Unfortunately, after two years of testing the project was cancelled, because the radar did not perform as expected and resolutions in range and azimuth were poor. The second operational US OTH radar was Over-the-horizon-Backscatter (OTH-B) that consists of two bistatic systems located at the East and the West Coast of the US, respectively. The east system transmitter was located at Moscow (Maine) and the receiver in Columbia (Maine). The west system had its transmitter in Christmas Valley (Oregon) and its receiver at Tulelake (California). In 2002, the West coast facilities were downgraded to "cold storage" status and finally, the antenna arrays have been pulled down and removed in 2007.

The US Navy created its own system, the Relocatable Over The Horizon Radar (ROTHR), which was originally intended to keep track of ship and aircraft movements over the Atlantic. The prototype was installed on the Aleutian Island and monitored the eastern coast of Russia between 1991 and 1993. Later, it was moved to Virginia (ROTHR-VA) in order to control illegal drug trade in central America and the Caribbean. Subsequently, additional ROTHRs were installed in Texas (ROTHR-TX) and in Puerto Rico (ROTHR-PR).

In Australia, Jindalee was the first bistatic OTH radar project and started in the 1960s. Transmitter and receiver are located separately in central Australia at Alice Springs. In 2003, Jindalee was combined with two other OTH systems. The first one near Laverton in Western Australia and the other one at Longreach, Queensland, in Eastern Australia. This system is now known as Jindalee Operational Radar Network (JORN). *Colegrove (2000)* reported on the development of JORN in detail.

In Russia, the first OTH radar system was DUGA-2 starting in 1971. It was followed by the operational system STEEL YARD, also known as Russian Woodspeeker due to its loud and repetitive pulses in the middle of the shortwave radio bands. In 2014, the Russians announced a new system named Container, capable of reaching the Netherlands *(Karpenko, 2014)*.

There are also OTH radars in China, Iran, and Japan, but only few details are known regarding these systems. According to a report, China may have as many as three OTH radars used in an early warning system and one OTH-B radar to provide surveillance of the South China Sea.

In France, two OTH radars are installed. One in Valensole that served for ionospheric research and operated by the University Pierre et Marie Curie *(Six et al., 1996)* and another one Nostradamus located south of Paris, which will be

described in detail in the following section.

3.3 SuperDARN

Super Dual Auroral Radar Network (SuperDARN) is a global scale network of HF and Very High Frequency (VHF) radars, constructed by engineers and scientists of a dozen different countries. Radars are located at mid-high latitudes with fields of view covering the polar regions for the study of ionospheric physics. The primary objective of SuperDARN is to provide direct continuous global-scale observations of the motion of irregularities in the ionospheric plasma density at middle ($30° - 55°$) and high ($> 55°$) latitudes *(Greenwald et al., 1995)*.

As of 2013, this network consists of 21 continuously operating radars in the northern hemisphere and 11 radars in the southern hemisphere (Figure 3.2). It is still expanding with new radars joining the network almost yearly. The view field of the radar network covers vast regions of ocean near the poles, which play a crucial role in global climate variability. All radars have an azimuthal resolution of approximately 4° for a transmission frequency of 12 MHz *(Greenwald et al., 1995)*.

SuperDARN has been successful in addressing a wide range of scientific questions concerning processes in the magnetosphere, ionosphere, and thermosphere. Radars measure the backscattered energy, the line-of sight Doppler velocity, and the width of the Doppler spectrum *(Baker et al., 2010)*. The Doppler motion of ionospheric irregularities can be used to infer the strength and direction of the ionospheric electric field as well as to calculate convection maps of ionospheric plasma (plasma convection velocity) on a global scale every 1-2 minutes. Convection maps are important diagnostics of energy transfer from the solar wind into the magnetosphere and ionosphere.

Apart from that, scientific objectives of SuperDARN are the investigation of large scale waves, for instance gravity waves, the motion of ionospheric plasma, ionospheric irregularities, and high-latitude plasma structures. In addition, the roughness at the Earth surface including ocean waves and ice cover can be investigated with SuperDARN, since the scatter from the ground is most intense from water-covered areas and almost extinguished over the ice cap of central Greenland.

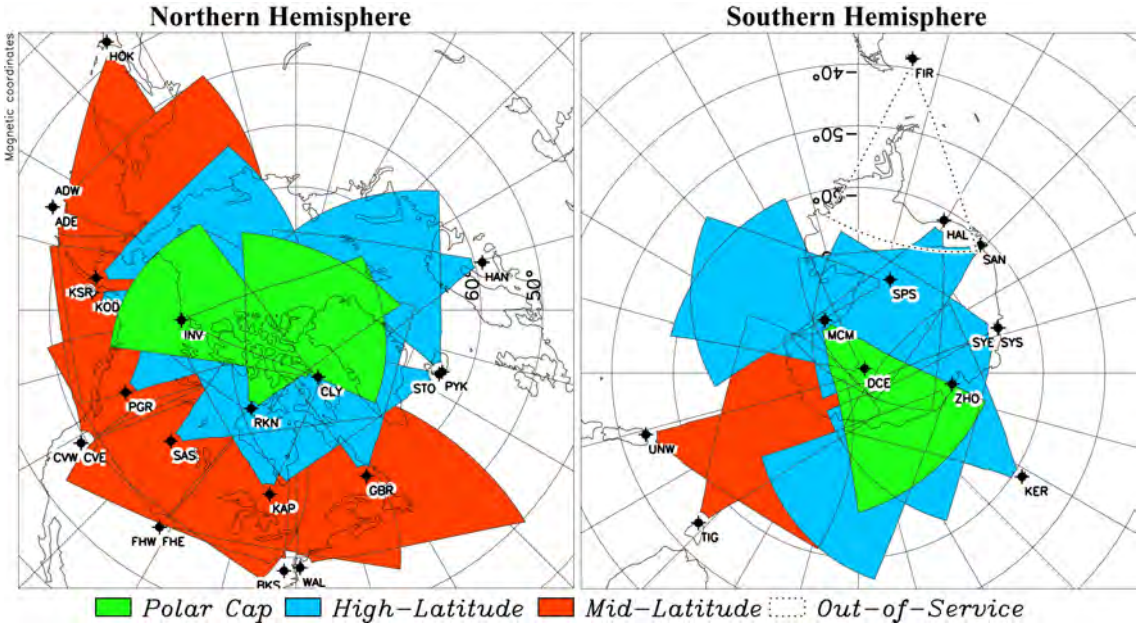


Figure 3.2: Fields of view of currently operational SuperDARN radars in the northern (left) and southern hemisphere (right). Source: `http://superdarn.org/tiki-index.php`.

3.4 Nostradamus

Nostradamus, an OTH radar operated by the Office National d'Etudes et de Recherche Aérospatiale (ONERA), is located 80 km south of Paris at Dreux. In contrast to other OTH radars *e.g.*, JORN or ROTHR, this radar is monostatic that means the antennas for emission and reception are at the same location.

This french OTH radar consists of 288 biconical antenna elements distributed randomly over the arms of a three-branch star spaced by 120°. Each antenna is 7 meter high and 6 meters wide (Figure 3.3). This star-shaped antenna arrangement allows a 360° coverage in azimuth, and the biconical form of each antenna enables the control of the transmission beam in elevation. These unique characteristics of Nostradamus permit the investigation of a very large area of more than 2000 km range (Figure 3.4) all around Europe. Only the central part of the array (96 antennas) is dedicated to transmission and reception, the entire array is used for reception allowing a great capability in receiving beam forming.

Despite the vast sounded area, the radar conserves good resolution in reception with, for example, approximately 2.35° in azimuth and 5.43° in elevation for a frequency of 11 MHz, because the number of antennas used for receiving is three times larger than that used for emission. The receiving array possesses a broad aperture allowing the scattered signals to be resolved into fine azimuth cells. In addition, timing of the received signal allows resolution of the signal into range cells. The resolution in azimuth and elevation is determined by the beam aperture at -

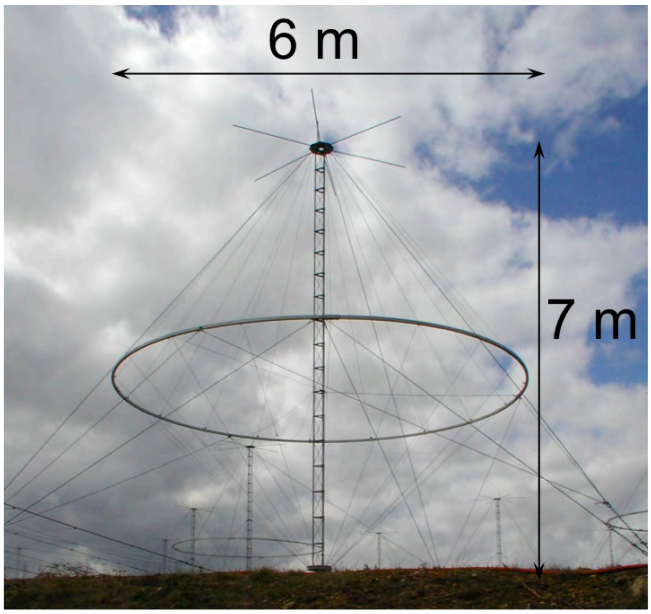


Figure 3.3: Example of a biconical antenna of Nostradamus. Source: *Molinié (2011)*

3 dB below the maximum. This angle characterizes the angle range of the antenna, into which at least half of the maximal power is radiated. Azimuth and range of the returned signal can be determined with a certain resolution depending on the frequency of the signal, since the antenna beam width is frequency depending. In general, resolution in elevation is better for higher frequencies.

The returned signal can therefore be separated into azimuth-range cells, as shown in Figure 3.4). The resulting range-azimuth resolution cell pattern (Figure 3.4, right) is then treated as a search plane for targets, which would manifest as local maxima of received signal power in a cell relative to the surrounding cells. Local maxima are declared as detections. To separate moving targets from the clutter echo a Doppler processing is used. Tracking the location of these detections over time provides target trajectories. Since the amount of data was very large at the time of construction, the antennas are assembled in sub-arrays in order to reduce the data quantity in the computer. The signal from each sub-array is digitalized and merged together giving a receiving beam.

Nostradamus has two different operating modes - the radar and the sounding mode. Due to its unique antenna arrangement, Nostradamus allows to do backscatter sounding by frequency sweep (**BSS**) and backscatter sounding by elevation scanning (**ELS**). In the **BSS** method, a signal is transmitted at a fixed azimuth and elevation angle and scanned in frequencies between 6 and 28 MHz. A measurement for one frequency and one elevation angle takes around 3.5 s, and a complete measurement (all frequencies and elevation angles) requires approximately 21 min. After digitalization of the received signals, cross-correlation with the emitted sig-

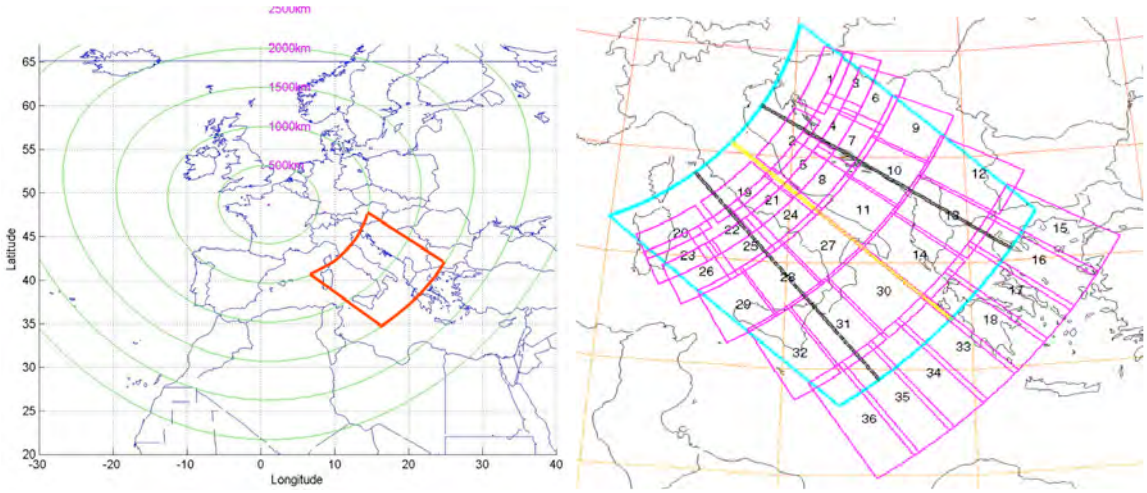


Figure 3.4: Simulations for the capacities of OTH radar Nostradamus: Reached distances ranges (left) and range-azimuth resolution cell pattern example for the red square in the left figure (right). Source: *Molinié (2011)*.

nal and a Doppler processing, a graph elevation angle as function of group path is obtained. Since Nostradamus is also used for military applications, information about the backscattered energy are confidential. In the **ELS** operating mode, the radar scans in elevation, but at fixed frequency and azimuth. The radar can perform a panoramic sounding, where the azimuth varies from 0° to 360° for a given frequency and elevation angle, as exemplified in Figure 3.5. In this radar mode, the first objective is the determination of range and distance of targets.

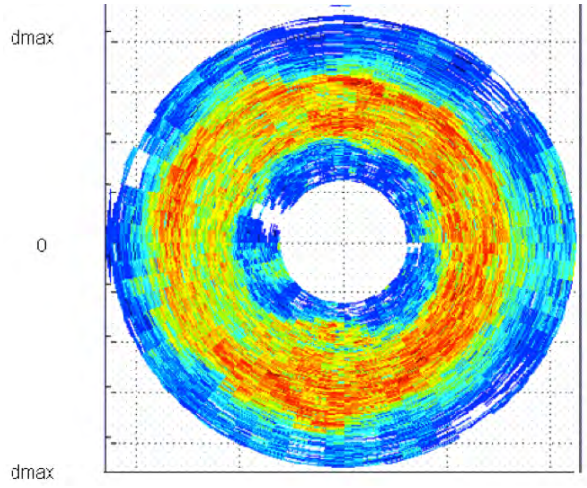


Figure 3.5: Variability of the ionosphere: Panoramic sounding for one frequency and one elevation angle. Colors represent the amplitude of the backscatter echo. Axis labels are not shown due to the confidentiality of the information about the distance reached by the radar. Source: *Bazin et al. (2006)*

3.5 Scientific applications of OTH radar

Although applications of OTH radar are mainly military and defense-orientated (tracking boats and planes, detection of targets), these radars have scientific relevance, too, as already demonstrated by SuperDARN. As mentioned above, the propagation of HF radio waves in the ionosphere depends on the properties of the ionospheric plasma, therefore, OTH radar can be used to study the ionosphere and phenomena occurring inside. Apart from ionospheric science, OTH radars have applications in several different scientific fields.

One of these fields is oceanography, where this radar is very useful to map surface wind directions and ocean currents. Winds over the ocean generate waves on the water that can be identified by HF OTH radar. The Doppler frequency spectrum produced by echoes from water waves, allows determination of the direction of the waves generated by the wind and hence the direction of the wind itself. In addition, the strength of the waves (indicating the state of the sea, or roughness) can be ascertained *(Young et al., 1997)*.

According to the same principle, it is possible to map ocean currents. This method is based on the fact that HF signals are primarily backscattered from *Bragg* resonant surface waves. These are waves having wavelengths which are exactly half of the wavelength of the radio signal. The echo from the sea surface consists of two sharp *Bragg* lines systematically centered on either side of the zero Doppler if no currents are present. The Doppler position corresponds to the velocity of the waves. Otherwise, in the case of underlying surface currents, the *Bragg* lines are no longer

symmetric to zero Doppler, since the current changes the wave velocity *(Georges et al., 1998)*.

Also in astrophysics OTH radar is useful. *Thomas and Netherway (1989)* showed the potential to use OTH radar for the observation of meteors. In ionospheric science, *Occhipinti et al. (2010)* and *Coïsson et al. (2011)* showed that perturbations in the ionospheric plasma related to Acoustic Gravity Waves (AGW) and Internal Gravity Wave (IGW) excited by tsunamis or earthquakes associated Rayleigh waves can be detected using backscatter echoes of monostatic OTH radar. In this thesis the possibility of OTH radar to produce a 3D image of electron density in the ionosphere is discussed.

Chapter 4

Ionospheric tomography

4.1 Tomography

The term tomography is derived from the ancient Greek words τοΐμος (slice) and γράΦω (to write). It denotes a technique that allows to reconstruct the volume of an object (a human body in the case of medical imaging or a geologic structure in geophysics) from a series of measurements taken from outside the object. It is a non-invasive technique enabling the visualization of internal structures without destroying the object. These measurements result in a reconstruction of certain properties within the object depending on the type of information provided by the sensors (sound pressure, attenuation of a light beam, variation in velocity or polarization of seismic waves).

From a mathematical point of view, tomography consists of two stages. At first, it requires the development of a forward model describing accurately the physical phenomena that are measured. Then, in a second step, the model or 3D distribution of the properties within the object is determined, based on the forward model.

A simple example explaining tomography is a Sudoku. The objective is to fill in a 9×9 grid with digits so that each column and each row contains all numbers from 1 to 9. The forward model is to calculate the sum of each row and each column. Knowing the sums of each row and column, the inverse model is to complete the grid.

In medical tomography physicians use X-rays ($\lambda = 0.01$-$10\,\mathrm{nm}$) sent in all directions through the body of a patient in order to have a look at organs and bones without surgery. X-rays passing through the patient are attenuated and its intensity is reduced. Repeating this procedure for different angles gives information on how the inside looks like. Computers then combine these images into a 3D picture of the body.

Tomography has many other applications, which are quite different from those

in medicine. An interesting example comes from archaeology, where tomography was used to determine the cause of death of the Pharao Tutankhamun *(Hawass, 2005)*. In seismology, seismic waves propagating through the Earth provide, for instance, information about its 3D velocity structure (*e.g., Wawerzinek et al. (2013)*, Appendix A).

In analogy to the examples given above, it is possible to monitor the state of the ionosphere using tomography. Radio waves propagating through the ionosphere are refracted due to the presence of free electrons, and the measured time delay with respect to an ionospheric model can be used to obtain information on the electron density distribution. Most of ionospheric tomography methods use signals emitted by GPS satellites and received on the ground. In the following section, this method as well as its limitations are discussed in detail, followed by a description of the new developed method of ionospheric tomography by OTH radar.

4.2 Ionospheric tomography and the role of GPS

Although initially intended for military applications, GPS is a dual-use technology, meaning that it has significant military and civilian applications. Since GPS data became available for scientific applications, it can be used for ionospheric tomography. Figure 4.1 illustrates the principle of GPS tomography. The GPS satellite emits two signals with the frequencies $f_1 = 1575.42\,\mathrm{MHz}$ and $f_2 = 1227.60\,\mathrm{MHz}$ that traverse the ionosphere and are recorded at the ground stations.

As described above and in Section 3.1, EM waves are perturbed when traveling through the ionosphere. The ionospheric refraction causes amongst others phase delays of the signals, which are measured at the ground GPS receivers. The differential phase shift encountered by dual-frequency satellite signals for each satellite location and receiver pair is proportional to the relative TEC along the corresponding ray-path. The phase difference $\Delta\phi$ between the two signals is correlates to the integral of electron density N_e along the ray path between satellite and receiver by

$$\Delta\phi \sim \left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right) \int N_e \, ds. \tag{4.1}$$

The sum of electron density N_e along the ray-path is called Slant Total Electron Content (STEC)

$$\mathrm{STEC} = \int N_e \, ds, \tag{4.2}$$

where ds is the ray-path. The measured TEC along with some *a priori* information about the background ionospheric model allows estimation of the local electron density of the ionosphere. Despite of its indirectness, this method has the capability

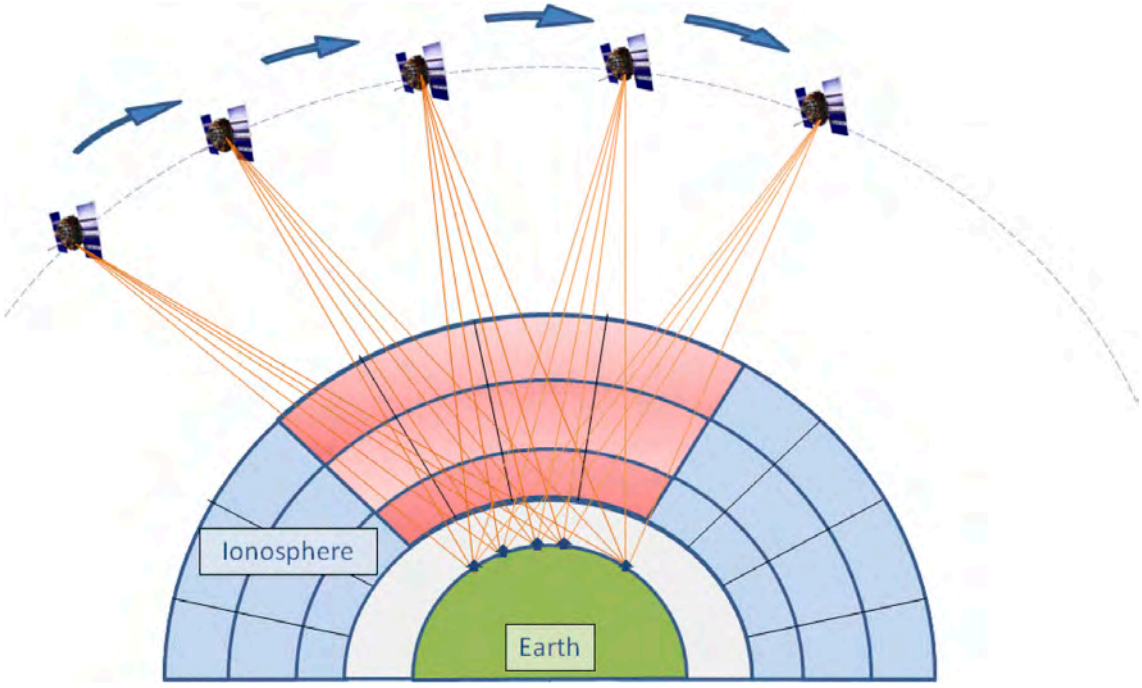


Figure 4.1: Principle of GPS ionospheric tomography. The GPS satellite orbiting around the Earth is emitting signals that are measured at several receivers at the ground. The measured STEC can be inverted for the local electron density perturbation in the red area of the ionosphere. Source: `http://gnss.be/ionosphere_tutorial.php`

to provide a complete multidimensional image. This is an advantage as compared to other measuring techniques explained before.

First simulations of the feasibility of GPS topography were carried out by *Austen et al. (1988)* using TEC values computed for a realistic satellite-receiver configuration and applying the simultaneous iterative reconstructon technique (SIRT) *(Gilbert, 1972)* algorithm for image reconstruction. He successfully showed that numerical tomography techniques can be used to produce two-dimensional vertical cross-sections of the electron density in the ionosphere. Several other numerical tests have been carried out by others trying different algorithm for image reconstruction *(Raymund et al., 1990; Raymund, 1994)*, until the first experimental measurements were reported.

Andreeva et al. (1990) published the first experimental results using TEC data collected at three receivers in Russia and reconstructed the first two-dimensional vertical cross-section of the ionosphere by satellite radio tomography. Since then, GPS-based computerized ionospheric tomography has been widely used to investigate the temporal and spatial variations of ionospheric structures *(Hansen et al., 1997; Hajj et al., 1994; Hernández-Pajares et al., 1998; Bust et al., 2004; Garcia and Crespon, 2008; Ma et al., 2005; Yizengaw et al., 2005; Wen et al., 2007)*. In addition, the installation of global (*e.g.*, IGS) and regional (*e.g.*, GEONET in Japan)

ground-based GPS networks enabled ionospheric tomography worldwide. In Europe *(Mitchell and Spencer, 2003)*, Russia *(Kunitsyn et al., 1994)*, Antartica *(Heaton et al., 1996)*, China *(Wen et al., 2007)* as well as at the polar caps *(Pokhotelov et al., 2011)*.

Although the amount of available data is huge due to a multitude of satellites and receivers, ionospheric tomography is an ill-posed problem. In general, the reconstruction of images from measurements is an inverse problem that is often ill-posed because measurements are both an incomplete sampling of information and often corrupted by noise. As a consequence, the solution is often unstable (because it is very sensitive to small data changes) and not unique (more than one reconstruction fits the observed data).

The literature of the last years reveals a lot of different reconstruction algorithms tested for ionospheric tomography to address these problems of instability and non-uniqueness. *Kersley et al. (1993); Heaton et al. (1995); Mitchell et al. (1995); Pryse et al. (1995); Vasicek and Kronschnabl (1995)* applied the iterative Multiplicative Algebraic Reconstruction Technique (MART) *(Gordon et al., 1970)* algorithm that attempts to minimize differences between measured and calculated TEC values by modifying the background ionosphere until the differences are acceptably small.

The Algebraic Reconstruction Technique (ART) incorporates some prior information on each pixel in an ionospheric grid and was used by *Kunitsyn et al. (1994,?)*. However, the solution obtained with this algorithm is, considerably limited, because it is very sensitive to the initial ionospheric model, used for those cells in the tomographic grid, which are not hit by any ray. To cope with this problem, *Wen et al. (2010)* recently proposed the constrained algebraic reconstruction technique (CART) algorithm, where cells not hit by any ray extract information from their neighbors. They validated their algorithm with numerical simulations and finally applied it for the reconstruction of the electron density distribution over China. *Kunitake et al. (1995)* successfully used a modified version of singular value decomposition to reconstruct the TEC over Japan in a magnetically disturbed period.

Validation of tomographic images

Tomographic images need to be verified to be sure that the obtained image corresponds to the reality. *Pryse and Kersley (1992)* reported the first observation regarding the verification of a tomographic image by another instrument. They used TEC measurements collected at two receivers in Scandinavia and verified their tomographic image obtained with a simple SIRT algorithm by comparison with measurements of the EISCAT radar. After this first successful implementation of the EISCAT radar in radio tomography, it served also in other experimental results for

validation for tomographic images of electron density in Scandinavia *(Nygren et al., 1997; Walker et al., 1997)*.

Other authors compared peak electron densities and the height of the layer of maximum electron density from tomographic images with ionosonde data. *Fehmers et al. (1998)* show in their ionospheric tomography experiment over midlatitude Europe that the tomography undervalues the maximum electron density $N_{e,max}$ by 20 % and overvalues the height of the layer of maximum electron density between $80 - 200\,\mathrm{km}$.

Dear and Mitchell (2007) observed that the peak electron density in simulation studies is underestimated by 15 % and up to 40 % for real GPS data by their tomographic reconstruction with Multi-Instrument Data Analysis System (MIDAS) compared to ionosonde data. They explain the underestimation by a poor representation of the vertical profile in the simplified set of basis functions they used in the reconstruction algorithm to constrain the result. They show that incorporation of the F2 peak height from two ionosondes and a range of scale heights as input in the reconstruction algorithm improves the result giving a better match with the peak electron density of the ionosondes.

Limitations

Although GPS is a powerful tool for studying the ionosphere, theoretical limitations of ionospheric tomography using a LEO satellite-to-Earth configuration have been discussed in detail by *Yeh and Raymund (1991)* and *Na and Sutton (1994)*. Apart from spatial and temporal limitations associated with the experimental arrangements, data sets are often incomplete because of the finite sampling interval in time and a limited view angle of each receiver. Furthermore, an invariant ionosphere during the time of measurement was assumed. The major problem of ionospheric tomography using GPS is the absence of horizontal ray-paths resulting in low vertical resolution. Finally, the high frequency of the GPS signal limits the sensitivity of the EM waves to the maximum of electron density in the ionosphere that is the F-region.

Three approaches are described to overcome the limitations in vertical resolution due to the absence of horizontal ray-paths. The first approach is to incorporate extra information from other experiments, for instance ionosonde data, which give information on the lower vertical profile. *Heaton et al. (1995)* tested the incorporation of scaled ionograms into the imaging and found improvement in the vertical profiles. *Kersley et al. (1993)* showed promising results with the incorporation of ionosonde data into the reconstruction algorithm. *Markkanen et al. (1995)* applied a Bayesian approach to simulated results and incorporated peak heights as a priori

information in the reconstruction algorithm.

The second approach is to make use of the general knowledge regarding the shape of ionospheric profiles to constrain the reconstruction results and fill the information gap. The most widely used solution is that proposed by *Fremouw et al. (1992)*, who applied a set of vertical orthonormal functions, created from ionospheric models to image the vertical profile. However, existing ionospheric models are partially incorrect, because they fail to predict the strong day-to-day variations. Therefore, reconstructions constrained with vertical profiles from ionospheric models are often inaccurate.

Fehmers et al. (1998) proposed a model-independent algorithm compensating the lack of horizontal ray paths with information that does not depend on a specific model. They impose that the electron density cannot be negative at high and low altitudes, and assumed it to be smooth and vertically stratified. These constraints are incorporated in the optimization problem, and the mathematical form of the problems allows additionally incorporation of ionosonde data as constraints in the optimization problem. Their tests show moderate sucess with an error in the layer-height estimation in the order of 90 km.

The third approach is to combine measurements from different sources in the inversion, *i.e.*, from ionosondes, radar or occultation measurements. For instance, *Rius et al. (1997)* showed that combination of GPS TEC and occultation data improved the vertical resolution, whereas the use of ground data alone is insufficient for vertical reconstruction of electron density.

Fridman and Nickisch (2001) developed a method for simultaneous inversion of vertical and backscatter ionograms from OTH radar with TEC measurements collected by receivers operated in the Caribbean to obtain a 3D, smooth ionospheric model over a fixed geographical region. *Fridman et al. (2006)* and *Fridman et al. (2009)* developed a 3D real-time reconstruction method of the ionosphere (GPSII), that is able to use absolute and relative TEC data from ground and space GPS receivers including occultation, vertical TEC from altimeters, and in situ measurements on LEO satellites such as CHAMP as well as electron density profiles from vertical sounders. Finally, *Fridman et al. (2012)* upgraded their GPSII method enabling the incorporation of oblique backscatter sounding data. The combined algorithm produces a dynamic model of electron density for a fixed geographical region. This model is consistent with backscatter leading edge data, vertical sounding data as well as with absolute and relative TEC data from a number of GPS/LEO receivers.

With the installation of global and regional GPS networks the amount of data increased and the angular as well as the ray path coverage improved, making the 3D reconstruction of the ionosphere with a time evolution (3D+1) possible. Over Scan-

dinavia, *van de Kamp (2013)* imaged the ionosphere in 3D for the entire December 2006 using TEC measurements performed at the GPS network GEOTRIM in Finland and together with EISCAT radar data as input to constrain the vertical profile. Their inversion results show that the inversion can better resolve vertical profiles of irregular structures when using the profiles from EISCAT instead of Chapman profiles in the inversion process.

Mapping of disturbances in the ionosphere

Mitchell et al. (1995) have shown the ability of tomography to image large-scale ionisation depletions, which are known as troughs and generally found on the nightside auroral mid-latitude boundary. Since then, tomographic imaging has been used by several authors to image troughs in the ionosphere. Thus, *Heaton et al. (1996)* imaged a trough over Antarctica for the first time and *Kersley et al. (1997)* used ionospheric tomography to map a trough over the U.K and Scandinavia.

However, several papers based on experimental results showing tomographic images of Traveling Ionospheric Disturbances (TID) have been published (*e.g.*, *Cook and Close (1995); Markkanen et al. (1995); Pryse et al. (1995).* TIDs are large-scale irregularities in electron density, which travel in the form of a wave over large distances without significant change of their shape. Large-scale TIDs have periods of the order of one hour, wavelengths of about 1000 km, and horizontal speeds greater than $250\,\mathrm{m\,s^{-1}}$ *(Schunk and Nagy, 2009, Chapter 11).*

They are gravity waves generated usually by temperature change, interaction with the wind with the topography, geomagnetic storms, tsunamis, tropical storms, etc. Although various authors pointed out that tomographic methods cannot resolve TID propagating in the ionosphere, *Nygren et al. (1997)* showed in their tomographic experiment at four receiving stations in Scandinavia that the tomography is unable to resolve wave fronts lying northward of the receiver chain and explained this with rays being perpendicular to the wave fronts. *Saksman et al. (1997)* demonstrated mathematically and by simulation how a wave-like TID may be partly invisible to transmitters on the ground due to the experimental setup.

4.3 Summary

Different methods of ionospheric tomography using GPS were described in the previous sections and their potential and limitations were discussed. Due to high availability of stations and data this method is very powerful and enables a worldwide imaging of the ionospheric plasma in 3D with time evolution and the detection of perturbations in the ionosphere. Nevertheless, ionospheric tomography by GPS is an ill-posed problem due to spatial and temporal limitations associated with the

experimental arrangements, incomplete data sets and limited view angle of each receiver that has been addressed in several papers in the past years.

Strong limitations of GPS tomography are the poor vertical resolution due to the absence of horizontal ray-path and its sensitivity to the region of maximum ionization in the ionosphere because of the high frequency of the GPS signal. In the next sections a new ionospheric tomography method based on OTH radar will be presented to overcome some limitations.

4.4 Ray tracing

The first step in developing a tomography method for OTH radar requires a modeling of the propagation of EM waves in the ionospheric plasma. There are a number of studies for modeling the propagation of rays in the plasma, all of them rely on different hypothesis. Most of these studies assume an isotropic ionosphere neglecting the Earth's magnetic field and losses in the ionosphere, others are limited to 2D modeling and based on the assumption that the azimuth does not change along the ray-path. There are two main types of ray tracing in the ionosphere, called numerical and analytical ray tracing.

Analytical ray tracing mostly applies to a spherically symmetrical ionosphere, whose profiles are described by QP or quasi-cubic (QC) segments. As its name suggests, analytical ray tracing uses explicit equations to define the ionosphere and to determine ray parameters such as ground range, reflection height, phase, and group path. Consequently, analytical ray tracing can only be applied to some ionospheric models, but it is much faster than numerical ray tracing. This makes the method particularly valuable in HF applications, where a large number of rays has to be traced. On the other hand, it is difficult to include the effects of the 3D Earth's magnetic field, horizontal gradients, and the ellipticity of the Earth in analytical ray tracing. Thus, analytic ray tracing is limited to simple and unrealistic ionospheric models. Nevertheless, *Norman and Cannon (1997)* developed an analytical ray tracing code that takes into account horizontal gradients by segmenting the ionosphere horizontally as well as vertically.

Numerical ray tracing usually integrates the position and the ray direction for each ray point from the equations, which describe the propagation of waves in a medium. The most widely used ray tracing technique in ionospheric science is that described by *Jones and Stephenson (1975)*. It allows to trace rays in several different ionospheric models taking into account the Earth's magnetic field, electron collision, and absorption. *Tsai et al. (2010)* developed a 3D numerical ray tracing technique on a phenomenological ionospheric model for Taiwan, considering the Earth's magnetic field and horizontal gradients in order to simulate synthetic ionograms. More

recently, *Azzarone et al. (2012)* developed a MATLAB software, IONORT, which allows ionospheric ray tracing in a geocentric spherical coordinate system, taking into account the geomagnetic field and particle-electron collisions. Other numerical ray-tracing programs have been developed by different groups, but they are not necessarily freely available.

Although a lot of methods exists, none of them has been used for this thesis, but the code Tracé De Rayons (TDR) developed by *Occhipinti (2006)* has been implemented and was improved and developed further during this research work. This code traces rays in a 3D heterogeneous ionospheric model starting at a given point on the ground or into the space in the WGS84 coordinate system *(National Imagery and Mapping Agency, 2000)*. In this work the starting point is always considered as the geographical position of the OTH radar Nostradamus (lat: 48.64°N, lon: 1.08°E). The code neglects the Earth's magnetic field, since its application was tailored for Nostradamus, but the resolution of this radar is not sufficient for resolving the two waves (ordinary and extraordinary). In addition, the radar cannot determine the polarization of the waves, making it impossible to distinguish them. Moreover, the tomography is designed for real time application, so ray tracing should run as fast as possible.

4.5 Theory of tomography method

4.5.1 v-method

As described in Section 2.5.1, EM waves can be approximated as rays, and their propagation in a medium is described by the *Eikonal equation*

$$\frac{d}{ds}\left(n\frac{d\vec{x}}{ds}\right) = \nabla n \tag{4.3}$$

that relates the change in ray geometry $\vec{x}$ along an arc length s to the variation in the refractive index n. For a stationary, isotropic, and horizontal stratified ionosphere, the refractive index depends only on the frequency f_e of the emitted EM signal and on the electron density $N_e(\vec{r})$ *(Davies, 1990)*

$$n\left(\vec{r}\right) = \sqrt{1 - \frac{e^2 N_e\left(\vec{r}\right)}{4\pi^2 \epsilon_0 m_e f_e^2}} = \sqrt{1 - \frac{80.6 N_e\left(\vec{r}\right)}{f_e^2}} = \sqrt{1 - \frac{f_p^2}{f_e^2}}, \tag{4.4}$$

where $\vec{r}$ is the location vector, ϵ_0 the vacuum permittivity, e and m_e are the charge and the mass of the electron, respectively. The electron density N_e is provided by the semi-empirical ionospheric model NeQuick. This model has been chosen for this

study, because there are only slight differences between the IRI and NeQuick models in the description of the ionospheric bottom side and ray tracing should run as fast as possible. IRI accounts for the magnetic field and therefore runs slower than the quick-run model NeQuick.

The solution of equation 4.3 is calculated numerically using the fourth order Runge-Kutta method *(Press et al., 1992)*, where the step size h in the Runge-Kutta method was adapted to the local elevation angle ϕ along the ray with

$$h = max\left(1000 \cdot \left|\frac{\sin\phi}{\sin 1^\circ}\right|, 10\right), \tag{4.5}$$

so that it varies between 1000 m in distance for local elevation angles larger than 1° and 10 m at the reflection point. This variable step size ensures a better sensitivity at the reflection points. This is of importance, since rays are most sensitive to the medium in the region around the reflection point, where the plasma frequency f_p and the signal frequency f_e (equation 4.4) are equal. The sensitivity of the rays to the medium is discussed in detail in Section 4.6.2. At their ends, rays are interpolated to altitude zero in order to avoid an artificial time delay introduced by different ray-path lengths.

Figure 4.2 shows traced rays in the ionospheric model NeQuick for six different frequencies and varying elevation angles. The ionosphere was generated for October at 12 UT and a solar flux of 198.1 solar flux units (SFU). Different ray types are clearly visible. Lower angle rays, reaching larger distances up to 2500 km with increasing frequency, and high angle rays that penetrate deeper into the ionosphere with higher frequencies. Additionally, guided rays exists, which get trapped in the ionosphere and move almost parallel to the Earth's surface (*e.g.*, ray for 12 MHz with elevation angle 13° in Figure 4.2). These rays are called *Pedersen Rays* and may be generated by large gradients in electron density.

Knowing the ray-path in the ionospheric model by solving equation 4.3 allows to calculate the propagation time of an EM wave from the radar through the ionosphere and to the ground, where it is backscattered.

$$T_{phase} = \frac{1}{c}\int_{s(n)} n\left(\vec{r}\right) ds, \tag{4.6}$$

where c is the speed of light and $n\left(\vec{r}\right)$ the refractive index of the medium. By replacing the refractive index in equation (4.6) by equation (4.4) and linearizing by first order Taylor series expansion, the propagation time of an EM wave can be

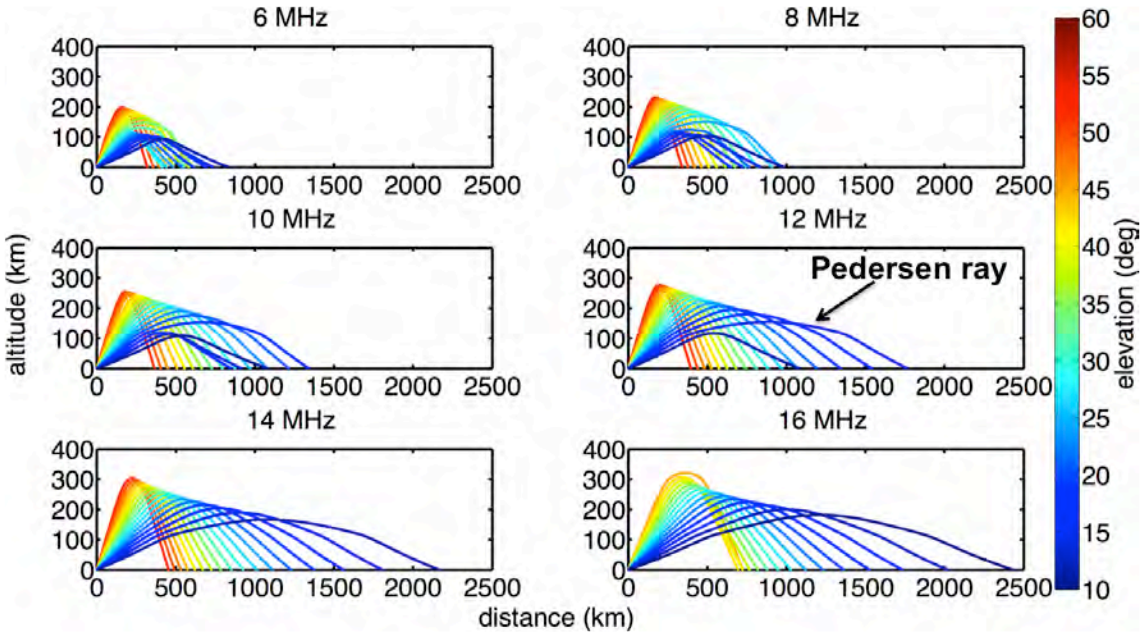


Figure 4.2: Example of rays traced in azimuth 89° in the ionospheric model NeQuick for October at 12 UT and a solar flux of 198.1 solar flux units. For each frequency (6 MHz, 8 MHz, 10 MHz, 12 MHz, 14 MHz and 16 MHz), the elevation angle varies between 10° and 60°.

separated into two parts *(Roy et al., 2014)*

$$\begin{aligned} T_{phase} &= \frac{1}{c}\int_{s(n)} \sqrt{1 - \frac{80.6 N_e\left(\vec{r}\right)}{f_e^2}}\, ds \\ &\approx \underbrace{\frac{1}{c}\int_{s(n)} ds}_{vacuum} \underbrace{- \frac{40.3}{c f_e^2}\int_{s(n)} N_e\left(\vec{r}\right)\, ds}_{ionosphere}, \end{aligned} \tag{4.7}$$

where the first integral in the second line describes the propagation in vacuo and the second one the delay introduced by the ionosphere.

Figure 4.3 illustrates the limit of linearization comparing the first and the second line of equation (4.7). The difference between exact and linearized propagation time is mostly negligible and is in order of few percent for high elevation angles. Another interesting feature evident in Figure 4.3 is the boundary between E and F layer that separates the down-left corner (E region) from the rest of the figure (F region). The lack of values in the right upper corner (white or navy-blue cells) is associated with rays that have not been reflected in the ionosphere that means they traverse the ionosphere and escape into space.

Equation (4.7) allows to calculate by ray tracing the propagation time T_{phase}^{synth} of any EM wave with a given frequency f_e in any given ionospheric model with electron density $N_e^0(\vec{r})$. The electron density in the real ionosphere can be described

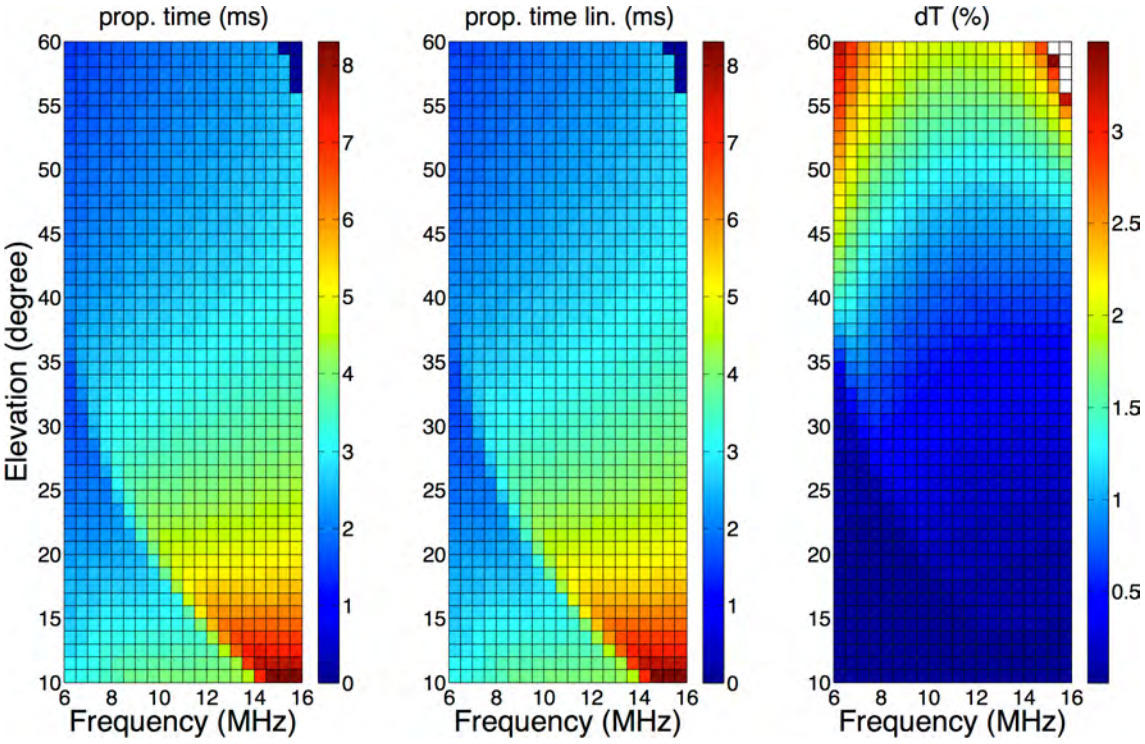


Figure 4.3: Validation of the linearization of the refractive index. Exact propagation time calculated with the first line of equation (4.7) (left), the linearized propagation time calculated with the second line of equation (4.7) (middle), and their difference in percent (right) for 1071 traced rays with elevation angles $10° - 60°$ and frequencies $6 - 16$ MHz. Source: *Roy et al. (2014)*

by the electron density of an *a priori* ionospheric model $N_e^0(\vec{r})$ plus a perturbation $\delta N_e(\vec{r})$ with respect to the model, *i.e.*, $N_e^{real} = N_e^0(\vec{r}) + \delta N_e(\vec{r})$. The ray-path is also perturbed in the real ionosphere compared to the ionospheric model. Assuming that s_0 is the ray-path in the *a priori model* Ne^0, the ray-path in the real ionosphere can be described as $s' = s_0 + \delta s$. The propagation time in the real ionosphere is given by the term

$$T_{phase}^{real} = \frac{1}{c}\int_{s_0+\delta s} ds - \frac{40.3}{cf_e^2}\int_{s_0+\delta s}\left(N_e^0\left(\vec{r}\right) + \delta N_e\left(\vec{r}\right)\right)\, ds, \tag{4.8}$$

which can be separated into

$$\begin{aligned} T_{phase}^{real} &= \frac{1}{c}\int_{s_0} ds - \frac{40.3}{cf_e^2}\int_{s_0}\left(N_e^0\left(\vec{r}\right) + \delta N_e\left(\vec{r}\right)\right)\, ds \\ &+ \frac{1}{c}\int_{\delta s} ds - \frac{40.3}{cf_e^2}\int_{\delta s}\left(N_e^0\left(\vec{r}\right) + \delta N_e\left(\vec{r}\right)\right)\, ds. \end{aligned} \tag{4.9}$$

The integrals in the second line are neglected if starting and endpoints of the ray-paths are fixed (*e.g.*, epicenter and seismometers in the case of seismic tomography).

Since rays are traced between two fixed points and Fermat's principle states that for fixed endpoints the traveltime along a ray-path is stationary with respect to perturbations in the path. The linearized synthetic propagation in an *a priori* ionospheric model along the unperturbed ray-path s_0 is given by

$$T_{phase}^{synth} = \frac{1}{c}\int_{s_0} ds - \frac{40.3}{cf_e^2}\int_{s_0} N_e^0\left(\vec{r}\right) ds. \tag{4.10}$$

The terms in the second line of equation (4.9) describe the ray-path deflection. Neglecting them, the difference δT_{phase} between real and synthetic propagation time (eq. 4.10) in an *a priori* ionospheric model takes the form *Roy et al. (2014)*

$$\delta T_{phase} = T_{phase}^{real} - T_{phase}^{synth} = -\frac{40.3}{cf_e^2}\int_{s_0} \delta N_e\left(\vec{r}\right) ds. \tag{4.11}$$

This approach is based on the hypothesis that the ray-path $s(n)$ in the real ionosphere and in the *a priori* model are the same. That means that the electron density variation δN_e only introduces variation in the speed of the EM waves and not in the ray-path. Therefore, this approach is named ***v-method***.

4.5.2 Damped least squares inversion

The inverse problem to be solved is to reconstruct the electron density perturbation $\delta N_e(\vec{r})$, which is linked to the time delay between real and synthetic propagation time δT_{phase} (equation 4.11). In order to solve numerically equation 4.11, the model is parametrized by N homogeneous, non-overlapping blocks indexed i (Figure 4.4). The electron density perturbation $\delta N_e(\vec{r})$ is expressed by a linear combination of electron density perturbation δm_i in block i and N known basis functions $B_i(\vec{r})$ by

$$\delta N_e\left(\vec{r}\right) = \sum_{i=1}^{N} \delta m_i \cdot B_i\left(\vec{r}\right), \tag{4.12}$$

where $B_i(\vec{r})$ is defined as

$$B_i\left(\vec{r}\right) = \begin{cases} 1 & \text{if } \vec{r} \text{ in block } i \\ 0 & \text{otherwise.} \end{cases}$$

Substituting equation (4.12) into equation (4.11), equation (4.11) becomes for jth-

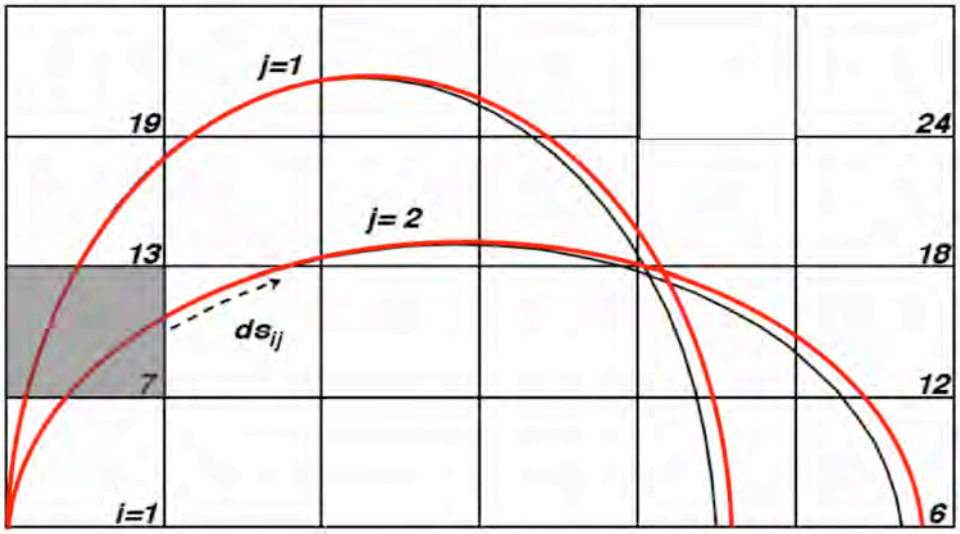


Figure 4.4: Schematic representation of the parametrization of j-rays and i-cells as well as of the ray-path deflection induced by a localized perturbation (e.g., the gray cell). This produces a perturbed ray (red) as compared to the unperturbed ray (black).

measurement of travel time perturbation

$$\delta T_j = -\frac{40.3}{f_{e_j}^2 c} \sum_{i=1}^{N} \delta m_i \, ds_{ij}, \tag{4.13}$$

where ds_{ij} is the length of ray-path segments within block i of ray j and δm_i is the electron density perturbation in block i (Figure 4.4). In this study, relative values of perturbations have been used, to facilitate the comparison of the obtained perturbation with the *a priori* ionospheric model. For relative values, equation (4.13) becomes

$$\frac{\delta T_j}{T_j^0} = -\frac{40.3}{f_{e_j}^2 \, c \, T_j^0} \sum_{i=1}^{N} m_i^0 \frac{\delta m_i}{m_i^0} \, ds_{ij}. \tag{4.14}$$

T_j^0 ($T_j^0 = T_{phase}^{synth}$) is the synthetic propagation time in the *a priori* ionospheric model m_i^0 and i is the block index. Introducing a matrix $\mathbf{A}$ of size $M \times N$ with M being the number of travel time measurements and N the number of basic functions (that is the number of blocks in the parametrization of the ionospheric model), equation (4.14) can be rewritten as

$$\frac{\delta T_j}{T_j^0} = \sum_{i=1}^{N} \frac{\delta m_i}{m_i^0} \, A_{ji} \tag{4.15}$$

or in tensor notation as

$$\frac{\delta \mathbf{T}}{\mathbf{T^0}} = \mathbf{A} \cdot \frac{\delta \mathbf{m}}{\mathbf{m^0}}. \tag{4.16}$$

$\delta\mathbf{m}$ is the model vector with N unknown electron density perturbations δm_i. The vector $\delta\mathbf{T}$ contains M observed travel time perturbations δT_j, and $\mathbf{A}$ is the geometric matrix containing $M \times N$ ray-path segments ds_{ij} of ray j in block i multiplied by a coefficient

$$A_{ji} = -\frac{40.3}{f_{e_j}^2 \, c \, T_j^0} m_i^0 ds_{ij}. \tag{4.17}$$

In order to find relative perturbations $\frac{\delta \mathbf{m}}{\mathbf{m^0}}$, the inverse of the matrix $\mathbf{A}$ has to be calculated according to the equation

$$\frac{\delta \mathbf{m}}{\mathbf{m^0}} = \mathbf{A^{-1}} \cdot \frac{\delta \mathbf{T}}{\mathbf{T^0}}. \tag{4.18}$$

Since there is no exact solution of this problem, the best thing to do is to estimate that model parameters giving the best approximate solution. According to *Menke (1989)*, the best approximate solution $\frac{\delta \mathbf{m}}{\mathbf{m^0}}$ of an inverse problem has to minimize the L_2 norm of data and model, nominally

$$\left|\left| \frac{\delta \mathbf{T}}{\mathbf{T^0}} - \mathbf{A} \cdot \frac{\delta \mathbf{m}}{\mathbf{m^0}} \right|\right|^2 = \min \qquad \left|\left| \frac{\delta \mathbf{m}}{\mathbf{m^0}} \right|\right|^2 = \min. \tag{4.19}$$

The first condition imposes the best fit to the data, the second one minimizes the discrepancy from the *a priori* model. The Euclidean length of the solution quantifies the simplicity of the solution. A solution is defined to be simple if it is small when measured under the L_2-norm. This is a prior assumption to the solution of the problem. Then, in least-squares sense the solution of equation (4.18) is

$$\begin{array}{lll} \frac{\delta \mathbf{m}}{\mathbf{m^0}} = \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T \cdot \frac{\delta \mathbf{T}}{\mathbf{T^0}} & \Leftrightarrow & \| \, \frac{\delta \mathbf{T}}{\mathbf{T^0}} - \mathbf{A} \cdot \frac{\delta \mathbf{m}}{\mathbf{m^0}} \, \|^2 = min. \\ \frac{\delta \mathbf{m}}{\mathbf{m^0}} = \mathbf{A}^T \cdot \left(\mathbf{A} \cdot \mathbf{A}^T\right)^{-1} \cdot \frac{\delta \mathbf{T}}{\mathbf{T^0}} & \Leftrightarrow & \| \, \frac{\delta \mathbf{m}}{\mathbf{m^0}} \, \|^2 = min. \end{array}$$

The minimum norm solution in the second line can be obtained using Lagrange multipliers and its derivation is described in detail by *(Menke, 1989, Chapter 3)*.

It is known that inverse problems in geophysics generally present a number N of parameters (here the vector $\delta\mathbf{m}$) larger than the number M of observations (here the vector $\delta\mathbf{T}$). These problems are called under-determined inverse problems. In practice, inverse problems are never purely under-determined, but mixed-determined,

meaning that some model parameters are over-determined and others are under-determined. For instance, in this tomography problem there may be cells passed through by several rays and others which are missed entirely (Figure 4.4). Consequently, the matrices $\mathbf{A}^T \cdot \mathbf{A}$ and $\mathbf{A} \cdot \mathbf{A}^T$ cannot be inverted, because they are almost singular. Even in the case of formally existing inverse matrices, these are often ill-conditioned that means small changes in the data vector ($\delta\mathbf{T}$) lead to large changes in the model estimation ($\delta\mathbf{m}$).

To find a stable solution ($\frac{\delta\mathbf{m}}{\mathbf{m^0}}$) balancing the sensitivity to the data as well as coherence with the *a priori* model, *Menke (1989)* suggests a *damped least squares solution*

$$\frac{\delta\mathbf{m}}{\mathbf{m^0}} = \left(\mathbf{A}^T \cdot \mathbf{A} + \lambda \cdot \mathbf{I}\right)^{-1} \cdot \mathbf{A}^T \cdot \frac{\delta\mathbf{T}}{\mathbf{T^0}}, \tag{4.20}$$

minimizing the cost function

$$F = \left|\left|\frac{\delta\mathbf{T}}{\mathbf{T^0}} - \mathbf{A} \cdot \frac{\delta\mathbf{m}}{\mathbf{m^0}}\right|\right|^2 + \lambda\left|\left|\frac{\delta\mathbf{m}}{\mathbf{m^0}}.\right|\right|^2. \tag{4.21}$$

λ is a regularization parameter and $\mathbf{I}$ is the identity matrix.

Simultaneous minimization of both terms is not possible, but the parameter λ controls the emphasis that is put on the conflicting requirement. If λ is set to zero, the misfit to the data is minimized, but no prior assumption to the solution is made to rule out the under-determined model parameters. The choice of this regularization parameter λ is a very sensitive issue and will be explained in Section 4.6.3.

4.5.3 Damped least squares inversion with constraints

The damped least squares inversion described before imposes the best fit to the data and minimizes the discrepancy from the *a priori* model, given by equations 4.19. However, this inverse problem is an under-determined problem, because the number of rays is smaller than the number of pixels ($M < N$), *e.g.,* 1056 rays and 2000 cells. In particular, the number of cells not hit by any ray $\sim$ 1600 is large as compared to the number of cells crossed by rays $\sim$ 400. As a result, the ray geometry defines a singular matrix, which makes the reconstruction of the electron density in the probed media difficult.

For those pixels without any traversing rays them, or cells with just a few number of rays, additionally constraints that impose smoothness to the solution can be included in the algorithm. These constraints should minimize the difference between the solution in one cell and its eight (or less in the case of border cells) adjacent

cells (Figure 4.5). For each of this eight constraints a matrix can be constructed.

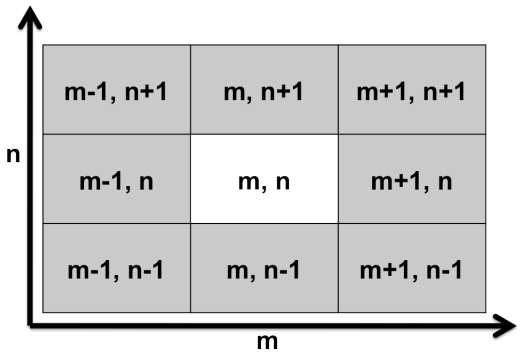


Figure 4.5: Sketch for creating eight constraint matrices minimizing the difference of the solution in cell m,n and its eight adjacent cells.

For instance, a constraint matrix minimizing the difference between the solutions in cell m, n and cell $m+1$, n can be defined as

$$\mathbf{R_1} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

Attention has to be paid at the boundaries in altitude and distance and the arrangement of the grid. Similarly, seven additional matrices $\mathbf{R}$ can be created with each matrix being different. Then, the cost function (equation 4.21) becomes

$$F = \underbrace{\left|\left|\frac{\delta \mathbf{T}}{\mathbf{T^0}} - \mathbf{A}\cdot\frac{\delta \mathbf{m}}{\mathbf{m^0}}\right|\right|^2}_{\text{data misfit}} + \underbrace{\lambda\left|\left|\frac{\delta \mathbf{m}}{\mathbf{m^0}}\right|\right|^2}_{\text{solution norm}} + \underbrace{\lambda_1\left|\left|\mathbf{R_1}\frac{\delta \mathbf{m}}{\mathbf{m^0}}\right|\right|^2 + \cdots + \lambda_8\left|\left|\mathbf{R_8}\frac{\delta \mathbf{m}}{\mathbf{m^0}}\right|\right|^2}_{\text{difference to the 8 adjacent cells}}. \quad (4.22)$$

Minimizing the cost function (that means $\frac{\partial F}{\delta m} = 0$) yields

$$\frac{\partial F}{\delta m} = 0 = 2\mathbf{A^T}\mathbf{A}\frac{\delta \mathbf{m}}{\mathbf{m^0}} - 2\mathbf{A^T}\frac{\delta \mathbf{T}}{\mathbf{T^0}} + 2\lambda\frac{\delta \mathbf{m}}{\mathbf{m^0}} + 2\lambda_1\mathbf{R_1^T}\mathbf{R_1}\frac{\delta \mathbf{m}}{\mathbf{m^0}} + \cdots + 2\lambda_8\mathbf{R_8^T}\mathbf{R_8}\frac{\delta \mathbf{m}}{\mathbf{m^0}}. \quad (4.23)$$

The best solution of this inverse problem is

$$\frac{\delta \mathbf{m}}{\mathbf{m^0}} = \left(\mathbf{A^T}\mathbf{A} + \lambda\mathbf{I} + \lambda_1\left[\mathbf{R_1^T}\mathbf{R_1} + \cdots \mathbf{R_8^T}\mathbf{R_8}\right]\right)^{-1}\cdot\mathbf{A^T}\cdot\frac{\delta \mathbf{T}}{\mathbf{T^0}}, \quad (4.24)$$

assuming that $\lambda_1 = \lambda_2 = \cdots = \lambda_8$. There are two regularization parameters to choose, namely λ and λ_1.

4.5.4 Consideration of the ray-path deflection

In monostatic OTH radars, the endpoints of the rays (where the signal is backscattered by the ground) are unknown, so the location of the scattering point at the ground can change for a constant elevation angle and depends on the electron density variation $\delta N_e(\vec{r})$ in the ionosphere (Figure 4.4). Thus, the ray-paths in the real ionosphere and in the ionospheric model are not equal, $s\,(n) \neq s\,(n_0)$, introducing an additional shift δT^{ray}_{phase} in the propagation time. *Snieder and Spencer (1993)* showed that both effects (velocity variation and ray-path deflection) can be combined in a single perturbation theory. At first-order, the two effects are simply additive *(Snieder and Spencer, 1993)* and the perturbation in the propagation time can be expressed as

$$\delta T^{real}_{phase} = \delta T^{velocity}_{phase} + \delta T^{ray}_{phase}. \tag{4.25}$$

The deduction of an equation for the propagation time that takes into account both effects, the velocity variation and the change of the ray-path, requires that the integrals in the second line of equation (4.9) are not neglected. Replacing δs in equation 4.9 with $\delta s = s' - s_0$, gives the following term for the propagation time in the real ionosphere *(Roy et al., 2014)*

$$\begin{aligned} T^{real}_{phase} &= \frac{1}{c}\int_{s'} ds - \frac{1}{c}\int_{s_0} ds \\ &- \frac{40.3}{cf_e^2}\Big[\int_s \left(N_e^0\,(\vec{r}) + \delta N_e\,(\vec{r})\right)\, ds - \int_{s_0} \left(N_e^0\,(\vec{r}) + \delta N_e\,(\vec{r})\right)\, ds\Big] \\ &+ \frac{1}{c}\int_{s_0} ds - \frac{40.3}{cf_e^2}\int_{s_0} \left(N_e^0\,(\vec{r}) + \delta N_e\,(\vec{r})\right)\, ds. \end{aligned} \tag{4.26}$$

That is

$$T^{real}_{phase} = \frac{1}{c}\int_{s'} ds - \frac{40.3}{cf_e^2}\int_{s'} \left(N_e^0\,(\vec{r}) + \delta N_e\,(\vec{r})\right)\, ds. \tag{4.27}$$

This equation describes both effects, velocity variation and ray-path deflection, on the propagation time of EM waves in the real ionosphere. It is obvious that this equation can be simply derived from equation 4.7 by replacing $N_e\,(\vec{r})$ with $N_e^0\,(\vec{r}) + \delta N_e\,(\vec{r})$ and $s\,(n)$ by s'. For the difference in the propagation time δT^{real}_{phase} between real (eq. 4.27) and synthetic propagation time (eq. 4.10) in the *a priori* ionospheric

model now follows

$$\begin{aligned}\delta T^{real}_{phase} &= T^{real}_{phase} - T^{synth}_{phase} \\ &= \frac{1}{c}\int_{s'} ds - \frac{1}{c}\int_{s_0} ds \\ &- \frac{40.3}{cf_e^2}\int_{s'} \left(N_e^0(\vec{r}) + \delta N_e(\vec{r})\right)\, ds + \frac{40.3}{cf_e^2}\int_{s_0} N_e^0(\vec{r})\, ds. \end{aligned} \quad (4.28)$$

This form of equation 4.25 allows to describe the difference between data and synthetics accounting for both effects (velocity and ray-path deflection). Consequently, this approach is called the ***v&r-method***. Now it follows that the two terms of equation 4.25, $\delta T^{velocity}_{phase}$ and δT^{ray}_{phase}, are given as

$$\begin{aligned}\delta T^{ray}_{phase} &= \frac{1}{c}\int_{s'} ds - \frac{1}{c}\int_{s_0} ds \\ &- \frac{40.3}{cf_e^2}\Big[\int_{s'} \left(N_e^0(\vec{r}) + \delta N_e(\vec{r})\right)\, ds - \int_{s_0} \left(N_e^0(\vec{r}) + \delta N_e(\vec{r})\right)\, ds\Big] \end{aligned} \quad (4.29)$$

and

$$\delta T^{velocity}_{phase} = -\frac{40.3}{cf_e^2}\int_{s_0} \delta N_e(\vec{r})\, ds. \quad (4.30)$$

δT^{ray}_{phase} depends on the unknown electron density perturbation δN_e, that is the solution of the inverse problem mentioned above. To set up an inverse problem, which allow the determination of δN_e based on observations of δT^{real}_{phase}, δT^{ray}_{phase} needs to be expressed in terms of known δN_e. The study of *Snieder and Spencer (1993)* also indicates that sensitivity kernels $K(s)$ can be defined so that

$$\delta T^{ray}_{phase} = \int_{s_0} K(\vec{r}) \cdot \delta Ne(\vec{r})\,, \quad (4.31)$$

where $K(\vec{r})$ is the data kernel, which in the v-method described above is just a delta function along the unperturbed ray-path s_0. Here, the kernel contains the Fréchet derivatives $\partial T/\partial m$, where ∂T is a perturbation in the propagation time caused by a perturbation in the model m. If the relation between the model m and the propagation time T is linear, the sensitivity function can be computed numerically $K(\vec{r})$ by

1. assuming a Dirac delta perturbation $\delta N_e = \delta(s - s_i)$, where i denotes the cell,
2. running the Tracé De Rayons (TDR) code mentioned before to obtain the travel time delay due to velocity variation, $\delta T^{velocity}_j$ for ray j (*i.e.,* equation 4.30) and the total time delay δT^{j}_{phase}, (equation 4.28)

3. calculating δT_{ray} by means of equation 4.25.

The resulting value can be substituted on the left-hand side of equation 4.31, while the right-hand side collapses to $K(s_i)$ by the properties of the Dirac delta. Iterating the procedure over the location of all cells s_i allows to compute K throughout the volume of interest. In practice, the Dirac delta is replaced by a localized perturbation of an arbitrary but finite known value δN_e^* and K is obtained by dividing the resulting value of δT_{phase}^{ray} by δN_e^*, so the kernel is given by

$$K_{ji} = \frac{\delta T_j^*}{\delta Ne_i^*}.$$

Once kernels have been computed, equation 4.25 can be replaced by

$$\delta T_{phase,j}^{real} = -\frac{40.3}{f_{e_j}^2 c} \sum_{i=1}^{N} \delta m_i \, ds_{ij} + \sum_{i=1}^{N} K_{ji} \, \delta m_i \tag{4.32}$$

or in a tensor formalism as

$$\delta \mathbf{T} = (\mathbf{A} + \mathbf{A}') \cdot \delta \mathbf{m}, \tag{4.33}$$

where (for absolute values)

$$A_{ji} = -\frac{40.3}{f_{e_j}^2 \, c} ds_{ij} \tag{4.34}$$

and

$$\mathbf{A}' = K_{ji} = \frac{\delta T_{i,j}^*}{\delta Ne_i^*}. \tag{4.35}$$

The damped least squares inversion described in Section 4.5.2 is applied to the matrix $\mathbf{A}$ in the case of the ***v-method***, and to the matrix $\mathbf{A} + \mathbf{A}' = \mathbf{M}$ in the case of the ***v&r-method***.

In order to proof the linearity of the problem and to choose the amplitude of the localized, imposed electron density perturbation δNe_i^* in the calculation of the ray-path deflection, the ratio $\frac{\delta T^*}{dm^*}$ was traced as a function of the imposed perturbation dm^* for one ray (*i.e.,* 11 MHz, 35° take-off angle) perturbing all crossed cells one after another. Figure 4.6 shows the linear relationship between the propagation time difference of the perturbed and the not-perturbed ray δT^* and the imposed perturbation dm^*. Furthermore, the figure proves that the ratio $\frac{\delta T^*}{dm^*}$ is independent of the imposed electron density δNe_i^* for most of the cells. Only for cells at altitude around 80 km the ratio varies slightly due to the strong variability of the *a priori* ionospheric model. In the following, the imposed electron density δNe_i^* was set to 1%.

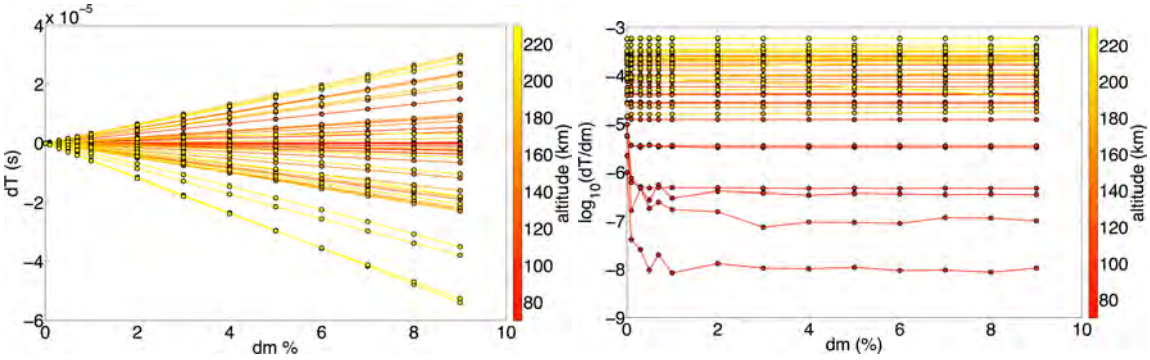


Figure 4.6: Proof of the linearity of the inverse problem. The propagation time difference of one ray traced in the *a priori* ionospheric model and one ray perturbed by a localized, imposed electron density of abritray amplitude dm^* (left) and the ratio $\frac{\delta T^*}{dm^*}$ as a function of imposed electron density perturbation dm^*. A column of points represents all cells crossed by the ray, and their color corresponds to the altitude.

4.6 Inversion results for synthetics

In order to validate and compare the two methods described in the previous sections, the method was tested for synthetic data generated by ray tracing TDR in the continuous *a priori* ionospheric model *NeQuick (Radicella and Leitinger, 2001)*. The data were simulated by tracing 1071 rays with elevation angles between 10° and 60° and in a frequency range of 6 −16 MHz in the ionospheric model plus an additional, known perturbation δNe^{target}. $\delta\mathbf{T}$ is then the difference in propagation time between the ray tracing in the perturbed ionospheric model, and the ray tracing in the *a priori* ionospheric model, *i.e.,* equation (4.11). The solution of the inverse problem, the obtained perturbation $\delta\mathbf{m}$, has to correspond to δNe^{target}. In the following sections, results obtained for a checkerboard perturbation *(Lévêque et al., 1993)* and for a localized perturbation of 0.1 % will be presented and discussed. The background ionosphere $Ne^{apriori}$ was generated for October at 12:00 UT with a solar flux of 198.1 SFU.

The ray-path deflection in the $v\&r$-method was calculated by adding 1% of electron density perturbation cell by cell, since the ratio of travel time perturbation to imposed electron density perturbation $\frac{\delta T}{dm^*}$ is independent of dm^* (see Section 4.5.4).

4.6.1 Effect of ray-path deflection

The theoretical limit of the developed method was investigated by calculation of a vector of traveltime perturbations $\delta\mathbf{T}^{frozen}$ satisfying exactly the hypothesis that the electron density perturbation δNe^{target} modifies only the velocity of EM waves,

i.e., rays are *frozen* in the *a priori* model configuration. This theoretical data set represents the idealized case of no ray-path perturbation. This resembles to classic seismic tomography (where both endpoints of the ray are known). It should be inverted to explore the resolution linked to the data coverage, as well as the numerical noise introduced by the discretization of the models (number and dimensions of the cells).

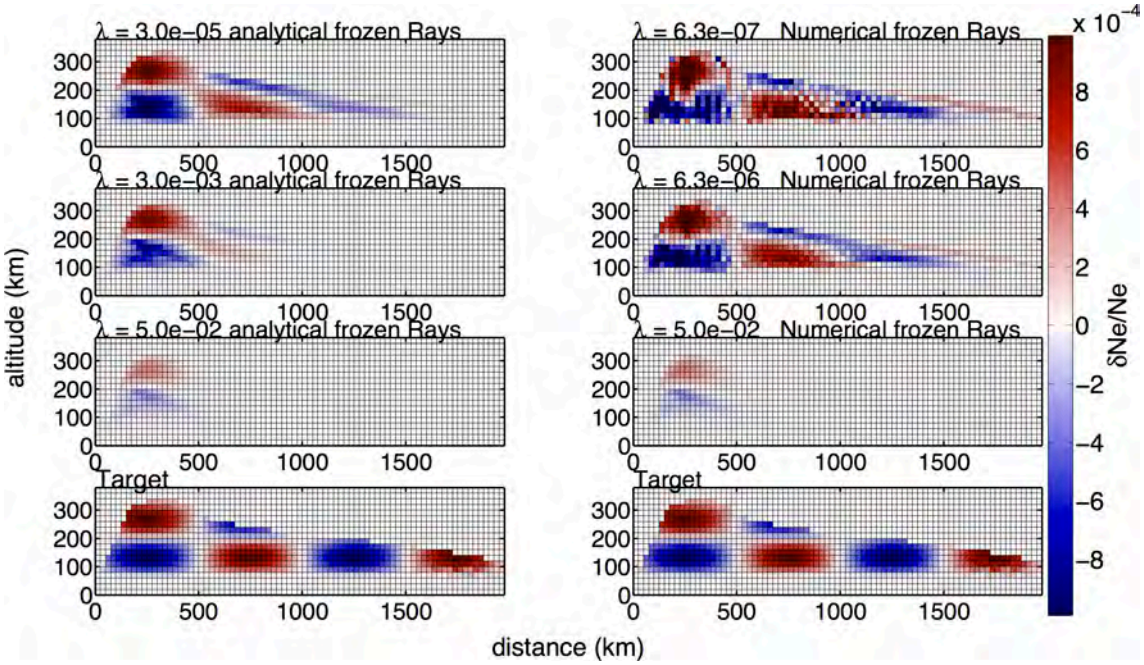


Figure 4.7: Inversion results for a checkerboard benchmark test using analytically (left) and numerically (right) calculated *frozen* rays. The second plot from the top of each column shows the solution for the best regularization parameter, chosen as described in Section 4.6.3.

Two tests were realized to explore the theoretical limit using numerical and analytical *frozen* rays. In the case of analytical *frozen* rays, the vector $\delta\mathbf{T}^{frozen}$ of travel time perturbations was calculated analytically to calculate the propagation time, according to

$$\left(\frac{\delta\mathbf{T}}{\mathbf{T}}\right)_{ana}^{frozen} = \mathbf{A_{s_0}} \cdot \left(\frac{\delta\mathbf{Ne}}{\mathbf{N_e^0}}\right)^{target}, \tag{4.36}$$

where $\mathbf{A_{s_0}}$ contains the ray-path in the unperturbed model $\mathbf{m_0}$ and is given by equation 4.17. Next using equation 4.20, $\frac{\delta\mathbf{T}}{\mathbf{T}}_{ana}^{frozen}$ was inverted for the relative electron density perturbation $\frac{\delta\mathbf{Ne}}{\mathbf{N_e^0}}$. This is, of course, just a mathematical operation, but proves the functionality of the code and allows to estimate the noise level introduced by ray tracing and the discretization of the model by cells.

Numerical *frozen* rays were traced in the *a priori* ionospheric model and a perturbed propagation time along the unperturbed ray-path s_0 was calculated, using

the equation

$$T_{phase} = \frac{1}{c}\int_{s_0} ds - \frac{40.3}{f_{e_j}^2 c}\int_{s_0}\left(N_e^0 + \delta N_e\left(\vec{r}\right)\right) ds. \tag{4.37}$$

Then for the relative time delay $\frac{\delta \mathbf{T}}{\mathbf{T}}^{frozen}_{num}$ follows

$$\left(\frac{\delta \mathbf{T}}{\mathbf{T}_{phase}^{synth}}\right)_{num}^{frozen} = \frac{\mathbf{T}_{phase} - \mathbf{T}_{phase}^{synth}}{\mathbf{T}_{phase}^{synth}} = -\frac{40.3}{f_{e_j}^2 \cdot c \cdot \mathbf{T}_{phase}^{synth}}\int_{s_0}\frac{\delta N_e\left(\vec{r}\right)}{N_e^0} ds, \tag{4.38}$$

were $\mathbf{T}_{phase}^{synth}$ is given by equation 4.10. The inversion is then performed by means of equation 4.20. Figures 4.7 and 4.8 show the inversion for three different regulariza-

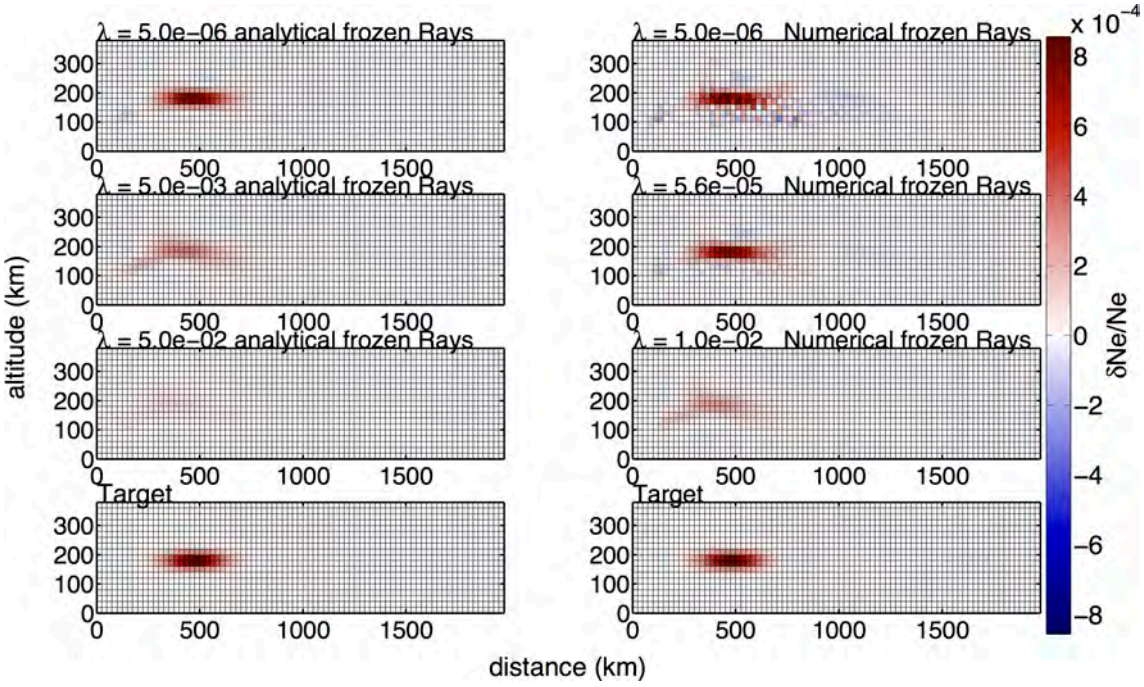


Figure 4.8: Inversion results for a localized perturbation using analytically (left) and numerically (right) calculated *frozen* rays. The second plot from the top of each column shows the solution for the best regularization parameter, chosen as described in Section 4.6.3.

tion parameters λ, using analytical and numerical *frozen* rays for a checkerboard and a localized perturbation. The inversion with numerically calculated *frozen* rays with the best regularization parameter (*i.e.*, $\lambda = 6.3 \cdot 10^{-6}$ for the checkerboard and $\lambda = 5.6 \cdot 10^{-5}$ for the localized perturbation) reproduces well the target model, although some noise is present in the solution. At distances larger than 1500 km, the targets are not resolved due to the sparse ray coverage (Figure 4.9). Both tests show that the inversions do not depend significantly on the regularization parameter. This

result is also observed for the inversion using the analytically calculated travel time perturbation $\delta\mathbf{T}_{ana}^{frozen}$, but the solution is less noisy and more damped. It reflects the noise introduced by tracing rays in a continuous ionospheric model like NeQuick instead of using a discretized model with the parametrization of homogenous blocks. The mean noise is around 12%.

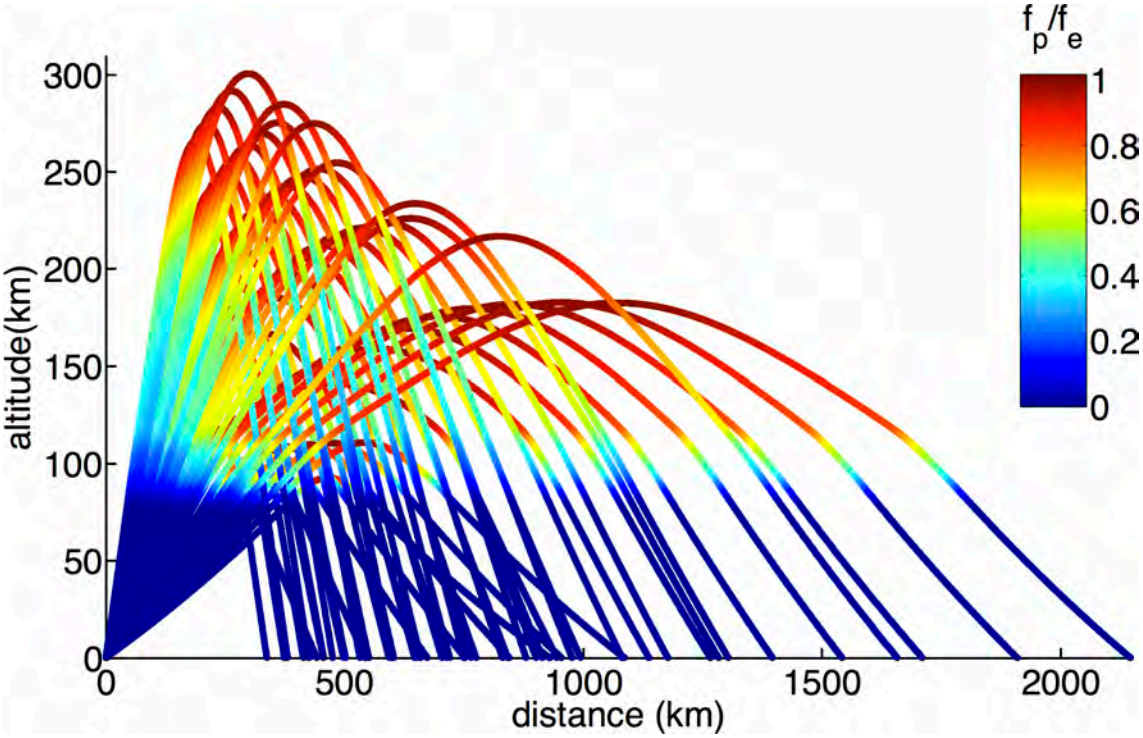


Figure 4.9: Ratio of plasma to signal frequency along the rays, calculated using equation 4.40 in order to show the sensitivity of the rays to the medium. Rays are most sensitive to the medium where the ratio is approximately 1.

