



รายงานวิจัยฉบับสมบูรณ์

โครงการ: การแก้ปัญหาย้อนกลับแบบผสมระหว่าง CG
และ Occam สำหรับข้อมูล Magnetotelluric
(A Hybrid CG/Occam Inversion for
Magnetotelluric Data)

รศ. ดร. วีระชัย สิริพันธ์วรารักษ์

ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

มิถุนายน 2553

รายงานวิจัยฉบับสมบูรณ์

โครงการ: การแก้ปัญหาย้อนกลับแบบผสมระหว่าง CG และ
Occam สำหรับข้อมูล Magnetotelluric
(A Hybrid CG/Occam Inversion for
Magnetotelluric Data)

รศ. ดร. วีระชัย สิริพันธ์วรารณ์

ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

บทคัดย่อ

ในโครงการนี้ เราเสนอโปรแกรมใหม่ที่ใช้ในการแปลความหมายข้อมูล Magnetotelluric โปรแกรมนี้เรียกว่า WSMIX3DMT เป็นโปรแกรมที่มีพื้นฐานความคิดมาจากสองโปรแกรมเก่า คือ data space conjugate gradient (WSDCG3DMT) และ data space Occam's inversion (WSINV3DMT) โปรแกรม WSMIX3DMT นี้เป็นโปรแกรมที่ดัดแปลงทางคณิตศาสตร์มาจากโปรแกรม WSDCG3DMT ดังนั้นหน่วยความจำที่ต้องใช้จึงเล็กน้อยเมื่อเทียบกับ WSINV3DMT เหมือนกับของ WSDCG3DMT แต่แทนที่เราจะใช้ trade-off parameter ตัวเดิมตลอด inversion เราเปลี่ยนแปลงไปเรื่อยๆ เหมือนกับที่ทำใน WSINV3DMT แต่การเปลี่ยนแปลงนี้เป็นไปตาม run-time ไม่ได้เป็นไปตาม data misfit กระบวนการที่เราใช้นี้ทำให้โปรแกรม WSMIX3DMT รันได้เร็วกว่าทั้ง WSDCG3DMT และ WSINV3DMT และในขณะเดียวกันก็ใช้หน่วยความจำน้อยกว่า สิ่งนี้ทำให้ WSMIX3DMT เป็นโปรแกรม inversion ที่มีประสิทธิภาพสูงที่สุด โปรแกรมนี้ได้ถูกทดสอบและเปรียบเทียบกับโปรแกรมเก่าทั้งจากข้อมูลเทียมและข้อมูลจริง

Abstract

In this project, we create a new inversion scheme (WSMIX3DMT) based a mixed of the data space conjugate gradient (WSDCG3DMT) and the data space Occam's inversion (WSINV3DMT) methods. WSMIX3DMT is mathematically a slight modification of WSDCG3DMT, its memory requirement is therefore a fraction of WSINV3DMT as in WSDCG3DMT. Instead of fixing the trade-off parameter, it is varied similar to WSINV3DMT. However, the variation is according to the run-time, not based on the data misfit. This strategy makes WSMIX3DMT faster than both WSDCG3DMT and WSINV3DMT, and at the same time requires least memory. This makes WSMIX3DMT as the most efficient inversions. Computational performances and comparisons of all three methods are demonstrated with both synthetic and field datasets.

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบคุณ สกว. ที่ให้โอกาสข้าพเจ้าได้ทำวิจัยในเรื่องที่ข้าพเจ้าถนัดต่อเนื่องมาตลอด ตั้งแต่จบปริญญาเอก ทุนวิจัยสกว. นี้ทำให้ข้าพเจ้าสามารถใช้เวลาทำงานวิจัยได้อย่างเต็มที่และทำงานวิจัยที่มีคุณภาพเป็นที่ยอมรับในกลุ่มคนทำวิจัยเรื่องเดียวกัน ซึ่งวัดได้จากการอ้างอิงใน SCOPUS หรือ ISI database รวมทั้งการที่ข้าพเจ้าได้รับเชิญไปบรรยายในการประชุมต่างประเทศหลายๆ แห่ง ทุนวิจัยนี้ยังสามารถทำให้ข้าพเจ้าได้ใช้เวลาเต็มที่ในการเตรียมบุคลากรและนักศึกษาในการทำวิจัยในสาขาธรณีฟิสิกส์ในอนาคตเพื่อเป็นประโยชน์ต่อประเทศ ซึ่งผลงานที่เกิดกับนักศึกษาก็จะมีปรากฏออกมาเรื่อยๆ

นอกจากนี้ ข้าพเจ้าต้องขอขอบคุณ Prof. Dr. Gary Egbert จาก Oregon State University, Associate Professor Dr. Makoto Uyeshima และ Professor Dr. Hisashi Utada จาก Earthquake Research Institute (ERI), University of Tokyo และ Professor Dr. Yasuo Ogawa จาก Tokyo Institute of Technology ที่สนับสนุนงานวิจัยข้าพเจ้ามาโดยตลอด และทั้งนี้ก็ต้องไม่ลืมที่จะขอบคุณ ผศ. ดร. ศรีสุตา วรามิตร หัวหน้าภาควิชาฟิสิกส์และผู้ร่วมงานคนอื่นๆ ในภาควิชาฟิสิกส์ที่ช่วยสนับสนุนข้าพเจ้ามาโดยตลอด รวมทั้งนักศึกษาในกลุ่มวิจัยธรณีฟิสิกส์ มหาวิทยาลัยมหิดล ที่ร่วมกันฝ่าฟันอุปสรรคต่างๆ เพื่อความเป็นเลิศในงานวิจัยด้านนี้ นอกจากนี้แล้วยังมีเพื่อนๆ จากภาควิชาอื่นๆ อีกด้วย รวมทั้งท่านคณบดีและทีมงานคณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ที่คอยกระตุ้นและสนับสนุนงานวิจัยมาตลอด

อีกกลุ่มหนึ่งที่ข้าพเจ้าจะไม่ขอบคุณไม่ได้ คือ บุคลากรของสกว. ฝ่ายวิชาการที่เป็นมิตรอันดี และคอยช่วยเหลือในเรื่องต่างๆ เป็นอย่างดี

สุดท้ายนี้ ข้าพเจ้าต้องขอขอบคุณบุคคลที่คอยเป็นกำลังใจ เข้าใจในทุกสิ่งทุกอย่าง บุคคลเหล่านี้คือคุณแม่ที่เพิ่งล่วงลับไปและครอบครัวพี่น้องของข้าพเจ้า ขอขอบคุณมากครับ

รศ. ดร. วีระชัย สิริพันธ์วรารณ์

เนื้อหางานวิจัย

บทนำ

Magnetotelluric เป็นเทคนิคหนึ่งทางธรณีฟิสิกส์ เทคนิคนี้เริ่มต้นจากการวัดสนามแม่เหล็กและสนามไฟฟ้าที่บริเวณพื้นผิวของโลก อัตราส่วนของสนามแม่เหล็กและสนามไฟฟ้าสามารถนำมาใช้เป็นตัวบ่งบอกถึงสภาพความต้านทานไฟฟ้า (electrical resistivity) หรือ ความสามารถในการนำไฟฟ้า (electrical conductivity) ภายใต้อพื้นโลกที่มีความลึกต่างๆ ได้ เราสามารถนำข้อมูล electrical resistivity นี้ไปใช้ในการอธิบายโครงสร้างของโลกเพื่ออธิบายการเกิดแผ่นดินไหว (Siripunvaraporn et al., 1998; Unsworth et al., 2000; Boonchaisuk et al., 2010) การศึกษาเทคนิคของพื้นที่ (Jones, 1992) หรือใช้ในการสำรวจหาทรัพยากรธรรมชาติ (Tuncer et al., 2006; Orange, 1989; Vozoff, 1972) และอื่นๆ

ข้อมูลคลื่นแม่เหล็กไฟฟ้าที่วัดได้มาจากการสำรวจแต่ละพื้นที่จะเป็นข้อมูลดิบที่ต้องนำมาผ่าน data processing เพื่อให้ได้ข้อมูลที่เราเรียกว่า apparent resistivity และ phase หรือ impedance tensor ซึ่งเป็นฟังก์ชันของความถี่หรือความถี่ เพื่อนำไปใช้ในการตีความหมายต่อไป การตีความหมายจากข้อมูลโดยตรงนั้นเป็นไปได้ยาก เนื่องจากข้อมูลที่ได้มาไม่ได้เป็นฟังก์ชันของความถี่ ดังนั้น inversion หรือการแก้ปัญหาย้อนกลับ คือกระบวนการที่นำเอาค่า apparent resistivity และ phase ที่เป็นฟังก์ชันของความถี่หรือความถี่ ไปแปลงให้เป็นค่า electrical resistivity กับความลึก โดยผ่านกระบวนการทางคณิตศาสตร์ที่สลับซับซ้อน ทั้งนี้ผลลัพธ์สุดท้ายที่ได้คือแบบจำลอง (model) สภาพความต้านทานไฟฟ้าแบบสามมิติ (3-D)

การพัฒนาโปรแกรม inversion สำหรับข้อมูล MT มีมาต่อเนื่องโดยผู้วิจัย เริ่มต้นจากการพัฒนา 2-D inversion (Siripunvaraporn and Egbert, 2000) โดยโปรแกรม 2-D นี้ที่มีชื่อว่า REBOCC มีนักวิจัยจากทั่วโลกนำไปใช้ในการแปลความหมายข้อมูลจริง (SCOPUS: อ้างอิง 97 ครั้ง as of 24 June 2010) เมื่อได้รับการสนับสนุนจากสกว. ผู้วิจัยก็ได้พัฒนาเป็นโปรแกรม 3-D (Siripunvaraporn et al., 2005) โดยมีชื่อโปรแกรมน่า WSINV3DMT ซึ่งสามารถทำงานได้แม้บนเครื่อง PC ธรรมดา ซึ่งถือว่าเป็นโปรแกรมแรกของโลกที่มีการ release สู่อุตสาหกรรมและได้มีการนำไปใช้จริง (SCOPUS: อ้างอิง 33 ครั้ง as of 24 June 2010) เทคนิคของ WSINV3DMT ได้ถูกนำไปประยุกต์ใช้กับข้อมูลประเภทอื่นๆ ด้วย เช่นข้อมูล Network-MT data (Siripunvaraporn et al., 2004) ข้อมูล 2-D DC Resistivity (Boonchaisuk et al., 2008) และข้อมูล Phase Tensor (Patro et al., 2010) เป็นต้น