The limited complexity of the solution can be explained by the varying sensitivity of the rays to the medium. The electron density in the ionosphere is strongly varying with altitude, from zero at around 80 km to $10^{11}-10^{12}e/m^3$ at around 300 km (Figure 2.4). Consequently, EM waves emitted by the radar are particularly sensitive to the zone where the rays are reflected that is where the plasma frequency f_p approaches the emission frequency f_e. A more exact equation for the ratio of plasma to emission frequency at the reflection point can be derived from Snell's spherical law,

$$R \cdot n(0) \cdot \cos\phi_0 = (R+h) \cdot n(h) \cdot \cos\phi_h, \tag{4.39}$$

where R is the Earth radius, $n\,(0)$ and $n\,(h)$ the refractive index at the Earth surface and at the altitude h, respectively, ϕ_0 the elevation angle of the ray, and ϕ_h the local elevation angle along the ray. Assuming that $n(0)$ is 1 and using equation 4.4 for

the refractive index, the ratio of plasma to emission frequency is given by

$$\frac{f_p}{f_e \cdot \sqrt{1 - \frac{\cos^2 \phi_0 \cdot R^2}{(R+h)^2}}} = 1. \tag{4.40}$$

Figure 4.9 shows this ratio along the traced rays color coded. The ratio equals 1 between 200 and 300 km altitude at around 500 km horizontal distance and at 1000 km distance. By contrast, the corner around 200 km of altitude and close to the radar (around 100 km distance) is not well reproduced. Indeed, this zone lacks of reflected rays and is only crossed by nearly vertically rays. Consequently, these regions are not well resolved in the inversion.

4.6.2 v-method vs. $v\&r$-method

The inversion results $\frac{\delta \mathbf{m}}{\mathbf{m}}$ for the checkerboard benchmark test and a localized perturbation are illustrated in Figured 4.11 and 4.10, respectively, where in the solutions obtained with the v-method and the $v\&r$-method are directly compared for three different regularization parameters λ. The model is parametrized by homogenous blocks of 25 km x 20 km, covering an area of 2500 km from the geographical coordinates of the radar Nostradamus and reaching up to 400 km altitude. As expected, the solution models depend significantly on the regularization parameters. This effect was less serious in the case of the idealized *frozen* ray (Figures 4.7 and 4.8).

With ray-path deflections (endpoint perturbations) taken into account in the inversion, resolution deteriorates and the results are affected more profoundly by the choice of λ (eq. 4.20). The correct location (500 km distance, 200 km altitude) of maximum anomaly of the localized perturbation is only reconstructed by the $v\&r$-method provided that an adequate value is assigned to λ; the v-method identifies a high $\delta Ne/Ne$ anomaly in the common area of $200-500$ km in horizontal distance and $100-200$ km altitude, but slightly mislocates it and does not reproduce its shape.

The better performance of the $v\&r$- compared to the v-method is evident in Figure 4.12, where the $v\&r$ L-curve *(Tikhonov, 1963)* has a more pronounced corner than that from the v- method. Compared to the ideal *frozen*-ray case (Figure 4.8), both the v- and $v\&r$-methods occasionally introduce large-scale negative $\delta Ne/Ne$ anomalies that do not correspond to any feature of the target model, where the sign of perturbation is always positive. These artifacts can be explained with nonlinear effects of ray-path deflections, which (endpoints not fixed) can, in principle, result in faster propagation times even if the velocity perturbation is negative. While the $v\&r$-method allows identification of more details of the target in the case of the checkerboard perturbation (Figure 4.11), the v-method identifies only the negative

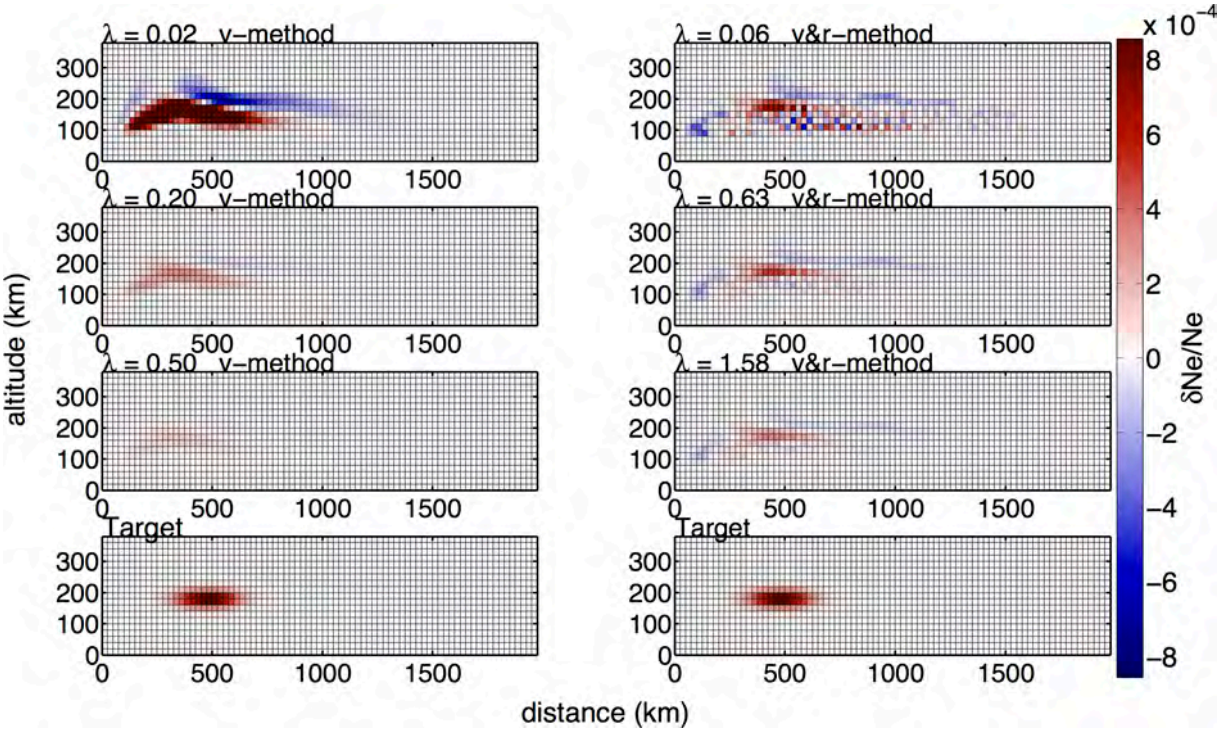


Figure 4.10: Inversion results for the v-method (left) and $v\&r$-method (right) for a localized perturbation of 0.1 % using three different regularization parameters. The inversions for the best regularization parameter are shown in the second plot from top.

and positive perturbation at 500 km distance, but cannot reproduce their shape and amplitude. This result is grid independent, as demonstrated in Figures 5.2 in Section 5, where the pixel dimensions were enhanced in distance and altitude, respectively. The plots show the inversion for the best value of regularization, chosen as described in the following Section. Independent of the grid, the $v\&r$-method is able to reproduce more details of the perturbations, even for larger distances. Resolution of the tomography method for different targets and grid sizes will be discussed more detailed in Section 5. The bad reconstruction at the edges and for larger distances is due to the sparse ray coverage, as demonstrated in Section 4.6.1, this is also due to the sparse ray coverage, not to the ray-path deflection.

4.6.3 Dependence of the solution on regularization

As *Roy et al. (2014)* showed, the regularization parameter λ strongly controls the noise level in the solution. This is also clearly visible in Figure 4.11. In Section 4.5.2 is discussed that the chosen regularization parameter λ has to minimize the misfit to the data $\mathbf{dT}$ as well as the solution norm quantifying the discrepancy with the *a priori* model (*i.e.,* equation 4.19).

To explore the influence of λ, the inverse problem (eq. 4.20) is solved for a set of regularization parameters, and the misfit is analyzed as function of the normalized solu-

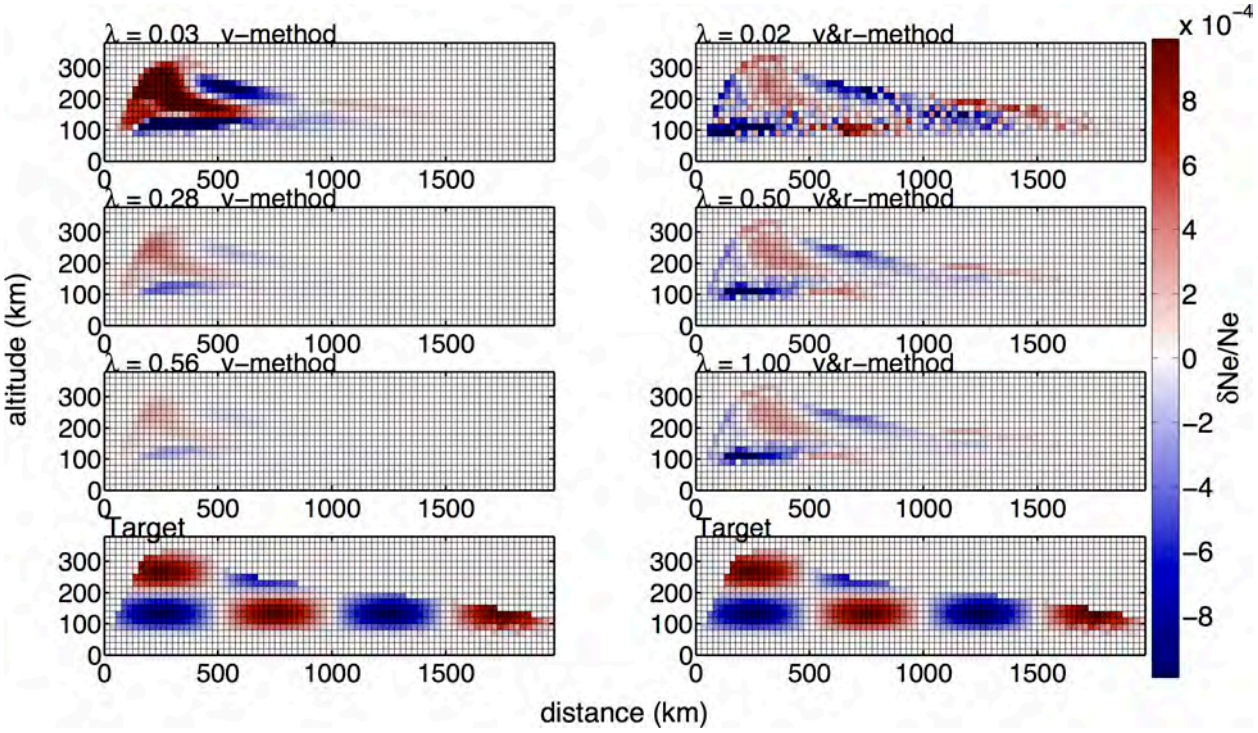


Figure 4.11: Inversion results for the v-method (left) and $v\&r$-method (right) for a checkerboard perturbation of 0.1 % using different regularization parameter. The inversions for the best regularization parameter are shown in the second plot from the top.

tion norm. Following *Tikhonov (1963)* the obtained plot (*e.g.*, Figure 4.12) is called trade-off curve, also commonly known as L-curve, because of its shape. The best compromise between the minimum misfit and the minimum solution norm is usually obtained geometrically at the maximum curvature of the L-curve *(Tikhonov, 1963)*. Examples for L-curves are given in Figure 4.12 for the localized and the checkerboard benchmark test using the $v\&r$ and the v-method. The best compromises are $\lambda_{best} = 0.5$ and $\lambda_{best} = 0.63$, respectively. Smaller regularization parameters for both tests and both inversion methods add noise in the inversion, so the solution norm is large, but the misfit is small. On the other hand, larger regularization parameters strongly damp the solution, therefore the solution norm is small, but the misfit is large.

It is important to notice that the best regularization parameter, chosen at the maximum curvature of the L-curve, strongly depends on the range of the explored parameters. Indeed, the minimum value λ_{min} of the explored set of λ strongly controls the resulting λ_{best}. This is clearly visible in Figure 4.13, showing the inversion results obtained for different ranges of regularization parameters ($\lambda = [\lambda_{min}, 5000$], with $\lambda_{min} = [\ 10^{-5},\ 0.1]$) for the checkerboard perturbation. For each of these four ranges, a L-curve has been calculated, and the best value of regularization has been determined geometrically at the maximum curvature of the corresponding L-

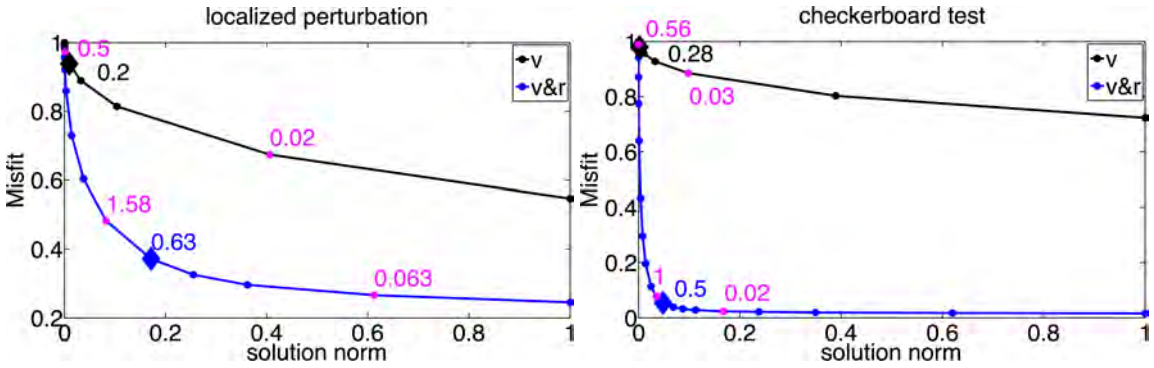


Figure 4.12: L-curves for the localized perturbation (a) and the checkerboard benchmark test (b) and the using the v- and $v\&r$-methods for inversion. Inversion results for the magenta points are shown in Figures 4.11 and 4.10. Diamonds represent the best regularization parameter. Misfit and solution norm are given by equation 4.19.

curve as described before. Obviously, the choice of a too small value of λ_{min} leads to a λ_{best} value that imposes very much noise on the result of the inversion, *e.g.*, $\lambda = 0.02$. At the opposite extreme, a larger value of λ_{min} results in a λ_{best} that reduces considerably the amplitude of the solved pattern, because the solution is damped too much.

In order to find the best range of regularization parameters for the calculation of the L-curve, an error curve has been calculated: For each λ-range $[\lambda_{min}, \lambda_{max}]$, with $\lambda_{min} = [10^{-5}, 10^{3}]$ and $\lambda_{max} = 5000$, the L-curve has been calculated, and the best value of regularization has been chosen from this curve. Then, the error *i.e.*, $\sum_{i=1}^{cells} \left(\left(\frac{\delta Ne}{Ne}\right)_i^{target} - \left(\frac{\delta Ne}{Ne}\right)_i \right)^2$, between the target model and the solution has been determined. The λ_{min} minimizing the error curve is $\lambda_{min} = 10^{-3}$ for the checkerboard perturbations. The corresponding λ_{best} is 0.5 and the inversion result using this regularization parameter is shown in Figures 4.11 in the second plot from top in the right column.

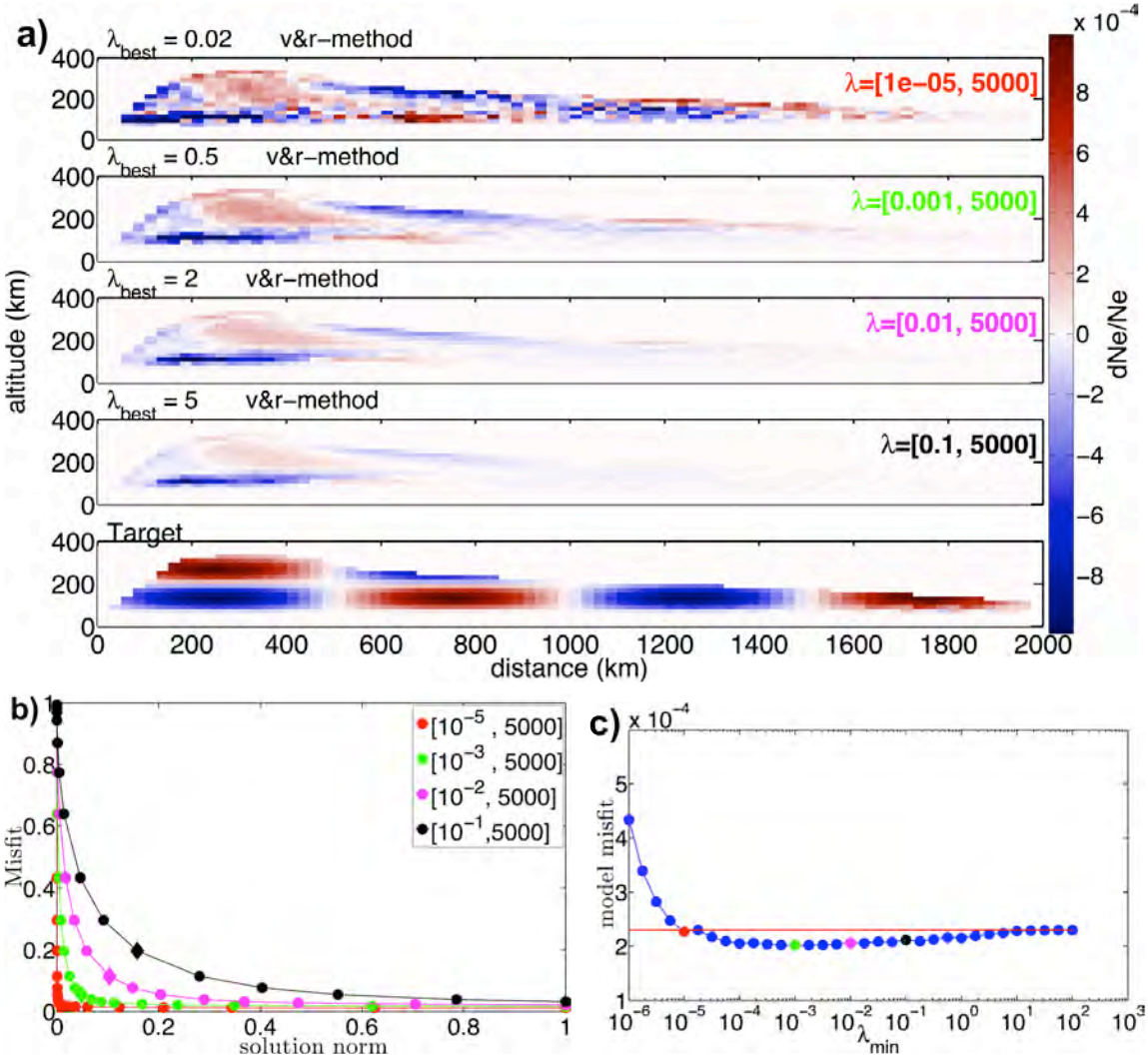


Figure 4.13: a) Inversion results for the checkerboard benchmark test using different ranges of regularization parameters for the calculation of the L-curve shown in b). b) L-curve for different λ-ranges for the checkerboard benchmark test. Diamonds represent the best value of regularization for each curve for which the inversion is shown in a). c) Corresponding error curve for the explored range of regularization parameters. Colored points represent the errors for inversions using the best regularization parameters from the L-curves. The red line is the sum over the target models, *i.e.*, $\sum_{i=1}^{cells} \left(\frac{\delta Ne}{Ne}\right)_i^2$. The errors converge to the red line when the solution is totally damped. Source: *Roy et al. (2014)*

4.6.4 Damped least squares inversion with constraints - Results

In order to reduce the noise in the inversions showed by *Roy et al. (2014)* in particular at low altitudes around 100 km, the damped least squares inversion with constraints (Section 4.5.3) was applied to the $v\&r$-method. This procedure is not expected to change radically the inversion result, but it smoothes the solution. As the previous inversions show, results obtained with the v-method are already smooth in the first

iteration.

The major problem is to choose the two regularization parameters for the minimum solution norm, λ and λ_1, for the constraint matrices. A useful starting point for at least getting 'on the map' is using the trace of the matrices as regularization parameters. That means, the regularization parameters are chosen as

$$\lambda = \frac{\text{trace}\left(\mathbf{M^T} \cdot \mathbf{M}\right)}{\text{trace}\left(\mathbf{I}\right)} \qquad \lambda_1 = \frac{\text{trace}\left(\mathbf{M^T} \cdot \mathbf{M}\right)}{\text{trace}\left(\mathbf{R}\right)}. \tag{4.41}$$

$\mathbf{I}$ is the identity matrix, $\mathbf{M}$ denotes the matrix $\mathbf{A} + \mathbf{A}'$ used in the $v\&r$-method, and $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{H^T H}, \tag{4.42}$$

where $\mathbf{H}$ is the sum of the eight constraint matrices minimizing the difference to the eight adjacent cells

$$\mathbf{H} = \mathbf{R_1} + \mathbf{R_2} + \mathbf{R_3} + \mathbf{R_4} + \mathbf{R_5} + \mathbf{R_6} + \mathbf{R_7} + \mathbf{R_8}. \tag{4.43}$$

Unfortunately, in the present case, this choose of regularization means λ is around

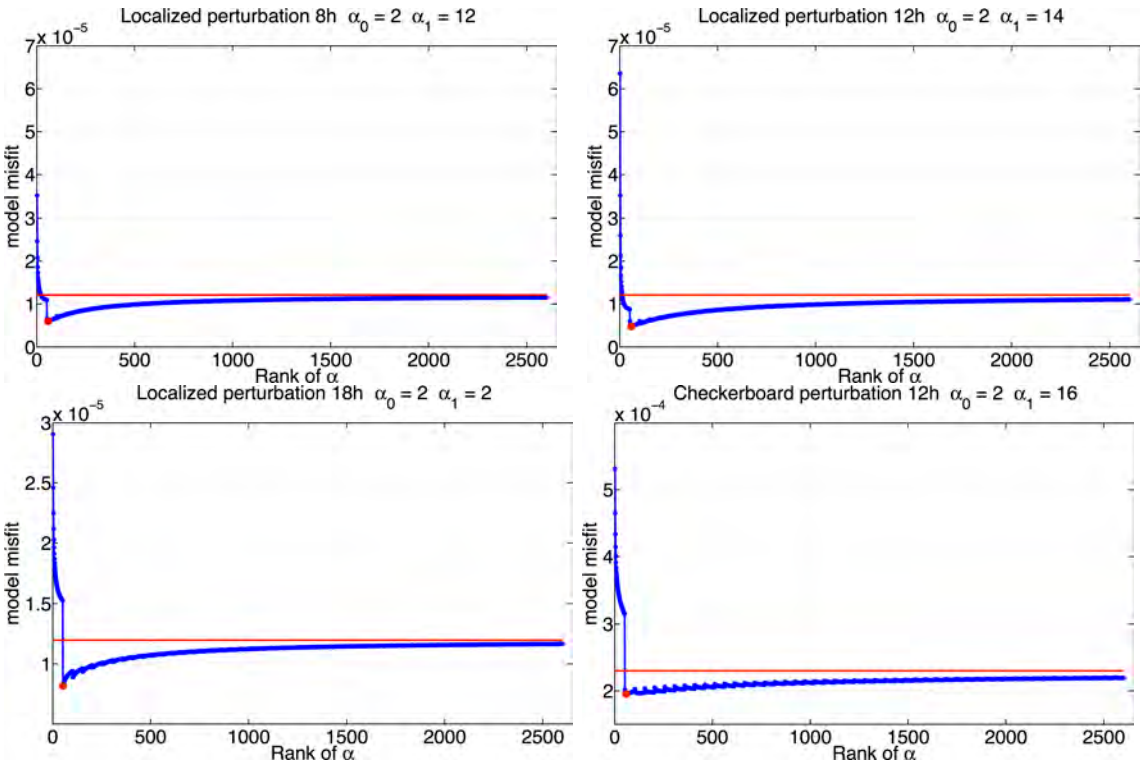


Figure 4.14: Errorcurves for the determination of the best combination of parameter α_0 and α_1 for a checkerboard perturbation and the localized perturbation for different daytimes. First, α_1 was modified for a fixed α_0. The red points mark the minimum errors, and the red lines are the sum over the cells of the target model, $\sum_i^{cells} \left(\frac{\delta Ne}{Ne^0}\right)_i^2$.

100 times larger than λ_1 and consequently, the weight given to the constraints would be negligible small. Therefore, the regularization parameters were fixed to the trace of the matrix as described in equation 4.41, but multiplied by two factors, α_0 and α_1.Thus, the best inversion result is given by

$$\frac{\delta \mathbf{m}}{\mathbf{m^0}} = \left(\mathbf{A^T A} + \alpha_0 \lambda \mathbf{I} + \alpha_1 \lambda_1 \mathbf{R}\right) \cdot \mathbf{A^T} \cdot \frac{\delta \mathbf{T}}{\mathbf{T^0}}. \tag{4.44}$$

The best combination of the parameters α are determined by calculation of an error curve keeping one parameter fixed and alternating the other. For each pair of these two parameters, the inversion was calculated with equation 4.44 and the difference between target and solution was plotted as a function of α. The best combination of these two parameters is defined by the minimum difference to the target model. The explored range for the two parameters is 0 to 100 with a sampling interval of 2.

Figure 4.14 shows these error curves as function of the parameters α for a localized perturbation at different daytimes ($8h$, $12h$, and $18h$), and the checkerboard perturbation at $12h$. For the localized perturbation at $12h$ and $8h$, as well as for the checkerboard test at $12h$, the error curves return the same value for α_0 and slightly different values for α_1. The difference in the second parameter is probably due to the sampling interval of 2. These results indicate the independence of the parameters α from perturbation and daytime.

However, at night, ($18h$), α_0 is 2 as for the other daytimes, but $\alpha_1 = 2$ is much smaller. At night, rays are reflected to a lesser extent, due to low ionization, therefore, the tomographic matrix has changed dramatically as compared to $12h$.

Figure 4.15 and Figure 4.16 compare the inversion results obtained with the damped least squares inversion without and with constraints for a checkerboard test and a localized perturbation at $12h$. In the first case, the regularization parameter was chosen by the L-curve criterion as described in Section 4.6.3, and in the second case by the error curve for the α parameters in Figure 4.14. The effect of smoothing is clearly visible for both target models. The noisy solution at altitude 100 km disappeares in the smoothed solution for the localized perturbation; and for the checkerboard perturbation, the shape of the negative perturbation at altitude 100 km and at the borders are better reconstructed.

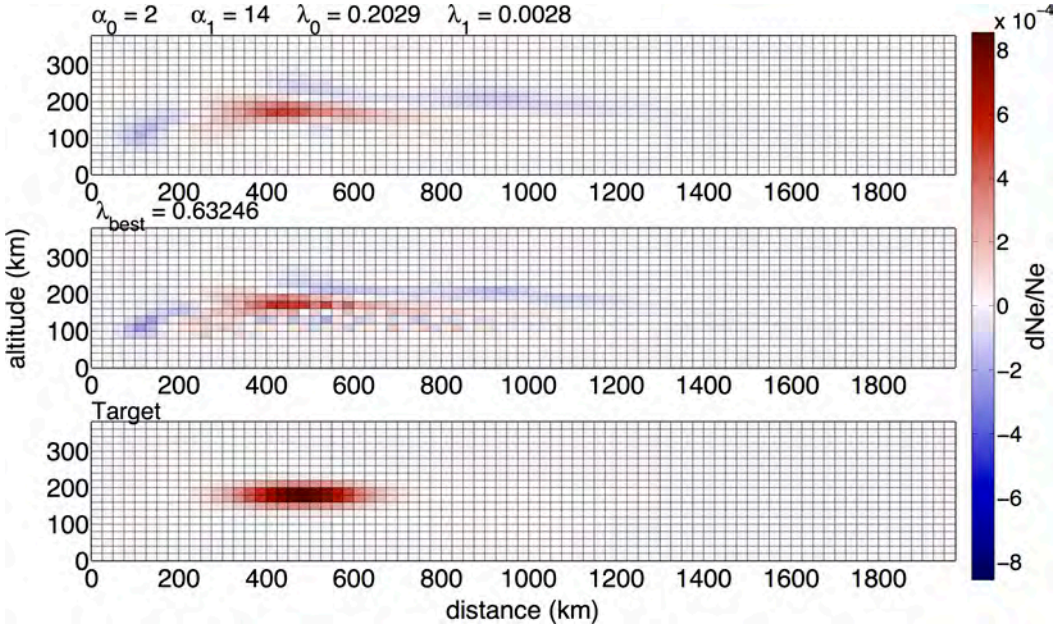


Figure 4.15: Damped least squares inversion vs. damped least squares with constraints for a localized perturbation. From bottom to top: Target model, damped least squares inversion with the best regularization parameter λ_{best} chosen from the L-curve, and damped least squares inversion with constraints.

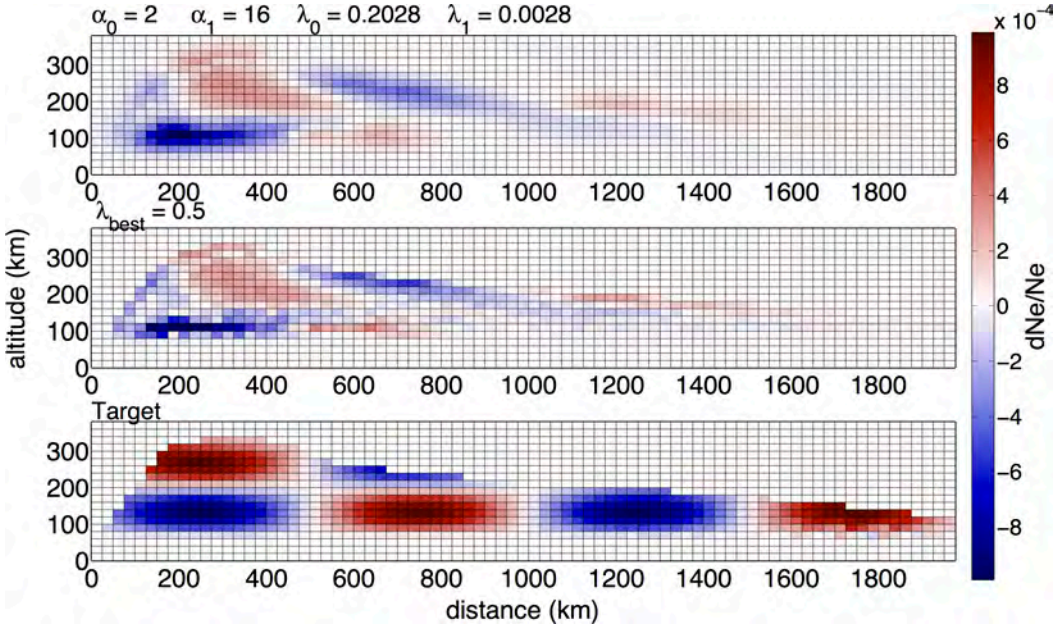


Figure 4.16: Damped least squares inversion vs. damped least squares with constraints for a checkerboard perturbation. From bottom to top: Target model, damped least squares inversion with the best regularization parameter λ_{best} chosen from the L-curve, and damped least squares inversion with constraints.

4.6.5 Conjugate Gradient algorithm

The results published by *Roy et al. (2014)* that were described above, are based on the solution of equation 4.18 by a damped least squares inversion. To take into account more the nonlinearity of the problem, introduced by the ray-path deflection, another method was applied, namely the Conjugate Gradient Method (CGM) *(Hestenes and Stiefel, 1952)*. This method is an iterative algorithm for the numerical solution of sparse systems of linear equations, like equation 4.16. In analogy to in the damped least square inversion, the minimum of a cost function F is searched, minimizing the error to the data and the discrepancy from the *a priori* model,

$$\begin{aligned} F(\mathbf{m}) &= \underbrace{\left(\mathbf{T}-\mathbf{T}^{obs}\right)^{T}\cdot\mathbf{W_d}\cdot\left(\mathbf{T}-\mathbf{T}_{obs}\right)}_{\text{data misfit}} \\ &+ \underbrace{\left(\mathbf{m}-\mathbf{m}_{prior}\right)^{T}\cdot\mathbf{W_m}\cdot\left(\mathbf{m}-\mathbf{m}_{prior}\right)}_{\text{model misfit}}. \end{aligned} \tag{4.45}$$

This equation a cost function similar to that in equation 4.21 except that the cost function is expressed on absolute values and the damped least squares solution was transformed into a more generally regularized least-squares solution. The latter avoids the choice of an optimum regularization parameter λ, a problem, which has no simple solution.

In general, the weight matrices $\mathbf{W_m}$ and $\mathbf{W_d}$ can be anything. In Bayesian statistics one treats the inverse problem from a statistical point of view combining a priori information about the data and the model with the data that are actually measured. The weight matrices reflect true physical *a priori* information for data and model. In such a Bayesian approach, the weight matrices are given by the covariance matrices for data and model,

$$\mathbf{W_m}=\mathbf{C_m^{-1}} \quad \mathbf{W_d}=\mathbf{C_d^{-1}} \tag{4.46}$$

The CGM is similar to the method of steepest descent, but with a small difference. The method of steepest descent starts at an arbitrary point, choosing the direction, in which F decreases most quickly, which is the direction opposite to the gradient of F. After a series of steps $x_{(1)}$, $x_{(2)}$ ... the errors are sufficiently small and the solution is the desired one. This method often takes steps in the same direction as earlier steps (Figure 4.17), undoing success achieved previously. It would be better to move in none interfering directions, exhausting the error in one direction and ensure that no additional error is introduced in that orientation in the next step.

Therefore, the directions in the CGM are *conjugate* (perpendicular in some sense) to all previous ones. The first iteration of CGM is identical to the first iteration of

Steepest Descent, meaning that one moves in the direction opposite to the gradient. The gradient of the cost function (equation 4.45) is given by *(Tarantola, 2005, Chapter 6)*,

$$\mathbf{g_0} = \mathbf{C_M} \cdot \frac{\partial F}{\partial \mathbf{m}} = \mathbf{C_M} \cdot \mathbf{G} \cdot \mathbf{C_D^{-1}} \cdot (\delta \mathbf{T} - \mathbf{A}, \cdot \delta \mathbf{m}) + \delta \mathbf{m}, \quad (4.47)$$

where $\mathbf{G}$ is a matrix that contains the partial derivatives $\frac{\partial \mathbf{T}}{\partial \mathbf{m}}$.

However, an important and critical point in the CGM is the choice of a step length, which quantifies how far one moves in one direction. Generally, this is done by performing a line search that means one searches a step length that minimizes the cost function in the search direction. In this work, a modified algorithm of the conjugate gradient was applied, following *Tape et al. (2007)*. Their work uses a polynomial line search. The CGM for k iterations can be summarized by the

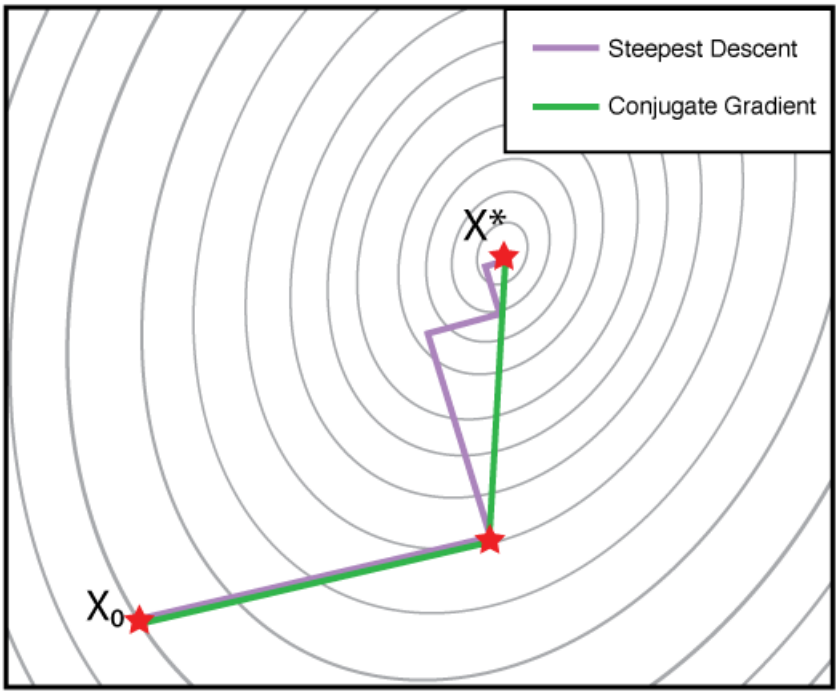


Figure 4.17: Conjugate gradient vs. Steepest Descent Method. Solid grey lines are contours of the quadratic form with constant value $F(x)$. Starting at the arbitrary point x_0, the conjugate gradient method (green) arrives at the solution x^* after two iterations, the steepest descent method (purple) after five iterations. Source: `http://visiblegeology.com/rowan/presentations/2013/MSc2PhD/#/4/2`

following steps.

1. For an initial arbitrary starting model $\mathbf{m}_0$, calculate the misfit $\mathrm{F}(\mathbf{m}_0)$ (equation 4.45) and the steepest ascent vector $\mathbf{g}_0$ (equation 4.47).

2. Calculate the search direction $\mathbf{p}$. In the first iteration the search direction is set

to minus the gradient. In subsequent iterations conjugate search directions are calculated based on the previous search directions, in order to avoid searching in directions that have been searched before and spoiling the minimization along another :

$$\mathbf{p}_k = \begin{cases} -\mathbf{g}_k & \text{k=1,} \\ -\mathbf{g}_k + \beta_k \mathbf{p}_{k-1} & \text{k>1} \end{cases}$$

The parameter β_k is given for instance by the Polak-Ribíre formula *(Tarantola, 2005, Chapter 6)*.

$$\beta_k = \frac{(\mathbf{g}_k - \mathbf{g}_{k-1}) \cdot \mathbf{C_M^{-1}} \cdot \mathbf{B_0} \cdot \mathbf{g}_k}{\mathbf{g}_{k-1} \cdot \mathbf{C_M^{-1}} \cdot \mathbf{B_0} \cdot \mathbf{g}_{k-1}}, \tag{4.48}$$

where $\mathbf{B}_0$ is a preconditioning operator. The simplest choice for $\mathbf{B}_0$ is the identity matrix. There exist other formulations of the parameter β, too.

3. Perform a polynomial line search *(Tape et al., 2007)* to choose a step length v_k that minimizes $F(\mathbf{m} + v_k \cdot \mathbf{p})$ that is the misfit in the search direction $\mathbf{p}$. The choice of this step size, which determines how far one should go in the search direction, is a key decision in this algorithm. To find a step size

   (a) Calculate at first a trial step, or test parameter v_t by interpolating the misfit function F with a quadratic polynomial

   (b) Update the model in the search direction $\mathbf{p}$ to obtain a test model $\mathbf{m}_t = \mathbf{m} + v_t \cdot \mathbf{p}$ and calculate the misfit for this test model $F(\mathbf{m}_t)$

   (c) Interpolate the misfit function F between the current misfit $F(\mathbf{m}_t)$ and the misfit for the test model $(\mathbf{m}_t)$ by a quadratic polynomial and choose the step size v_k that gives the minimum of the polynomial.

4. Update the model $\mathbf{m}_{k+1} = \mathbf{m_k} + v_k \cdot \mathbf{p}$

5. If the misfit $F(\mathbf{m}_{k+1})$ is suitable small, $\mathbf{m}_{k+1}$ is the desired model, otherwise restart from 2.

4.6.6 Results

The above described CGM was applied to synthetic data generated by ray tracing TDR in the continuous *a priori* ionospheric model *NeQuick (Radicella and Leitinger, 2001)* for October at 12 UT with a solar flux of 198.1 SFU. The data was simulated by tracing 1071 rays with elevation angles between 10° and 60° and in a frequency range of $6-16$ MHz in the ionospheric model plus an additional, known perturbation

δNe^{target}. Again, the solution of the inverse problem, the obtained perturbation $\delta\mathbf{m}$, has to correspond to δNe^{target}.

The CGM was first validated with the analytical frozen rays (Section 4.6.1), satisfying the hypothesis that the electron density perturbation δNe^{target} modifies only the velocity of EM waves, *i.e.*, rays are *frozen* in the *a priori* model configuration. This theoretical data set represents the idealized case of no ray-path perturbation.

In order to choose the data and model covariance matrix, the idealized CGM was applied to a localized perturbation of 5%. The model variance signifies the allowed deviation of the solution from the *a priori* ionospheric model, the data variance signifies the precision in the data. A small model variance does not allow the solution to differ much from the background ionospheric model. The model covariance was modified from 1% to 10% and the data covariance was fixed to $8.3 \cdot 10^{-6} s$.

Figure 4.18 and 4.19 illustrate the evolution of data and model misfit as a function of iterations.

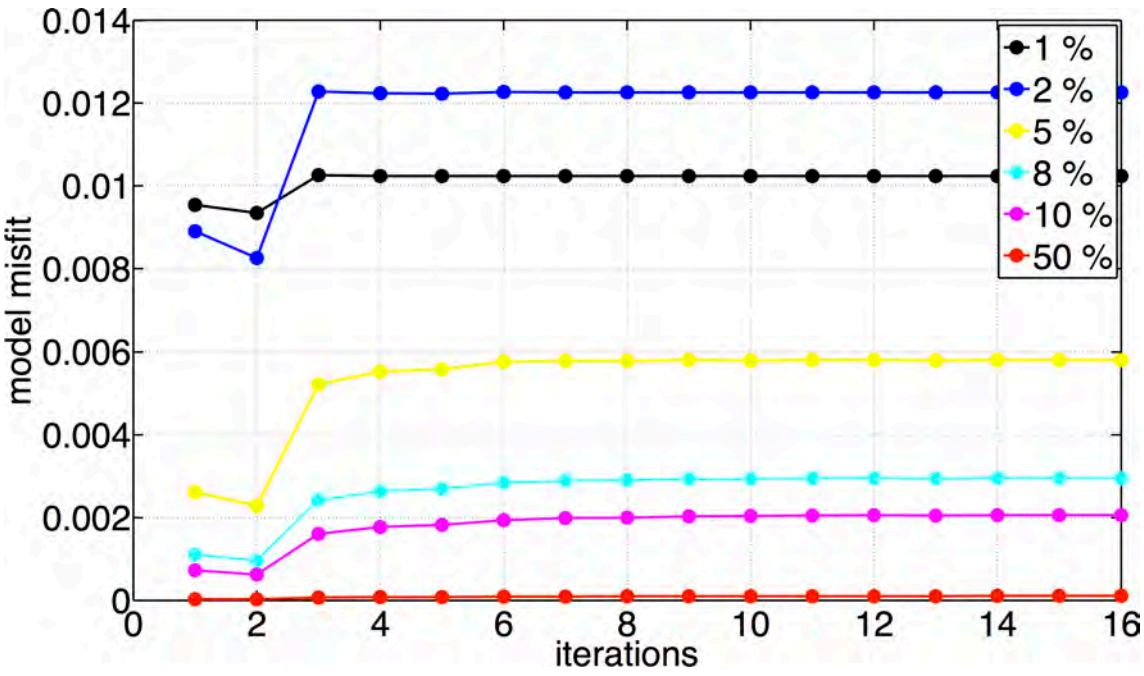


Figure 4.18: Model misfit as a function of iterations for different model covariances and a fixed data covariance of $8.3 \cdot 10^{-6} s$. The CGM was applied to a localized perturbation of 5% in the idealized case of frozen rays.

Smaller model variances do not allow the solution to differ much from the *a priori* ionospheric model, therefore, data and model misfit are quickly converging for 1% and not much changing with iterations. The larger the model variance, the more the model can change, and the longer it takes to converge. After 16 iterations, the data misfit is smallest for the largest covariance.

The CGM method was now applied to two synthetic, known electron density

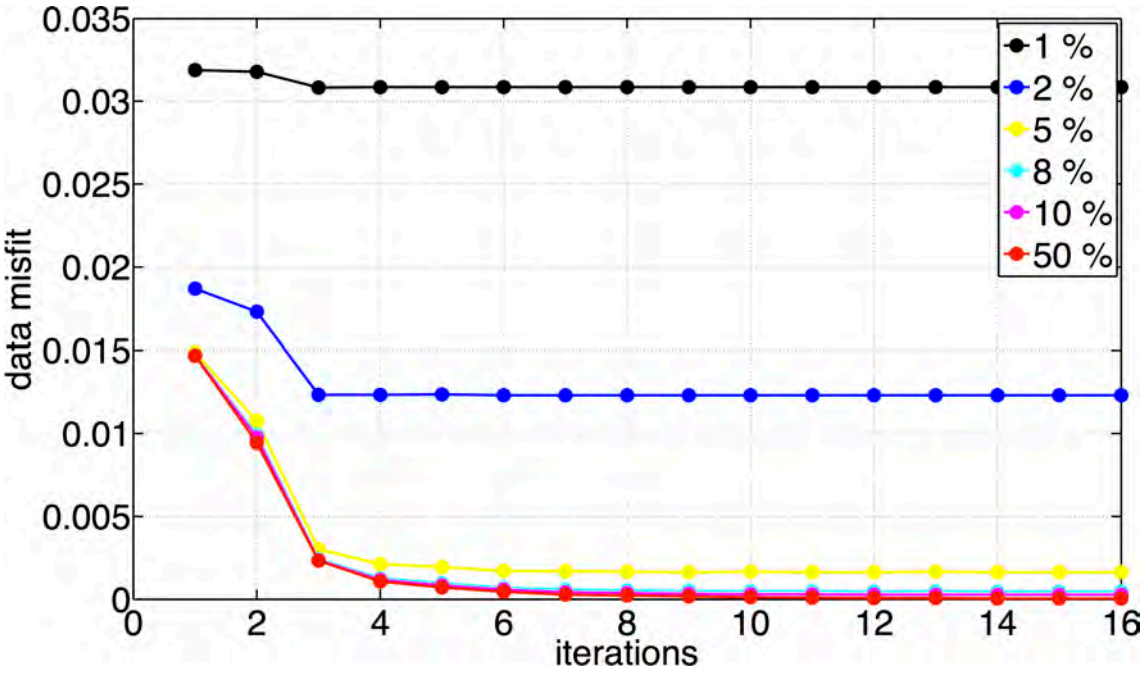


Figure 4.19: Model misfit as a function of iterations for different model covariances and a fixed data covariance of $8.3 \cdot 10^{-6} s$. The CGM was applied to a localized perturbation of 5% in the idealized case of frozen rays.

perturbations that are a localized perturbation and a checkerboard perturbation of 5%. The covariance matrix for the model and data, $\mathbf{C_M}$ and $\mathbf{C_M}$ are diagonal matrices with 10% model variance and $8.3 \cdot 10^{-}6s$ data variance, respectively. Figure 4.20 shows the evolution of the solution with iterations for the localized perturbation. Already after four iterations, the perturbation is at the right horizontal and vertical place and with further iterations, the surrounding noise is reduced. The data misfit (equation 4.45) is reduced with iterations and converges, as shown in Figure 4.21. The model misfit (equation 4.45) increases, because the solution is a difference to the *a priori* ionospheric model. After iteration seven, it is constant.

Similar results are obtained for the checkerboard perturbation (Figure 4.22). After six iterations all negative and positive perturbations up to 500 km are well recovered. In larger distances the perturbations are poorly reproduced due to the sparse ray coverage (Figure 4.9). For the checkerboard perturbation, the data misfit decreases and converges, the model misfit increases and converges to a constant value, too (Figure 4.23).

These preliminary tests show that the Conjugate Gradient method is working in the idealized case of the frozen rays. In a next step, this method has to be applied to synthetics considering that the ray-path in the *a priori* ionospheric model is not the same as in the perturbed ionosphere (Section 4.5.4).

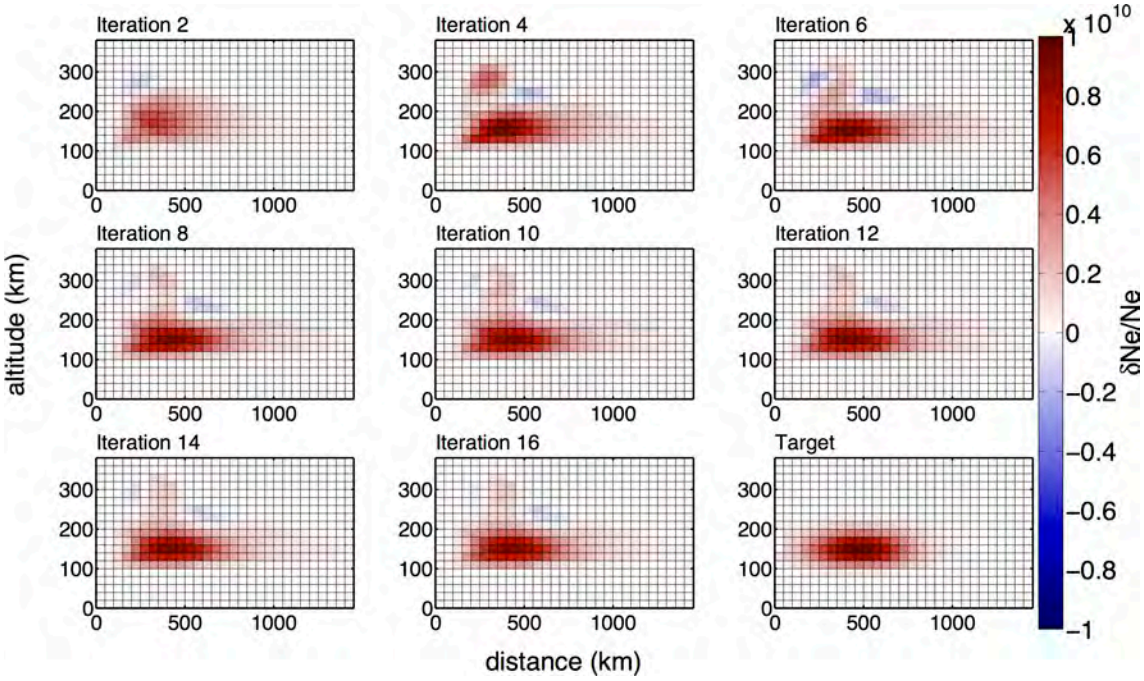


Figure 4.20: Solution of the Conjugate Gradient Method in the idealized case of the frozen rays for a localized perturbation.

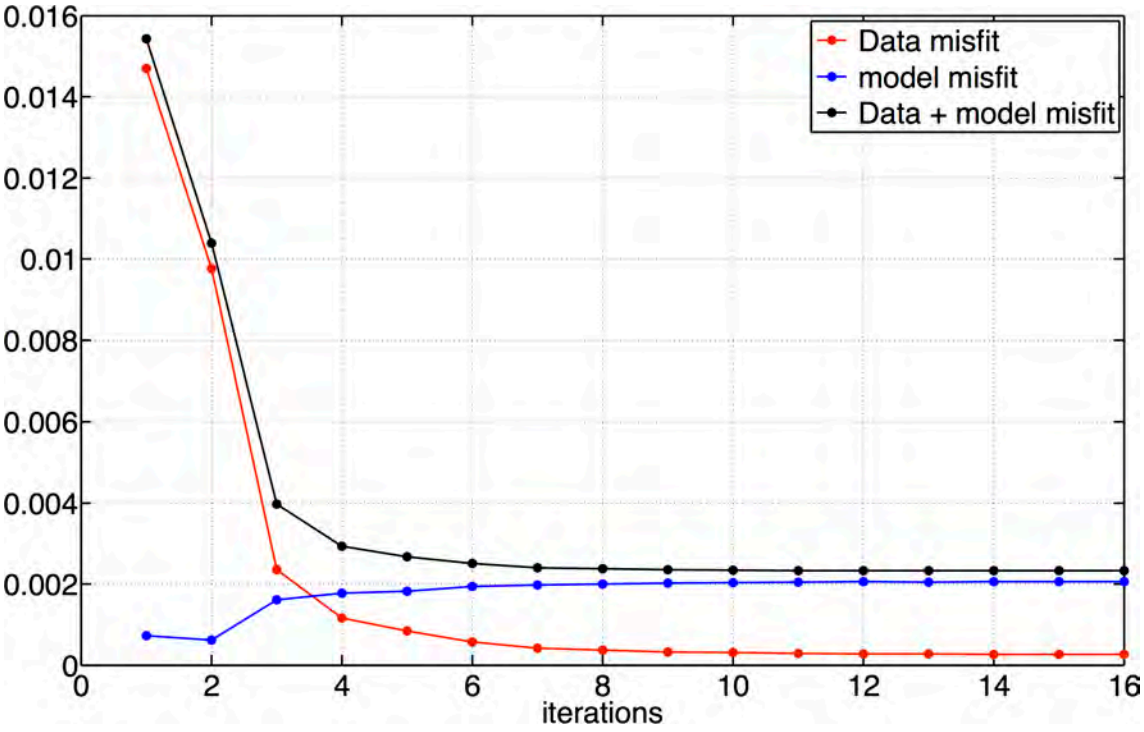


Figure 4.21: Data and model misfit (equation 4.45) as a function of iterations for the Conjugate Gradient Method for a localized perturbation.

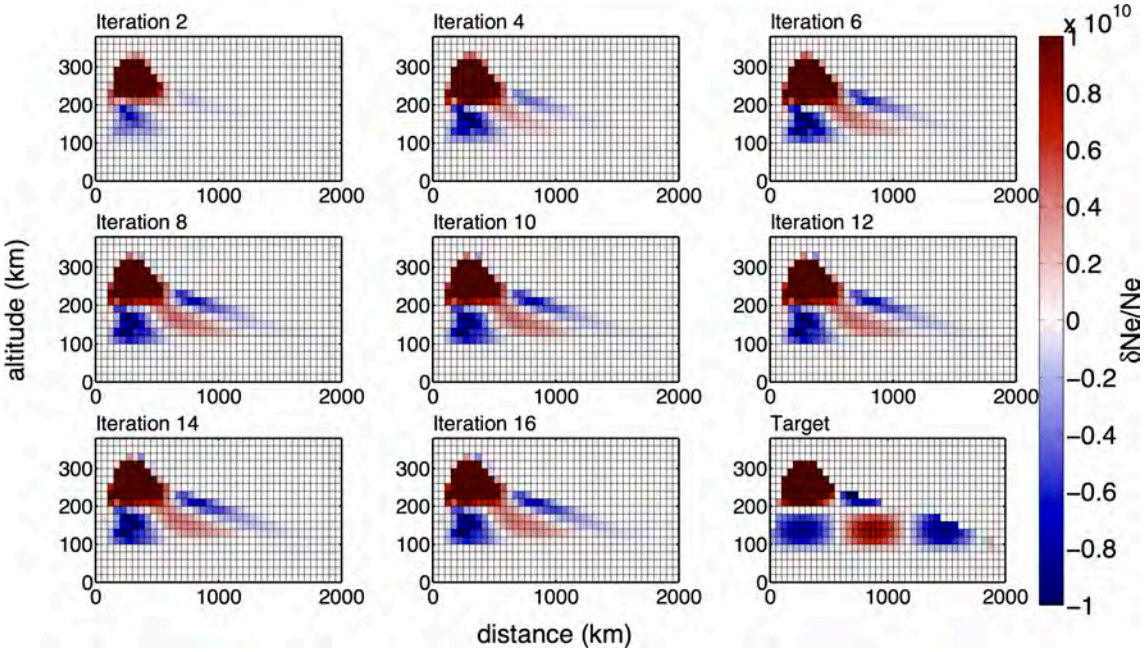


Figure 4.22: Solution of the Conjugate Gradient Method in the idealized case of the frozen rays for a checkerboard perturbation.

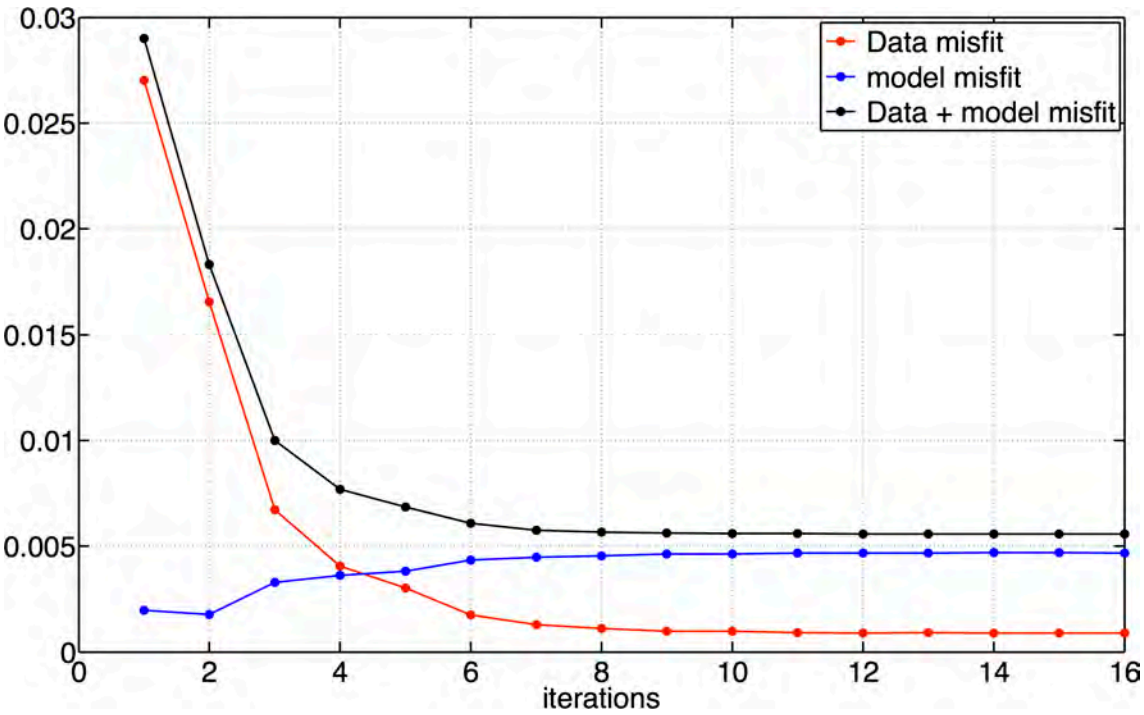


Figure 4.23: Data and model misfit (equation 4.45) as a function of iterations for the Conjugate Gradient Method applied to a checkerboard perturbation.

Chapter 5

Resolution tests

In order to evaluate the resolution of the tomography method, tests were performed for several checkerboard tests and localized perturbations of different size and amplitude. In this chapter, inversion results obtained for a selection of target models are discussed. This selection includes large scale perturbations of several hundred kilometers as well as very small perturbations of only few kilometers. Localized perturbations were created using gaussian functions and their size was estimated at the 3 dB level of the maximum amplitude.

5.1 Effect of cell dimension on inversion results

In order to gain insight into the possible resolution using this tomography method, a first test was performed with cell dimensions being increased in altitude for a fixed distance of 50 km and vice versa (for a fixed altitude of 20 km cell dimensions are enlarged in distance). All of these test sets use the same amplitude of electron density perturbation of 0.1%.

In the second test, the amplitude of the electron density perturbation was scaled up for a fixed grid, to find out the order of still resolvable perturbations. The v and the $v\&r$-method were used for inversion of both test sets, and the best value of regularization was chosen as described in Section 4.6.3. In summary, these tests show the independence of the inversion results of the grid and demonstrate the possibility of this method to reconstruct small scale anomalies and separate them spatially.

5.1.1 Localized perturbation

Two resolution tests of localized perturbations with dimension 470×71 km and 235×47 km were performed. The results are illustrated in Figure 5.1 and 5.2 respectively.

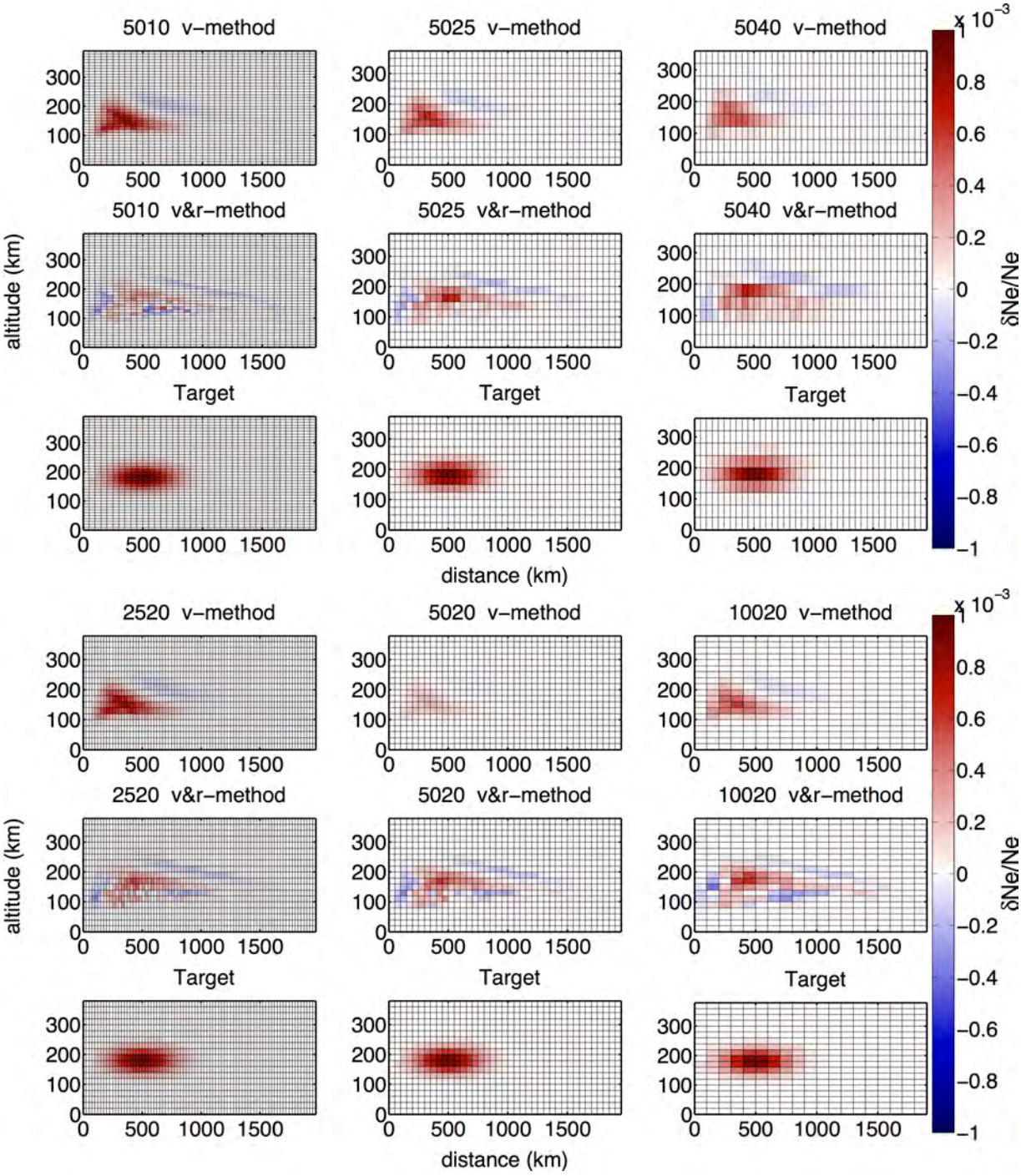


Figure 5.1: Inversion results for a localized perturbation of 0.1% using the v-method and the $v\&r$-method. The cell dimensions are 50 km in distance and varied in altitude from 10 km to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are 20 km in altitude and varied from 25 km to 50 km and 100 km in distance.

In both cases, the correct location (500 km distance, 180 km altitude) of maximum anomaly in the target model is only reconstructed by the $v\&r$-method. The v-method identifies a high $\delta Ne/Ne$ anomaly in the general area of 200 −500 km in

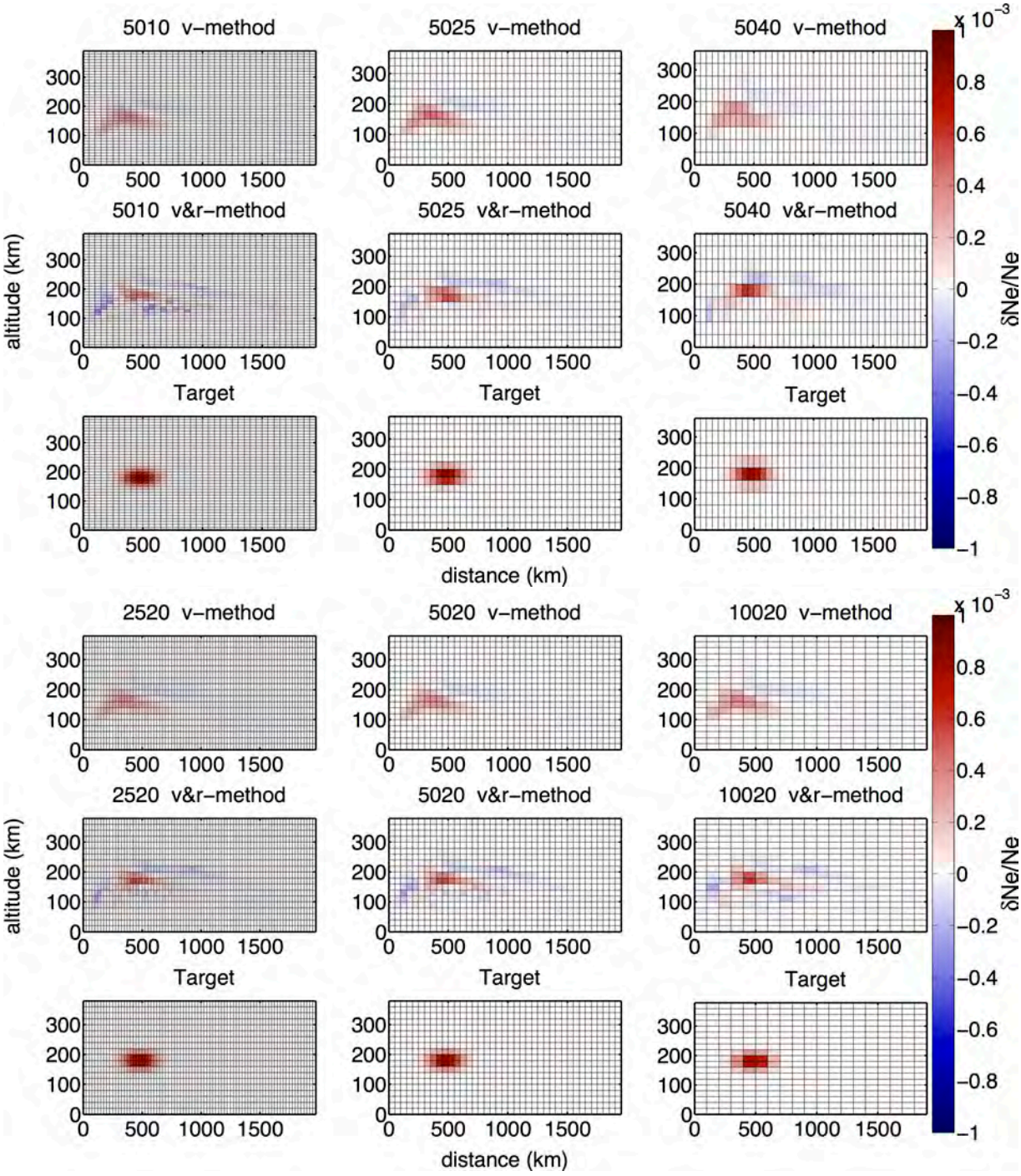


Figure 5.2: Inversion results for a localized perturbation of 0.1% using the v and the $v\&r$-method. The cell dimensions are fixed to 50 km in distance and increased in altitude from 10 km to 25 km and 40 km in the top panel. In the bottom panel cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance.

horizontal distance but slightly mislocates it at 100 −200 km in altitude and does not reproduce its shape. Both v- and $v\&r$-method introduce large-scale low $\delta Ne/Ne$ anomalies that do not correspond to any feature of the target model, where the

sign of theperturbation is always positive. As discussed in section 4.6.2, these are artifacts to the nonlinear effects of ray-path deflections, which (endpoints not being fixed) can in principle result in a faster propagation time even if the velocity perturbation is negative. These are more clearly visible for larger cell dimensions. The grid sizes does not influence the result.

5.1.2 Checkerboard tests

The inversion results for an unrealistic checkerboard test with alternately changed perturbation (positive and negative) are illustrated in Figure 5.3 and Figure 5.4, where the size of the perturbations is the same, but the second checkerboard test was shifted downwards by 60 km.

The v-method only recognizes a positive and negative perturbation at around 200 km and 300 km altitude at 500 km horizontal distance. The two negative perturbations at the top and at the right boundary cannot be separated by the v-method. The shape of the positive anomaly is not well reconstructed and mislocated. While $v\&r$-method cannot separate the two negative perturbations at the top neither, more details of the perturbation, also in larger horizontal distances are recognized in the inversion. These resolution tests show that the grid size is not influencing the inversion results. Results are not worse or better with a smaller or larger grid size.

The test with the downwards shifted checkerboard test confirms that the $v\&r$-method performs quite better than the v-method, because all perturbations are recognized, although not clearly separated by each other. In the v-method the missing of details in the inversion result is obvious, and only two perturbations at between 200 km and 300 km altitude are visible.

In the unrealistic case of the radical checkerboard test (Figure 5.5) the sign of the perturbation is alternating from cell to cell. While the v-method can partly reproduce the pattern of alternating positive and negative perturbation between 200 and 300 km altitude, the $v\&r$-method finds large anomalies at 100 km altitude that do not correspond to the pattern in the target and mask the perturbations at larger altitudes.

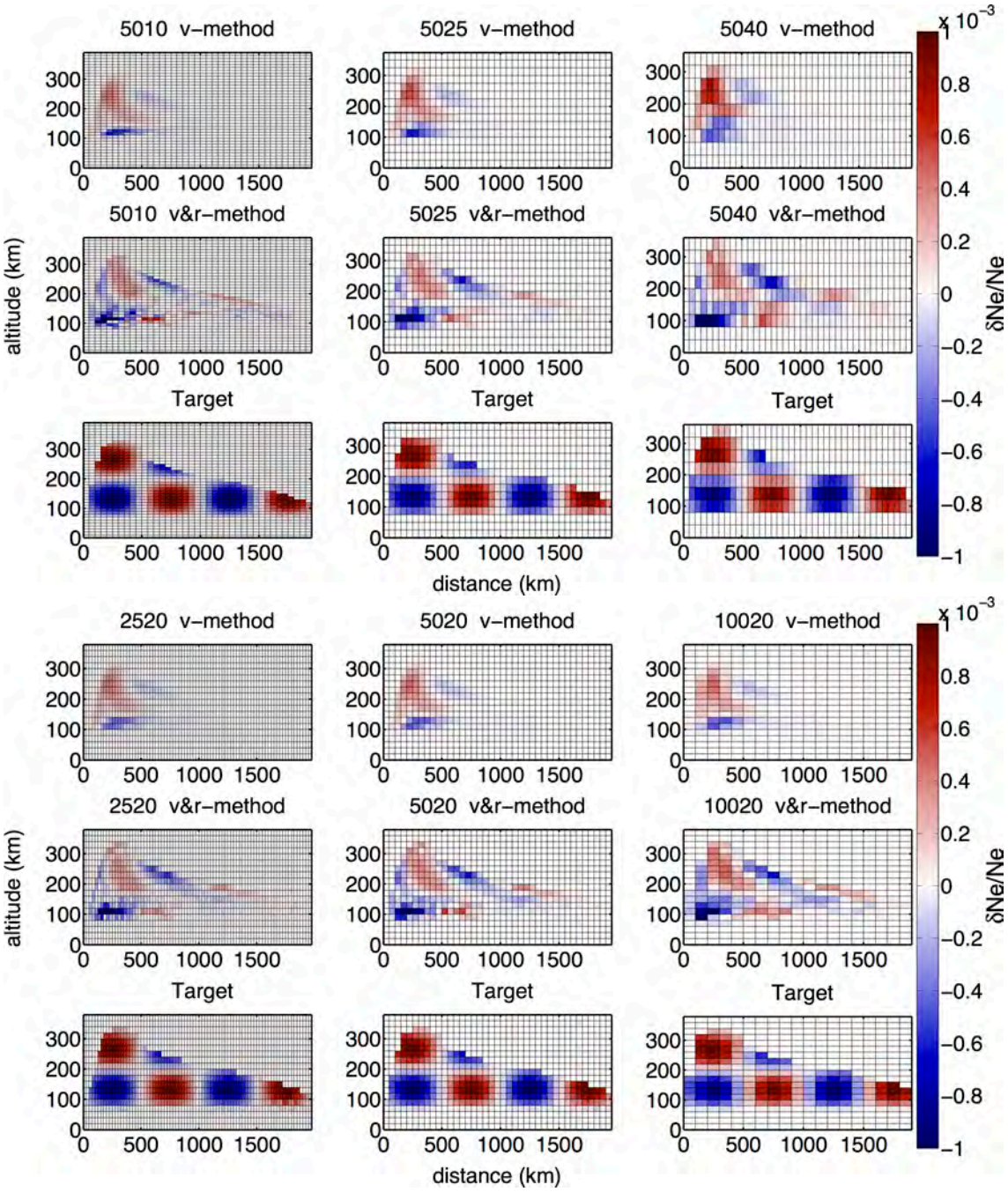


Figure 5.3: Inversion results for a checkerboard perturbation of 0.1% using the v- and the $v\&r$-method. The cell dimensions are fixed to 50 km in distance and enlarged in altitude from 10 km, to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance.

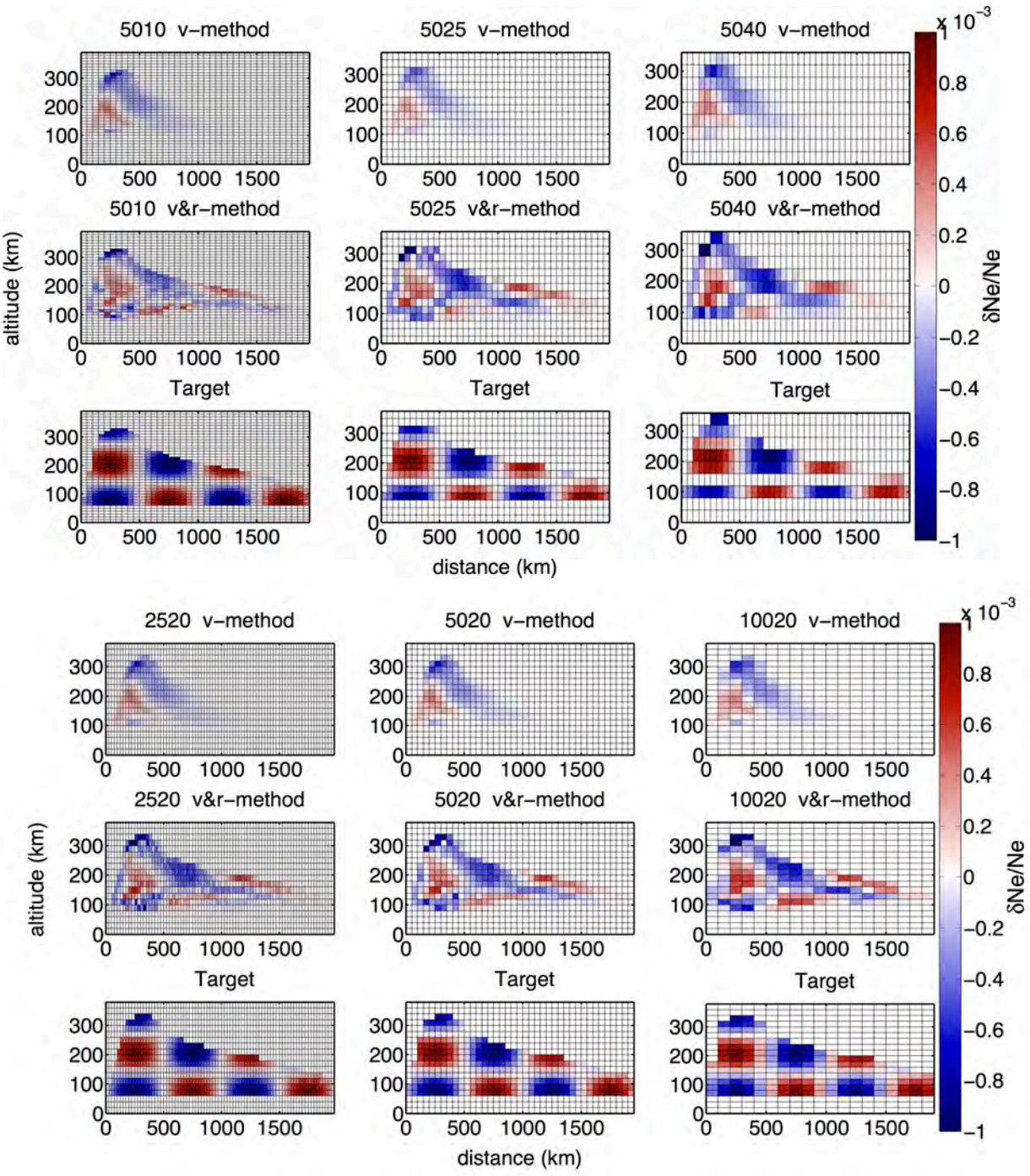


Figure 5.4: Inversion results for the checkerboard perturbation in Figure 5.3, but shifted downward at around 60 km. The v- and $v\&r$-method were used for inversions. The cell dimensions are fixed to 50 km in distance and enlarged in altitude from 10 km to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance.

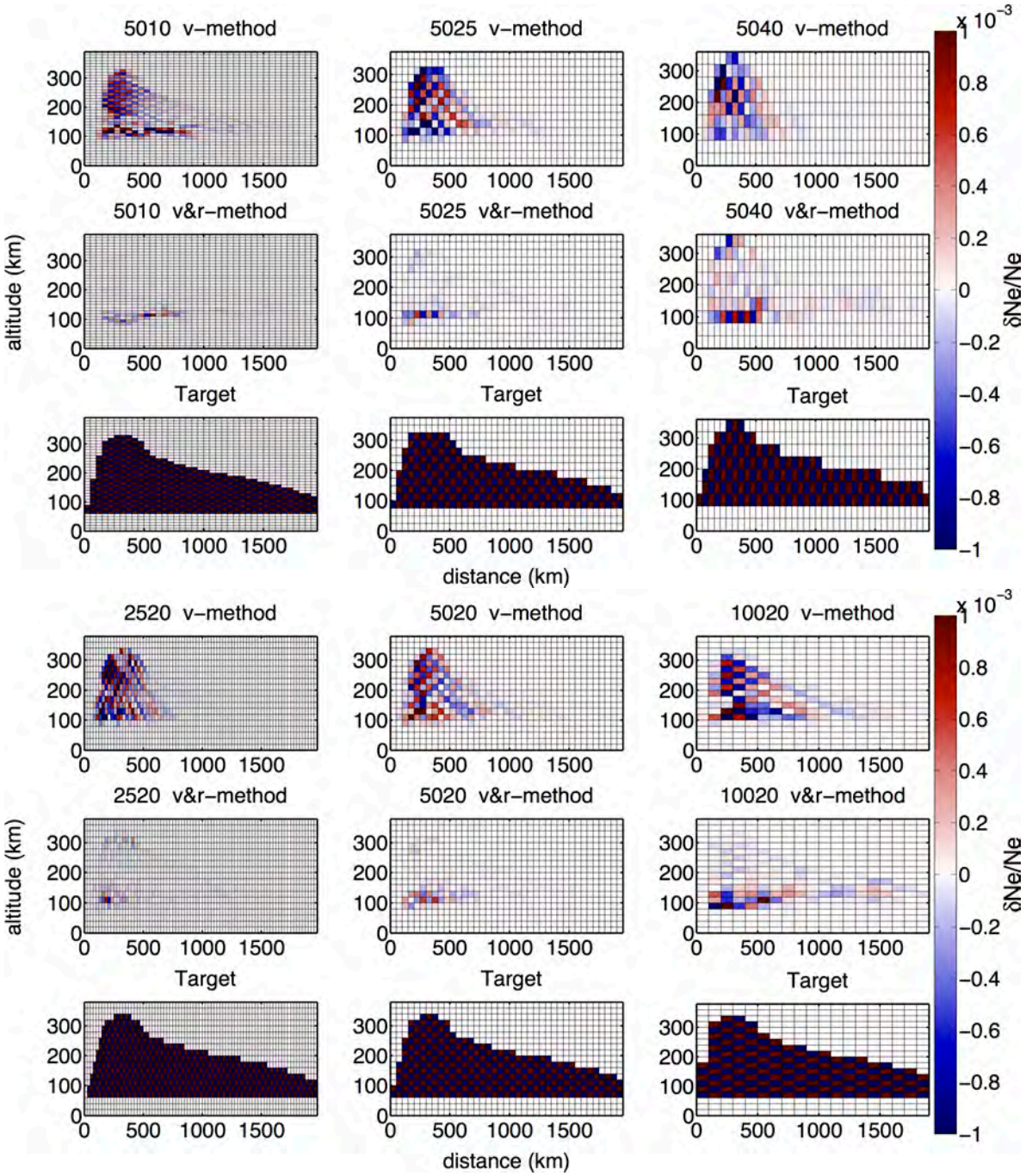


Figure 5.5: Inversion results for a checkerboard perturbation of 0.1% using the v and the $v\&r$-method. The cell dimensions are fixed to 50 km in distance and enlarged in altitude from 10 km, to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance. The sign of the perturbations is alternating from cell cell (radically).

5.1.3 Two localized perturbations ($165 \times 47\,\mathrm{km}$)

The tomography method is also able to recognizes and separate smaller scale anomalies, *e.g.,* Figure 5.6. These tests represents a realistic case of a depletion and an enhancement of electron density in the ionosphere. The two localized (positive and negative) perturbations are located at 300 and 500 km respectively, in 250 km altitude. The $v\&r$-method identifies both perturbations and is able to separate them, independent of the grid. Both maxima anomalies are reproduced well. In the v-method the negative anomaly is always correctly located, but the positive localized perturbations is smeared and not well recovered.