อย่างไรก็ตาม แม้ว่า WSINV3DMT จะทำงานได้บนเครื่อง PC ทั่วไป แต่ก็ยังมีปัญหาอยู่ โดยปัญหาหลัก คือ ยังคงต้องใช้หน่วยความจำของเครื่องคอมพิวเตอร์ในปริมาณมากเมื่อใช้กับข้อมูลที่มีขนาดใหญ่ สิ่งนี้คือข้อจำกัดของตัวโปรแกรม วิธีแก้ไขก็คือการเพิ่มหน่วยความจำของคอมพิวเตอร์ให้มากที่สุดเท่าที่เครื่องจะรับได้ ซึ่งก็จะทำให้ต้นทุนการทำงานสูงหรือแพงมากขึ้น

ต่อมาผู้วิจัยก็ได้รับการสนับสนุนจากสกว. เพื่อแก้ไขข้อบกพร่องนี้โดยพัฒนาเป็น algorithm ใหม่ขึ้นมาเพื่อลดปริมาณหน่วยความจำนี้ วิธีหนึ่งที่เราใช้ก็คือ การแก้ระบบสมการด้วยวิธี conjugate gradient (CG) แทนที่จะแก้แบบโดยตรง คือใช้ Cholesky decomposition เหมือนที่ทำได้ใน WSINV3DMT การใช้วิธี CG ทำให้เราไม่ต้องเก็บ sensitivity matrix (J) ซึ่งมีขนาดใหญ่ในหน่วยความจำของคอมพิวเตอร์ซึ่งทำให้เราลดปริมาณการใช้หน่วยความจำได้เป็นอย่างมาก โปรแกรมใหม่นี้เราเรียกว่า data space conjugate gradient method (DCG) หรือ WSDCG3DMT จากการทดลองของ Siripunvaraporn and Egbert (2007) และ Siripunvaraporn and Sarakorn (2010) สำหรับข้อมูล 2-D และ 3-D พบว่า ข้อดีของเทคนิคนี้ก็คือ ไม่ต้องใช้หน่วยความจำในปริมาณมากเหมือน WSINV3DMT แต่ทว่าข้อเสียของเทคนิคใหม่นี้คือเวลาที่ใช้รันโปรแกรมนั้นมากกว่า ดังนั้นมันจึงเป็น trade-off ซึ่งกันและกันระหว่างเวลากับหน่วยความจำ

ในข้อเสนอโครงการนี้ เราเสนอที่จะผสมโปรแกรมทั้งสองอันไว้ด้วยกัน เพื่อคงไว้ในข้อดี นั่นคือใช้หน่วยความจำน้อย ในขณะที่เดียวกันก็ใช้เวลาในการรันน้อยด้วย ในรายงานฉบับนี้เราจะเริ่มต้นจากการบรรยายโปรแกรม WSINV3DMT แล้วตามด้วย WSDCG3DMT จากนั้นก็เสนอแนะเทคนิคใหม่ที่เรียกว่า WSMIX3DMT รวมทั้งผลที่ได้จากการรันโปรแกรม

Inversion : Overview

การทำ inversion คือการหาแบบจำลอง ($\mathbf{m}$) ที่สามารถให้ค่า model responses $\mathbf{F}[\mathbf{m}]$ ที่ fit ข้อมูล $\mathbf{d}$ ที่มีทั้งหมด N ค่าได้สมเหตุสมผล ซึ่งสามารถเขียนเป็นสมการคณิตศาสตร์ได้ดังนี้

$$U(\mathbf{m}, \lambda) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \lambda^{-1} \{ (\mathbf{d} - \mathbf{F}[\mathbf{m}])^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{F}[\mathbf{m}]) - X^2_* \} \quad (1)$$

เมื่อ $\mathbf{C}_d$ คือ data covariance และ T คือ transpose of matrix, $\mathbf{m}$ คือ model ที่มีทั้งหมด M ค่า ส่วน $\mathbf{m}_0$ คือ base model และ $\mathbf{C}_m$ คือ model covariance และ λ^{-1} คือ Lagrange multiplier

สมการที่ (1) นี้มีความหมายว่าเรากำลังทำการ search หาแบบจำลอง (model) ที่มีลักษณะแบบ minimum structure โดยมีข้อแม้ว่าแบบจำลองที่ได้จะต้อง fit ข้อมูลได้เป็นอย่างดีซึ่งถูกกำหนดโดยค่า X^2 . การกำหนดในลักษณะนี้ทำให้ inversion นั้น stable มากขึ้น

การ minimize สมการนี้ คือ การคำนวณหา stationary point ของสมการที่ (1) นี้เมื่อเทียบกับ λ และ $\mathbf{m}$ ซึ่งคำนวณได้ยาก วิธีหนึ่งคือการแก้สมการ penalty functional แทน ซึ่งมีลักษณะดังนี้

$$\Phi_{\lambda}^m = (\mathbf{d} - \mathbf{F}[\mathbf{m}])^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{F}[\mathbf{m}]) + \lambda (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0), \quad (2)$$

เนื่องจากเมื่อ λ นั้นคงที่ หรือ fixed ไว้ เราจะได้ว่า $\partial \Phi / \partial \mathbf{m} = \partial \Phi_{\lambda} / \partial \mathbf{m}$ ดังนั้นเราสามารถแก้สมการ (2) แทนที่สมการที่ (1) ได้แต่ต้อง vary ค่า λ ไปเรื่อยๆ เพื่อให้ได้ค่า misfit ที่น้อยที่สุดหรือตามที่ตั้งเอาไว้

สมการที่ (2) เป็นสมการใน model space ซึ่ง Siripunvaraporn et al. (2005) and Siripunvaraporn and Egbert (2000) แสดงให้เห็นว่าการแก้ปัญหาใน model space นั้นมีข้อเสียคือใช้เวลานานมากๆ และใช้หน่วยความจำสูงมากๆ Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2005) จึงเสนอให้แก้ปัญหาใน data space แทน

ดังนั้นขั้นตอนแรกคือการแปลงสมการที่ (2) จาก model space ให้อยู่ใน data space ซึ่งสามารถทำได้ดังนี้ โดยการเขียน แบบจำลอง $\mathbf{m}$ ให้เป็นฟังก์ชันของ sensitivity matrix ดังนี้ $\mathbf{m} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}^T \boldsymbol{\beta}$ เมื่อ $\boldsymbol{\beta}$ คือ unknown expansion coefficient vector ดังนั้นสมการที่ (2) จะกลายเป็น

$$\Phi_{\lambda}^d = \lambda^{-1} (\mathbf{d} - \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta})^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta}) + (\boldsymbol{\beta}^T \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta}), \quad (3)$$

เมื่อ $\mathbf{J} = [\partial \mathbf{F} / \partial \mathbf{m}]$ คือ $N \times M$ sensitivity matrix ซึ่งเป็นตัวอธิบายการเปลี่ยนแปลงของข้อมูลเนื่องจากการเปลี่ยนแปลงของ model และ $\mathbf{d} = \mathbf{d} - \mathbf{F}[\mathbf{m}] + \mathbf{J}(\mathbf{m} - \mathbf{m}_0)$

เนื่องจาก $\mathbf{F}[\mathbf{m}]$ นั้นเป็น non-linear problem ดังนั้น iterative solutions จึงจำเป็น (Constable et al., 1987) model response $\mathbf{F}[\mathbf{m}]$ จึงจำเป็นต้องถูก linearized ก่อนโดยใช้ first order Taylor's series expansion,

$$\mathbf{F}[\mathbf{m}_{k+1}] = \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_{k+1} - \mathbf{m}_k), \quad (4)$$

เมื่อ k คือ iteration number ในการหา stationary points ของ (3) เราทำได้โดยการ differentiate (3) with respect to β เราได้ว่าในแต่ละ iteration จะมี solution ดังนี้

$$\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2} [\lambda \mathbf{I} + \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}]^{-1} \mathbf{C}_d^{-1/2} \mathbf{d}_k, \quad (5)$$

ข้อดีของการแก้สมการ (1) ใน data space ก็คือ matrix ที่ต้องทำการ invert มีขนาดเพียง $N \times N$ เท่านั้น ไม่ใช่ $M \times M$ เหมือนในกรณีของ model space เมื่อ N คือจำนวนข้อมูลและ M คือขนาดของแบบจำลอง สำหรับข้อมูลเพิ่มเติมศึกษาใน Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2005).

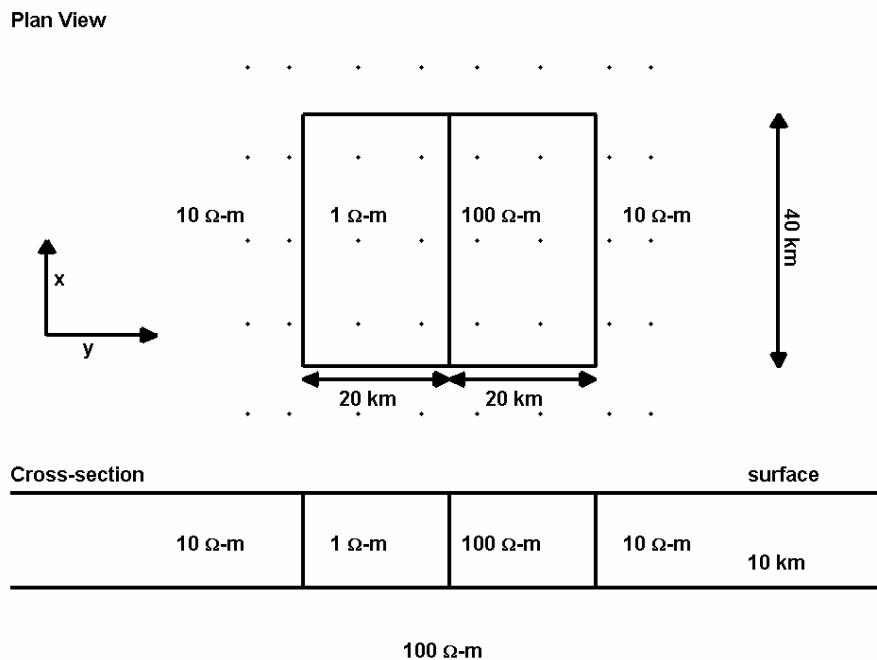
การแก้สมการที่ (5) สามารถทำได้สองวิธี วิธีแรกใช้ใน WSINV3DMT ส่วนวิธีที่สองถูกนำไปใช้ใน WSDCG3DMT

WSINV3DMT : Data Space Occam's Inversion

วิธีแรกคือสร้าง matrix $\mathbf{J}$ และ $\mathbf{R} = [\lambda \mathbf{I} + \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}]$ และเก็บเมตริกซ์เหล่านี้ไว้ในหน่วยความจำ จากนั้นก็ใช้วิธี Cholesky decomposition ในการแก้สมการที่ (5) วิธีนี้เป็นวิธีที่ใช้ใน WSINV3DMT (Siripunvaraporn et al., 2005; Siripunvaraporn and Egbert, 2009) และ DASOCC (Siripunvaraporn and Egbert, 2000) วิธีนี้จะเปลืองหน่วยความจำเนื่องจากต้องเก็บเมตริกซ์ $\mathbf{J}$ และ $\mathbf{R}$ ซึ่งมีขนาด $N \times M$ และ $N \times N$ ซึ่งอาจมีค่าสูงมากก็ได้ถ้าจำนวนข้อมูลมาก

WSDCG3DMT : Data Space Conjugate Gradient Algorithm

อีกวิธีหนึ่งที่ใช้แก้สมการที่ (5) คือการใช้วิธี conjugate gradient วิธีนี้ทำให้ไม่ต้องสร้างเมทริกซ์ J และ R ที่ต้องเก็บไว้ในหน่วยความจำอีกต่อไป เทคนิคนี้จึงประหยัดหน่วยความจำไปได้มาก ด้วยเทคนิคนี้เราไม่ได้สร้างเมทริกซ์ J โดยตรงแต่เราคำนวณผลคูณของเมทริกซ์ J กับเวกเตอร์ใดๆ เช่น Jx หรือ $J^T y$ เทคนิคนี้เป็นเทคนิคที่ใช้ใน WSDCG3DMT ข้อเสียของโปรแกรมนี้คือใช้เวลารันนานกว่า WSINV3DMT ซึ่งจะแสดงให้เห็นในตอนถัดไป



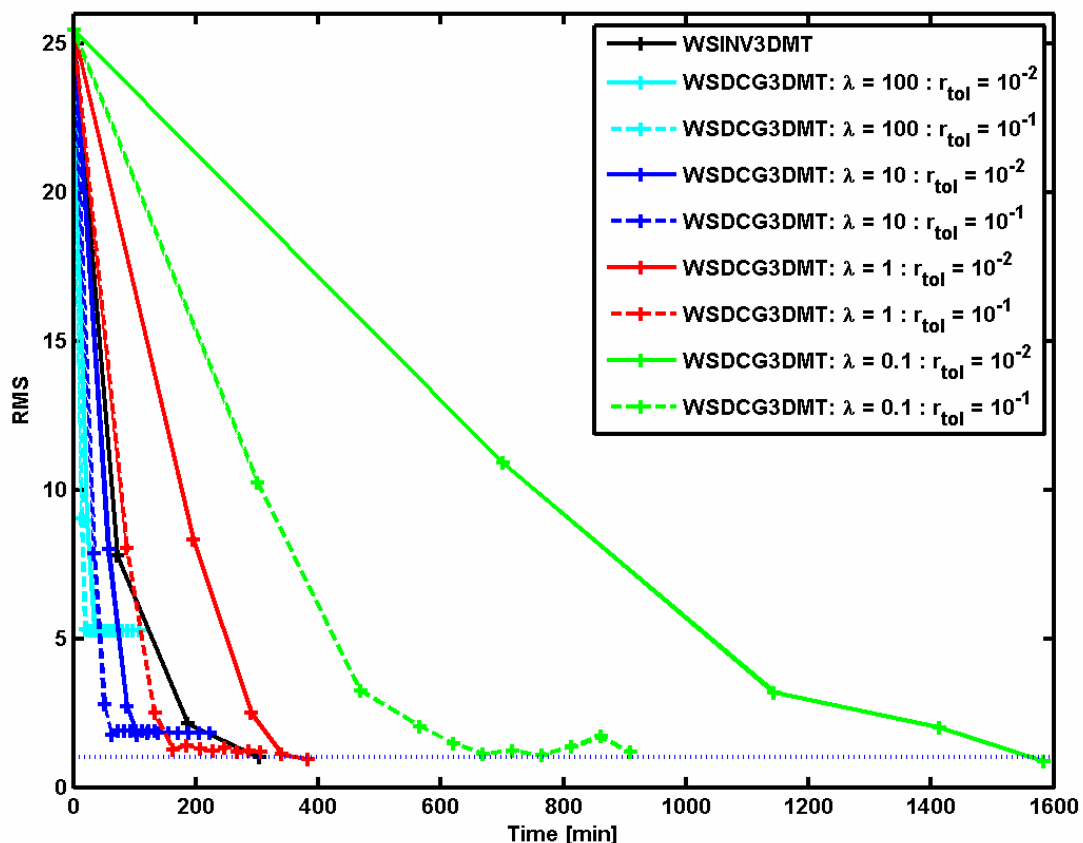
รูปที่ 1 รูปแสดงแบบจำลองเทียมที่ใช้สร้างข้อมูลเทียมเพื่อใช้ในการทดสอบโปรแกรม

การประเมินผลโปรแกรม WSINV3DMT และ WSDCG3DMT และผลการทดสอบ

การทดลองของเราเริ่มต้นจากทดสอบทั้งสองโปรแกรม WSINV3DMT และ WSDCG3DMT กับข้อมูลเทียม (synthetic data) โดยใช้แบบจำลองตามรูปที่ 1 ข้อมูลเทียมประกอบไปด้วย impedance tensor ทั้งสี่ components มีทั้งหมด 40 สถานีวัดและมีความถี่ทั้งหมด 16 ความถี่ ขนาดของแบบจำลองเท่ากับ $28 \times 28 \times 21$ ดังนั้นข้อมูลนี้มี $N = 40 \times 16 \times 8 = 5,120$ และ $M = 28 \times 28 \times 21 = 16,464$ การทดลอง

ต่อไปนี้เป็นบนเครื่องเดียวกัน คือบนเครื่อง Intel Core Two Duo 6400, 2.13 GHz จากจำนวนข้อมูลนี้ เราสามารถประมาณหน่วยความจำของทั้งสองโปรแกรมได้ว่า WSINV3DMT ต้องใช้หน่วยความจำถึง 1 GByte ในขณะที่ WSDCG3DMT จะใช้เพียง 0.4 Gbyte ซึ่งน้อยกว่าเกือบครึ่งหนึ่ง

การทดลองแรก เรารันโปรแกรม WSINV3DMT กับ WSDCG3DMT ที่หลายค่า $\lambda = 100, 10, 1, 0.1, 0.01$ ผลการทดลองแสดงในรูปที่ 2 ซึ่งแสดงให้เห็นสำหรับค่า λ ของ WSDCG3DMT ที่ converge สู่ 1 RMS นั้นจะใช้เวลาในการทำงานช้ากว่า WSINV3DMT ซึ่งใช้เวลาเพียง 300 นาที ในขณะที่ WSDCG3DMT สำหรับค่า $\lambda = 1$ และ 0.1 จะใช้เวลาถึง 400 นาทีและ 1600 นาที



รูปที่ 2 แสดงการลู่เข้าหาคำตอบของ WSINV3DMT (สีดำ) และ WSDCG3DMT ที่หลากหลายค่า λ ผลการทดลองแสดงให้เห็นว่า WSDCG3DMT ใช้เวลานานกว่า WSINV3DMT

การทดลองนี้มีข้อสังเกตที่น่าสนใจมากๆ อันดับแรกก็คือ WSDCG3DMT ที่มีค่า λ สูงจะใช้เวลาไว้มากในแต่ละ iteration ในขณะที่มีค่า λ ต่ำจะใช้เวลานานกว่ามาก เช่นที่ $\lambda = 100$ iteration แรกใช้เวลาไม่ถึง 20 นาที ในขณะที่ $\lambda = 0.1$ iteration แรกใช้เวลามากถึง 700 นาที อย่างที่สอง สำหรับค่า λ

สูงแม้ว่าจะใช้เวลาไวแต่ก็ไม่สามารถหาคำตอบได้เลย เช่นที่ $\lambda = 100$ และ 10 แต่สำหรับค่า λ ต่ำจะสามารถหาคำตอบได้ เช่นที่ $\lambda = 1$ และ 0.1 อย่างที่สามคือ iteration แรกจะใช้เวลาในการรันนานที่สุด iteration ถัดๆ ไปจะใช้เวลาในการรันน้อยลงไปเรื่อยๆ เช่นสำหรับ $\lambda = 0.1$, iteration ที่หนึ่ง สอง สามและสี่ จะใช้เวลารันประมาณ 700 นาที่ 500 นาที่ 400 นาที่ และ 200 นาที่ ตามลำดับ