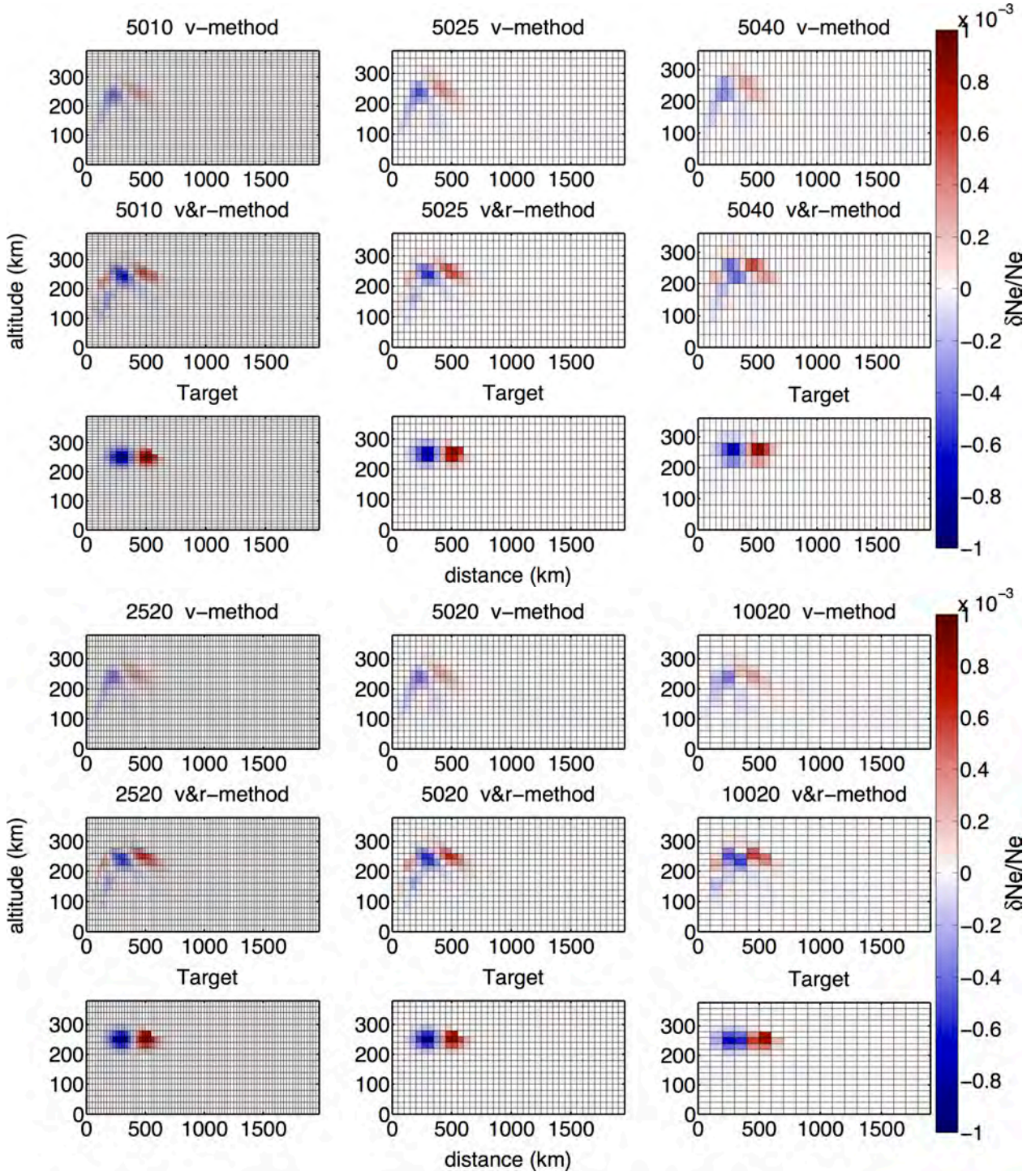


Figure 5.6: Inversion results for two localized perturbation of 0.1% using the v- and the $v\&r$-method. The cell dimensions are fixed to 50 km in distance and increased in altitude from 10 km, to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance.

5.1.4 Cell perturbations

Supposing the unrealistic case of a perturbation in only one cell (Figure 5.7), the v-method performs much better than the $v\&r$-method, because only the v-method is able to reproduce the correct location of the perturbed cell, even for the smallest cells of 25×20 km in distance and altitude. By contrast to the checkerboard tests and the localized perturbations, in this tests, the target model changes when the dimension of the cells are enlarged.

The $v\&r$-method does not find the cell perturbation in any case, even for the largest pixel of 100 km horizontal distance, and the inversion result is strongly damped and blurred over the model space, so that no maximum anomaly can be found.

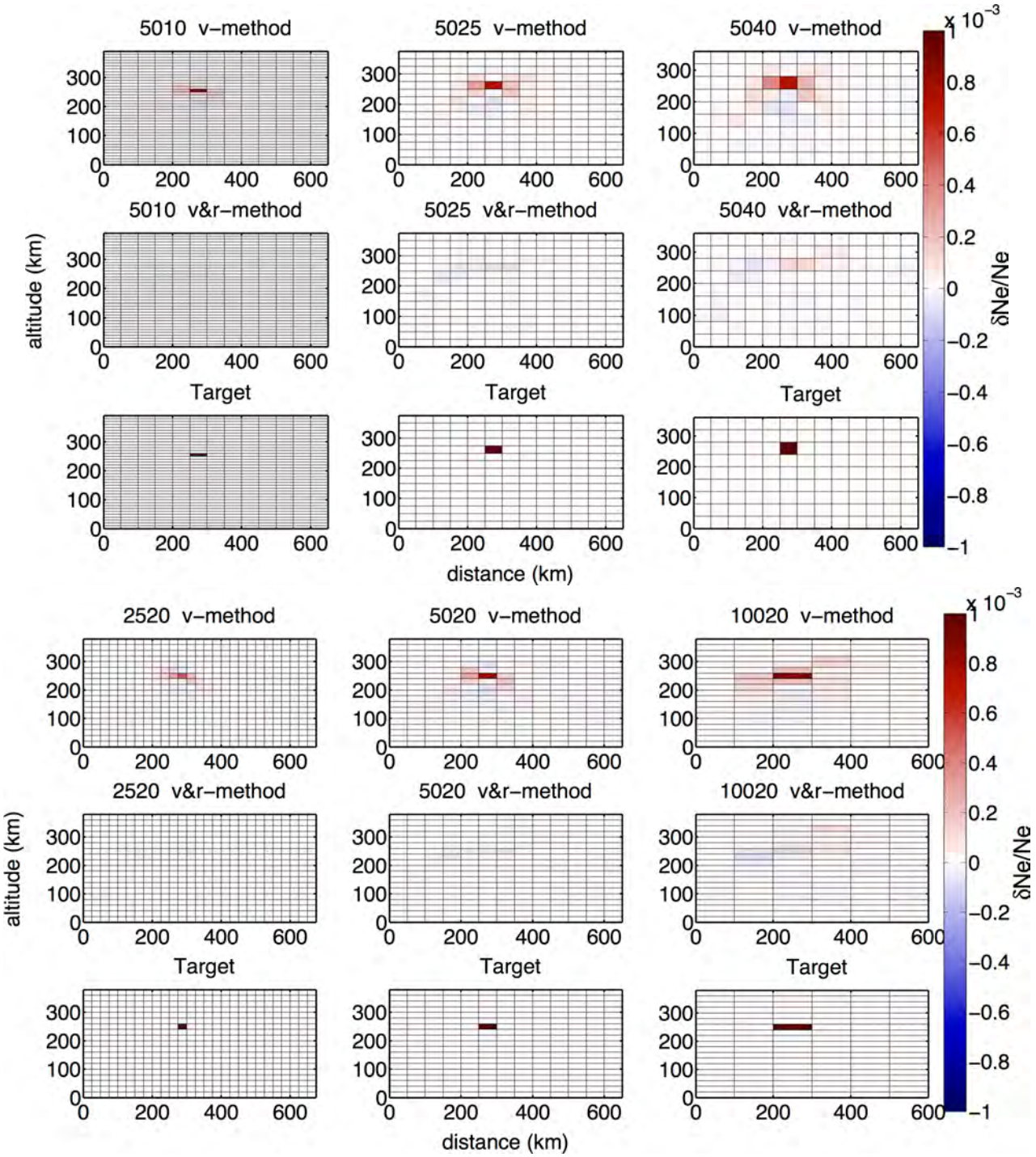


Figure 5.7: Inversion results for a localized perturbation in one cell of 0.1% using the v and the $v\&r$-method. The cell dimensions are fixed to 50 km in distance and increased in altitude from 10 km, to 25 km and 40 km in the top panel. In the bottom panel, cell dimensions are fixed to 20 km in altitude and enlarged from 25 km to 50 km and 100 km in distance. The size of perturbations is changing with the grid.

5.2 Which amplitude of perturbation is still resolvable ?

In the second test set, cell dimensions are fixed to 25 km × 20 km and only the perturbation amplitude is increased. Figure 5.8 shows the inversion results for the v and $v\&r$-method for the same localized perturbation as in Figure 5.2. The results indicate that small perturbations of electron density down to 10^{-3} % and up to 20% can still be detected by both inversion methods. Perturbation of 10^{-4}% are too small to be resolved by either of the inversion methods. While these small perturbations are still somehow visible in the v-method, the $v\&r$-method detects nothing.

Independent of the amplitude, the v-method misallocate the perturbation, and in both method these large scale negative perturbations discussed previously are visible. Both inversion methods cannot solve localized perturbation larger than 50%, neither the amplitude nor the shape of the perturbation are resolved.

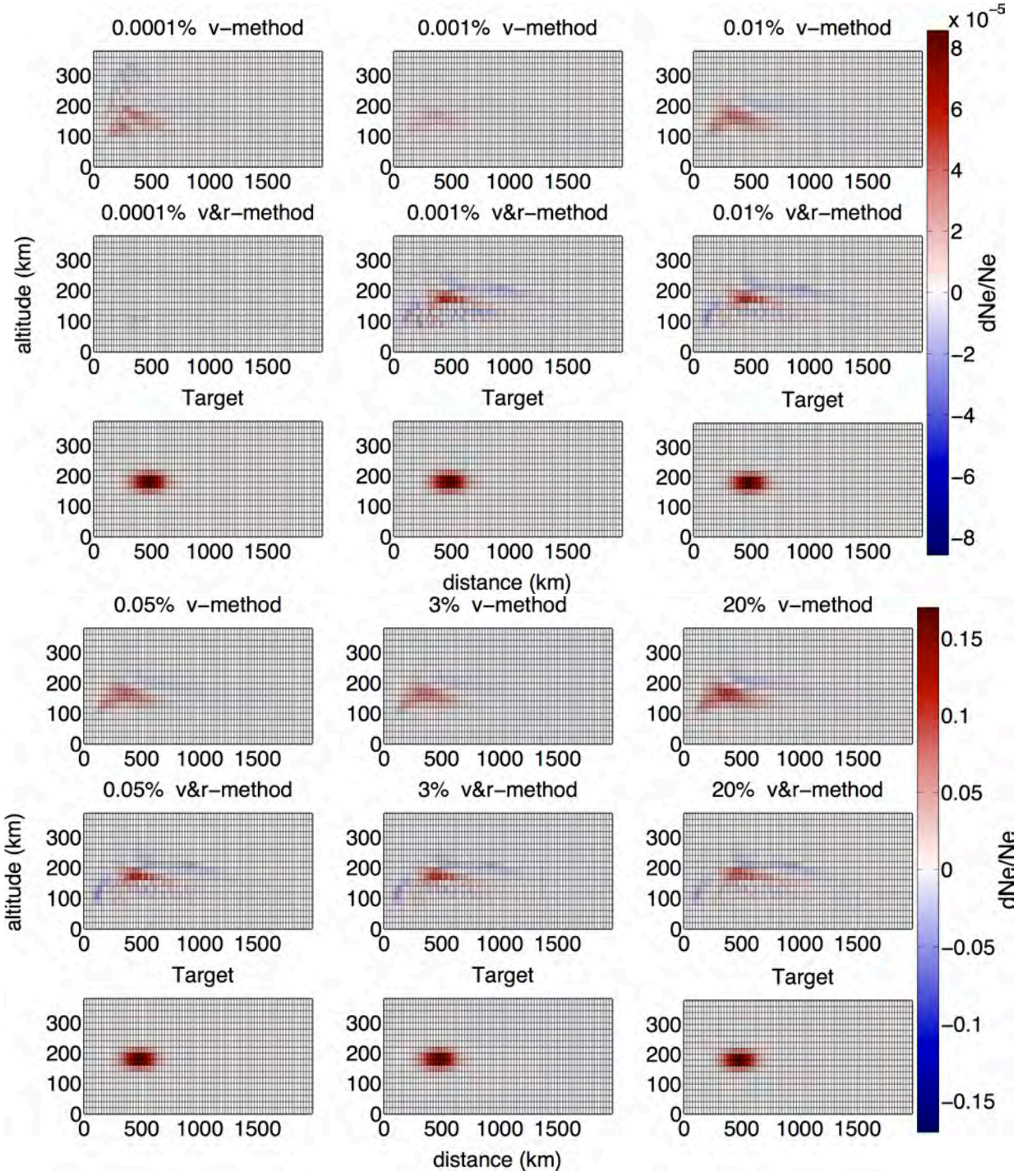


Figure 5.8: Inversion results for a localized perturbation of 0.1% using the v and the $v\&r$-method. The cell dimensions are fixed to 25 km × 20 km and the amplitude of the perturbation is increased.

5.3 Conclusions

The resolved area in all tests is limited by the uneven ray coverage, corresponding to a single OTH radar. Due to the strong variation of the ionospheric background

(electron density equal to zero at 80 km, but in the order of $10^{11} - 10^{12} e/m^3$ at 300 km), EM waves emitted by the radar are particularly sensitive to the zone where the rays are reflected, where the plasma frequency f_p and the emission frequency f_e are approximately the same. This defines the area of good coverage (independent of inversion method), which is the region between 100 and 300 km altitude and 500 −2500 km in range (Figure 5.9).

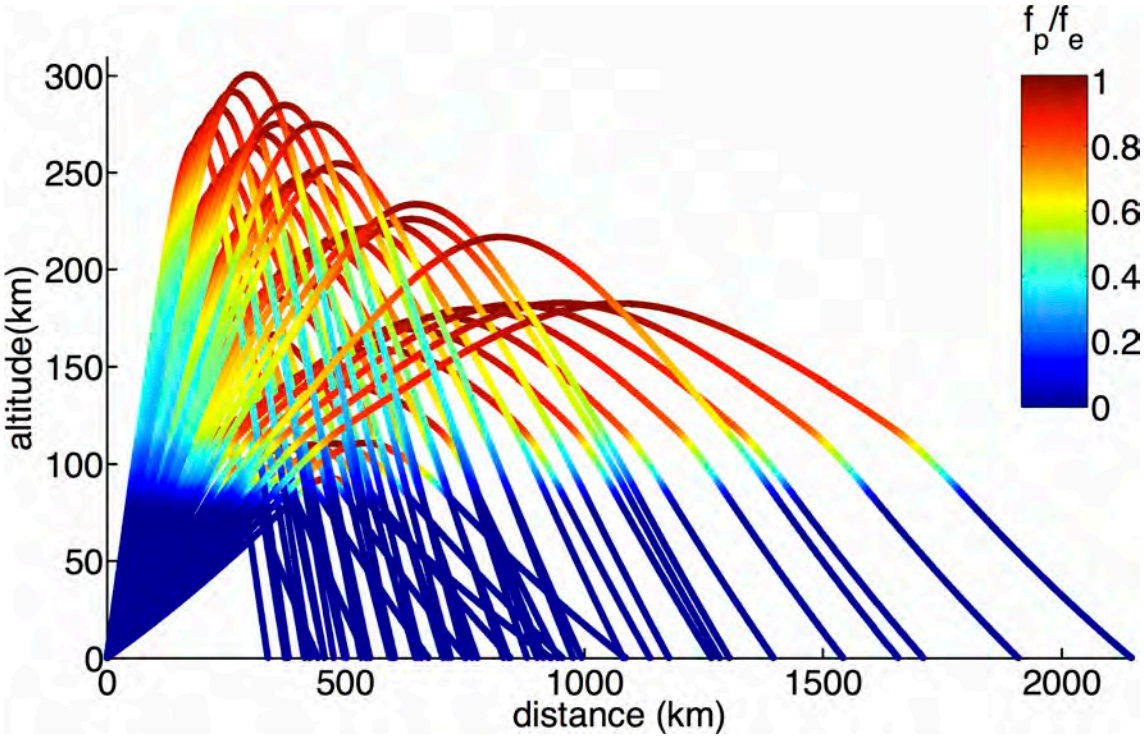


Figure 5.9: Ratio of plasma f_p to emission frequency f_e, calculated along the ray-path illustrating the sensitivity of the rays to the medium. The highest sensitivity is the zone where the ratio is 1.

In summary, all grid tests show that the v- and the $v\&r$-method generally enables the identification of perturbations above 100 km altitude. Furthermore, these tests indicate independence of the inversion results of the grid, or, in other words, the grid does not not influence drastically the inversion results.

In general, for large scale perturbations, the $v\&r$-method seems to perform better as compared to the v-method, which is illustrated in Figures 5.4 and 5.1 for the checkerboard and the localized perturbations. The checkerboard tests demonstrate that the $v\&r$-method can reproduce more details of the target model, whereas the v-method is strongly damped and only reproduces the target up to 800 km horizontal distance.

As an even better result, both methods are able to recognize and separate small scale anomalies of 165 km × 20 km (Figure 5.6) and reproduce the correct sign of the perturbations. This works well for anomalies located at 200 − 300 km altitude, where

rays are reflected and therefore are most sensitive to the medium. The checkerboard perturbation tests (Figures 5.3 and 5.4) confirm this and in addition clearly indicate that anomalies at lower altitude are not well reconstructed, neither using the v-method nor with the $v\&r$-method.

Figure 5.10 compares the model misfit for the v and the $v\&r$-method for all performed tests, except for the second checkerboard test, as a function of the grid size for the v and the $v\&r$-method.

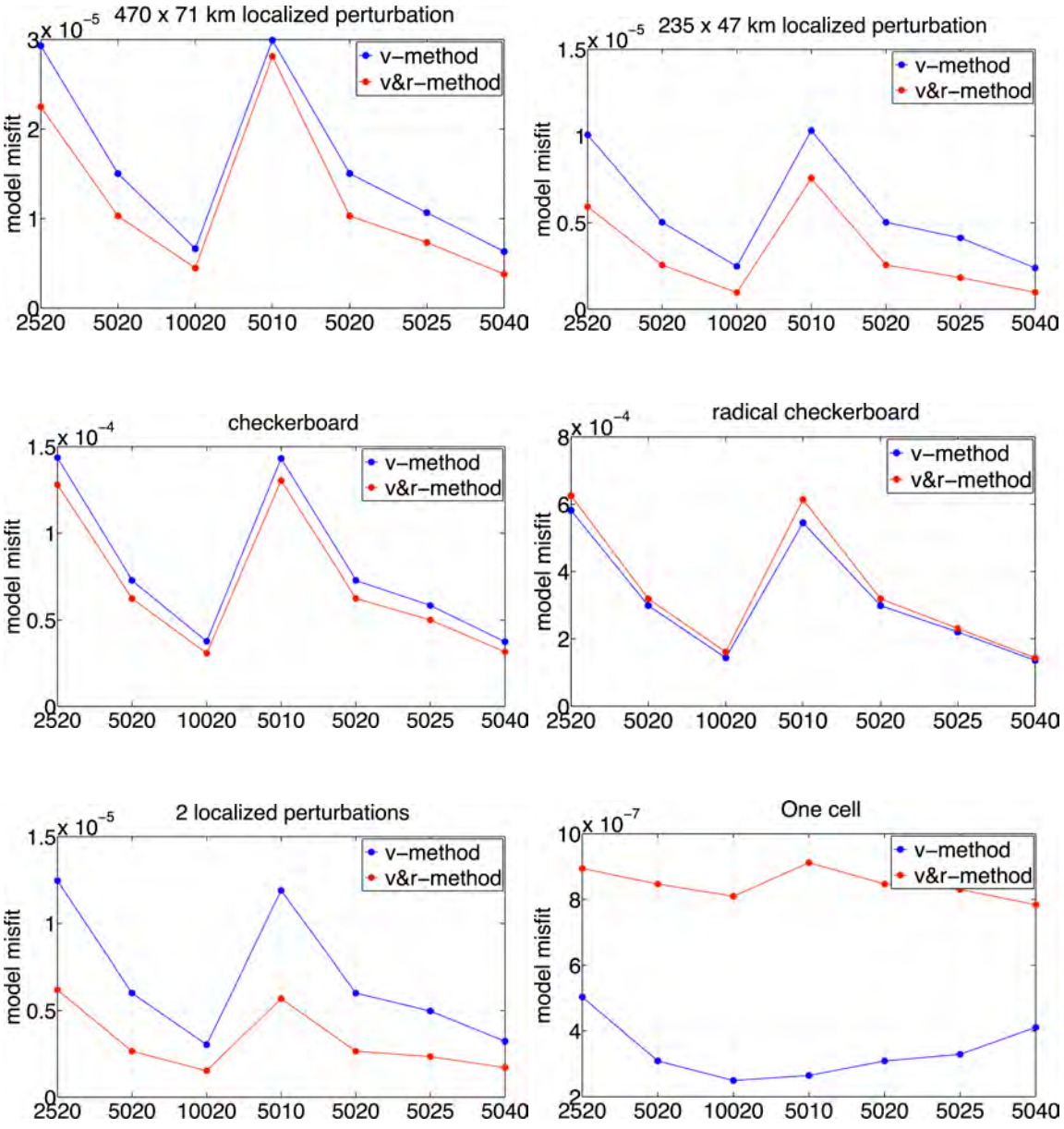


Figure 5.10: Model misfit as a function of the grid size for the v and the $v\&r$-method. The x-axis display the grid size in distance $\times$ altitude in km, respectively.

In four of the six performed tests, the model misfit is smaller for the $v\&r$-method than for the v-method. Only for the radical checkerboard test and the perturba-

tion in one cell, the v-method performs better in terms of model misfit. This was confirmed by the inversion results in Figure 5.5 and 5.7.

Chapter 6

Iterative approach

In order to improve the inversion results obtained with a single inversion, an iterative method was implemented. To this end, the solution found after each iteration can be used as a new *a priori* model for following inversions. Then, after a certain number of iterations, the synthetic target model should be obtained. In each iteration, rays are traced again in the new, updated model, the tomography matrix is recomputed accordingly for both the v and $v\&r$-method described in Chapter 4, and the differences between observed (or, in the present case, synthetic) and computed propagation times are calculated and inverted.

Based on the previous synthetic tests, the grid was fixed to 25×20 km in distance and altitude, respectively, and the background ionosphere was generated for October at noon with a solar flux of 198 SFU. The following tests were performed for a localized electron density perturbation of 0.1 %. Unless stated otherwise, data were simulated by tracing 1071 rays with elevation angles between 10° and 60° and in a frequency range of $6 - 16$ MHz. For evaluation of the difference between target and obtained solution, data and model misfits are calculated throughout iterations. The data misfit is the sum over the propagation time difference of all rays, given by

$$\text{data misfit} = \sum_{j=1}^{rays} \left(\frac{\delta T_j}{T_j} \right)^2, \tag{6.1}$$

where δT is the difference between the propagation time in the target and the propagation time T after each iteration in the updated model. The model misfit is given as the difference between target and solution summed over all cells, that is

$$\text{model misfit} = \sum_{i=1}^{cells} \left[\left(\frac{\delta Ne}{Ne^0} \right)_i^{target} - \sum_{k=1}^{iter} \frac{\frac{\delta Ne}{Ne_{k,i}} \cdot Ne_{k,i}}{Ne_i^0} \right]^2. \tag{6.2}$$

The model misfit can only be calculated for synthetic tests and not for real data.

The use of relative perturbations in this work means that the electron density perturbation δNe in each iteration is linked to the previous, updated ionospheric model Ne_k. For comparison with the target model that is with respect to the *a priori* ionospheric model Ne^0, the obtained perturbation δNe has to be renormalized by Ne^0 after each iteration.

6.1 Computational Costs

The iterative approach is associated with a very long execution times (especially the $v\&r$-method), due to the calculation of the ray-path deflection that needs recalculation of the ray-path after each perturbation of a cell. Reduction of the execution time required parallelization of the code to take advantage of multi-core architectures of modern computers. Therefore, the TDR FORTRAN code was modified using Open Multi-Processing (OpenMP) allowing to run it in parallel on a multiprocessor computer. In the following, the performances of the parallel code will be analyzed briefly.

In Figure 6.2 the time required to trace $\sim 10^3$ rays on a grid with 1000 pixels for the v-method are compared, using a Mac laptop with four cores and a Desktop computer with 32 cores. However, the Mac laptop, can execute eight threads in parallel, although it has only four cores. This feature is called hyper-threading, meaning that each core runs two threads at the same time. As expected, the

Figure 6.1: Computational costs for numerical calculation of ray-path deflection in the $v\&r$-method for $\sim 10^3$ rays on a grid with 1000 pixels. The execution times in parallel (OpenMP) and on a single processor are compared using a Mac laptop with four cores and a 32 core desktop computer.

OpenMP code is 32 times faster in parallel on the 32 cores desktop computer, but on the Mac laptop this code is only around 3 times faster. This may be associated

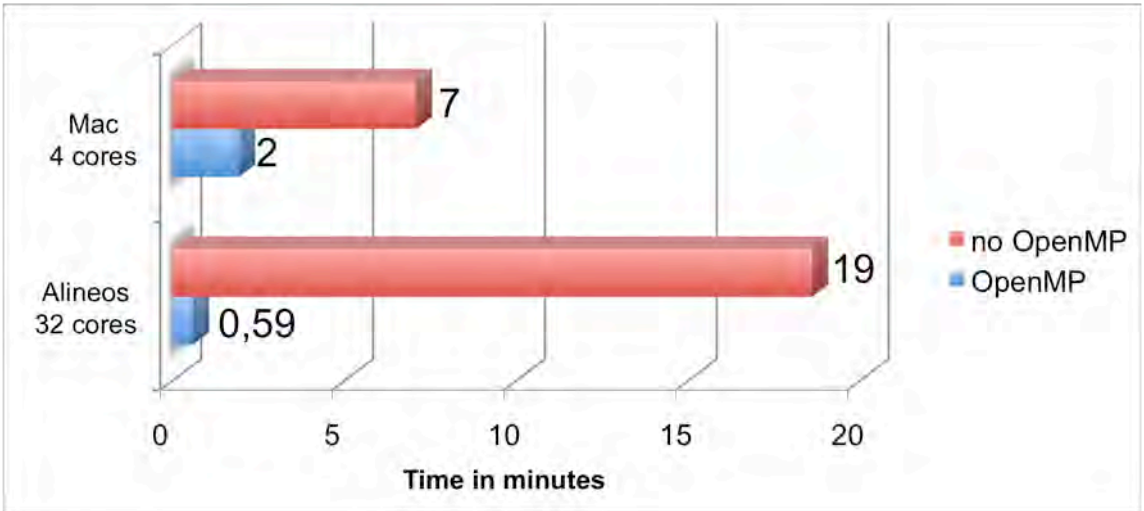


Figure 6.2: Computational costs for tracing $\sim 10^3$ rays on a grid with 1000 pixels for the v-method. The execution time in parallel (OpenMP) and on a single processor are compared using a Mac laptop with four cores and a 32 core desktop computer.

with the hyper-threading, which allows to run two threads on one processor and leads to slowing-down.

Figure 6.1 illustrates the time needed to calculate the ray-path deflection for $\sim 10^3$ rays on a grid with 1000 pixels. Certainly, the execution time is grid dependent, because the number of pixels perturbed varies with the pixel dimensions. Again, an expected gain of 32 is reached for the desktop computer, while the gain is around 3 for the Mac computer. The above performed analysis shows that the v-method performed on the 32 core desktop computer much faster in a single iteration as compared to the $v\&r$-method in parallel on a grid with 1000 pixels and $\sim 10^3$ rays, with $\sim 36s$, compared to ~ 9 minutes. **That means, for one iteration with the $v\&r$-method, 15 iterations with the v-method can be calculated in the same time**.

6.2 Preliminary results

Figure 6.3 shows some preliminary results obtained with the iterative v and $v\&r$-method applied to a localized perturbation of 0.1% *(Roy et al., 2014)*. In the interest of computational speed, no separate L-curve analysis was conducted (which would involve several inversions) at each iteration, but the regularization parameter λ was fixed to the trace of the matrix $\mathbf{A^T} \cdot \mathbf{A}$ divided by 10 (v-method) or the trace of $\mathbf{M^T} \cdot \mathbf{M}$ divided by 100 ($v\&r$-method). The effect of this regularization will be discussed in detail in Sections 6.3.2 and 6.4.2. Additionally, after each iteration the solution was not re-injected entirely, but multiplied with a feedback value of 0.5. The impact of this value is explained subsequently in Section 6.2.1.

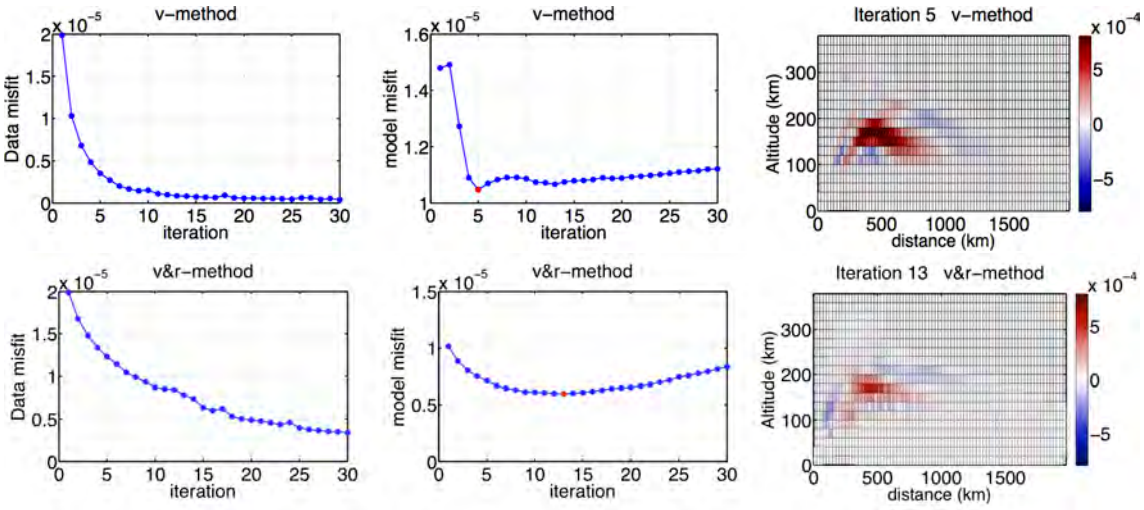


Figure 6.3: Data misfit $\sum_{j=1}^{rays} \left(\frac{\delta \mathbf{T}}{\mathbf{T}}\right)_j$ and model misfit as a function of iterations for the v-method (top) and the $v\&r$-method (bottom). The inversion results for the minimum (red point) of the model misfit (iteration 5 for v-method and iteration 13 for $v\&r$-method) are shown on the right panels.

Method	**iteration**	**optimum model misfit**	**corresponding data misfit**
v	5	$\sim 1.1 \cdot 10^{-5}$	$\sim 0.4 \cdot 10^{-5}$
v&r	13	$\sim 0.5 \cdot 10^{-5}$	$\sim 0.8 \cdot 10^{-5}$

Table 6.1: Model misfit and corresponding data misfits for results in Figure 6.3.

For both, the v and the $v\&r$-method, data and model misfit decrease with iterations, indicating that the model improves with iterations. The difference between target and solution is less than 30% for the $v\&r$-method at iteration 13 and less than 40% for the v-method at iteration 5 *(Roy et al., 2014)*. The values for data and model misfit at the model misfit minima are summarized in Table 6.3. For the v&r-method the model misfit at its minimum makes up nearly half of that obtained with the v-method, whereas the data misfit at iteration 13 for the $v\&r$-method is twice the value computed with the v-method. However, the $v\&r$-method reproduces better the shape of the target model and the smaller value of the model misfit indicates the better performance of this method compared to v-method.

After a critical number of iterations, the discrepancy between solution and target model starts to increase for both inversions, indicating the system to become very unstable with iterations. Nevertheless, the data misfit decreases further and converges in the case of the v-method. The increase of the model misfit can be rationalized by two explanations.

First, the unknown endpoints of the ray allow the algorithm to fit the data by strongly deflecting the ray or by highly perturbing the electron density to fit the

propagation time of the data. Any given observation can be explained by a combination of both negative and positive heterogeneity whose amplitude might grow indefinitely with further iterations performed. A similar problem is the choice of the regularization parameter λ. The following tests can serve to determine adequate values for such a parameter, which are used in real inversions with the same data coverage.

In the second explanation it is assumed that after a critical number of iterations, the propagation time difference vector which is inverted is in the same order of magnitude as the numerical noise introduced by discretization of the problem. Hence, the inversion just adds noise to the solution. To explore the reason for the instability of the system, the effects of regularization and feedback were checked with iterations for both methods.

6.2.1 What is the feedback?

Initial iteration tests published by *Roy et al. (2014)* showed that this inverse problem is very unstable and the solution diverges from the target model with iterations (Figure 6.3). One possible explanation is that the re-injected perturbation strongly perturbs rays upon adding a large perturbation at the reflection points of the rays, where they are most sensitive to the medium (Chapter 4).

The present inverse problem can be considered as a closed loop control system *(Åström and Murray, 2008)* and is illustrated in Figure 6.4. The general objective in a control system is to induce a output behavior in the desired way by manipulating the input. This can be achived by using a feedback. In addition, a feedback can stabilize an unstable system and also reduce the sensitivity to external disturbances or to changing parameters in the system itself. In a car, for instance, a controller sensing the current speed can maintain the current speed in the presence of disturbances such as hills and headwinds. The term feedback simply means that a certain part of the output is "fed back" as input into the process.

Here, the electron density perturbation δNe, obtained from the inversion of the propagation time difference δT between data and synthetics, is multiplied by a feedback value α before it is added to the electron density Ne_{k-1}. Ne_{k-1} is the sum of previous electron density perturbation, namely

$$Ne_{k-1} = Ne^0 + \alpha \sum_{j=1}^{k-1} \delta Ne_j, \tag{6.3}$$

with iteration k and Ne^0 as the electron density from the *a priori* ionospheric model.

The present problem is, however, a nonlinear nested closed loop system, because the electron density perturbation δNe depends on the inversion of the propagation

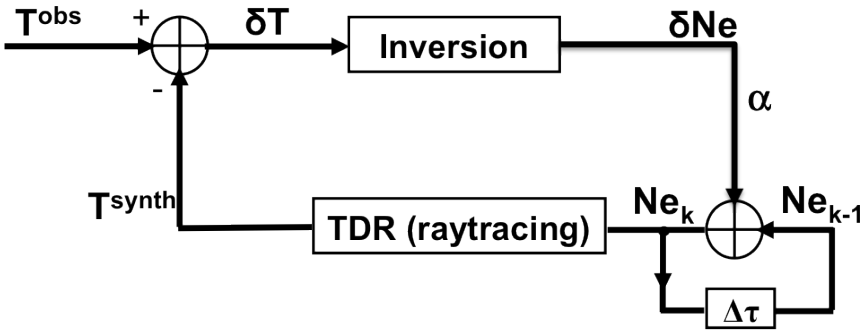


Figure 6.4: The inverse problem represented as closed loop control system. The propagation time difference δT is inverted to obtain the electron density perturbation δNe. δNe is multiplied by a feedback α and added to the electron density Ne_{k-1}. Then, rays are traced in the updated model Ne_k. The propagation time T^{synth} is calculated, which is then compared with the observed propagation time T^{obs}. The parameter α is the feedback and $\Delta\tau$ means a delay of an iteration.

time difference δT, meaning $\delta Ne = \text{inv}(\delta T)$. δT itself is a function of the raytracing in δNe, meaning $\delta T = \text{TDR}(\delta Ne)$. Here, the feedback α applies to the electron density perturbation δNe.

6.3 Iterating the $v\&r$-method

The preliminary results described previously show that the iterative $v\&r$-method improves the inversion result with iterations. Data and model misfit decrease and are smaller than in the first iteration. However, the severe ray-path deflection destabilizes the system and after a critical number of iterations the solution diverges from the target model. Therefore, to reduce the effects of severe ray-path deflection in the $v\&r$-method and to achieve convergence of the algorithm, the effect of feedback and regularization were analyzed. The corresponding results are discussed in detail in the following subsections.

6.3.1 Effect of feedback on the iterative $v\&r$-method

Different feedback values α (0.1, 0.3, 0.5, and 1) were tested for the inversion using the $v\&r$-method. 30 iterations were performed for each value, except for the feedback of 1, which is already totally dominated by noise at iteration 11 and, therefore, the solution disagrees with the target model in amplitude and shape. (Figure 6.6). After each iteration, the obtained solution was multiplied by the feedback in order to stabilize the following inversion. In this way, the update of the ionospheric model occurs slightly and not abruptly, and the rays are not perturbed as strongly as if they are traced in the model plus the entire perturbation. This approach should assure convergence.

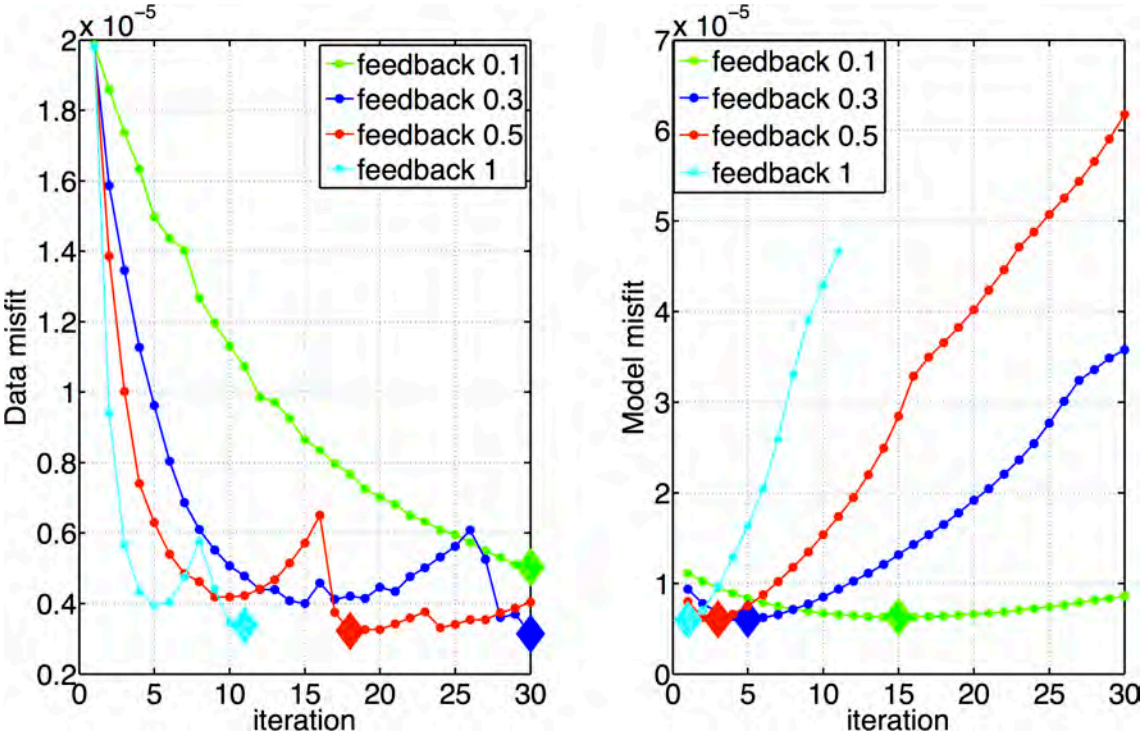


Figure 6.5: Data misfit (left) and model misfit (right) as a function of iterations using different feedback values in the $v\&r$-method. Diamonds mark minima.

Figure 6.5 shows the evolution of data and model misfits in the iterative $v\&r$-method. Both misfits decrease and increase after a critical number of iterations. The increase depends on the feedback value used and starts later for smaller feedback values. The minima of data and model misfits do not occur at the same iteration number. In the present case, the model misfit starts increasing after 15 iterations ($\alpha = 0.1$), 5 iterations ($\alpha = 0.3$), and 3 iterations ($\alpha = 0.5$). For all four feedbacks, the minima of the model misfit are nearly the same (Table 6.2 and Figure 6.5). This also applies for the data misfits at the corresponding iterations, except of the very large data misfit for a feedback of 1 at the first iteration.

Figure 6.5 illustrates that the increase of the model misfit is less rapid for the smallest feedback (α=0.1) as compared to the others, indicating some stabilization effect of the feedback on the method, because the model is updated less radically. Nevertheless, the data misfit curve for feedback 0.1 suggests that the minimum is not yet reached and further iterations are necessary to reach the same minimum level of other feedback values.

Method	iteration	optimum model misfit	corresponding data misfit
$\alpha = 0.1$	15	$0.63 \cdot 10^{-5}$	$0.86 \cdot 10^{-5}$
$\alpha = 0.3$	5	$0.607 \cdot 10^{-5}$	$0.96 \cdot 10^{-5}$
$\alpha = 0.5$	3	$0.605 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
$\alpha = 1$	1	$0.594 \cdot 10^{-5}$	$2 \cdot 10^{-5}$

Table 6.2: Mininum model misfit and corresponding data misfit for results in Figure 6.5 using different feedback values α in the iterative $v\&r$-method. The L-curve criterion was used to choose the best regularization parameter.

As mentioned above, the increasing model misfit can be explained by the trade-off between velocity heterogeneity and ray-path deflection (Section 4.6.2) or by the numerical noise introduced by the discretization of the medium. The data misfit decreases for a certain number of iterations, indicating that the data are fitted by the combination of both negative and positive heterogeneities.

Subsequently, it increases again, followed by a large step which is ascribed to the fact that rays with δT larger than in the first iteration were removed from the data set for inversion. Since ray end points are not fixed and unknown in the present problem, the ray-path can be greatly perturbed as compared to Earth seismic tomography.

Figure 6.6 illustrates the inversion results for the minima in the model (top panel) and the data misfits (bottom panel) marked by diamonds in Figure 6.5. All solutions at the model misfit minima display negative anomalies that do not exist in the target model, but the position of maximum perturbation anomaly (at around 500 km horizontal distance and 200 km altitude) is reproduced. The shape of the perturbation does not change with iterations, only the amplitude of the solutions increases. These negative perturbation anomalies can be explained by the above mentioned nonlinear effects of ray-path deflections, which (endpoints not being fixed) can in principle, result in a faster propagation time even if the velocity perturbation is negative.

In summary, the inversion results are very similar, as confirmed by nearly the same values for the data and model misfits given in Table 6.2. The inversion result after iteration 15 with feedback $\alpha = 0.1$ resembles that after the first iteration using $\alpha = 1$. Therefore, this parameter did not lead to convergence in the model misfit, but prevented the model misfit from increasing very fast. Furthermore, it slowed down the model update and did not lead to significant improvement of the final result.

The tests on the feedback described above show that, for the $v\&r$-method, a feedback value of 0.1 is too small, because additional iterations are needed to obtain

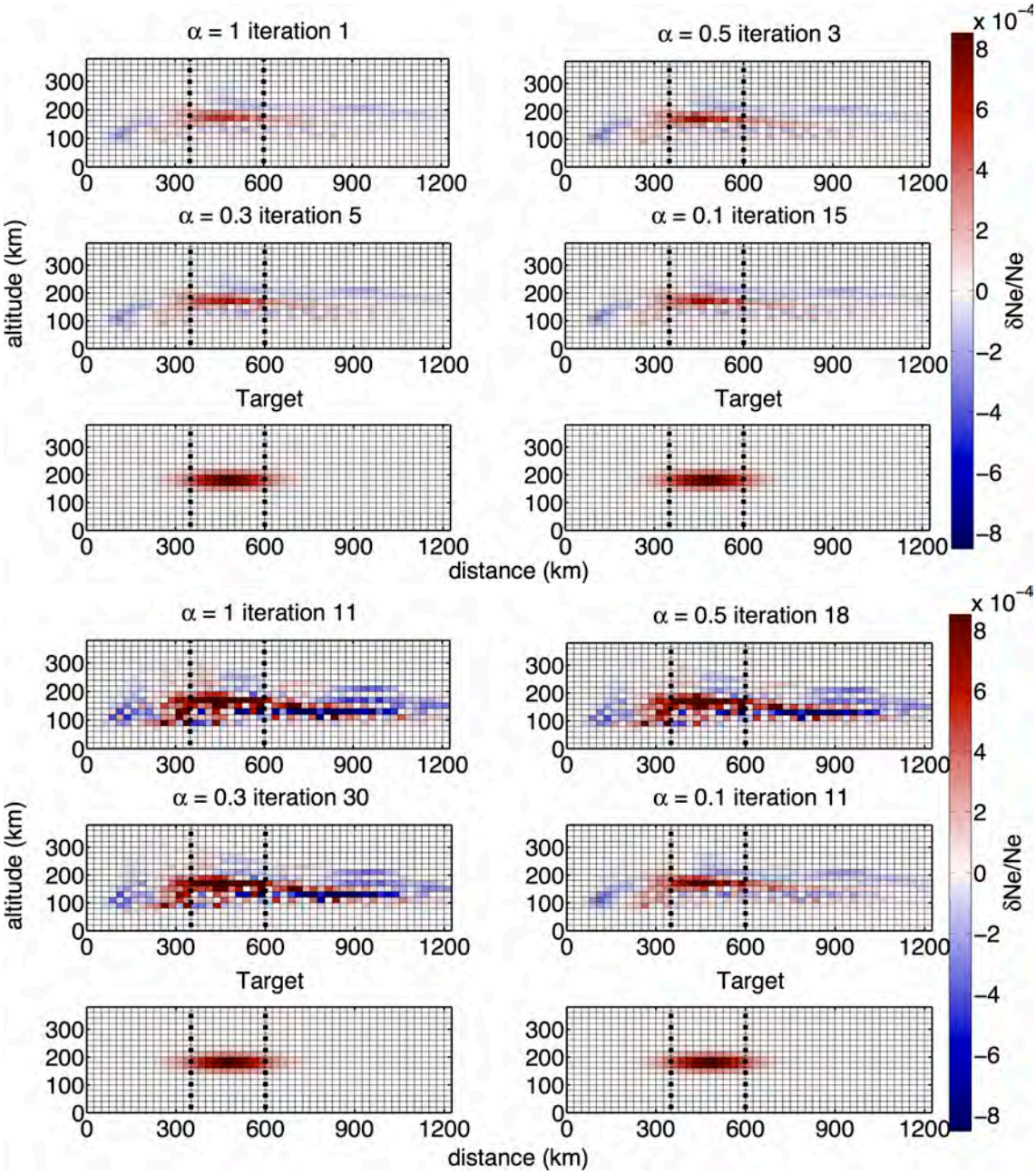


Figure 6.6: Inversion results at the model misfit minima (top panel) and the data misfit minima (bottom panel) marked in Figure 6.5 by diamonds. Different feedback values (α) were used, and the L-curve as regularization was applied in the iterative $v\&r$-method. The bottom line of each panel shows the target model.

the same inversion result with a larger value of α. On the other hand, re-injecting the entire solution after each iteration seems to destabilize the system and leads to a rapid increase of data and model misfits. In terms of iterations, the tests showed that using the L-curve, maximal 5 iterations are necessary to obtain a satisfying result.

6.3.2 Effect of the regularization on the iterative $v\&r$-method

As discussed in Section 4.5.2, the present inverse problem is an under-determined one and the matrix to be inverted ($\mathbf{A^T A}$) is sparse due to ray coverage. As a result, the determinant of $\mathbf{A^T A}$ is close to zero and the matrix cannot be inverted. However, this inverse problem can be solved by regularization.

In the iteration results presented above, the L-curve criterion was applied to select

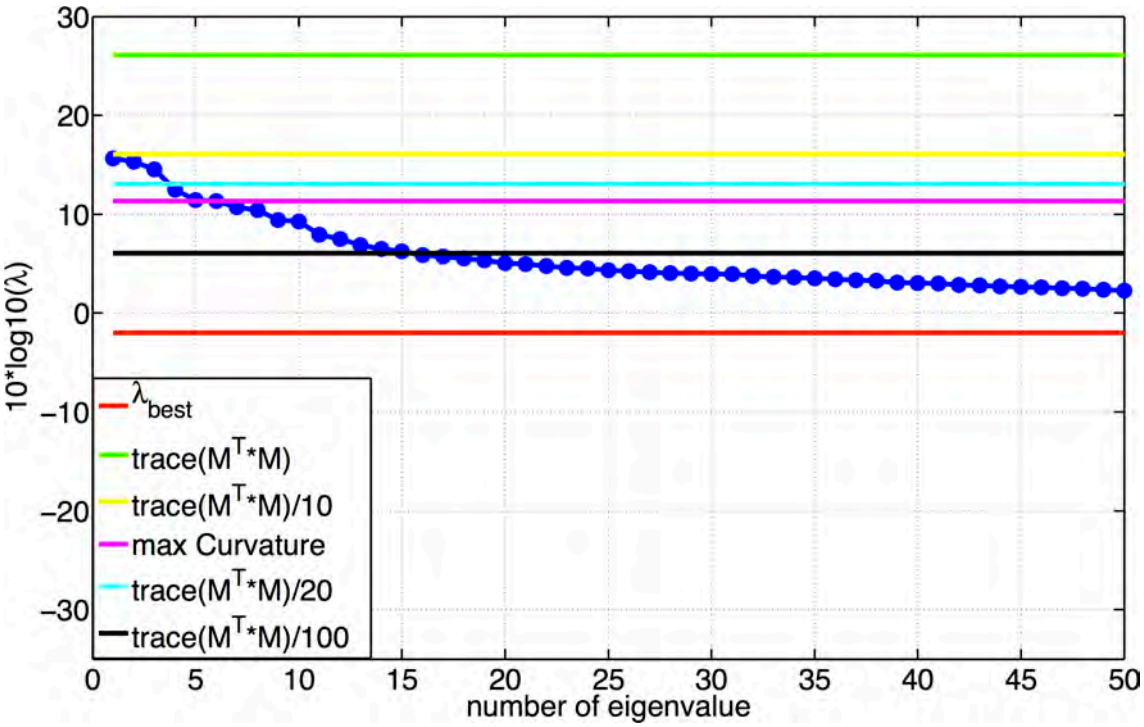


Figure 6.7: 50 largest eigenvalues λ of matrix $\mathbf{M^T} \cdot \mathbf{M}$ in descending order and different values of regularization. λ_{best} is the regularization parameter selected by the L-curve, *trace* means the trace of the matrix $\mathbf{M^T} \cdot \mathbf{M}$ and *max Curvature* the maximum curvature of the eigenvalue curve (dark blue).

the best regularization parameter. This criterion is rather arbitrary, because there is no uniform definition of the x-axis of a L-curve plot, which depends on the user's choice. In addition, different dimensions of x and y axis lead to the result that the curvature of this curve as well as its corner are not uniquely defined. Therefore, the L-curve criterion is only a rough rule to obtain a regularization parameter.

Due to the limitations of the L-curve selection criterion, another means of selecting a regularization parameter was tested. The ill-conditioning of the inverse problem manifests itself in a large dynamic of eigenvalues of the matrix. The trace of a matrix corresponds to the sum of eigenvalues and allows for a quick estimation of well determined values, which are the largest ones. Based on these results, the regularization can be adjusted *(Press et al., 1992, Chapter 18.5)*.

Figure 6.7 shows the 50 largest eigenvalues of the matrix to invert for the $v\&r$-method at the first iteration, as well as the levels of different regularization parameters relative to the eigenvalues. Obviously, the best regularization parameter selected by the L-curve criterion (red line) is too small compared to the eigenvalues. That means many values in the matrix, which correspond to numerical noise without useful information, are inverted and not suppressed.

On the other hand, the trace of the matrix $\mathbf{M^T \cdot M}$ (green line) as regularization parameter is too large, suppressing all useful information and strongly damping the solution. The value of the trace of the matrix divided by 100 (black line) correlates exactly with the corner of the eigenvalue curve that separates well estimated from poorly estimated eigenvalues. The maximum curvature of the eigenvalue curve (purple line) lies at a local maximum and is therefore of no physical relevance.

Using a feedback of 0.5, different regularizations (trace divided by 10, by 100, and the maximum curvature of the eigenvalue curve) were tested in 30 iterations. The regularization parameter changes in iterations, because it is determined by the trace of the tomographic matrix that is recalculated in each iteration. The evolution of data and model misfits is shown in Figure 6.8, and the inversion results for each method at the model and the data misfit minima are illustrated in Figure 6.9. The model misfit increases most quickly when the L-curve is used in iterations (after iteration 3), for the other methods it increases slowly after iterations 13 and 16 using the trace/100 and the maximum curvature, respectively. The larger the regularization, the later is the increase of the model misfit.

The model misfits and corresponding data misfits are summarized in Table 6.3. As for the feedback test described above, the model misfit minima are very similar. This also applies for the data misfits at the corresponding iterations, except for the iterations with the trace of the matrix divided by 10. Consequently, the inversion results shown in Figure 6.9 are very similar.

Method	**iteration**	**optimum model misfit**	**corresponding data misfit**
L-curve	3	$0.61 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
trace$(M^T \cdot M)/100$	13	$0.6 \cdot 10^{-5}$	$0.78 \cdot 10^{-5}$
max curv	16	$0.61 \cdot 10^{-5}$	$0.76 \cdot 10^{-5}$
trace$(M^T \cdot M)/10$	30	$0.78 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$

Table 6.3: Minimum model misfit and corresponding data misfit for results in Figure 6.8 using different regularization methods and a feedback of $\alpha = 0.5$ in the iterative $v\&r$-method.

The black and magenta curves for the model misfit in Figure 6.8 are similar to that in Figure 6.5 using the L-curve and a feedback of 0.1 (green curve). Fur-

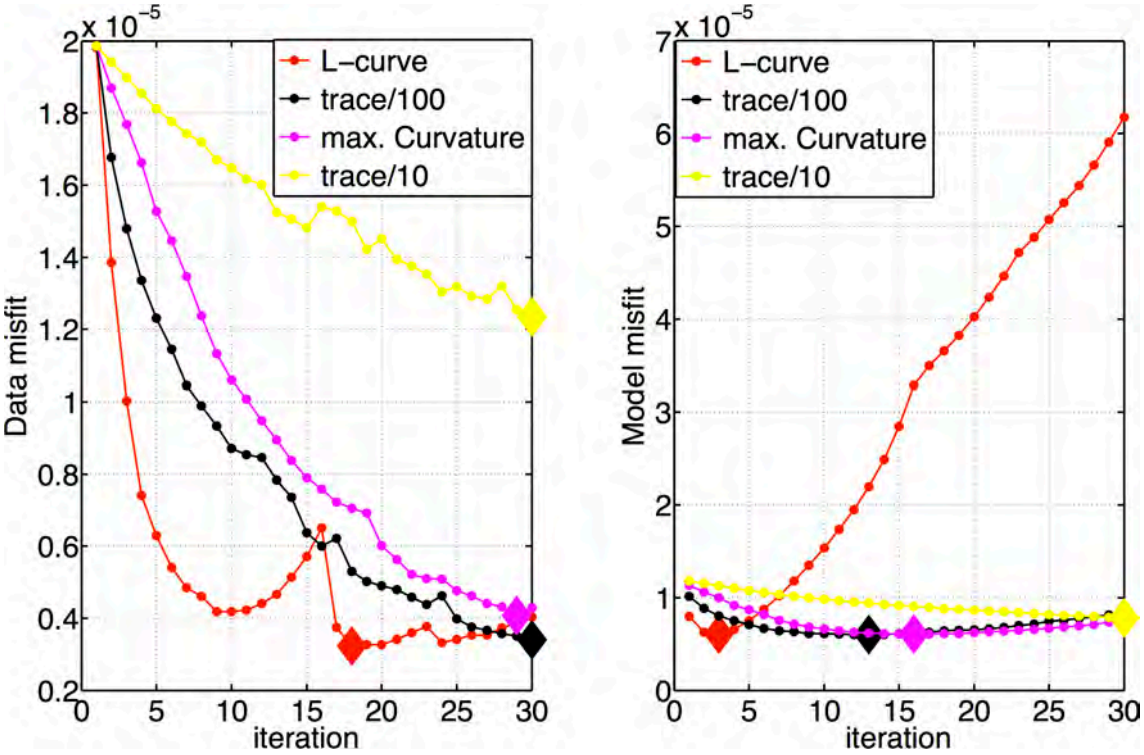


Figure 6.8: Data (left) and model misfit (right) evolution with iterations using the iterative $v\&r$-method and different approaches to identify a regularization parameter. The feedback value was $\alpha = 0.5$.

thermore, data and model misfits are similar at there minima (Table 6.2 and 6.3). This indicates that feedback and regularization are linked together. The same inversion result as for the L-curve and a feedback of 0.1 can be obtained with a larger regularization and a larger feedback after nearly the same number of iterations.

Upon iterating with the trace of the matrix divided by 10, the regularization parameter is so large (Figure 6.7) that the solution is strongly damped and the model is not significantly updated after each iteration. For this reason, the data misfit is still large and decreases at iteration 30. Probably more iterations are required to reach the same level of misfit as for the other regularization methods.

As a consequence, the inversion results shown in Figure 6.9 have a very small amplitude and do not match the target model. Depending on the chosen regularization, maximal 16 iterations are necessary to reach the minimum of the model misfit, not considering the iterations of the trace of the matrix divided by 10.

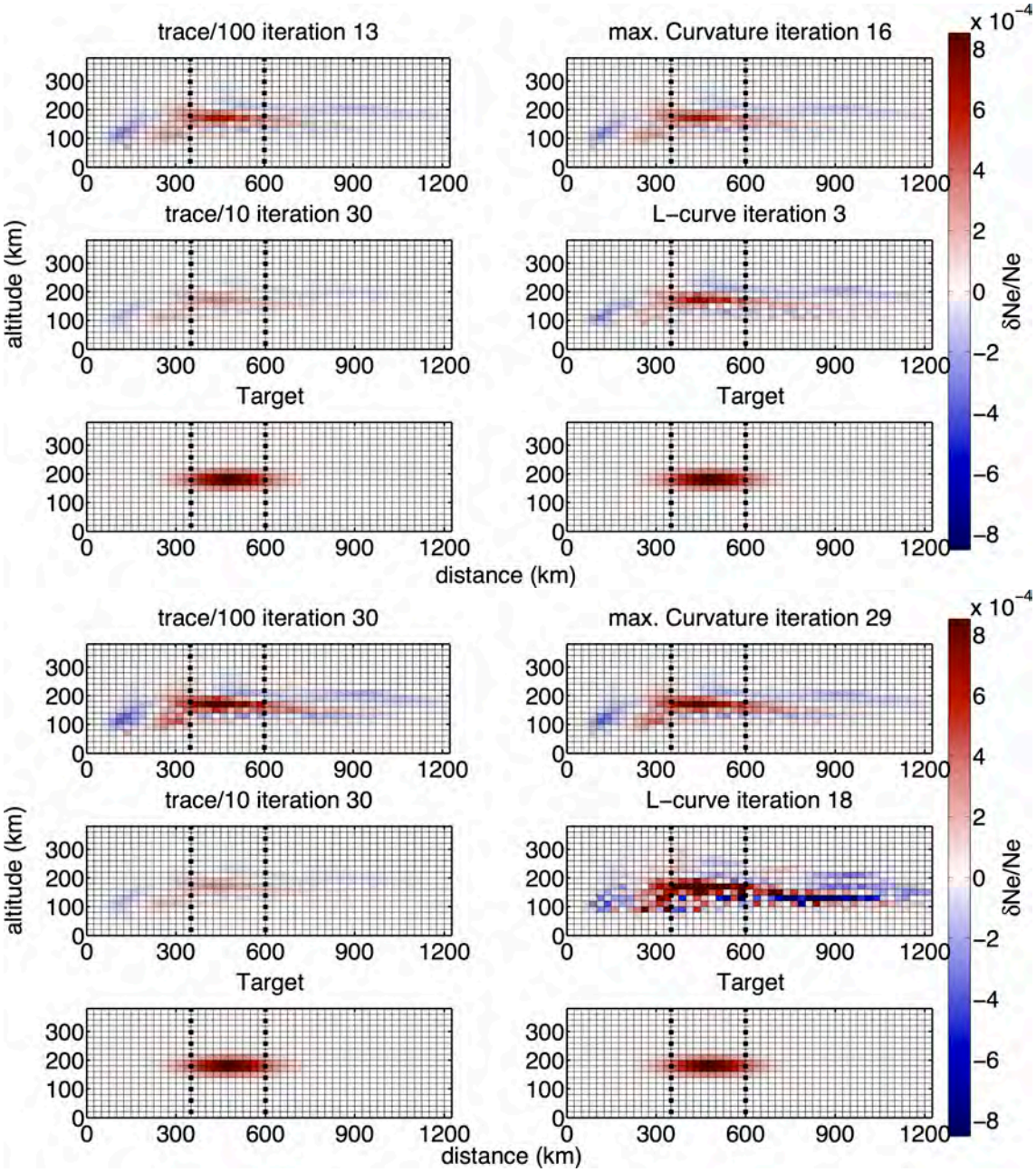


Figure 6.9: Inversion results at the model misfit minima (top panel) and the data misfit minima (bottom panel) marked in Figure 6.8 by diamonds. The feedback was set to 0.5. The bottom line of each panel shows the target model.

6.3.3 Iterative damped least squares inversion with constraints

The noise occurring in the damped least squares inversion and visible in particular at a low altitude of 100 km for all regularization methods (Figure 6.9), may strongly affect rays when passing from one cell into another. In order to avoid this effect and the resulting increasing noise in the solution, the use of a smoothed solution re-

injected after each iteration could provide better results in the iterative approach. Thus, the damped least squares inversion with constraints (4.5.3) was applied to the iterative $v\&r$-method. This method does not use a L-curve criterion to choose a regularization parameter. Instead, the factors λ and λ_1 given by equation 4.41, were initialized as described in Section 4.5.3. The other two parameters were fixed to $\alpha_0 = 2$ and $\alpha_1 = 14$ based on the tests described in Section 4.6.4. λ and λ_1 may change slightly with iterations, because the matrices ($\mathbf{A}$ and $\mathbf{M}$) evolve with iterations. Nevertheless, the best regularization parameter chosen from the L-curve ($\lambda = 0.63$) is very close to $\alpha_0 \cdot \text{trace}\left(\mathbf{M^T M}\right) / \text{trace}(\mathbf{I}) = 0.4$. Therefore, a smoothed solution is obtained with nearly the same regularization.

13 iterations were calculated applying the constraint damped least squares inversion with the feedback value set to 0.5. The evolution of data and model misfits as well as the inversion result for the minima model misfit are compared to 13 iterations of the $v\&r$-method using the L-curve and the same feedback value. As shown in Figure 6.10, the data misfits decrease in both cases, and the model misfits increase after a critical number of iterations, but the increase is much slower using the constraints. Additionally, the minimum reached by model and data misfit is smaller in the case of constraints matrices. On the contrary, the data misfit is identically for both cases identically at iterations 3 and 4 (the model misfit minima).

Comparing the inversion results obtained for iterations with and without constraint matrices (Figure 6.10), the effect of smoothing is clearly evident. The noise around the solution, in particular at low altitude (100 km) and larger distances from 600 km has completely disappeared and the solution is smooth. As compared to the iterations using the L-curve criterion, only one additional iteration is necessary to obtain a smoothed solution with a smaller model misfit.

Method	**iteration**	**optimum model misfit**	**corresponding data misfit**
without constraints	3	$0.61 \cdot 10^{-5}$	$0.74 \cdot 10^{-5}$
with constraints	4	$0.4 \cdot 10^{-5}$	$0.74 \cdot 10^{-5}$

Table 6.4: Minimum model misfit and corresponding data misfit for the results shown in Figure 6.10.

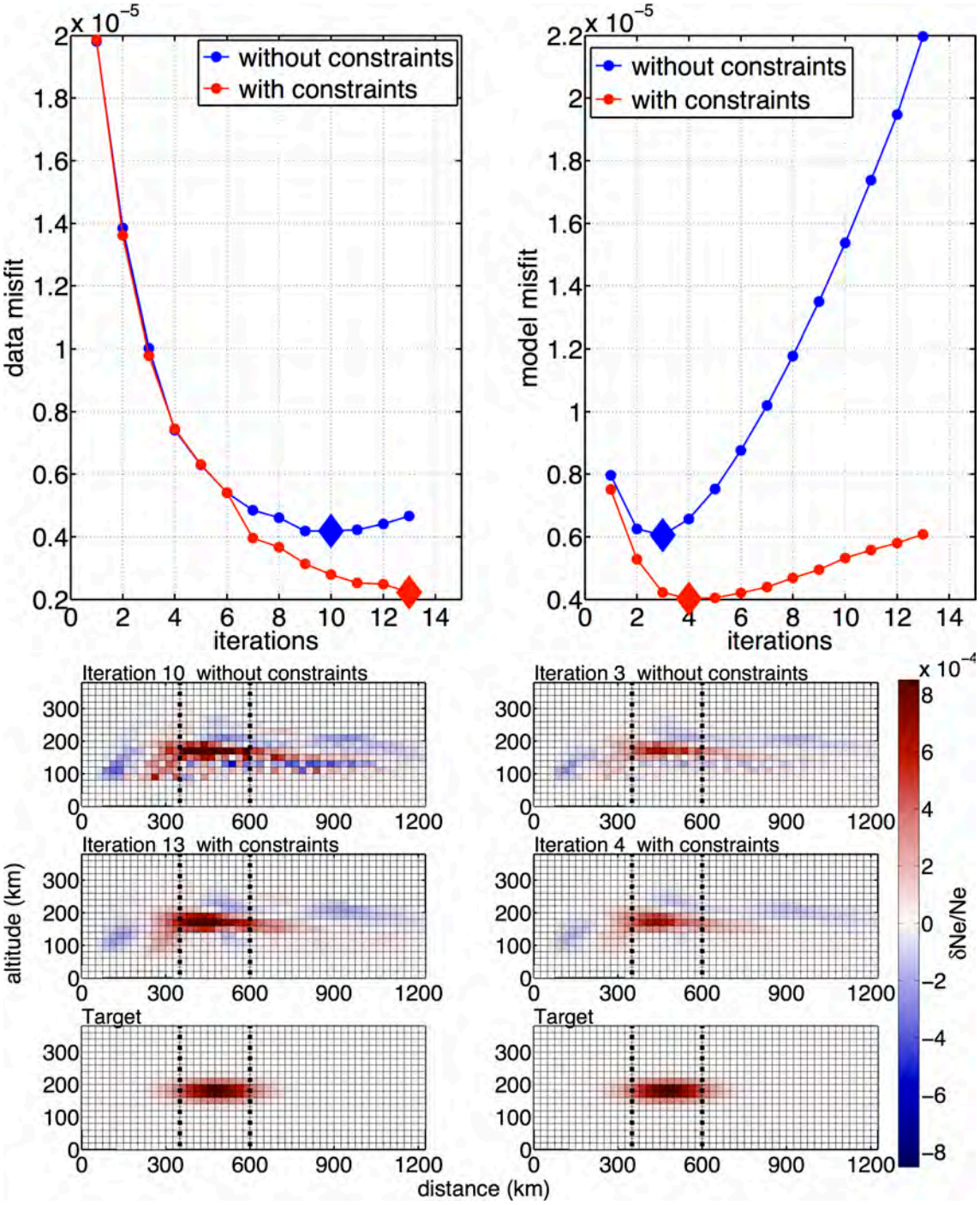


Figure 6.10: Iterative damped least squares inversion using the L-curve criterion as regularization vs. iterative damped least squares inversion with constraints. The top panel illustrates the evolution of data and model misfits as a function of iterations for both inversions and the bottom panel the inversion results at data misfit minima (left) and model misfit (left). The feedback was set to 0.5.

6.4 Iterating the v-method

In the following, the results obtained with the iterative v-method are discussed in detail. This method was applied to the same target model (localized perturbation), grid and ionospheric conditions. Again, the effects of feedback and regularization on the convergence of the solution were tested. Since solutions after the first iteration in the v-methods are already smooth, smoothing was omitted.

6.4.1 Effect of the feedback on iterative v-method

As for the $v\&r$-method, 30 iterations where calculated for the v-method using feedback values of 1, 0.5, 0.3, and 0.1. In each iteration, the L-curve was computed again to choose the best regularization parameter. The evolution of data and model misfits, as well as the inversion results at the data misfit minima are shown in Figure 6.11.

For all feedback values, both misfits decrease without a subsequent increase as seen for the $v\&r$-method. Here again, the decreasing rate depends on the feedback α: the smaller the feedback, the slower the misfits decrease. Thus, the solution evolves more quickly with a larger feedback value. This is also seen in the inversion results which display a horizontal movement of the maximum of electron density perturbation towards the correct position of the target model with further development of the solution. While the maximum of electron density perturbation is located around 350 km horizontal distance for $\alpha = 0.1$, it has moved to approximately 400 km when re-injecting the entire solution after each iteration.

Nevertheless, it will take probably many additional iterations to move the solution to the correct position using the large regularization imposed by the L-curve. Interestingly, the minima of model and data misfits for all feedback values are at the 30. iteration, except for $\alpha = 0.5$, where the model misfit minimum is at iteration 24. This may indicate that the real misfit minima have not been reached yet and further iterations are necessary.

Compared to the iterative $v\&r$-method described in Section 6.3.1, regardless of the feedback value used more iterations are necessary to reach a model misfit minimum with the v-method. A preliminary test (not shown here) illustrates that a minimum is not reached even after 60 iterations with a feedback of 0.1. Nevertheless, 30 iterations for the v-method are calculated faster (18 minutes) than 3 iterations for the $v\&r$-method (27 minutes).

The values for the model misfit minima and the corresponding data misfits are summarized in Table 6.5. Again, the model misfit minima are similar for all feedback values, just like the corresponding data misfits, except for the feedback of 0.1 giving

a larger data misfit. Figure 6.11 suggests that in this case further iterations are necessary to reach the misfit level of the other feedback values.

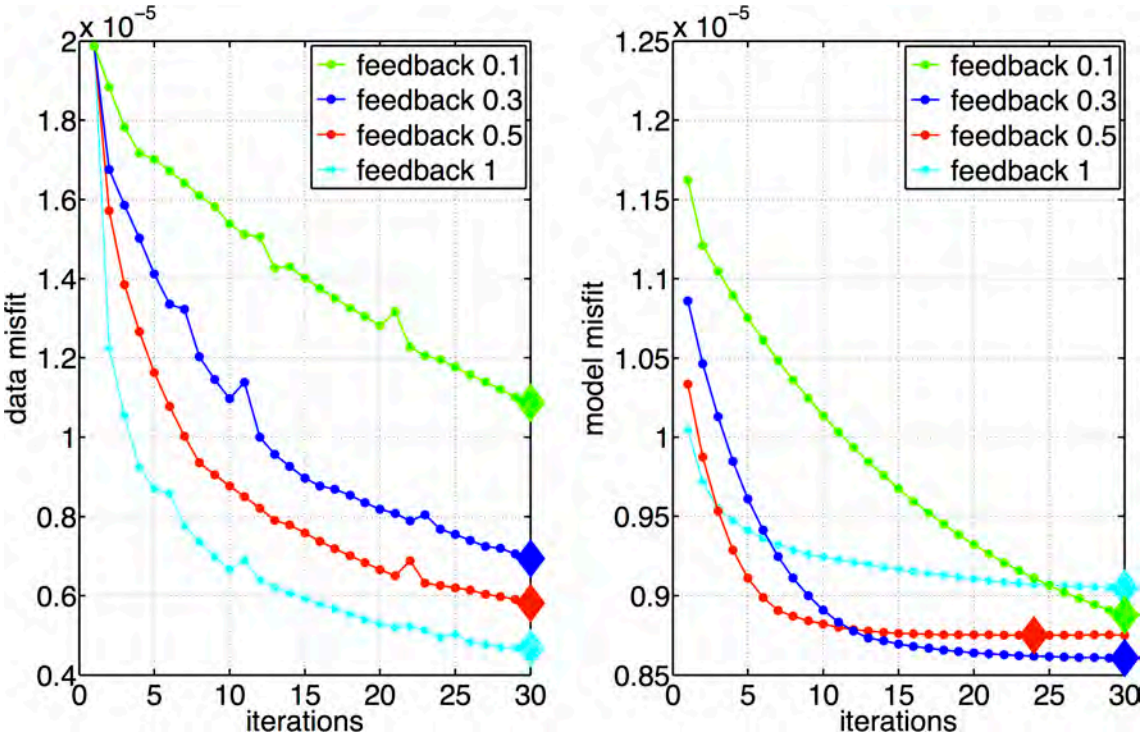


Figure 6.11: Effect of feedback value α on the inversion results of the v-method with the L-curve as regularization. The left panel shows the evolution of data misfit and the right panel the evolution of the model misfit.

Method	iteration	optimum model misfit	corresponding data misfit
$\alpha = 0.1$	30	$0.89 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$
$\alpha = 0.3$	30	$0.86 \cdot 10^{-5}$	$0.69 \cdot 10^{-5}$
$\alpha = 0.5$	24	$0.88 \cdot 10^{-5}$	$0.63 \cdot 10^{-5}$
$\alpha = 1$	30	$0.9 \cdot 10^{-5}$	$0.47 \cdot 10^{-5}$

Table 6.5: Model misfit minima and corresponding data misfits for results shown in Figure 6.11, depending on different feedback values α in the iterative v-method.

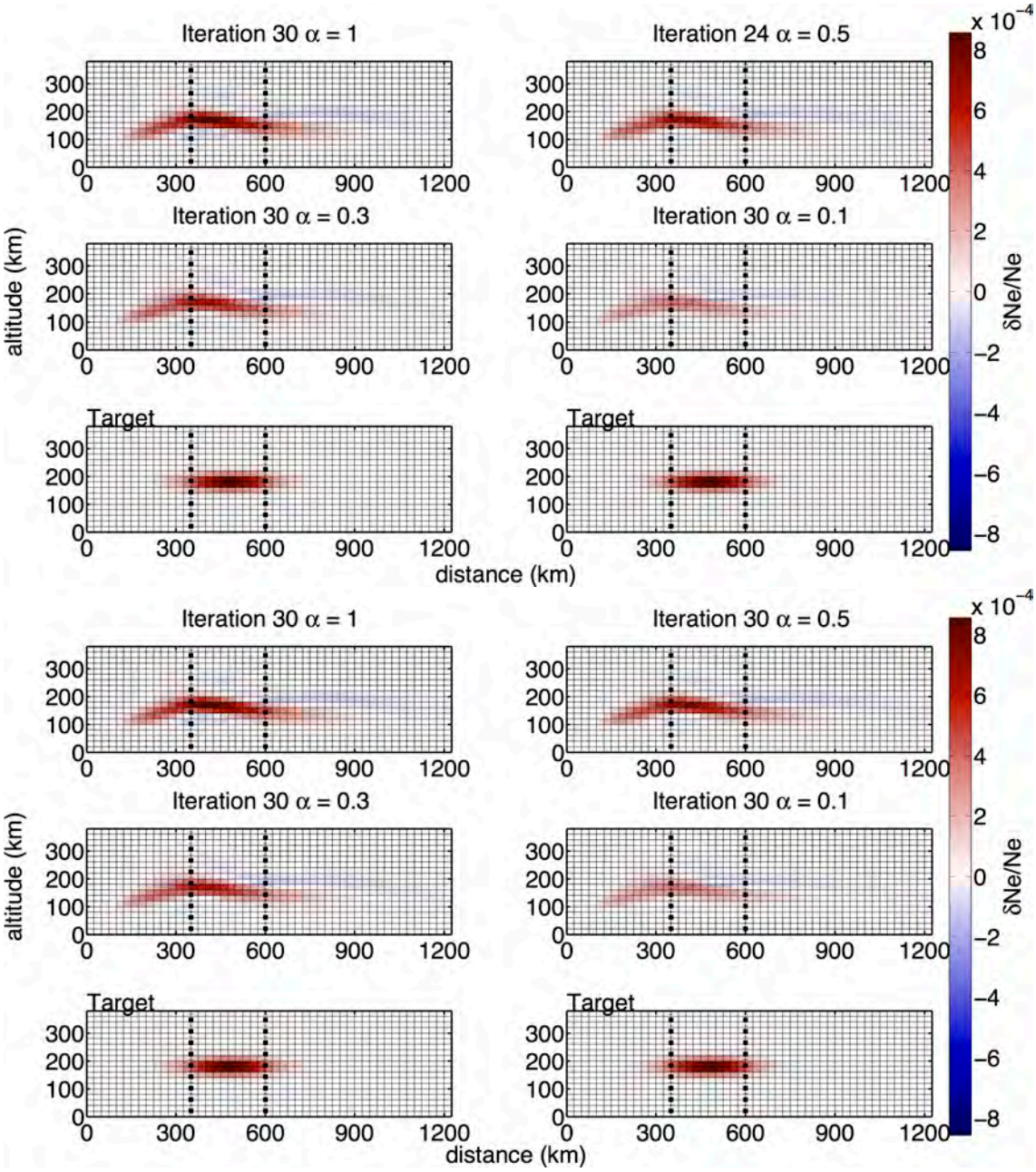


Figure 6.12: Inversion results at the model misfit minima (top panel) and the data misfit minima (bottom panel) marked in Figure 6.5 by diamonds. Different feedback values (α) were used and the L-curve as regularization was applied in the iterative v-method. The bottom line of each panel shows the target model.

6.4.2 Effect of regularization on the iterative v-method

The tests described above showed lead to the result that the L-curve is not only unsuitable to choose a regularization parameter for the iterative v-method, but it strongly damps the solution. Moreover, the model misfit minimum is not reached after 30 iterations, regardless of the feedback value used. Therefore, other regularizations were tested as described for the iterative $v\&r$-method in Section 6.3.1.

Figure 6.13 shows this regularization parameters in the first iteration compared to the 50 largest eigenvalues of the matrix $\mathbf{A^T} \cdot \mathbf{A}$. Clearly, the values selected by

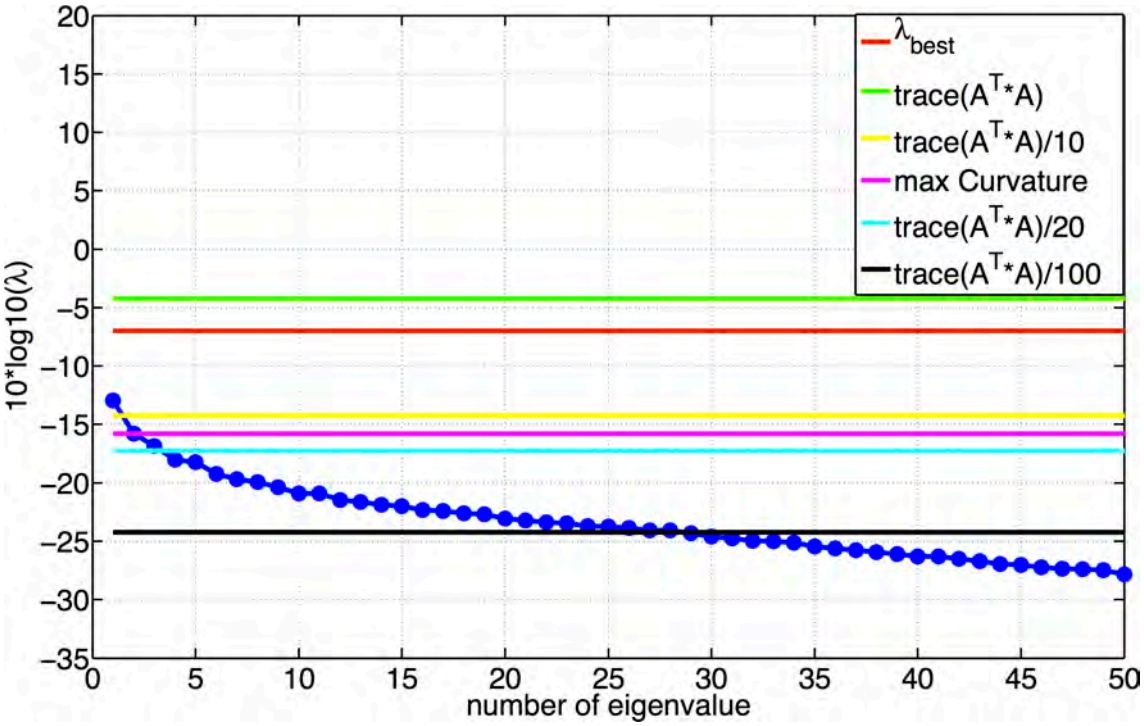


Figure 6.13: 50 largest eigenvalues λ of matrix $\mathbf{A^T} \cdot \mathbf{A}$ in descending order and different values of regularization. λ_{best} is the regularization parameter selected by the L-curve, *trace* means the trace of the matrix $\mathbf{A^T} \cdot \mathbf{A}$ and *max Curvature* the maximum curvature of the eigenvalue curve.

the L-curve (red line) and the trace of the matrix (green line) are too large for regularization. They are larger than the largest eigenvalue of the matrix, so they strongly damped the solution and no useful information is added to it, just white noise. Therefore, in the case of the L-curve criterion, only very few information is added, and the feedback does not destabilize the system in iterations.

On the other hand, the trace of the matrix divided by 100 (black line) is very small and may introduce much noise into the inversion, but it increases the freedom of the solution with regard to change. Here, a careful choice of the feedback value α is essential. The smaller the regularization parameter, the smaller α has to be chosen in order to change the solution slightly and to avoid destabilization of the system

with iterations. Test with iterations for different feedback values for the trace of the matrix divided by 10, 20, and 100 as regularization parameter proved: The system becomes unstable for the smallest regularization parameter (trace divided by 100) already for α larger than 0.1, while for the trace of the matrix divided by 10 the iterations are stable for a feedback up to 0.5.

Figure 6.14 compares the evolution of data and model misfits for four regularization methods (L-curve, trace divided by 10, 20, and 100). In all methods, a feedback value of 0.1 was used. Since the L-curve criterion chooses the largest regularization parameter, the model misfit decreases very slowly with iterations and has not yet reached a minimum at iteration 30, while for the other regularization methods a minimum is reached based on a quick evolution of the solution.

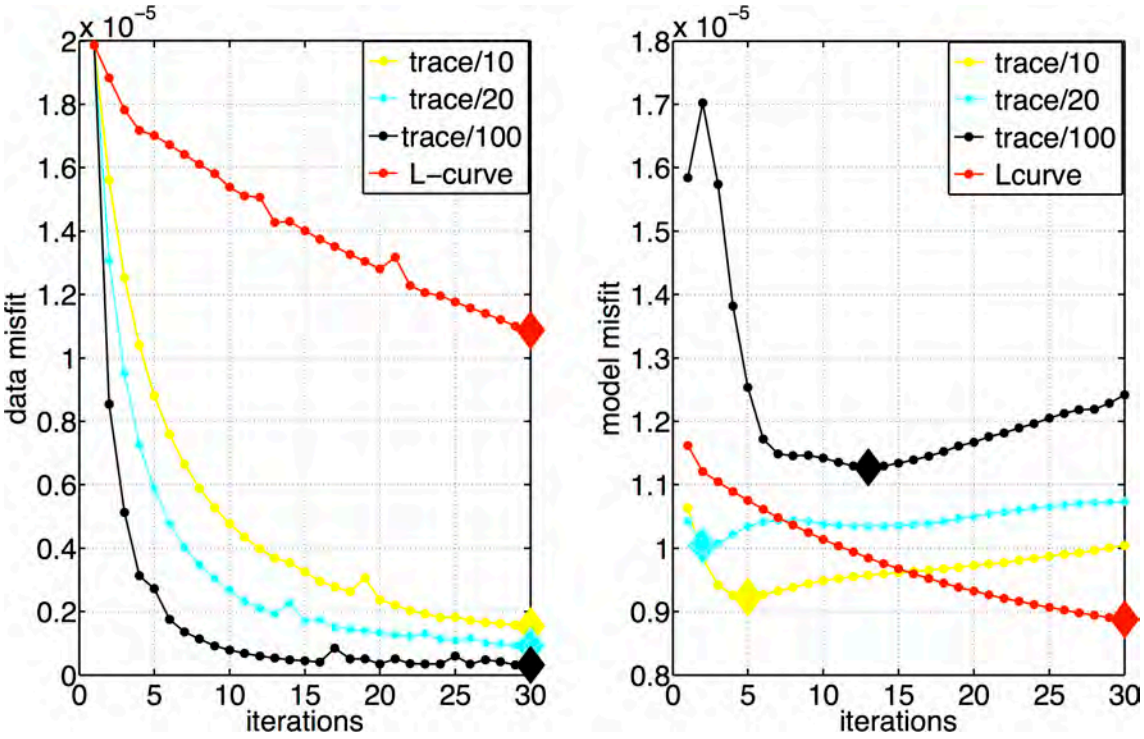


Figure 6.14: Data and model misfits evolution with iterations for four regularization parameters applied to the v-method and a feedback value of 0.1.

Model misfit minima and the corresponding data misfits are summarized in Table 6.6. Model misfits for the regularizations without L-curve are similar, but the corresponding data misfits are quite different, especially the very small data misfit for the trace divided by 100 is remarkable.