จากผลการทดลองครั้งนี้ทำให้เราสามารถนำไปขยายผลเพื่อสร้าง Algorithm ใหม่ขึ้นมา

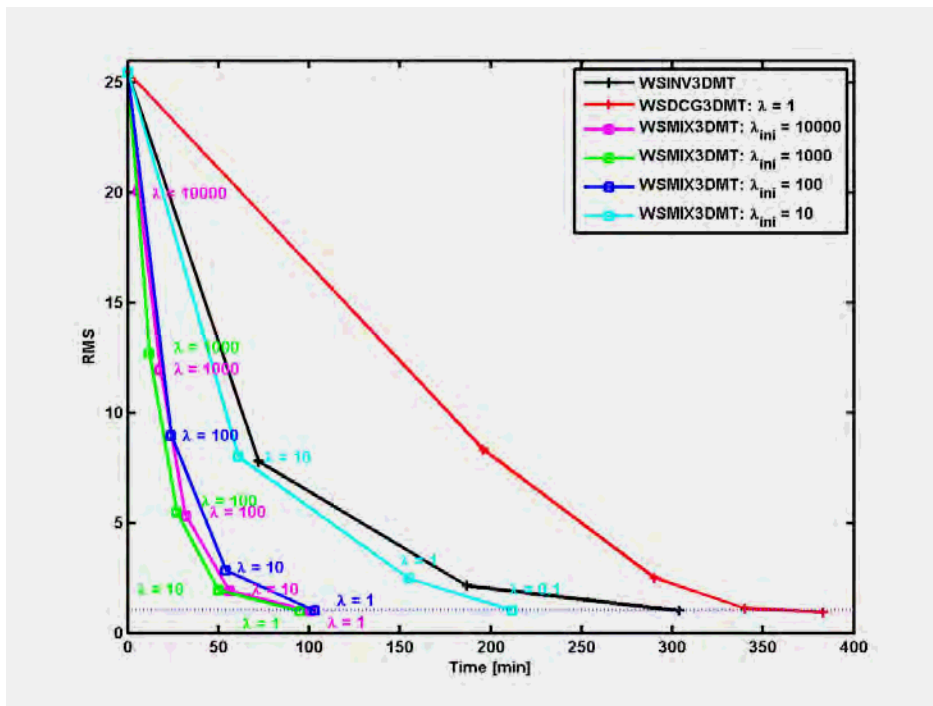
WSMIX3DMT : a Mixed Scheme of DCG and Occam's inversion

จากการทดลองเบื้องต้น เราพบว่าเราสามารถสร้างโปรแกรมขึ้นมาใหม่ โดยใช้หน่วยความจำเหมือน WSDCG3DMT แต่ทว่ามีความเร็วเร็วกว่าทั้ง WSDCG3DMT และ WSINV3DMT

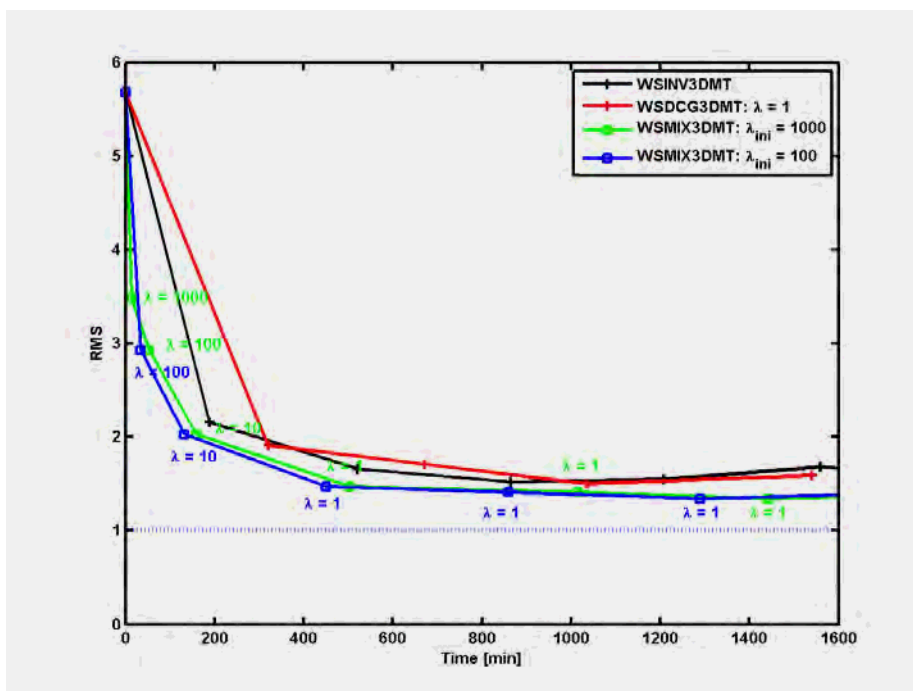
หลักการของโปรแกรมใหม่นี้มาจากการผสมกันของ WSDCG3DMT และ WSINV3DMT ตัวคณิตศาสตร์แทบจะเหมือนกับ WSDCG3DMT แต่หลักการของโปรแกรมใหม่นี้จะง่าย ๆ คือการเปลี่ยนค่า λ ในแต่ละ iteration ที่คล้ายคลึงกับ WSINV3DMT แต่ทว่าในครั้งที่เราจะเริ่ม iteration แรกด้วยค่า λ ที่มีค่ามาก เนื่องจากเมื่อ λ มาก เวลาทำงานก็จะสั้นลง โดยเฉพาะใช้กับ iteration แรกๆ ที่ต้องใช้เวลาในการทำงานสูง จากนั้นเราจะลดค่า λ ใน iteration ถัดไปเรื่อยๆ ตามแต่ที่กำหนด เช่นลดลง 10 เท่า เป็นต้น ตัวอย่างของการรันโปรแกรม เริ่มต้นจาก $\lambda = 1000$ ใน iteration แรก จากนั้นก็จะลดลงเป็น $\lambda = 100$ ใน iteration ที่สอง และเป็น $\lambda = 10$ ใน iteration ที่สาม ไปเรื่อยๆ จนถึงค่าน้อยที่สุดซึ่งในที่นี้เรากำหนดไว้ที่ $\lambda = 0.1$

การประเมินผลโปรแกรม WSMIX3DMT และ WSINV3DMT และ WSDCG3DMT

เราทดสอบโปรแกรมใหม่ WSMIX3DMT กับข้อมูลเทียมเดิม และเทียบผลที่ได้กับสองโปรแกรมเก่า เนื่องจากคณิตศาสตร์ของโปรแกรมใหม่นั้นเหมือนกับ WSDCG3DMT ดังนั้นหน่วยความจำจึงเท่ากัน แต่เมื่อดูที่เวลาแล้วจะเห็นว่าโปรแกรมใหม่ WSMIX3DMT นั้นไวที่สุด คือใช้เวลาน้อยกว่า 100 นาที่ไม่ว่าจะเริ่มต้นด้วยค่า λ ที่เท่าไรก็ตาม ส่วน WSDCG3DMT ใช้เวลา 400 นาที่ ส่วน WSINV3DMT ใช้เวลา 300 นาที่ ดังแสดงในรูปที่ 3



รูปที่ 3 แสดงการลู่เข้าหาคำตอบของ WSINV3DMT (สีดำ) และ WSDCG3DMT ที่ $\lambda = 1$ (สีแดง) และ WSMIX3DMT ที่เริ่มต้นจาก $\lambda = 10000$ (สีชมพู), 1000 (สีเขียวอ่อน), 100 (สีน้ำเงิน) และ 10 (สีฟ้าอ่อน) กับข้อมูลเทียมรูปที่ 1 ผลการทดลองแสดงให้เห็นว่า WSMIX3DMT ใช้เวลาไวกว่าทั้ง WSDCG3DMT และ WSINV3DMT



รูปที่ 4 แสดงการลู่เข้าหาคำตอบของ WSINV3DMT (สีดำ) และ WSDCG3DMT ที่ $\lambda = 1$ (สีแดง) และ WSMIX3DMT ที่เริ่มต้นจาก $\lambda = 1000$ (สีเขียวอ่อน) และ 100 (สีน้ำเงิน) กับข้อมูล EXTECH จริง (Tuncer et al, 2006) ผลการทดลองแสดงให้เห็นว่า WSMIX3DMT ใช้เวลาไวกว่าทั้ง WSDCG3DMT และ WSINV3DMT

นอกจากนี้ทั้งโปรแกรมใหม่และเก่าถูกนำไปทดสอบกับข้อมูลจริง EXTECH data (see Tuncer et al., 2006) พบว่าให้ผลที่เหมือนกัน คือ WSMIX3DMT (สีน้ำเงินและเขียว) ไวกว่าทั้ง WSINV3DMT (สีดำ) และ WSDCG3DMT (สีแดง) และยังใช้หน่วยความจำเท่ากับ WSDCG3DMT ซึ่งน้อยกว่า WSINV3DMT มากๆ ดังแสดงในรูปที่ 4

สรุปผล

เราได้พัฒนาโปรแกรมใหม่ขึ้นมา WSMIX3DMT โปรแกรมมีหลักการจากสองโปรแกรมเก่าคือ WSINV3DMT และ WSDCG3DMT คณิตศาสตร์ของโปรแกรมใหม่จะเหมือนกับ WSDCG3DMT แต่หลักการจะคล้ายกับ WSINV3DMT คือ vary λ ไปในแต่ละ iteration แต่กรณีนี้จะเลือกเริ่มต้นที่ λ มาก ก่อนที่จะค่อยๆ ลดลง ผลการทดลองทั้งกับข้อมูลเทียมและข้อมูลจริงพบว่า โปรแกรม WSMIX3DMT ใช้หน่วยความจำเท่ากับ WSDCG3DMT ซึ่งน้อยกว่า WSINV3DMT มากๆ แต่ในขณะเดียวกันก็ใช้เวลาในการรันน้อยกว่าโปรแกรมเก่าสองถึงสามเท่า