Figure 6.15 illustrates the inversion results at data and model misfit minima. The results obtained after 30 iterations with the L-curve are less noisy as compared to the other inversion results after 30 iterations as well. As mentioned before, this is due to the large regularization, which exceeds the largest eigenvalue of the matrix and suppresses all useful information in the matrix to invert and just adds white

Method	iteration	optimum model misfit	corresponding data misfit
L-curve	30	$0.89 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$
trace$(M^T \cdot M)/100$	13	$1.1 \cdot 10^{-5}$	$0.05 \cdot 10^{-5}$
trace$(M^T \cdot M)/20$	2	$1 \cdot 10^{-5}$	$1.3 \cdot 10^{-5}$
trace$(M^T \cdot M)/10$	5	$0.92 \cdot 10^{-5}$	$0.88 \cdot 10^{-5}$

Table 6.6: Data and model misfit for results in Figure 6.15 depending on different regularization methods and a feedback of $\alpha = 0.1$ in the iterative v-method.

noise. Moreover, the solution does not correspond with the right position of the target model. By contrast, the solutions for the three other methods are at the correct positions, and, a low regularization corresponds with a higher noise level in the solution. This is clearly confirmed by the model misfit curves in Figure 6.14.

Again, the solution moves slightly horizontally with iterations, when the inversion results at model misfit minima are compared. (Figure 6.15, bottom panel). While the solution for the L-curve regularization is still mislocated at iteration 30, with smaller regularization (trace divided by 100) it is already after iteration 13. Subsequently, the algorithm only adds noise to the solution,inducing the model misfit to increase again (Figure 6.14). The trace of the matrix divided by 10 imposes a larger regularization to the solution, so the maximum of electron density perturbation has only slightly moved horizontally at iteration 5, however, the difference to iteration 30 with the L-curve is clearly evident.

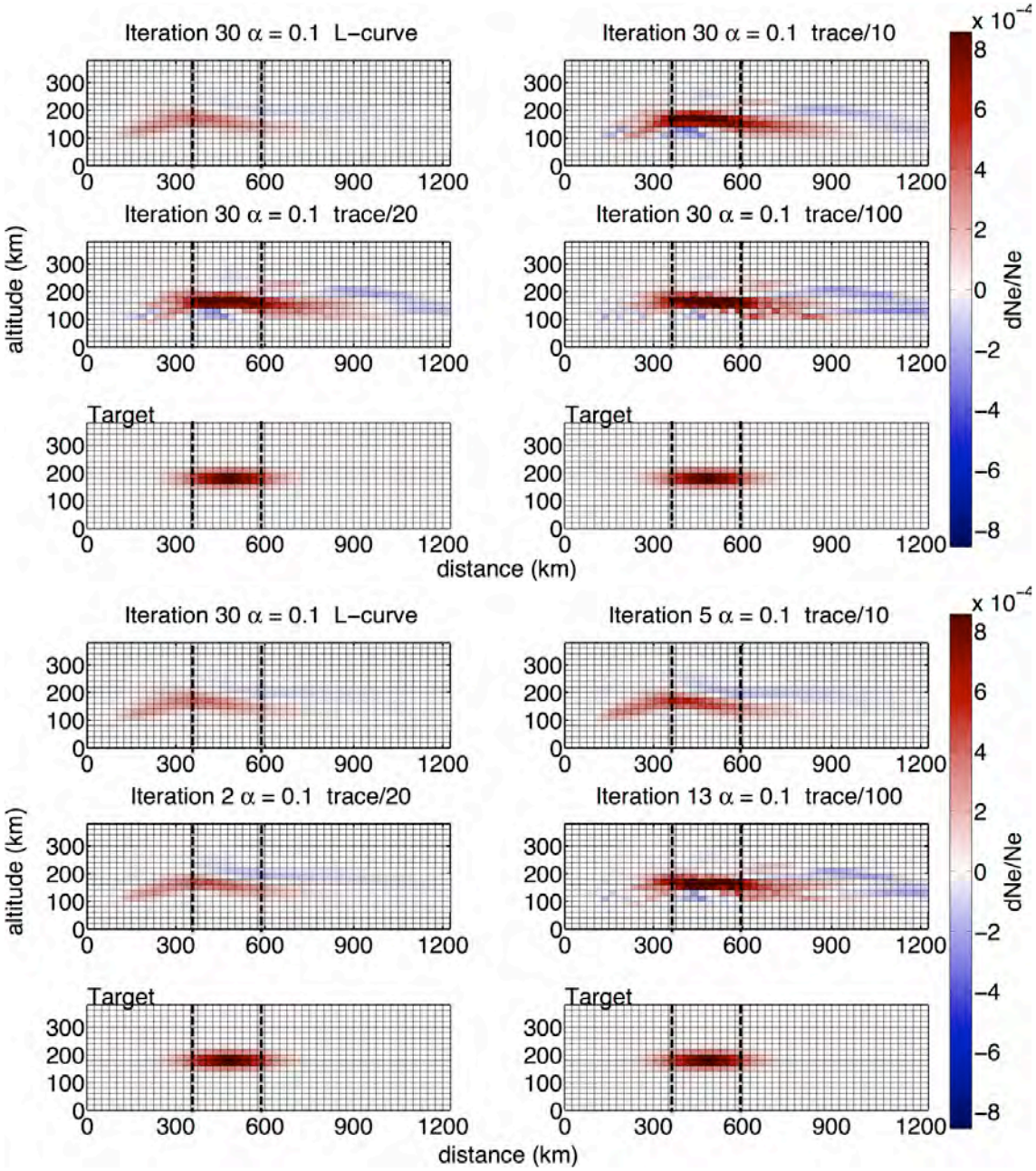


Figure 6.15: Effect of regularization on the iterative v-method. Inversion results with the v-method at model misfit (bottom panel) and data misfit minima (top panel) in Figure 6.14. For all four methods, a feedback value of 0.1 was applied.

6.5 Summary and Conclusions

In this chapter both, the v-method and the $v\&r$-method were iterated in order to improve the inversion results. Preliminarily results published by *Roy et al. (2014)* demonstrate that this is achieved by decreasing model and data misfits. At their minima, they are smaller as compared to the first iteration. Additionally, they illustrate the better performance of the iterative $v\&r$-method in terms of model

misfit (30% for $v\&r$, 40% for v-method). However, while the data misfits decrease for both methods, the model misfit increase after a critical number of iterations. This indicates that the system becomes unstable and the solution diverges from the target model. This instability can be explained by the numerical noise, introduced by discretization of the model by blocs or by non-linearity of the ray-path deflection.

To address with this problem and reduce the instability in the iterative approach, the effects of feedback and regularization were tested for both methods. The first parameter controls the amplitude, the second one damps the re-injected solution. For both methods, feedback values of 0.1, 0.3, 0.5, and 1, as well as another regularization based on the eigenvalues of the matrices in addition to the L-curve criterion were tested.

In the iterative $v\&r$-method with the L-curve as regularization (Section 6.3.1), the solutions have similar model misfit minima and corresponding data misfits, but depending on the feedback value used, more iterations are needed to obtain similar inversion results. This means, the feedback slows down the convergence, but the test also showed a more slowly increase in the model misfit. By contrast, in the v-method a minimum in the model misfit is not yet reached after 30 iterations, therefore probably more iterations are necessary. In terms of execution time this means, that the $v\&r$-method with the L-curve as regularization, will faster reach a model misfit minimum, regardless which feedback is used.

In a second test, the regularization method was tested in the iterative approach. The analysis of the eigenvalues of the matrices to invert compared to the regularization revealed for the $v\&r$-method that the L-curve returns to a value being too small and introducing noise into the inversion, whereas for the v-method the value by the L-curve chosen is to large, suppressing all useful information in the matrix. Regardless of the regularization method used in the v- and the $v\&r$-method, model and data misfit minima are similar and reached later, the larger the regularization. While the maximum of electron density perturbation is already at the right position in the first iteration of the $v\&r$-method, it is mislocated in the v-method, but moves horizontally with further iterations, improving the result.

In conclusion, the $v\&r$-method is computationally more expensive and numerically more complex, but all in all more rapid, because it reaches model misfit minima generally faster than the v-method. This is probably linked to the fact that the $v\&r$-method takes into account the ray-path deflection starting from the first iteration, whereas the v-method considers ray-path deflection from the second iteration on, because rays are traced in an updated, new ionospheric model. Contrary to the v-method, the $v\&r$-method needs a smoothing in order to reduce the noise in the solution at low altitude.

Chapter 7

Real data inversion

This chapter deals with the application of the developed tomography method to real data measured by the OTH radar Nostradamus. Thus synthetic data were replaced by real data. Section 7.1 describe the preprocessing used to prepare the data for the inversion. Subsequently, the real data were inverted and vertical profiles of electron density were calculated (Section 7.2). At the end of this chapter, some perspectives for further applications and improvements are discussed.

7.1 Data preprocessing

The operation mode of the OTH radar Nostradamus is explained in Section 3.4. The radar transmits a chirp signal (linear frequency modulation) with a transmission time of 3 ms and a recurrence of 30 ms. After digitalization of the received signal, cross-correlation with the emitted signal, Doppler processing and digital beam forming for different elevation angles, a curve of the elevation angle as a function of group path or group propagation time is obtained (*e.g.,* Figure 7.1). A complete measurement with Nostradamus (all frequencies and elevation angles) takes around 8 Minutes and it is assumed that the ionosphere is stable and constant during this time of measurement.

Figure 7.1 shows real data obtained on March 14, 2006 at 1855 UT measured by Nostradamus. Obviously, the data contain much scatter, and reflections from different ionospheric layers are visible. Clearly these data need some preprocessing before being inverted. Orange and red data points at elevation angles larger than 50° are probably multi-path data, meaning that the signal is reflected twice or more often at the ground. Generally, only rays with a single reflection are considered in the ray tracing methods, (*e.g.* Figure 7.6).

Preprocessing is necessary to remove multi-path data, side lobes and noise peaks automatically from the dataset. This can be accomplished using the virtual height

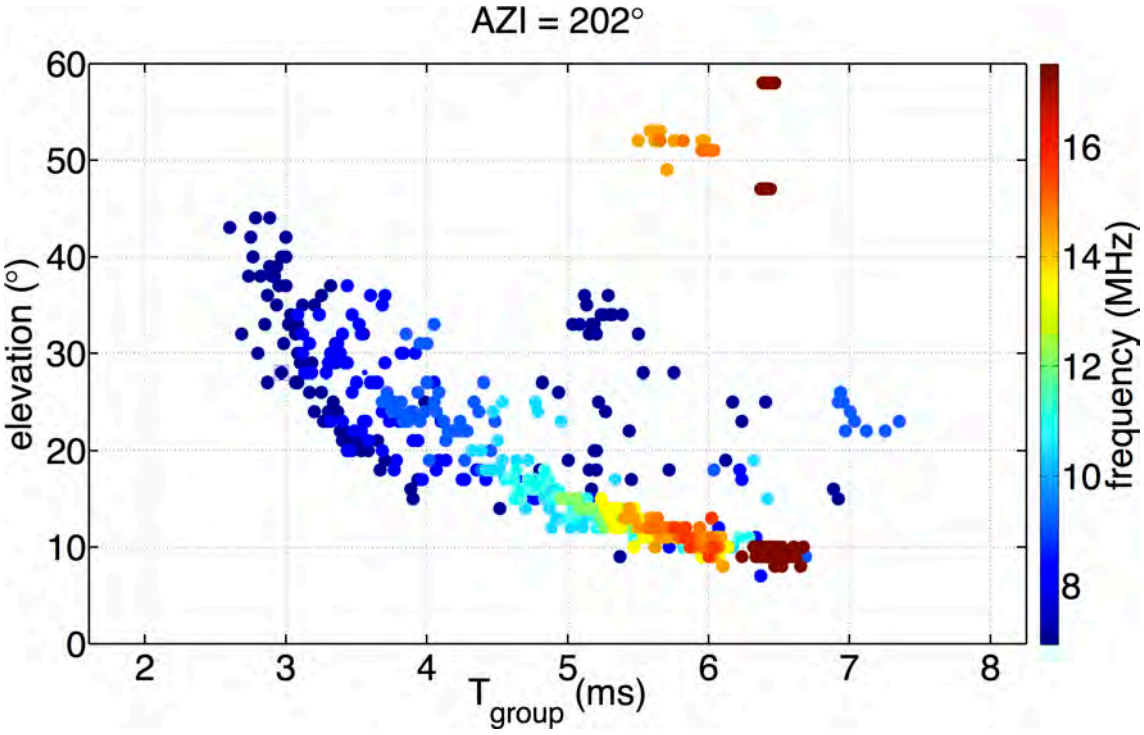


Figure 7.1: Real data (group propagation time as a function of elevation angle) obtained from the OTH radar Nostradamus on March 14, 2006 at 1855 UT in azimuth 202°. The frequency is color coded.

of reflection h_v, which can be easily calculated from the ground distance d_g and the elevation angle ϕ, using the theorem of Martyn *(Martyn, 1935)*

$$h_v = \sqrt{\left(\frac{d_g}{2}\right)^2 + R_0^2 + 2R_0 \cdot \frac{d_g}{2} \cdot \sin\phi} - R_0, \tag{7.1}$$

with R_0 being the Earth radius. This theorem is a simple approximation for the propagation of radio waves in the ionosphere, allowing pre-filtering of points corresponding to the propagation in the same ionospheric layer (Figure 7.2). Multi-path rays normally have virtual heights of reflections being twice the virtual height of single-path rays and are often larger than 500 km. Consequently, the virtual height can be used to remove these data from the dataset. Even in the case of wrong virtual heights (for they are only an approximation), reflections corresponding to the same ionospheric layer must have very similar virtual heights.

Figure 7.3 shows the relationship between virtual height of reflection, elevation angle, and group distance for rays with a frequency of 6 MHz and elevation angles between 10° and 20°, traced in the *a priori* ionospheric model NeQuick generated for October at 12 UT. This period was chosen because it allows to observe the reflection in different ionospheric layers, whereas for the data shown in Figure 7.1, reflections are only observable in the F-layer, but not in the E-layer. One distinguishes between

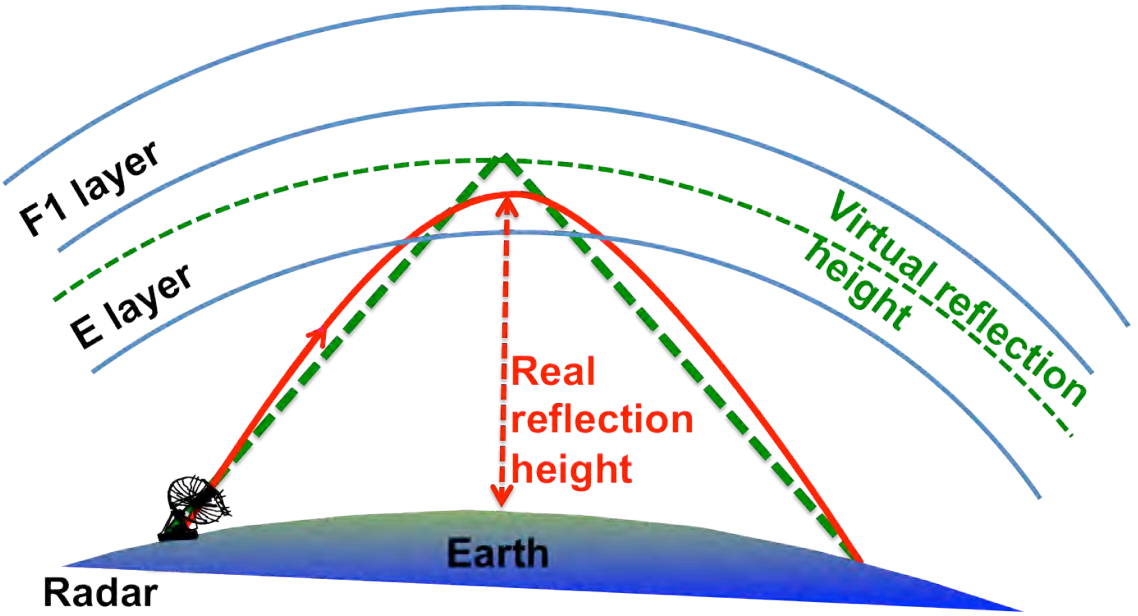


Figure 7.2: Explanation of Martyn's theorem. The path of the signal (red) is approximated by the green colored path. The virtual height is always larger than the true height of reflection.

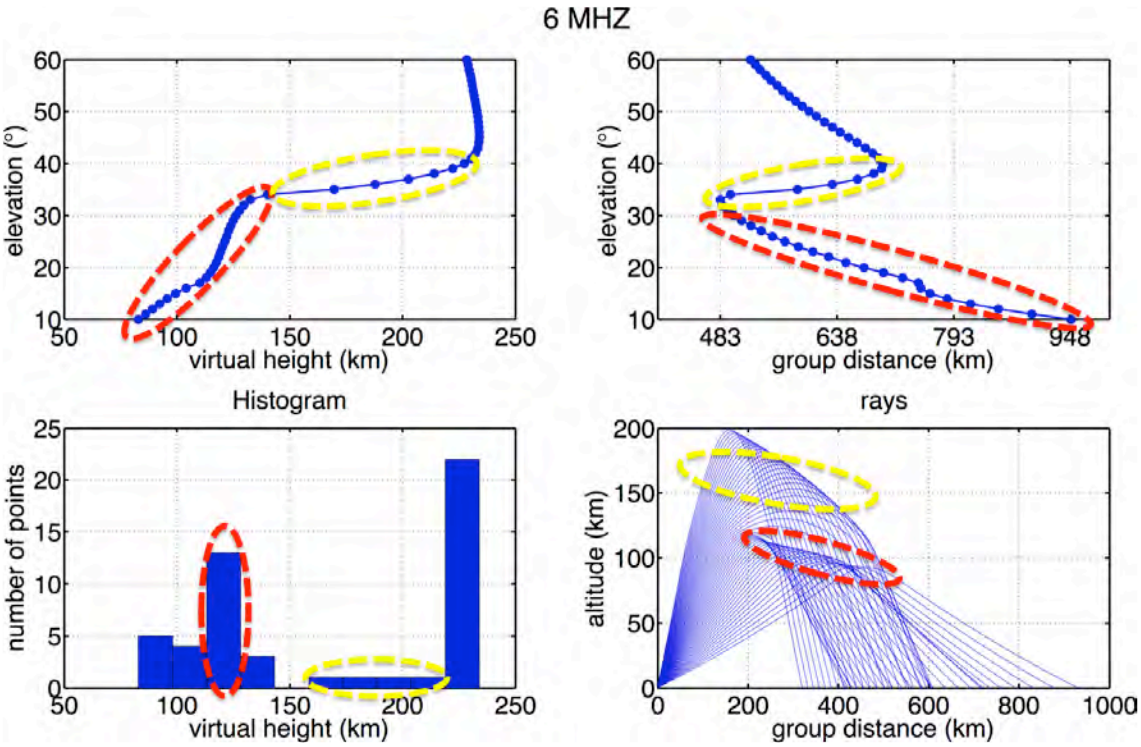


Figure 7.3: Relationship between virtual height of reflection, group distance and elevation angle for rays with a frequency of 6 MHz and elevation angles between 10°-20°, traced in the *a priori* ionospheric model NeQuick generated for October at 12 UT. Yellow: high-angle ray. Red: low-angle ray.

high- and low-angle rays. For rays contributing to the low-angle ray (red in Figure 7.3), the virtual height of reflection varies only slightly, but for the high-angle ray (yellow), a small variation in the elevation angle of the ray leads to a significant increase of the group distance and, consequently, of the virtual height of reflection (Equation 7.1).

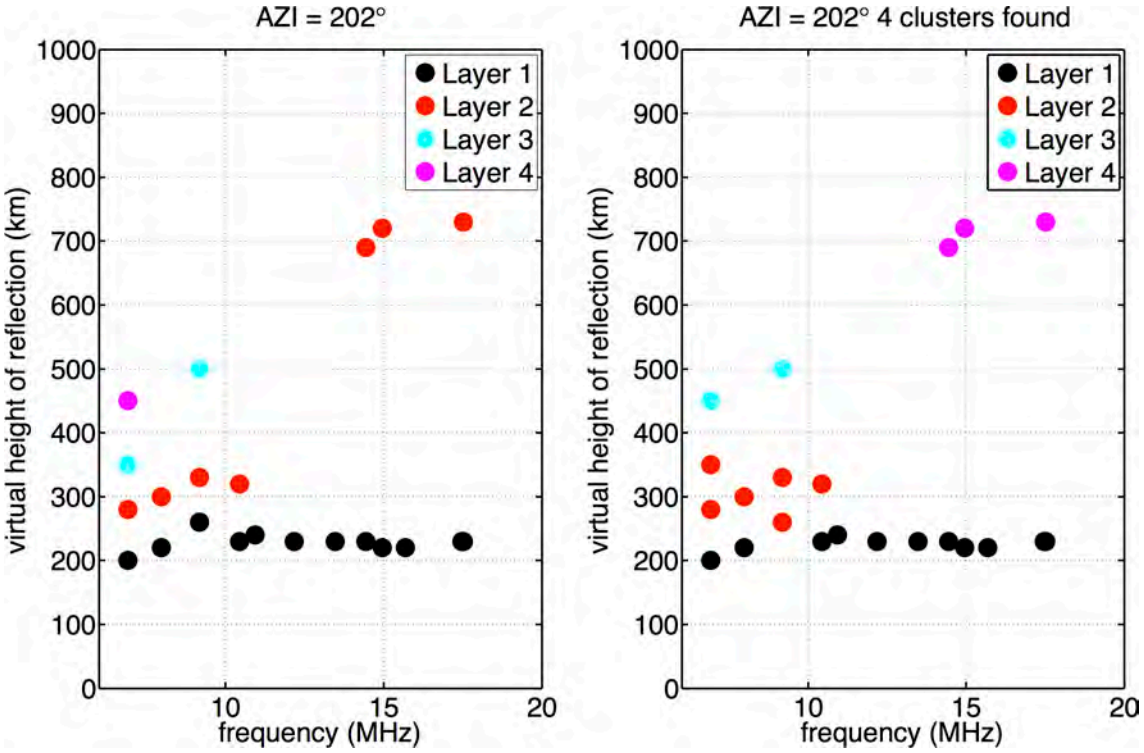


Figure 7.4: Virtual height of reflection as a function of signal frequency. The color denotes the number of ionospheric layers given in the preprocessing of the radar data to each data point (left). Data points corresponding to the same ionospheric layer are arranged in clusters (right).

The virtual height of reflection is calculated already during preprocessing of the radar data. It is included in the data set for each azimuth and one frequency, and the data points are assigned to ionospheric layers based on histogram analysis. Ionospheric layers are numbered from 1 to 5. These numbers do not correspond to the real numbers of ionospheric layers for the day of measurement (5 ionospheric layers do not exist) and are not comparable between frequencies (Figure 7.4, left) because they vary with frequency.

In a first step of real data analysis, a k-means cluster analysis *(Forgy, 1965)* was applied to the virtual height of reflection allowing an automatic assignment of points with the same virtual height of reflection to the same ionospheric layer. The aim of this method is to cluster the data in k portions, with the sum of the squared difference between the cluster barycenters being minimal. For the present data example, Figure 7.4 shows the virtual height of reflection as function of fre-

quency before and after clustering. Data points with the same virtual height are

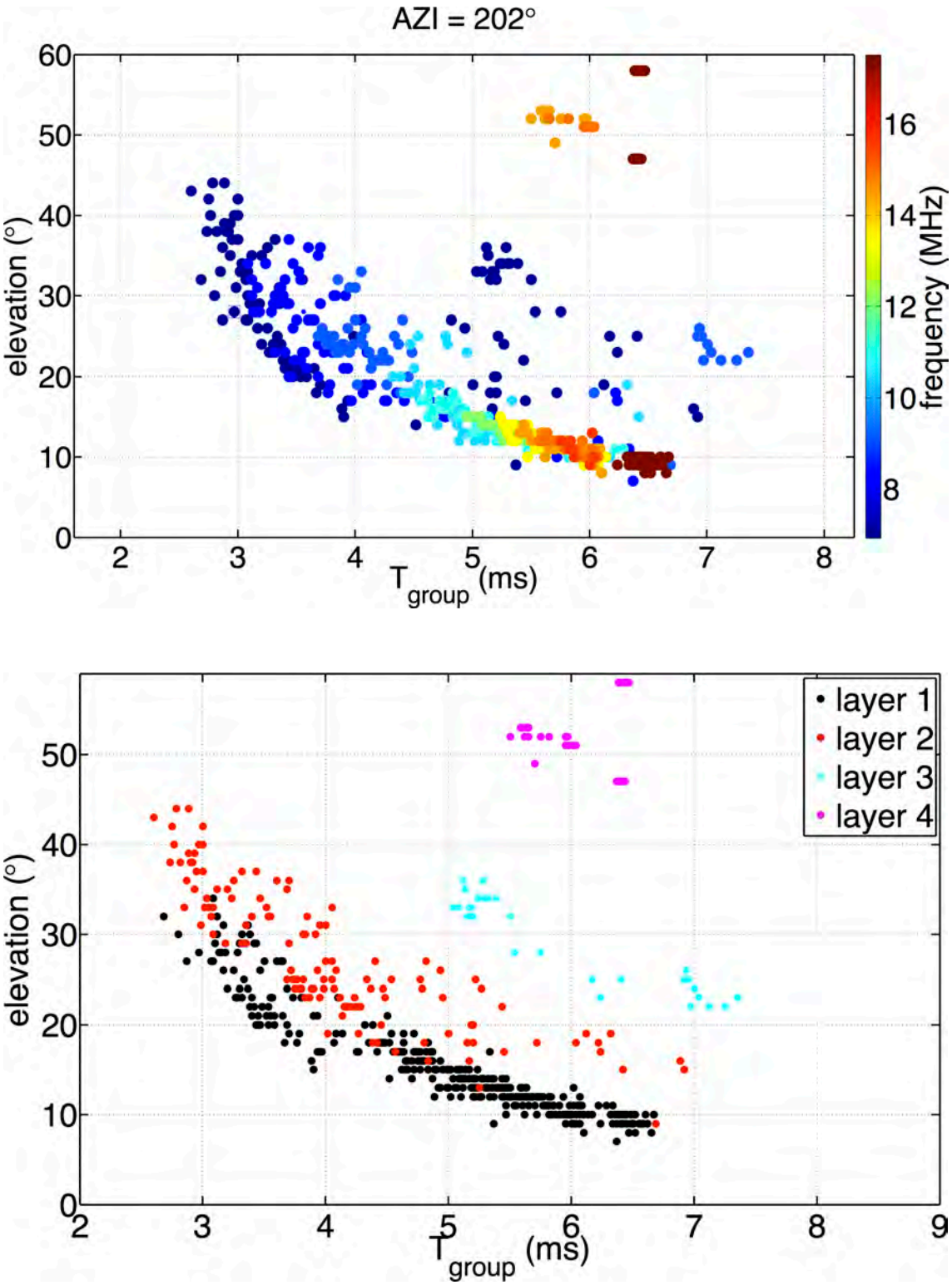


Figure 7.5: Raw data from March 14, 2006 in azimuth 202° (top) and assigned to different ionospheric layers using a clustering algorithm (bottom).

now clustered together. In this plot, the reflections from the F-layer (at 200 km) are

clearly visible, whereas those not from the E-layer are not observable, because it has already disappeared at night (1855 UT). The cyan and purple colored points having twice the virtual height of the black and red points may be multi-path reflections from the F-layer.

Based on this analysis, the real data can be color coded as ionospheric layers, in order to identify the data, which should be excluded from the dataset (Figure 7.5). The data have been assigned to 4 different layers, where data with larger propagation times that probably represents multi-path data, have been assigned to layers 3 and 4. In this way, the multi-path data can be excluded from the data set.

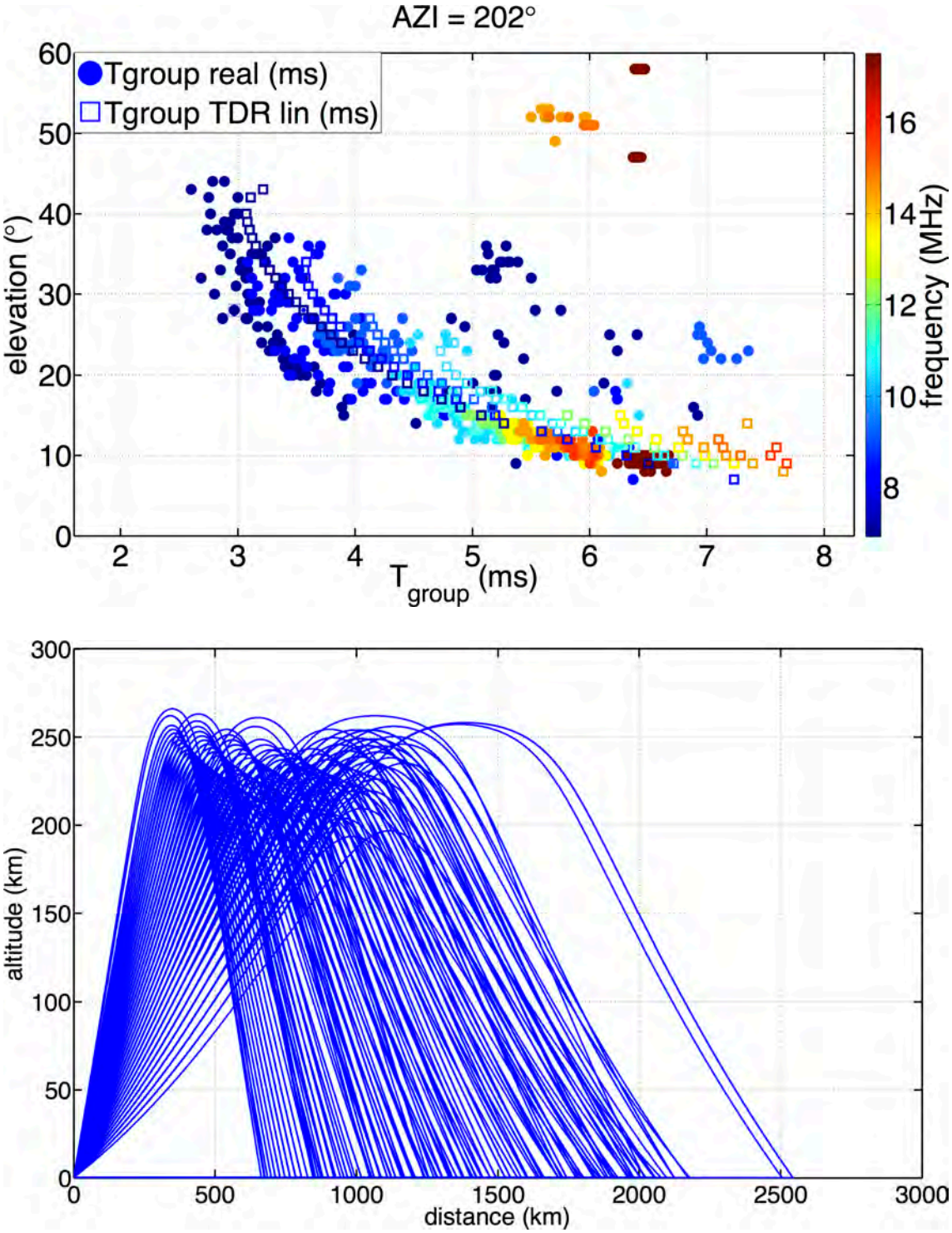


Figure 7.6: Group propagation time measured by the radar on March 14, 2006 in azimuth 202° (squares) compared to the propagation time obtained by ray tracing (dots) in the *a priori* ionospheric model (top) and corresponding rays (bottom).

Measurements performed by Nostradamus using the signal frequencies and elevation angles Nostradamus together with corresponding synthetic data allow calculation of a propagation time difference between data and synthetics. In Figure 7.6

(top), the group propagation times as functions of elevation are compared for real and synthetic data. The ray tracing was performed in the *a priori* ionospheric model NeQuick generated for the period and solar flux of the data. The corresponding rays are shown in Figure 7.6 in the bottom plot.

All rays are reflected between 200 and 250 km altitude, thus in the F region and not in the E region that has already disappeared. The ray tracing does not reproduce the multi-path data at elevation angles larger than 50°. In addition, the radar emits a beam in a certain azimuth with a specific frequency and elevation angle. This signal is backscattered at several points, so multiple propagation time measurements are possible or one frequency and one elevation angle. On the contrary, the ray tracing returns only one propagation time. Figure 7.6 illustrates a general overestimation of the data in terms of propagation time by the ray tracing.

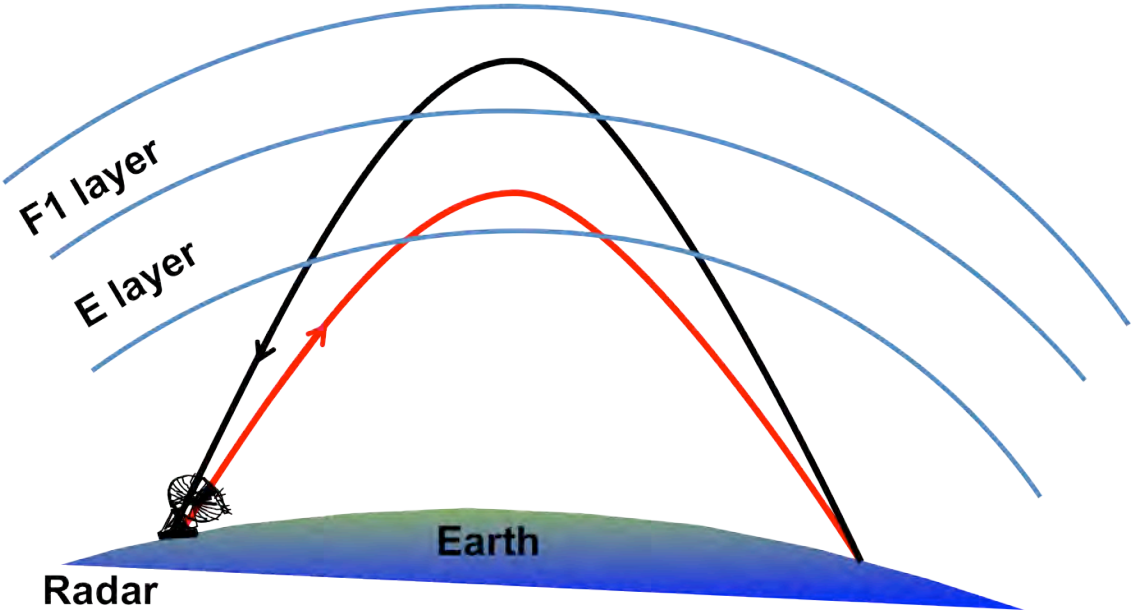


Figure 7.7: Explanation of path effects in the ionosphere. The signal propagates from the radar to the ground along a path in the E-layer (red), but along a path in the F-layer on its way back to the radar (black).

This might been explained by the fact that path propagated by the signal from the radar to the ground, and back to the radar are not the same (Figure 7.7). While propagating in the E-layer in the forward run, the signal might propagate in the F-layer when returning to the radar. This effect is not taken into account in the ray tracing in the *a priori* model.

7.2 Inversion results

7.2.1 Spring data set

The tomography method was tested on real data using the data set of march 14, 2006, at 1855 UT. It contains measurements in eight azimuths of 22°, 67°, 112°, 157°, 202°, 247°, and 337°. All data were preprocessed in the described way before inversion. After that, ray tracing of the remaining data was performed in the *a priori* ionospheric model NeQuick given for the appropriate day, time and solar flux was performed. The solar flux data is measured daily at local noon by the radio telescope in Penticon, Canada at 2800 MHz (10.7 cm wavelength), it is freely available on the NOAA FTP server. For March 14, 2006, the solar flux is 73.5 SFU.

Figure 7.8 shows the comparison between real data and synthetics used to compute the vector $\delta\mathbf{T} = T^{real} - T^{synth}$ of travel time perturbations for the azimuths 67°, 157°, 247° and 337°. The group propagation time differs between +20% and -40%. After ray tracing for the calculation of the synthetic data, the tomographic

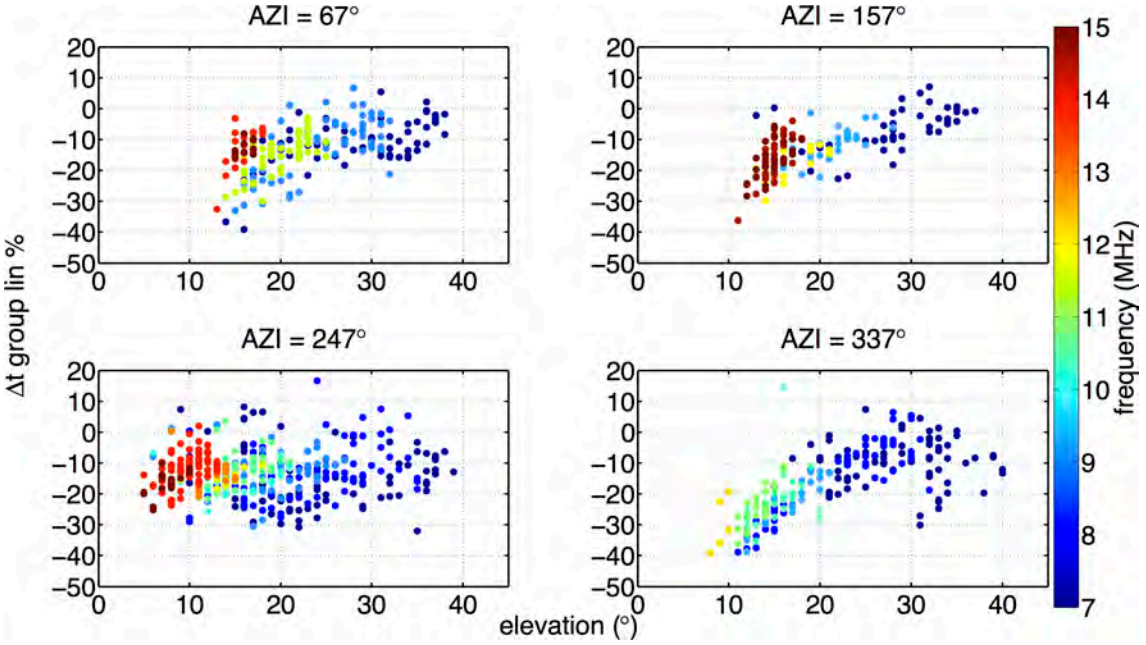


Figure 7.8: Difference in percent between the propagation time obtained by ray tracing in an *a priori* ionospheric model and real data of March 14, 2006 at 1855 UT for four different azimuths. The frequency is color coded.

matrices for the v- and the $v\&r$-method were calculated, followed by the inversions using the damped least squares inversion method as described in Section 4.5.

The inversion results of the real data after the first iteration in a selected azimuth (247°) for three different values of regularization in the v- and the $v\&r$-method are shown in Figure 7.9. The data set consists of 428 rays. Both methods find electron

density perturbations located between 200 and 300 km altitude and stretching up to 1500 km distance. It is not surprising that the inversion detects only a perturbation at that altitude, because at the time of measurement (1855 UT) the E-layer has nearly disappeared and the EM signals are reflected mainly in the F-layer. How-

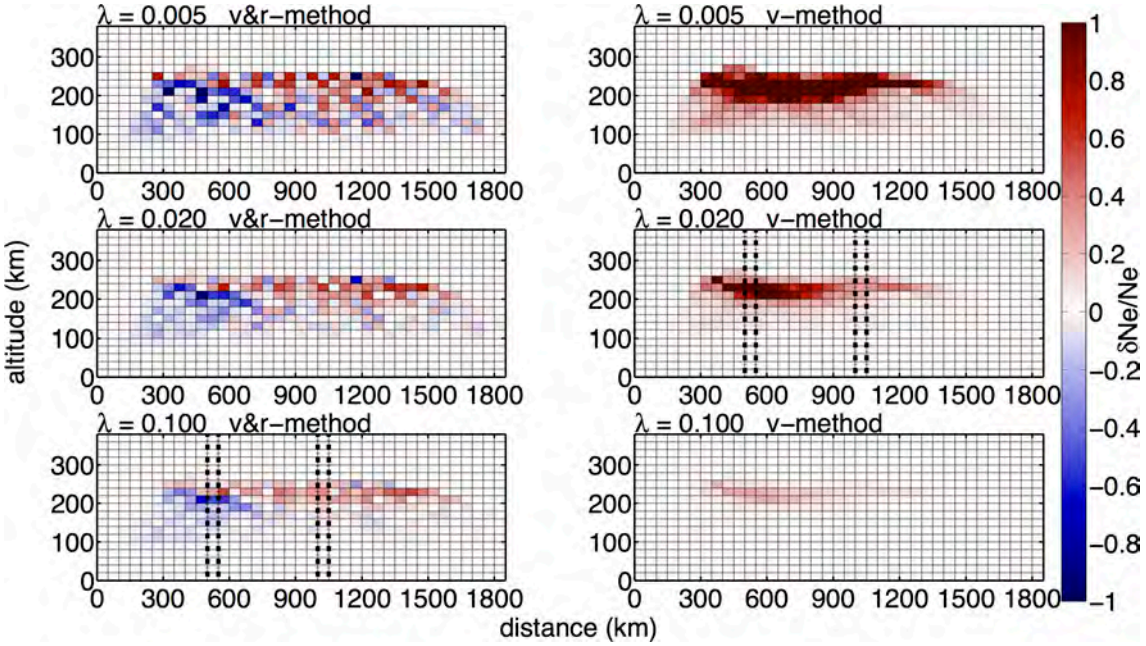


Figure 7.9: Real data inversion for data of March 14, 2006 at 1855 UT in azimuth 247°. In the $v\&r$-method (left) and the v-method (right), different values of regularization λ were used. The grid size is 50 × 20 km in distance and altitude. The results for the best regularization parameter are shown in the bottom left and the middle right plot, respectively.

ever, while the $v\&r$-method reconstructs positive and negative perturbations, the v-method finds an entirely positive and unrealistic perturbation of large amplitude. For the real data also applies that the regularization parameter λ does not change the inversion results for both methods, but a larger regularization leads to a more damped solution, reducing the amplitude of the solution. For the largest regularization $\lambda = 0.1$, the perturbation is mainly concentrated above 200 km altitude, whereas for smaller regularizations, other perturbations at 100 km are visible.

The best regularization parameters for both methods ($\lambda = 0.1$ and λ =0.02, respectively) were again calculated from the L-curve, where the λ range obtained for the checkerboard $\lambda = [10^{-3},\ 10^{3}]$ was used. Then a vertical profile of electron density can be calculated for any distance from the radar based on an inversion result like that in Figure 7.9. This allows correction of a vertical profile given from an *a priori* model.

Figure 7.10 illustrates the vertical profiles of electron density in 500 km distance from the radar in the *a priori* ionospheric model NeQuick as well as the profiles

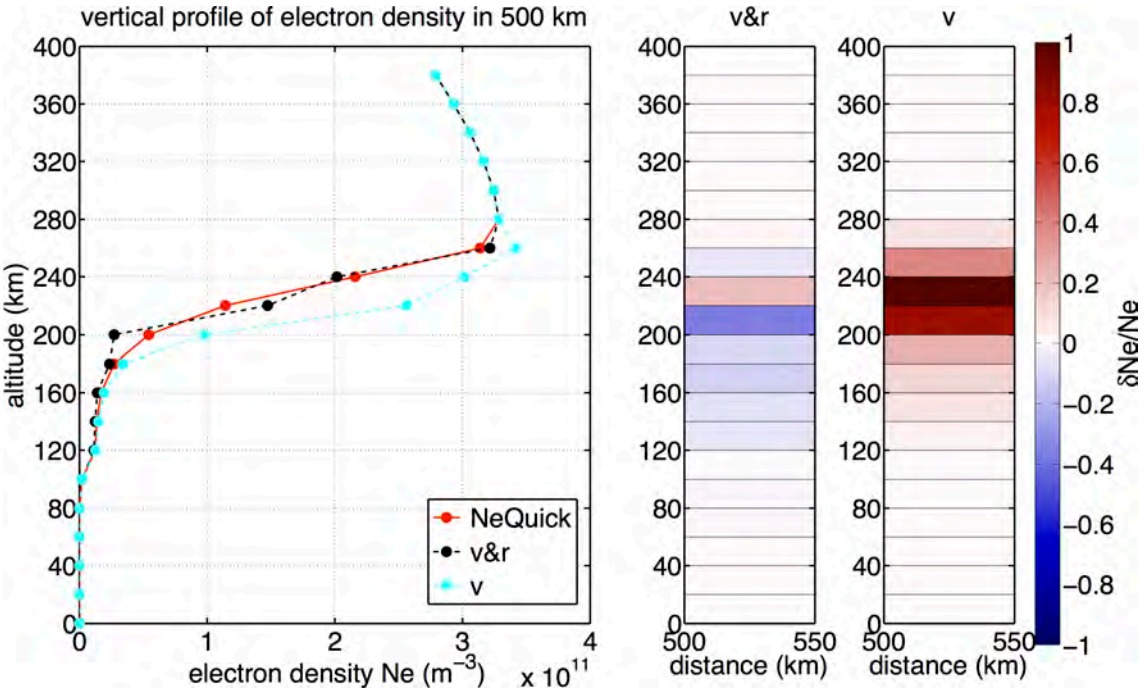


Figure 7.10: Vertical profile of electron density in 500 km distance from the radar calculated from real data inversion of OTH radar Nostradamus. The data were collected on March 14, 2006, in azimuth 247°. The electron density perturbations $\delta Ne/Ne$ obtained by inversion are shown in the two subplots on the right. The electron density perturbation of one point in the left plot corresponds to the block starting from the altitude of the point in the right plot.

corrected with the perturbation obtained from the $v\&r$ and the v-method. The vertical profile for NeQuick was calculated for the corresponding period and solar flux in azimuthal direction of 247° in 500 km distance and then parametrized like the grid with a value every 20 km. After that, it was corrected with the electron density perturbations found by the inversion marked by vertical dashed lines in Figure 7.9. The resulting vertical profiles show sensitivity to variations located between 180 km and 260 km altitude *(Roy et al., 2014)*.

In both methods, there are no corrections of the NeQuick electron density profile for altitudes lower than 160 km, because the inversion only finds perturbations in the F-layer and none for altitudes above 260 km. The electron density perturbation for the $v\&r$-method averages around 7% and for the v-method around 15% with a large perturbation at 220-240 km altitude.

Figure 7.11 illustrates the same three vertical profiles of electron density at 1000 km distance from the radar *(Roy et al., 2014)*. Here, both methods find entirely positive perturbations of electron density in the order of 40% an altitude range of 180-260 km. Two inversion methods (v and $v\&r$) do not differ considerably. Nevertheless, the vertical profiles show the strong variability of the ionosphere between two local

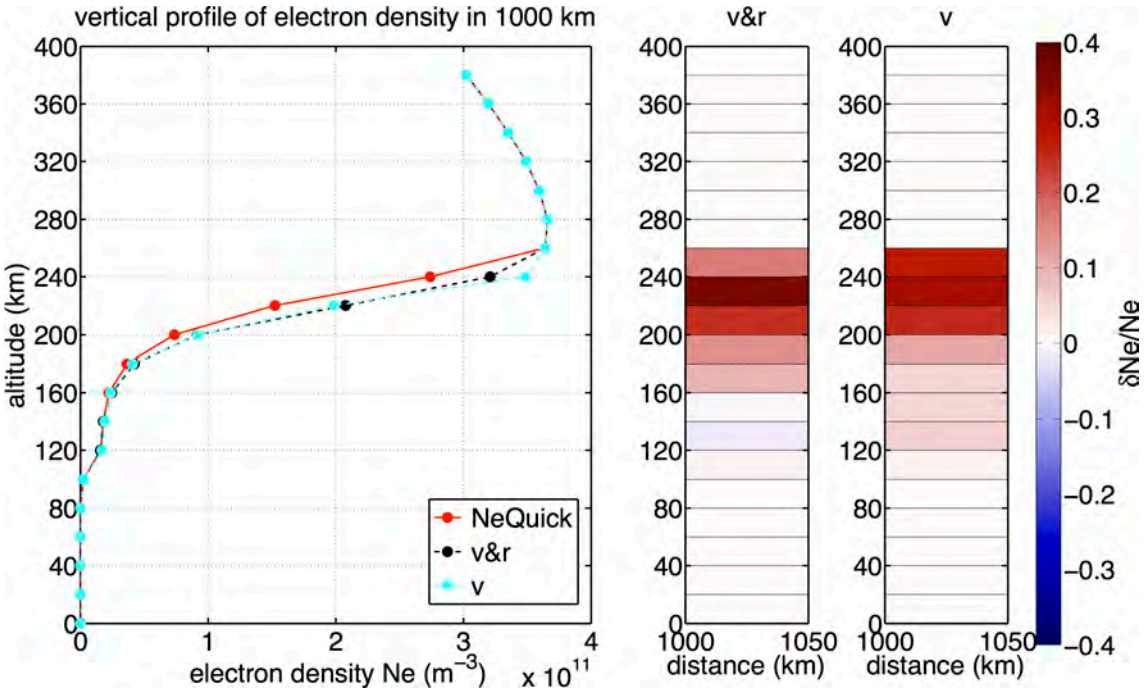


Figure 7.11: Vertical profile of electron density in 1000 km distance from the radar calculated from real data inversion of OTH radar Nostradamus. The data were collected on March 14, 2006 in azimuth 247°. The electron density perturbations $\delta Ne/Ne$ obtained by inversion are shown in the two subplots in the right. The electron density perturbation of one point in the left plot corresponds to the block starting from the altitude of the point in the right plot.

positions (500 km and 1000 km).

Comparison of the vertical profiles of electron density obtained by inversion at 500 km distance from the radar to vertical profiles over ionosonde stations in Europe are illustrated in Figure 7.12. The latter are reconstructed using Multi-Quasi-Parabolic (MQP) parametrization based on the ionospheric parameters given by the operators of the ionosondes. Clearly, the inversion result obtained with the $v\&r$-method is closer to most ionosondes in Europe and performs better than the v-method which is far away from the profiles over the ionosondes.

In order to demonstrate potential applications of the OTH radar Nostradamus in ionospheric tomography, the data for azimuths 67°, 157°, and 337° were also inverted for the same day and time. The inversion was performed with the damped least squares method using the $v\&r$-method, and for each azimuth the best regularization parameter was chosen from the L-curve in the range $\lambda = [10^{-3},\ 10^{3}]$ which has proven as the best in the checkerboard test. The results are plotted in Figure 7.13 in the corresponding azimuthal direction over Europe. Positive and negative electron density perturbations are visible in all azimuthal directions in an altitude range of $200 - 300$ km and a distance range up to 1500 km. As expected, the largest

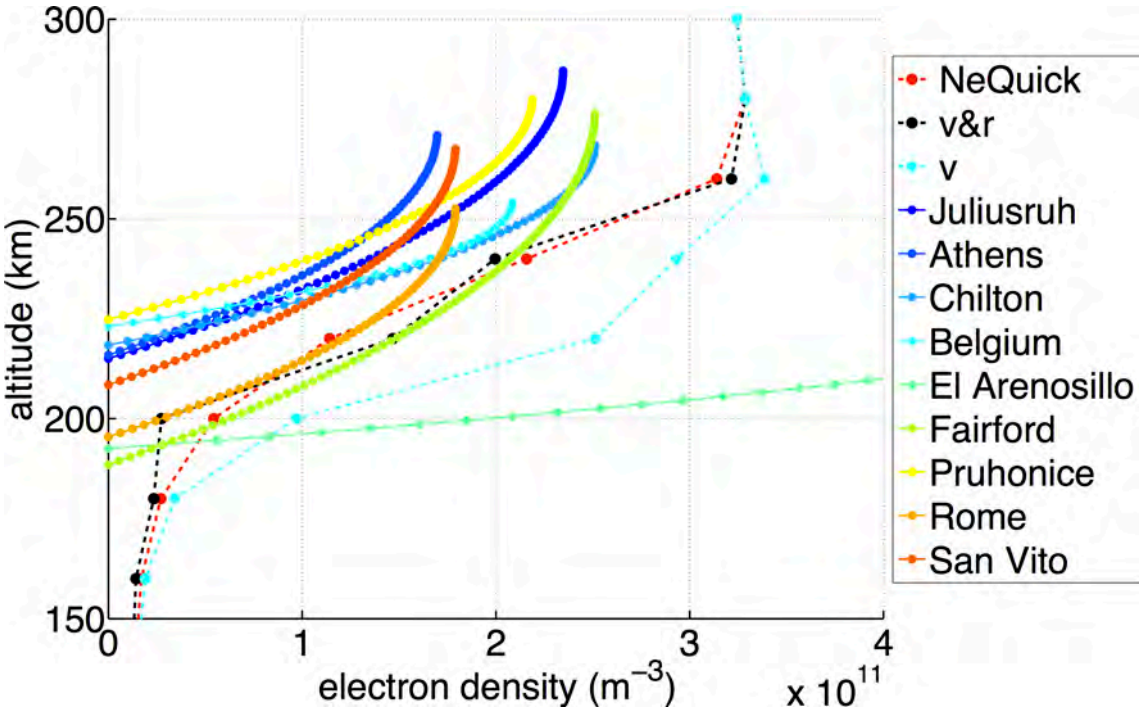


Figure 7.12: Comparison of vertical profiles of electron density above ionosondes in Europe with vertical profiles obtained by inversion of Over-the-horizon radar data. The data were collected march 14, 2006 in azimuth 247°.

electron density perturbation is seen in azimuth directions 247° and 337°, due to the position of the Sun at the time of measurement (1855 UT). In addition, Figure 7.13 illustrates the potential of the developed ionospheric tomography method. This method enables the calculation of a vertical profile of electron density in any desired distance and in any azimuth direction where the radar took measurements. In an intermediate azimuth, a vertical profile can be calculated by interpolation.

Further perspectives, based on the potential of this method, for instance the combination with GPS tomography are discussed in Section 7.3.2.

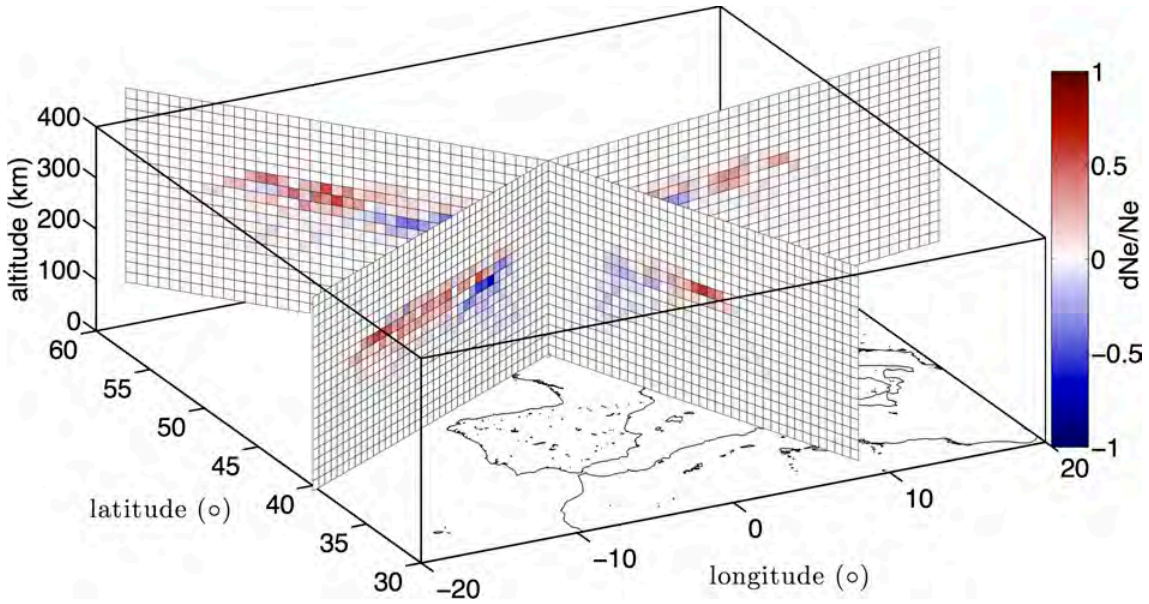


Figure 7.13: Electron density perturbation obtained by real data inversion. The data are collected at OTH radar Nostradamus on March 14, 2006, at 1855 UT in four azimuth directions, nominally 67°, 157°, 247°, and 337°. Source: *Roy et al. (2014)*

7.2.2 Winter data set

The tomography method was also applied to another data set recorded in winter (December) 2008. In that year, the solar flux was still small ∼60-70 SFU, because the Sun underwent a solar minimum. Therefore, ionization in the ionosphere was smaller and less rays were reflected. This time, a data set was chosen for noon, where the ionization is highest, and possibly reflections from the E and the F-layer are observable.

In Figure 7.14 the measured group propagation time in azimuth 67° and the propagation time obtained by ray tracing in the *a priori* model are compared. As already observed as for the spring data set, there are only very few reflections from the E layer in the synthetic data. This can be explained by fact that the E layer is generally not well estimated in ionospheric models. The propagation time measured by OTH radar Nostradamus is larger than that obtained from the ray tracing, probably due to the path effects mentioned above.

Figure 7.15 shows the inversion results using the data measured on December 13, 2008, at noon with different regularization parameters λ. The inversion results with the best regularization parameter chosen from the L-curve are $\lambda = 0.1$ for the $v\&r$- and $\lambda = 0.05$ for the v-method. The perturbations located between 100 and 300 km altitude are in the order of 20%. They appear in a distance range up to 1500 km for the $v\&r$-method, whereas the v-method seems to be strongly damped. Moreover, such large-scale entirely negative perturbations are hardly realistic and

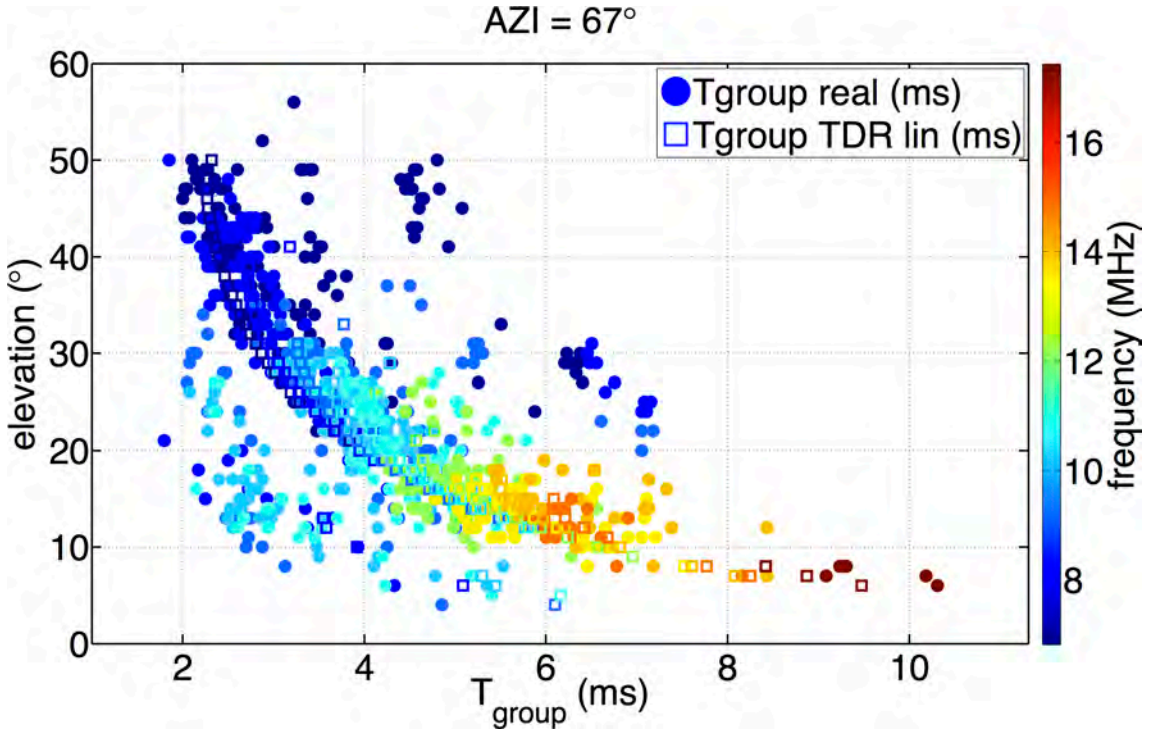


Figure 7.14: Group propagation time measured by OTH radar Nostradamus on December 13, 2008 in azimuth 67° compared to the propagation time obtained by ray tracing (dots) in the *a priori* ionospheric model.

indicate a better performance of the $v\&r$-method.

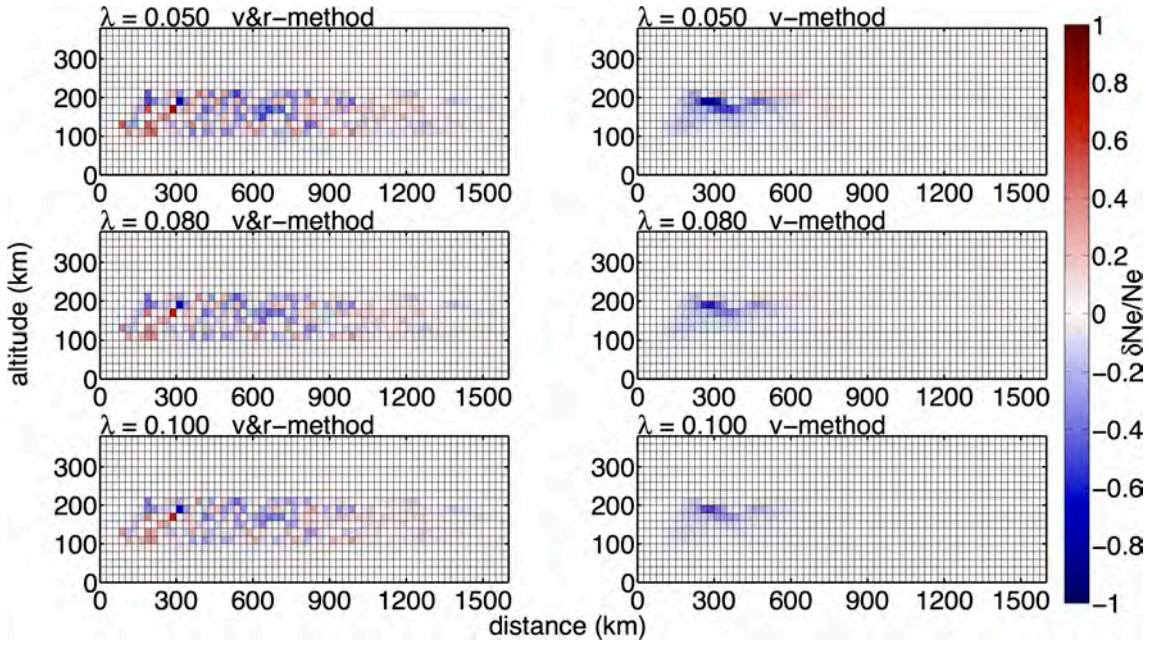


Figure 7.15: Real data inversion for data of December 13, 2008 at 1154 UT in azimuth 67°. In the v&r-method (left) and the v-method (right), different values of regularization λ were used. The grid size is 25×50 km in distance and altitude, respectively.

Perturbations in the v-method inversion results are only detectable up to 600 km. In a next step, vertical profiles of electron density were calculated for 200 and 500 km distance from the radar in 67° azimuth. These profiles are shown in Figure 7.16 and 7.17. At 200 km distance from the radar, the perturbations found by the two inversion between 100 and 180 km altitude are contradictory. While the $v\&r$-method finds positive perturbations, those detected with the v-method inversion are negative. Above 180 km altitude, both methods give constant results and reconstruct negative anomalies. In higher altitudes, perturbations are found no longer due to the lack of reflected rays, and the vertical profiles coincide with the NeQuick model.

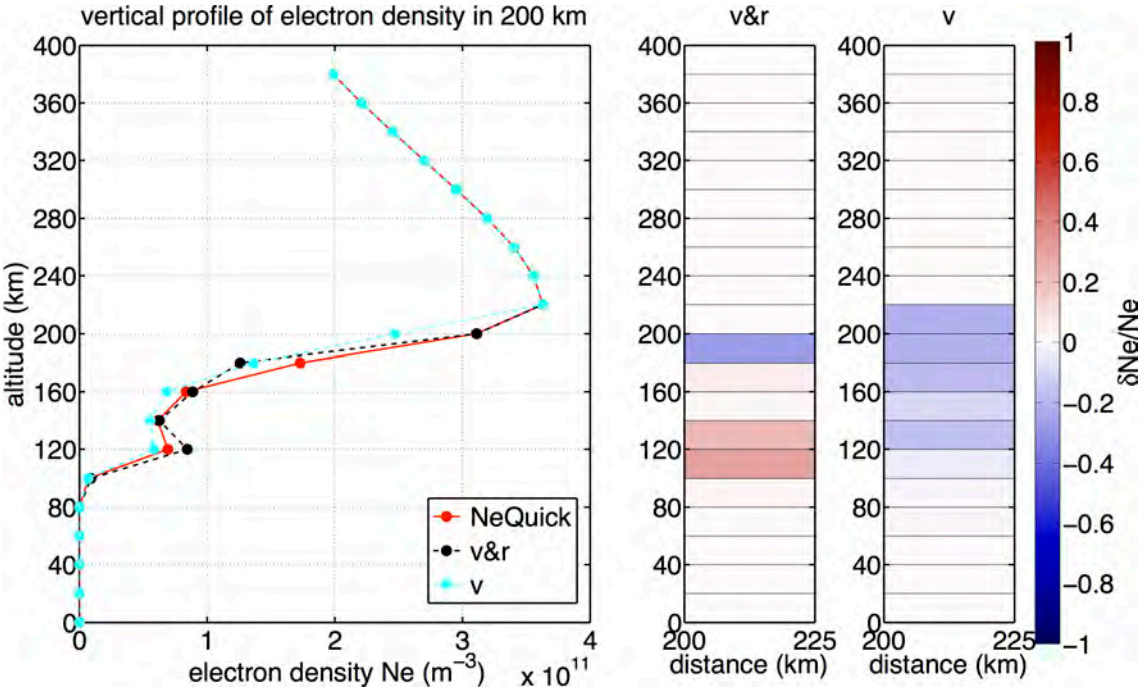


Figure 7.16: Vertical profiles of electron density at 200 km distance from the radar using data from December 13, 2008, measured in 67° azimuth. The electron density perturbations are obtained from the inversion in Figure 7.15 for the best regularization parameter. The electron density perturbation of one point in the left plot corresponds to the block starting from the altitude of the point in the right plot.

At 500 km distance from the radar (Figure 7.17), both inversion methods coincide finding negative perturbations (Figure 7.17) up to 160 km altitude, with those reconstructed by the $v\&r$-method being larger (20%) compared to the v-method (<10%).

Above 220 km altitude, no electron density perturbations are detected, and the vertical profiles for v and $v\&r$-method fit the profile from the *a priori* model.

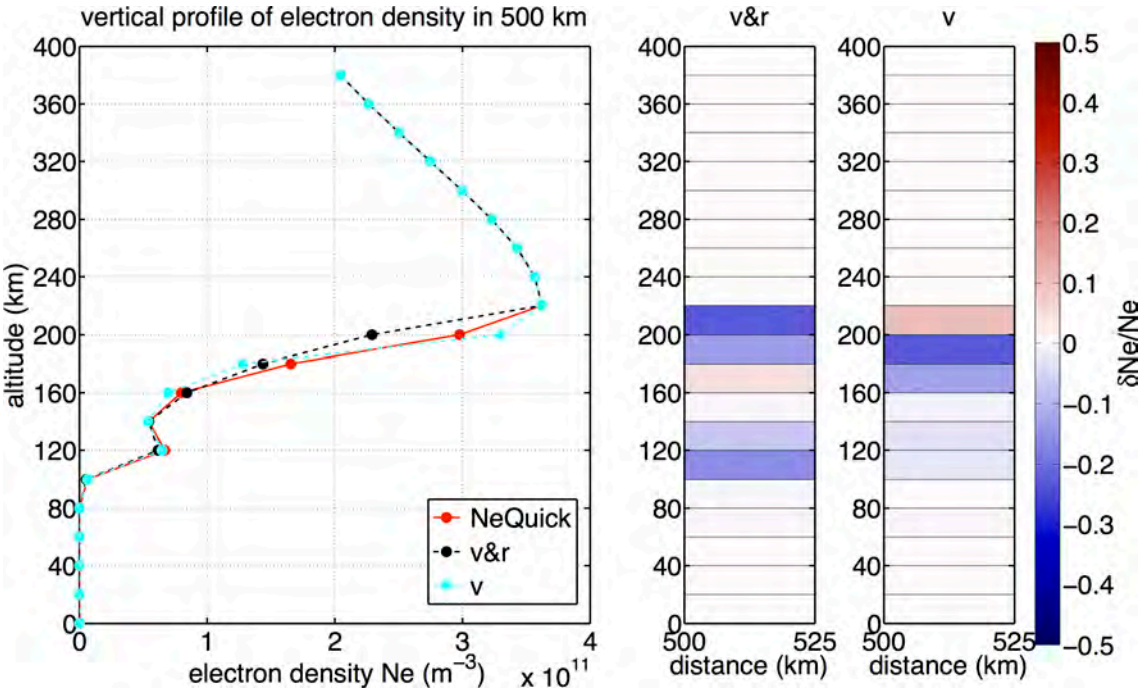


Figure 7.17: Vertical profiles of electron density at 500 km distance from the radar using data from December 13, 2008, measured in 67° azimuth. The electron density perturbation of one point in the left plot corresponds to the block starting from the altitude of the point in the right plot.

7.3 Perspectives

7.3.1 Comparison with ionosonde data

As shown, this ionospheric tomography method allows to calculate vertical profiles of electron density at any desired distance reached by the radar. Nevertheless, this results need to be validated by other instruments that provide vertical profiles, such as ionosondes. Some of them supply profiles of the electron density above the station calculated after ionogram inversion, and these profiles can be downloaded directly from the European DIgital upper Atmosphere Server (DIAS) homepage. Other stations provide only the three ionospheric parameters f_c, h_m, and y_m, but on the basis of these data, a vertical profile can be calculated, using for instance, the MQP modelization *(Croft and Hoogasian, 1968)*.

However, the choice of an ionosonde for comparison with the inversion results from Nostradamus requires consideration of the distance of the ionosonde from the radar and the azimuth. Nostradamus data were collected step-by-step in eight different azimuths, starting at 22° with steps of 45°. This is the normal operation mode of this radar. The radar range depends on the frequency and is in the order of 800 to 1600 km. Consequently, ionosondes to be considered for comparison should

be sited in one of the azimuths Nostradamus took measurements and within the half distance range that is 400 to 800 km. Data from Nostradamus are available for the years 2006-2012. Table 7.1 lists some ionosondes in Europe as well as their distance to Nostradamus and their azimuth. The ionosondes fulfilling the conditions in distance and azimuth mentioned above are highlighted in Table 7.1. These are the stations in Chilton, Dourbes, Fairford, Pruhonice, and Ebre.

Ionosonde Station	**latitude** (◦)	**longitude** (◦)	**azimuth** (◦)	**distance** (km)
Juliusruh, Germany	54.60	13.40	47.36	1075
Athens, Greece	38.00	23.50	114.97	2151
Chilton, U.K.	**51.50**	**359.40**	**339.96**	**339.5**
Dourbes, Belgium	**50.10**	**4.6**	**56.20**	**301.82**
El Arenosillo, Spain	37.1	353.30	209.03	1429
Fairford, U.K.	**51.70**	**358.50**	**332.59**	**386**
Pruhonice, Czech Rep.	**50.00**	**14.60**	**76.14**	**989.47**
Rome, Italy	41.8	12.50	126.21	1171
Ebre, Spain	**40.8**	**0.50**	**183.2**	**872.65**
San Vito, Italy	40.60	17.80	117.9	1591

Table 7.1: Ionosonde stations in Europe with distances and azimuths relative to OTH radar Nostradamus. Blue highlighted stations can be considered for comparison with inversion results from Nostradamus. Source: `http://car.uml.edu/common/DIDBFastStationList`

The position of the other ionosondes is too far away from Nostradamus or their azimuths differ too much. In a preliminary test, the vertical profiles of electron density obtained by the ionospheric tomography with Nostradamus data were compared to the vertical profile of the ionosonde station Chilton. Data from Chilton, sited at azimuth 340° and in 340 km distance (Table 7.1) can be compared with measurements of Nostradamus taken in 337° azimuth. The ionogram of Chilton at December 19, 2008, at noon (Figure 7.18) illustrates that only an E and a F2 layer observable. Based on the ionospheric parameters listed in the left column of the ionogram, the vertical profile (black line in Figure 7.18) can be reconstructed using MQP modeling.

The inversion results of the Nostradamus data using the v and the $v\&r$-method are shown in Figure 7.19. For both methods, the L-curve was used to choose the best regularization parameter λ. The electron density perturbation is located between 100 and 300 km altitude and in the order of 50% for the $v\&r$ method, but very small for the v-method (5%). This can be explained by the large regularization parameter λ that strongly damps the solution. For this reason, all electron density perturbations are located at 200 km altitude in the v-method.

For comparison with the vertical profile of Chilton, at first a vertical profile

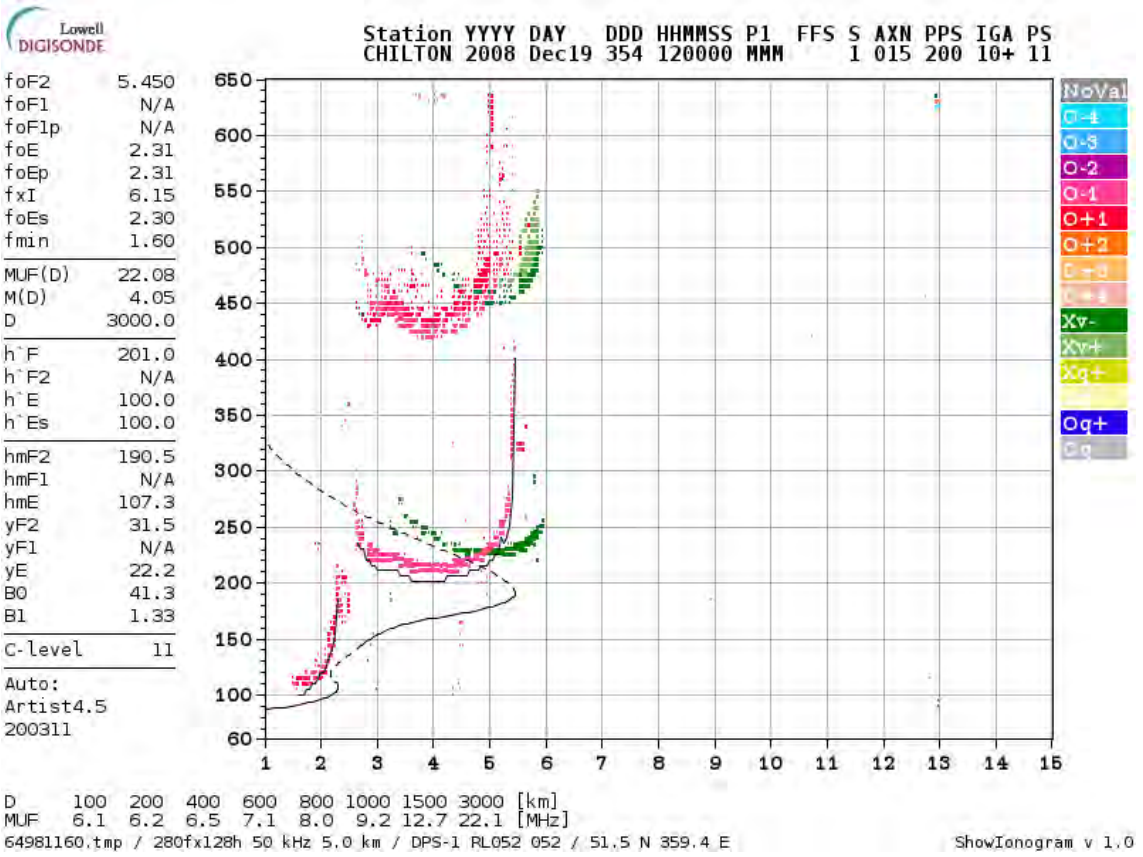


Figure 7.18: Ionogram measured at Chilton on December 19, 2008, at noon and inverted vertical profile of plasma frequency over the station (black line). Source: `http://car.uml.edu/common/DIDBDayStationStatistic?ursiCode=RL052&year=2008&month=12&day=19`

for December 19, 2008 was computed in the *a priori* ionospheric model NeQuick at the position of Chilton (Table 7.1). This profile was subsequently corrected by the perturbations found by inversions with the v- and the $v\&r$-method at the corresponding distance between 325 and 350 km marked by dotted lines in Figure 7.19.

The vertical profiles of electron density are compared in Figure 7.20. The profile from the *a priori* model NeQuick is shifted upwards in altitude as compared to the profile from Chilton. Additionally, the electron density derived from the model NeQuick is generally smaller than that obtained from the vertical sounder. In the selected distance range (325 and 350 km), the $v\&r$-method reconstructs positive and negative electron density perturbations that modify the *a priori* ionospheric model to even smaller perturbations at altitudes of 120 to 160 km (E-layer). Consequently, the corrected profile shifts away from the Chilton profile. By contrast, beyond those altitudes the profile corrected by the perturbation that is more similar by the $v\&r$-method is closer to the profile derived from Chilton.

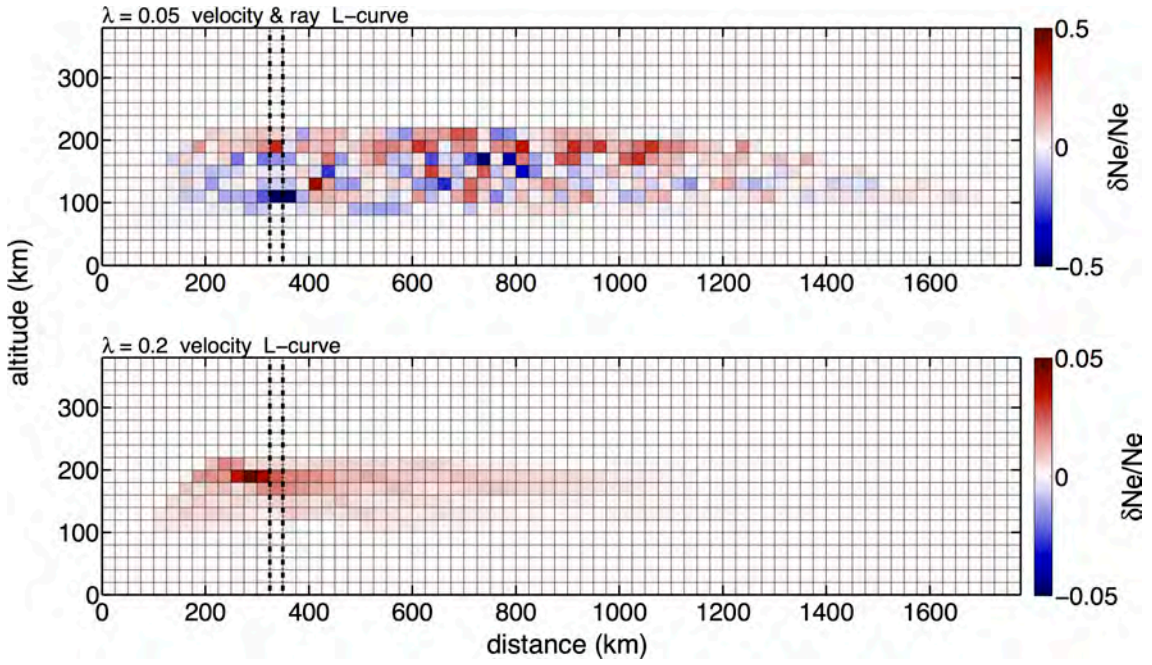


Figure 7.19: Inversion results using real data measured by OTH radar Nostradamus on December 19, 2008, at noon in 337° azimuth. The perturbations found at distances marked by dashed lines were used to correct a vertical profile of electron density from the *a priori* ionospheric model and shown in Figure 7.20.

However, the profile from Chilton is not obtained by direct measurement, but by inversion of the measured ionogram, using an inversion method that is unknown in detail. Since the inversion result found by the v-method is strongly damped and the perturbations are very small (5%), the profile corrected by the v-method coincides with the profile in the *a priori* model NeQuick. Synthetic tests have shown (Chapter **??**) that, in general, a satisfying solution needs more iterations using the v-method. Additional investigations on this subject are required.

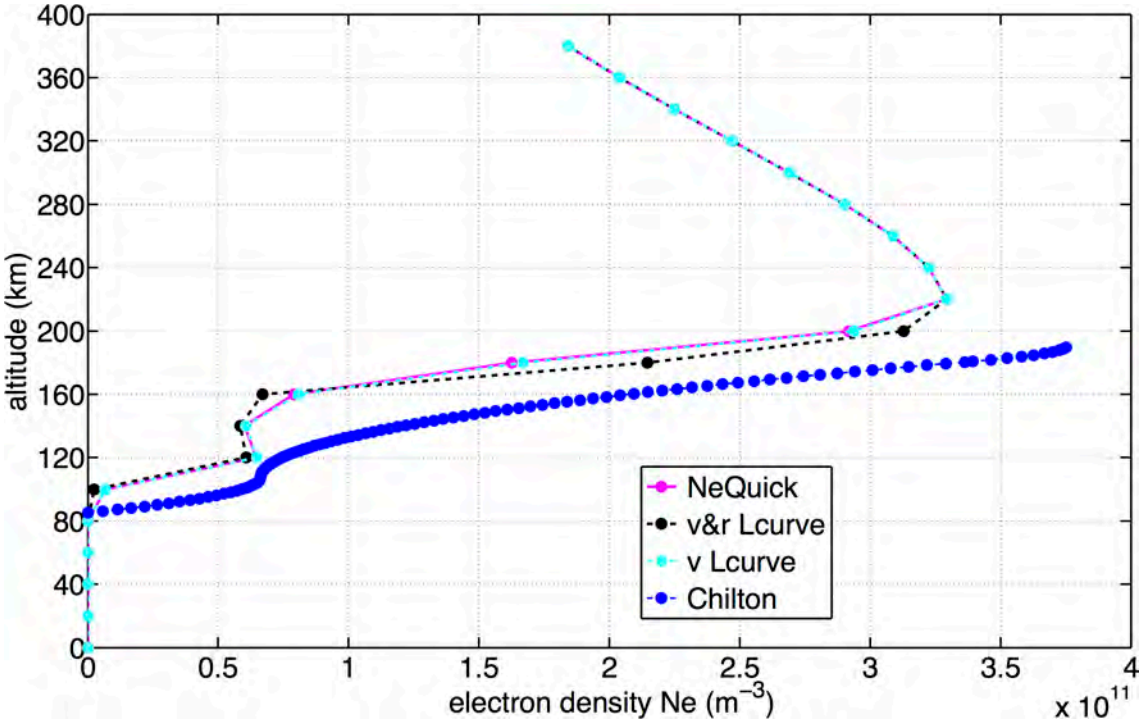


Figure 7.20: Comparison of the vertical profiles of electron density obtained at the ionosonde station Chilton (blue) on December 19, 2008, and by ionospheric tomography using the v- and the $v\&r$-method.

7.3.2 Combination with GPS

As discussed in Chapter 4, ionospheric tomography based on GPS is limited to the region of maximum ionization in the ionosphere (F2-layer), because the GPS works with high signal. On the contrary, ionospheric tomography with OTH radar allows good reconstruction of the ionospheric plasma up to 300 km altitude within the radar range, as the tests with synthetics and real data demonstrated. In order to reconstruct the ionospheric plasma completely over the entire altitude scale, a combination of GPS and OTH radar in one inversion should be realized.