หนังสืออ้างอิง

- Boonchaisuk, S., Vachiratiengchai, C., **Siripunvaraporn, W.**, 2008, Two-dimensional direct current (DC) resistivity inversion: Data space Occam's approach, *Physics of the Earth and Planetary Interiors* 168 (3-4), pp. 204-211
- Boonchaisuk, S., Satitpitakul, A., Vachiratiengchai, C., Nualkhow, P., Amatyakul, P., Unhapipat, S., Rung-Arunwan, T., Sarakorn, W., Siripunvaraporn, W., and Ogawa, Y., 2010, Three-dimensional crustal resistivity structure beneath Kanchanaburi province, Western part of Thailand., SPC 2010, Kanchanaburi, Thailand, 25-27 March.
- Constable, C. S., Parker, R. L., and Constable, C.G., 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: *Geophysics*, **52**, 289-300.
- Jones, A. G., 1992, Electrical conductivity of the continental lower crust, in Fountain, D. M., Arculus, R. J., and Kay, R. W., Eds., *Continental Lower Crust*: Elsevier Science Publ. Co., Inc., 81-143.
- Orange, A. S., 1989, Magnetotelluric Exploration for Hydrocarbons: *Proc. IEEE*, **77**, 287-317.
- Patro P., Uyeshima M., and **W. Siripunvaraporn**, 2010, Three-dimensional inversion of magnetic phase tensor, submitted *Geophysical Journal International*,
- Siripunvaraporn, W., Egbert, G. D., Eisel, M., and Unsworth, M., 1998, A High Resolution EM Survey of the San Andreas Fault (SAF): Local Conductivity Structure in a Regional Context: *Eos, Trans. Am. Geophys. Union*, **79**, 45, Fall Meet. Suppl., F227.
- Siripunvaraporn W. and Egbert, G. D., 2000, An Efficient Data-Subspace Inversion for Two-Dimensional Magnetotelluric Data, *Geophysics*, 65, 791-803.
- Siripunvaraporn, W., M. Uyeshima and G. Egbert, 2004, Three-dimensional inversion for Network-Magnetotelluric data, *Earth Planets Space*, 56, 893-902.
- Siripunvaraporn, W., G. Egbert, Y. Lenbury and M. Uyeshima, 2005, Three-Dimensional Magnetotelluric Inversion: Data Space Method, *Phys. Earth and Planet. Interior.*, 150, 3-14.
- Siripunvaraporn W. and G. Egbert, 2007, Data Space Conjugate Gradient Inversion for 2-D Magnetotelluric Data, *Geophysical Journal International*, **170**, 986-994.
- Siripunvaraporn W. and G. Egbert, 2009, WSINV3DMT: Vertical Magnetic Field Transfer Function Inversion and Parallel Implementation, *Physics of the Earth and Planetary Interiors* 173 (3-4), pp. 317-329
- Siripunvaraporn W., and W. Sarakorn, 2010, An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method, submitted to *Geophysical Journal International*.
- Tuncer V., M. Unsworth, W. Siripunvaraporn, J. Craven, 2006, Audio-magnetotelluric exploration for unconformity type uranium deposits at the McArthur River Mine, northern Saskatchewan (Canada), *Geophysics*, **71**, B201-209.
- Unsworth M., P. Bedrosian, M. Eisel, G. Egbert and W. Siripunvaraporn, 2000, Along strike variations in the electrical structure of the San Andreas Fault at Parkfield, California, *Geop. Res. Lett.*, **27**, 3021-3024.
- Vozoff, K., 1972, The Magnetotelluric Method in the Exploration of Sedimentary Basins: *Geophysics*, **37**, 98-141.

Output ที่ได้จากโครงการ

ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (submitted and revised)

- Siripunvaraporn W., and W. Sarakorn, 2010, An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method, submitted to *Geophysical Journal International*.
- Rung-Arunwan T. and W. Siripunvaraporn, 2010, An efficient modified hierarchical domain decomposition for 2-D Magnetotelluric forward modeling, submitted *Geophysical Journal International*, moderate revision.

ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (published)

- Siripunvaraporn W. and G. Egbert, 2009, WSINV3DMT: Vertical Magnetic Field Transfer Function Inversion and Parallel Implementation, *Physics of the Earth and Planetary Interiors* 173 (3-4), pp. 317-329
- Kalscheuer, T., Pedersen, L.B., Siripunvaraporn, W., 2008, Radiomagnetotelluric two-dimensional forward and inverse modelling accounting for displacement currents, *Geophysical Journal International* 175 (2), pp. 486-514
- Boonchaisuk, S., Vachiratienchai, C., Siripunvaraporn, W., 2008, Two-dimensional direct current (DC) resistivity inversion: Data space Occam's approach, *Physics of the Earth and Planetary Interiors* 168 (3-4), pp. 204-211

Software

- **Siripunvaraporn W.**, 2010, WSMIX3DMT; Three-Dimensional inversion program based on a mixed of the data space conjugate gradient technique and the Occam's method. The code is written with Fortran 77/95.

การเสนอผลงานในที่ประชุมวิชาการนานาชาติบางส่วน

- Weerachai Siripunvaraporn, 2010, A combined DCG and OCCAM algorithm for three-dimensional Magnetotelluric Data, Japan Geosciences Union Meeting, 2010, Chiba, Japan, May 23 – 28.
- Noriko Tada, Kiyoshi Baba, Weerachai Siripunvaraporn, Makoto Uyeshima and Hisashi Utada, 2010, 3-D inversion of marine Magnetotelluric data, Japan Geosciences Union Meeting, 2010, Chiba, Japan, May 23 – 28.
- Noriko Tada, Kiyoshi Baba, Weerachai Siripunvaraporn, Makoto Uyeshima and Hisashi Utada, 2009, Toward 3-D inversion of seafloor MT data, Japan Geosciences Union Meeting 2009, Chiba, Japan, May 16 – 21.
- Thomas Kalscheuer, Laust B. Pedersen and Weerachai Siripunvaraporn, 2008, Radiomagnetotelluric two-dimensional forward and inverse modeling accounting for displacement currents, Induction Workshop #19, Beijing, China, October 23-29.
- Noriko Tada, Kiyoshi Baba, Weerachai Siripunvaraporn, Makoto Uyeshima and Hisashi Utada, 2008, Modification of a 3-D Magnetotelluric forward code considering topography effect in order to estimate 3-D conductivity structures of oceanic upper mantle using inversion method, Induction Workshop #19, Beijing, China, October 23-29.
- Uyeshima M, Ogawa T, Yamaguchi S, Murakami H, Toh H, Yoshimura R, Oshiman N, Tanbo T, Koyama S, Mochizuki H, Marutani Y, Usui Y, and Siripunvaraporn W, 2008, 3-D resistivity structure in the vicinity of the Atotsugawa fault revealed by a Network-MT survey, Induction Workshop #19, Beijing, China, October 23-29.
- Weerachai Siripunvaraporn, Gary Egbert and Hishi Utada, 2008, Current version of WSINV3DMT, Induction Workshop #19, Beijing, China, October 23-29.
- Weerachai Sarakorn, Weerachai Siripunvaraporn and Gary Egbert, 2008, Three-dimensional Magnetotelluric modeling using finite element method: Tetrahedral and

hexahedral elements, comparison to finite difference method, Induction Workshop #19, Beijing, China, October 23-29.

- Gary D Egbert and **Weerachai Siripunvaraporn**, 2008, Hybrid Occam/Conjugate Gradient Methods for Practical Electromagnetic Inversion, Induction Workshop #19, Beijing, China, October 23-29.
- Puwis Amatyakul and **Weerachai Siripunvaraporn**, 2008, A joint inversion of direct-current resistivity and Magnetotelluric data, Induction Workshop #19, Beijing, China, October 23-29.
- Tawat Rung-Arunwan and **Weerachai Siripunvaraporn**, 2008, Domain decomposition for 2D Magnetotelluric forward modeling, Induction Workshop #19, Beijing, China, October 23-29.
- Noriko Tada, **Weerachai Siripunvaraporn**, Makoto Uyeshima, Kiyoshi Baba, Hisashi Utada, 2008, Development of marine 3D MT inversion scheme, Japan Geoscience Union Meeting 2008, May 25-30, Chiba, Japan.
- Uyeshima, M., R. Yoshimura, T. Ogawa, S. Yamaguchi, H. Murakami, H. Toh, N. Oshiman, S. Koyama, T. Tanbo, Y. Usui, H. Mochizuki, W. Siripunvaraporn, and Research Group for Crustal Resistivity Structure in the NKTZ Concentrated Deformation Zone, 2008, 3-D resistivity structure beneath the Atotsugawa Fault zone revealed by the Network-MT observations, The Japan Geoscience Union Meeting 2008, May 25-30, Chiba, Japan

ภาคผนวก

ภาคผนวก ก. Manuscript

- Siripunvaraporn W., and W. Sarakorn, 2010, An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method, submitted to *Geophysical Journal International*.

ภาคผนวก ข. Manuscript

- Rung-Arunwan T. and W. Siripunvaraporn, 2010, An efficient modified hierarchical domain decomposition for 2-D Magnetotelluric forward modeling, submitted *Geophysical Journal International*, moderate revision.

ภาคผนวก ค. Reprint

- Siripunvaraporn W. and G. Egbert, 2009, WSINV3DMT: Vertical Magnetic Field Transfer Function Inversion and Parallel Implementation, *Physics of the Earth and Planetary Interiors* 173 (3-4), pp. 317-329

ภาคผนวก ง. Reprint

- Kalscheuer, T., Pedersen, L.B., Siripunvaraporn, W., 2008, Radiomagnetotelluric two-dimensional forward and inverse modelling accounting for displacement currents, *Geophysical Journal International* 175 (2), pp. 486-514

ภาคผนวก จ. Reprint

- Boonchaisuk, S., Vachiratienchai, C., Siripunvaraporn, W., 2008, Two-dimensional direct current (DC) resistivity inversion: Data space Occam's approach, *Physics of the Earth and Planetary Interiors* 168 (3-4), pp. 204-211

ภาคผนวก ก. Manuscript

- Siripunvaraporn W., and W. Sarakorn, 2010, An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method, submitted to *Geophysical Journal International*.

An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method

Weerachai Siripunvaraporn^{1,2} and Weerachai Sarakorn³

¹Department of Physics, Faculty of Science, Mahidol University, Rama 6 Rd., Rachatawee, Bangkok 10400, THAILAND.

²ThEP, Commission of Higher Education, 328 Si Ayuthaya Road, Bangkok 10400, THAILAND

³Department of Mathematics, Faculty of Science, Mahidol University, Rama 6 Rd., Rachatawee, Bangkok, 10400, THAILAND.

Abstract

In this paper, we start with the implementation and extension of the data space conjugate gradient (DCG) method previously developed for the two-dimension (2-D) to the three-dimension (3-D) Magnetotelluric (MT) data, and will be referred to as WSDCG3DMT. Synthetic experiments show that WSDCG3DMT usually spends computational time longer than the data space Occam's inversion (WSINV3DMT). However, memory requirement of WSDCG3DMT is only a fraction of WSINV3DMT. Knowledge and information gained from the synthetic studies of WSDCG3DMT has led to a creation of a mixed scheme (WSMIX3DMT) of the data space conjugate gradient and the data space Occam's methods. WSMIX3DMT is a slight modification of WSDCG3DMT but enhancing so that its computational time is several factors lower than both WSINV3DMT and WSDCG3DMT. Because WSMIX3DMT is a modification of WSDCG3DMT, its memory requirement is therefore a fraction of WSINV3DMT as in WSDCG3DMT. This makes WSMIX3DMT as the most efficient inversions. Computational performances and comparisons of all three methods are demonstrated with both synthetic and EXTECH field datasets.