A first synthetic test was conducted to check the feasibility of such an approach. Virtual GPS stations were put on the ground along a line in 89° azimuth and a satellite at 20 200 km altitude is assumed to move in that direction. Then, an electron density perturbation δNe in form of a checkerboard test of 5% was added to the *a priori* model, followed by the calculation of the vector with differential TEC for the GPS inversion and the vector of propagation time difference $\frac{\delta \mathbf{T}}{\mathbf{T}}$ for the OTH radar. The radar data was simulated for an ideal case by tracing 1071 rays with elevation angles between 10° and 60° and in a frequency range of 6 −16 MHz. The differential

TEC, $dTEC$, is given by

$$dTEC = \int \delta Ne \cdot ds, \tag{7.2}$$

with ds being the ray-path between a satellite and a ground station. The problem to be solved is given by Equation 7.3. The two matrices for the radar and the GPS form one matrix that has to be inverted.

$$\begin{pmatrix} \frac{\delta T}{T}_1 \\ \frac{\delta T}{T}_2 \\ \vdots \\ \frac{\delta T}{T}_i \\ --- \\ dTEC_1 \\ dTEC_2 \\ \vdots \\ dTEC_k \end{pmatrix} = \begin{pmatrix} A_{11}^{OTH} & A_{12}^{OTH} & \dots & A_{1j}^{OTH} \\ A_{21}^{OTH} & A_{22}^{OTH} & \dots & A_{2j}^{OTH} \\ \vdots & \vdots & \vdots & \vdots \\ A_{i1}^{OTH} & A_{i2}^{OTH} & \dots & A_{ij}^{OTH} \\ ---- & ---- & ---- & ---- \\ A_{11}^{GPS} & A_{12}^{GPS} & \dots & A_{1i}^{GPS} \\ A_{21}^{GPS} & A_{22}^{GPS} & \dots & A_{2i}^{GPS} \\ \vdots & \vdots & \vdots & \vdots \\ A_{k1}^{GPS} & A_{k2}^{GPS} & \dots & A_{kj}^{GPS} \end{pmatrix} \cdot \begin{pmatrix} \frac{\delta Ne}{Ne}_1 \\ \frac{\delta Ne}{Ne}_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \frac{\delta Ne}{Ne}_j \end{pmatrix} \tag{7.3}$$

$\mathbf{A}^{OTH}$ is the matrix $\mathbf{A} + \mathbf{A'}$ for the $v\&r$-method (Section 4.5.4). The matrix for the GPS is given by

$$\mathbf{A_{ij}}^{GPS} = Ne_i^0 \cdot ds_{ij}, \tag{7.4}$$

with Ne^0 being the electron density from the *a priori* ionospheric model.

The two data sets have very different values, being larger in the GPS $dTEC$ vector ($\sim 10^{15}$) compared to the OTH propagation time difference vector (~ 0.25). Additionally, as a result of the GPS station configuration chosen, the number of $dTEC$ measurements doubles the radar data. In order to perform a joint inversion considering both data sets without masking one, the two datasets need to be weighted:

$$|d_{OTH}| \sim \text{weight}^2 \cdot |d_{GPS}| \,. \tag{7.5}$$

$|d_{OTH}|$ is the L_2-norm of the vector with travel time differences $\frac{\delta \mathbf{T}}{\mathbf{T}}$ and $|d_{GPS}|$ is the norm of the differential TEC values. The weight is applied to the latter and the matrix for the GPS.

Figure 7.21 illustrates the joint inversion result for the simulated checkerboard perturbation of 5% on a grid of 50×20 km in distance and altitude, respectively. The best regularization parameter was chosen in the way described in Section 4.6.3.

Due to the GPS, reconstruction of the target model is quite well between 300 and 400 km altitude and negative and positive anomalies are clearly separated. Between 100 and 300 km altitude, the OTH radar contributes to reconstruction of the target model up to 800 km distance and negative as well as positive electron density perturbations are indicated. Probably, this inversion result can be improved further by applying a damped least squares inversion with constraints (Section 4.5.3). For larger distances, no perturbation at low altitude is detected by the GPS based on its high-frequency signal. However, the high-altitude electron density irregularities are slightly enlarged in the direction of the top boundary of the grid and do not end distinctly as in the target model.

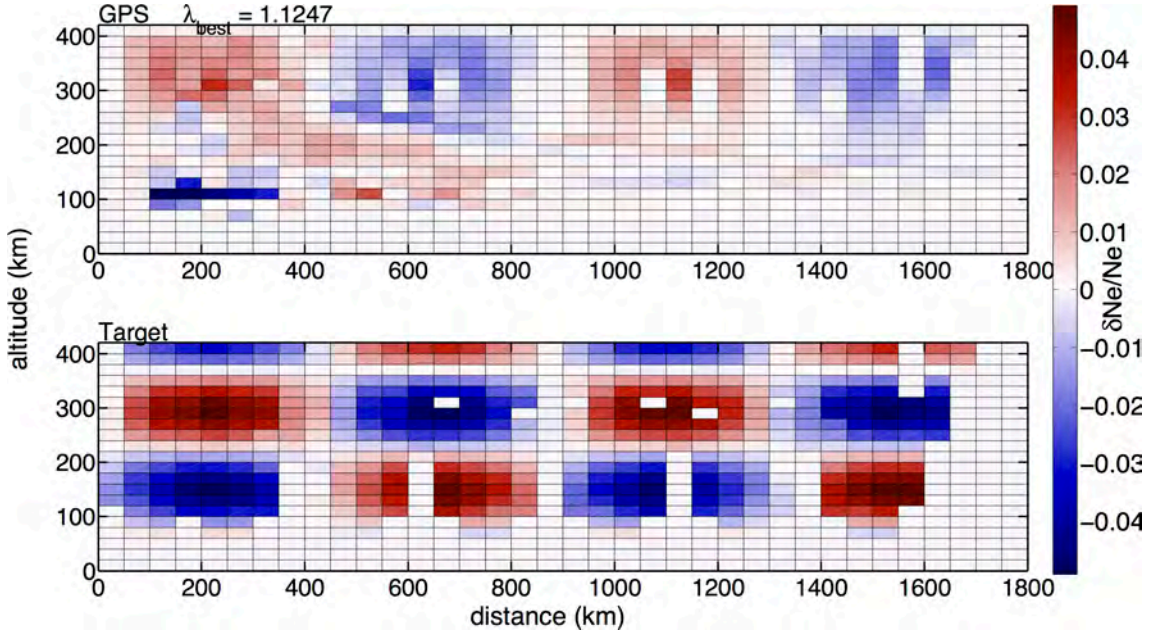


Figure 7.21: Joint inversion of OTH radar and GPS (top) using simulated data for a checkerboard perturbation of 5%. Perturbation gaps in the target model are due to missing ray-paths. The grid size is 50×20 km in distance and altitude, respectively.

The test described above showed that a joint inversion of GPS and OTH radar can improve the inversion result. However, this test is based on an unrealistic GPS station distribution along a line in the azimuth direction of the radar.

7.4 Conclusions

In this chapter, the ionospheric tomography method, previously tested for synthetic data, was applied to real data measured by OTH radar Nostradamus. The performance was analyzed on two data sets, obtained on early evening in spring 2006 and at noon in winter 2008. Contrary to the synthetic data, inversion of real data needs some preprocessing of the data, to consider multi-path and path effects in

the real ionosphere as compared to the *a priori* ionospheric model. A good tool to accomplish that was found in the virtual height of reflection, easily calculated by Martyns theorem.

For both data sets, the v and the $v\&r$-method were used for inversion, and their performances were compared after one iteration. Generally, for the best regularization parameter chosen from the L-curve, the v method inversion results are strongly damped, their amplitudes reduced and the electron density perturbations very localized. This can be explained, as discussed in Chapter 6, with the best value of regularization being larger than the eigenvalues in the matrix, suppressing all useful information. With the $v\&r$-method and the same technique to choose a regularization parameter, anomalies are found within the range of the radar up to 1500 km distance. To go a step further, the electron density profiles from the *a priori* ionospheric model NeQuick generated for the corresponding period and solar flux, were corrected with the electron density perturbations found by inversion.

However, the real data suffers from a reduced number of rays as compared to the previous synthetic tests, making the problem less unique and ill-posed. Additionally, the Sun underwent a deep solar minimum in 2006 and still in 2008, so the solar flux and ionization in the ionosphere were small. This reduced the number of reflected rays and the ray coverage, which became evident in the inversions, where electron density anomalies are mainly reconstructed in the F-layer, even at noon when an E-layer is present, too. Therefore, *a priori* vertical profiles of electron density were mainly corrected in the altitude range between 180 and 240 km and coincide with the vertical profile from the *a priori* ionospheric model.

As discussed in the Section 7.3, the vertical profiles of electron density obtained by inversion can be compared with those from ionosonde stations in Europe. A preliminary test demonstrated this to be practicable for data from the station in Chilton and the $v\&r$-method showing a better performance. Nevertheless, this approach needs further development, with more iterations, since the v-method generally needs more iterations than the $v\&r$-method. Additionally, for a comparison, the methodology used to invert the Chilton data needs a deep understanding to minimize the effect of methodical differences on the results. The raw data derived from the Chilton ionosonde could be included into our general joint inversion using several instruments (i.e., GPS, OTH radar, ionosondes).

With the inversion of real data in four different azimuths, the impressive potential of the tomography method was illustrated, showing the general possibility of a 3D tomography over Europe. As discussed in the last section of this chapter, and demonstrated in a synthetic test, this ionospheric tomography method could be combined with the tomography done by GPS. This would improve the lack of sensitivity of the GPS to the lower ionosphere at 100 km altitude and the two methods

would complement one another. Since the OTH radar Nostradamus is not available at the moment, only data from the previous years can be used, measured in the 8 azimuths in the normal operation mode. For this data, corresponding GPS orbits need to be found. Another option is to use radar data from one of the other OTH radar in the U.S, Australia or the SuperDARN. In order to apply the tomography method to these other radars, they need to provide measurements of the elevation angle and the frequency, or the tomography method needs to be adapted at their measurements.

Chapter 8

Conclusions and Perspectives

In the first part of this chapter, the conclusions of the research described here will be summarized. There are several areas of this study, which can be further explored or where further developments based on these results are promising. These perspectives will be summarized in the second part of this chapter.

8.1 Conclusions

In this thesis, a method for ionospheric tomography for monostatic OTH radar (receiver and transmitter are at the same place) was developed for the first time. All previous established ionospheric tomography methods mainly base upon GPS. During the last decade, many efforts were made to improve these methods and to overcome limitations due to ray-coverage and the non-uniqueness of the inverse problem (Section 4.2). However, due to the high frequency of the GPS signal, the attempts to ionospheric tomography were only sensitive to the region of maximum ionization (the F2-layer).

Validation of the method

The tomography method developed for OTH radar was validated on a number of different synthetic benchmark tests (Chapter 4) that consist of checkerboard and localized perturbations. The v-method, considering only the effect of electron density perturbation on the propagation time of the EM waves, was found to be insufficient in a first iteration, because it does not reconstruct the target models well. Therefore, the $v\&r$-method was established, taking into account the ray-path deflection introduced by the unknown endpoints of the rays.

The concept of ray-path deflection was validated by the idealized case of *frozen rays*, where the ray-path is not perturbed, as if the ray end-points had been known. In this case, synthetic target models are well reconstructed in the area of high ray

coverage within the range of the radar.

The $v\&r$-method can be improved further, reducing the noise by smoothing with a damped-least squares inversion with constraints (Section 4.6.4). The v-method does not need a smoothing in the first iteration.

Resolution analysis

The v- and the $v\&r$-methods were applied to a bunch of different synthetic benchmark tests of different size and amplitude. This detailed sensitivity and accuracy study allowed to investigate the resolution of the tomography method (Chapter 5).

In these tests, the influences of the grid, cell dimensions, regularization parameter as well as smallest perturbation amplitude where analyzed. The main conclusions of these resolution tests are that the inversion results for both methods are independent of the grid size using checkerboard or localized perturbations. In an even better result, two small localized perturbations of 165×20 km can be identified and separated by both methods.

However, the $v\&r$-method seems to perform better in the first iteration for checkerboard perturbations and localized perturbations because it identifies the correct position of the maximum anomaly and reconstructs more details of the checkerboard perturbation in larger distance. The v-method mislocates the perturbation, is not able to reproduce its shape and is strongly damped in the case of the checkerboard.

Both inversion results (v and $v\&r$) show large-scale negative anomalies for the localized perturbation that are not present in the target model and do not appear in the *frozen ray* inversion. These artifacts can be explained with nonlinear effects of ray-path deflections, which (endpoints being not fixed and unknown), in principle, can result in a faster propagation even if the velocity perturbation is negative.

All synthetic tests illustrated the dependence of the solution on the ray-coverage and on the zone of sensitivity of rays to the medium, where the plasma frequency approximates the signal frequency. The checkerboard perturbation demonstrated these dependences clearly, because perturbations are neither at large distances or zones with few rays, nor at low altitude well resolved.

Development of an iterative methodology

The inversion results obtained after the first iteration for a localized perturbation showed that for the $v\&r$-method the amplitude of the solution is too small, whereas the v-method mislocates the perturbation. In order to improve these inversion results, an iterative methodology was developed for the v and the $v\&r$-method. The preliminary results for an iterative approach published by *Roy et al. (2014)*

show a decrease in data and model misfits with iterations and a subsequent increase in the model misfit after a critical number of iterations. At the model misfit minima the difference between target and solution is less than 30% for $v\&r$ at iteration 13 and less than 40% for the v-method at iteration 5 *(Roy et al., 2014)*. This lead to the conclusion that the iteration approach generally improves the inversion results, because data and model misfits are reduced as compared to the first iteration, but the system becomes unstable after a critical number of iterations. To address with this problem and reduce the instability, the effects of feedback and regularization were tested for both methods. The first parameter controls the amplitude, the second one damps the re-injected solution.

Both parameters were found to slow down convergence in both inversion methods, but they do not reduce the instability. However, the $v\&r$-method reaches model misfit minima generally faster than the v-method. This may be explained by the fact that the $v\&r$-method takes into account the ray-path deflection starting from the first iteration, while the v-method seems to consider it indirectly with the ray tracing in an updated model after the first iteration. This became obvious with the horizontal movement of the electron density perturbation to the correct position with iterations in the v-method. At their minima, the model misfits for the v and the $v\&r$-method are in the same order. The instability in the inversion, being more severe in the $v\&r$ method than in the v-method, may be explained by the numerical noise induced by the cells discretization or by the nonlinearity of the ray-path deflection.

Since the $v\&r$-method is computationally very expensive, the code was parallelized to take advantage of multi-core architectures of modern computers. In general, for one iteration with the $v\&r$-method, 15 iterations with the v-method can be calculated in the same time. However, often the v-method did not reach a model misfit minimum even after 30 iterations.

In conclusion, the $v\&r$-method is computationally more expensive, numerically more complex and needs smoothing to reduce the noise in the solution at low altitude, on the other hand, it is faster as compared to the v-method.

Real data inversion

The v- and the $v\&r$-methods were applied successfully to real data collected with the OTH radar Nostradamus. Two data sets derived from spring 2006 in the early evening and from of winter 2008 at noon were tested. The inversion provided a map of electron density perturbation in the selected azimuth for the given day and time. In the first data set, only electron density perturbations in the F-layer are visible, the second data set also presents an E-layer.

Based on maps like these, electron density profiles from an *a priori* ionospheric model can be corrected in any desired distance within the range of the radar. This is an enormous improvement compared to, for instance, vertical profiles calculated on the base of backscattered ionogram inversions, which allows only the calculation of the electron density in the middle of the ray-path and its validity for all other distances is assumed or obtained by interpolation.

For the first time, a reconstruction of the real electron density distribution in the ionosphere was obtained from OTH radar tomography, revealing vertical irregularities in the region between 200 and 300 km height. Performing the inversion in different azimuths showed that these irregularities are also visible in other directions at the same day. Additionally, this first test illustrated the impressive potential of the ionospheric tomography method to reconstruct the electron density over the entity of Europe using Nostradamus.

In a preliminary synthetic test, for validation of the inversion results vertical profiles of electron density obtained after the first iteration were compared with those obtained from the ionosonde in Chilton, U.K. The *v&r*-method corrects the profile from the *a priori* ionospheric model towards the profile derived from the ionosonde, but additional iterations are necessary to reduce the difference between them.

However, the Sun underwent a solar minimum in 2006 and 2008 and as a consequence, solar flux and ionization in the ionosphere were weak. Therefore, less rays are reflected in the *a priori* ionospheric model in the ray tracing, and the number of rays is reduced further by using the frequency and elevation angles of Nostradamus measurements. That makes the inverse problem more ill-posed and non-unique compared to the inversion of synthetic data.

A second preliminary synthetic test showed that the tomography obtained by OTH radar can be combined with the tomography by GPS and both two methods can complement each other. The tomography with GPS is mainly sensitive to the region of maximum ionization (F2 layer) because of the high signal frequency. On the other hand, reconstruction of the electron density in the lower ionosphere up to 300 km altitude by the OTH radar is quite well.

8.2 Perspectives

Some remaining issues retain open, they are interesting subjects of future work. The ionospheric tomography method presented here was developed for monostatic OTH radar. These kind of radars can be found also in the USA (ROTHR), Australia (JORN), and the North-Pole (SuperDARN), and the method can be applied to previously collected data from the French radar Nostradamus. If the resolution of

other OTH radars is sufficient and they are able to measure the polarization of a wave, the effect of the magnetic field need to be included in the ray tracing to distinguish between ordinary and extraordinary waves.

Moreover, this tomography method is not limited to monostatic OTH radars and can easily be adapted to bistatic OTH radars. Although here the ray-path deflection needs to be considered, because the signal is still backscattered at the ground, at an unknown point, and returns to the receiver along another way through the ionosphere.

This tomography approach can be improved further by including other ionospheric sounding techniques, in particular TEC measured by GPS at the ground with dense arrays or by occultation with on-boarded GPS receivers. The latter idea is of particular interest, because both methods complement each other: While OTH ionospheric tomography is sensitive to the lower ionosphere (up to 300 km altitude) during daytime, GPS ionospheric tomography is only sensitive to the region of maximum ionization ($\sim$ 300 km) due to the high frequency of the signal. In addition, TEC GPS data can compensate for the lack of reflected HF rays during night. The feasibility of such a joint inversion was demonstrated in a preliminary synthetic test (Section 7.3.2).

In previous works, GPS ionospheric tomography was already combined with other sounding techniques, such as ionosonde data and backscattered ionogram inversions (see Chapter 4), but combination with an OTH radar tomography method has not yet been reported. This opens exciting new perspectives, improving the vertical resolution and supplying more information on electron density at the lower ionosphere.

In this thesis, the developed ionospheric tomography method was applied successfully to real data from OTH radar Nostradamus (Chapter 7), showing the potential to obtain 3D maps of electron density throughout Europe. This tomographic approach can be developed further to four dimensions, including the time and providing the evolution of electron density during day and night or over a full year.

The obtained vertical profiles of electron density need to be validated by ionosonde data in Europe or by other inversion methods. The validation with the data of the vertical sounder in Chilton has been tested, but needs further development with additional iterations.

The developed iterative approach showed the general improvement of the solution, but also an instability of the system after a critical number of iterations. This problem could be explained by the numerical noise introduced by the discretization of the model, it can be addressed by a more non-linear inversion method, for instance the CGM (Section 4.6.5).

Acknowledgments

In the following, I want to thank some people for supervision and support, both technically or morally, through the many ups and downs but nevertheless exciting and informative last three years. First of all, I thank the **ONERA** for giving me the possibility and means to realize this thesis as well as the **Planetary and Space Sciences group** at the IPGP.

I express my sincere gratitude to **Giovanni Occhipinti** for giving me the chance to continue his work in such an exciting field. Our lively discussions have always led to new insights and ideas and I acknowledge his always positive and encouraging attitude. Additionally, I thank him for the possibility to participate in numerous conferences, as well as his support regarding invited seminars.

I thank **Gilbert Auffray** for his warm welcome in his research team at the ONERA, his care and help to bring my thesis to a good direction. Many thanks to all my **ONERA colleges of the RBF team**. Thanks to them, I enjoyed working at the ONERA. Without their friendship, help with my thesis, in administrative or technical problems, and their humor in everyday life, this thesis would have been much more difficult to finalize. In this context, it was a pleasure for me sharing an office with **Nicolas, Jean-François** and later with **Marie-José**, giving the distraction I needed every now and then.

I deeply thank **Jean-Philippe Molinié** for supervising me continuously during the past three years. His ideas and suggested solutions have significantly contributed to the success of this work and I appreciate the informative and, due to his immense knowledge, very helpful discussions. I also appreciate his availability to listen to any of my personal problems, his patience especially with my explanations in french, and his advice.

Moreover, I thank **Mark Wieczorek** for having been my official thesis director till the end, as well as for his support, advise and helpful suggestions during my thesis.

A special thank goes to **Lapo Boschi** for his help in developing the conjugate gradient method and his endless and patient explanations. He always took the time to answer my questions, even outside office-hours, and his e-mails and photo expla-

nations helped me a lot.

Further thanks goes to **Véronique Rannou** for fruitful discussions on real data inversions and all issues concerning the ionosphere. I thank **Pierdavide Coïsson** for his explanations of the ionosphere, the discussions he had, and for providing me a (genius !!) LaTeX thesis template.

Last but not least, I am grateful to everybody who generously offered me help during my quest of Ph.D degree and to all people who spent time in reading my manuscript for improvement and error search, in particular my **mother**, **Veronique, Pierdavide, Valérie and Alain**.

Finally, I thank my **parents and sisters** for their encouragement during bad times, and their support in any situation when I needed them. Without them, I would never have finished this thesis. They gave me the strength to go further when I doubted in my possibilities.

List of Figures

List of Tables

Appendices

Appendix A

Article: New constraints on the 3D shear wave velocity structure of the upper mantle underneath Southern Scandinavia revealed from non-linear tomography

I strongly contributed to this work as a student assistant at the Geophysical Institute at the Karlsruhe Institute of Technology (KIT) in Germany. The collaboration continued during my Ph.D thesis at the IPGP, and the article was finished and published in January 2013.

Abstract

In this work, the 3D shear-wave velocity underneath Southern Scandinavia was revealed from seismic tomography, using the MAntle investiGations of Norwegian Uplift Structures (MAGNUS) seismic station network in Norway. The travel time residuals are corrected for the known crustal structure of Southern Norway and weighted to account for data quality and pick uncertainties.

The resulting residual pattern of subvertically incident waves is very uniform and simple. It shows delayed arrivals underneath Southern Norway compared to fast arrivals underneath the Oslo Graben and the Baltic Shield. The 3D upper mantle v_s structure underneath the station network is determined by performing non-linear travel time tomography. As expected from the residual pattern the resulting tomographic model shows a simple and continuous v_S perturbation pattern: a negative v_S anomaly is visible underneath Southern Norway relative to the Baltic Shield in the east with a contrast of up to 4 % v_S and a sharp W-E dipping transition zone. Reconstruction tests reveal besides vertical smearing a good lateral reconstruction of the dipping v_S transition zone and suggest that a deep-seated anomaly at depth is real and not an inversion artifact.

The upper part of the reduced v_S anomaly underneath Southern Norway (down to 250 km depth) might be due to an increase in lithospheric thickness from the Caledonian Southern Scandes in the west towards the Proterozoic Baltic Shield in Sweden in the east. The deeper-seated negative v_S anomaly (300 - 410 km depth) could be caused by a temperature anomaly possibly combined with effects due to fluids or hydrous minerals. The determined simple 3D v_S structure underneath Southern Scandinavia indicates that mantle processes might influence and contribute to a Neogene uplift of Southern Norway

Tectonophysics 602 (2013) 38–54

Contents lists available at SciVerse ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

New constraints on the 3D shear wave velocity structure of the upper mantle underneath Southern Scandinavia revealed from non-linear tomography

B. Wawerzinek *, J.R.R. Ritter, C. Roy [1]

Karlsruhe Institute of Technology, Geophysical Institute, Hertzstr. 16, 76187 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 25 May 2012
Received in revised form 2 December 2012
Accepted 23 December 2012
Available online 12 January 2013

Keywords:
Mantle processes
Seismic tomography
Shear waves
Relative travel time residuals
Southern Scandinavia

ABSTRACT

We analyse travel times of shear waves, which were recorded at the MAGNUS network, to determine the 3D shear wave velocity (v_S) structure underneath Southern Scandinavia. The travel time residuals are corrected for the known crustal structure of Southern Norway and weighted to account for data quality and pick uncertainties. The resulting residual pattern of subvertically incident waves is very uniform and simple. It shows delayed arrivals underneath Southern Norway compared to fast arrivals underneath the Oslo Graben and the Baltic Shield. The 3D upper mantle v_S structure underneath the station network is determined by performing non-linear travel time tomography. As expected from the residual pattern the resulting tomographic model shows a simple and continuous v_S perturbation pattern: a negative v_S anomaly is visible underneath Southern Norway relative to the Baltic Shield in the east with a contrast of up to 4% v_S and a sharp W–E dipping transition zone. Reconstruction tests reveal besides vertical smearing a good lateral reconstruction of the dipping v_S transition zone and suggest that a deep-seated anomaly at 330–410 km depth is real and not an inversion artefact. The upper part of the reduced v_S anomaly underneath Southern Norway (down to 250 km depth) might be due to an increase in lithospheric thickness from the Caledonian Southern Scandes in the west towards the Proterozoic Baltic Shield in Sweden in the east. The deeper-seated negative v_S anomaly (330–410 km depth) could be caused by a temperature anomaly possibly combined with effects due to fluids or hydrous minerals. The determined simple 3D v_S structure underneath Southern Scandinavia indicates that mantle processes might influence and contribute to a Neogene uplift of Southern Norway.

1. Introduction

At the western edge of the Baltic Shield, Northern Europe, the Scandinavian Mountains (Scandes) form an about 1400 km long mountain range (Fig. 1). The Scandes are the second largest mountain range in Europe and extend from 59°N to 63°N. The present high topography of the Scandes is still under debate (Lidmar-Bergström and Bonow, 2009; Nielsen et al., 2009). The main mountain building phase occurred during the Scandinavian Caledonian orogeny 440–420 Ma ago when Baltica and Laurentia collided (Torsvik and Cocks, 2005). Today the mountains still reach maximum heights of 2.5 km and the Southern Scandes in Norway have roughly a dome-like shape (Fig. 1). This dome has a high-level low-relief landscape with a base at about 1000 m altitude on which the mountain massifs are sited (Lidmar-Bergström and Bonow, 2009; Lidmar-Bergström et al., 2000). There are different hypotheses which try to explain the present high topography of the Scandes far away from current plate boundaries and in the absence of active compressional tectonics (Cloetingh et al., 2007). Some studies come to the conclusion that there were later uplift phases after the Caledonian orogeny, for example during the Neogene (e.g. Lidmar-Bergström et al., 2000; Rohrman and van der Beek, 1996; Smelror et al., 2007). Japsen and Chalmers (2000) and Anell et al. (2009) summarise widespread high topography areas, Cenozoic uplift and their geodynamic driving forces around the North Atlantic. Another point of view is that the Caledonides were formed during one main orogenic phase and that the present topography is a result of the interaction between erosion, climate and ice (Nielsen et al., 2009). However, this hypothesis is in dispute (Gabrielsen et al., 2010; Lidmar-Bergström and Bonow, 2009 and references therein).

The debate on this controversial issue is mainly due to sparse and insufficient data which limit our knowledge on the deep structure of the Scandes and thus prevent the determination of a consistent geodynamic model. However, recent seismological experiments improved this situation. Using receiver functions to map the Moho, Svenningsen et al. (2007) found a 10–12 km thick crustal root underneath the Southern Scandes and claimed this root as Airy-type compensation for the mountains. Stratford and Thybo (2011a) reviewed existing crustal models, presented the results from recently measured seismic refraction lines (MAGNUS-REX project) and compiled a Moho

* Corresponding author at: Leibniz Institute for Applied Geophysics, Stilleweg 2, 30655 Hannover, Germany. Tel.: +49 511 643 3389.
E-mail address: britta.wawerzinek@liag-hannover.de (B. Wawerzinek).
[1] Now at: Institut de Physique du Globe de Paris, Université Paris Diderot, 5 rue Thomas Mann, 75205 Paris Cedex 13, France.

0040-1951/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.tecto.2012.12.033

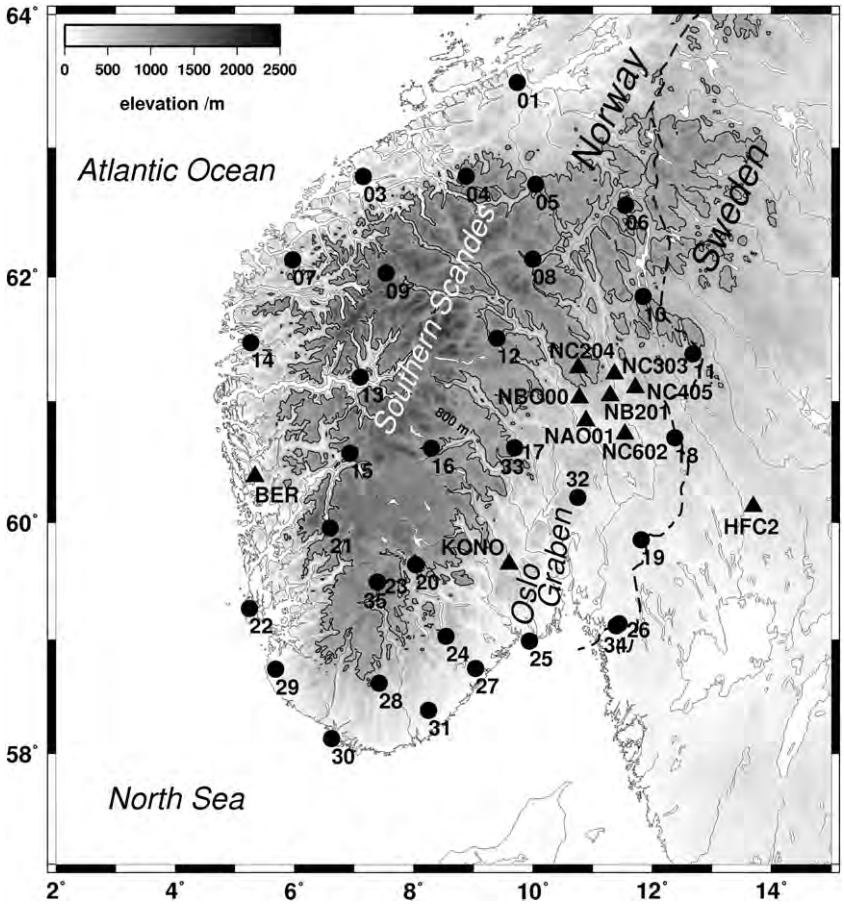


Fig. 1. Map of the study region and the MAGNUS station network. Numbers correspond to mobile stations (circles, the leading NWG station code is omitted for clarity), while permanent stations (triangles) are labelled with complete station code. The elevation is colour-coded and the 800 m contour line is plotted to visualise the dome-like shape of the Southern Scandes. For global context of the study area see Fig. 2.

map. Their model has a thinner crustal root and its 40 km thick crust is regarded as typical shield-type crust. Stratford and Thybo (2011a) interpreted the observed 60 km lateral offset between the deepest Moho topography and the highest surface topography by a possible lithospheric flexure. The increased seismic velocities (up to 5%) around the Palaeozoic Oslo Graben, which were determined from the MAGNUS-REX data, are explained by mafic intrusions (Stratford and Thybo, 2011b).

Below the 22 NORSAR (Fig. 1) subarrays, Aki et al. (1977) found a relatively low seismic velocity region to the west compared to a higher seismic velocity region to the east. This difference in seismic P-wave velocity (v_P) is interpreted as Baltic Shield material in the east, which is undisturbed by the Caledonian orogeny, and the rifting of the Oslo Graben or its related volcanism. Medhus et al. (2009) determined travel time residuals of teleseismic P-waves which were recorded during field experiments and at permanent stations in western Scandinavia. Their residual pattern of late arrivals indicates reduced seismic velocity in the upper mantle below the Southern Scandes and the Norwegian-Danish Basin, whereas fast arrivals are observed towards east at the Oslo Graben and into Sweden (Medhus et al., 2009). Two teleseismic tomography studies by Medhus et al. (2012) present relative and absolute P-wave velocity perturbations in the upper mantle below the Southern Scandes and their surroundings: low v_P is recovered underneath the Southern Scandes and the Danish and German basins. The contrast to the faster upper mantle v_P under the Baltic Shield is up to 5% and its western boundary is clearly mapped (Medhus et al., 2012). Reduced S-wave velocity (v_S) is also known below the Scandes from surface wave analyses: Weidle and Maupin (2008) determined a low v_S anomaly reaching from Iceland across the eastern North Atlantic below southern Norway at about 70–150 km depth. Maupin (2011) confirmed the low v_S anomaly in the mantle underneath Southern Norway. The influence of these low seismic velocity anomalies on the topography on the Scandes is still unclear. Although dynamic topography may play a role for generation of high elevation (for a review see Braun, 2010), its role in Scandinavia is disputed. Pascal and Olesen (2009) demonstrated with integrated gravity and thermal modelling that Cenozoic uplift of the Scandes cannot be explained only by an asthenospheric diapir.

In the above context we want to address the following questions: How does the mantle structure look like at the transition between the offshore and onshore Norwegian lithosphere including the Southern Scandes? How does the mantle structure change further east towards the Baltic Shield? Are there structural elements at depth which may help to describe the mountain building processes in the region and what is the lateral and depth extension of possible anomalies? The presence of seismic anomalies and interpretations concerning their origin help to explain the building process at depth. For example, a low velocity anomaly may be caused by increased temperature and relatively shallow asthenospheric material. Such knowledge can indicate that a fraction of the present high topography is due to dynamic topography (Braun, 2010). If no velocity anomalies are present underneath the Scandes a completely different mountain building process without active mantle processes would be the cause for the present topography. The idea of this work is to measure travel times of teleseismic shear waves recorded at the stations of the MAGNUS network and to determine travel time residuals relative to a laterally homogeneous standard Earth model, e.g. *iasp91* (Kennett and Engdahl, 1991). These residuals are inverted into 3D shear

wave velocity perturbations underneath the station network. Our inversion technique is very similar to the famous first seismic tomography by Aki et al. (1977) which was also conducted in Southern Norway.

2. The MAGNUS experiment

The MAGNUS experiment (MAntle investiGations of Norwegian Uplift Structures) was a broadband seismological field experiment to record teleseismic and local earthquakes across Southern Norway (Fig. 1). Thirty-one mobile broadband recording stations of the KArlsruhe BroadBand Array (KABBA) were deployed in and around the Southern Scandes between September 2006 and June 2008. The KABBA stations together with the permanent broadband stations KONO (Kongsberg), BER (Bergen), HFC2 (Hagfors) and 7 stations of the Norwegian Seismic Array (NORSAR) covered a nearly equally spaced network across the Southern Scandes. The MAGNUS dataset was integrated into the EUROCORE programme TOPO-EUROPE (Cloetingh et al., 2007) of the European Science Foundation and therein specifically into the Collaborative Research Project *TopoScandiaDeep* (http://www.mn.uio.no/geo/english/research/projects/toposcandiadeep/, state 29th November 2012) which concentrates on the deep structure of the Scandes. For details on MAGNUS see Weidle et al. (2010).

A relatively good azimuthal coverage of teleseismic earthquakes was achieved due to continuous recording with very few station failures. Seismic phases from 179 earthquakes were studied in order to determine travel time residuals across the MAGNUS network. 128 of these events (Fig. 2) with signal-to-noise ratios larger than 2 were suitable for the shear wave tomography, because seismic phases of these events could be clearly seen at most stations. The moment magnitude M_w of the earthquakes was mostly above 6. However, weaker events underneath Africa were specifically examined to fill observational gaps in the south.

3. Analysis of travel time residuals

3.1. Determination of relative, weighted, crust corrected travel time residuals

At first the raw waveforms are pre-processed including a removal of possible offsets and trends in the recordings. The individual instrument response is removed by deconvolution to obtain true amplitude velocity-proportional seismograms. Finally all seismograms are bandpass filtered from 8 s to 20 s (0.05–0.125 Hz). This frequency band is identified as most suitable, because the seismic phases can be seen best at most stations in this band. Generally, the signal-to-noise ratio (SNR) is better at inland stations compared to stations close to the Atlantic and North Sea coasts. In order to enhance the polarised shear waves a coordinate rotation is done which transforms the horizontal N–S and E–W recording components into transverse (T) and radial (R) components. The theoretical backazimuth to the epicentres is used for the rotation.

We determine travel time residuals for 128 teleseismic events which we use as input data for the tomography. Following Evans and Achauer (1993), we pick the first clearly visible phase (up or down in a seismogram) of a shear wave front across the whole network. The arrival time picking was done by hand at magnified waveforms (Douglas et al., 1997) of the direct S or SKS wave fronts. Both R and T component waveforms could be used from 79 events (direct S-waves), only T component waveforms from 28 events (direct S-waves) and only R component waveforms from 21 events (mainly SKS-waves). Altogether 8348 pick times are determined

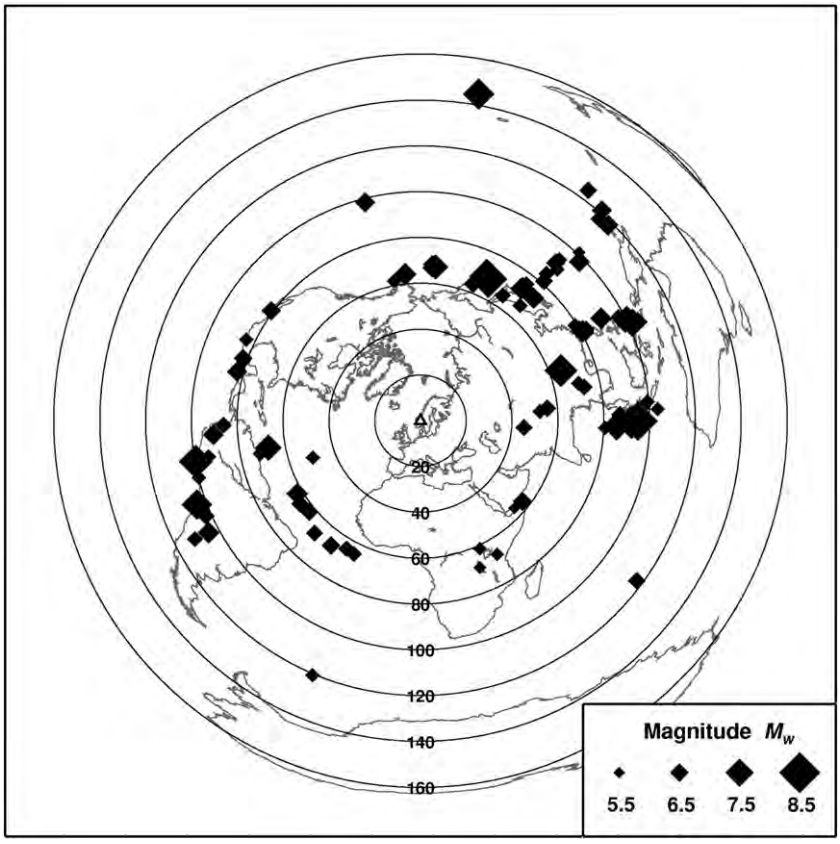


Fig. 2. Distribution of the analysed 128 earthquakes (diamond symbols) in dependence of their epicentral distance (in degree; radial axis) and backazimuth. The location of the MAGNUS network is indicated by a triangle in the centre. The size of the symbols indicates the moment magnitude M_w of the events. Hypocentral parameters are taken from the International Seismological Centre and the U.S. Geological Survey.

(Table 1). Depending on the shape of the S or SKS wavelet and its SNR, a quality factor is attributed to each pick. High-quality picks with a high SNR and symmetric waveform are "A"-class, less clear picks are "B"-class, acceptable picks are "C"-class and doubtful picks are "X"-class (see supplement by Kirschner et al., 2011 for data picking and quality issues). "X"-class picks are removed from the further analysis. In this way 6196 travel time data remain for the residual analysis, including 2859 T component picks and 3337 R component picks. The quality distribution is 942 "A", 3109 "B" and 2145 "C" class picks. If both component recordings, R and T, can be used for picking, only the picks of the component recording with the better data quality are kept and the picks of the other component recording are removed. The component selection (R or T) is done event wise by determining the better pick quality, e.g. the larger number of "A" picks, of the component recordings. This sorting reduces our dataset to 3999 high-quality travel time residuals including 757 "A", 2041 "B" and 1201 "C" class picks.

To account for station elevation and known variations of the gross crustal structure, we apply a 1D crustal travel time correction following Martin et al. (2005). This correction takes into account the travel time between sea level and true station height, sediment effects (here not necessary due to missing sediments below the MAGNUS stations) and Moho topography relative to the *iasp91* reference Earth model (Moho at 35 km depth; Kennett and Engdahl, 1991). Ray tracing is done with the theoretical slowness of the wave. The 1D crustal travel time correction is isotropic, so that travel times on R and T components are affected in the same way. For Norway the Moho topography was taken from Stratford et al. (2009) and it varies between 28 km depth (station NWG22 at the coast) and 40 km depth (stations NWG05 and NWG08 in the NE of the MAGNUS network). In Fig. 3 the crustal travel time corrections are shown for each station site based on an average slowness of 7 s/° for teleseismic shear waves. Along the coast teleseismic shear waves arrive about 0–0.5 s earlier relative to *iasp91* (35 km Moho depth; Kennett and Engdahl, 1991) due to the thinner crust (about 28–30 km). Inside the Southern Scandes the arrival time is delayed by up 0.4 s due to a thin crustal root and the high station elevation. Compared to the residuals described below, the crustal correction is about 20%–30% and is needed to avoid smearing of crustal travel time anomalies into the underlying mantle during the inversion.

The calculation of the residuals is done in the following standardised way (see also Evans and Achauer, 1993 or Kirschner et al., 2011): Hypocentral parameters including origin time were mostly taken from the International Seismological Centre, preferably updated Engdahl et al. (1998) solutions. The theoretical travel time is determined based on the *iasp91* reference Earth model (Kennett and Engdahl, 1991). The measured travel time is corrected for first order crustal travel time effects due to surface and Moho topography (Fig. 3). The difference between the measured and the theoretical travel times is weighted according to the quality classes of the picks (weights: "A" 1.0, "B" 0.5 and "C" 0.25), see Evans and Achauer (1993) or Kirschner et al. (2011). Finally, the average residual value of each event phase is subtracted to eliminate far-field effects such as hypocentre uncertainties and source side structural anomalies.

Table 1
Overview on number of determined S- and SKS-wave residuals; the overall analysed residuals are given on the left side; the residuals used for the tomographic inversion are given on the right side.

Phase	Comp.	No. overall A / B / C / X quality	No. for inversion A / B / C quality
S	R	242 / 945 / 614 / 685	126 / 362 / 198
S	T	594 / 1698 / 1045 / 1002	556 / 1365 / 712
SKS, SKKS	R	106 / 466 / 486 / 465	75 / 314 / 291
		Σ 942 / 3109 / 2145 / 2152	Σ 757 / 2041 / 1201
Sum		8348	3999

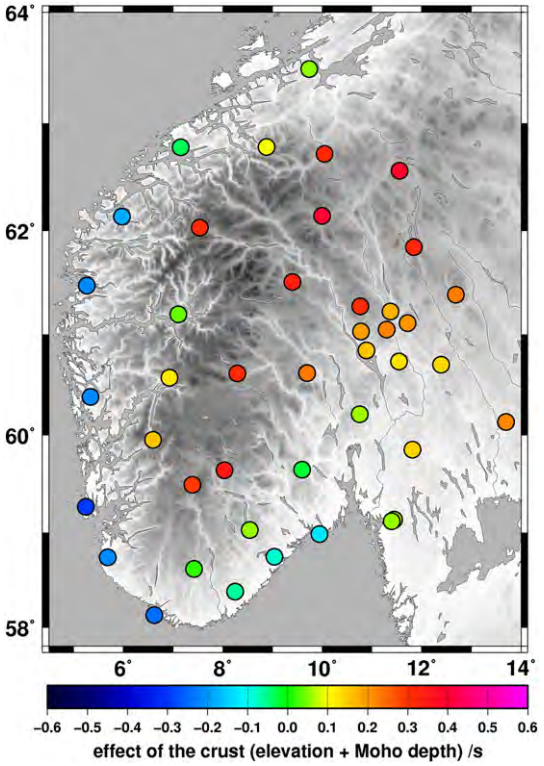


Fig. 3. Crustal travel time correction terms for an average teleseismic shear wave slowness of 7 s/°. The correction time is relative to the *iasp91* reference Earth model (Kennett and Engdahl, 1991) and includes station elevation (Weidle et al., 2010) and Moho topography (Stratford et al., 2009) effects on the travel time.

In doing so, the knowledge of the absolute velocities at depth is removed whereby only relative velocity changes will be relevant in the following. This procedure results in first-order crust corrected, weighted, relative travel time residuals, just called residuals in the following. These residuals are caused by seismic velocity anomalies underneath the station network and we use them as input data to invert for a 3D image of the shear wave velocity contrasts.

A comparison of the R and T component residual pattern at the stations shows that there is hardly any difference between both. Both components display a quite similar distribution of early and late arrival times (Fig. 4). E.g. at station NWG28, which is located in the southern part of the MAGNUS network (Fig. 1), S-waves from NW, N, NE, E and SE and SKS-waves are delayed (reddish diamonds) relative to S-waves from SW and W (bluish diamonds) both on R and T component recordings (Fig. 4). The only exception is found at stations NWG18, NWG32, NC602, NAO01 and NB201 for events in a backazimuthal range of 200° to 270° and with about 11–13 s/° slowness (epicentres at the South Atlantic ridge, see Fig. 2). These stations are all placed just north of the Oslo Graben (Fig. 1). There the 36 R component arrivals are faster compared to the 45 T component arrivals. This is the only observation of a dependence on polarisation on the residuals and hence a very weak indication of anisotropy in our entire residual dataset. However, all other measurements (approx. 98% of the dataset) do not show any difference between R and T component residuals. This overwhelming coincidence is astonishing, because a shear wave splitting analysis of SKS phases revealed clear shear wave splitting below the Southern Scandes (Roy and Ritter, in press). The SKS splitting, which is depth- and backazimuth dependent, must be related to a complex anisotropic structure at depth and seems to average out in our residual measurements. The reason for this difference between residuals measurements and SKS splitting is not understood yet, but also observed in the Eifel (Keyser et al., 2002). However, we infer that an isotropic travel time

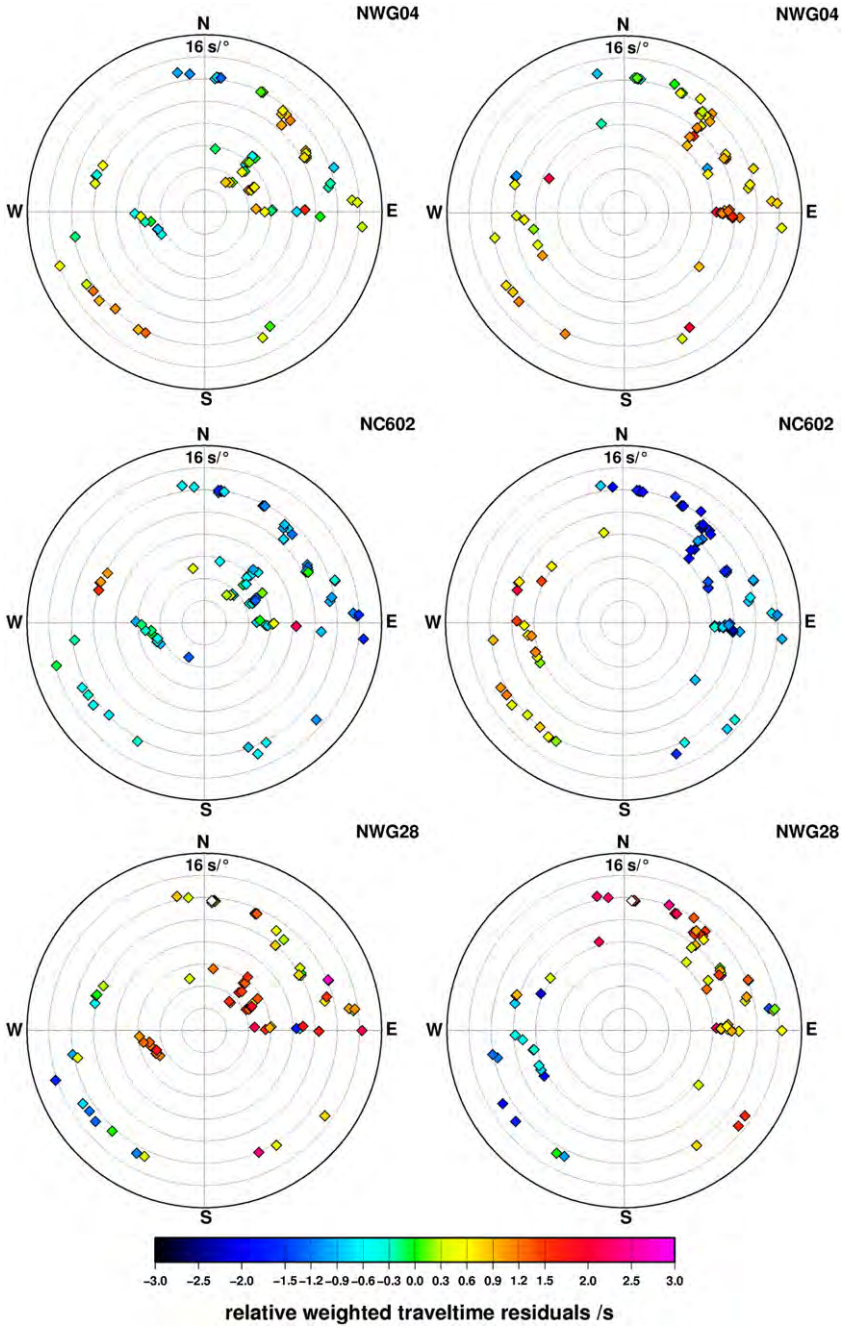


Fig. 4. Radial (left column) and transverse (right column) component residuals at selected stations of the MAGNUS network (Fig. 1). The azimuthal direction depends on the backazimuth of the events (Fig. 2). The radial axis represents slowness in 2 s/° intervals. The travel time residuals are relative to their average across the network, weighted by quality and crust corrected (see Section 3).

inversion for a 3D tomography can be justified in our case due to the prevailing isotropic residuals.

3.2. Residual pattern beneath the MAGNUS network

For the tomography we choose for each event phase the residuals of either the R or T component recordings depending on the better data quality which is estimated from the weighting procedure (see above). In this way only the most reliable residual measurements are used for the tomography. In Fig. 5 we summarise all these 3999 residuals (Table 1) in a spatial view. These are averaged per station and colour-coded per station (circles) as well as smoothly interpolated between the station sites. This residual plot displays very obvious large-scale trends in the residual pattern and indicates their related shear wave velocity anomalies (Δv_S) at depth: waves arriving from the NE or SE (epicentres in the NW Pacific or Asia, Fig. 2) arrive earlier at stations east of ~10°E on the Baltic Shield (Fig. 5a and b) whereas waves from SW (South American epicentres) arrive mostly delayed at the western Baltic Shield (Fig. 5c). The four stations NWG19, NWG26, NWG34 and HFC2 in the SE of the MAGNUS network show only fast arrivals, with the exception of NE incident waves that propagate underneath the Southern Scandes.

In the centre of the MAGNUS network, inside the Southern Scandes, mainly delayed arrivals are measured (Fig. 5). Since first order travel time effects due to surface and Moho topography are already excluded, mantle anomalies are the main reason for the displayed residuals assuming that errors in the crustal model as well as lateral heterogeneities in the crust cause only second order

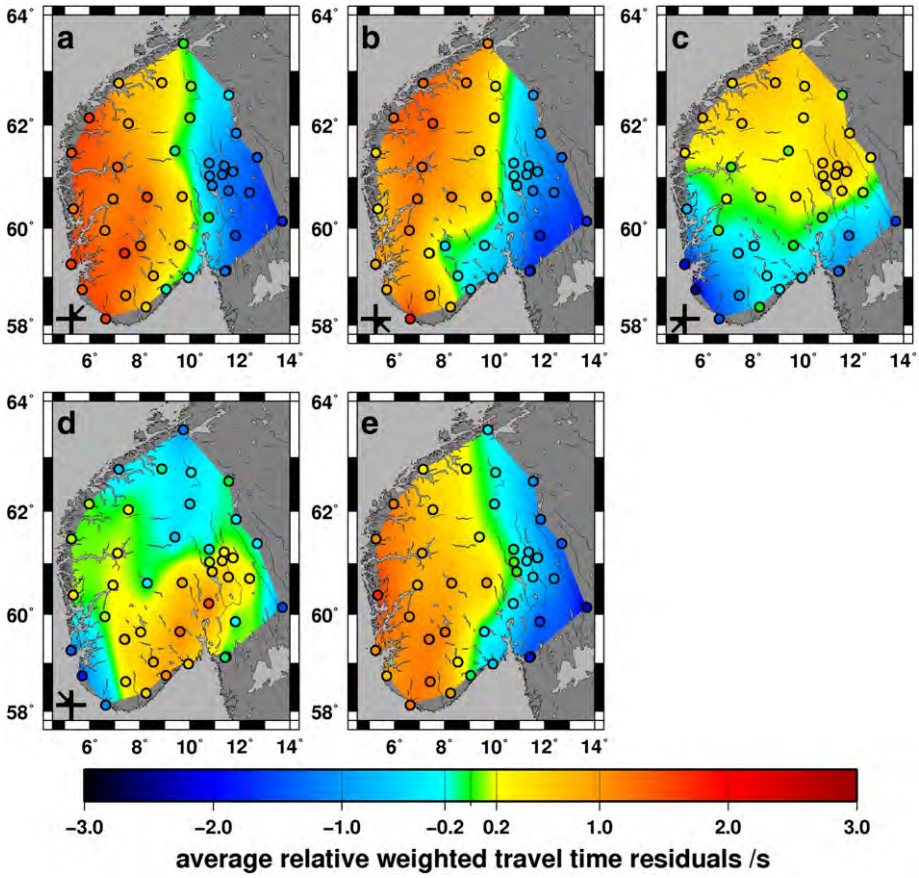


Fig. 5. Spatial distribution of direction-dependent residuals. The travel time residuals are relative to their average across the network, weighted by quality and crust corrected (see Section 3). a) Average residuals for events from NE backazimuth (see arrow in lower left corner), b) average residuals for events from SE backazimuth, c) average residuals for events from SW backazimuth, d) average residuals for events from NW backazimuth and e) average residuals for events with steeply incident core phases.

travel time effects. In the south of the network, residuals are delayed if the waves propagated from N and NE and they arrive earlier for waves approaching from SW. The stations in the NW of the network (NWG03, NWG04, NWG07 and NWG09) have mostly delayed arrivals with the biggest delays for waves propagating from SE (Fig. 5b).

The overall residual pattern in Fig. 5 has one major common feature: waves propagating underneath the Southern Scandes are always delayed whereas waves travelling underneath the Baltic Shield arrive early. Along the west coast of Southern Norway waves arriving from the Atlantic side are faster at the southern stations compared to the northern stations. The backazimuthal direction has a greater influence on the residuals compared to the slowness (or incidence angle) as can be seen for example in Fig. 4.

The transition between slow seismic velocity at depth in the west and fast seismic velocity at depth in the east can be visualised best by plotting only residuals of steeply arriving SKS phases (Fig. 5e). The maximum residual contrast reaches nearly 4 s. The change between fast and delayed arrivals occurs at 10°E longitude in Southern Norway. This change is not very sharp but appears as transition between about 9°E and 11°E in a 100 km wide zone. The biggest SKS delays do not coincide with the highest surface topography of the Southern Scandes. These are found towards SW Norway at the North Sea coast. The area around the Oslo Graben is characterised by fast arrivals relative to the Southern Scandes. On the Baltic Shield the SKS phases arrive increasingly earlier towards east (Fig. 5e).

The scatter of the picked arrival times can be assessed in Fig. 4. The scatter is mainly due to picking uncertainties and timing problems at the station. Recordings with known timing problems due to missing GPS time synchronisation had been excluded from this analysis. The remaining data scatter can be assessed by comparing residuals from epicentre clusters, e.g. Sumatra, Japan or South America. The residuals from such clusters show similar values in Fig. 4 and their scatter is mostly less than 0.1 s. Larger residuals variations may be related to identified picking uncertainties which are treated by weighting the data quality (see Sections 3.1 and 4.1). Thus picks with possible larger uncertainties are down-weighted during the inversion.

Based on the residual pattern we expect the following shear wave velocity anomalies Δv_S in the upper mantle: positive Δv_S in the east of the MAGNUS network and hence faster v_S underneath the Baltic Shield in contrast to negative Δv_S in the centre of the network and hence a pronounced low v_S anomaly underneath the Southern Scandes. Towards south and southwest of the network negative Δv_S are expected. Compared to previous residual datasets that we analysed in Europe and Africa our residuals at the MAGNUS network in Norway display a very simple low-velocity anomaly underneath its centre.

3.3. Residual pattern of former studies in Norway and adjacent areas

Our S-wave residuals can be compared to published P-wave residuals. Using teleseismic P-waves, Berteussen (1975) found relative residuals of ±0.8 s across the NORSAR stations and a general trend for late arrivals in the west and early arrivals in the east. He also demonstrates that a variation in Moho depth alone can only explain about 20% of these observations and would require an increase of more than 10 km in Moho depth towards west (that is thicker crust below the south-eastern part of the Southern Scandes relative to the Baltic Shield in the border region between Norway and Sweden). Thus Berteussen (1975) realised that more complicated seismic models were necessary to explain the residuals and Aki et al. (1977)

presented such a 3D model in their pioneering work on seismic tomography.

The ±0.8 s P-wave residual by Berteussen (1975) corresponds to roughly ±1.4 s for S-wave residuals based on a v_P/v_S of about 1.73. We find weaker residual changes (<1 s) across the NORSAR stations (Fig. 5) which might be caused by our computation of weighted residuals or due to a reduced v_P/v_S in Southern Norway. Waves propagating from SE to NW and nearly vertically incident SKS-waves show residual changes of only 0.6–0.75 s across the NORSAR stations. We observe later arrivals at the western NORSAR stations compared to the eastern NORSAR stations for vertically incident waves (Fig. 5e) as well as delayed arrivals at the north-western NORSAR stations relative to the south-eastern NORSAR stations for waves propagating from SE to NW (Fig. 5b). In contrast, waves with incidence from NW to SE show faster arrivals (approx. 1 s) at northern and western NORSAR stations compared to the southern and eastern NORSAR stations (Fig. 5d). However, no significant residual variations (<0.3 s) between the NORSAR stations are found from waves propagating from NE to SW and vice versa (Fig. 5a, c). The residual variations of nearly vertically incident waves, which are caused by upper mantle velocity perturbations directly underneath the station network, show the same residual distribution pattern as Berteussen (1975): late arrivals in the west and early arrivals in the east of the NORSAR network. However the residual variation across the NORSAR stations is smaller (0.6 s) compared to the observation of Berteussen (1975).

Analysing teleseismic P-waves, Medhus et al. (2009) found delayed arrivals underneath Southern Norway compared to the Baltic Shield in the east. The sharp transition between delayed and faster arrivals appears in the area of the Oslo Graben and the residual variations are smaller than 1 s. Medhus et al. (2009) assumed that a change in lithospheric thickness might cause the observed residual pattern. According to this idea, an increase in lithospheric thickness towards the Baltic Shield and therefore faster velocities underneath the Baltic Shield compared Southern Norway might lead to faster arrivals in the area of the Baltic Shield and delayed arrivals in Southern Norway. The residual pattern of Medhus et al. (2009) is in good agreement with our S-wave residual pattern (Fig. 5), although the transition zone between faster and delayed arrivals is located further west in our S-wave residual pattern compared to the P-wave residual pattern of Medhus et al. (2009).

A comparison to other travel time analyses in the region shows related relative residual patterns: Shomali et al. (2006) analysed P- and S-wave travel time data of the TOR experiment and found a residual pattern linked to the Trans European Suture Zone (TESZ): South of TESZ the residuals are positive (late arrivals) whereas negative residuals (early arrivals) are observed north of the TESZ. Eken et al. (2007) analysed P-wave travel times which were recorded at the Swedish National Seismological Network (SNSN). Their pattern of averaged relative residuals illustrates early wave arrivals in the southern part of the network compared to later wave arrivals in the northern part of the network (Eken et al., 2007) and their resulting mean P-wave velocity pattern illustrates higher seismic velocity in the southern part of Sweden compared to northern Sweden. Eken et al. (2007) presume that the change of the sign of the travel time residuals coincides with the Archean–Proterozoic boundary. Although these travel time residuals are not directly comparable to our study (different networks and mean values), they show that early wave arrivals are observed on the Baltic Shield in southern Sweden.

4. Travel time tomography

4.1. Inversion method

To compute the teleseismic travel time tomography model and hence determine the 3D upper mantle v_S structure, we use the non-linear JI-3D inversion method (Jordan, 2003). JI-3D is a further development of the ACH-type inversion scheme (Aki et al., 1977). For a detailed description of the JI-3D inversion method see Jordan (2003) or Wawerzinek et al. (2008).

The model space is composed of layers which are subdivided into variably wide blocks to take into account the station distribution, the ray density and criss-crossing rays (Jordan, 2003). In the centre of the model the block size corresponds roughly to the station distance whereas the block size is enlarged at the model edges where only few criss-crossing rays are observed (Fig. 6a). Thereby, the values of the diagonal elements of the resolution matrix are almost the same which leads to a stable inversion result (Jordan, 2003). The ray paths are computed iteratively by using a 3D ray tracer (Steck and Prothero, 1991). To include a priori information, e.g. uncertainties of the residuals or the allowed variation of the individual model parameters, a Bayes algorithm is applied (Zeyen and Achauer, 1997). To account for uncertainties of onset picking and therefore residuals picks of quality class “C” are weighted as 0.25, “B”-picks as 0.5 and “A”-picks as 1. Comparable to model damping the admitted model parameter variation is set variably for each block (Jordan, 2003). For the starting model the admitted parameter variation is determined by performing a trade-off analysis, during iterative inversion the admitted model parameter variation is computed by the inversion code (Jordan, 2003). To account for crustal 3D travel time effects, which are not eliminated by applying the 1D crustal correction (see Section 3), the model parameters of the crustal blocks are not totally fixed but only small variations are allowed (Martin et al., 2005).

The crust and upper mantle parameterisation underneath our study area (Table 2) is subdivided into 9 layers down to 500 km depth (Fig. 6b) and 272 blocks with a maximum horizontal model space extension of 1200 km N–S and 1200 km E–W. The vertical model extension is chosen in dependence of the network extension and the observed criss-crossing rays. In addition, the relative P-wave tomography study of Medhus et al. (2012) image no significant seismic velocity contrast underneath 400 km depth. Therefore, the vertical model extension was set to 410 km depth to image the complete upper mantle velocity structure. The upper first and second layers contain the crust and account for station site effects (first layer) and not yet corrected 3D crustal effects (second layer). The deepest layer (410–500 km) is required to absorb effects of underlying shear wave velocity anomalies or heterogeneities (e.g. Evans and Achauer, 1993). Thereby the absorbing effect is dependent on the spatial resolution in the overlying layers. However, these three layers (the two crustal layers and the deepest layer) will not be interpreted later on, because they are mainly used to absorb inversion artefacts. The interpretation will only include the upper mantle structure from 35 km to 410 km depth (layers 3–8, Table 2).

The starting model comprises the average v_S of the *iasp91* Earth reference model (Kennett and Engdahl, 1991) as background velocities. To account for non-linear effects the inversion is performed iteratively and we achieve a stable result by computing two iterations which leads to a residual variance reduction of almost 63% which means that the data are underfitted. The offset and averaging technique (Evans and Zucca, 1988) is performed to reduce the influence of the block parameterisation. Thereby the parameter grid is shifted along the block edges with an offset of 1/3 of the smallest block size and for each of these nine parameter grids an individual inversion is performed. To obtain the final inversion result the individual nine inversion results are averaged afterwards (Evans and Achauer, 1993).

4.2. Inversion results

Our inversion results are presented as horizontal (Fig. 7) and vertical (Figs. 8 and 9) cross sections through the model with Δv_S anomalies. Following the description of the model we discuss its significance based on reconstruction tests. As we use relative residuals across the station network with a mean value of zero, no knowledge

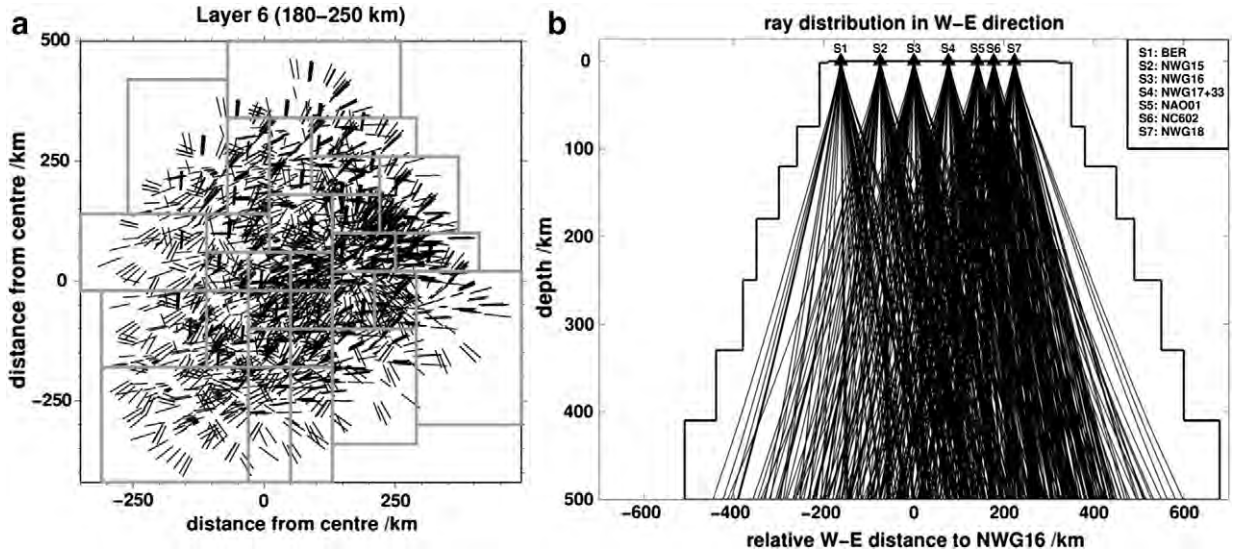


Fig. 6. Ray distribution and parameterisation of the tomography model. a) Horizontal cross section through the model layer at 180–250 km depth with ray path segments and the block parameterisation (Wawerzinek, 2012). b) Vertical W–E cross section between 60°N and 61°N with ray geometry to the seismic stations and the vertical layer parameterisation.

of absolute velocities can be given, therefore only velocity contrasts between the anomalies are meaningful. Hence, absolute velocity perturbations cannot be determined but these are mostly in the same percentage range when compared with absolute velocity models. Five major Δv_S anomalies can be identified in the tomography model:

I. The shallow mantle (<180 km depth) along the Norwegian coast has negative Δv_S of about −2% compared to the continental region (35–180 km depth slices in Fig. 7 and W–E cross sections in Fig. 9).
II. At shallow depth (35–75 km depth slice in Fig. 7) the mantle underneath the Oslo Graben area has positive Δv_S of about 1.5% relative to the surrounding lithosphere.
III. The mantle below the Baltic Shield has positive Δv_S compared to the mantle underneath the Southern Scandes; this boundary is not vertical but dips underneath the Baltic Shield between 35 km and 180 km depth (W–E cross sections in Fig. 9).
IV. Below the Southern Scandes there is no specific v_S anomaly in the upper 180 km depth (35–180 km in Fig. 7 and cross sections in Fig. 9) besides the general contrast to the Baltic Shield lithosphere which is characterised by higher v_S.
V. In the lowermost upper mantle, at about 250–410 km depth, there is a clear low v_S anomaly with a high contrast in v_S of maximum 3% relative to the mantle underneath the Atlantic towards SW and the Baltic Shield in the east.

Table 2
Parameterisation of the inversion model; the used background shear wave velocities v_S are taken from *iasp91* (Kennett and Engdahl, 1991).

Layer	Depth range in km	Node depth in km	v_S background in km/s
1	0–2	0	3.36
2	2–35	34	3.75
3	35–75	36	4.47
4	75–120	95	4.49
5	120–180	150	4.51
6	180–250	215	4.53
7	250–330	290	4.66
8	330–410	409	4.87
9	410–500	411	5.07
Half space			5.26

4.3. Reconstruction of the inversion

In order to access the reliability of the inversion results we first compare the spatial distribution of the measured residuals with the anomalies in the inversion model. Then we evaluate numerical reconstruction tests in order to test the confidence and limitations of our 3D model as well as the smearing of anomalies along the ray paths.

A comparison of the inversion result (Fig. 7) with the distribution of residuals of nearly vertically incident waves (Fig. 5e) reveals a very good correlation at 120–410 km depth: The inversion result shows negative Δv_S (reduced v_S) in the area of delayed arrivals (Southern Norway) relative to positive Δv_S (increased v_S) in the area of fast arrivals (Baltic Shield). The transition between delayed and fast arrivals coincides well with the transition zone between slower and faster velocity at 180–250 km depth. Thus, the inversion result visualises a simple v_S distribution as expected by analysing the distribution of the travel time residuals.

Four numerical reconstruction tests are performed in the following way: first a 3D model with synthetic Δv_S anomalies is constructed. 3D ray tracing is used to calculate travel times in these models with the same ray distribution as for the real MAGNUS tomography. Afterwards synthetic residuals are determined. Then the synthetic residuals are inverted with the same parameterisation as the real model in order to estimate the resolution properties of the input anomalies. We display the results of a checkerboard test (CBT in Fig. 10) and three anomaly tests (AT1, AT2 and AT3 in Figs. 11, 12 and Supp. Fig. S4, respectively). Each reconstruction test displays the inversion result after applying the offset and averaging technique (Evans and Zucca, 1988) as interpolated velocity pattern. Additionally, the inversion results of the nine individual inversion models are displayed as circles in the depth slices in Figs. 10–12.

The input velocity model of the checkerboard test (CBT, Fig. 10) comprises alternating anomalies with v_S amplitudes of +1.5% (blue outlines) and −1.5% (red outlines) at 75–125 km and 180–250 km depth. The synthetic input anomalies have extensions of 100 km by 100 km at 75–120 km depth and 200 km by 200 km at 180–250 km depth. The inversion result of the CBT (Fig. 10) recovers the input velocity anomalies but also illustrates the vertical smearing both upwards and downwards. At first, we note that in some model areas the single inversion results (circles) reproduce the synthetic anomalies

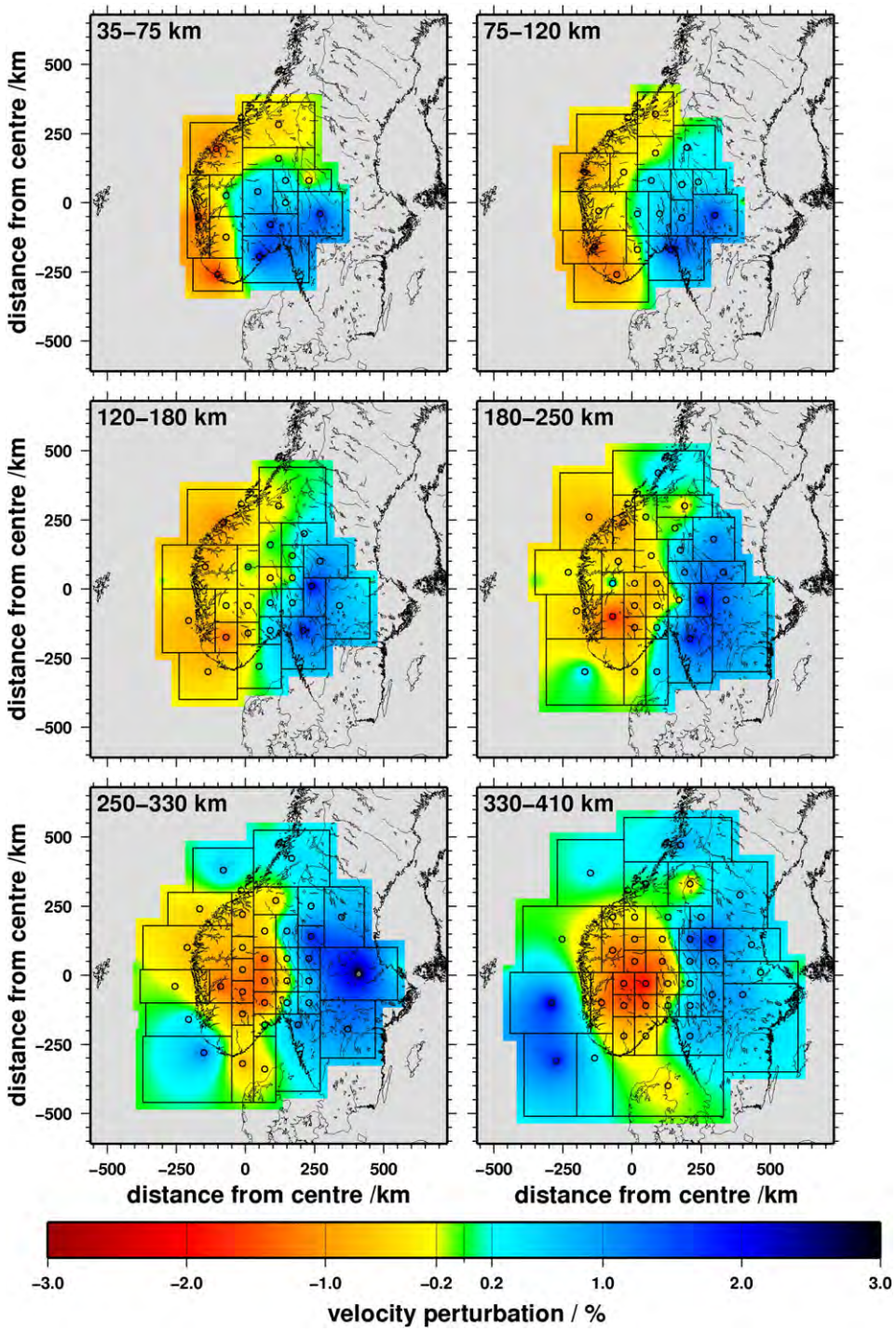


Fig. 7. Horizontal slices through the inversion result after applying the offset and averaging technique. The S-wave velocity contrast is colour-coded and the model parameterisation is indicated by block boundaries and node positions (circles). The S-wave velocity is reduced underneath Southern Norway compared to its surroundings especially to the Baltic Shield in the east. The sharp transition zone between slower and faster v_S is displaced towards east with increasing depth.

better than the final offset and averaging inversion result. However, in other areas the image of the finally averaged inversion model better represents the synthetic anomalies. This outcome demonstrates the influence of parameterisation on the inversion result. Since sharp vertical boundaries are not expected in the upper mantle and the influence of the parameterisation on the inversion result has to be mitigated, we apply offset and averaging technique (Evans and Achauer, 1993; Evans and Zucca, 1988). Second, the mapping of the synthetic anomalies is good in the centre of the model space, but there is reduced resolution towards the model edges. This variation in resolution is caused by the ray distribution (see Fig. 6) which has a decreasing ray density and few criss-crossing rays towards the model edges. Third, the synthetic anomalies at 180–250 km depth are better resolved than those at 75–120 km depth which may be caused by the larger extension of the input anomalies at 180–250 km depth and therefore more traversed criss-crossing rays inside the parameterisation blocks. Fourth, the amplitudes of the synthetic anomalies are underestimated. The CBT indicates that the vertical resolution is reduced due to strong vertical smearing (Fig. 10) whereas the lateral resolution is reduced by applying the

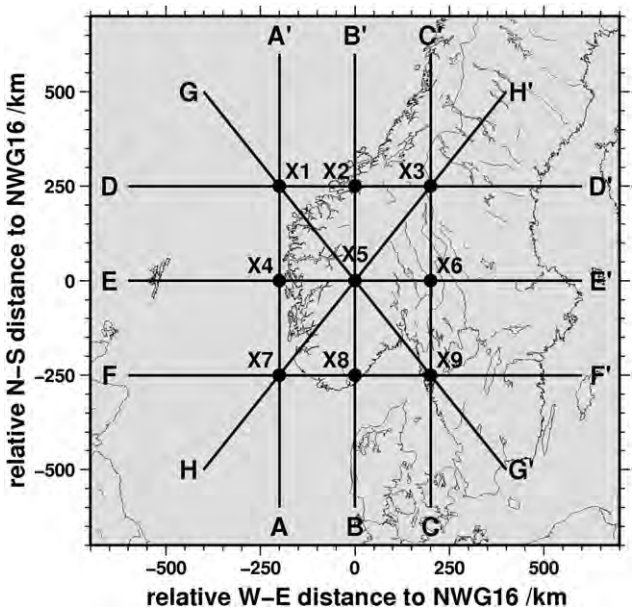


Fig. 8. Location map of the vertical cross sections (A–H) through the inversion result which are shown in Fig. 9. The intersection points of the profiles are labelled X1–X9.

offset and averaging technique. However, large anomalies (approx. 200 km) are well reproduced in the centre of the model, but there is a decrease in resolution towards the model edges.

In reconstruction test AT1 (Fig. 11) synthetic anomalies are pre-set which are similar to those of the actual inversion result (Fig. 7) with amplitudes of +1.5% v_S (blue outlines) and −1.5% v_S (red outlines) at 75–250 km depth and a W–E dipping velocity transition. Reconstruction test AT1 (Fig. 11 and Supp. Fig. S2) shows moderate vertical smearing (upwards and downwards) across two layers (e.g. no smearing into the uppermost two layers). The synthetic anomalies (75–180 km depth) are recovered very well just as the W-E transition zone between slower and faster velocity anomalies (see vertical cross sections in Supp. Fig. S2), taking into account the lateral smearing due to the application of the offset and averaging technique. The amplitudes of the synthetic anomalies are underestimated. Based on test AT1 we infer that the deep negative v_S anomaly in Fig. 7 is not caused by vertical smearing, because the amplitude contrast in AT1 at 330–410 km depth (Fig. 11) is much smaller than the v_S contrast in the actual model (Fig. 7).

In order to learn more about the resolution of the deep anomaly, we perform an additional reconstruction test AT2 (Fig. 12 and Supp. Fig. S3) which includes a velocity anomaly with an amplitude of −2.5% v_S (red outline) at 330–410 km depth. This reconstruction test AT2 recovers the negative v_S anomaly at 330–410 km depth, but its perturbation amplitude (about −1.5% v_S) is underestimated. Instead, an increase in amplitude (approx. ±1.5% v_S) is observed at 75–250 km depth due to an increase in upward vertical smearing compared to test AT1 (Fig. 11). The reconstruction tests AT1 (Fig. 11 and Supp. Fig. S2) and AT2 (Fig. 12 and Supp. Fig. S3) show that besides the W–E v_S contrast a deep additional anomaly is required to explain the negative v_S anomaly at 330–410 km depth and that the transition zone between faster velocity in the east and slower velocity in the west can be resolved well with our dataset.

To show that the upper v_S anomaly is not caused by vertical smearing of a strong deep-seated low velocity anomaly another reconstruction test AT3 (Supp Fig. S4) is performed. A low velocity anomaly is pre-set at 330–410 km depth with an amplitude of −2.5% v_S. The inversion result of this reconstruction test illustrates a well reproduced low velocity anomaly at 330–410 km depth whose amplitude is underestimated and vertical upward smearing across the model is found. However, no similarities between the inversion result of AT3 and the actual model in Fig. 7 are observed at 35–330 km depth. This confirms our assumption that the low velocity anomaly at 75–250 km depth may not be explained as inversion artefact due to a strong deep-seated anomaly.

From the four reconstruction tests (CBT, AT1, AT2 and AT3) we conclude: The ray distribution (ray density and criss-crossing rays) causes a decrease in resolution from the centre towards the edges of the model which leads to a reduced lateral resolution close to the model edges. By applying the offset and averaging technique weak lateral smearing occurs but the influence of the parameterisation is reduced. However, large anomalies (~200 km) are reproduced well. Due to strong vertical smearing the vertical resolution is low. Nevertheless, the reconstruction test AT1 (Fig. 11) and AT2 (Fig. 12) show that larger anomalies and their transitions (Supp. Figs. S2 and S3) can be mapped well and that an additional deep-seated negative v_S anomaly can be recovered. Such a negative v_S anomaly at 330–410 km depth explains the observed residuals at the MAGNUS network. Both, the upper and deeper-seated anomaly can be resolved independently (AT1 and AT3). The deeper-seated anomaly might continue below 410 km depth in the mantle but this is beyond our resolution. The positive anomaly at 250–410 km depth in the south-western part of the inversion model may partly be due to the condition for a mean value of 0% per layer. This assumption may be supported by a comparison of the reconstruction tests AT1 and AT2 (Figs. 11 and 12). A positive anomaly appears in the south-western part of the model by including a strong negative anomaly in the centre of the model at 330–410 km depth. This effect demonstrates that only velocity contrasts are meaningful in regional tomography models based on normalised residuals. In addition, the anomaly is located at the edge of our inversion volume where the resolution is low caused due to only few criss-crossing rays.

5. Discussion and interpretation

A teleseismic travel time tomography is performed to determine the 3D v_S structure beneath Southern Scandinavia. As relative travel time residuals are used as input data, v_S contrasts can be interpreted in our tomographic images of the upper mantle (Figs. 7 and 9). The observed v_S structure contains a stable v_S pattern at 35–410 km depth: v_S is reduced underneath Southern Norway compared to its surroundings, especially the Baltic Shield in the east. The transition between slower and faster v_S is sharp and it dips towards east with increasing depth. The velocity contrast is maximum 4% v_S and on average up to 3% v_S. In Fig. 13 a schematic illustration summarises the main features which are recovered by our teleseismic shear wave analysis. The reconstruction tests (AT1 and AT2) imply that the W–E dipping transition zone as well as the deep anomaly down to 410 km depth are well resolved.

There are also global and regional tomography studies which are based on body waves (P) and surface waves and cover the area of Southern Scandinavia. Weidle and Maupin (2008) analysed group velocities of surface waves (Rayleigh and Love waves) to derive the 3D v_S distribution underneath Northern Europe. They revealed a negative velocity anomaly beneath Iceland that stretches out towards Southern Norway. In their model v_S is reduced up to −10% of v_{S_ak135} at 70–150 km depth (Weidle and Maupin, 2008). Across Southern Norway the velocity contrast is about 5%–7% of v_{S_ak135} (Weidle and Maupin, 2008) what is significantly larger than our modelled v_S reduction of −3% to −4% of v_{S_iasp91}. This difference may be caused by performing different tomography methods with diverse sensitivities and resolutions. As we determine relative travel time residuals, a regionally broad v_S anomaly would be removed and only variations within the anomaly would be left over. Thereby our model can only resolve these more local v_S contrasts underneath the MAGNUS network and may underestimate a broader anomaly.