1. Introduction

Recently, number of three-dimensional (3-D) magnetotelluric (MT) surveys is substantially increased worldwide (e.g., Tuncer et al., 2006; Patro and Egbert, 2008, among many others). This might be due to the fact that MT has increasingly accepted by many geophysicists and seismologists. Another factor is the improvements of the data acquisition units, the measurement sensors and their accessories. Examples of MT uses are for geothermal explorations (e.g., Heise et al., 2008; Árnason et al., 2010), volcanoes and tectonic studies (Uyeshima, 2007; Patro and Egbert, 2008; Hill et al., 2009; Ingham et al., 2009) and ore explorations (Tuncer et al., 2006; Queralt et al., 2007; Farquharson and Craven, 2008; Türkoğlu et al., 2009; Goldax and Kosteniuk, 2010). All of these have led to a higher demand for 3-D MT inversion codes for interpretation.

Currently, a number of 3-D MT inversion algorithms have been developed (e.g. Mackie & Madden 1993; Newman & Alumbaugh 2000; Zhdanov et al. 2000; Sasaki 2001; Mackie, personal communication 2002; Siripunvaraporn et al. 2004, 2005; Sasaki and Meju, 2006; Han et al., 2008; Lin et al., 2008,2009; Farquharson and Craven, 2008; Adveed and Adveed, 2009; Siripunvaraporn et al., 2009). All algorithms are designed to find “best” model that fits the data but also “geologically” interpretable. One of the 3-D algorithms (and the only one currently available to the MT communities) is the WSINV3DMT program by Siripunvaraporn et al. (2005; 2009). The algorithm’s idea was based on the Occam’s style inversion introduced for 1-D MT data by Constable et al. (1987). Occam’s inversion is known for its robust calculation and its efficiency. However, its disadvantage is the large memory requirements, and the extensive computational time, particularly when applying to 2-D and 3-D modeling (Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005).

To reduce both storage and calculation time, Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2004; 2005) transformed the original Occam’s inversion which is a model space method into the data space Occam’s algorithm. The transformation makes it practical for 3-D MT inversion on most computers. However, WSINV3DMT still requires substantial memory to store the $N \times M$ sensitivity matrix, where N and M are the data and model parameters, respectively. Siripunvaraporn and Egbert (2007) used 2-D MT data to show that the large storage can be avoid by using a data space conjugate gradient (DCG) approach.

From the 2-D studies, Siripunvaraporn and Egbert (2007) concluded that the DCG method can significantly reduce the memory usage. However, its computational time can be longer than that of the data space Occam's algorithm. Computational time of the DCG method is controlled by the stopping criteria used inside the conjugate gradient (CG) algorithm when solving the normal equation ($\mathbf{R}\mathbf{x} = \mathbf{b}$). The CG solver is terminated when the relative error ($r = \|\mathbf{R}\mathbf{x} - \mathbf{b}\|/\|\mathbf{b}\|$) reaches a given tolerance r_{tol} . Smaller r_{tol} (e.g., $r_{tol} < 10^{-2}$) requires many number of CG iterations, while larger r_{tol} (e.g., $r_{tol} = 10^{-1}$) requires significantly less but can cause the inversion to fail to converge to the target misfit. Large number of CG iterations translates into longer CPU time. Our 2-D studies also showed that $r_{tol} = 10^{-2}$ is the optimal tolerance value. The model generated with $r_{tol} = 10^{-2}$ differs less than a percent from that generated with $r_{tol} = 10^{-8}$ but requires significantly less CPU time.

In addition, convergence rate of the DCG inversion also depends on the regularization parameter λ , which acts as a trade-off between the data norm and the model norm. Larger λ ($\lambda > 10$) demands small number of CG iterations per inversion iteration. However, the inversion could not bring the misfit down to the desired misfit because large λ produces very smooth model. Smaller λ ($0.1 \leq \lambda \leq 10$) can reach the desired level of misfit but normally requires large number of CG iterations per inversion iteration. However, if λ is too small ($\lambda < 0.1$), DCG can break down. If it converges, it requires significantly large number of CG iterations and also produces "very rough and spurious" structures which is not geologically interpretable.

Here, we directly implement and extend the data space conjugate gradient (DCG) algorithm for the 3-D MT data. Hereafter, we will refer to the 3-D DCG method as WSDCG3DMT. Numerical experiments are performed on a synthetic data in a similar way as conducted in the 2-D experiments (Siripunvaraporn and Egbert, 2007). The objective is to verify whether the conclusions learned from the 2-D cases remain the same or different for the 3-D data. Knowledge gained from the synthetic studies has led us to a creation of a mixed scheme of the Occam's inversion and the DCG method. We will refer to a mixed scheme as WSMIX3DMT.

We start the paper with a brief review of the data space conjugate gradient method (WSDCG3DMT) and its necessary mathematics. More details on the data space Occam's inversion and the data space conjugate gradient method can be found in many previous publications (Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005; Siripunvaraporn

and Egbert, 2007; and Boonchaisuk et al., 2008). Later, a mixed scheme (WSMIX3DMT) between the DCG method and the Occam method is introduced. Numerical experiments on both synthetic data and EXTECH data are performed with these three algorithms (WSINV3DMT, WSDCG3DMT and WSMIX3DMT). Comparisons in terms of computational time and memory are analyzed and discussed. A conclusion is given at the end.

2. Review of Data Space Conjugate Gradient Inversion

Consider a general objective functional Φ^m ,

$$\Phi^m = \Phi_d + \lambda\Phi_m = (\mathbf{d} - \mathbf{F}[\mathbf{m}])^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{F}[\mathbf{m}]) + \lambda (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0), \quad (1)$$

where Φ_d a data norm, Φ_m a model norm, $\mathbf{m}$ the resistivity model of dimension M , $\mathbf{m}_0$ the prior model, $\mathbf{C}_m$ the model covariance matrix, $\mathbf{d}$ the observed data with dimension N , $\mathbf{F}[\mathbf{m}]$ the forward model response, $\mathbf{C}_d$ the data covariance matrix, and λ a regularization parameter.

To minimize (1) in a data space method, we start with the transformation of the model space objective functional (1) to a data space objective functional (2) by expressing a model as a linear combination of rows of the smoothed sensitivity matrix (Parker, 1994), or $\mathbf{m} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}^T \boldsymbol{\beta}$. Then, (1) becomes

$$\Phi^d = (\mathbf{d} - \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta})^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta}) + \lambda (\boldsymbol{\beta}^T \mathbf{J} \mathbf{C}_m^T \mathbf{J}^T \boldsymbol{\beta}), \quad (2)$$

where $\mathbf{J} = \partial \mathbf{F} / \partial \mathbf{m}$ is an $N \times M$ sensitivity matrix, and $\mathbf{d} = \mathbf{d} - \mathbf{F}[\mathbf{m}] + \mathbf{J}(\mathbf{m} - \mathbf{m}_0)$. To minimize (2), $\mathbf{F}[\mathbf{m}_{k+1}]$ is linearized with the first order Taylor series expansion, as $\mathbf{F}[\mathbf{m}_{k+1}] = \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k (\mathbf{m}_{k+1} - \mathbf{m}_k)$, when k is an inversion iteration number. Differentiating (2) with respect to $\boldsymbol{\beta}$ and rearranging, an iterative sequence of approximate solutions can be obtained as,

$$\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2} [\lambda \mathbf{I} + \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}]^{-1} \mathbf{C}_d^{-1/2} \mathbf{d}_k, \quad (3)$$

where $\mathbf{I}$ is an identity matrix.

There are two methods to solve (3). First method is to explicitly form $\mathbf{J}$ and $\mathbf{R} = [\lambda \mathbf{I} + \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}]$ and store them in the computer memory. $\mathbf{R}$ will be factorized into lower and upper matrices (LU-factorization), and then solved with backward and forward substitutions.

This method is used in WSINV3DMT program for 3-D MT data (Siripunvaraporn et al., 2005; Siripunvaraporn and Egbert, 2009) and DASOCC for 2-D MT data (Siripunvaraporn and Egbert, 2000). This scheme requires substantial amount of RAM to store $N \times M$ $\mathbf{J}$ and also $N \times N$ $\mathbf{R}$ matrices. This could prohibit a run on very large data sets, particularly for 3-D cases.

Instead of forming and decomposing $\mathbf{R}$ as in WSINV3DMT, an alternative method is to solve (3) with an iterative solver. Because $\mathbf{R}$ is theoretically symmetric, (3) is commonly solved with a conjugate gradient (CG) method as in many MT inversion algorithms (see Mackie and Madden, 1993; Siripunvaraporn and Egbert, 2007; Lin et al., 2008). One clear advantage of using CG to solve (3) is that the large $N \times M$ sensitivity matrix $\mathbf{J}$ is not explicitly formed and stored in the computer memory. Only a product of $\mathbf{J}$ or $\mathbf{J}^T$ with an arbitrary vector is required by solving one forward problem per period (see Mackie and Madden, 1993; Newman and Alumbaugh, 2000; Rodi and Mackie, 2001; Siripunvaraporn and Egbert, 2007; Lin et al., 2008). Two routines to compute $\mathbf{J}\mathbf{p}$ and $\mathbf{J}^T\mathbf{q}$ are therefore implemented here for the 3-D problem, where $\mathbf{p}$ and $\mathbf{q}$ are general $M \times 1$ and $N \times 1$ vectors, respectively. This method is used in WSDCG3DMT.

The data space conjugate gradient algorithm and the routines to explicitly form $\mathbf{J}$ and to compute $\mathbf{J}\mathbf{p}$ and $\mathbf{J}^T\mathbf{q}$ are briefly described in the following sub-sections.