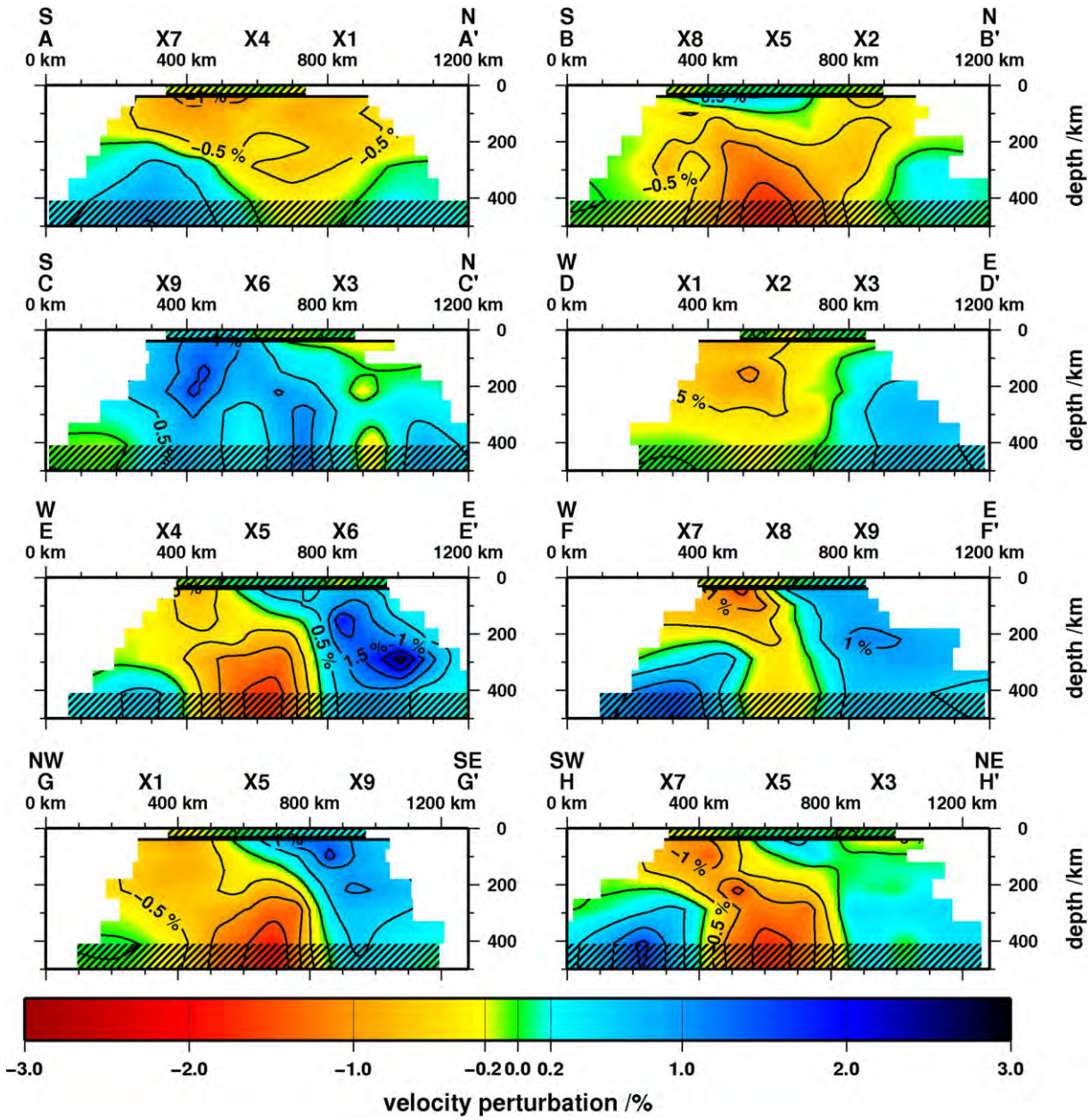


Fig. 9. Vertical slices through the inversion result after applying the offset and averaging technique. The S-wave velocity contrast is colour-coded with the same colour scale as in Fig. 7. Underneath Southern Norway the S-wave velocity is reduced compared to the Baltic Shield in the east. In west–east direction (profiles DD′, EE′, FF′, GG′ and HH′) a sharp transition zone between slower and faster v_S is visible which is displaced towards east with increasing depth.

By analysing the phase velocity of Rayleigh waves at the MAGNUS stations, Maupin (2011) determined an average 1D $v_S(z)$ profile underneath Southern Norway. In this model the upper mantle v_S is reduced compared to the reference Earth model *ak135* (Kennett et al., 1995). At 150 km depth v_S is 2% slower than v_{S_ak135} (Maupin, 2011) and approximately 0.3 km/s slower than v_S expected in cratonic areas (Pedersen et al., 2009) as found underneath Sweden (Maupin, 2011). Maupin's (2011) Rayleigh wave model is in good agreement with our tomographic body wave model which shows a velocity change of −2% v_{S_iasp91} below Southern Norway and a velocity contrast of up to 4% v_{S_iasp91} between the Southern Norway in the west and the Baltic Shield in the east.

P-wave tomography models (Bijwaard and Spakman, 1999; Spakman published in Jones et al., 2010; Medhus et al., 2012) also show a reduction of mantle v_P underneath Southern Norway compared to the Baltic Shield in the east. The global model of Bijwaard and Spakman (1999) contains a v_P reduction of −2% v_{P_ak135} at 300 km depth, whereas a more recent study of Spakman (published in Jones et al., 2010) image a v_P reduction of 0.2–0.3 km/s relative to the Baltic Shield. In both studies v_P increases towards the Baltic Shield and the transition between faster and slower velocity is located in the area of the Oslo Graben (Bijwaard and Spakman, 1999; Spakman published in Jones et al., 2010). Medhus et al. (2012) performed an absolute and a relative v_P travel time tomography whose v_P distributions show a strong reduction of v_P underneath Southern Norway compared to the Baltic Shield. At a depth range of 100–300 km the v_P contrast is up to 3%–4% of v_{P_ak135} in the absolute velocity model and also 3%–4% of v_{P_iasp91} in the relative velocity model (Medhus et al., 2012).

In contrast to the above mentioned studies, Ritsema et al. (2004) and Megnin and Romanowicz (2000) found no indication of a velocity reduction beneath Southern Norway compared to its surroundings.

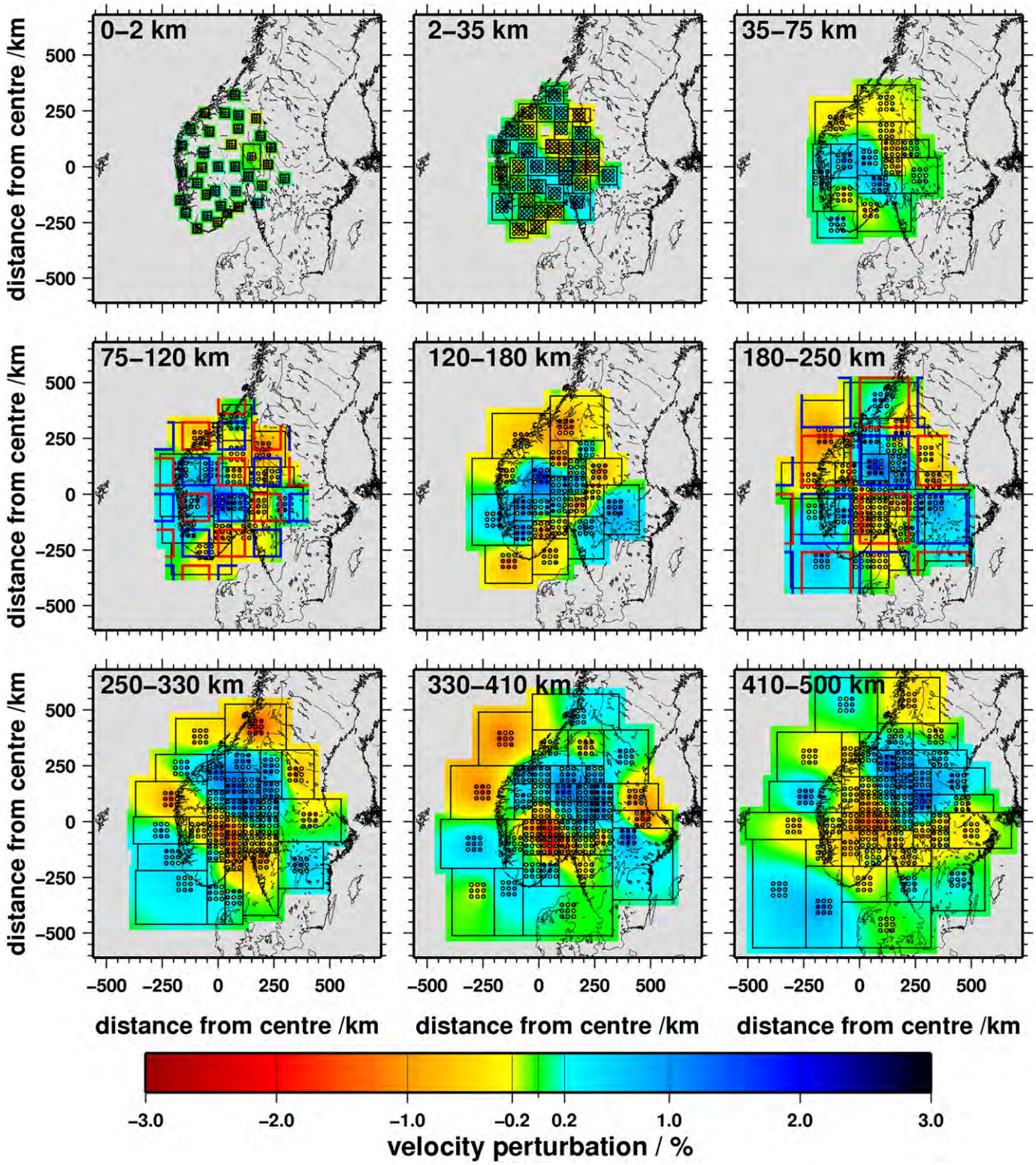


Fig. 10. Reconstruction test CBT (checkerboard pattern test). The pre-set synthetic anomalies (blue outline: +1.5% v_S, red outline: −1.5% v_S) are placed in the velocity model at 75–120 km and 180–250 km depth. The interpolated S-wave velocity perturbations visualise the corresponding inversion result after applying the offset and averaging technique. Additionally, the circles show the block-wise inversion results of the individual nine offset inversion models before averaging. There is strong vertical smearing which leads to reduced vertical resolution within the model space.

This might be due to a lower resolution in these global tomography studies caused by the sparse covering with permanent stations in Scandinavia.

The conclude, the seismic velocity (v_P and v_S) studies, which are based on dense regional station networks, show a major common feature: the reduced seismic velocity underneath the Scandinavian Mountains and relatively increased seismic velocity underneath the Baltic Shield (this study; Bijwaard and Spakman, 1999; Maupin, 2011; Weidle and Maupin, 2008; Spakman published in Jones et al., 2010; Medhus et al., 2012). However, the seismic velocity contrast

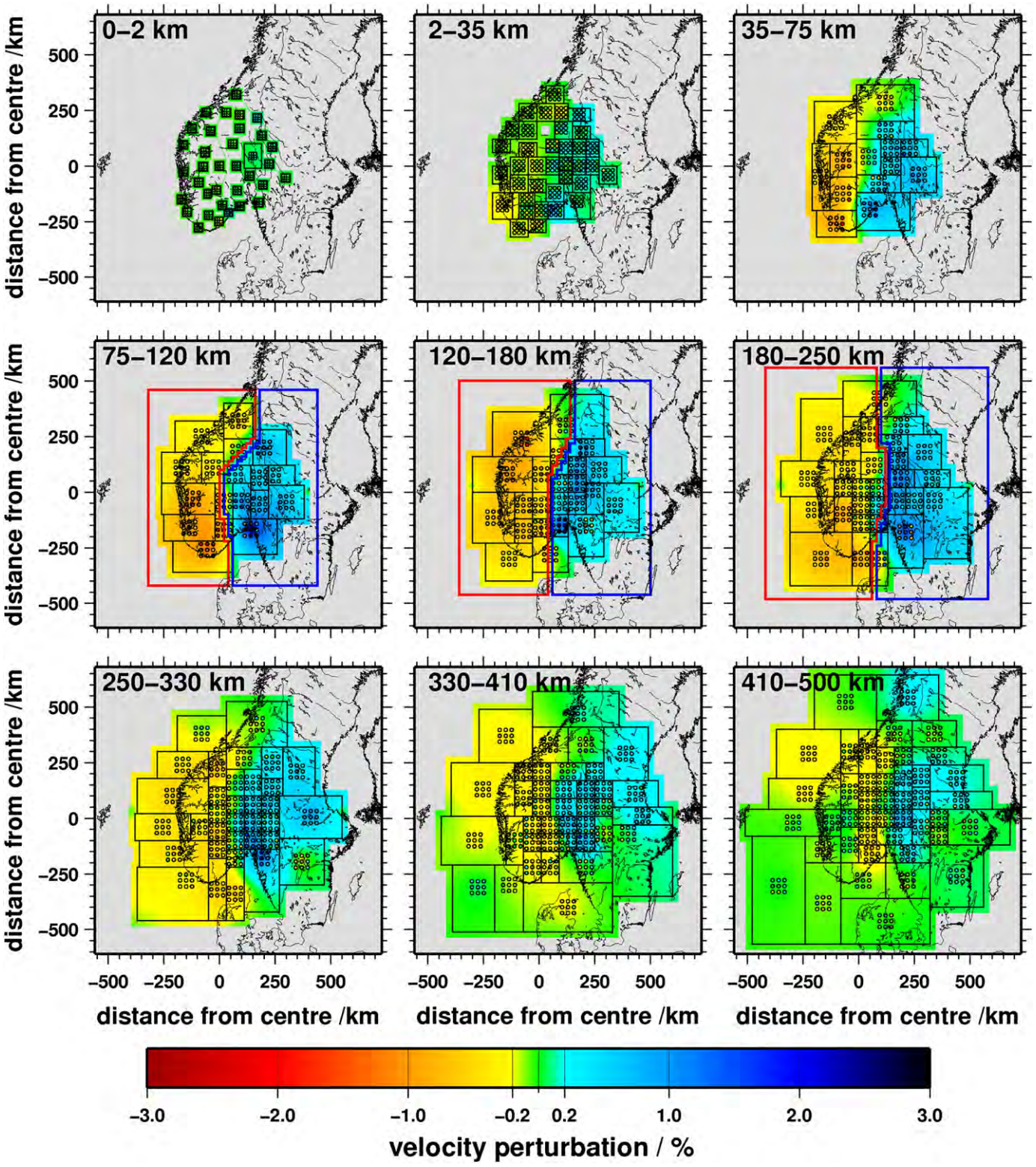


Fig. 11. Reconstruction test AT1 (anomaly pattern test 1). The pre-set synthetic anomalies are placed in the velocity model at 75–250 km depth (blue outline: +1.5% v_S, red outline: −1.5% v_S) and these represent a dipping W–E contrast as found in the real model (Fig. 7). The interpolated S-wave velocity perturbations visualise the corresponding inversion result after applying the offset and averaging technique (vertical cross sections see Supp. Fig. S2). Additionally, the circles show the inversion results of the individual nine offset inversion models before averaging. This reconstruction test illustrates a good reproduction (position and amplitude) of the synthetic anomalies, although weak vertical smearing is visible across two layers.

and especially the amplitude of the velocity reduction vary as well as the exact location of the anomaly. Differences between these v_P and v_S structures are related mainly to the perturbation amplitude which might be caused by analysing P- and S-waves in different period ranges, by different ray paths or propagation directions, different resolution properties and different model space parameterisations. Additionally, a variation in the v_P/v_S ratio might affect differences between v_P and v_S tomographic structures.

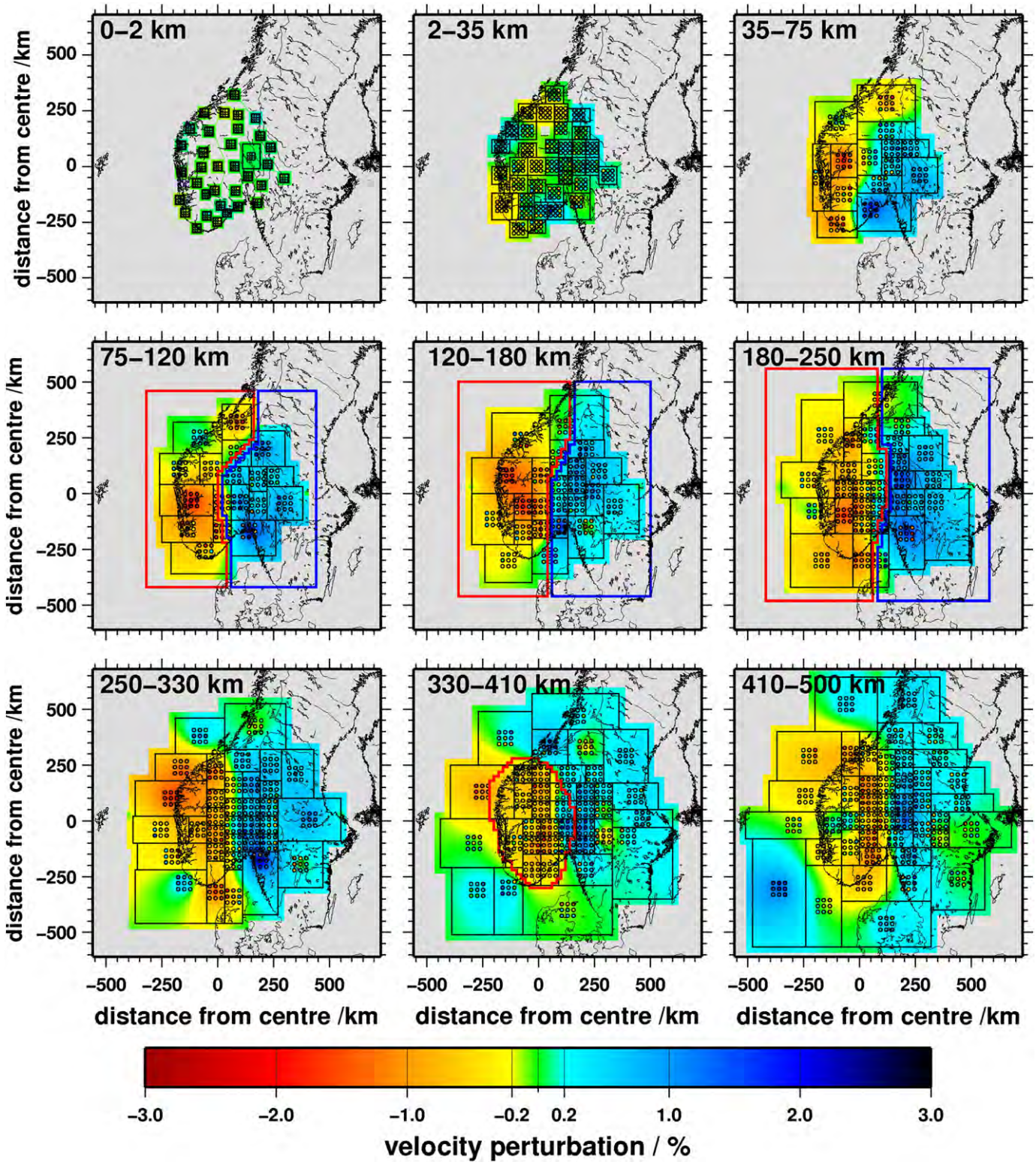


Fig. 12. Reconstruction test AT2 (anomaly pattern test 2). The pre-set synthetic anomalies are placed in the velocity model at 75–250 km depth (blue outline: +1.5% v_S, red outline: −1.5% v_S) in addition to a stronger synthetic anomaly (red outline: −2.5% v_S) at 330–410 km depth. These anomalies represent v_S contrasts as found in the real model (Fig. 7). The interpolated S-wave velocity distribution visualises the corresponding inversion result after applying the offset and averaging technique (vertical cross sections see Supp. Fig. S3). Additionally, the circles show the inversion results of the individual nine offset inversion models before averaging. This reconstruction test illustrates a good reproduction (position and amplitude) of the synthetic anomalies, although some vertical smearing is visible.

In Scandinavia other regional tomography studies also imaged deep reaching velocity contrasts in the upper mantle across major tectonic boundary zones. E.g. the Tornquist Zone, the suture between the Proterozoic NE Europe and the Phanerozoic Central Europe, is characterised by up to 5% P-wave velocity (v_P) contrast by non-linear tomography (Shomali et al., 2002). Sandoval et al. (2004) recovered an inclined velocity contrast at the suture of the Archean Karelian terrane and the Proterozoic Svecofennian terrane in Finland

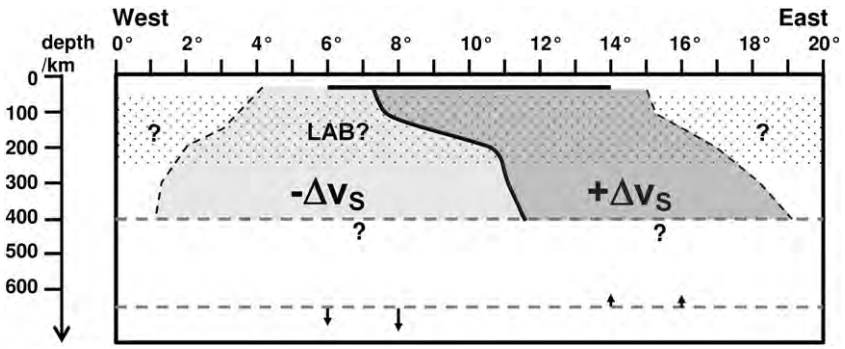


Fig. 13. Vertical cross section through a schematic model of the shear wave structure underneath Southern Scandinavia at 60.5°N (Wawerzinek, 2012). At 35–410 km depth, v_S is continuously reduced underneath Southern Norway (west of 11°E,-Δv_S, light grey) compared to the Baltic Shield (+Δv_S, dark grey). In addition, the schematic model contains discontinuities which were recovered by common conversion point stacks of S-receiver functions (Wawerzinek, 2012). Several transitions to low velocity zones are visible in a depth range of 50–250 km (dashed zone), but a unique identification of the lithosphere–asthenosphere-boundary is not possible. Observed depth variations of the 660 km discontinuity (lower grey dashed line) are marked by arrows which are not scaled correctly. The 660 km discontinuity is deeper underneath Southern Norway compared to the Baltic Shield. In the area of Southern Norway, the 410 km discontinuity is not detected in our S-RF (upper grey dashed line) which might be caused by the observed v_S anomaly at 330–410 km depth.

and Russia. Further in the west, in Sweden, Eken et al. (2007) determined a v_P contrast of 2%–3% across the Proterozoic–Archean lithospheric domains. Their tomographic image shows a slab-like shape. In Eken et al. (2008) P- and S-wave velocity contrasts are displayed for the Proterozoic–Archean suture as well as for the transition between the Transcandinavian Igneouos Belt and the Svecofennian domain. These examples and our new results indicate that very old collisional processes from Proterozoic and even Archean times are still preserved in the upper mantle.

In the following, we discuss several reasons that may cause the velocity reduction beneath Southern Norway relative to the Baltic Shield: a variation in lithospheric thickness, a variation in mantle composition, a temperature variation, the presence of hydrated rocks or partial melts or a combination of some or all these possible causes. For petrophysical impacts on seismic velocity we refer to e.g. Cammarano et al. (2003), Goes et al. (2004) or Ritter (2007). The following interpretations may be valid for our v_S mantle model:

I. We infer an increase in lithospheric thickness from the Scandes towards the Baltic Shield as presented e.g. by Plomerová et al. (2008). They compiled regional body-wave data to create a model of the lithospheric thickness of Fennoscandia. In their model the lithospheric thickness increases from the western coast of Southern Norway (<90 km) towards the Baltic Shield in the east (>130 km; Plomerová et al., 2008). Shear wave velocity is slightly faster in the lower lithosphere compared to the asthenosphere (Eaton et al., 2009) and slightly faster inside older shield lithosphere compared to younger Phanerozoic orogens (Jones et al., 2010). Thus a lithospheric thickness variation might explain the lateral velocity variations at a depth range of 75–250 km (Fig. 9). Thereby, the observed slower v_S underneath Southern Norway could be interpreted as v_S in a shallower asthenosphere relative to the faster v_S in a deeper lithospheric root underneath the Baltic Shield. This lithosphere–asthenosphere transition seems not to be a steep boundary, but it is inclined with a dip towards east (Figs. 9 and 13).
II. The composition of the upper mantle certainly differs between the Phanerozoic Southern Norwegian region and the Baltic Shield in Southern Sweden, because the past geodynamic evolution of Southern Norway was strongly influenced by the Caledonian orogeny, the Cenozoic opening of the Northern Atlantic and the Neogene basin building in the North Sea (Ramberg et al., 2008; Torsvik and Cocks, 2005). In the mantle a related variation in mineral composition, fluid and melt content might affect the seismic velocity structure beneath Southern Scandinavia. However, we expect only small seismic velocity changes due to variations in the mantle composition (Cammarano et al., 2003; Sobolev et al., 1997).
III. Since temperature variations cause larger seismic velocity perturbations than compositional variations in an anhydrous mantle (Goes et al., 2004; Sobolev et al., 1997), temperature differences could perhaps better explain our velocity perturbations in the upper mantle. A positive temperature anomaly, which may involve partial melts, leads to a reduction of the elastic moduli and therefore to a reduction of the seismic velocity (Berckhemer et al., 1982). If the v_S reduction is caused by a pure temperature anomaly underneath Southern Norway, a temperature increase of up to 200–330 K (Cammarano et al., 2003) is required to explain a velocity contrast of up to 4% v_S at 75–250 km depth based on an average mineral composition of the mantle. A deeper anomaly at 330–410 km depth, which has a velocity contrast of about 4% v_S, may be explained by a pure temperature anomaly of about 400 K (Cammarano et al., 2003). This relatively high temperature anomaly might be reduced by taking into account hydrous minerals at 330–410 km depth. Thus a smaller temperature anomaly of less than 300 K is enough to explain the observed contrast of about 4% v_S (Ritter, 2007). However, the presence of fluids, partial melts or hydrous minerals inside the deep anomaly below Southern Norway cannot be confirmed at the moment.

In summary, our S-wave tomography study contributes the 3D shear wave velocity structure in the upper mantle to the on-going discussion on the origin of the Scandinavian Mountains. Especially our results should help to define the possible dynamic contribution of the mantle structure. The improved knowledge of the v_S structure of the upper mantle is an important proxy for the determination of the driving forces for the uplift and high topography. If one accepts increased temperature and the presence of shallow asthenospheric material underneath the mountains as reason for the seismic velocity reduction, then recently induced dynamic topography (Braun, 2010) may be an option to explain at least a fraction of the present topography in Southern Norway. However, the mantle domain with reduced seismic velocities (S-wave model in Fig. 7 and P-wave model in Medhus et al., 2012) is wider than the area of high topography, thus this low velocity mantle domain may affect the lithosphere in Southern Scandinavia across a wider region (Medhus et al., 2012). The uplift mechanism of Southern Norway can be better tested now by performing geodynamic modelling to explore different hypothesis, mechanisms and processes based on proxies from seismology. To verify or exclude different processes the observed shear wave velocity structure has to be taken into account as boundary condition. In the second part of *TopoScandiaDeep*, such a geodynamic modelling will be

performed to determine the cause of uplift in Southern Norway (Weidle et al., 2010).

6. Conclusions

To determine the 3D shear wave velocity structure underneath Southern Scandinavia, we analyse S-waves of 128 teleseismic events which were recorded at the MAGNUS network in Southern Norway. In the first step, we compute relative travel time residuals (Section 3) by taking into account 1D travel time effects due to the known crustal structure. These residuals are caused by velocity perturbations underneath the station network within the upper mantle. The resulting residual pattern of nearly vertically incident waves shows a simple residual distribution: late arrivals are visible in Southern Norway and early arrivals are observed in the area of the Oslo Graben and the Baltic Shield. The travel time difference is maximum 4 s across the MAGNUS network. S-wave residuals with flat incident ray paths depend significantly on backazimuth: the area of late arrivals is shifted systematically in the propagation direction of the teleseismic waves. This leads to the prediction that a low shear velocity zone underneath Southern Norway is the cause of the observed residual pattern.

In the second step, we compute a non-linear travel time tomography (Section 4) by using the JI-3D inversion code (Jordan, 2003). Taking into account pick uncertainties and data quality, weighted residuals are used as input data. The resulting v_S structure shows a continuous velocity pattern: slower v_S in the west compared to the east. The transition between fast and slow velocity dips towards east with increasing depth (Fig. 13). Below 180 km depth the area of slower velocity comprises Southern Norway compared to the area of faster velocity underneath the Baltic Shield. The low v_S anomaly reaches as deep as the mantle transition zone at 410 km depth. Because relative travel time residuals are used as input data for the tomography, only the velocity contrast can be interpreted which is up to 4% of v_S and on average 3% of v_S across Southern Scandinavia. Four reconstruction tests (Section 4.3) indicate a good resolution in the centre of the model space, e.g. the area of the lateral velocity transition. Furthermore, the reconstruction tests suggest that the observed low velocity anomaly at 330–410 km depth can be recovered and that it is not an inversion artefact.

Finally, the observed 3D v_S structure can be explained by laterally varying mantle properties. An increase in lithospheric thickness towards east can be related with different lithologies and/or lateral ambient temperature conditions. A shallower asthenosphere beneath Southern Norway would cause a positive temperature anomaly relative to colder lithosphere of the Baltic Shield in the east. In the lowermost upper mantle the deep-seated low v_S anomaly might be due to a combination of a temperature anomaly with fluids, partial melts and/or hydrous minerals at 330–410 km depth and might be due to deeper reaching geodynamic processes in the mantle transition zone or even the lower mantle.

The variable upper mantle structure indicates that mantle processes might influence or even support a Neogene uplift of Southern Norway. A more detailed geodynamic modelling analysis will be conducted including our upper mantle v_S structure. This analysis of possible uplift mechanisms will provide more insights on the regional uplift processes such as dynamic topography in Southern Norway.

Acknowledgements

MAGNUS waveforms were recorded with the mobile KArlsruhe BroadBand Array of the Karlsruhe Institute of Technology, Germany as well as with permanent stations of the NORSAR array and the Norwegian National Seismological Network. Financial support for the MAGNUS experiment was provided by the Universities of Aarhus, Copenhagen, Karlsruhe and Oslo as well as NORSAR. This work has been done in the framework of the ESF EUROCORES TOPO-EUROPE Programme 07-TOPO-EUROPE-FP-014: The Scandinavian mountain chain: deep processes (*TopoScandiaDeep*). We acknowledge financial support from the Deutsche Forschungsgemeinschaft (RI1133/8). We thank the many helping hands during the experiment (station deployment and station service) as well as data preparation. We thank the TopoScandiaDeep working group for helpful discussions. Figures were created with Generic Mapping Tools (Wessel and Smith, 1998) and in MATLAB®. Data pre-processing was done in MATLAB® using KaSP-Toolbox (Groos, 2010), S onset picking was done in SeismicHandler (Stammler, 1993) and the inversion was done by using JI-3D inversion code (Jordan, 2003). We thank the reviewers W. Spakman and T. Eken for helpful comments to improve the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.tecto.2012.12.033.

References

Aki, K., Christoffersson, A., Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere. Journal of Geophysical Research 82, 277–296.
Anell, I., Thybo, H., Artemieva, I.M., 2009. Cenozoic uplift and subsidence in the North Atlantic region: geological evidence revisited. Tectonophysics 474, 78–105.
Berckhemer, H., Kampfmann, W., Aulbach, E., Schmeling, H., 1982. Shear modulus and Q of forsterite and dunite near partial melting from forced oscillation experiments. Physics of the Earth and Planetary Interiors 29, 30–41.
Berteussen, K.A., 1975. Crustal structure and P-wave travel time anomalies at NORSAR. Journal of Geophysics 41, 71–84.
Bijwaard, H., Spakman, W., 1999. Tomographic evidence for a narrow whole mantle plume below Iceland. Earth and Planetary Science Letters 166, 121–126.
Braun, J., 2010. The many surface expressions of mantle dynamics. Nature Geoscience 3, 825–833.
Cammarano, F., Goes, S., Vacher, P., Giardini, D., 2003. Inferring upper-mantle temperatures from seismic velocities. Physics of the Earth and Planetary Interiors 138, 197–222.
Cloetingh, S.A.P.L., Ziegler, P.A., Beekman, F., Ben-Avraham, Z., Brun, J.-P., Bunge, H.P., Burov, E.B., Carbonell, R., Facenna, C., Friedrich, A., Gallert, J., Green, A.G., Heidbach, O., Jones, A.G., Matenco, L., Mosar, J., Oncken, O., Pascal, C., Peters, G., Sliaupa, S., Soesoo, A., Spakman, W., Stephenson, R.A., Thybo, H., Torsvik, T., de Vicente, G., Wenzel, F., Wortel, M.J.R., TOPO-EUROPE Working Group, 2007. TOPO-EUROPE: the geoscience of coupled deep Earth-surface processes. Global and Planetary Change 58, 1–118.
Douglas, A., Bowers, D., Young, J.B., 1997. On the onset of P seismograms. Geophysical Journal International 129, 681–690.
Eaton, D.W., Darbyshire, F., Evans, R.L., Grütter, H., Jones, A.G., Yuan, X., 2009. The elusive lithosphere–asthenosphere boundary (LAB) beneath cratons. Lithos 109, 1–22.
Eken, T., Shomali, Z.H., Roberts, R., Bödvarsson, R., 2007. Upper-mantle structure of the Baltic Shield below the Swedish National Seismological Network (SNSN) resolved by teleseismic tomography. Geophysical Journal International 169, 617–630.
Eken, T., Shomali, Z.H., Roberts, R., Hieronymus, C.F., Bödvarsson, R., 2008. S and P velocity heterogeneities within the upper-mantle below the Baltic Shield. Tectonophysics 462, 109–124.
Engdahl, E.R., van der Hilst, R., Buland, R., 1998. Global teleseismic earthquake relocation with improved traveltimes and procedures for depth determination. Bulletin of the Seismological Society of America 88, 722–743.
Evans, J.R., Achauer, U., 1993. Teleseismic velocity tomography using the ACH method: theory and application to continental-scale studies. In: Iyer, H.M., Hirahara, K. (Eds.), Seismic Tomography: Theory and Practice. Chapman and Hall, London, pp. 319–360.
Evans, J.R., Zucca, J., 1988. Active high-resolution seismic tomography of compressional wave velocity and attenuation structure at Medicine Lake volcano, northern California Cascade range. Journal of Geophysical Research 93, 15015–15036.
Gabrielsen, R.H., Faleide, J.I., Pascal, C., Braathen, A., Nystuen, J.P., Etzelmuller, B., O'Donnell, S., 2010. Reply to discussion of Gabrielsen et al. (2010) by Nielsen et al. (this volume): latest Caledonian to present tectonomorphological development of southern Norway. Marine and Petroleum Geology 27, 709–723.
Goes, S., Cammarano, F., Hansen, U., 2004. Synthetic seismic signature of thermal mantle plumes. Earth and Planetary Science Letters 218, 403–419.
Groos, J., 2010. Broadband seismic noise: Classification and Green's function estimation. Dissertation, Geophysical Institute, Karlsruhe Institute of Technology (KIT).
Japsen, P., Chalmers, J.A., 2000. Neogene uplift and tectonics around the north Atlantic: overview. Global and Planetary Change 24, 165–173.
Jones, A., Plomerova, J., Korja, T., Sodoudi, F., Spakman, W., 2010. Europe from bottom up: a statistical examination of the central and northern European lithosphere–asthenosphere boundary from comparing seismological and electromagnetic observations. Lithos 120, 14–29.
Jordan, M., 2003. JI-3D. A new approach to high resolution regional seismic tomography: Theory and applications, PhD thesis. Universität Göttingen, Göttingen, Germany.
Kennett, B.L.N., Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification. Geophysical Journal International 105, 429–465.

Kennett, B.L.N., England, E., Buland, R., 1995. Constraints on seismic velocities in the Earth from travel times. Geophysical Journal International 122, 108–124.
Keyser, M., Ritter, J.R.R., Jordan, M., 2002. 3D shear wave velocity structure of the Eifel plume, Germany. Earth and Planetary Science Letters 203, 59–82.
Kirschner, S., Ritter, J., Wawerzinek, B., 2011. Teleseismic wave front anomalies at a Continental Rift: no mantle anomaly below the central Upper Rhine Graben. Geophysical Journal International 186, 447–462.
Lidmar-Bergström, K., Bonow, J.M., 2009. Hypotheses and observations on the origin of the landscape of southern Norway — a comment regarding the isostasy-climate-erosion hypothesis by Nielsen et al. 2008. Journal of Geodynamics 48, 95–100.
Lidmar-Bergström, K., Ollier, C.D., Sulebak, J.R., 2000. Landforms and uplift history of southern Norway. Global and Planetary Change 24, 211–231.
Martin, M., Ritter, J.R.R., the CALIXTO working group, 2005. High-resolution teleseismic body-wave tomography beneath SE Romania. — I. Implications for three-dimensional versus one-dimensional crustal correction strategies with a new crustal velocity model. Geophysical Journal International 162, 448–460.
Maupin, V., 2011. Upper-mantle structure in southern Norway from beamforming of Rayleigh wave data presenting multipathing. Geophysical Journal International 185, 985–1002.
Medhus, A.B., Balling, N., Holm Jacobsen, B., Kind, R., England, R.W., 2009. Deep-structural differences in southwestern Scandinavia revealed by P-wave travel time residuals. Norwegian Journal of Geology 89, 203–214.
Medhus, A.B., Balling, N., Jacobsen, B.H., Weidle, C., England, R.W., Kind, R., Thybo, H., Voss, P., 2012. Upper-mantle structure beneath the Southern Scandes Mountains and the Northern Tornquist Zone revealed by P-wave traveltime tomography. Geophysical Journal International 189, 1315–1334.
Megnin, C., Romanowicz, B., 2000. The three-dimensional velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms. Geophysical Journal International 143, 709–728.
Nielsen, S.B., Gallagher, K., Leighton, C., Balling, N., Svenningsen, L., Jacobsen, B.H., Thomsen, E., Nielsen, O.B., Heilmann-Clausen, C.H., Egholm, D.L., Summerfield, M., Clausen, O.R., Piotrowski, J.A., Thorsen, M.R., Huuse, M., Abrahamsen, N., King, C., Lykke-Andersen, H., 2009. The evolution of western Scandinavian topography: a review of Neogene uplift versus the ICE (isostasy-climate-erosion) hypothesis. Journal of Geodynamics 47, 72–79.
Pascal, C., Olesen, O., 2009. Are the Norwegian mountains compensated by a mantle thermal anomaly at depth? Tectonophysics 475, 160–168.
Pedersen, H., Fishwick, S., Snyder, D., 2009. A comparison of cratonic roots through consistent analysis of seismic surface waves. Lithos 109, 81–95.
Plomerová, J., Babuška, V., Kozlovskaya, E., Hyvönen, L.V.L., 2008. Seismic anisotropy — A key to resolve fabrics of mantle lithosphere of Fennoscandia. Tectonophysics 462, 125–136.
Ramberg, I.B., Bryhni, I., Nøttvedt, A., Rangnes, K., 2008. The Making of a Land — Geology of Norway. Norsk Geologisk Forening, Trondheim.
Ritsema, J., van Heijst, H.J., Woodhouse, J.H., 2004. Global transition zone tomography. Journal of Geophysical Research 109, B02302.
Ritter, J.R.R., 2007. The seismic signature of the Eifel Plume. In: Ritter, J.R.R., Achauer, U. (Eds.), Mantle Plumes: A Multidisciplinary Approach. Springer, Heidelberg, pp. 379–404.
Rohrman, M., van der Beek, P., 1996. Cenozoic postrift domal uplift of North Atlantic margins: an asthenospheric diapirism model. Geology 24, 901–904.
Roy, C., Ritter, J.R.R., in press. Complex deep seismic anisotropy below the Scandinavian Mountains. Journal of Seismology. http://dx.doi.org/10.1007/s10950-012-9325-4.
Sandoval, S., Kissling, E., Ansorge, J., SVEKALAPKO Seismic Tomography Working Group, 2004. High-resolution body wave tomography beneath the SVEKALAPKO array – II. Anomalous upper mantle structure beneath the central Baltic Shield. Geophysical Journal International 157, 200–214.
Shomali, Z.H., Roberts, R.G., TOR Working Group, 2002. Non-linear body wave teleseismic tomography along the TOR array. Geophysical Journal International 148, 562–574.
Shomali, Z.H., Roberts, R.G., Pedersen, L.B., the TOR Working Group, 2006. Lithospheric structure of the Tornquist Zone resolved by nonlinear P and S teleseismic tomography along the TOR array. Tectonophysics 416, 133–149.
Smelror, M., Dehls, J., Ebbing, J., Larsen, E., Lundin, E.R., Nordgulen, O., Osmundsen, P.T., Olesen, O., Ottesen, D., Pascal, C., Redfield, T.F., Rise, L., 2007. Towards a 4D topographic view of the Norwegian sea margin. Global and Planetary Change 58, 382–410.
Sobolev, S.V., Zeyen, H., Granet, M., Achauer, U., Bauer, C., Werling, F., Altherr, R., Fuchs, K., 1997. Upper mantle temperatures and lithosphere–asthenosphere system beneath the French Massif Central constrained by seismic, gravity, petrologic and thermal observations. Tectonophysics 275, 143–164.
Stammler, K., 1993. SeismicHandler — Programmable multichannel data handler for interactive and automatic processing of seismological analyses. Computers and Geosciences 19, 135–140.
Steck, L.K., Prothero, W.A., 1991. A 3D raytracer for teleseismic bodywave arrival times. Bulletin of the Seismological Society of America 81, 1332–1339.
Stratford, W., Thybo, H., 2011a. Seismic structure and composition of the crust beneath the southern Scandes, Norway. Tectonophysics 502, 364–382.
Stratford, W., Thybo, H., 2011b. Crustal structure and composition of the Oslo Graben, Norway. Earth and Planetary Science Letters 304, 441–442.
Stratford, W., Thybo, H., Faleide, J.I., Olesen, O., Trygvasson, A., 2009. New Moho map of onshore southern Norway. Geophysical Journal International 178, 1755–1765.
Svenningsen, L., Balling, N., Jacobsen, B., Kind, R., Wylegalla, K., Schweitzer, J., 2007. Crustal root beneath the highlands of southern Norway resolved by teleseismic receiver functions. Geophysical Journal International 170, 1129–1138.
Torsvik, T.H., Cocks, L.R.M., 2005. Norway in space and time: a centennial cavalcade. Norwegian Journal of Geology 85, 73–86.
Wawerzinek, B., 2012. Untersuchung der elastischen Scherwellenstruktur unter dem Südskandinavischen Gebirge. Dissertation, Karlsruhe Institute of Technology, (in German).
Wawerzinek, B., Ritter, J.R.R., Jordan, M., Landes, M., 2008. An upper-mantle upwelling underneath Ireland revealed from non-linear tomography. Geophysical Journal International 175, 253–268.
Weidle, C., Maupin, V., 2008. An upper-mantle S-wave velocity model for Northern Europe from Love and Rayleigh group velocities. Geophysical Journal International 175, 1154–1168.
Weidle, C., Maupin, V., Ritter, J., Kværna, T., Schweitzer, J., Balling, N., Thybo, H., Faleide, J.I., Wenzel, F., 2010. MAGNUS — a seismological broadband experiment to resolve crustal and upper mantle structure beneath the Southern Scandes Mountains in Norway. Seismological Research Letters 81 (1), 76–84.
Wessel, P., Smith, W.H.F., 1998. New, improved version of the Generic Mapping Tools released. EOS, Transactions, American Geophysical Union 79, 579.
Zeyen, H., Achauer, U., 1997. Joint inversion of teleseismic delay times and gravity anomaly for regional structures. In: Fuchs, K. (Ed.), Upper Mantle Heterogeneities from Active and Passive Seismology. Kluwer Academic Publishers, Dordrecht, pp. 155–169.

Appendix B

Article: Complex deep seismic anisotropy below the Scandinavian Mountains

This article contains mainly my work of my diploma thesis at the Geophysical Institute (GPI) at the Karlsruhe Institute of Technology (KIT) in Germany. Additionally, results of my research on "Application of array processing to measure teleseismic shear wave splitting" during a one month research visit in 2011 at the Norwegian Seismic Array (NORSAR), supported by a NERA grant, are included. The collaboration with the GPI continued during my Ph.D thesis at the IPGP and the article was finished and published in August 2012.

Abstract

To study seismic anisotropy the birefringence of SKS and SKKS phases were analyzed. These phases, which should be polarized radially, are split into an additional transverse component if they propagate through an anisotropic medium. The results are directions ϕ of the apparent fast shear wave polarization and delay times δt between the split phases.

For station KONO in Southern Norway, ϕ and δt are frequency-dependent ϕ and δt values, indicating a depth-dependent anisotropy. Additionally, ϕ and δt values vary with epicentre backazimuths in Norway, indicating a complex anisotropic structure in the crust and upper mantle.

Stacking of the SKS/SKKS waveforms improves the signal-to-noise ratio along one station line and allows us to better determine the splitting parameters. A unique and complete model of the complex anisotropy cannot be obtained due to the limited observed backazimuth range. Near-surface tectonic structures correlate with the splitting pattern and thus the crust is one anisotropic layer in the region. Partly preferred orientations in the rock fabric at the surface can be correlated with ϕ.

J Seismol
DOI 10.1007/s10950-012-9325-4

ORIGINAL ARTICLE

Complex deep seismic anisotropy below the Scandinavian Mountains

Corinna Roy · Joachim R. R. Ritter

Received: 28 January 2012 / Accepted: 16 July 2012

Abstract Several seismological projects focused on the deep structure of the Scandinavian Mountains, in Norway and neighbouring Sweden. We use these recordings to study seismic anisotropy by analysing the birefringence of SKS and SKKS phases. These phases, which should be polarised radially, are split into an additional transverse component if they propagate through an anisotropic medium. Our results are directions Φ of the apparent fast shear wave polarisation and delay times δt between the split phases. For station KONO in Southern Norway, we find frequency-dependent Φ and δt values, indicating a depth-dependent anisotropy. Additionally, Φ and δt values vary with epicentre backazimuths in Norway, indicating a complex anisotropic structure in the crust and upper mantle. Stacking of the SKS/SKKS waveforms improves the signal-to-noise ratio along one station line and allows us to better determine the splitting parameters. A unique and complete model of the complex anisotropy cannot be obtained due to the limited observed backazimuth range. Near-surface tectonic structures correlate with the splitting pattern and thus the crust is one anisotropic layer in the region. Partly preferred orientations in the rock fabric at the surface can be correlated with Φ. Below one or more anisotropic layers must exist to explain the backazimuth- and frequency-dependent observations, as well as the long δt values (>2 s) which cannot be explained with crustal anisotropy alone. The spatial distribution of the splitting results indicates that different tectonics units, e.g. the Sveconorwegian, the Central and Northern Svecofennian and the Caledonian nappes, are each characterised by specific anisotropic signatures.

Keywords Scandinavia · SKS splitting · Anisotropy · Lithosphere

1 Introduction

The present high topography of the Scandinavian Mountain range (also called Scandes; Fig. 1) cannot be explained uniquely by existing geodynamic models (e.g. Pascal and Olesen 2009; Lidmar-Bergström and Bonow 2009; Gabrielsen et al. 2010 and references therein). At the same time, seismic anisotropy, which is the direction-dependent propagation velocity of seismic waves and often due to geodynamic deformational processes, is still unknown below the Scandes. Recently, Weidle and Maupin (2008) found a low shear wave velocity zone of −6 % relative to the surrounding mantle of about 70–150 km in depth which reaches from the Iceland plume across the NE Atlantic to Southern Norway. This anomaly might play a role in the Cenozoic uplift phase of the Scandes if it acts as a buoyancy source. Using Rayleigh wave

C. Roy · J. R. R. Ritter (✉)
Geophysical Institute, Karlsruhe Institute of Technology,
Hertzstr. 16,
76187 Karlsruhe, Germany
e-mail: joachim.ritter@kit.edu

Published online: 03 August 2012

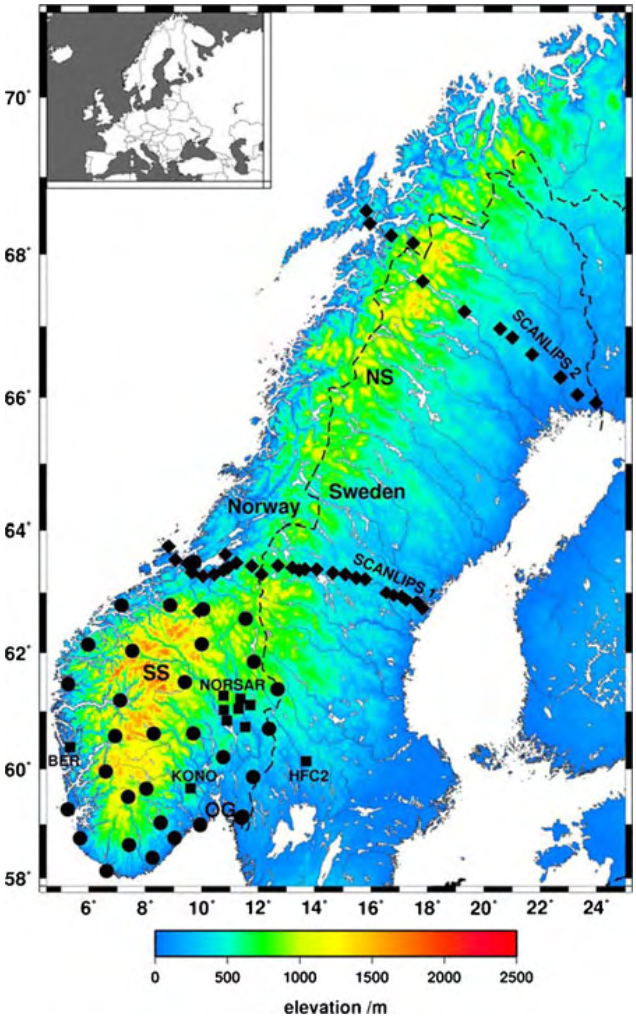


Fig. 1 Relief map displaying the study area with geographical locations and stations used (*squares* permanent broadband stations, *circles* MAGNUS temporary stations, *diamonds* SCANLIPS1 and SCANLIPS2 stations). *SS* Southern Scandes, *NS* Northern Scandes, *OG* Oslo Graben

measurements from the MAGNUS experiment (Weidle et al. 2010), Maupin (2011) confirmed the upper mantle low shear wave velocity anomaly at 150 km in depth directly underneath Southern Norway. If there exists presently an asthenospheric mantle flow from Iceland towards Norway, then one might expect a W–E fast direction for shear wave propagation inside this anomaly due to shear induced alignment of olivine crystals (Tommasi 1998). Another explanation for possible mantle anisotropy is flow of asthenospheric material around the still debated lithospheric root of the mountain range. Observed range parallel fast anisotropy is sometimes interpreted as such a flow pattern in the mantle (Nicolas 1993). The apparent plate motion direction, which is mostly parallel to asthenospheric flow, may be related also to possible anisotropic patterns. Therefore, observation of seismic anisotropy may be due to several reasons including crustal, lower lithospheric and upper mantle rock fabrics (Plomerová et al. 2008 and references therein).

The Scandinavian Mountains or Scandes at the west of the Baltic Shield are the second largest mountain range in Europe. Together the Southern and Northern Scandes form a 1,400-km-long mountain range at the NW edge of the European continent (Fig. 1). The Southern Scandes at about 59° to 63° N make up a dome-like massif, whereas the Northern Scandes form a SW–NE elongated range at about 65° to 70° N (Fig. 1). In between, the Central Scandes with lower topography are situated. Most of the Scandes' crust was built up and heavily deformed during the Scandinavian Caledonide orogeny 440–420 Ma ago during the collision of Baltica and Laurentia (Torsvik and Cocks 2005). The highly deformed nappes, which were thrusted onto the underlying Proterozoic lithosphere, still make up the prevailing mass of the Scandes. Today, the highest peaks reach as high as 2.4–2.6 km above sea level, and there are numerous hints that Cenozoic uplift added topography to the ancient Caledonian mountain range (Smelror et al. 2007; Anell et al. 2009). Japsen and Chalmers (2000) and Anell et al. (2009) summarise widespread high topography areas, Cenozoic uplift and their geodynamic driving forces around the Northern Atlantic. However, there are two contrasting hypotheses which try to explain the present high topography of the Scandes far away from current plate boundaries and in the absence of active compressional tectonics. The first type of models comprises two or more stages of uplift starting with the initial Caledonian stage and then followed by some 300 Ma of erosion possibly close to sea level until the end of the Mesozoic. The second uplift stage then formed the present topography by a dome-like rising of the Southern Scandes in the Paleogene and Neogene (Smelror et al. 2007). Rohrman et al. (1995) proposed an asthenospheric updoming as buoyancy force for the Scandes which may be maintained by intraplate stresses. Pascal and Olesen (2009) demonstrated with integrated gravity and thermal modelling that the Cenozoic uplift of the Scandes cannot be explained only by an asthenospheric diapir.

The second type of models contains only one uplift event, the Caledonide orogeny with subsequent continuous erosion. The recently proposed isostasy–climate–erosion hypothesis (Nielsen et al. 2009)

favours this single stage uplift hypothesis, but it is debated (Lidmar-Bergström and Bonow 2009). This controversy is mainly due to our limited knowledge on the deep structure of the Scandes (see "Section 2") due to sparse data, which prevents the determination of a consistent geodynamic model.

Seismic anisotropy is caused by preferred alignment of anisotropic minerals in rocks due to deformation processes, layering of rocks or alignment of cracks as well as preferred orientation of pores in the rock matrix (Babuska and Cara 1991; Maupin and Park 2007). Anisotropy is a hint for deformation pattern in the lithosphere (Silver 1996) and asthenosphere (Savage 1999) and thus anisotropy measurements help to unravel geodynamic processes. Besides direction-dependent seismic velocity, birefringence of shear waves is observed and used to detect seismic anisotropy. Especially the splitting of mantle shear waves into orthogonally polarised waves is analysed and the recovered anisotropic structures are interpreted as indicators for geodynamic processes (Silver 1996; Savage 1999, Karato et al. 2008). We analyse the splitting of teleseismic shear waves (SKS and SKKS phases; see "Section 3") in order to determine the elastic anisotropy (apparent direction of fast shear wave polarisation, which is expressed as angle Φ against north, and the apparent delay time δt between the fast and slow shear wave arrivals) below the Scandes.

Up to now, seismic anisotropy below the Scandes is hardly known. Only some SKS phases and their birefringence analysis at station KONO were published yet (Evans et al. 2003). At KONO, the fast shear wave propagation direction is in N–S direction and the average delay time of 1–1.5 s indicates anisotropy also below the crust. To the east, in Sweden, anisotropy was studied in detail at the national Swedish network (Eken et al. 2010). Crossing crustal seismic refraction profiles across the Southern Scandes do not show any signs for anisotropy (Stratford and Thybo 2011), although the lithosphere was heavily deformed during the past collisional events (Torsvik and Cocks 2005).

In this study, we combine seismological measurements at several networks in the region and apply the transverse component minimisation method of Silver and Chan (1991) to determine the splitting parameters δt and Φ. In order to improve the signal-to-noise ratio (SNR) at some selected nearby stations, we stack the waveforms and thus achieve waveforms with higher SNR and more stable splitting results. Frequency dependent splitting is analysed in order to estimate whether small- or large-scale anisotropic regions at depth cause the recovered complicated splitting pattern. The splitting results are compared to major tectonic units of the Scandes and the Baltic Shield.

2 Regional seismic structure and station distribution

Several geophysical studies were specifically designed to search for a crustal root or other structural elements which could isostatically sustain the Scandes. Svenningsen et al. (2007) found an up to 10-km-thick crustal thickening under the Southern Scandes based on P wave receiver function modelling along two station lines. This crustal root may sustain the high topography of the mountains by an Airy-type isostasy model. Using seismic refraction measurements, Stratford et al. (2009) determined a 3-D Moho map of Southern Norway based on three new seismic refraction profiles and earlier published models. Their updated Moho map shows a clear increase of crustal thickness from the coast (about 28–30 km) to a maximum of 38–40 km under the Southern Scandes and a thinning to about 36 km underneath the Oslo Graben. This crustal model by Stratford et al. (2009) was interpreted to mean that crustal *and* mantle compensation is required to sustain the high mountains.

The Central Scandes, with up to 1-km-high mountains, do not have a crustal root (England and Ebbing 2012). For the Northern Scandes, the crustal structure is not known in detail yet.

Global mantle tomography models contain a transition from low seismic velocity underneath Norway and the Southern Scandes relative to faster seismic velocity to the east underneath the Baltic Shield (Bijwaard and Spakman 2000). A new regional P wave tomography model has a sharp boundary with low seismic velocity in the upper mantle west of the Oslo Graben and fast velocity towards east under the Baltic Shield. This difference is interpreted as thin lithosphere and shallow hot asthenosphere in the west, contrasting with cold thicker shield lithosphere in the east (Medhus et al. 2012). Using surface waves, Weidle and Maupin (2008) imaged a low S wave velocity zone reaching from the Iceland plume towards Norway and below Southern Sweden. Maupin (2011) confirmed the presence of a low S wave anomaly at ~150 km in depth directly underneath Southern Norway.

None of the cited P or S wave studies found any hint for seismic anisotropy. The crossing seismic refraction models as well as the P and S wave mantle models are all based on isotropic seismic wave propagation, and it seems that anisotropy is not required to explain the travel time observations. Especially the high resolution seismic refraction models, which cover three different azimuths, have the same dominant seismic velocity values at different crustal layers (Stratford et al. 2009). Lateral velocity variations can be explained with lateral heterogeneities correlating with known crustal units. In Norway, only Evans et al. (2003) studied anisotropy using automatic teleseismic shear wave splitting at the permanent station KONO in Southern Norway (see also http://www.isc.ac.uk/SKS/index2.html). They recovered a N–S fast shear wave propagation direction and a split time of 1–1.5 s which requires a significant mantle contribution. In east of Norway, Eken et al. (2010) analysed teleseismic body waves at the Swedish National Seismic Network. They recovered at least four different domains of upper mantle anisotropy. These domains are related to different lithospheric blocks of the Baltic Shield and are interpreted as fossil anisotropy representing early plate tectonics. In "Section 6", our results are compared with the results of Eken et al. (2010).

Our main data source is the MAGNUS experiment (*MA*ntle investi*G*ations of *N*orwegian *U*plift *S*tructures) which is described in detail by Weidle et al. (2010). Within MAGNUS, 30 mobile broadband recording stations of the KArlsruhe BroadBand Array were installed in and around the Southern Scandes from September 2006 until May 2008. The mobile stations were placed in such a way to fill gaps between the few permanent broadband stations (Fig. 1): KONO (Kongsberg), BER (Bergen), HFC2 (Hagfors) and seven stations of the Norwegian Seismic Array (NORSAR). In 2009, the MAGNUS dataset was integrated into the EUROCORE programme TOPO-EUROPE of the European Science Foundation and therein specifically into the Collaborative Research Project *TopoScandiaDeep*. *TopoScandiaDeep* aims at developing a geophysical model for the lithosphere–asthenosphere system under Southern Norway, explaining the mechanisms that led to mountain formation far away from plate boundaries. Within *TopoScandiaDeep*, we could add selected teleseismic recordings from the SCANdinavian LIthosphere P and S wave (SCANLIPS1 and 2) experiments (recording times April–October 2006 and July 2008 to September 2009) across the Central and Northern Scandes (Fig. 1). SCANLIPS was organised by Leicester University (UK) and the Geological Survey of Norway (England and Ebbing 2008, 2012). More details on the seismic stations are listed in Table 1.

3 SKS and SKKS waveforms

We selected those earthquakes which should emit potential SKS and SKKS phases during the available recording time periods of the different stations. As selection criteria, we chose 80–140° as epicentre distance and a moment magnitude M_w of at least 6. In a first step, the recordings were visually inspected to search for SKS and SKKS phases with a sufficient SNR of about 3–4 on the horizontal components. Recordings without visible or very weak SKS or SKKS phases were removed from the dataset at this early stage. The finally analysed events are displayed in Fig. 2 and listed in Table 2. The suitable events occurred in four preferred backazimuth (BAZ) ranges: South America (red symbols in Fig. 2, BAZ~240–266°), Hawaii (green, BAZ~345°), South of Fiji (black, BAZ~9°) and NW Pacific (blue, BAZ~40–83°).

The basic data processing for the SKS splitting analysis comprises the following steps: deconvolution of the individual instrument response function, bandpass filtering from 5 s to 15 s, integration to ground displacement, rotation to radial (R) and transverse (T) components with the theoretical BAZ and slowness values and extraction of time windows with the SKS and SKKS phases. In Fig. 3, we present a waveform example from the MAGNUS dataset as a distance-dependent record section. On the R seismograms (Fig. 3a), we can clearly distinguish four main teleseismic phases (SKSac, SKKSac, ScS and SP) which arrive with different slowness values. The T seismograms (Fig. 3b) contain only three clear phases (SKS, SKKS and ScS). The observation of SKS and SKKS phases on the T component is an indication for anisotropy in the mantle, but wave scattering at heterogeneities may be another mechanism. The wavelet on the T component is similar to the derivative of the R signal (e.g. Fig. 4a). Our modelling in the succeeding discussion shows that the wavelets on the T component can be best explained by generation by birefringence of an S wave that is originally polarised only in the radial direction (due to P-to-S conversion at the core mantle boundary). The missing SP phase on the T

Table 1 Recording stations used in this study. Sensors since 2006: STS-2: Streckeisen STS-2; CMT-40 T, CMG-3 T, CMG-6TD: Güralp Systems; KS2000, KS54000 (borehole): Geotech Instruments

Station code	Latitude (°)	Longitude (°)	Elevation (m)	Sensor
Permanent				
HFC2	60.13	13.69	297	STS-2
BER	60.38	5.33	21	STS-2
KONO	59.65	9.59	−124	CMG3-T
NORSAR				
NAO01	60.84	10.89	426	KS54000
NBO00	61.03	10.78	529	KS54000
NB201	61.05	11.29	613	KS54000
NC204	61.28	10.76	851	KS54000
NC303	61.23	11.37	401	KS54000
NC405	61.11	11.72	496	KS54000
NC602	60.74	11.54	305	CMG3-T
MAGNUS mobile				
NWG01	63.49	9.74	88	STS-2
NWG03	62.78	7.15	181	STS-2
NWG04	62.78	8.88	215	STS-2
NWG05	62.72	10.04	564	STS-2
NWG06	62.56	11.55	842	STS-2
NWG07	62.13	5.97	43	STS-2
NWG08	62.14	9.99	822	CMG-40 T
NWG09	62.03	7.53	938	CMG-40 T
NWG10	61.84	11.85	691	STS-2
NWG11	61.39	12.68	441	KS2000
NWG12	61.51	9.39	672	STS-2
NWG13	61.19	7.10	19	STS-2
NWG14	61.48	5.26	17	STS-2
NWG15	60.58	6.92	174	STS-2
NWG16	60.62	8.29	613	STS-2
NWG17	60.62	9.69	511	KS2000
NWG18	60.70	12.38	212	STS-2
NWG19	59.85	11.82	289	KS2000
NWG20	59.64	8.03	829	KS2000
NWG21	59.95	6.59	410	STS-2
NWG22	59.27	5.24	66	STS-2
NWG23	59.49	7.39	834	STS-2
NWG24	59.03	8.54	398	KS2000
NWG25	58.99	9.94	25	KS2000
NWG26	59.14	11.44	131	STS-2
NWG27	58.75	9.04	102	STS-2
NWG28	58.62	7.42	343	STS-2
NWG29	58.74	5.68	24	STS-2
NWG31	58.38	8.24	164	STS-2
NWG32	60.21	10.75	567	STS-2
NWG33	60.62	9.69	481	KS2000

Table 1 (continued)

Station code	Latitude (°)	Longitude (°)	Elevation (m)	Sensor
NWG34	59.12	11.39	81	STS-2
NWG35	59.49	7.39	839	STS-2
SCANLIPS1				
N6001	63.74	8.83	19	CMG-6TD
N6002	63.53	9.06	56	CMG-6TD
N6003	63.47	9.49	0	CMG-6TD
N6004	63.33	9.65	305	CMG-6TD
N6005	63.27	10.04	69	CMG-6TD
N6006	63.29	10.44	158	CMG-6TD
N6007	63.37	10.77	179	CMG-6TD
N6008	63.41	10.93	150	CMG-6TD
N6009	63.48	11.23		CMG-6TD
N6010	63.43	11.77	290	CMG-6TD
N6011	63.30	12.12	631	CMG-6TD
N6013	63.44	12.70	478	CMG-6TD
N6014	63.40	13.21	565	CMG-6TD
N6016	63.37	13.45	388	CMG-6TD
N6017	63.38	13.66	531	CMG-6TD
N6018	63.38	14.07	409	CMG-6TD
N6019	63.32	14.64	281	CMG-6TD
N6021	63.24	15.47	351	CMG-6TD
N6022	63.22	15.79	389	CMG-6TD
N6024	62.99	16.54	222	CMG-6TD
N6025	62.95	16.81	345	CMG-6TD
N6026	62.93	17.09	377	CMG-6TD
N6027	62.88	17.24	294	CMG-6TD
N6028	62.84	17.62	246	CMG-6TD
N6029	62.74	17.82	101	CMG-6TD
N6130	62.69	9.90	675	CMG-6TD
N6132	63.62	10.84	146	CMG-6TD
N7001	68.4	15.95	15	CMG-3 T
SCANLIPS2				
N7001	68.40	15.95	15	CMG-3 T
N7002	68.56	15.83	8	CMG-3 T
N7003	68.25	16.72	60	CMG-3 T
N7005	68.15	17.49	628	CMG-3 T
N7010	67.63	17.83	463	CMG-40 T
N7014	67.21	19.31	404	CMG-3 T
N7017	66.97	20.57	419	CMG-3 T
N7018	66.85	21.01	375	CMG-3 T
N7020	66.61	21.71	396	CMG-3 T
N7023	66.28	22.71	0	CMG-3 T
N7025	66.04	23.31	70	CMG-40 T
N7027	65.92	23.97	50	CMG-3 T

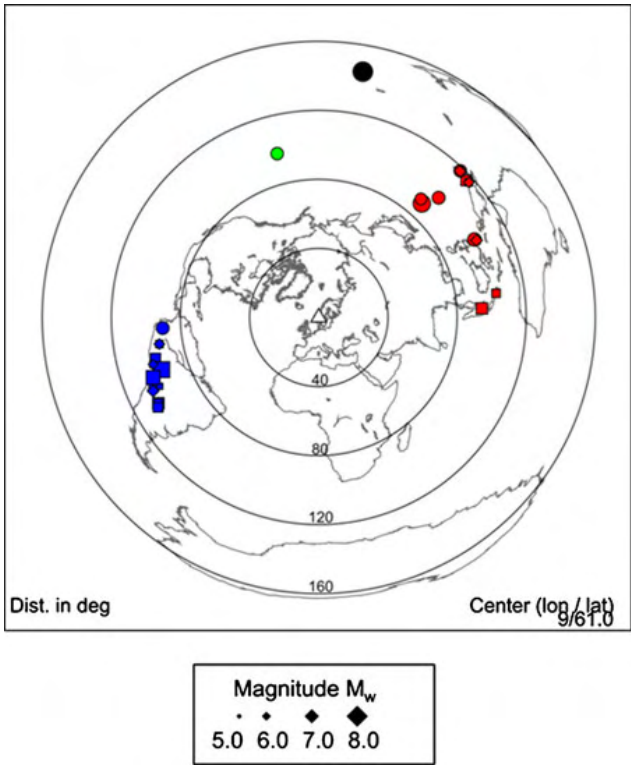


Fig. 2 Epicentre distribution of all analysed events. The *colour code* indicates similar backazimuthal regions relative to the study area (*triangle* in the centre)

component in Fig. 3b indicates the successful component rotation routine because SP should arrive as a radially polarised wave only. During our data analysis, slowness was measured for each SKS and SKKS phase and compared to the theoretical value of the *iasp91* reference Earth model (Kennett et al. 1991) in order to verify the identification of the phase.

The SNR of the split SKS and SKKS phases on the T component (Fig. 3b) is about 3–4 on average, which is enough for a stable splitting analysis (Vecsey et al. 2008). The SKS and SKKS recordings of the other events were plotted in a similar way and only clearly visible split phases were further analysed.

4 Methodology and splitting analysis

4.1 Splitting measurement

We applied the minimisation method by Silver and Chan (1991) to analyse the splitting properties of the observed SKS and SKKS phases. In a layered homogeneous isotropic Earth, SKS and SKKS phases should be polarised in radial direction (R) only because these phases are generated by compressional waves leaving the outer core. Along their travel paths through an anisotropic mantle, the SKS and SKKS phases are split and thus a fraction of their amplitudes appears on the transverse component (T). This splitting can be identified in the T seismograms (Figs. 3b and 4a) or in the particle motion diagram as an elliptical polarisation in the R–T plane (Fig. 4a). Due to the seismic velocity anisotropy, the two split waves propagate with slightly different shear wave velocities, and thus a split time or delay time δt evolves between both waves. Under the assumption of a single homogeneous anisotropic layer with transverse isotropy, the length of δt depends on the strength of the anisotropy, the wave propagation direction relative to the direction of fast anisotropic velocity and the length of the travel path inside the anisotropic region. The angle Φ is defined as the angle between the direction of fast shear wave polarisation and north. In the case of two or more anisotropic layers, δt and Φ become backazimuth and frequency dependent and only apparent δt and Φ can be measured.

The method by Silver and Chan (1991) determines the two splitting parameters δt and Φ by correcting for the effect of anisotropy. The aim is to find those δt and Φ values which minimise the SKS or SKKS amplitude on the transverse component (short: transverse component minimisation). Using a cross-correlation and a grid search in the δt–Φ domain (Fig. 4d), synthetic seismograms are calculated and the energy (squared amplitude) of the anisotropy-corrected T seismogram $(\mathrm{Tc})^2$ is determined (Fig. 4b). $(\mathrm{Tc})^2$ is plotted in a δt–Φ diagram which should reach a minimum for the correct δt and Φ parameters (Fig. 4d). We chose 3 s as upper limit for δt because greater values are unrealistic. As control, the polarisation in the R–Tc plane should become linear in radial direction for the corrected Tc waveform (Fig. 4b). We considered only those splitting measurements as useful results which show a clear linear particle motion in the R–T plane after correction and a well-determined minimum in the grid search. Based on the correct δt and Φ anisotropy parameters, a synthetic split T component seismogram can be computed and compared to the measured T recording (Fig. 4c). This method was used in numerous studies (Savage 1999) and is quite robust as long as the SNR at the T component is larger than 2. For a comparison with other methods, see Long and van der Hilst (2005) or Vecsey et al. (2008).

Table 2 Event list with SKS/SKKS splitting results at 5–15-s dominant period. M_w is the moment magnitude; event details are taken from the International Seismological Centre

Date (day month year)	Time (hh:mm:ss)	Latitude (°)	Longitude (°)	Depth (km)	M_w	Region
23 06 1991	21:22:29	−26.80	−63.35	558	7.1	Argentina
06 07 1991	12:19:49	−13.11	−72.19	104	7.0	Peru
10 05 1994	06:36:28	−28.50	−63.10	600	6.9	Argentina
09 06 1994	00:33:16	−13.84	−67.55	631	8.2	Bolivia
19 08 1994	10:02:52	−26.64	−63.42	563	6.5	Argentina
28 11 1997	22:53:42	−13.74	−68.79	586	6.7	Peru–Bolivia border
28 09 1998	13:34:30	−8.19	112.41	151	6.6	Java
23 04 2000	09:27:23	−28.31	−62.99	608	7.0	Argentina
27 07 2003	11:41:27	−20.13	−65.18	345	6.0	Bolivia
25 07 2004	14:35:19	−2.43	103.98	582	7.3	Sumatra
13 06 2005	22:44:34	−19.99	−69.20	115	7.8	Chile
17 11 2005	19:26:54	−22.37	−67.93	154	6.8	Chile–Bolivia border
25 08 2006	00:44:46	−24.40	−67.03	184	6.6	Chile–Argentina border
15 10 2006	17:14:12	20.12	−156.16	20	6.0	Hawaii
17 10 2006	01:25:12	−5.91	151.03	32	6.6	New Britain region
30 01 2007	21:37:44	21.07	144.86	20	6.6	Mariana Islands
26 07 2007	05:40:16	2.85	127.51	25	6.9	Halmahera
28 09 2007	13:38:59	21.99	142.71	275	7.5	Volcano Islands
16 11 2007	03:13:00	−2.36	−77.84	123	6.7	Peru–Ecuador border
09 12 2007	07:28:20	−26.16	−177.34	152	7.7	South Fiji Islands
09 05 2008	21:51:30	12.51	143.18	76	6.6	Mariana Islands
08 07 2008	09:13:07	−15.99	−71.75	123	6.2	Peru
26 08 2008	21:00:37	−7.64	−74.38	154	6.4	Peru–Brazil border
30 08 2008	06:54:08	−6.15	147.26	75	6.3	E New Guinea region
11 09 2008	00:00:00	1.88	127.36	96	6.5	Halmahera
12 07 2009	06:12:47	−15.04	−70.44	198	6.1	Peru

4.2 Frequency dependence

Our splitting results partly vary at single stations for different backazimuths as well as between nearby stations in the Southern Scandes. Such variations may be related to frequency-dependent effects: small-scale heterogeneity (Grechka and McMechan 1995) or depth-dependent anisotropy (Rümpker et al. 1999). Synthetic finite frequency studies show that the sensitivity region for anisotropic structure increases with SKS wave length and hence SKS period (Sieminski et al. 2008). Therefore, we determine splitting parameters by varying the frequency bandwidth of the SKS waveforms. This analysis can be done best at station KONO which provides most recordings with a good SNR. The tested period band 5–50 s corresponds to wavelengths of about 20–200 km in the upper mantle.

At first, the frequency or period content of the SKS recordings is tested by looking at the waveforms of the split phase of the T component (Fig. 5). In the unfiltered recording, the split SKS phase can be clearly identified (Fig. 5a), including a long period noise phase (about 80 s) and numerous short period arrivals (<5 s). The short period range of 5–10 s is displayed in Fig. 5b which has a clear transverse energy on the SKS waveform. At shorter periods, the amplitude of the SKS phase is too small for a splitting analysis. At periods of 5–15 s, the SKS phase is best observed in this example as well as for most other events and stations (Roy 2010). Therefore, the 5–15-s period band is selected for standard analysis in this paper.

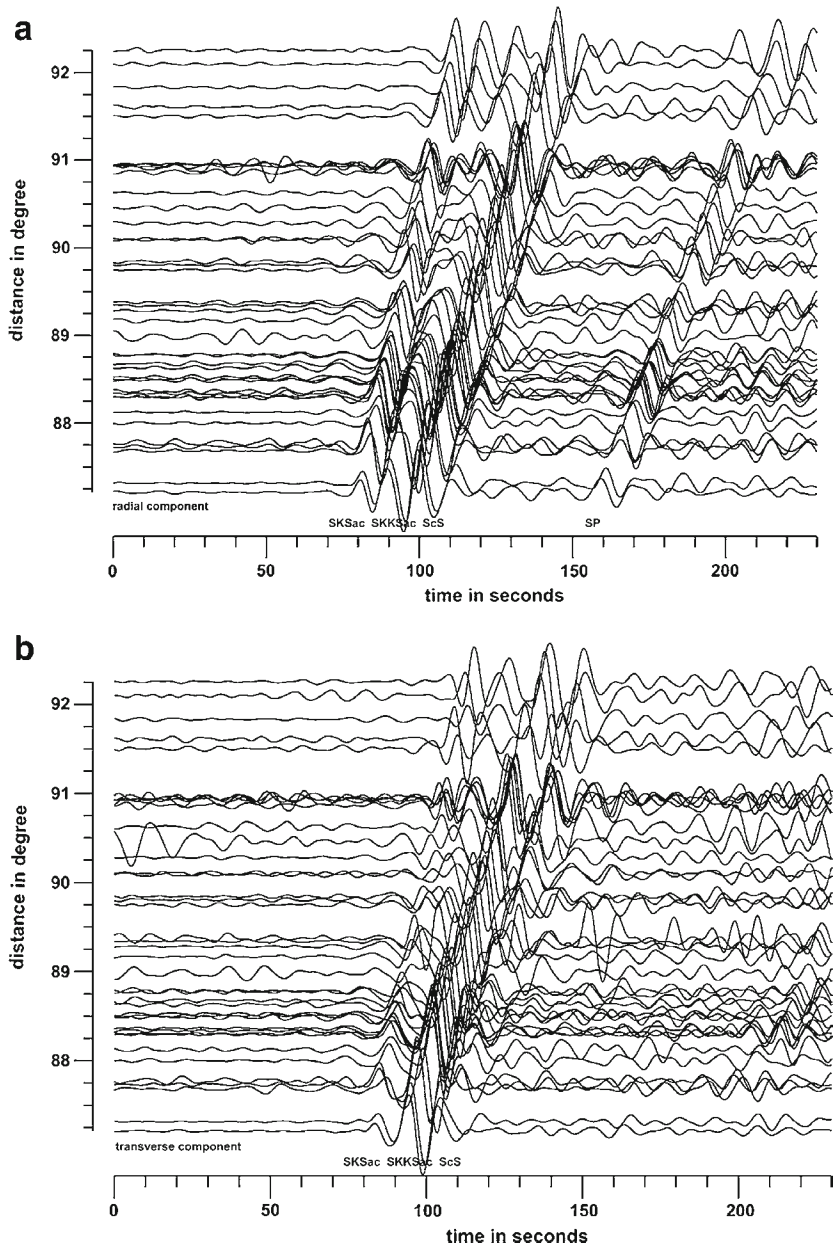


Fig. 3 Record section from the MAGNUS network. Shear wavefronts of the Mariana Islands region earthquake, 28th September 2007 (275 km in depth, m_b=6.7, M_w=7.7), are shown in the period band 5–15 s with trace normalisation: **a** radial component with four main phases, **b** transverse component with split transverse SKS and SKKS waves

SKS is still visible in the 10–50-s period range (Fig. 5d); however, for periods longer than 15 s, SKS has a small amplitude (Fig. 5e), especially for periods exceeding 20 s (Fig. 5f). Splitting parameters were determined for 5–10, 5–15 and 15–50 s recordings (Fig. 6). For KONO, we find that the apparent fast polarisation direction Φ changes systematically anticlockwise from +9° (Fig. 6a) to +5° (Fig. 6b) and −9° (Fig. 6c). These changes in Φ are smaller than the error bars of about 3–4° and much smaller than the variation that we find among neighbouring stations (see Fig. 8). The delay time δt increases systematically with increasing period from 1.2 s (Fig. 6a) to 1.5 s (Fig. 6b) and 2.5 s (Fig. 6c). This measured δt increase with period is significant and also known from other SKS analyses in New Zealand (Marson-Pidgeon and Savage 1997), Japan (Long and van der Hilst 2005; Wirth and Long 2010) or the Gulf of California (Long 2010); further results and interpretation are given in "Section 6".

4.3 Array processing

Array methods are used to measure the backazimuth and slowness of a wavefield. Based on these parameters, stacking of individual waveforms from neighbouring seismic stations can be done in order to improve the SNR of small coherent amplitude phases (Rost et al. 2006). Here we try such a data processing approach to

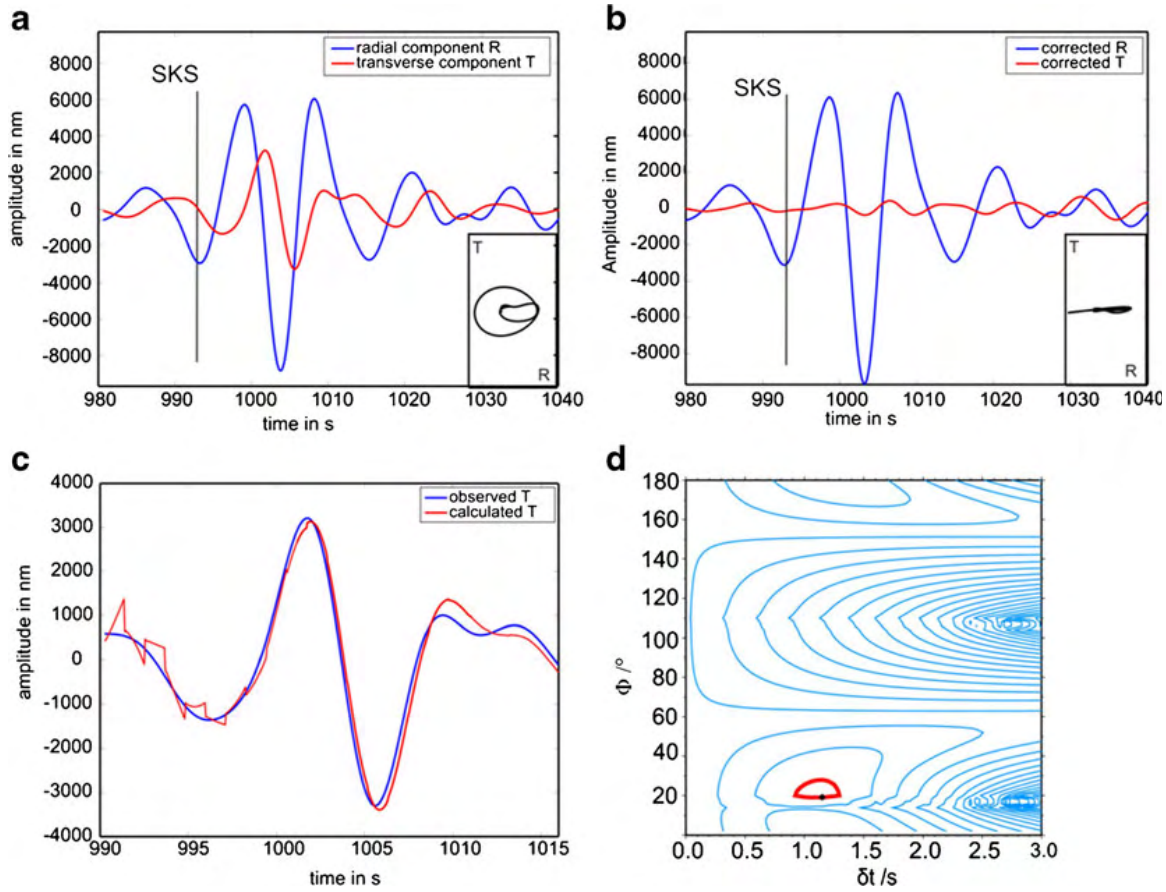


Fig. 4 Determination of splitting parameters at station KONO using waveforms of an earthquake underneath Argentina on 23rd June 1991 (556 km in depth, m_b=6.1). **a** Measured waveforms, **b** corrected waveforms using determined splitting parameters, **c** comparison between observed and calculated transverse component, **d** minimisation function for the transverse component; the minimum is found at δt=1.2±0.3 s and Φ=19±4° (*red line* indicates 95 % confidence region)

SKS splitting in a similar way as that of Gledhill and Gubbins (1996). Waveform stacking before the splitting analysis requires that very similar waveforms are observed at neighbouring stations and that the anisotropic structure does not vary below these stations in order to avoid waveform blurring or even deterioration of the splitting parameters.

The SCANLIPS1 station line across Northern Norway and Sweden (Fig. 1) is ideal to apply array processing for SKS phases because the stations are close together (about 15–20 km in distance; Figs. 1 and 9a) which is less than the dominant wavelength of 20–60 km of the analysed SKS and SKKS recordings. Across the SCANLIPS1 line, the few splitting results at single-station measurements do not change much between neighbouring stations (Fig. 9a; Table 3); however, due to low SNR, these results must be interpreted with caution. In order to derive more stable splitting measurements, we improve the SNR of the SKS and SKKS signals by stacking the R and T seismograms of neighbouring stations.

Figure 7 displays the SKS and SKKS waveforms across the SCANLIPS1 line for an event underneath the Chile–Argentina border region (25 August 2006, 00:44:47 UT). The recordings in Fig. 7a, d display the single (unstacked) seismograms of the R and T components, respectively. In Fig. 7b, c, we show the R component of the SKS and SKKS phases after stacking and alignment to the SKS and SKKS phases for better visibility.

Stacking was performed in the following way: first, we corrected the move-out of the recordings with the theoretical slowness values, which are 4.3 s/° for SKS and 6.9 s/° for SKKS, and the backazimuth of 249°. Then, the recordings of three neighbouring stations were stacked along the recording line in a gliding manner. Stations N6005, N6030 and N6032 were not used for this stacking method because the recordings

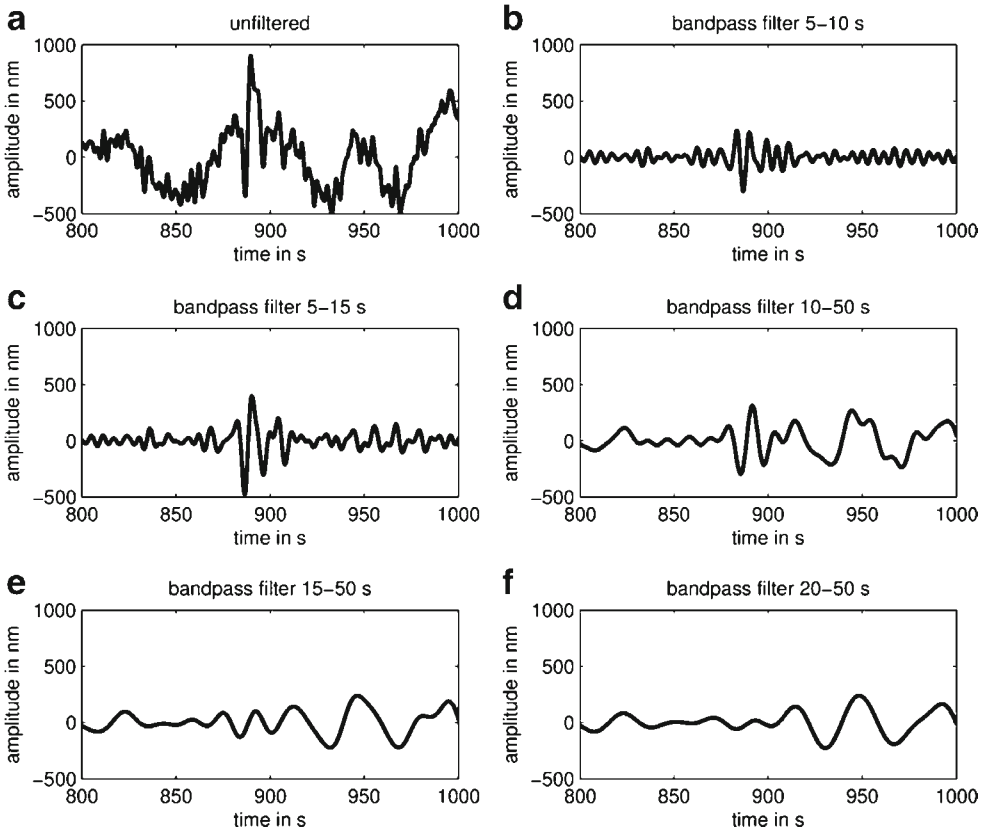


Fig. 5 SKS waveform on the tangential component of a 200-km-deep earthquake underneath Peru (12 July 2009, 06:12 UT, m_b=5.7, M_w=6.1) at station KONO. **a** Unfiltered recording, **b** bandpass-filtered 5–10 s, **c** 5–15 s, **d** 10–50 s, **e** 15–50 s, **f** 20–50 s

of N6005 were corrupt and the other two stations are not in line with the others. The station names given in Fig. 7b, c refer to the middle station of the three stacked recordings.

These stacked transverse waveforms are displayed in Fig. 7e, f. The stacked seismograms indicate that there is a clear SKS signal on the T recordings and hence we can expect anisotropy below the Central Scandes as well. The stacked seismograms gain more clear split SKS and even SKKS measurements due to the better SNR compared to the single measurements (Fig. 7a, d). Thus, stacking SKS or SKKS waveforms can allow a more detailed study of anisotropy compared to single recordings if suitable nearby stations are available.

5 Results

Our splitting analysis provides 154 results (Φ–δt pairs) from clearly split SKS or SKKS waveforms (Table 3; Figs. 8 and 9). We removed noisy data as well as possible null splits because the latter could not be clearly distinguished from noisy waveforms. In general, the presented Φ values are well resolved mostly within ±20° (2–σ level), while δt has quite large error limits of about ±1 s.

5.1 Station KONO

Station KONO in Southern Norway provides the most useful waveforms for our splitting analysis (Table 3). The apparent Φ is oriented N–S with a slight change toward NNW/SSE for increasing wave period (Fig. 6). The apparent split time δt increases with period: it is mostly below 1.4 s for periods of less than 10 s and reaches ~3 s for periods above 15 s (Fig. 6). The splitting analyses with the 12 SKS recordings from events below South America provide very reproducible results (blue arrows in Fig. 8). Splitting results based on one event below Java (red arrow, Φ~19°) and one event below Fiji (black arrow, Φ~37°) have a NE–SW tendency. The splitting results in the database of the International Seismological Centre (http://www.isc.ac.uk/SKS/; accessed 1 July 2012) have a NNE–SSW direction of Φ which coincides with our measurements.

5.2 Southern Norway

At the MAGNUS network in Southern Norway, the measured Φ directions appear quite variable at first sight (Fig. 8). Splitting results for events from the

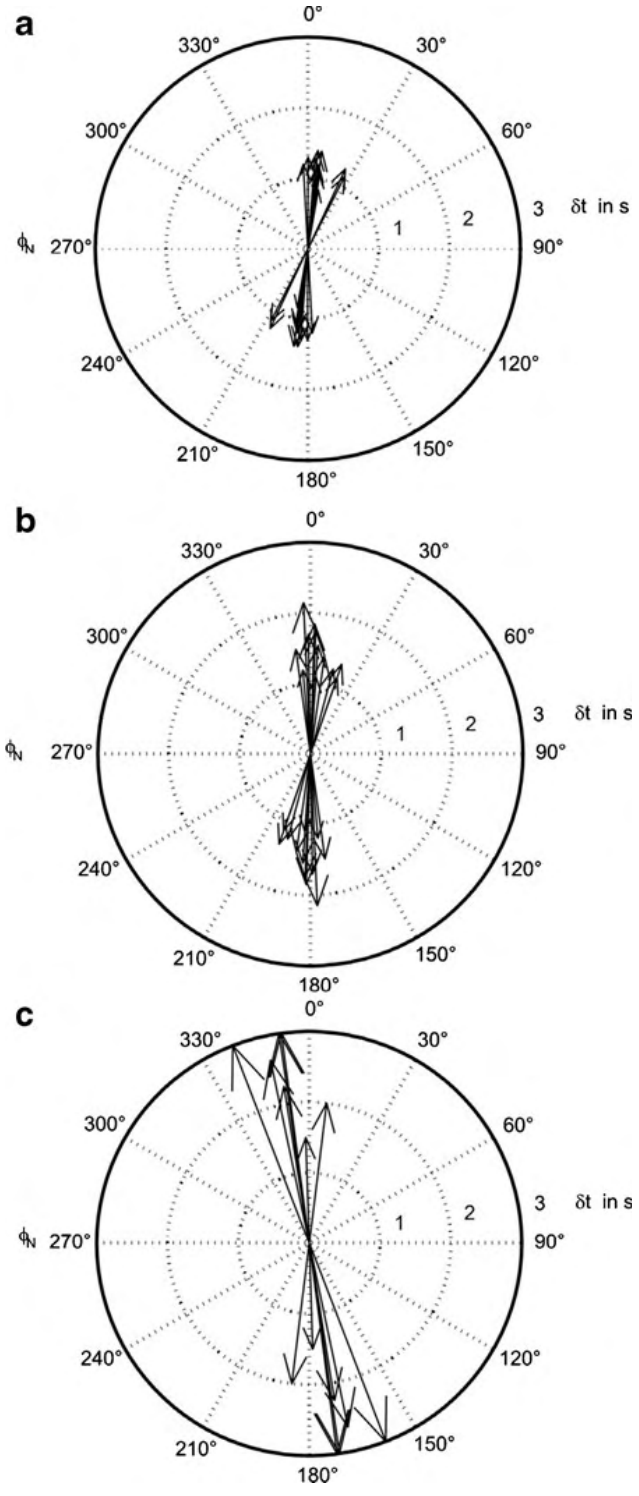


Fig. 6 Splitting results at KONO for different period bands: **a** 5–10 s, **b** 5–15 s, **c** 10–50 s

same backazimuth region are displayed with the same colour coding as the epicentre distribution in Fig. 2. Φ values measured for events with a similar epicentre backazimuth are mostly consistent at the same station —thus, it is inferred that the splitting measurements are stable and reproducible. At some stations, Φ varies with backazimuth (e.g. stations NWG01, 03, 04, 19, 22, 27, 29, 31 and NORSAR) and at a few stations Φ is less dependent on the backazimuth (NWG04, 18, 21, 30 and KONO). However, better backazimuthal coverage is needed to confirm the latter cases.

In the north of the MAGNUS network, station NWG01 has three different apparent Φ for three different backazimuths and station NWG03 has at least two different Φ, indicating a complex anisotropic underground. The apparent split time δt is above 1 s at most stations, but there is no systematic trend in the spatial distribution of δt in Southern Norway. The value of δt is mostly similar for different SKS phases from the same epicentre backazimuth but varies between different stations.

5.3 Central Scandes and Sweden

In the transitional region between the Southern and Northern Scandes, the splitting results are relatively homogeneous across the SCANLIPS1 line (Fig. 9a, b). We were able to determine nine splitting results from single-station measurements and 26 additional results after the gliding station stack (see "Section 4.3"). In the western part of SCANLIPS1, close to the coast, apparent Φ is oriented N–S (stations N6002-N6004). Just to the east, Φ is directed NE–SW at stations N6006–N6009. To the east of 11°E, Φ is uniform in NW–SE direction across the low mountains and their eastern foreland in Sweden. Only one SKKS splitting measurement at N6025 has Φ with NE–SW direction. Along the SCANLIPS1 line, Φ does not vary much between SKS (dark blue) and SKKS (light blue) measurements, indicating a possible common birefringence origin in the upper mantle and/or crust. The apparent split time δt is generally high, with many values exceeding 2 s. Although δt is not determined very precisely (typical uncertainties of up to 1 s), these high δt values are very uniform for a South American event (blue in Fig. 9a, b). The SKS and SKKS splitting results from the event under the New Britain region (red and orange) are more E–W-oriented with a smaller δt, indicating complex anisotropic layer geometry. The smooth results may become more complex when observations from more backazimuths can be obtained as in Southern Norway.

Table 3 SKS/SKKS splitting results from this study. The results for station KONO were measured at 5–15-s period as for the other stations

Station code	Event date	δt (s)	Error δt (s)	Φ (°)	Error Φ (°)	Phase
Permanent						
BER	09052008	1.1	0.7	20	12	SKS
HFC2	28092007	1.0	1.3	18	18	SKS
KONO	09122007	1.9	0.9	37	13	SKKS
	15102006	1.8	1.1	2	10	SKS
	12072009	1.9	0.7	2	8	SKS
	17112005	1.6	1.0	4	15	SKS
	23042000	1.5	0.7	172	13	SKS
	27072003	1.3	0.4	15	18	SKS
	28111997	1.1	0.7	4	15	SKS
	28091998	1.4	0.6	19	10	SKS
	08072008	1.8	1.8	2	10	SKS
	23061991	1.2	0.3	19	3	SKS
	19081994	1.4	0.4	8	15	SKS
	09061994	1.7	1.0	179	10	SKS
	10051994	1.2	0.5	175	13	SKS
	13062005	2.2	1.0	177	13	SKS
NORSAR						
NAO01	09122007	2.1	0.7	27	4	SKKS
	28091998	1.6	1.0	90	0.4	SKS
	13062005	1.8	1.3	170	7	SKS
	23061991	2.6	1.0	162	2	SKS
NB201	28092007	0.9	1.5	2	20	SKS
	28091998	0.8	1.0	106	39	SKS
	13062005	2.7	1.4	166	2	SKS
NBO00	28092007	1.6	1.5	32	20	SKS
	28091998	0.9	0.5	154	12	SKS
	30082008	1.1	1.0	118	31	SKS
NC204	16112007	1.4	0.9	110	15	SKS
	28092007	1.4	1.5	152	20	SKS
	28091998	0.6	0.5	118	34	SKS
	12092009	2.5	0.8	85	2	SKS
	27072003	2.8	1.4	75	4	SKS
NC303	28092007	0.9	1.5	2	20	SKS
	28111997	1.1	1.5	89	42	SKS
	28091998	0.7	0.7	116	15	SKS
	25072004	1.7	1.2	157	25	SKS
	08072008	1.4	1.1	86	15	SKS
	06061991	1.2	1.5	96	43	SKS
NC405	28092007	1.7	1.4	28	20	SKS
	26082008	0.8	0.5	112	20	SKS
NC602	09052008	3.0	1.1	145	5	SKS
	16112007	2.2	1.4	8	11	SKS
	28092007	1.7	1.2	148	12	SKS

Table 3 (continued)

Station code	Event date	δt (s)	Error δt (s)	Φ (°)	Error Φ (°)	Phase
	30012007	2.8	1.4	143	6	SKS
	28111997	1.4	1.0	9	17	SKS
	08072008	2.5	2.1	176	18	SKS
	12092009	1.9	1.3	179	20	SKS
	27072003	0.8	1.4	1	35	SKS
MAGNUS mobile						
NWG01	09052008	3.0	1.5	36	5	SKS
	09122007	1.0	0.7	87	13	SKKS
	15102006	1.4	0.5	1	5	SKS
NWG03	09052008	1.1	0.9	32	5	SKS
	15102006	0.9	0.8	12	15	SKS
	17102006	1.3	1.2	143	18	SKS
	26072007	1.4	1.4	163	15	SKS
	28092007	1.6	1.4	29	20	SKS
NWG04	09122007	2.4	0.6	26	3	SKKS
	28092007	0.9	1.5	0	20	SKS
NWG05	15102006	2.2	1.1	3	17	SKS
NWG07	09052008	2.5	1.8	142	10	SKS
	28092007	1.7	0.6	78	4	SKS
NWG08	17102006	0.8	1.0	63	20	SKS
NWG09	09052008	1.7	1.3	142	10	SKS
NWG11	09052008	1.0	1.4	150	20	SKS
	28092007	1.2	0.8	5	15	SKS
	30012007	1.6	1.1	154	15	SKS
NWG12	26072007	2.9	2.7	146	20	SKS
NWG13	09052008	2.4	1.0	37	2	SKS
NWG15	09052008	3.0	0.8	31	3	SKS
NWG16	09052008	0.9	0.6	30	13	SKS
	26072007	3.0	2.0	57	3	SKS
	28092007	1.7	1.3	32	12	SKS
NWG17	17102006	2.3	2.2	37	10	SKS
NWG18	09052008	1.2	1.5	42	6	SKS
	28092007	1.9	1.6	31	19	SKS
NWG19	09122007	1.2	0.6	79	13	SKKS
	15102006	2.4	1.0	66	4	SKS
	17102006	2.2	0.7	127	9	SKS
NWG20	28092007	0.5	1.1	169	20	SKS
NWG21	09052008	1.9	0.6	23	7	SKS
	17102006	1.2	0.6	20	13	SKS
	28092007	1.3	0.6	14	11	SKS
	30012007	1.8	0.6	1	5	SKS
NWG22	06112007	2.6	0.8	163	7	SKS
	28092007	2.6	0.6	27	4	SKS
	30012007	2.3	0.3	21	1	SKS

Table 3 (continued)

Station code	Event date	δt (s)	Error δt (s)	Φ (°)	Error Φ (°)	Phase
NWG23	17102006	1.5	0.7	155	13	SKS
	26072007	1.5	1.3	170	23	SKS
	28092007	1.2	0.5	149	7	SKS
NWG24	17102006	1.1	1.4	152	20	SKS
NWG25	15102006	1.7	1.0	61	10	SKS
NWG27	09052008	1.2	1.1	147	16	SKS
	09122007	2.6	0.6	23	4	SKKS
NWG28	09052008	3.0	1.2	152	8	SKS
	30012007	2.0	0.5	159	7	SKS
NWG29	09052008	2.7	1.4	138	3	SKS
	16112007	1.9	1.9	93	20	SKS
	17102006	2.5	0.5	155	4	SKS
NWG30	09052008	1.1	0.2	167.19	12	SKS
	17102006	0.9	0.4	0.5	17	SKS
	28092007	3.0	0.5	138.1	3	SKS
NWG31	09122007	1.3	1.8	4	28	SKKS
	16112007	2.7	1.0	149	9	SKS
	30012007	3.0	1.5	30	5	SKS
NWG34	09122007	0.9	0.6	57	18	SKKS
NWG35	09052008	1.7	1.0	152	14	SKS
SCANLIPS1 unstacked						
N6002	25082006	1.9	1.2	167	9	SKS
N6003	25082006	1.3	0.5	5	12	SKS
N6008	17102006	3.0	1.5	35	5	SKKS
N6010	17102006	0.8	1.1	106	38	SKS
N6014	25082006	1.5	1.3	148	13	SKS
	25082006	2.7	1.6	154	4	SKKS
	17102006	2.7	1.1	123	14	SKS
N6017	25082006	3.0	1.5	153	4	SKS
N6026	25082006	3.0	1.4	157	3	SKS
SCANLIPS1 stacked						
N6002	25082006	1.3	0.7	173	10	SKS
	25082006	0.8	1.4	1	20	SKKS
N6003	25082006	1.6	0.9	173	7	SKS
	25082006	1.6	1.1	175	20	SKKS
N6004	25082006	1.0	0.8	175	15	SKS
	25082006	1.1	0.6	25	15	SKKS
N6006	25082006	1.5	1.3	49	15	SKKS
N6007	25082006	3.0	0.4	60	3	SKKS
N6008	25082006	1.8	1.4	150	15	SKS
	17102006	0.6	1.1	87	30	SKS
N6009	25082006	1.8	0.9	150	3	SKS
N6010	17102006	1.2	0.5	119	8	SKS
N6011	25082006	1.8	0.6	149	4	SKKS

Table 3 (continued)

Station code	Event date	δt (s)	Error δt (s)	Φ (°)	Error Φ (°)	Phase
	17102006	1.3	1.4	122	30	SKS
N6013	25082006	2.5	0.5	153	2	SKS
	17102006	1.2	1.2	118	33	SKS
N6014	25082006	2.3	0.9	152	2	SKS
	17102006	1.0	1.3	119	35	SKS
N6016	25082006	2.8	0.9	152	5	SKS
N6017	17102006	0.6	0.7	83	38	SKS
N6018	25082006	2.3	1.5	153	5	SKS
	25082006	2.9	0.7	153	14	SKKS
N6019	25082006	2.7	1.5	153	5	SKS
N6020	25082006	2.2	1.3	151	5	SKS
N6021	25082006	2.3	1.4	154	10	SKS
N6025	25082006	2.0	0.6	53	8	SKKS
SCANLIPS2						
N7001	11092008	1.2	0.9	53	15	SKS
N7003	11092008	2.0	1.4	56	14	SKS
	30082008	1.4	0.8	27	14	SKS
N7005	11092008	0.8	0.3	3	10	SKS
	30082008	3.0	1.9	156	8	SKKS
N7014	11092008	2.4	1.0	63	4	SKS
	30082008	1.3	0.8	44	15	SKKS
N7020	11092008	1.1	0.6	3	10	SKS
	26082008	3.0	1.3	11	3	SKS
N7023	11092008	1.0	0.2	50	9	SKS
	26082008	1.8	1.0	83	8	SKS
	08072008	2.5	0.7	79	2	SKS
N7025	11092008	1.1	0.4	54	9	SKS
N7027	11092008	1.2	0.6	179	8	SKS

5.4 Northern Scandes and Sweden

Fourteen splitting results obtained from four earthquakes could be determined with the SCANLIPS2 dataset. Along the SCANLIPS2 line, Φ is mainly in NE–SW direction (Fig. 9c). This Φ direction is found close to the coast (N7001, N7003) and in Sweden on the eastern part of the line. There is one significant anomaly at station N7005 with Φ~NW–SE and a long δt of 3 s from an SKKS phase of an event under New Guinea. It is unclear whether this splitting result is due to a local feature underneath station N7005 because another SKS splitting measurement (Halmahera event) has N–S orientation with a much shorter δt of just 0.8 s. Both measurements have a similar SNR on the T component. The observed splitting results in the north may become more complex when observations from more backazimuths can be included as in Southern Norway.

6 Discussion and interpretation

Our observations of split SKS and SKKS waves and their derived apparent splitting parameters are strong indicators for seismic anisotropy below the Scandes in Norway and parts of Sweden. However, the splitting parameters Φ and δt vary for different epicentre backazimuths at some stations as well as between station clusters. Such variations indicate a complex

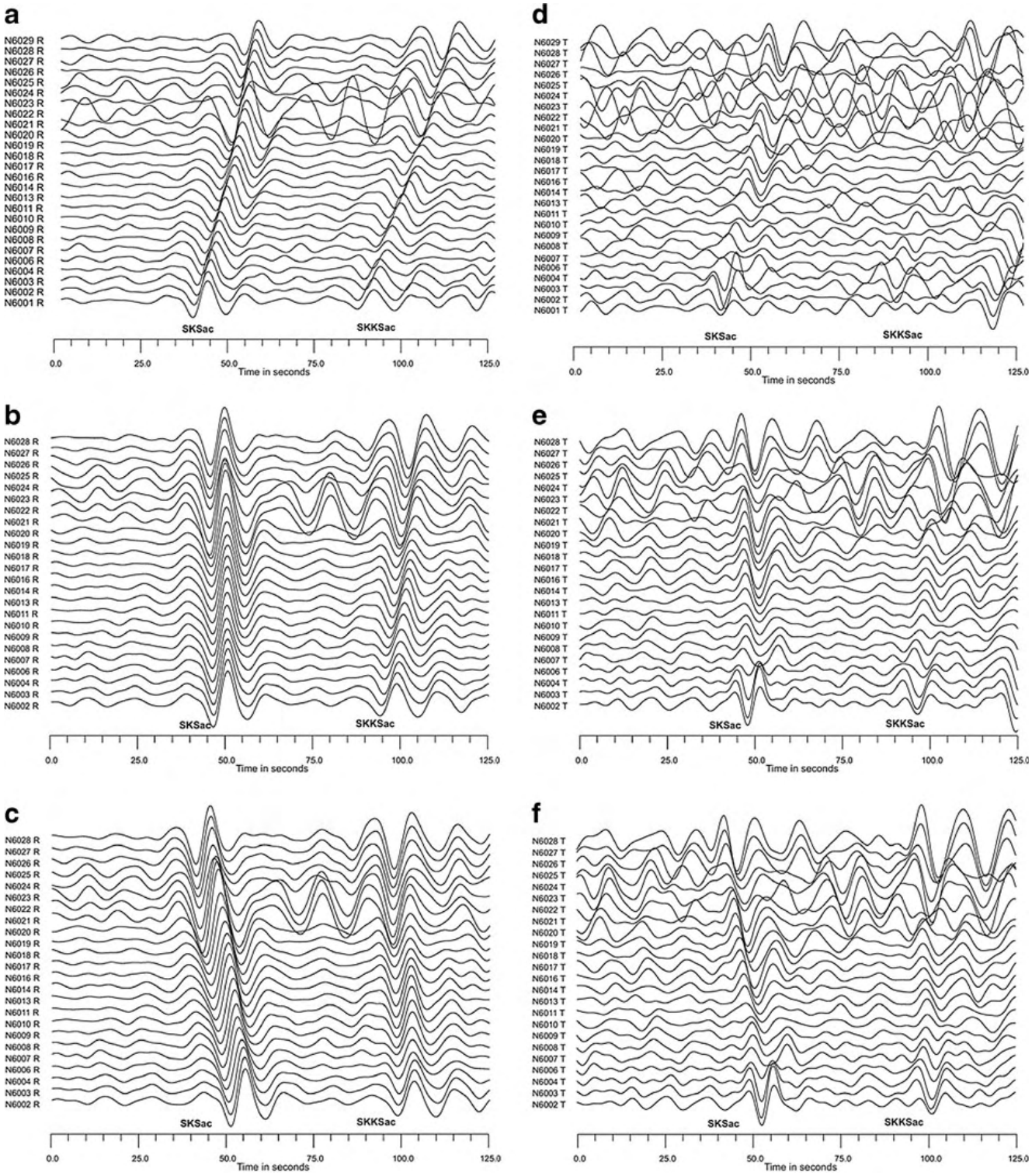


Fig. 7 SKS and SKKS waveforms at SCANLIPS1 stations of an event under the Chile–Argentina border region (25 Aug. 2006, 00:44:47 UT at 184 km in depth, M_w=6.6): **a** radial component, filtered at 5–15 Hz and amplitude normalisation, **b** stacked radial waveforms from three neighbouring stations aligned along the slowness of 4.3 s/° of the SKS phase, **c** stacked and aligned radial SKKS phase (along 6.9 s/° slowness), **d** filtered and unstacked transverse component, **e** stacked transverse SKS phase, **f** stacked transverse SKKS phase

anisotropic structure at depth. The observed Φ and δt variations do not allow an interpretation with a simple model such as *one* anisotropic layer with transverse anisotropy and *one* horizontal fast polarisation axis.

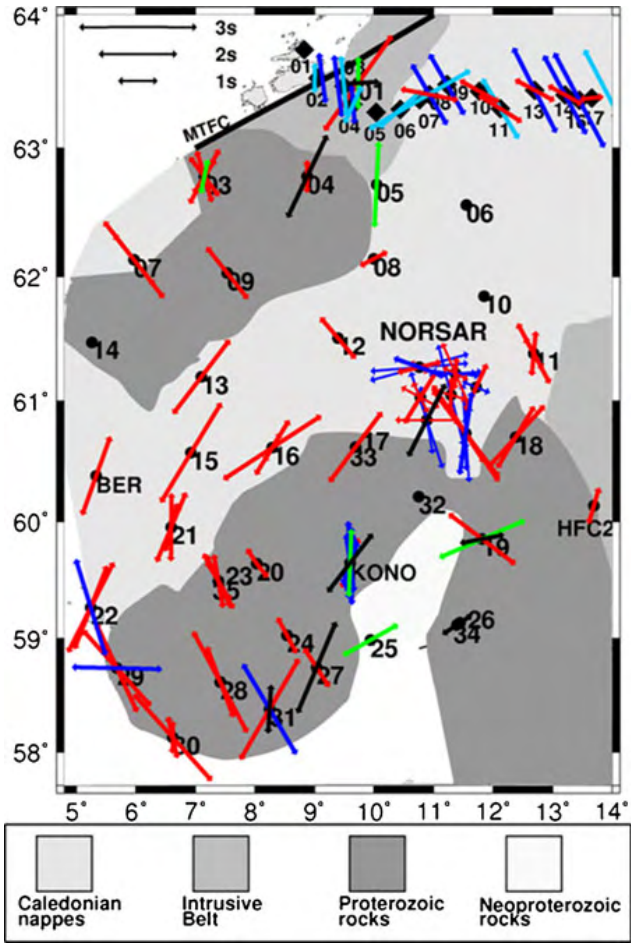


Fig. 8 Splitting results at the MAGNUS network in Southern Norway. The *colour coding* corresponds to the backazimuthal regions in Fig. 2. The main tectonic units are given after Ramberg et al. (2008). *MTFC* Møre–Trøndelag Fault Complex. Note that only the last two digits of the station names are given

More complex models are required, including lateral variation of anisotropy or isotropic crustal domains, multiple anisotropic layers, dipping fast axes, variable layer depths as well as variable layer thicknesses. However, complex anisotropic models cannot be determined uniquely with our dataset. Such complex models predict apparent Φ and δt values with strong backazimuth dependence, e.g. often including a characteristic 90° pattern (see Silver and Savage (1994) or Rümpker and Silver (1998)) and with frequency dependence (Rümpker et al. 1999). Our epicentre distribution (Fig. 2) limits such a backazimuth-dependent analysis. Even worse, most epicentres are in the NW Pacific and South America regions (Fig. 2), which have a backazimuth difference of about 180° at our stations. Such oppositely propagating shear waves should split to similar apparent Φ and δt values in horizontally transverse anisotropic structures and cannot resolve complex structures in many cases. Further constraints on seismic anisotropy are not available for our study region. Our dataset is presently the only one to describe seismic anisotropy below the Scandes.

There are several indications that our observations are mainly caused by deep-seated anisotropy in the upper mantle underneath the Scandes. First, the observed split times (δt>1 s) are too large to be explained solely with crustal anisotropy. Barruol and Mainprice (1993) argue that only as much as δt~0.3 s can be explained with reasonable and known crustal rock fabrics. Second, new, high-quality seismic refraction models (Stratford et al. 2009; Stratford and Thybo 2011) and ambient noise Rayleigh wave tomography models (Köhler et al. 2011) do not indicate that there is widespread crustal seismic anisotropy in Southern Norway. Third, the frequency bandwidth of the observed SKS splitting at station KONO indicates that wavelengths exceeding the crustal thickness are affected by splitting and that very large regions have an impact on the splitting of the SKS/SKKS wavefields. We also infer that small-scale heterogeneity does not influence much the splitting parameters at KONO because short-period (5–15 s) and long-period (10–50 s) splitting results have very similar Φ angles (Fig. 6). Fourth, there is mostly no obvious difference between SKS and SKKS splitting results from the same event at the SCANLIPS experiments (Table 3; Fig. 9). Thus, we assume that the shear wave splitting is generated along a common ray path for SKS and SKKS phases, which means that anisotropy is located within the upper mantle and the crust (Niu and Perez 2004).

6.1 Southern Norway

In the Southern Scandes, there is an obvious spatial correlation between similar splitting results and the major tectonic units at the surface (Fig. 8). The apparent directions of Φ are nearly NW–SE on the Proterozoic outcrop regions and NE–SW on the Caledonian nappes. At the boundaries between these units, the Φ pattern is more variable, for example, at the NORSAR array and at the northern stations (NWG01 and NWG03) of the MAGNUS network. This observation is an indication for laterally varying anisotropy. As argued before, crustal anisotropy alone cannot explain our observations. The backazimuthal variation of Φ indicates that two or more layers with different anisotropic fabrics or an inclined fast axis contribute to shear wave splitting. Thus, it

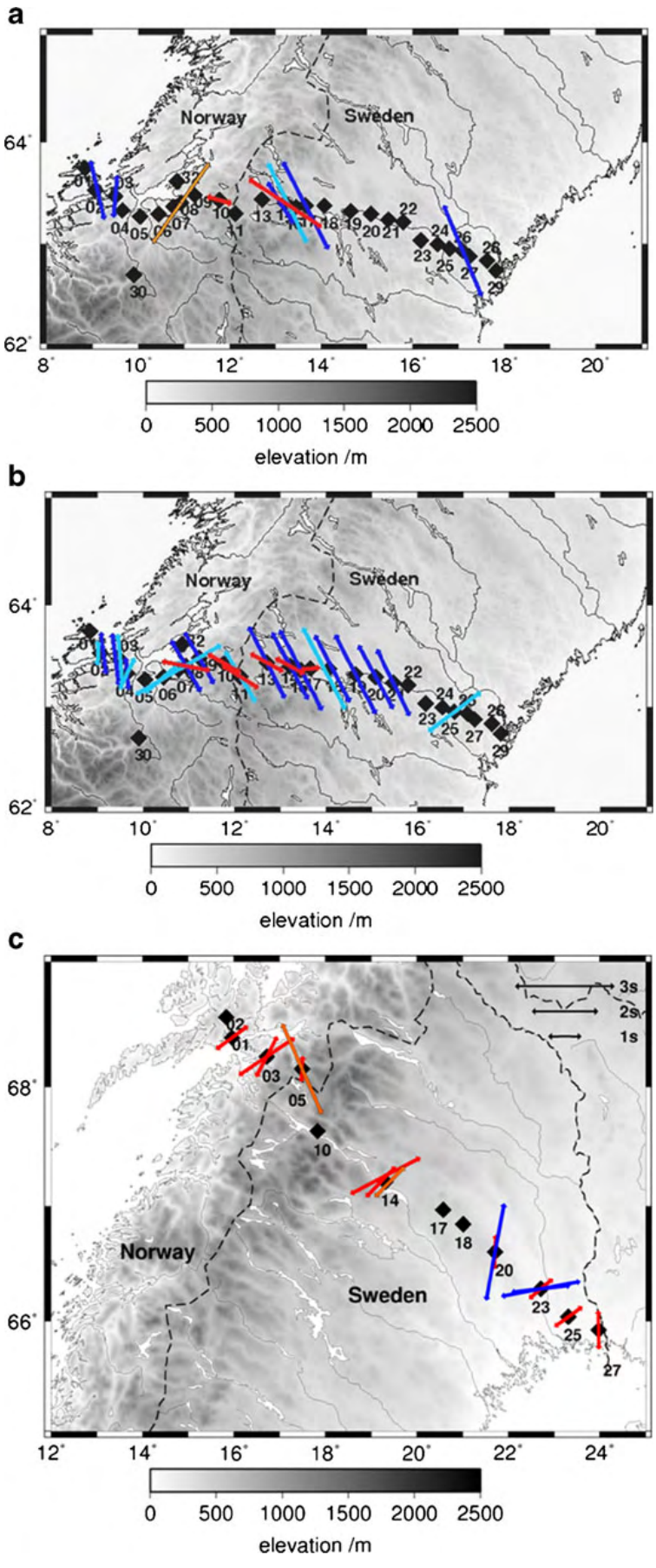


Fig. 9 Splitting results in Central and Northern Norway and Sweden. *Colours* correspond to the backazimuthal regions in Fig. 2. *Arrows* in *dark blue* (South American events) and *red* (Pacific events) are SKS results; arrows in *light blue* and *orange* (with *black boundaries*) are SKKS results from these epicentral regions. **a** Splitting results at SCANLIPS1 obtained from single-station measurements, **b** splitting results after gliding station stack, **c** results at SCANLIPS2 obtained from single-station measurements. Note that only the last two digits of the station names are given for clarity

seems that a near-surface anisotropic layer causes the finally measured direction of Φ. Such a layer could be inherited from a local tectonic event which left behind a local anisotropic rock fabric. Kaviani et al. (2011) demonstrate that even a thin isotropic layer, which changes laterally, can have a significant influence on the splitting parameters. Our observations can also indicate a multilayer model similar to the synthetic modelling scenarios by Rümpker et al. (1999). They study depth-dependent anisotropy and demonstrate that in the upper layer the fast axis orientation controls the apparent orientation of Φ (possibly crustal fabrics in our case), whereas in the lower layer the angle between the fast axis orientation and the epicentre backazimuth control the split time δt.

Our frequency-dependent splitting parameters for station KONO (Fig. 6) are similar to the synthetic waveform modelling and splitting results for models M3 and M4 in Rümpker et al. (1999). These synthetic examples contain a change in the fast axis orientation of about 50° and 60° between an upper and a lower anisotropic layer. This configuration leads to a change in apparent Φ of up to 40° relative to the true fast axis orientation in the upper layer and up to 60° (mostly <20°) between short-period (5 s) and long-period (20 s) SKS phases. However, in such cases, Φ depends on the backazimuth which we cannot resolve properly. The frequency-dependent apparent delay time δt increases with increasing period by up to 2–3 s in Rümpker et al. (1999), which is also observed in our data; this parameter is also backazimuth dependent. Our high δt of up to 3 s at KONO for events with backazimuths of 240–260° (South American epicentres) may be a hint that this propagation direction is favourable for strong shear wave splitting in a deep layer below the Southern Scandes. A unique model of anisotropy cannot be determined with the limited observed backazimuths. In the following text, only regions with different splitting characteristics are discussed.

At the stations on the Caledonian nappes in Southern Scandes (Fig. 8), the NE–SW orientation of Φ is interpreted as evidence for a considerable change of the rock fabric in the lithosphere compared to the neighbouring Proterozoic blocks where Φ is preferably NW–SE-oriented. The same Proterozoic lithosphere, and possibly trend in rock fabric, may be present below the Caledonian nappes, but the upper part of the lithosphere there was heavily deformed during the collisional processes as well as during the postcollisional collapse which again deformed it (Andersen 1998; Torsvik and Cocks 2005; Ramberg et al. 2008). These deformational processes produced rock fabrics that may now control the NE–SW orientation of Φ which is different to the surrounding older lithosphere units. The Caledonian nappes originate from an eastward nappe translation, whereas the postcollisional collapse had a top-westerly directed sense of shear (Andersen 1998). Thus, mostly NW–SE-oriented rock fabrics are found in the Caledonian mountains, which do not correlate with our fast polarisation directions. Also, some NE–SW striking extensional structures have been identified in the area of the Faltungsgraben (Andersen 1998) where our stations BER, NWG13, 15, 16, 21 and 22 were situated and which measured a NE–SW-oriented Φ (Fig. 8). The NE–SW trends postdate the more dominant top-west reactivated faults (Andersen 1998) and also other tectonic structures at depth may contribute to the splitting pattern in the Caledonian mountains. There is an uncertainty on how deep the surface tectonics and its related rock fabric reach down to depth. However, we infer that the crustal structure, such as the Caledonian nappes, has an influence on the observed SKS/SKKS splitting similar to the influence of even isotropic sediment layers (Kaviani et al. 2011). Due to the large δt values, it is obvious that shear wave splitting has an additional lower lithospheric or even asthenospheric origin, but the fast polarisation direction cannot be resolved at this depth range. Our findings are similar to the results of Bastow et al. (2007) who analyse SKS splitting in the Caledonian mountains in Northern Scotland and its surroundings. They also find variation of Φ across short station distances, which is interpreted as fossil anisotropic signature of lithospheric blocks.

The surface projection of the ray paths in the anisotropic medium are outlined in Fig. 10. For this projection, we assume a uniform 3 % anisotropy at depth to determine an estimate for the length of the SKS/SKKS ray path which is calculated with an average shear wave velocity of 4.5 km/s and the average δt value for a backazimuth region at a station site. The surface projection is then determined based on the average incidence angle of the rays. The colour coding in Fig. 10 corresponds to the apparent Φ and the map visualises the spatial distribution of Φ and δt. There are several domains with similar splitting parameters indicating small-scale coherence of anisotropy, e.g. in the area of the Caledonian nappes and the Proterozoic basement region in the south.

In the northwestern part of the MAGNUS network, the Møre–Trøndelag fault complex (MTFC; Fig. 8)

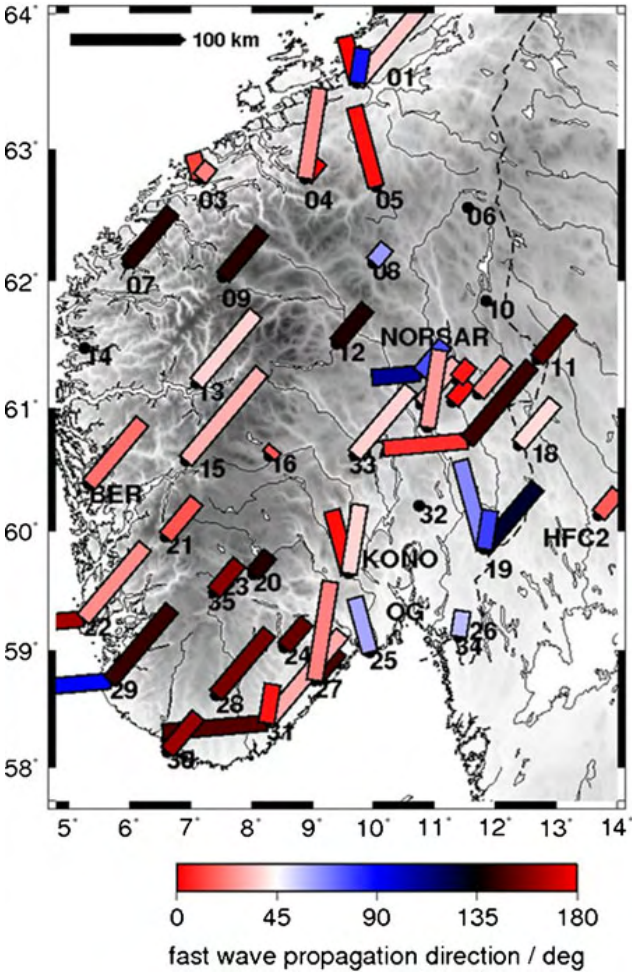


Fig. 10 Surface projection of split ray paths with fast anisotropic direction in colour. These ray path segments point into the backazimuth direction and they indicate the data coverage at depth. The segments are based on average ray paths for SKS phases from the main source regions (see Fig. 2 and text)

may play an additional role concerning deformation and generation of rock fabrics (Gilotti and Hull 1993; Redfield et al. 2005; Olsen et al. 2007). The NE–SW striking MTFC is about 50 km in width, about 300 km in length and consists partly of penetratively deformed rock masses. However, seismic anisotropy related to the MTFC has not yet been studied in detail. Our varying splitting results at stations NWG01, 03 and 04 indicate that anisotropy is complex in the vicinity of the MTFC and should be studied in more detail.

Around the Oslo Graben in southernmost Norway, the splitting results are also not unique. The ray path distribution in Fig. 10 shows that the split SKS/SKKS waves did not propagate directly underneath the Oslo Graben; however, the whole region is characterised by small-scale variations in geology, including faults and shear zones related to the rifting processes (Neumann et al. 1995). Our Φ and δt results vary at stations NWG19, 25, 27 and 34 (Fig. 8) with mainly apparent Φ in ENE–WSW direction. At other nearby stations, Φ has a more NW–SE trend in the Proterozoic lithospheric domain (Figs. 8 and 10) which belongs to the Sveconorwegina in this region. This variability in Φ may be due to locally changing dominant rock fabrics related to the rifting processes of the Oslo Graben. The Φ pattern gets even more complex at the NORSAR stations (Figs. 8 and 10) which are just at the northern end of the Oslo Graben and extend into the Caledonian nappes (named Eocambrian rocks; see Bungum et al. (1971)). At the southern stations at NORSAR, Φ is oriented mainly N–S with backazimuth-dependent deviations, whereas at the northern station at NORSAR Φ prevails in E–W direction (Fig. 8). This change possibly reflects the transition from the Caledonian nappes in the north with rock fabrics due to E–W postcollisional extension relative to the rifted Sveconorwegian lithosphere in the south. Anyhow, we interpret the small-scale variability as a significant influence of near-surface, crustal structure on the shear wave splitting results.

In southern Sweden, at station HFC2, we find a NNE–SSW-oriented Φ which is the same result as at the nearby station UDD in Eken et al. (2010). They interpret their splitting measurements in this region (58–62° N) as scattered compared to more uniform splitting results in the north and south of Sweden.

6.2 Central Scandes and Sweden

Across the SCANLIPS1 line at about 63° N, three domains with similar splitting parameters can be distinguished (Fig. 9a, b). To the west of 10° E longitude, close to the Atlantic coast, Φ is N–S-oriented for SKS and SKKS splits from South America. At about 10° to 11.5° E longitude, Φ varies for different backazimuths as well as SKS and SKKS. There seems to be a complex anisotropic structure at depth which may be related to the northern end of the MTFC as observed further south. The rock fabric of this fault zone is characterised by laterally varying directions in foliation and lineations (Gilotti and Hull 1993). In the Caledonian nappes, mainly NW–SE-oriented foliation is found, indicating top-to-the-NW transport, whereas NE–SW structures including deformed rock matrixes dominate close to the shear faults (Gilotti and Hull 1993). This deformation causes locally significant and

variable seismic anisotropy. Below this crustal zone, in the mantle, there may be another direction of the fast polarisation axis; however, due to few observed backazimuths, this cannot be resolved with the presently available datasets.

Along the eastern part of SCANLIPS1 (east of 11.5° E) up to the coast of the Baltic Sea in Sweden, the apparent Φ is very stable in NW–SE direction for the analysed South American event. SKS and SKKS measurements of this event are nearly the same, with the exception of one SW–NE Φ result at station N0025. Measurements at SKS phases from Pacific events (red arrows in Fig. 9) have a more E–W trend, indicating that a single layer model of anisotropy cannot explain the measurements. Splitting results at the permanent Swedish stations (NOR, SOL and HEM) also retrieved a WNW–ESE apparent Φ which is attributed to the anisotropic signature of the Central Svecofennian province (Eken et al. 2010). Structural maps indicate a NW–SE orientation of ductile shear zone and tectonic foliation at about 15° to 16° E. Further east, this structural trend changes to NE–SW foliation (Högdahl et al. 2009), coinciding with Φ at station N0025. This coincidence is a strong hint that upper crustal rock fabrics can influence the fast polarisation direction of SKS waves. At present, we cannot determine the fast polarisation direction at deeper levels in the crust or mantle. The δt values in this region are high with an average δt of 1.7 s and many single measurements exceeding 2 s (Table 3), which may be due to either a long travel path in the mantle with uniform moderate anisotropy (<3 %) or a short travel path with a highly anisotropic rock fabric.

6.3 Northern Scandes and Sweden

The Φ results across SCANLIPS2 line (66° to 68°N) clearly differ from SCANLIPS1 (~63.5° N). In the Northern Scandes and across its eastern foreland in Northern Sweden, the prevailing direction of Φ is NE–SW. Only at station N7005 inside the Northern Scandes is Φ oriented N–S (SKS) to NW–SE (SKKS). In Sweden, at the permanent network, apparent Φ is also NE–SW (Eken et al. 2010), which is interpreted to be due to a North Scevofennian–Karelian domain. Our Φ results show a clear difference in anisotropic signature between the Central and Northern Scecofennian domains as was also found by Eken et al. (2010). The δt values in the northern part are similar on average (~1.7 s) to the average δt in the central part of the Scandes and Sweden.

7 Summary

We present SKS/SKKS splitting results from broadband stations in Norway and Sweden which cover the Scandinavian Mountains as well as their eastern foreland. The splitting parameters Φ and δt can be determined well and are stable at most stations and lead to reproducible results for similar epicentre regions. The splitting results are not consistent with a simple one-layer model of anisotropy. We find frequency-dependent splitting parameters at KONO, which indicate depth-dependent anisotropy. At many stations, the splitting parameters depend on the backazimuth of the events which also indicates multiple-layer anisotropy or inclined fast axis alignment. The split times are mostly well above 1 s, indicating a mantle contribution. However, a correlation of the apparent fast direction Φ with mantle fabric in the lower lithosphere or flow in the asthenosphere cannot be deduced yet. Therefore, the non-uniqueness of the splitting results from different backazimuths still impedes an interpretation towards the origin of the uplift processes of the Scandinavian Mountains.

The recovered complicated pattern of Φ and δt cannot be explained with a unique model of the anisotropic structure at depth because splitting is observed from few backazimuths. However, characteristic signatures of Φ can be observed for different tectonic domains and partly correlate with preferred mineral orientations of rocks at the surface. In the south of Norway, Φ is mainly NW–SE-oriented in the Sveconorwegian Proterozoic domain. On the Caledonian nappes, Φ is NE–SW-oriented and Φ appears to vary significantly around the Oslo Graben and Møre–Trøndelag fault complex. In these two regions, the lithosphere is deeply faulted and deformed, which may lead to a small-scale pattern of rock fabrics and hence varying splitting results. In Central Sweden, on the Central Svecofennian lithosphere, Φ is preferably in NW–SE direction, clearly different from the NE–SW orientation of Φ in the Northern Svecofennian province further north.

The determination of the splitting results was partly improved by stacking the SKS/SKKS waveforms using array techniques. This approach should be pursued in the future to achieve more splitting parameters, especially from backazimuth regions without present

observations. Such improved backazimuth coverage is necessary to better interpret the observed splitting parameters in terms of anisotropic structure at depth.

Acknowledgments We thank especially Dr. Richard England, Dr. Johannes Schweitzer, and Dipl.-Geophys. Britta Wawerzinek for providing SCANLIPS, NORSAR, and MAGNUS waveforms, respectively. Two anonymous reviewers provided helpful and critical comments. MAGNUS waveforms were recorded with the mobile KArlsruhe BroadBand Array of the Universität Karlsruhe (TH) (now KIT), Germany, as well as with permanent stations of the NORSAR array and the Norwegian National Seismological Network. Financial support for the MAGNUS experiment was provided by the University of Aarhus, University of Copenhagen, University of Karlsruhe, and University of Oslo as well as NORSAR. SCANLIPS was conducted using seismic stations from the NERC Geophysical Equipment Facility (SEIS-UK) and NERC funding. This work has been done in association with the partners of the ESF EUROCORES TOPO-EUROPE Programme 07-TOPO-EUROPE-FP-014 "The Scandinavian mountain chain: deep processes (TopoScandiaDeep)". It was supported by the Deutsche Forschungsgemeinschaft through grant RI1133/8-1. SeismicHandler (Stammler 1993) was used for seismic waveform processing and GMT software (Wessel and Smith 1998) was used for plotting maps.

References

Andersen TB (1998) Extensional tectonics in the Caledonides of southern Norway, an overview. Tectonophysics 285:333–351

Anell I, Thybo H, Artemieva IM (2009) Cenozoic uplift and subsidence in the North Atlantic region: geological evidence revisited. Tectonophysics 474:78–105

Babuska V, Cara M (1991) Seismic anisotropy in the Earth. Kluwer Academic, Dordrecht

Barruol G, Mainprice D (1993) A quantitative evaluation of the contribution of crustal rocks to the shear-wave splitting of teleseismic SKS waves. Phys Earth Planet Inter 78:281–300

Bastow ID, Owens TJ, Helffrich G, Knapp JH (2007) Spatial and temporal constraints on sources of seismic anisotropy: evidence from the Scottish highlands. Geophys Res Lett 34:L05305

Bijwaard H, Spakman W (2000) Nonlinear global P-wave tomography by iterated linearized inversion. Geophys J Int 141:71–82

Bungum H, Husebye ES, Ringdal F (1971) The NORSAR array and preliminary results of data analysis. Geophys J R astr Soc 25:115–126

Eken T, Plomerová J, Roberts R, Vecsey L, Babuška V, Shomali H, Bodvarsson R (2010) Seismic anisotropy of the mantle lithosphere beneath the Swedish National Seismological Network (SNSN). Tectonophysics 480:241–258

England RW, Ebbing J (2008) SCANLIPS—a seismological study of epeirogenic uplift of Scandinavia. Geophys Res Abstr 10:EGU2008-A-02842

England RW, Ebbing J (2012) Crustal structure of central Norway and Sweden from integrated modelling of teleseismic receiver functions and the gravity anomaly. Geophys J Int, in press

Evans MS, Kendall J, Willemann RJ (2003) Development of automated SKS splitting measurement—an additional parameter to be provided by the ISC. Eos Trans Am Geophys Union 84(46) Fall Meet Suppl Abstract S32C-03

Gabrielsen RH, Faleide JI, Pascal C, Braathen A, Nystuen JP, Etzelmuller B, O'Donnell S (2010) Reply to discussion of Gabrielsen et al. (2010) by Nielsen et al. (this volume): latest Caledonian to present tectonomorphological development of southern Norway. Marine Petroleum Geol 27:1290–1295

Gilotti JA, Hull JM (1993) Kinematic stratification in the hinterland of the central Scandinavian Caledonides. J Struc Geol 15:629–646

Gledhill K, Gubbins D (1996) SKS splitting and the seismic anisotropy of the mantle beneath the Hikurangi subduction zone, New Zealand. Phys Earth Planet Inter 95:227–236

Grechka VY, McMechan GA (1995) Anisotropy and non-linear polarization of body waves in exponentially heterogeneous media. Geophys J Int 123:959–965

Högdahl K, Sjöström H, Bergman S (2009) Ductile shear zones related to crustal shortening and domain boundary evolution in the central Fennoscandian Shield. Tectonics 28:TC1003

Japsen R, Chalmers JA (2000) Neogene uplift and tectonics around the North Atlantic: overview. Global Planet Chang 24:165–173

Karato S, Jung H, Katayama I, Skemer P (2008) Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies. Annu Rev Earth Planet Sci 36:59–95

Kaviani A, Rümpker G, Weber M, Asch G (2011) Short-scale variations of shear-wave splitting across the Dead Sea basin: evidence for the effects of sedimentary fill. Geophys Res Lett 38:L04308

Kennett BLN, Engdahl ER (1991) Traveltimes for global earthquake location and phase identification. Geophys J Int 105:429–465

Köhler A, Weidle C, Maupin V (2011) Directionality analysis and Rayleigh wave tomography of ambient seismic noise in southern Norway. Geophys J Int 184:287–300

Lidmar-Bergström K, Bonow JM (2009) Hypotheses and observations on the origin of the landscape of southern Norway—a comment regarding the isostasy–climate–erosion hypothesis by Nielsen et al. 2008. J Geodynamics 48:95–100

Long ML (2010) Frequency-dependent shear wave splitting and heterogeneous anisotropic structure beneath the Gulf of California region. Phys Earth Planet Inter 182:59–72

Long ML, van der Hilst RD (2005) Estimating shear-wave splitting parameters from broadband recordings in Japan: a comparison of three methods. Bull Seism Soc Am 95:1346–1358

Marson-Pidgeon K, Savage MK (1997) Frequency-dependent anisotropy in Wellington, New Zealand. Geophys Res Lett 24:3297–3300

Maupin V (2011) Upper-mantle structure in southern Norway from beamforming of Rayleigh wave data presenting multipathing. Geophys J Int 185:985–1002

Maupin V, Park J (2007) Theory and observations: wave propagation in anisotropic media. In: Romanowicz B, Dziewonski A (eds) Treatise on geophysics, vol 1, Seismology and structure of the earth. Elsevier, Amsterdam, pp 289–321

Medhus AB, Balling N, Jacobsen BH, Weidle C, Voss P, England RW, Kind R, Thybo H (2012) Upper mantle structure beneath the Southern Scandes Mountains and the Northern

Tornquist Zone revealed by P-wave travel time tomography. Geophys J Int 189:1315–1334
Neumann ER, Olsen KH, Baldridge WS (1995) The Oslo rift. In: Olsen KH (ed) Continental rifts: evolution, structure, tectonics. Elsevier, Amsterdam, pp 345–373
Nicolas A (1993) Why fast polarization directions of SKS seismic waves are parallel to mountain belts. Phys Earth Planet Interiors 78:337–342
Nielsen SB et al (2009) The evolution of western Scandinavian topography: a review of Neogene uplift versus the ICE (isostacy–climate–erosion) hypothesis. J Geodynamics 47:72–95
Niu F, Perez AM (2004) Seismic anisotropy in the lower mantle: a comparison of waveform splitting of SKS and SKKS. Geophys Res Lett 31:L24612. doi:10.1029/2004GL021196
Olsen E, Gabrielsen RH, Braathen A, Redfield TF (2007) Fault systems marginal to the Møre–Trøndelag fault complex, Osen–Vikna area, Central Norway. Norwegian J Geol 87:59–73
Pascal C, Olesen O (2009) Are the Norwegian mountains compensated by a mantle thermal anomaly at depth? Tectonophysics 475:160–168
Plomerová J, Frederiksen AW, Park J (2008) Preface: seismic anisotropy and geodynamics of the lithosphere–asthenosphere system. Tectonophysics 462:1–6
Ramberg IB, Bryhni I, Nøttvedt A, Rangnes K (2008) The making of a land. Geology of Norway. The Norwegian Geological Association, Oslo
Redfield TF, Braathen A, Gabrielsen RH, Osmundsen PT, Torsvik TH, Andriessen PAM (2005) Late Mesozoic to Early Cenozoic components of vertical separation across the Møre–Trøndelag fault complex, Norway. Tectonophysics 395:233–249
Rohrman M, van der Beek P, Andriessen P, Cloetingh S (1995) Meso-Cenozoic morphotectonic evolution of southern Norway: Neogene domal uplift inferred from apatite fission track thermochronology. Tectonics 14:704–718
Rost S, Thorne MS, Garnero EJ (2006) Imaging global seismic phase arrivals by stacking array processed short-period data. Seis Res Lett 77:697–707. doi:10.1785/gssrl.77.6.697
Roy C (2010) SKS-Doppelbrechung und Anisotropie unter dem Skandinavischen Gebirge. Diploma thesis, KIT, Geophysical Institute (in German)
Rümpker G, Silver PG (1998) Apparent shear-wave splitting parameters in the presence of vertically varying anisotropy. Geophys J Int 135:790–800
Rümpker G, Tommasi A, Kendall J-M (1999) Numerical simulations of depth-dependent anisotropy and frequency-dependent wave propagation effects. J Geophys Res 104:23141–23153
Savage MK (1999) Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting? Rev Geophys 37:65–106
Sieminski A, Paulssen H, Trampert J, Tromp J (2008) Finite-frequency SKS splitting: measurement and sensitivity kernels. Bull Seism Soc Am 98:1797–1810
Silver PG (1996) Seismic anisotropy beneath the continents: probing the depth of geology. Annu Rev Earth Planet Sci 24:385–432
Silver PG, Chan WW (1991) Shear wave splitting and subcontinental mantle deformation. J Geophys Res 96:16429–16454
Silver PG, Savage MK (1994) The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers. Geophys J Int 119:949–963
Smelror M, Dehls J, Ebbing J, Larsen E, Lundin ER, Nordgulen O, Osmundsen PT, Olesen O, Ottesen D, Pascal C, Redfield TF, Rise L (2007) Towards a 4D topographic view of the Norwegian sea margin. Global Planet Chang 58:382–410
Stammler K (1993) SeismicHandler—programmable multichannel data handler for interactive and automatic processing of seismological analyses. Comp Geosci 19:135–140
Stratford W, Thybo H (2011) Seismic structure and composition of the crust beneath the southern Scandes, Norway. Tectonophysics 502:364–382
Stratford W, Thybo H, Faleide JI, Olesen O, Tryggvason A (2009) New Moho map for onshore southern Norway. Geophys J Int 178:1755–1765
Svenningsen L, Balling N, Jacobsen BH, Kind R, Wylegalla K, Schweitzer J (2007) Crustal root beneath the highlands of southern Norway resolved by teleseismic receiver functions. Geophys J Int 170:1129–1138
Tommasi A (1998) Forward modeling of the development of seismic anisotropy in the upper mantle. Earth Planet Sci Lett 160:1–13
Torsvik TH, Cocks LRM (2005) Norway in space and time: a centennial cavalcade. Norwegian J Geol 85:73–86
Vecsey L, Plomerová J, Babuŝka V (2008) Shear-wave splitting measurements—problems and solutions. Tectonophysics 462:178–196
Weidle C, Maupin V (2008) An upper-mantle S-wave velocity model for Northern Europe from Love and Rayleigh group velocities. Geophys J Int 175:1154–1168
Weidle C, Maupin V, Ritter J, Kværna T, Schweitzer J, Balling N, Thybo H, Faleide JI, Wenzel F (2010) MAGNUS—a seismological broadband experiment to resolve crustal and upper mantle structure beneath the southern Scandes mountains in Norway. Seism Res Lett 81:76–84. doi:10.1785/gssrl.811.76
Wessel P, Smith WHF (1998) New, improved version of generic mapping tools released. Eos Trans Am Geophys Union 79:579
Wirth E, Long MD (2010) Frequency-dependent shear wave splitting beneath the Japan and Izu-Bonin subduction zones. Phys Earth Planet Inter 181:141–154

Appendix C

Article: Effect of ray and speed perturbations on Ionospheric Tomography by Over-the-horizon radar: A new method

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE

10.1002/2014JA020137

Key Points:
- A new 3-D ionospheric tomography method for monostatic OTH radar
- The raypath deflection effect is not negligible
- Three-dimensional reconstruction of the electron density over Europe is possible

Correspondence to:
C. Roy,
roy@ipgp.fr

Citation:
Roy, C., G. Occhipinti, L. Boschi, J.-P. Moliné, and M. Wieczorek (2014), Effect of ray and speed perturbations on ionospheric tomography by over-the-horizon radar: A new method, *J. Geophys. Res. Space Physics*, *119*, doi:10.1002/2014JA020137.

Received 28 APR 2014
Accepted 18 AUG 2014
Accepted article online 22 AUG 2014

Effect of ray and speed perturbations on ionospheric tomography by over-the-horizon radar: A new method

Corinna Roy[1,2], Giovanni Occhipinti[1], Lapo Boschi[3], Jean-Philippe Moliné[2], and Mark Wieczorek[1]

[1]Institut de Physique du Globe de Paris, Université Paris Diderot, Paris, France, [2]Office National d'études et de recherches aérospatiales, Palaiseau, France, [3]Laboratoire iSTeP, Université Pierre et Marie Curie, Paris, France

Abstract Most recent methods in ionospheric tomography are based on the inversion of the total electron content measured by ground-based GPS receivers. As a consequence of the high frequency of the GPS signal and the absence of horizontal raypaths, the electron density structure is mainly reconstructed in the *F*2 region (300 km), where the ionosphere reaches the maximum of ionization, and is not sensitive to the lower ionospheric structure. We propose here a new tomographic method of the lower ionosphere, based on the full inversion of over-the-horizon (OTH) radar data. Previous studies using OTH radar for ionospheric tomography inverted only the leading edge echo curve of backscatter ionograms. The major advantage of our methodology is taking into account, numerically and jointly, the effect that the electron density perturbations induce not only in the speed of electromagnetic waves but also on the raypath geometry. This last point is extremely critical for OTH radar inversions as the emitted signal propagates through the ionosphere between a fixed starting point (the radar) and an unknown end point on the Earth surface where the signal is backscattered. We detail our ionospheric tomography method with the aid of benchmark tests. Having proved the necessity to take into account both effects simultaneously, we apply our method to real data. This is the first time that the effect of the raypath deflection has been quantified and that the ionospheric plasma density has been estimated over the entirety of Europe with an OTH radar.

1. Introduction

Most recent methods in ionospheric tomography are based on the inversion of the total electron content (TEC) derived from dual-frequency receivers of the Global Navigation Satellites System (GNSS), mainly the American Global Positioning System (GPS). The TEC is the integral of electron density along the raypath between the satellite and the receiver, computed from the phase difference between two signals emitted by the GPS satellite and measured at the receiver [*Mannucci et al.*, 1993]. TEC is usually visualized as global or regional two-dimensional maps to show the state of the ionosphere [*Mannucci et al.*, 1998]. The measurement of TEC on the raypath between satellites and receivers, jointly with some a priori information about the background ionospheric model, allows additionally to estimate the local electron density of the ionosphere using the inverse problem theory, [e.g., *Tarantola*, 2005].

GNSS ionospheric tomography was first tested in simulations by *Austen et al.* [1988], showing that numerical tomography techniques can be used to produce two-dimensional vertical cross sections of the electron density in the ionosphere. *Andreeva et al.* [1990] published the first experimental result using TEC data collected at three receivers in Russia and reconstructed a two-dimensional vertical cross section of the ionosphere by satellite radio tomography for the first time.

The installation of global (e.g., the International GNSS Service) and regional (e.g., GNSS Earth Observation Network (GEONET) in Japan) ground-based GPS networks greatly increased the amount of available data and raypath coverage, making three-dimensional reconstruction of the ionosphere with time evolution (3-D + 1) possible. Since then, GPS-based computerized ionospheric tomography has been widely used to investigate the temporal and spatial variations of ionospheric structures [*Hansen et al.*, 1997; *Hajj et al.*, 1994; *Hernández-Pajares et al.*, 1998; *Bust et al.*, 2004; *Garcia et al.*, 2005; *Ma et al.*, 2005; *Yizengaw et al.*, 2005; *Wen et al.*, 2007].

Although GPS is a powerful tool for studying the ionosphere, theoretical limitations of ionospheric tomography using GNSS satellite-to-Earth configuration have been discussed in detail in the literature [*Yeh and*

Raymund, 1991; *Na and Sutton*, 1994]. Apart from the spatial and temporal limitations associated with the experimental arrangements, the data sets are often incomplete because of the finite sampling interval and the limited angle of view of each receiver. Additionally, most of the tomographic methods make the assumption of an invariant ionosphere during the time of measurement.

Notwithstanding those previous limitations, the major problems of ionospheric tomography using GPS are the following. First, the absence of horizontal raypaths, resulting in a low vertical resolution. Second, the high frequency of the GPS signal limits the sensitivity of the electromagnetic waves to the maximum of electron density in the ionosphere, nominally the *F* region at around 300 km altitude.

To overcome the major limitations of low vertical resolution, three approaches are possible. The first approach is to incorporate, in the reconstruction, extra information from other instruments, for instance ionosonde data, which allow for the determination of electron density in the lower ionospheric layers, such as the E (∼100 km) and *F*1 region (∼250 km). *Heaton et al.* [1995] tested the incorporation of scaled ionograms into the imaging and found improvement in the vertical profiles. *Kersley et al.* [1993] showed promising results with the incorporation of ionosonde data into the reconstruction algorithm and better agreement with the EISCAT (European Incoherent Scatter Scientific Association) measurements in the vicinity of the layer maximum. *Markkanen et al.* [1995] applied a Bayesian approach to simulated results and incorporated peak heights as a priori information in the reconstruction algorithm for calculating the *F* region electron density, which are compared with the EISCAT radar observations.

The second approach is to use general knowledge of the shape of ionospheric profiles to constrain the reconstruction results and to fill the information gap: The most widely used solution is the one proposed by *Fremouw et al.* [1992], who used a set of vertical orthonormal functions, created from ionospheric models, to model the vertical profile. Unfortunately, existing ionospheric models are partially incorrect, because they fail to predict the strong day-to-day variations, consequently the reconstructions constrained with vertical profiles from ionospheric models are often inaccurate. Consequently, *Fehmers et al.* [1998] proposed a model-independent algorithm that compensates for the lack of horizontal raypaths with information that is not related to some specific model: basically, they impose that the electron density cannot be negative at high and low altitudes and is additionally smooth and vertically stratified. Their tests show moderate success with an error in the layer-height estimation on the order of 90 km.

Another recently developed method uses occultation data. The technique is based on using the radio signals continuously broadcast by the GNSS satellites (GPS/Global Navigation Satellite System/Galileo) that are measured by a receiver on a low Earth orbit satellite. Along its way through the ionosphere the signal has been refracted due to free electrons that modify the refraction index. Since both satellites move over time, this technique allows a vertical scanning of successive layers of the atmosphere. *Rius et al.* [1997] showed that the combined measurement of GPS TEC by ground stations and occultations improved the vertical resolution. Additionally, the authors proved that the use of ground stations alone is insufficient for vertical reconstruction of the electron density.

To deal with the instabilities of the solution, introduced by noise and the ill conditioning of the problem, different reconstruction algorithm have been tested for ionospheric tomography. Several works [*Kersley et al.*, 1993; *Heaton et al.*, 1995; *Mitchell et al.*, 1995; *Pryse et al.*, 1995; *Vasicek and Kronschnabl*, 1995] applied the iterative multiplicative algebraic reconstruction technique [*Gordon et al.*, 1970] that attempts to minimize the difference between measured and calculated TEC values by modifying the background ionosphere until the differences are acceptably small. *Kunitsyn et al.* [1994a, 1994b] tried an iterative algebraic reconstruction technique algorithm that incorporates some prior information into each pixel in the ionospheric grid that they wish to solve. The solution obtained with this algorithm, however, is severely limited because it is very sensitive to the initial ionospheric model. To cope with this problem, *Wen et al.* [2010] recently proposed the constrained algebraic reconstruction technique, where cells not hit by any ray extract information from neighbor cells.

The purpose of our work is to present a new tomographic method of the lower ionosphere (≤300 km) based on the full inversion of over-the-horizon (OTH) radar data. Our method can integrate easily GPS TEC observations of ground-based or onboard receivers (occultation), as well as ionosonde measurements when such data are available.

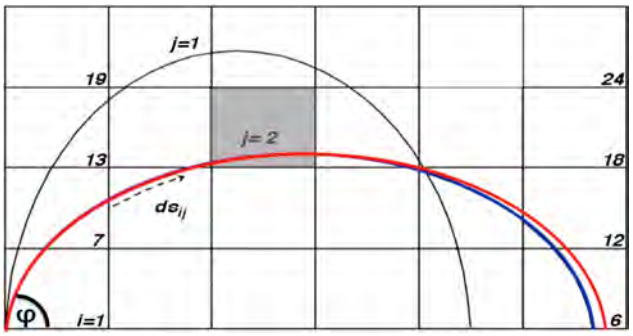


Figure 1. Schematic representation of our parametrization of *j* rays with elevation angle φ and *i* cells with the raypath deflection induced by a localized perturbation (e.g., gray cell) producing a perturbed ray (red) compared to the unperturbed ray (black).

OTH radar takes advantage of the refraction properties of the ionosphere, where the presence of free electrons causes the electromagnetic (EM) wave deflection in the ionosphere. Consequently, the emitted signal, after deflection in the ionosphere, can reach the ground beyond the horizon, typically up to thousands of kilometers away from the transmitter. There, the signal is backscattered and goes back to the receiver following the same raypath (Figure 1). The received signal contains all information about the propagating medium. We emphasize that the point where the signal is backscattered is not fixed and changes with the ionization of the propagating medium.

Previous studies in ionospheric tomography by OTH radar are all based on the inversion of the leading edge echo curve, which contains, for each frequency, only the information of the ray with minimal group delay (measured from the emitter to a point of first contact with the Earth). To estimate the three major ionospheric parameters (the critical frequency f_c, the peak height, and the semithickness for each layer), two approaches to solve the inverse problem are possible. Either fitting the observed leading edge with a quasi-parabolic ionospheric layer [*Rao*, 1974; *Bertel and Cole*, 1988; *Ruelle and Landeau*, 1994; *Landeau et al.*, 1997]; or using ray tracing to numerically simulate the leading edge [*Coleman*, 1998; *Fridman and Fridman*, 1994; *Fridman*, 1998].

A major drawback of using only the leading edge is that valuable information present in the data is neglected. To overcome this limitation, we set up an inverse problem taking into account the complete radar data set. We use the ray-tracing tool TDR (Tracé de Rayon) [*Occhipinti*, 2006] to calculate the synthetic propagation time. This code traces rays in an a priori heterogenic 3-D ionospheric model NeQuick [*Radicella and Leitinger*, 2001] and in the elliptical WGS84 coordinate system [*National Imagery and Mapping Agency*, 2000], solving numerically the Eikonal equation, describing the propagation of rays in a medium, by a fourth-order Runge-Kutta method. It neglects the Earth's magnetic field, in order to apply the method to all OTH radars that are not able to determine the wave polarization and consequently cannot discriminate between the ordinary and extraordinary mode, induced by the geomagnetic field. In section 4, the calculated propagation time is compared with the propagation time measured by the OTH radar. In our inverse problem, the electron density in the ionosphere is directly estimated from the difference between the calculated and measured propagation time (section 4 and Figure 10).

As a consequence of the measurement characteristics and geometry of OTH radar, the major challenge of this methodology is imposed by the measurement geometry. EM waves emitted by OTH radars are reflected in the ionosphere then arrive at the ground where they are backscattered to the receiver following the same raypath. Though we treat only the monostatic case in this work, the method could be easily generalized to bistatic cases, where the receiver and emitter are not at the same location.

We present here a new ionospheric tomography method that takes into account, jointly, the velocity variation of electromagnetic waves induced by the electron density variation, as well as the induced perturbation in the raypath (section 2). The developed methodology will be tested using a set of synthetic benchmark tests (section 3) and will be applied to real data from the OTH radar Nostradamus (New Transhorizon Decametric System Applying Studio Methods) [*Bazin et al.*, 2006] in section 4.

2. Theory

2.1. The Forward Problem

The propagation of high-frequency (HF) radio waves in the ionosphere can be described using ray theory. Assuming an isotropic ionospheric medium, neglecting the Earth magnetic field

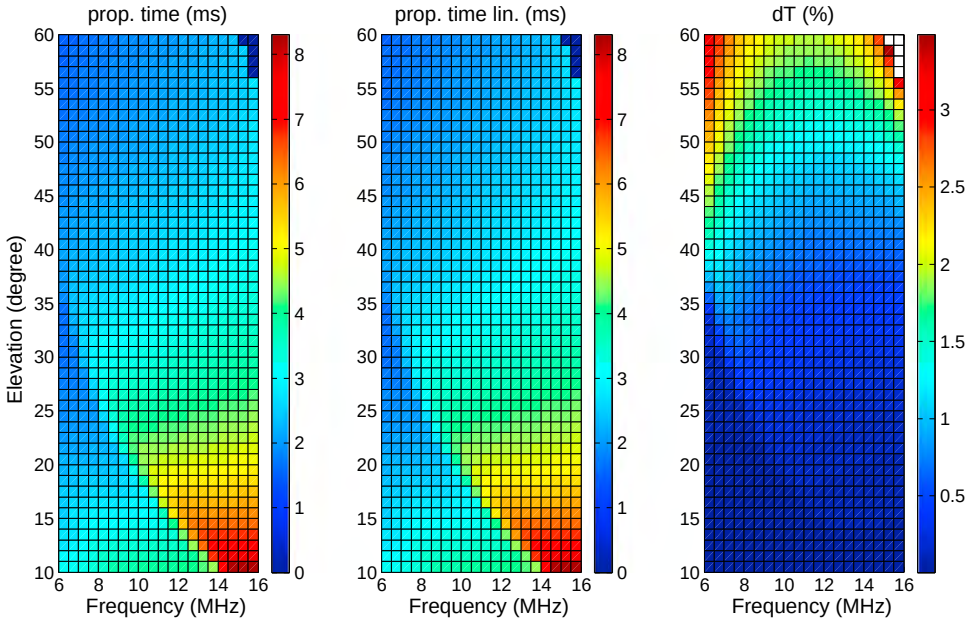


Figure 2. (left) Propagation time calculated without approximation (first line of equation (3)) versus (middle) linearized propagation time (calculated with the second line of equation (3)), and the (right) difference between them in percent. Computation for 1071 rays traced with elevation angle 10°–60° and frequency 6–16 MHz.

and losses in the ionosphere, the propagation time along the raypath $s\left(n\left(\vec{r}\right)\right)$ is given by Fermat's principle:

$$T_{\text{phase}} = \frac{1}{c}\int_{s(n)} n\left(\vec{r}\right) \mathrm{d}s, \tag{1}$$

where c is the speed of light, $n\left(\vec{r}\right)$ the refractive index of the medium, and $\vec{r}$ position.

Assuming the ionosphere is a stationary, isotropic, and horizontal stratified medium, the unmagnetized refractive index depends only on the frequency f_e of the emitted EM wave and the electron density $N_e(\vec{r})$,

$$n\left(\vec{r}\right) = \sqrt{1 - \frac{e^2 N_e\left(\vec{r}\right)}{4\pi^2 \epsilon_0 m_e f_e^2}} \approx \sqrt{1 - \frac{80.6 N_e\left(\vec{r}\right)}{f_e^2}}, \tag{2}$$

where ϵ_0 is the vacuum permittivity and e and m_e are the charge and the mass of the electron, respectively. The layered structure of the ionosphere produces a change of the refractive index as a function of altitude and bends EM waves emitted at high frequency (HF) toward the ground to locations beyond the horizon, typically up to thousands of kilometers away from the transmitter. By replacing the refractive index (2) in equation (1) and linearizing by a first-order Taylor series expansion, the propagation time of an EM wave can be separated into two parts:

$$\begin{aligned} T_{\text{phase}} &= \frac{1}{c}\int_{s(n)} \sqrt{1 - \frac{80.6 N_e\left(\vec{r}\right)}{f_e^2}}\mathrm{d}s \\ &\approx \frac{1}{c}\int_{s(n)} \mathrm{d}s - \frac{40.3}{c f_e^2}\int_{s(n)} N_e\left(\vec{r}\right)\, \mathrm{d}s, \end{aligned} \tag{3}$$

where the first integral of the second line describes the propagation in the vacuum and the second one the delay introduced by the ionosphere. Figure 2 shows the limit of the linearization comparing the first and second line of equation (3). The difference between the exact and linearized propagation time is always smaller than 3.5% when elevation angle is less than 60°.

Equation (3) allows to calculate, using the ray-tracing code TDR described by *Occhipinti* [2006], the propagation time $T_{\text{phase}}^{\text{synth}}$ of any EM wave with a given frequency f_e in any given ionospheric model with electron density $N_e^0(\vec{r})$.

$$T_{\text{phase}}^{\text{synth}} = \frac{1}{c}\int_{s(n)} \mathrm{d}s - \frac{40.3}{cf_e^2}\int_{s(n)} N_e^0\left(\vec{r}\right)\,\mathrm{d}s.$$

We introduce an electron density perturbation $\delta N_e\left(\vec{r}\right)$ of the a priori model $Ne^0(\vec{r})$ to describe the real ionosphere $Ne^0(\vec{r}) + \delta N_e(\vec{r})$. Then the propagation time in the real ionosphere is given by

$$\begin{aligned} T_{\text{phase}}^{\text{real}} &= \frac{1}{c}\int_{s(n)} \mathrm{d}s - \frac{40.3}{cf_e^2}\int_{s(n)} \left(N_e^0\left(\vec{r}\right) + \delta N_e\left(\vec{r}\right)\right)\,\mathrm{d}s \\ &= T_{\text{phase}}^{\text{synth}} - \frac{40.3}{cf_e^2}\int_{s(n)} \delta N_e\left(\vec{r}\right)\mathrm{d}s. \end{aligned}$$

We then obtain a travel time perturbation δT_{phase} that is the difference between propagation times in the real ionosphere and in the a priori model (e.g., Figure 10):

$$\delta T_{\text{phase}} = T_{\text{phase}}^{\text{real}} - T_{\text{phase}}^{\text{synth}} = -\frac{40.3}{cf_e^2}\int_{s(n)} \delta N_e\left(\vec{r}\right)\mathrm{d}s \tag{4}$$

This solution allows us to compute the travel time perturbation δT_{phase} as a function of an electron density perturbation δN_e of the a priori model Ne^0. The difference between the propagation time measured by the OTH radar and the one computed by TDR is directly linked to the electron density perturbation δN_e.

We emphasize that this approach is based on the hypothesis that the raypath $s(n)$ in the real ionosphere and in the a priori model is the same. That means that we are supposing that the electron density variation δN_e only introduces variations in the speed of the EM waves and not in the raypath. Consequently, we call this approach the *v* method.

2.2. The Inverse Problem

In order to solve numerically our problem, we choose a parametrization of N homogeneous, nonoverlapping blocks, indexed i, where the electron density perturbation $\delta N_e(\vec{r})$ can be expressed as

$$\delta N_e\left(\vec{r}\right) = \sum_{i=1}^{N} \delta m_i \cdot B_i\left(\vec{r}\right), \tag{5}$$

where the electron density perturbation in the block i is δm_i and N known basis functions $B_i(\vec{r})$ are

$$B_i\left(\vec{r}\right) = \begin{cases} 1 & \text{if } \vec{r} \text{ in block } i, \\ 0 & \text{otherwise} \end{cases}$$

Consequently, equation (4) for the jth measurement of travel time perturbation takes the following form:

$$\delta T_j = -\frac{40.3}{f_{e_j}^2 c}\sum_{i=1}^{N} \delta m_i\,\mathrm{d}s_{ij}. \tag{6}$$

where $\mathrm{d}s_{ij}$ is the length of raypath segments within block i of ray j.

Introducing the matrix $\mathbf{A}$ of size $M \times N$, where M is the number of travel time measurements and N is the number of basis functions (i.e., the number of blocks in the parametrization of our ionospheric model), equation (6) can be rewritten as

$$\delta T_j = \sum_{i=1}^{N} \delta m_i A_{ji}. \tag{7}$$

In tensor notation equation (7) takes the form

$$\delta \mathbf{T} = \mathbf{A} \cdot \delta \mathbf{m}, \tag{8}$$

where $\delta\mathbf{m}$ is the vector of model parameters with N unknown electron density perturbations δm_i, $\delta\mathbf{T}$ is the vector containing the M observed travel time perturbations δT_j, and $\mathbf{A}$ is the geometric

matrix containing the $M \times N$ raypath segments ds_{ij} of the jth ray in block ith, as well as the ray coefficient,

$$A_{ji} = -\frac{40.3}{f_{e_j}^2 c} ds_{ij}.$$

2.3. Solution of the Inverse Problem Regularization

Following *Menke* [1989] we solve equation (8) seeking the model that minimizes the L_2 norm for both data and model, namely, $\| \delta\mathbf{T} - \mathbf{A} \cdot \delta\mathbf{m} \|^2 = \min$ and $\| \delta\mathbf{m} \|^2 = \min$. The first condition imposes the best fit to the data, and the second condition minimizes the discrepancy from the a priori model. These two conditions are equivalent to solving the following two equations:

$$\delta\mathbf{m} = \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T \cdot \delta\mathbf{T} \Leftrightarrow \quad \| \delta\mathbf{T} - \mathbf{A} \cdot \delta\mathbf{m} \|^2 = \min$$
$$\delta\mathbf{m} = \mathbf{A}^T \cdot \left(\mathbf{A} \cdot \mathbf{A}^T\right)^{-1} \cdot \delta\mathbf{T} \Leftrightarrow \quad \| \delta\mathbf{m} \|^2 = \min$$

It is known that inverse problems in geophysics generally present a number N of parameters (here the vector $\delta\mathbf{m}$) larger than the number M of observations (here the vector $\delta\mathbf{T}$). Consequently, the problem is underdetermined, and the matrices $\mathbf{A}^T \cdot \mathbf{A}$ and $\mathbf{A} \cdot \mathbf{A}^T$ cannot be inverted. The matrices are close to singular, and even if the inverse matrices formally exist, they are often ill conditioned; that is, small changes in the data vector ($\delta\mathbf{T}$) lead to large changes in the model estimation ($\delta\mathbf{m}$).

To find a more stable solution ($\delta\mathbf{m}$) balancing the sensitivity to the data, as well as the coherence with the a priori model, *Menke* [1989] suggests the following damped least squares solution

$$\delta\mathbf{m} = \left(\mathbf{A}^T \cdot \mathbf{A} + \lambda \cdot \mathbf{I}\right)^{-1} \cdot \mathbf{A}^T \cdot \delta\mathbf{T}, \tag{9}$$

which minimizes the cost function $\| \delta\mathbf{T} - \mathbf{A} \cdot \delta\mathbf{m} \|^2 + \lambda \| \delta\mathbf{m} \|^2$, where λ is a regularization parameter and $\mathbf{I}$ the identity matrix.

It is not possible to minimize both terms simultaneously, but the parameter λ controls the emphasis that we put on the conflicting requirements. In the section 3.2, we highlight the way how to select the best value of λ satisfying our problem.

2.4. Taking the Raypath Deflection Into Account

In monostatic OTH radars, the end points of the rays (where the signal is backscattered by the ground) are not known. Consequently, the location of the scattering point can change for a constant elevation angle and depends on the electron density variation $\delta N_e(\vec{r})$ in the ionosphere (Figure 1). As a result, the raypath deviation, introduced by the variation of the scattering point, can introduce an additional shift $\delta T_{\text{phase}}^{\text{ray}}$ in the propagation time. We extend the theory described above, in order to take into account jointly not only the speed variation (v method) but also the raypath deflection induced by the variation of the scattering point, what we call the v and r method. *Snieder and Spencer* [1993] showed that both effects can be combined in the perturbation approach, and the two effects are simply additive to first order. Based on equation (31a) of *Snieder and Spencer* [1993] we can write the total observed time delay as

$$\delta T_{\text{phase}} = -\frac{40.3}{c f_e^2} \int_{s(n)} \delta N_e\left(\vec{r}\right) ds + \delta T_{\text{phase}}^{\text{ray}}. \tag{10}$$

If $s(n)$ is the raypath in the a priori model $N_e^0(\vec{r})$ and $s^*(n)$ is the raypath in the perturbed model $N_e^0(\vec{r}) + \delta N_e(\vec{r})$, $\delta T_{\text{phase}}^{\text{ray}}$ is described by

$$\begin{aligned} \delta T_{\text{phase}}^{\text{ray}} = {} & \frac{1}{c} \int_{s^*(n)} ds^* - \frac{1}{c} \int_{s(n)} ds \\ & + \frac{40.3}{c f_e^2} \left[\int_{s(n)} N_e^0(\vec{r}) ds - \int_{s^*(n)} N_e^0(\vec{r}) ds^* \right. \\ & \left. + \int_{s(n)} \delta N_e(\vec{r}) ds - \int_{s^*(n)} \delta N_e(\vec{r}) ds^* \right]. \end{aligned} \tag{11}$$

Figure 3 compares the values of $\delta T_{\text{phase}}^{\text{velocity}}$ and $\delta T_{\text{phase}}^{\text{ray}}$ for a selected set of rays and shows that the raypath deflection is not negligible. This is discussed in more detail in section 3.1.