2.1 Data Space Conjugate Gradient Algorithm (WSDCG3DMT)

The data space conjugate gradient algorithm denoted as WSDCG3DMT has two iterative loops. The outer loop which is a main inversion loop is to minimize (2), while the inner loop is to minimize $\mathbf{R}\mathbf{x} = \mathbf{b}$ in (3) with a conjugate gradient (CG) method where $\mathbf{R} = [\lambda \mathbf{I} + \mathbf{C}_d^{-1/2} \mathbf{J} \mathbf{C}_m \mathbf{J}^T \mathbf{C}_d^{-1/2}]$, $\mathbf{b} = \mathbf{C}_d^{-1/2} \mathbf{d}$ and $\mathbf{x} = \mathbf{C}_d^{1/2} \mathbf{\beta}$ (see Barrett et al., 1994 for Preconditioned Conjugate Gradient algorithm). The algorithm was summarized in Figure 2 of Siripunvaraporn and Egbert (2007), and is repeatedly presented below with more explanations.

Reading inputs and initializing variables.

Start DCG “outer” loop to minimize (2) : iteration k

1. Compute $\mathbf{d}_k = \mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_k - \mathbf{m}_0)$
2. Start DCG “inner” loop by using CG to solve $\mathbf{R}_k \mathbf{x} = \mathbf{b}$

2.1 Initialization: $\mathbf{x}_{(0)} = 0$; $\mathbf{r}_{(0)} = \mathbf{b}$, where $\mathbf{r} = \|\mathbf{R}\mathbf{x} - \mathbf{b}\|/\|\mathbf{b}\|$.

for $icg = 1, 2, \dots, ncgmax$ or $\|\mathbf{r}^T \mathbf{r}\| < r_{tol}$, where icg a CG iteration number, $ncgmax$ a maximum number of CG iterations, and r_{tol} a stopping tolerance level.

$$2.2 \quad \mathbf{z}_{(icg-1)} = \mathbf{r}_{(icg-1)}$$

$$2.3 \quad \delta_{(icg-1)} = \mathbf{r}_{(icg-1)}^T \mathbf{z}_{(icg-1)}$$

$$2.4 \quad \text{if } (icg = 1) \quad \mathbf{p}_{(1)} = \mathbf{z}_{(0)}$$

else

$$\beta_{(icg-1)} = \delta_{(icg-1)} / \delta_{(icg-2)}$$

$$\mathbf{p}_{(icg)} = \mathbf{z}_{(icg-1)} + \beta_{(icg-1)} \mathbf{p}_{(icg-1)}$$

endif

$$2.5 \quad \mathbf{q}_{(icg-1)} = \mathbf{R}_k \mathbf{p}_{(icg)}$$

$$2.6 \quad \alpha_{(icg-1)} = \delta_{(icg-1)} / \mathbf{p}_{(icg)}^T \mathbf{q}_{(icg)}$$

$$2.7 \quad \mathbf{x}_{(icg)} = \mathbf{x}_{(icg-1)} + \alpha_{(icg)} \mathbf{p}_{(icg)}$$

$$2.8 \quad \mathbf{r}_{(icg)} = \mathbf{r}_{(icg-1)} - \alpha_{(icg-1)} \mathbf{q}_{(icg)}$$

2.9 if $(\|\mathbf{r}^T \mathbf{r}\| < r_{tol})$ or $(icg > ncgmax)$, then stop CG iteration and go to 3, else go to 2.2.

end icg

3. Compute $\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k \mathbf{C}_d^{-1/2} \mathbf{x}$

4. Compute $\mathbf{F}[\mathbf{m}_{k+1}]$ and RMS misfit $\|\mathbf{C}_d^{-1/2}(\mathbf{d} - \mathbf{F}[\mathbf{m}_{k+1}])\|$
5. Check condition;
 - 5.1 exit if misfit below the desired level, go to 6;
 - 5.2 continue if misfit is greater than the desired level, go to 1;
6. End DCG outer loop.

Step 1 requires calling one forward routine for $\mathbf{F}[\mathbf{m}_k]$, and another call to compute $\mathbf{J}_k(\mathbf{m}_k - \mathbf{m}_0)$. On step 2.1, system (3) is already normalized, therefore there is no preconditioner here. Step 2.5 is a “key” for the CG solver. It requires two forward modeling calls to compute $\mathbf{s} = \mathbf{J}_k^T \mathbf{C}_d^{-1/2} \mathbf{p}_{(icg)}$ and $\mathbf{J}_k \mathbf{C}_m \mathbf{s}$. Step 3 demands one forward modeling call to compute $\mathbf{J}_k \mathbf{C}_d^{-1/2} \mathbf{x}$. Step 4 requires another forward modeling call to compute the model responses $\mathbf{F}[\mathbf{m}_{k+1}]$. Overall, numbers of forward modeling calls to compute the model response is two per outer loop iteration per period, and to compute a multiplication of $\mathbf{J}$ or $\mathbf{J}^T$ with a vector is $2 + 2N_{cg}$ per outer loop iteration per period, where N_{cg} is a number of CG iterations. A total number of forward modeling calls would therefore be $4 + 2N_{cg}$ per period per outer loop iteration.

2.2 Forward Modeling and Sensitivity Calculation

Given an electrical conductivity (σ) or resistivity (ρ) model, to yield MT responses at the surface, the electric fields ($\mathbf{E}$) are computed from the second order Maxwell's equation,

$$\nabla \times \nabla \times \mathbf{E} = i\omega\mu\sigma\mathbf{E}, \quad (4)$$

where ω is an angular frequency and μ the magnetic permeability. Discretizing the model and applying the staggered grid finite difference approach to (4), we obtain a system of equations for a given period or frequency,

$$\mathbf{S}\mathbf{e} = \mathbf{b}, \quad (5)$$

where $\mathbf{e}$ represents the unknown internal electric fields, $\mathbf{b}$ a vector containing the terms associated with the boundary electric fields, and $\mathbf{S}$ a large sparse symmetric and complex coefficient matrix. System of equations (5) is solved with a quasi-minimum residual (QMR) method per period and per polarization as in Siripunvaraporn et al. (2002). Surface responses can then be obtained from a linear combination of a vector $\mathbf{a}$ associated at a measurement site and the computed electric fields,

$$\mathbf{F}[\mathbf{m}] = \mathbf{a}^T \mathbf{e} = \mathbf{a}^T \mathbf{S}^{-1} \mathbf{b}. \quad (6)$$

To compute for the sensitivity $\mathbf{J} = \partial \mathbf{F} / \partial \mathbf{m}$ at a given period, equation (6) is differentiated with respect to the model $\mathbf{m}$,

$$\mathbf{J} = \partial \mathbf{F} / \partial \mathbf{m} = \partial (\mathbf{a}^T \mathbf{e}) / \partial \mathbf{m} = \mathbf{a}^T \mathbf{S}^{-1} \mathbf{\Theta} + \mathbf{\Psi}, \quad (7)$$

where $\mathbf{\Theta} = \partial \mathbf{b} / \partial \mathbf{m} - (\partial \mathbf{S} / \partial \mathbf{m}) \mathbf{e}$ and $\mathbf{\Psi} = (\partial \mathbf{a}^T / \partial \mathbf{m}) \mathbf{e}$. The process to form $\mathbf{J}$ is straightforward by first constructing $\mathbf{\Theta}$, solving $\mathbf{S}^{-1} \mathbf{\Theta}$, multiplying the result with $\mathbf{a}^T$ and finally adding with $\mathbf{\Psi}$. With this technique, calculating $\mathbf{S}^{-1} \mathbf{\Theta}$ would require solving the system of equations (5) M times per period and per polarization (Rodi, 1976). This calculation can be very significant, particularly in 3-D cases.

To reduce number of forward callings, reciprocity property of the electromagnetic fields (see Rodi, 1976; Mackie and Madden, 1993; Siripunvaraporn and Egbert, 2000) is applied to (7). With the reciprocity, the process of computing $\mathbf{J}$ is modified by first solving $(\mathbf{a}^T \mathbf{S}^{-1})^T$, then multiplying the result with $\mathbf{\Theta}^T$ before finally adding with $\mathbf{\Psi}^T$. Using the reciprocity technique, computing $(\mathbf{a}^T \mathbf{S}^{-1})^T$ would require solving the system of equations (5) only N_s times per period and per polarization (Rodi, 1976; Siripunvaraporn and Egbert, 2000), where N_s is the number of observed stations which is typically a lot smaller than M , particularly in 3-D cases. The reciprocity theorem helps significantly decreasing the computational time of the program (Siripunvaraporn and Egbert, 2000).

2.3 Multiplication of $\mathbf{J}$ or $\mathbf{J}^T$ to any vectors

To compute the product of $\mathbf{J}$ with a given vector $\mathbf{p}$, equation (7) becomes

$$\mathbf{Jp} = \mathbf{a}^T \mathbf{S}^{-1} \mathbf{\Theta p} + \mathbf{\Psi p}. \quad (8)$$

The process is started with a multiplication of $\mathbf{\Theta p}$, then solving $\mathbf{S}^{-1} \mathbf{\Theta p}$, multiplying the result with $\mathbf{a}^T$, and finally adding them with the product of $\mathbf{\Psi p}$. Similarly, to compute the product of $\mathbf{J}^T$ with a given vector $\mathbf{q}$, equation (7) also becomes

$$\mathbf{J}^T \mathbf{q} = \mathbf{\Theta}^T [\mathbf{S}^T]^{-1} \mathbf{a q} + \mathbf{\Psi}^T \mathbf{q}. \quad (9)$$

The process here is also straightforward. It starts with a multiplication of $\mathbf{a q}$, because $\mathbf{S} = \mathbf{S}^T$, then solving $\mathbf{S}^{-1} \mathbf{a q}$ and multiplying them with $\mathbf{\Theta}^T$, finally adding the result with $\mathbf{\Psi}^T \mathbf{q}$. Equation (8) and (9) show that each process requires solving the system of equations (5) only one times per period and per polarization. Storage for $\mathbf{J}$ matrix is not necessary for (8) and (9) but required for (7).

2.4 Theoretical Comparisons for Forming $\mathbf{J}$ and Its Multiplications

Both forming $\mathbf{J}$ and its multiplications ($\mathbf{Jp}$ or $\mathbf{J}^T \mathbf{q}$) require solving the same system of equations (5), but with different right hand sides. As in section 2.2 and 2.3, forming $\mathbf{J}$ requires solving (5) with $\mathbf{a}$ as the right hand side, while computing $\mathbf{Jp}$ and $\mathbf{J}^T \mathbf{q}$ have $\mathbf{\Theta p}$ and $\mathbf{a q}$, as their right hand sides, respectively. All vectors ($\mathbf{a}$, $\mathbf{\Theta p}$ and $\mathbf{a q}$) are sparse, but $\mathbf{\Theta p}$ and $\mathbf{a q}$ involve more non-zero terms than $\mathbf{a}$. Consequently, solving (5) with $\mathbf{\Theta p}$ and $\mathbf{a q}$ as the right hand sides will require larger number of QMR iterations than with just $\mathbf{a}$ as the right hand side to converge to the same accuracy level. Similar behavior was also occurred in 2-D cases. Because system of equations for 2-D cases is small, the difference is therefore not significant. However, for 3-D case, the difference in CPU time is noticeable and will be shown in the numerical experiments.