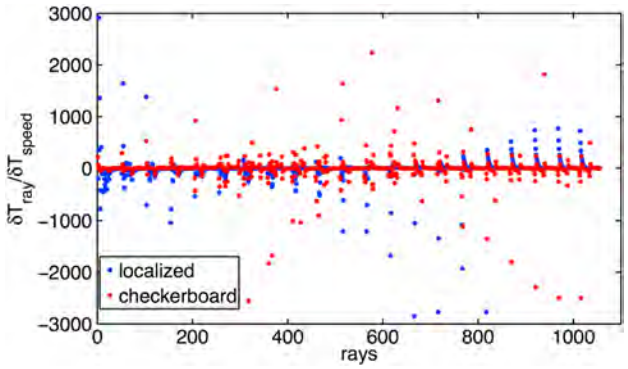


Figure 3. The ratio of δT_{ray} (equation (11)) to δT_{speed} (equation (4)) computed for the localized perturbation (Figure 4) and checkerboard (Figure 5) target models. The number of raypath are arranged in such a way that for one frequency (6–16 MHz), the elevation angle has been varied between 10° and 60°. The peaks correspond to low elevation angles between 10° and 30° for each frequency.

In this section, we extend the theory described above, in order to take into account jointly, not only the speed variation (*v* method) but also the raypath deflection induced by the variation of the scattering point (*v* and *r* method).

δT^{ray}_{phase} depends on δN_e, since raypath deflections are caused by electron density perturbations. To set up an inverse problem, allowing us to determine δN_e based on observations of δT_{phase}, we therefore write δT_{phase} as a linear function of δN_e. The study of *Snieder and Spencer* [1993] also indicates that sensitivity kernels $k(\vec{r})$ can be defined such that

$$\delta T^{ray}_{phase} = \int_{s(n)} k(\vec{r}) \cdot \delta Ne(\vec{r}) \tag{12}$$

where $k(\vec{r})$ is the data kernel [*Menke*, 1989], which in the *v* method described above is just a delta function along the unperturbed ray $s(n)$. Here the kernel contains the Fréchet derivatives $\partial T/\partial m$, where ∂T is a perturbation in the propagation time caused by a perturbation in the model m. If the relation between the model m and the propagation time T is linear, the sensitivity function $k(\vec{r})$ can be computed numerically. In practice, using the parametrization described above, we impose a localized electron density perturbation of arbitrary amplitude δNe_i^* at only the ith cell; consequently, we can compute by ray tracing the partial time perturbation δT_j^* along the jth ray induced by δNe_i^* in the ith cell:

$$\delta T_j^* = k_{ij} \cdot \delta Ne_i^*. \tag{13}$$

This allows us to create our base function

$$k_{ji} = \frac{\delta T_j^*}{\delta Ne_i^*},$$

that, following our linear hypothesis, is valid for a general case. We emphasize for additional clarity that the perturbed propagation time δT_j^* is computed following equation (11), where the perturbed raypath $s^*(n)$ is traced in the a priori model plus the perturbation δNe_i^* only in the ith cell. We can finally express equation (10) as

$$\delta T_j = -\frac{40.3}{f_{e_j}^2 c} \sum_{i=1}^{N} \delta m_i \, ds_{ij} + \sum_{i=1}^{N} k_{ji} \, \delta m_i, \tag{14}$$

or, in a tensor formalism as

$$\delta \mathbf{T} = (\mathbf{A} + \mathbf{A}') \cdot \delta \mathbf{m}, \tag{15}$$

where $A'_{ij} = k_{ij}$. The general inverse method solution described in the section 2.3 is applied to the matrix $\mathbf{A}$ in the case of *v* method, and to the matrix $\mathbf{A} + \mathbf{A}'$ in the case of *v* and *r* method.

3. Inversion Results for Synthetics

In order to validate and compare the two methods described in the previous sections, we generate synthetic data by the ray-tracing TDR in a known a priori ionospheric model, *NeQuick* [*Radicella and Leitinger*, 2001] that we call $Ne^{a\ priori}$, plus an additional perturbation that we call δNe^{target}. The $\delta \mathbf{T}$ is calculated as a difference of propagation time in $Ne^{a\ priori} + \delta Ne^{target}$ minus the propagation time in $Ne^{a\ priori}$. We emphasize that the ray geometry is different in the two models as the perturbation δNe^{target} introduces variation of the raypath.

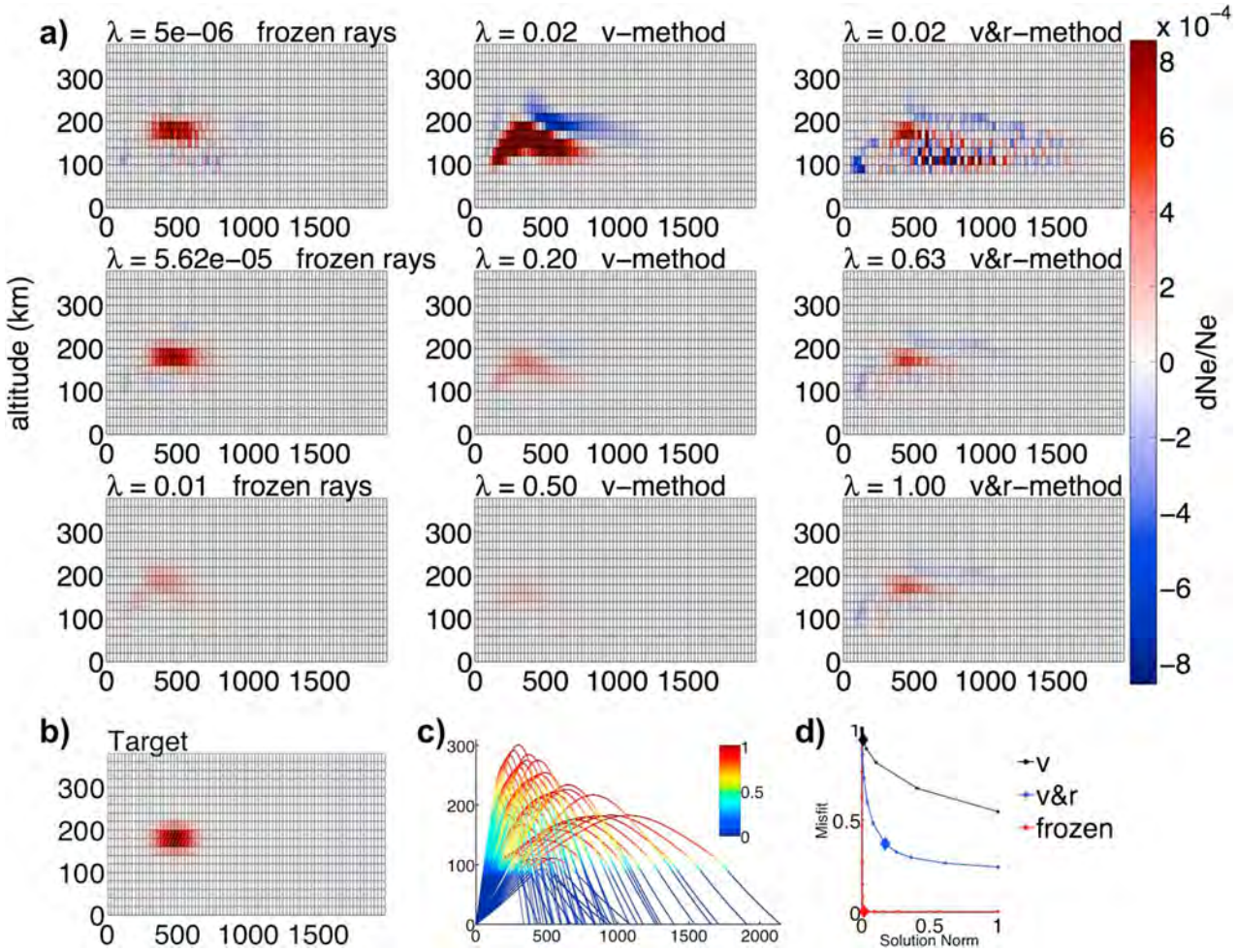


Figure 4. (a) Inversion results for the localized electron density perturbation benchmark with three different methods and for three different values of λ. Inversions for the best regularization parameter λ_{best} (middle row) for each method: frozen rays (left column), *v* method (middle column), and *v* and *r* method (right column). (b) Target model. (c) Ratio between plasma frequency and emission frequency along the raypath quantifying the sensitivity of each ray to the propagating medium. Rays are most sensitive to the medium where the ratio is approximately 1. (d) *L* curves for the frozen rays, *v* method, and *v* and *r* method. Diamonds correspond to the best values of regularization λ, i.e., $5.62 \cdot 10^{-5}$, 0.2, and 0.63 respectively.

Additionally, we compute the vector of travel time perturbations $\delta\mathbf{T}^{frozen}$ satisfying exactly the hypothesis that the electron density perturbation δNe^{target} modifies only the speed of EM waves (equation (4)); rays are *frozen* in the a priori model configuration. This data set represents the idealized case of no raypath perturbation, as if the ray end points were known; we shall invert it to separate the effects of poor data coverage and unknown raypath deflection or model resolution.

Independent of the method used (*v* or *v* and *r*), as well as for the synthetic data set ($\delta\mathbf{T}$ or $\delta\mathbf{T}^{frozen}$), the solution of our inverse problem $\delta\mathbf{m}$ has to correspond to δNe^{target}. We emphasize that all synthetic data are computed numerically by TDR in electron density continuous models ($N_e^{a\ priori}$ and δN_e^{target}) and not in discretized models following our parametrization. This is equivalent to introduce a noise in the synthetics in the order of 12%. The target model δN_e^{target} is parametrized using the above described parametrization for comparison with the solution $\delta\mathbf{m}$.

Quantitatively, we simulate rays with elevation angles between 10° and 60° and in the frequency range 6–16 MHz, as this is the operating capacity of the Nostradamus OTH radar [*Bazin et al.*, 2006]. The background ionosphere $Ne^{a\ priori}$ was generated for October at 12:00 UT with a solar flux of 198.1 solar flux units. The model is parametrized by a grid starting from the geographical coordinates of the Nostradamus radar

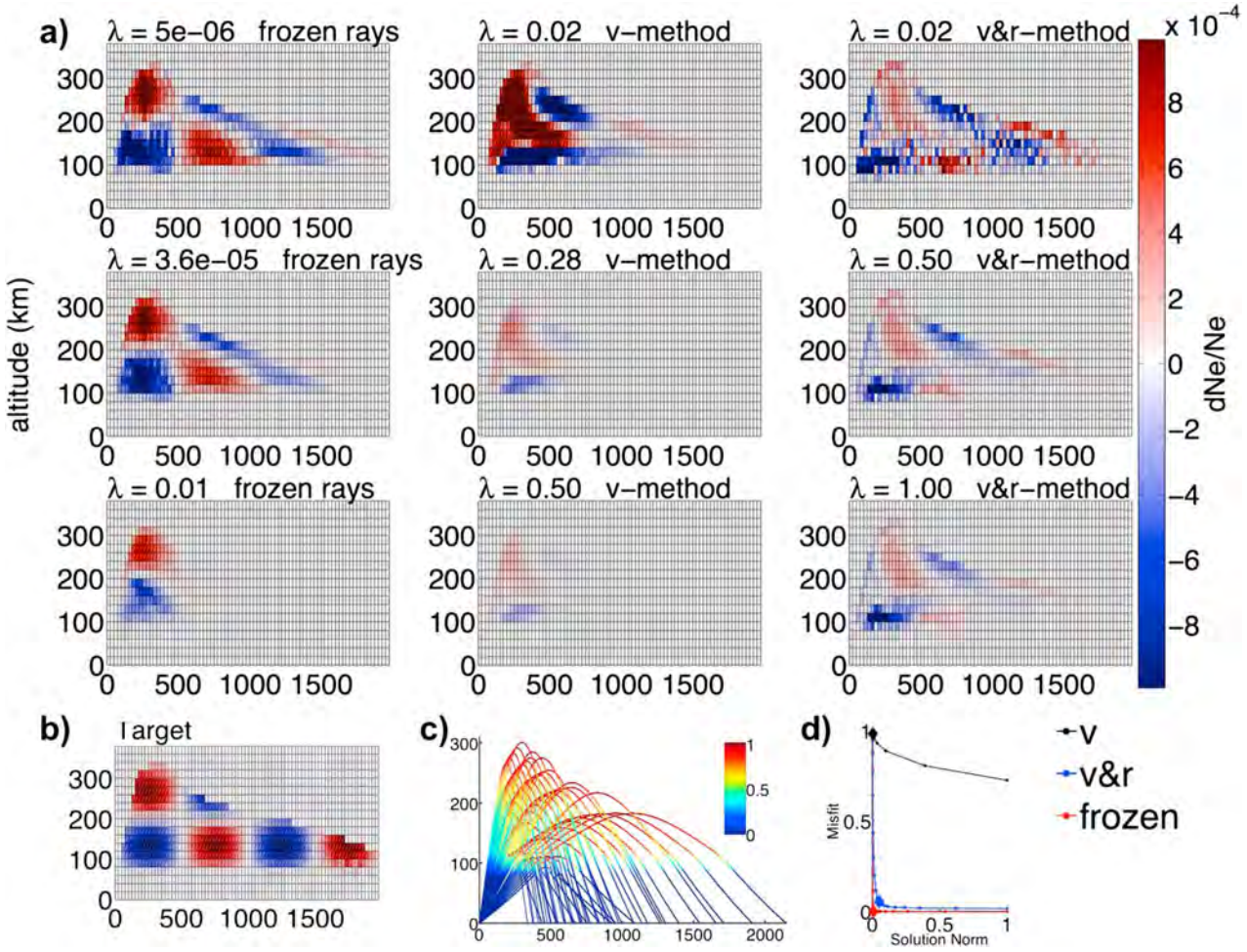


Figure 5. Same as Figure 4 but for a checkerboard perturbation. Diamonds of the *L* curves correspond to the best values of regularization λ, i.e., $3.6 \cdot 10^{-5}$, 0.28, and 0.5, respectively, for the frozen ray, *v* method, and *v* and *r* method.

covering an area of 2500 km in distance and reaching up to 400 km altitude. Each pixel of the grid has a dimension of 25 km in distance and 20 km in altitude. We traced here 1071 rays, using $\delta\phi = 1°$ in the elevation angle and $\delta f = 0.5$ MHz for the emission frequency, in accord with the capability of the radar Nostradamus.

We apply here the described methods to two different $\delta Ne^{\text{target}}$. The first is an ionospheric perturbation of a localized electron density perturbation of 0.1% of the background model $Ne^{\text{a priori}}$ (Figure 4), and the second is a checkerboard perturbation with the same order of amplitude (Figure 5). In the next section we comment in detail on the results of our synthetic tests.

The synthetic data set inverted in this section includes $\sim 10^3$ rays. Ray tracing for the entire synthetic data set in parallel on eight processors takes around 2 min on a grid with spatial resolution of 25 × 20 km and 2000 cells. The 2-D inversion result for the *v* method is obtained after 5 min for the first iteration. The performance of the *v* and *r* method is grid dependent, because of the number of cells to perturb. For a grid with spatial resolution of 25 × 20 km and $\sim 10^3$ rays, it takes around 40 min. For comparison with GPS ionospheric tomography we note that *Seemala et al.* [2014] can construct the 3-D electron density over Japan within 55 min on a grid with spatial resolution of 1° in latitude/longitude using data of 748 GEONET stations. Reducing the horizontal resolution of their grid to 2°, they obtain a 3-D image within 15 min.

3.1. The *v* Method Versus *v* & *r* Method

The inversion results for $\delta\mathbf{m}$ of our first test are summarized in Figure 4, where the solutions obtained with the *v* method and the *v* and *r* method are directly compared with the idealized frozen rays solution

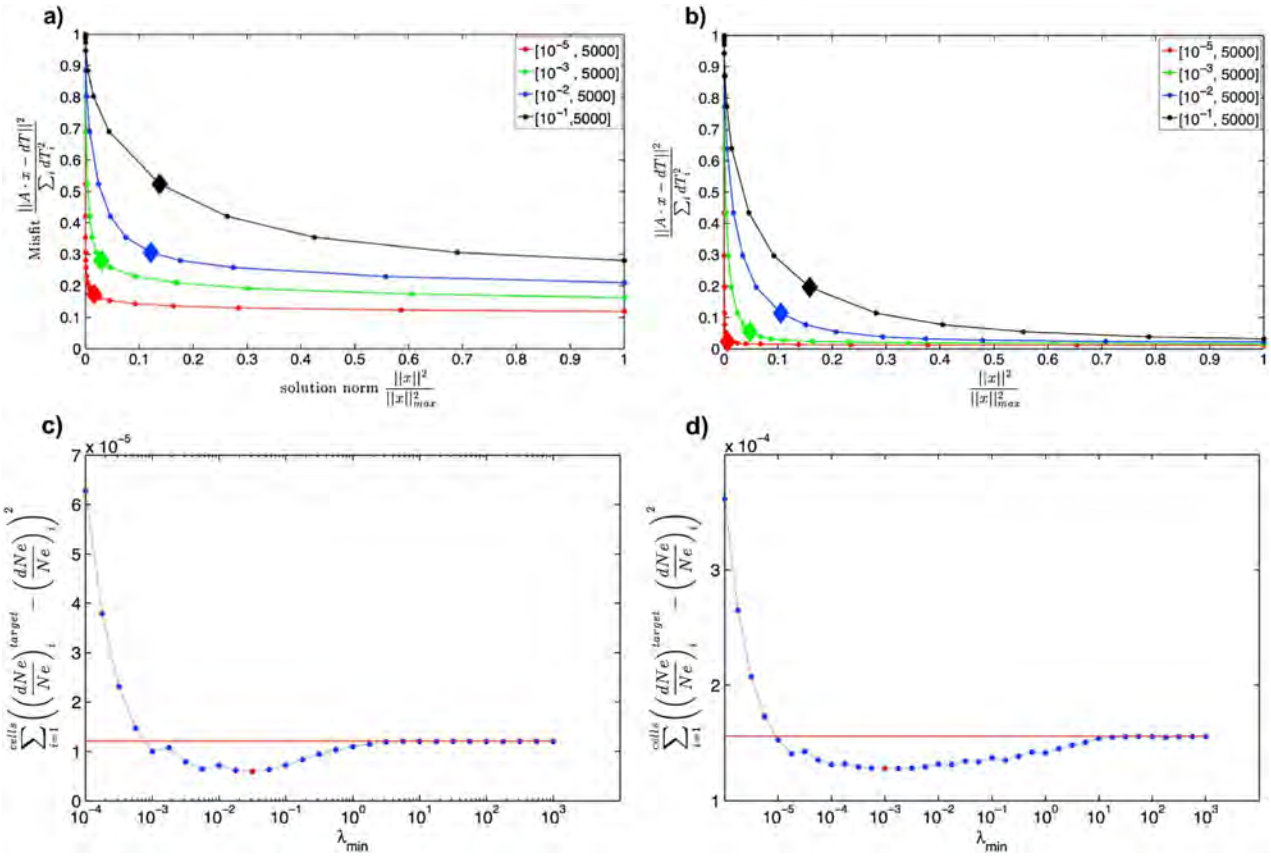


Figure 6. *L* curves for different λ ranges indicated in the top left for the (a) localized and the (b) checkerboard benchmark test using the *v* and *r* method for inversion. (c and d) Corresponding error curves for the explored range of regularization parameters. The red point represents the minimum error, and the red line is the sum over the target models, i.e., $\sum_{i=1}^{\text{cells}} \left(\frac{dNe}{Ne}\right)_i^2$. The errors converge to the red line when the solution is totally damped.

(subsection 3). Figure 4a shows the results for three different regularization parameters λ where the inversion results for the best regularization parameter (section 3.2) are shown in the middle of each column. The best λ is obtained by performing the inversion changing the regularization parameter and plotting the data misfit against the model misfit (i.e., *L* curve). The best regularization parameter lies at the maximum curvature of this curve, as this represents a compromise between small misfit and small solution norm (section 3.2).

As a consequence of the strong variation of the ionospheric background (electron density equal to zero at around 80 km, and of the order of $10^{11} - 10^{12} e/m^3$ at around 300 km) and the emission frequency dependence of the refraction index, EM waves emitted by the radar are particularly sensitive to the zone where the rays are reflected, where the plasma frequency f_p is approximately equal to the emission frequency f_e (Figure 5c). This defines the area of good coverage (independent of inversion method), as illustrated by Figure 5c with comparison with the frozen-ray inversion.

As is to be expected, solution models depend significantly on the regularization parameter. This effect is less severe in the idealized frozen-ray case that resembles GPS ionospheric tomography (both end points of the ray are known). As soon as raypath deflections (end point perturbations) are taken into account in our synthetic data, the resolution deteriorates and the choice of λ (equation (9)) affects more profoundly our results. The correct location (500 km distance, 200 km altitude) of the maximum anomaly in the target model is only reconstructed by the *v* and *r* method provided that an adequate value is assigned to λ. Furthermore, the *v* method identifies a high d*Ne*/*Ne* anomaly in the general area of 200–500 km in horizontal distance and 100–200 km in altitude but slightly mislocates it and does not reproduce its shape. The better performance of the *v* and *r* method compared to *v* method is confirmed by Figure 4d, where the *v* and *r L* curve [*Tikhonov*, 1963] has a more pronounced corner than the *L* curve obtained with the *v* method.

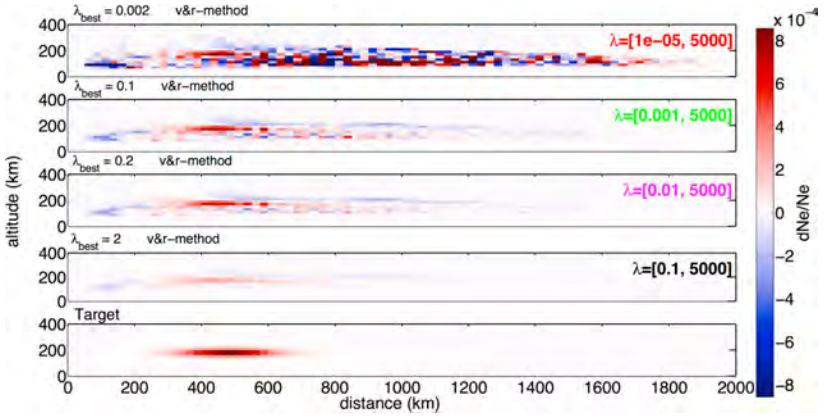


Figure 7. Inversion results using the *v* and *r* method with the best value of regularization chosen from the *L* curves in Figure 6a for the localized perturbation in order to explore the sensitivity to the λ range. The bottom row shows the target model.

Compared to the result obtained with the ideal frozen rays, both the *v* and *v* and *r* methods occasionally introduce large-scale low d*Ne*/*Ne* anomalies that do not correspond to any feature of the target model, where the sign of perturbation is always positive. We ascribe these artifacts to the nonlinear effects of ray-path deflections, which (end points not being fixed) can in principle result in a faster propagation time even if the velocity perturbation is negative.

The inferences made from Figure 4 are confirmed by results illustrated in Figure 5, where the same set of inversions is conducted after replacing the target model (and associated synthetic data) with a checkerboard. General comparison from Figures 4a and 5a, with Figures 4c and 5c, clearly show that the model is mainly reproduced in the zone of sensitivity where the emission frequency f_e is close to the plasma frequency f_p. This clearly emphasize the role of the coverage in our solution.

The selection of the value of λ in order to choose the best solution is detailed in the next section. The following discussions are only applied to the *v* and *r* method.

3.2. Solution Dependence on Regularization

To quantify the robustness of our solutions obtained with both the *v* and *v* and *r* methods, we explore their dependence on our choice of regularization parameter λ. This analysis is limited to the synthetic data discussed in section 3.1 so that obtained solution can be compared to a target model.

We first observe that while the *L* curve method (Figures 4d and 5d) is useful to monitor qualitatively the trade-off between data misfit and model quality, it is not guaranteed to provide the "best" solution, i.e., the one closest to the real world. This is confirmed by Figures 6a and 6b, where we show how the choice of model-norm normalization results in a different curvature of the *L*-curve and thus a different choice of preferred model. Each plot of Figures 7 and 8 illustrates the result for λ_{best} obtained from the corresponding *L* curve in Figures 6a and 6b. Figures 6c and 6d show the discrepancy between solution and target model for the same set of inversions and confirms that the model quality can be significantly affected by a inadequate choice of λ. Inspection of Figures 6c and 6d allows to identify the value of λ corresponding to the minimum error (discrepancy between solution and target model).

The corresponding λ_{best} are 0.63 and 0.5, respectively, and the inversion results corresponding to these regularization parameters are shown in Figures 4 and 5 in the middle of the right columns. The difference between target and model is less than 40% for the localized and less than 57% for the checkerboard perturbation, using the *v* and *r* method.

3.3. Iterative Approach

Since the results presented before are obtained after only one inversion, we attempt to improve the solution by iterating the synthetic inverse problem. That means that the solution models $\delta\mathbf{m}$ found in section 3.1 can be used as starting models of new inversions. Raypaths are traced, the tomography matrix accordingly recomputed, and the differences between observed (or, in the present case, synthetic) and computed (in the new model) travel times replace the data to be inverted. In the interest of computational speed, we do

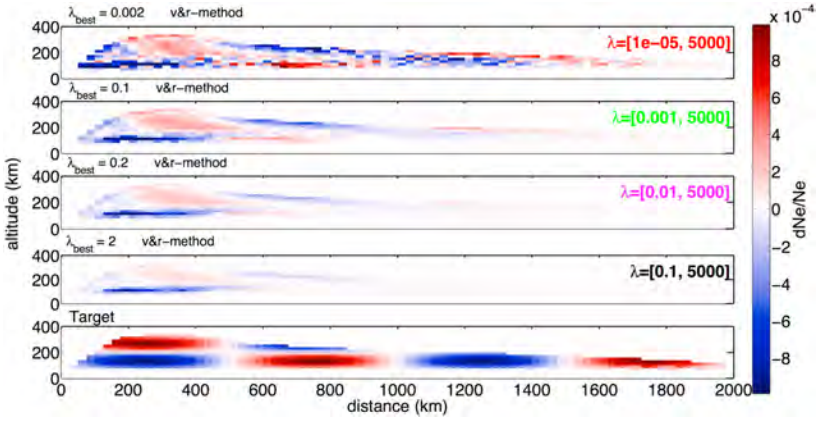


Figure 8. Inversion results using the *v* and *r* method with the best value of regularization chosen from the *L* curves in Figure 6b for a checkerboard perturbation in order to explore the sensitivity to the λ range. The bottom row shows the target model.

not conduct a separate *L* curve analysis (which would involve several inversions) at each iteration, but we fix λ to the trace of $\mathbf{A}^T \cdot \mathbf{A}$ divided by 10 (*v* method) or the trace of $(\mathbf{A} + \mathbf{A}')$ divided by 100 (*v* and *r* method) [*Press et al.*, 1992, chapter 18]. These values were determined with a few preliminary test to properly tune convergence speed. The trace of the matrix is equal to the sum of eigenvalues and allows for a quick estimation of the eigenvalues that are well determined. Based on the them, the regularization can be adjusted. The inspection of eigenvalues of the matrices showed that the λ selected by the *L* curve is to large in the case of the *v* method and too small for the *v* and *r* method compared to the largest eigenvalues. In the first case this is imposing large restriction to the solution, in the second case it is adding too much noise to the solution.

The results of this exercise are summarized in Figure 9. The difference between target and solution is less than 30% for the *v* and *r* at iteration 13 and less than 40% for the *v* method at iteration 5. Nevertheless, after a critical number of iterations, the discrepancy between solution and target model starts growing for both the *v* and *v* and *r* inversions. We interpret this as an effect of the mentioned trade-off between velocity heterogeneity and raypath deflection (section 3.1). Entries of the tomography matrix can be either negative or positive. Any given observation can be explained by a combination of both negative and positive

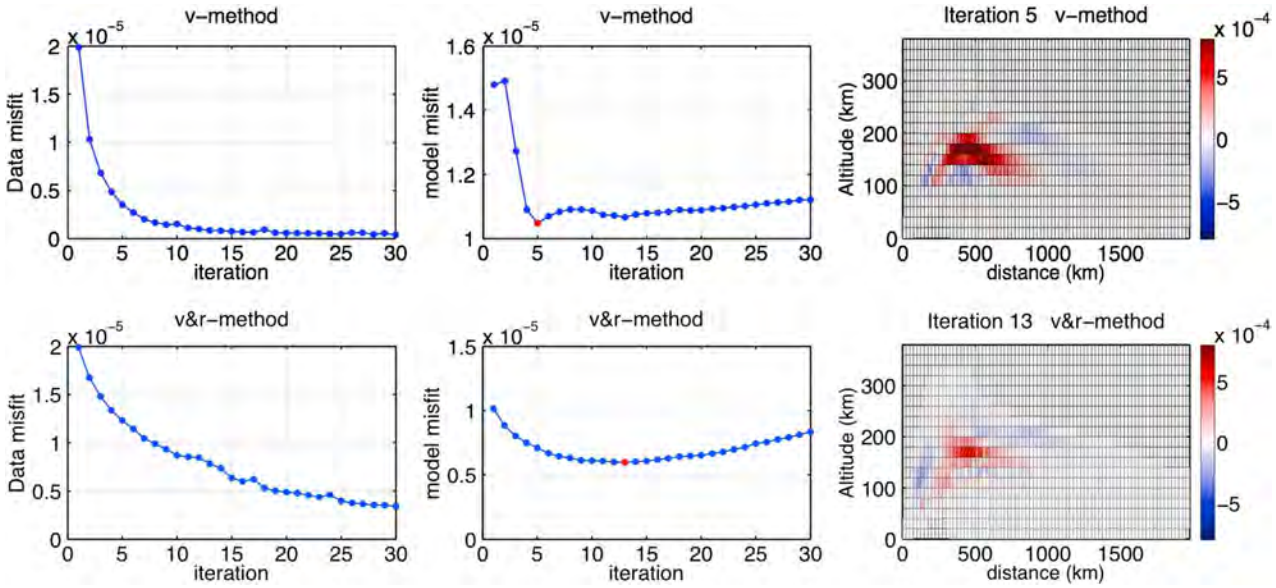


Figure 9. Data misfit $\sum_{j=1}^{\text{rays}} \left(\frac{\mathbf{dT}}{\mathbf{T}}\right)_j$ and model misfit as function of the number of iterations for the (top) *v* method and the (bottom) *v* and *r* method. The inversion results for the minimum (red point) of the model misfit (iteration 5 for *v* method and iteration 13 for *v* and *r* method) are shown on the right column.

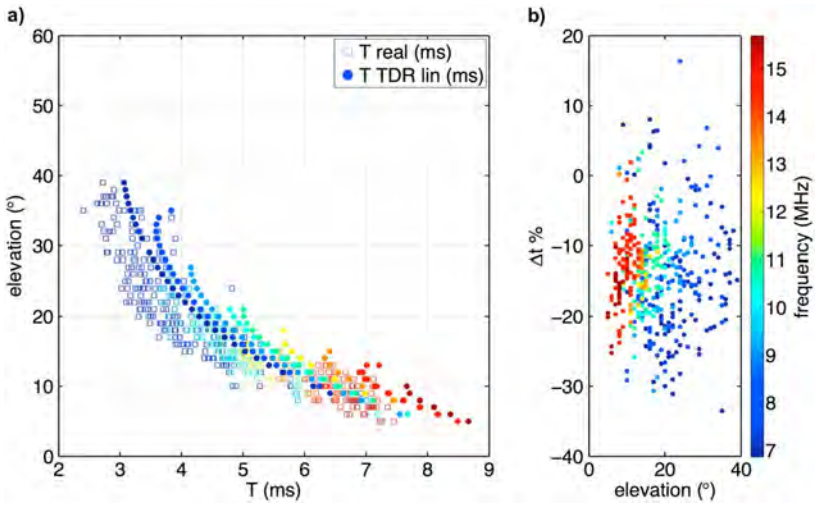


Figure 10. Propagation time of the real data and synthetic data simulated by ray tracing as (a) function of frequency and elevation angle, and the (b) difference (%).

heterogeneity whose amplitude might grow indefinitely as further iterations are performed. This problem is similar to that of selecting λ, and we suggest that the synthetic tests presented here can serve to determine adequate values for such parameters, to be used in real inversions with the same data coverage. This approach rests on the assumption that the effect of limitations in data coverage is more important to model resolution than that of data noise.

4. OTH Radar Nostradamus and Real Data Inversion

In the following section we apply the developed *v* and *r* method to real data of the OTH radar Nostradamus exploring the regularization parameter range with a $\lambda_{min} = 10^{-3}$ in accord with the results of the previous

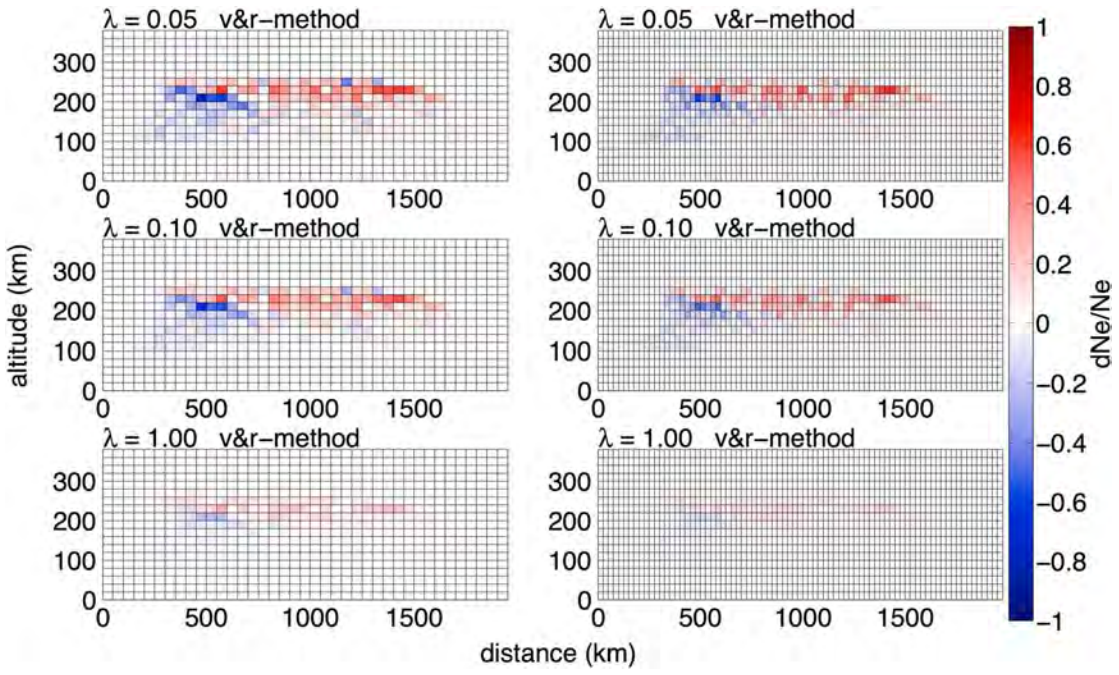


Figure 11. Inversion of real data collected 14 March 2006, at 18:55 UT with azimuth 247°, and for two different grids: (left) pixel size 50 × 20 km and (right) pixel size 25 × 20 km in distance and altitude. The middle row shows the result for the best regularization parameter λ_{best}.

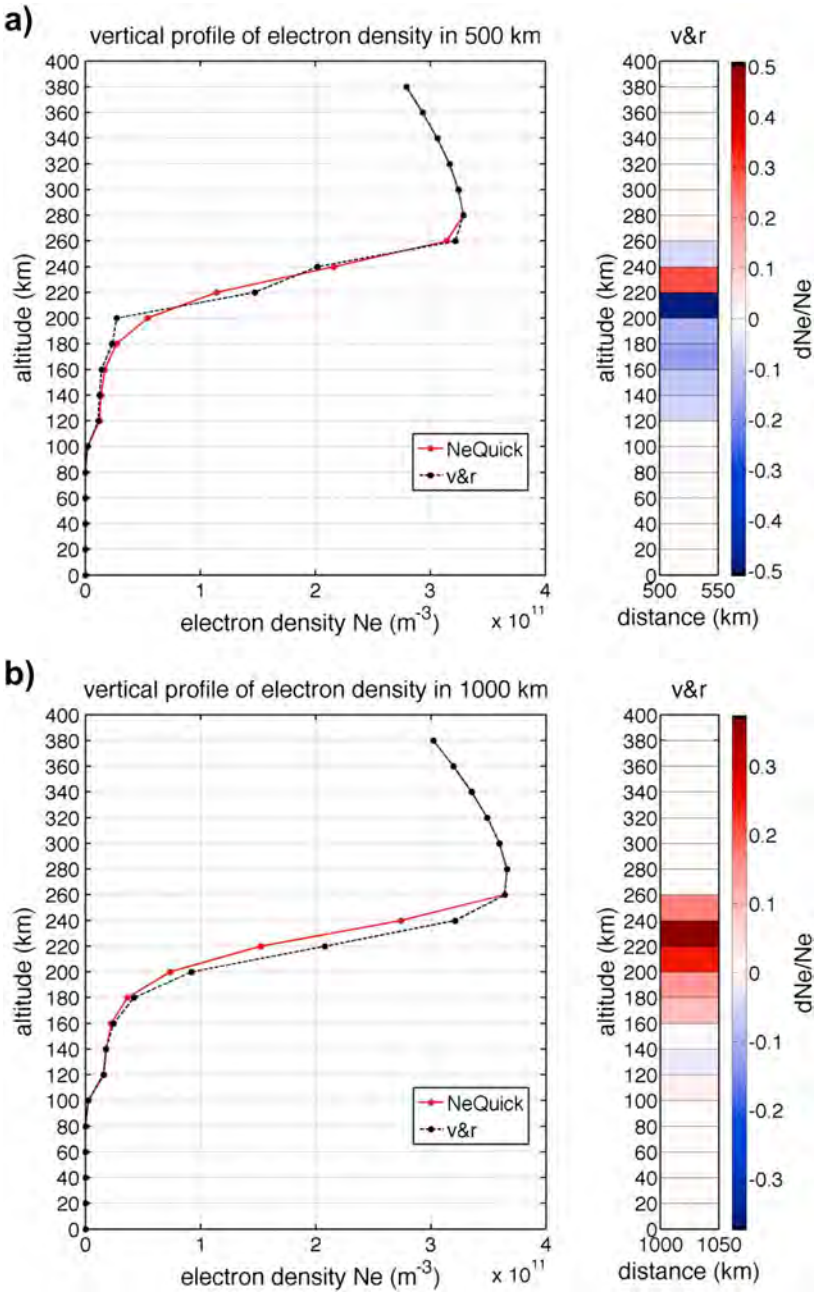


Figure 12. Vertical profiles of electron density at (a) 500 km distance and (b) 1000 km distance from the radar. The red line shows the ionospheric a priori model NeQuick, and the black line is the solution obtained by the *v* and *r* method (the a priori model plus the perturbation).

synthetic tests. Nostradamus is a monostatic radar that consists of 288 biconical antenna elements distributed over the arms of a three-branch star. The choice of this antenna arrangement allows beam forming with a coverage of 360° in azimuth and elevation control with a resolution of approximately 2.35° in azimuth and 5.43° in elevation for a frequency of 11 MHz. The central part of the array (96 antennas) is dedicated to transmission and reception, and the entire array is used for reception allowing a greater capability in receiving beam forming. The Nostradamus configuration allows to investigate a very large area of more than thousands of kilometer in range all around Europe.

We use here data that were collected on 14 March 2006, at 18:55 UT, for a frequency range 6–16 MHz and scanning in elevation from 10° to 60°. We traced rays with frequency and elevation angle corresponding to

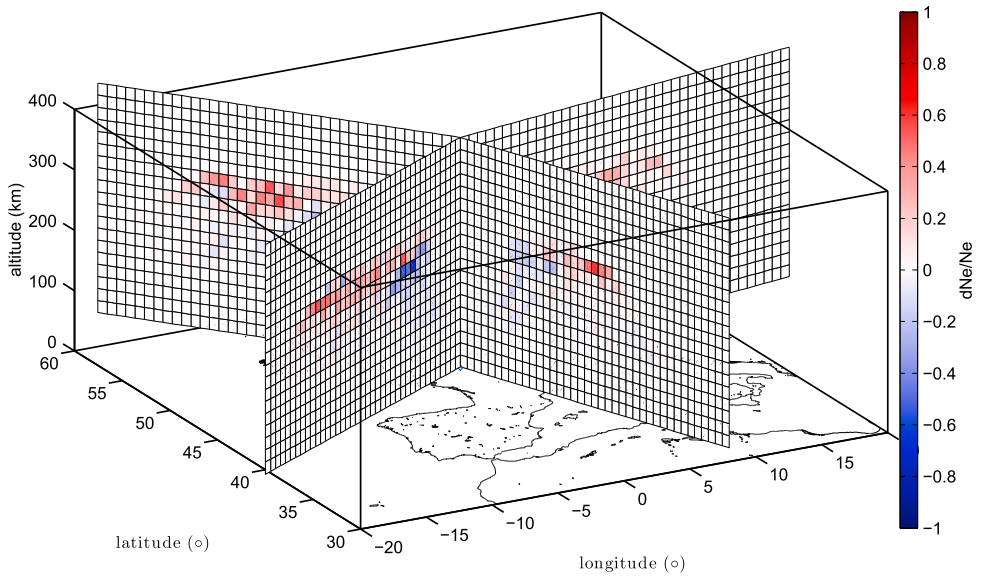


Figure 13. Electron density perturbation obtained by real data inversion measured by the OTH radar Nostradamus the 14 March 2006, at 18:55 UT in four azimuth directions, nominally 67°, 157°, 247°, and 337°.

the radar measurements in the a priori ionospheric model NeQuick given for this day and time. Figure 10 shows the comparison between data and the synthetics used to compute the vector $\delta\mathbf{T}$ of travel time perturbations for an azimuth of 247°. The difference is around 30% for rays with low frequency of 7–9 MHz, and around 20% for the highest frequencies of 16 MHz.

Figure 11 shows an example of the inversion of the real data in one selected azimuth (247°) for different values of regularization for the *v* and *r* method for two different grid sizes, i.e., (50 × 20 km and 25 × 20 km distance and altitude). There is no significant difference in the inversion results for this range of parameters, and the resolved perturbation is mainly located at 200 km altitude with a maximum of 50%. It is not surprising that the perturbation is located around 200 km altitude, since at this time (18:55 UT) the *E* layer has nearly disappeared and only the *F* layer remains strongly visible. The *v* and *r* method shows stable results for all regularization parameters, validating the reliability of the solution.

To further explore the inversion results, we calculated vertical profiles of electron density from the inversion with the best regularization parameter for two different distances from the radar (500 km and 1000 km), and we compare them with the electron density from the a priori model NeQuick. The resulting vertical profiles (Figure 12) show a sensitivity to variations located between 120 km and 260 km of altitude in accord with the sensitivity of the EM waves emitted by the radar. Indeed, the plasma frequency below 120 km altitude is too small to affect the emitted signal, and the maximum altitude of reflection is located around 260 km. The recovered electron density perturbation is of the order of 10%.

In order to show the potential applications of the OTH radar Nostradamus in ionospheric tomography, we inverted here the data for the same day and time measured for different azimuths: 67°, 157°, and 337°. The obtained tomographic images are plotted in Figure 13 over Europe, in the corresponding azimuthal directions. The electron density perturbation observed at 247° azimuth at 200 km altitude is visible in all azimuthal directions.

We highlight that the developed methodology could easily include other kinds of ionospheric data, in particular TEC measurements by ground-located or onboard GPS receivers. Consequently, we note the possibility of creating regional and/or global ionospheric tomography based on joint inversions of various ionospheric monitoring data.

5. Conclusions

We developed a new linear ionospheric tomography method for over-the-horizon radar (OTH) that takes into account not only the effect of the electron density perturbation on the velocity of electromagnetic waves (*v* method) but also the effect on the raypath deflection (*v* and *r* method). The characteristic uncertainty of end points of the raypath followed by EM waves emitted by OTH radar makes this second effect comparable to the first. Based on synthetic tests, we showed the necessity of taking into account the raypath deflection for tomographic inversions of the ionosphere. This is the first time that this problem is explored and emphasized in the OTH radar ionospheric tomography. Notwithstanding the methodological advance, the quality of the solution depends on the ray coverage, as well as, on the sensitivity of the rays to the medium that is the zone where the plasma frequency is close to the emission frequency of the propagating signal. This zone is strongly reduced at night, where only the high-frequency rays are reflected. The difference between target and solution for the localized perturbation is about 40% and 59% for the *v* and *r* and *v* methods, respectively, but can be reduced to 30% and 40% with iterations. For comparison we note that *Seemala et al.* [2014] archived a difference between target and solution of less than 10% with their GPS tomography over Japan during night.

Since the problem is underdetermined, meaning that the number of parameters to estimate is larger than the number of observations, our inversions have to be regularized. In order to find the best regularization parameter λ, we calculate a trade-off curve that quantifies the conflicting requirements between satisfying the data and the coherence with the a priori model. We highlight that the best regularization parameter λ_{best} is strongly dependent of the explored λ range. Consequently, we define a synthetic protocol to determine the λ range to explore, in order to minimize the solution error. The selected λ range can be applied to real data inversion in order to maximize the quality of the solution.

Application of the developed methodology on real data gives stable solutions, showing the quality of the inversion method and the reliability of the solution. A 3-D ionospheric tomography over Europe, based on the inversion of the OTH radar Nostradamus data, has been presented here in order to show the potential application of the developed method. Although our method has been developed for OTH radar, the number of these radars worldwide is limited to France (Nostradamus), United States (relocatable over-the-horizon radar), Australia (Jindalee), and North Pole (Super Dual Auroral Radar Network); consequently, we emphasize that this method can be developed further by including other ionospheric sounding techniques, in particular TEC measured by GPS, both measured at the ground with dense GPS array or by occultation with onboard GPS receivers. The latter is an interesting idea, because the two methods complement each other, where our method is sensitive to the lower ionosphere (up to 300 km altitude) during daytime, and the GPS because of its high frequency is sensitive to the region of maximum ionization (~300 km). Additionally, GPS data can compensate for the lack of reflected high-frequency (HF) rays during night.

Acknowledgments

This project is supported by the Programme National de Télédétection Spatiale (PNTS), grant PNTS-2014-07 and by the CNES grant SI-EuroTOMO. The Nostradamus data used in this work were supplied by ONERA. We thank the Aeronomy and Radiopropagation Laboratory of the Abdus Salam International Centre for Theoretical Physics (ICTP) for providing NeQuick model. The authors thank P. Lognonné for helpful theoretical discussions, as well as V. Rannou for discussions about real data inversions. We thank the reviewers for their constructive remarks.

Michael Balikhin thanks the reviewers for their assistance in evaluating the paper.

References

Andreeva, E. S., A. V. Galinov, V. E. Kunitsyn, Y. A. Mel'Nichenko, E. E. Tereschenko, M. A. Filimonov, and S. M. Chernyakov (1990), Radiotomographic reconstruction of ionization dip in the plasma near the Earth, *Soviet. J. Exp. Theo. Phys. Lett.*, *52*, 145–148.

Austen, J. R., S. J. Franke, and C. H. Liu (1988), Ionospheric imaging using computerized tomography, *Radio Sci.*, *23*(3), 299–307.

Bazin, V., J. Molinie, J. Munoz, P. Dorey, S. Saillant, G. Auffray, V. Rannou, and M. Lesturgie (2006), A general presentation about the OTH-radar Nostradamus, in *Radar, 2006 IEEE Conference on*, pp. 634–642, Electromagnetism and Radar Dept., ONERA, Palaiseau, France.

Bertel, D. G., and R. F. Cole (1988), The inversion of backscatter ionograms IPS radio and space services, *Technical Report, IPS-TR-88-03 (1988)*.

Bust, G. S., T. W. Garner, and T. L. Gaussiran (2004), Ionospheric data assimilation three-dimensional (IDA3D): A global, multisensor, electron density specification algorithm, *J. Geophys. Res.*, *109*, A11312, doi:10.1029/2003JA010234.

Coleman, C. J. (1998), A ray tracing formulation and its application to some problems in over-the-horizon radar, *Radio Sci.*, *33*(4), 1187–1197.

Fehmers, G. C., L. P. J. Kamp, F. W. Sluijter, and T. A. T. Spoelstra (1998), A model-independent algorithm for ionospheric tomography: 1. Theory and tests, *Radio Sci.*, *33*(1), 149–163.

Fremouw, E. J., J. A. Secan, and B. M. Howe (1992), Application of stochastic inverse theory to ionospheric tomography, *Radio Sci.*, *27*(5), 721–732.

Fridman, S. V. (1998), Reconstruction of a three-dimensional ionosphere from backscatter and vertical ionograms measured by over-the-horizon radar, *Radio Sci.*, *33*(4), 1159–1171.

Fridman, O. V., and S. V. Fridman (1994), A method of determining horizontal structure of the ionosphere from backscatter ionograms, *J. Atmos. Terr. Phys.*, *56*(1), 115–131.

Garcia, R., F. Crespon, V. Ducic, and P. Lognonné (2005), 3D Ionospheric tomography of post-seismic perturbations produced by Denali earthquake from GPS data, *Geophys. J. Int.*, *163*, 1049–1064, doi:10.1111/j.1365-246X.2005.02775.x.

Gordon, R., R. Bender, and G. T. Herman (1970), Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, *J. Theor. Biol.*, *29*(3), 471–481.
Hansen, A. J., T. Walter, and P. Enge (1997), Ionospheric correction using tomography, in *Proc. Inst. Nav. ION GPS-97, Kansas City*, pp. 249–260, Institute of Navigation, Manassas, Va. [Available at http://www.ion.org/publications/abstract.cfm?articleID=2796.]
Hajj, G. A., R. Ibanez-Meier, E. Kursinski, and L. Romans (1994), Imaging the ionosphere with the Global Positioning System, *Int. J. Imaging Syst. Technol.*, *5*(2), 174–187.
Heaton, J., S. Pryse, and L. Kersley (1995), Improved background representation, ionosonde input and independent verification in experimental ionospheric tomography, *Ann. Geophys.*, *13*(12), 1297–1302.
Hernández-Pajares, M., J. M. Juan, J. Sanz, and J. G. Solé (1998), Global observation of the ionospheric electronic response to solar events using ground and LEO GPS data, *J. Geophys. Res.*, *103*(A9), 20,789–20,796.
Kersley, L., J. A. T. Heaton, S. E. Pryse, and T. D. Raymund (1993), Experimental ionospheric tomography with ionosonde input and EISCAT verification, *Ann. Geophys.*, *11*, 1064–1074.
Kunitsyn, V. E., E. S. Andreeva, O. G. Razinkov, and E. D. Tereshchenko (1994a), Phase and phase-difference ionospheric radio tomography, *Int. J. Imaging Syst. Technol.*, *5*(2), 128–140.
Kunitsyn, V. E., E. S. Andreeva, E. D. Tereshchenko, B. Z. Khudukon, and T. Nygrén (1994b), Investigations of the ionosphere by satellite radiotomography, *Int. J. Imaging Syst. Technol.*, *5*(2), 112–127.
Landeau, T., F. Gauthier, and N. Ruelle (1997), Further improvements to the inversion of elevation-scan backscatter sounding data, *J. Atmos. Sol. Terr. Phys.*, *59*(1), 125–138.
Ma, X. F., T. Maruyama, G. Ma, and T. Takeda (2005), Three-dimensional ionospheric tomography using observation data of GPS ground receivers and ionosonde by neural network, *J. Geophys. Res.*, *110*, A05308, doi:10.1029/2004JA010797.
Mannucci, A. J., B. D. Wilson, and C. D. Edwards (1993), A new method for monitoring the Earth's ionospheric total electron content using GPS global network, in *Proceedings of the 6th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 1993)*, pp. 1323–1332, Salt Lake City, Utah, September 1993. [Available at http://www.ion.org/publications/abstract.cfm?articleID=4319.]
Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge (1998), A global mapping technique for GPS-derived ionospheric total electron content measurements, *Radio Sci.*, *33*(3), 565–582.
Markkanen, M., M. Lehtinen, T. Nygren, J. Pirttilä, P. Henelius, E. Vilenius, E. Tereshchenko, and B. Khudukon (1995), Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, *Ann. Geophys.*, *13*(12), 1277–1287.
Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory/William Menke*, Academic Press, San Diego, Calif.
Mitchell, C., D. Jones, L. Kersley, S. E. Pryse, and I. Walker (1995), Imaging of field-aligned structures in the auroral ionosphere, *Ann. Geophys.*, *13*(12), 1311–1319.
Na, H., and E. Sutton (1994), Resolution analysis of ionospheric tomography systems, *Int. J. Imaging Syst. Technol.*, *5*(2), 169–173.
National Imagery and Mapping Agency (2000), "Department of defense world geodetic system 1984: Its definition and relationships with local geodetic systems", *Tech. Rep., TR8350.2*, National Imagery and Mapping Agency, St. Louis, Mo.
Occhipinti, G. (2006), Observations multi-parametres et modélisation de la signature ionosphérique du grand séisme de Sumatra, PhD thesis, Institut de Physique du Globe de Paris, Paris, France.
Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992), *Numerical Recipes in Fortran: The Art of Scientific Computing* 2nd ed., Cambridge Univ. Press, New York.
Pryse, S., C. Mitchell, J. Heaton, and L. Kersley (1995), Travelling ionospheric disturbances imaged by tomographic techniques, *Ann. Geophys.*, *13*(12), 1325–1330.
Radicella, S. M., and R. Leitinger (2001), The evolution of the DGR approach to model electron density profiles, *Adv. Space Res.*, *27*(1), 35–40.
Rao, N. N. (1974), Inversion of sweep-frequency sky-wave backscatter leading edge for quasiparabolic ionospheric layer parameters, *Radio Sci.*, *9*(10), 845–847.
Rius, A., G. Ruffini, and L. Cucurull (1997), Improving the vertical resolution of ionospheric tomography with GPS occultations, *Geophys. Res. Lett.*, *24*(18), 2291–2294.
Ruelle, N., and T. Landeau (1994), Interpretation of elevation-scan HF backscatter data from Losquet Island radar, *J. Atmos. Terr. Phys.*, *56*(1), 103–114.
Seemala, G. K., M. Yamamoto, A. Saito, and C.-H. Chen (2014), Three-dimensional GPS ionospheric tomography over Japan using constrained least squares, *J. Geophys. Res. Space Physics*, *119*, 3044–3052, doi:10.1002/2013JA019582.
Snieder, R., and C. Spencer (1993), A unified approach to ray bending, ray perturbation and paraxial ray theories, *Geophys. J. Int.*, *115*(2), 456–470.
Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Pa.
Tikhonov, A. (1963), Solution of incorrectly problems and the regularization method, *Dokl. Akad. Nauk*, *4*, 1035–1038.
Vasicek, C., and G. Kronschnabl (1995), Ionospheric tomography: An algorithm enhancement, *J. Atmos. Terr. Phys.*, *57*(8), 875–888.
Wen, D., Y. Yuan, and J. Ou (2007), Monitoring the three-dimensional ionospheric electron density distribution using GPS observations over China, *J. Earth Syst. Sci.*, *116*(3), 235–244.
Wen, D., S. Liu, and P. Tang (2010), Tomographic reconstruction of ionospheric electron density based on constrained algebraic reconstruction technique, *GPS Solut.*, *14*, 375–380.
Yeh, K. C., and T. D. Raymund (1991), Limitations of ionospheric imaging by tomography, *Radio Sci.*, *26*(6), 1361–1380.
Yizengaw, E., P. Dyson, E. Essex, and M. Moldwin (2005), Ionosphere dynamics over the southern hemisphere during the 31 March 2001 severe magnetic storm using multi-instrument measurement data, *Ann. Geophys.*, *23*, 707–721.

Appendix C. Article: Effect of ray and speed perturbations on Ionospheric Tomography by Over-the-horizon radar: A new method

Bibliography

Åström, K. J. and R. M. Murray (2008). *Feedback Systems: An Introduction for Scientists and Engineers*.

American Institute of Aeronautics and Astronautics (1999). *Guide to Reference and Standard Ionosphere Models*. ANSI/AIAA G. American Institute of Aeronautics & Astronautics.

Andreeva, E. S., A. V. Galinov, V. E. Kunitsyn, Y. A. Mel'Nichenko, E. E. Tereschenko, M. A. Filimonov, and S. M. Chernyakov (1990, August). Radiotomographic reconstruction of ionization dip in the plasma near the earth. *Soviet Journal of Experimental and Theoretical Physics Letters 52*, 145–148.

Arbesser-Rastburg, B. (2006). The GALILEO single frequency ionospheric correction algorithm.

Austen, J. R., S. J. Franke, and C. H. Liu (1988). Ionospheric imaging using computerized tomography. *Radio Science 23*(3), 299–307.

Azzarone, A., C. Bianchi, M. Pezzopane, M. Pietrella, C. Scotto, and A. Settimi (2012). IONORT: A Windows software tool to calculate the HF ray tracing in the ionosphere. *Computers & Geosciences 42*(0), 57 – 63.

Baker, J., J. Ruohoniemi, A. Ribeiro, L. Clausen, R. Greenwald, N. Frissell, and K. Sterne (2010). Monitoring ionospheric space weather with the Super Dual Auroral Radar Network (SuperDARN). pp. 1414–1417.

Bazin, V., J. Molinie, J. Munoz, P. Dorey, S. Saillant, G. Auffray, V. Rannou, and M. Lesturgie (2006). A general presentation about the OTH-Radar Nostradamus. *Radar, 2006 IEEE Conference on*, 634–642.

Benito, E., A. Bourdillon, S. Saillant, V. Rannou, and J. Molinié (2008). Inversion of HF backscatter ionograms using elevation scans. *Journal of Atmospheric and Solar-Terrestrial Physics 70*(15), 1935 – 1948.

Bilitza, D. (1990). International reference ionosphere 1990.

Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. Reinisch (2014). The International Reference Ionosphere 2012 – a model of international collaboration. *J. Space Weather Space Clim. 4*, A07.

Budden, K. G. (1985). *The Propagation of Radio Waves*. Cambridge University Press.

Bust, G. S., T. W. Garner, and T. L. Gaussiran (2004). Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm. *Journal of Geophysical Research: Space Physics 109*(A11).

Chapman, S. (1931). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *proceedings of the Physical Society 43*(1), 26.

Coïsson, P., G. Occhipinti, P. Lognonné, J.-P. Molinié, and L. M. Rolland (2011). Tsunami signature in the ionosphere: A simulation of OTH radar observations. *Radio Sci. 46*.

Coïsson, P., S. M. Radicella, R. Leitinger, and B. Nava (2006). Topside electron density in IRI and NeQuick: Features and limitations. *Advances in Space Research 37*(5), 937 – 942.

Colegrove, S. (2000). Project Jindalee: from bare bones to operational OTHR. In *Radar Conference, 2000. The Record of the IEEE 2000 International*, pp. 825–830.

Cook, J. and S. Close (1995). An investigation of TID evolution observed in MACE'93 data. *Annales geophysicae 13*(12), 1320–1324.

Croft, T. A. and H. Hoogasian (1968). Exact ray calculations in a quasi-parabolic ionosphere with no magnetic field. *Radio Science 3*, 69.

Daniell, R. E., L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka, and R. W. Schunk (1995). Parameterized ionospheric model: A global ionospheric parameterization based on first principles models. *Radio Science 30*(5), 1499–1510.

Davies, K. (1965). *Ionospheric radio propagation*. Monograph. U. S. Dept. of Commerce, National Bureau of Standards; for sale by the Superintendent of Documents, U. S. Govt. Print. Off.

Davies, K. (1990). *Ionospheric Radio*, Volume 31 of *IEE electromagnetic wave series*. London: Peter Peregrinus Ltd.

Dear, R. M. and C. N. Mitchell (2007). Ionospheric imaging at mid-latitudes using both GPS and ionosondes. *Journal of atmospheric and solar-terrestrial physics 69*(7), 817–825.

Di Giovanni, G. and S. Radicella (1990). An analytical model of the electron density profile in the ionosphere. *Advances in Space Research 10*(11), 27 – 30.

Fehmers, G. C., L. P. J. Kamp, F. W. Sluijter, and T. A. T. Spoelstra (1998). A model-independent algorithm for ionospheric tomography: 1. Theory and tests. *Radio Science 33*(1), 149–163.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics 21*, 768–769.

Fremouw, E. J., J. A. Secan, and B. M. Howe (1992). Application of stochastic inverse theory to ionospheric tomography. *Radio Science 27*(5), 721–732.

Fridman, O. V. and S. V. Fridman (1994). A method of determining horizontal structure of the ionosphere from backscatter ionograms. *Journal of Atmospheric and Terrestrial Physics 56*(1), 115 – 131.

Fridman, S. V. (1998). Reconstruction of a three-dimensional ionosphere from backscatter and vertical ionograms measured by over-the-horizon radar. *Radio Sci. 33*(4), 1159–1171.

Fridman, S. V., L. Nickisch, and M. Hausman (2012). Inversion of backscatter ionograms and TEC data for over-the-horizon radar. *Radio Science 47*.

Fridman, S. V. and L. J. Nickisch (2001). Generalization of ionospheric tomography on diverse data sources: Reconstruction of the three-dimensional ionosphere from simultaneous vertical ionograms, backscatter ionograms, and total electron content data. *Radio Sci. 36*(5), 1129–1139.

Fridman, S. V., L. J. Nickisch, M. Aiello, and M. Hausman (2006). Real-time reconstruction of the three-dimensional ionosphere using data from a network of GPS receivers. *Radio Science 41*(5).

Fridman, S. V., L. J. Nickisch, and M. Hausman (2009). Personal-computer-based system for real-time reconstruction of the three-dimensional ionosphere using data from diverse sources. *Radio Science 44*(3).

Garcia, R. and F. Crespon (2008). Radio tomography of the ionosphere: Analysis of an underdetermined, ill-posed inverse problem, and regional application. *Radio Sci. 43*.

Georges, T. M., J. A. Harlan, T. N. Lee, and R. R. Leben (1998). Observations of the Florida Current with two over-the-horizon radars. *Radio Science 33*(4), 1227–1239.

Gilbert, P. (1972). Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology 36*(1), 105 – 117.

Gillmor, C. S. (1982). Wilhelm Altar, Edward Appleton, and the Magneto-Ionic Theory. *Proceedings of the American Philosophical Society 126*(5), pp. 395–440.

Gordon, R., R. Bender, and G. T. Herman (1970). Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology 29*(3), 471–481.

Greenwald, R. A., K. B. Baker, J. R. Dudeney, M. Pinnock, T. B. Jones, E. C. Thomas, J.-P. Villain, J.-C. Cerisier, C. Senior, C. Hanuise, R. D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A. D. M. Walker, N. Sato, and

H. Yamagishi (1995, February). Darn/Superdarn: A Global View of the Dynamics of High-Lattitude Convection. *71*, 761–796.

Hagfors, T. and K. Schlegel (2001). Earth's ionosphere. In J. Bleeker, J. Geiss, and M. Huber (Eds.), *The Century of Space Science*, pp. 1559–1584. Springer Netherlands.

Hajj, G. A., R. Ibanez-Meier, E. Kursinski, and L. Romans (1994). Imaging the ionosphere with the Global Positioning System. *International Journal of Imaging Systems and Technology 5*(2), 174–187.

Hansen, A. J., T. Walter, and P. Enge (1997). Ionospheric correction using tomography. *Proceedings of Institute of Navigation ION GPS-97, Kansas City*, 249–260.

Hatfield, V. (1970). Derivation of ionospheric parameters from backscatter data. In I. Forecasting (Ed.), *AGARD Conf. Proc.*, Volume 49, pp. 16–9 – 16.9.

Hawass, Z. (2005). Tutankhamun CT Scan. `http://www.guardians.net/hawass/press_release_tutankhamun_ct_scan_results.htm`.

Headrick, J. M. and S. J. Anderson (2008). HF Over-the-Horizon Radar. In M. I. Skolnik (Ed.), *Radar Handbook*, Chapter 20. McGraw-Hill.

Heaton, J., G. Jones, and L. Kersley (1996). Toward ionospheric tomography in Antarctica: first steps and comparison with dynasonde observations. *Antarctic Science 8*, 297–302.

Heaton, J., S. Pryse, and L. Kersley (1995). Improved background representation, ionosonde input and independent verification in experimental ionospheric tomography. *Annales geophysicae 13*(12), 1297–1302.

Hernández-Pajares, M., J. M. Juan, J. Sanz, and J. G. Solé (1998). Global observation of the ionospheric electronic response to solar events using ground and LEO GPS data. *Journal of Geophysical Research: Space Physics 103*(A9), 20789–20796.

Hestenes, M. R. and E. Stiefel (1952). Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards 49*, 409–436.

Huba, J. D., G. Joyce, and J. A. Fedder (2000). Sami2 is Another Model of the Ionosphere (SAMI2): A new low-latitude ionosphere model. *Journal of Geophysical Research: Space Physics 105*(A10), 23035–23053.

Jones, R. M. and J. J. Stephenson (1975). A versatile three-dimensional ray tracing computer program for radio waves in the ionosphere. Report 75-76 PB2488567, OT.

Karpenko (2014). The radar station (ZGRLS) of 29b6 below the horizon "CONTAINER". http://bastion-karpenko.ru/container/.

Kersley, L., J. A. T. Heaton, S. E. Pryse, and T. D. Raymund (1993). Experimental ionospheric tomography with ionosonde input and EISCAT verification. *Annales Geophysicae 11*, 1064–1074.

Kersley, L., S. Pryse, I. Walker, J. Heaton, C. Mitchell, M. Williams, and C. Willson (1997). Imaging of electron density troughs by tomographic techniques. *Radio Science 32*(4), 1607–1621.

Kunitake, M., K. Ohtaka, T. Maruyama, M. Tokumaru, A. Morioka, and S. Watanabe (1995). Tomographic imaging of the ionosphere over Japan by the modified truncated SVD method. *Annales geophysicae 13*(12), 1303–1310.

Kunitsyn, V. E., E. S. Andreeva, O. G. Razinkov, and E. D. Tereshchenko (1994). Phase and phase-difference ionospheric radio tomography. *International Journal of Imaging Systems and Technology 5*(2), 128–140.

Kunitsyn, V. E., E. S. Andreeva, E. D. Tereshchenko, B. Z. Khudukon, and T. Nygrén (1994). Investigations of the ionosphere by satellite radiotomography. *International Journal of Imaging Systems and Technology 5*(2), 112–127.

Landeau, T., F. Gauthier, and N. Ruelle (1997). Further improvements to the inversion of elevation-scan backscatter sounding data. *Journal of Atmospheric and Solar-Terrestrial Physics 59*(1), 125 – 138.

Lay, T. and T. Wallace (1995). *Modern Global Seismology*. International Geophysics. Elsevier Science.

Leitinger, R., M. L. Zhang, and S. M. Radicella (2005). An improved bottomside for the ionospheric electron density model NeQuick. *Annals of Geophysics 48*(3), 535–534.

Lévêque, J.-J., L. Rivera, and G. Wittlinger (1993). On the use of the checkerboard test to assess the resolution of tomographic inversions. *Geophysical Journal International 115*(1), 313–318.

Liu, B.-Y. (2007). *HF Over-The-horizon Radar system Performance analysis*. Ph. D. thesis, Naval Postgraduate School.

Luhmann, J. (1995). Ionospheres. In M. G. Kivelson and C. T. Russell (Eds.), *Introduction to space physics*. Cambridge university press.

Ma, X. F., T. Maruyama, G. Ma, and T. Takeda (2005). Three-dimensional ionospheric tomography using observation data of GPS ground receivers and ionosonde by neural network. *Journal of Geophysical Research: Space Physics 110*(A5).

Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge (1998). A global mapping technique for GPS-derived ionospheric total electron content measurements. *Radio Science 33*(3), 565–582.

Markkanen, M., M. Lehtinen, T. Nygren, J. Pirttilä, P. Henelius, E. Vilenius, E. Tereshchenko, and B. Khudukon (1995). Bayesian approach to satellite radiotomography with applications in the Scandinavian sector. *Annales geophysicae 13*(12), 1277–1287.

Martyn, D. F. (1935). The propagation of medium radio waves in the ionosphere. *Proceedings of the Physical Society 47*(2), 323.

Menke, W. (1989). *Geophysical data analysis : discrete inverse theory / William Menke*. Academic Press San Diego.

Mitchell, C., D. Jones, L. Kersley, S. E. Pryse, and I. Walker (1995). Imaging of field-aligned structures in the auroral ionosphere. *Annales geophysicae 13*(12), 1311–1319.

Mitchell, C. N. and P. S. Spencer (2003). A three-dimensional time-dependent algorithm for ionospheric imaging using GPS. *Annals of Geophysics 46*(4).

Molinié, J.-P. (2011). NOSTRADAMUS, over the horizon radar for early warning and long range detection. confidential personal communication.

Na, H. and E. Sutton (1994). Resolution analysis of ionospheric tomography systems. *International Journal of Imaging Systems and Technology 5*(2), 169–173.

Najman, P. and T. Kos (2014). Performance Analysis of Empirical Ionosphere Models by Comparison with CODE Vertical TEC Maps.

National Imagery and Mapping Agency (2000). Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems. Technical Report TR8350.2, National Imagery and Mapping Agency, St. Louis, MO, USA.

Nava, B., P. Coïsson, and S. M. Radicella (2008). A new version of the NeQuick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics 70*(15), 1856 – 1862.

Norman, R. J. and P. S. Cannon (1997). A two-dimensional analytic ray tracing technique accommodating horizontal gradients. *Radio Sci. 32*(2), 387–396.

Nygren, T., M. Markkanen, M. Lehtinen, E. Tereshchenko, B. Khudukon, O. Evstafiev, and P. Pollari (1997). Comparison of F-region electron density observations by satellite radio tomography and incoherent scatter methods. *Annales Geophysicae 14*(12), 1422–1428.

Occhipinti, G. (2006). *Observations multi-parametres et modelisation de la signature ionospherique du grand seisme de Sumatra*. Ph. D. thesis, Institut de Physique du Globe de Paris.

Occhipinti, G., P. Dorey, T. Farges, and P. Lognonné (2010). Nostradamus: The radar that wanted to be a seismometer. *Geophys. Res. Lett. 37*.

Pokhotelov, D., P. Jayachandran, C. Mitchell, J. MacDougall, and M. Denton (2011). GPS tomography in the polar cap: comparison with ionosondes and in situ spacecraft data. *GPS Solutions 15*(1), 79–87.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992). *Numerical Recipes in Fortran: The Art of Scientific Computing* (2 ed.). Cambridge University Press.

Pryse, S. and L. Kersley (1992). A preliminary experimental test of ionospheric tomography. *Journal of Atmospheric and Terrestrial Physics 54*(7–8), 1007 – 1012.

Pryse, S., C. Mitchell, J. Heaton, and L. Kersley (1995). Travelling ionospheric disturbances imaged by tomographic techniques. *Annales geophysicae 13*(12), 1325–1330.

Radicella, S. M. and R. Leitinger (2001). The evolution of the DGR approach to model electron density profiles. *Advances in Space Research 27*(1), 35 – 40.

Rao, N. N. (1974). Inversion of sweep-frequency sky-wave backscatter leading edge for quasiparabolic ionospheric layer parameters. *Radio Science 9*(10), 845–847.

Raymund, T. D. (1994). Ionospheric tomography algorithms. *International Journal of Imaging Systems and Technology 5*(2), 75–85.

Raymund, T. D., J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker (1990). Application of computerized tomography to the investigation of ionospheric structures. *Radio Science 25*(5), 771–789.

Rees, M. H. (1989). *Physics and Chemistry of the Upper Atmosphere*, Volume 1. Cambridge University Press.

Reinisch, B. W. and H. Xueqin (1983). Automatic calculation of electron density profiles from digital ionograms: 3. Processing of bottomside ionograms. *Radio Science 18*(3), 477–492.

Rius, A., G. Ruffini, and L. Cucurull (1997). Improving the vertical resolution of ionospheric tomography with GPS occultations. *Geophysical Research Letters 24*(18), 2291–2294.

Roy, C., G. Occhipinti, L. Boschi, J.-P. Molinié, and M. Wieczorek (2014). Effect of ray and speed perturbations on ionospheric tomography by over-the-horizon radar: A new method. *Journal of Geophysical Research: Space Physics*.

Ruelle, N. and T. Landeau (1994). Interpretation of elevation-scan HF backscatter data from Losquet Island radar. *Journal of Atmospheric and Terrestrial Physics 56*(1), 103 – 114.

Saksman, E., T. Nygrén, and M. Markkanen (1997). Ionospheric structures invisible in satellite radiotomography. *Radio Science 32*(2), 605–616.

Schunk, R. (2013). *Ionospheric Models for Earth*. American Geophysical Union.

Schunk, R. and A. Nagy (2009). *Ionospheres*. Cambridge University Press.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, M. Hairston, and B. M. Howe (2004). Global Assimilation of Ionospheric Measurements (GAIM). *Radio Science 39*(1).

Seemala, G. K., M. Yamamoto, A. Saito, and C.-H. Chen (2014). Three-dimensional GPS ionospheric tomography over Japan using constrained least squares. *Journal of Geophysical Research: Space Physics 119*(4), 3044–3052.

Six, M., J. Parent, A. Bourdillon, and J. Delloue (1996). A new multibeam receiving equipment for the Valensole skywave HF radar: description and applications. *Geoscience and Remote Sensing, IEEE Transactions on*.

Snieder, R. and C. Spencer (1993). A unified approach to ray bending, ray perturbation and paraxial ray theories. *Geophysical Journal International 115*(2), 456–470.

Tape, C., Q. Liu, and J. Tromp (2007). Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves. *Geophysical Journal International 168*(3), 1105–1129.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Thomas, R. M. and D. J. Netherway (1989). Observations of meteors using over-the-horizon radar. *Proceedings of the Astronomical Society of Australia 8*, 88–93.

Thomason, J. F. (2003). Development of Over-the-Horizon Radar in the United States. In I. R. . C. 2003 (Ed.), *IEEE Radar 2003 Conference 2003, Adelaide, Australia, September 3-5*, pp. 599–601.

Tikhonov, A. (1963). Solution of incorrectly problems and the regularization method. *Doklady Akademii nauk SSSR 4*, 1035–1038.

Titheridge, J. (1985). Ionogram analysis with the generalised program POLAN. Technical report, World Data Center A for Solar-Terrestrial Physics, Boulder, CO (USA).

Tsai, L. C., C. H. Liu, and J. Y. Huang (2010). Three-dimensional numerical ray tracing on a phenomenological ionospheric model. *Radio Sci. 45*(5).

van de Kamp, M. M. J. L. (2013). Medium-scale 4-D ionospheric tomography using a dense GPS network. *Annales Geophysicae 31*, 75–89.

Vasicek, C. and G. Kronschnabl (1995). Ionospheric tomography: an algorithm enhancement. *Journal of Atmospheric and Terrestrial Physics 57*(8), 875 – 888.

Walker, I., J. Heaton, L. Kersley, C. Mitchell, S. Pryse, and M. Williams (1997). EISCAT verification in the development of ionospheric tomography. *Annales Geophysicae 14*(12), 1413–1421.

Wawerzinek, B., J. Ritter, and C. Roy (2013). New constraints on the 3D shear wave velocity structure of the upper mantle underneath Southern Scandinavia revealed from non-linear tomography. *Tectonophysics 602*(0), 38 – 54. TOPO-EUROPE III.

Wen, D., S. Liu, and P. Tang (2010). Tomographic reconstruction of ionospheric electron density based on constrained algebraic reconstruction technique. *GPS Solutions 14*, 375–380.

Wen, D., Y. Yuan, and J. Ou (2007). Monitoring the three-dimensional ionospheric electron density distribution using GPS observations over China. *Journal of earth system science 116*(3), 235–244.

Yeh, K. C. and T. D. Raymund (1991). Limitations of ionospheric imaging by tomography. *Radio Science 26*(6), 1361–1380.

Yizengaw, E., P. Dyson, E. Essex, and M. Moldwin (2005). Ionosphere dynamics over the Southern Hemisphere during the 31 March 2001 severe magnetic storm using multi-instrument measurement data. In *Annales Geophysicae*, Volume 23. European Geosciences Union.

Young, G., J. Harlan, and T. Georges (1997). Application of over-the-horizon radar observations to synoptic and mesoanalysis over the Atlantic. *Weather and forecasting 12*(1), 44–55.

Acronyms

NO^+ ionized nitric oxide. 23

N_2 molecular nitrogen. 21

O^+ ionized oxygen. 23

O_2 dioxygen. 21

O_2^+ ionized molecular oxygen. 23

3D 3 Dimensions. 39, 42, 51, 53, 54, 58–61, 156, 163

AEROS Aeronomy Satellite. 29

AGW Acoustic Gravity Waves. 51

ART The Algebraic Reconstruction Technique. 56

ARTIST Automatic Real-Time Ionogram Scaler with True height. 36

CART constrained algebraic reconstruction technique. 56

CCIR Comité Consultatif International pour la Radio. 30

CGM Conjugate Gradient Method. 85–89, 163

CHAMP CHAllenging Minisatellite Payload. 41, 58

CME Coronal Mass Ejections. 26

COSMIC Constellation Observing System for Meteorology Ionosphere and Climate. 41

COSPAR Commitee on Space Research. 29

DIAS European DIgital upper Atmosphere Server. 149

Digisonde Digital Ionospheric Goniometric Ionosonde. 35–37

EISCAT European Incoherent Scatter. 29, 56, 59

EM electromagnetic. 19, 25, 26, 32–35, 43, 54, 57, 60–62, 65, 70, 73, 76, 88, 106, 142, 159

EPN EUREF Permanent Network. 42

EUV extreme ultraviolet. 23, 25, 29

GAIM Global Assimilation of Ionospheric Measurements. 28

GEONET GPS Earth Observation Network. 42

GLONASS Global'naya Navigatsionnaya Sputnikovaya Sistema. 41

GNSS Global Navigation Satellite System. 27, 41, 42

GPS Global Positioning System. 26, 28, 35, 40–42, 54–60, 145, 153–157, 159, 162, 163, 168

HF High Frequency. 22, 24, 25, 36, 43, 46, 50, 60, 163

ICTP International Centre for Theoretical Physics. 30

IGS International GNSS Service. 42, 55

IGW Internal Gravity Wave. 51

IMF Interplanetary Magnetic Field. 26

IRI International Reference Ionosphere. 28–30, 62

IRNSS Indian Regional Navigation Satellite System. 41

ISIS International Satellites for Ionospheric Studies. 29

ISR Incoherent Scatter Radar. 28

ITU-R International Telecommunication Union-Radiocommunication. 29, 30

JORN Jindalee Operational Radar Network. 45, 47, 162

LEO Low Earth Orbit. 41, 42, 57, 58

LOS Line Of Sight. 41

MADRE Magnetic-Drum Radar Equipment. 44, 45

MAGNUS MAntle investiGations of Norwegian Uplift Structures. 175

MART Multiplicative Algebraic Reconstruction Technique. 56

MIDAS Multi-Instrument Data Analysis System. 57

MQP Multi-Quasi-Parabolic. 144, 149, 150

NIMS Navy Ionospheric Monitoring System. 40

NO nitric oxide. 22

NOAA National Oceanic and Atmospheric Administration. 36, 38, 141, 167

Nostradamus NOuveau Système TRAnshorizon Décamétrique Appliquant les Méthodes Utilisées en Studio. iii, 15–17, 45, 47, 48, 61, 77, 133, 139, 143, 144, 146, 147, 149, 150, 155, 157, 161–163

NRL Naval Research Laboratory. 44

ONERA Office National d'Etudes et de Recherche Aérospatiale. 47

OpenMP Open Multi-Processing. 110

OTH Over-the-horizon. 26, 34, 38–40, 43–45, 47, 49–51, 54, 58, 60, 61, 70, 105, 133, 143, 144, 146, 150, 152–155, 157, 159, 161–163, 167, 170, 171

OTH-B Over-the-horizon-Backscatter. 45

PIM Parametrized Ionosphere Model. 28

POLAN POLynomial ANalysis. 36

QC quasi-cubic. 60

QP quasi-parabolic. 39, 60

QVI quasi-vertical-incident. 39

QZSS Quasi-Zenith Satellite System. 41

ROTHR Relocatable Over The Horizon Radar. 45, 47, 162

SAMI2 Sami2 is Another Model of the Ionosphere. 28

SFU solar flux units. 62, 73, 87, 109, 141, 146

SID sudden ionospheric disturbance. 27

SIRT simultaneous iterative reconstructon technique. 55, 56

STEC Slant Total Electron Content. 54, 55

SuperDARN Super Dual Auroral Radar Network. 46, 47, 50, 157, 162, 167

TDR Tracé De Rayons. 71, 73, 87, 110

TEC Total Electron Content. 25, 26, 28, 40, 41, 54–56, 58, 59, 153, 154, 163

TID Traveling Ionospheric Disturbances. 19, 59

URSI Union Radio-Scientifique Internationale. 29

USA United States of America. 44, 162

UV Ultraviolet. 20, 22, 27

VHF Very High Frequency. 46

WARF Wide-Aperture Research Facility. 45

Ionospheric tomography by Over-The-Horizon radar

Most recent methods in ionospheric tomography are based on the inversion of the total electron content measured by ground-based GPS receivers. As a consequence of the high frequency of the GPS signal and the absence of horizontal raypaths, the electron density structure is mainly reconstructed in the F2 region (300 km), where the ionosphere reaches the maximum of ionization, and is not sensitive to the lower ionospheric structure.
To overcome these limitations, a new tomographic method of the lower ionosphere, based on the full inversion of over-the-horizon (OTH) radar data has been developed in this thesis.
This method takes into account the effect of the electron density perturbation on the velocity of the electromagnetic waves, as well as the ray-path deflection. This last point is extremely critical for OTH radar inversions as the emitted signal propagates through the ionosphere between a fixed starting point (the radar) and an unknown end point on the Earth surface where the signal is backscattered.
The first part of this work is based on the theoretical development of the tomography method of Over-the-Horizon radar, as well as the validation on synthetic benchmark tests and resolution analysis.
After validation, the method has been applied to real data from the Over-the-horizon radar Nostradamus, showing the potential of a 3D ionospheric tomography over Europe.

Mots-clés : IONOSPHERE ; OVER THE HORIZON RADAR ; NOSTRADAMUS ; TOMOGRAPHY ; INVERSE PROBLEM

Tomographie de l’ionosphère par radar transhorizon à onde de ciel

La plupart des méthodes récentes de tomographie de l’ionosphère sont basées sur l'inversion du contenu électronique total mesuré par des récepteurs GPS basés au sol. A cause de la haute fréquence du signal GPS et l'absence de rayon horizontal, la structure de la densité électronique est principalement reconstruite dans la région F2 (300 km d’altitude), où l'ionosphère atteint le maximum d'ionisation. Cette reconstruction est peu sensible à la structure de l’ionosphère aux altitudes inférieures.

Pour surmonter ces limitations, une nouvelle méthode de tomographie des couches basses de l'ionosphère, basée sur l'inversion complète de données de radar transhorizon a été développée dans cette thèse.

Cette méthode prend en compte l'effet des perturbations de la densité électronique sur la vitesse des ondes électromagnétiques, ainsi que la déviation du trajet des ondes due à ces perturbations. Ce dernier point est extrêmement important pour l’inversion des données de radar transhorizon. En effet le signal émis se propage par l'ionosphère entre un point de départ fixe (le radar) et un point d’arrivée inconnu sur la surface de la Terre où le signal est rétrodiffusé.
La première partie de ce travail est basée sur le développement théorique de la méthode de tomographie pour radar transhorizon, ainsi que la validation sur des données synthétiques et l'analyse des performances obtenues.
Après la validation, la méthode a été appliquée à des données obtenues avec le radar transhorizon Nostradamus, montrant ainsi le potentiel d'une tomographie en 3 dimensions de l’ionosphère sur l'Europe.

Keywords : IONOSPHERE ; RADAR TRANSHORIZON ; NOSTRADAMUS ; TOMOGRAPHIE ; PROBLEME INVERSE

Office National d'Études et de Recherches Aérospatiales
BP 72 - 29 avenue de la Division Leclerc
92322 CHATILLON CEDEX
Tél. : +33 1 46 73 40 40 - Fax : +33 1 46 73 41 41
http://www.onera.fr