2.5 Parallel Implementation

Similar to WSINV3DMT (Siripunvaraporn and Egbert, 2009), we also implement our 3-D DCG code on a parallel system. Although memory is not an issue for the DCG method, its extensive runtime is still a big concern due to its numerous calls to the forward modeling routine. As in WSINV3DMT, we parallelize WSDCG3DMT over frequencies via MPI (Message Passing Interface) libraries. For DCG, the parallelization is relatively simple, just distributing the forward

modeling call of each period to each processor node when computing the forward response $\mathbf{F}[\mathbf{m}]$, and calculating $\mathbf{J}\mathbf{p}$ and $\mathbf{J}^T\mathbf{q}$. The simplicity occurs because there is no need to form and store the cross-product $\mathbf{R}$ as in WSINV3DMT (Siripunvaraporn and Egbert, 2009).

3. Numerical Experiments on a Synthetic Data : WSDCG3DMT & WSINV3DMT

Here, before we introduce a mixed scheme of the data space conjugate gradient method and the Occam's inversion; we start with the repetitions of the same experiments we conducted with the 2-D MT data but now with the 3-D MT data. The goal of the experiments is to check whether the same conclusions derived from the 2-D studies can be gained. In addition, we also compare the results with WSINV3DMT in terms of computational time and memory.

Similar to Siripunvaraporn et al. (2005) and Siripunvaraporn and Egbert (2009), we use the same synthetic model to generate a synthetic dataset for testing our codes. The synthetic model consists of two anomalies, 1 Ω -m and 100 Ω -m buried next to each other inside a 10 Ω -m layer lying on top of a 100 Ω -m half-space as illustrated in Figure 1 (Figure 4 in Siripunvaraporn et al., 2005; Figure 3b in Siripunvaraporn and Egbert, 2009). The model mesh for the inversion was discretized at $28 \times 28 \times 21$ (+7 air layers) in x , y and z , respectively. The full complex impedance data (Z_{xx} , Z_{xy} , Z_{yx} and Z_{yy} ; i.e. $N_m = 4$) is generated for 40 MT sites ($N_s = 40$) located regularly covering the two anomalies (solid dots in Figure 1) and 16 periods from 0.031 to 1000 second ($N_p = 16$). Five percent Gaussian noise calculated from the data magnitude ($|Z_{xy}Z_{yx}|^{1/2}$) was added to the impedance data. With this configuration, model parameter M would be equal to $28 \times 28 \times 21 = 16,464$, while data parameter N would be equal to $40 \times 16 \times 8 = 5,120$. In this experiment, all runs can be performed on a serial machine; an Intel Core Two Duo 6400, 2.13 GHz machine with 2 GBytes of RAM. Bigger model mesh or dataset would prohibit a run on this serial machine for WSINV3DMT.

Our first test is to perform the WSDCG3DMT program with various λ ($\lambda = 100, 10, 1, 0.1, 0.01$) and two r_{tol} (10^{-1} and 10^{-2}) for the DCG inner loop or the CG loop. Convergence behaviors of WSDCG3DMT for various λ and different r_{tol} as a function of time are shown in Figure 2 in comparison to WSINV3DMT. An inverted model after four iterations from WSDCG3DMT ($\lambda = 1$ and $r_{tol} = 10^{-2}$) is shown in Figure 3. The inversion can recover both

anomalies and the underlying layer similar to the inverted result from WSINV3DMT (Figure 6 of Siripunvaraporn et al., 2005).

For larger λ (10 and 100) with $r_{tol} = 10^{-2}$, DCG cannot converge to the desired level of 1 RMS. It can only lower the misfit down in the first two iterations before idling. Similar to the 2-D tests, larger λ requires smaller number of CG iterations to solve the normal equation (3) per outer loop iteration. This is reflected in a small amount of computing time as shown in Figure 2 (cyan and blue colors). For smaller λ (1 and 0.1) with $r_{tol} = 10^{-2}$, DCG is able to converge to the desired 1 RMS in four iterations. However, in contrast to larger λ , it demands significantly large number of CG iterations to solve (3) per one outer loop iteration. This is shown by a large amount of computational time in Figure 2 (red and green), particularly for the first iteration.

Reducing number of CG iterations per main iteration would help decreasing a computer runtime. One way is to set r_{tol} to a larger value. Here, at 10^{-1} . In all λ cases with $r_{tol} = 10^{-1}$, DCG has difficulty to converge to the target misfit of 1 RMS as seen in dash-lines of Figure 2. Larger r_{tol} would only help reducing computing time but not the convergence. In contrast, setting r_{tol} to smaller values (e.g., at 10^{-3} or less), number of inversion iterations to converge to the desired misfit is the same as in the case of $r_{tol} = 10^{-2}$. Inverted model is also less than a percent difference. Major difference is at the number of CG iterations per main inversion iteration which is significantly larger for smaller r_{tol} . These experiments show that $r_{tol} = 10^{-2}$ is appeared to be an optimal tolerance level for terminating the CG iterations in the DCG inner loop.

For $\lambda = 0.01$ or smaller, DCG fails to converge from the start. The sign of the divergence can be observed or detected inside the CG solver after some number of CG iterations. This becomes a very important and useful information. We can use it as a criterion to decide the termination of the WSDCG3DMT code. Whenever a divergence inside the CG loop takes place, program is stopped. The cause for the divergence behavior inside the CG loop is probably due to the loss of the orthogonality of matrix $\mathbf{R}$.

From all of these experiments, we can infer that both 2-D studies from Siripunvaraporn and Egbert (2007) and 3-D studies here yield almost the same conclusions. Optimal convergence occurs in the λ ranges between 0.1 and less than 10, and also with $r_{tol} = 10^{-2}$.

Computational performance in term of memory and CPU time of WSDCG3DMT is then compared with those from WSINV3DMT. Majority of the memory requirements for

WSINV3DMT is to store $\mathbf{J}$ and $\mathbf{R}$ matrices which can be approximated from $8NM+8NN$ with double precisions. This is about 1 GBytes in our test case. The code also requires less than 0.3 GBytes for storing $\mathbf{S}$, $\partial\mathbf{S}/\partial\mathbf{m}$, and other parts for miscellaneous computations. For WSDCG3DMT, we do not store $\mathbf{J}$ and $\mathbf{R}$ in the memory. One GBytes of RAM is therefore not needed as in the case of WSINV3DMT. WSDCG3DMT requires only about 0.4 GBytes to store many different matrices and vectors. This is about the same as the memory used for the miscellaneous computations in WSINV3DMT.

In term of computational time, WSINV3DMT converges to the desired misfit within three iterations in about 300 minutes as shown in a black line of Figure 2, while WSDCG3DMT with $\lambda = 1$ and $\lambda = 0.1$ uses about 400 and 1600 minutes, respectively. This again shows that computational time of WSINV3DMT is less than that of converged WSDCG3DMT. Thus, in term of computational performance, one can clearly see that WSDCG3DMT has advantage in terms of memory. However, its computational time can be significantly greater than that of WSINV3DMT. A trade-off between computational time and memory used would be a factor for users to decide. This is also similar to the 2-D studies (Siripunvaraporn and Egbert, 2007).

In 2-D studies, we did not compare CPU time, but number of forward modeling calls of each algorithm. Here, similar analysis are performed for the 3-D cases. WSINV3DMT requires a fix number of callings at $N_p N_s N_m + N_p(N_\lambda + 1)$ per inversion iteration to form the sensitivity and compute the misfit, where N_λ is a number of λ varied to search for the minimum misfit in each iteration of the Occam's inversion. In our experiments, for the first iteration, $N_\lambda = 5$, number of forward modeling calls for WSINV3DMT is therefore at 2,656. For WSDCG3DMT, in each iteration, number of forward modeling calls depends on a number of CG iterations (N_{cg}) in the DCG inner loop, and equal to $4N_p + 2N_p N_{cg}$ per inversion iteration as we previously discussed. In our experiments, for the case $\lambda = 1$ and $r_{tol} = 10^{-1}$, $N_{cg} = 47$ for the first iteration, number of forward modeling calls is then at 1,568.

Although number of forward modeling calls of WSDCG3DMT is about 1,000 less than WSINV3DMT, computational time is actually slightly longer for the first iteration of both methods as shown in Figure 2. This indicates that for each forward modeling call, WSDCG3DMT requires averagely longer runtime than that of WSINV3DMT. Because of more complicated right hand sides in the system of equation (5) when computing $\mathbf{Jp}$ or $\mathbf{J}^T \mathbf{q}$ than

forming $\mathbf{J}$, as already stated in Section 2.4, it requires larger number of QMR iterations to converge to the solution. This study shows that to test the efficiency of the inversion, just counting number of forward modeling calls can be misleading (see Newman and Alumbaugh, 1997; Siripunvaraporn and Egbert, 2007).

Another interesting point for WSDCG3DMT is the reduction of the number of CG iterations per outer loop iteration when misfit becomes lower. For example, in the case $\lambda = 1$ and $r_{tol} = 10^{-2}$, $N_{cg} = 108, 48, 25$ and 21 , respectively, from the first to forth iteration of the main inversion loop. This is reflected and shown with lesser CPU time for successive iterations in Figure 2. The reduction of number of CG iterations occurs on every case in our examples. When inverted solution gets closer to the “true” solution, normal equation (3) is probably lesser stiff and therefore become easily to solve.

4. The mixed scheme of the DCG and Occam’s inversions (WSMIX3DMT)

Because DCG does not explicitly form and store the sensitivity matrix, DCG therefore requires significantly less memory than the Occam’s inversion. However, the major drawback of the DCG method is its computational time which could be longer than the Occam’s inversion. Here, we propose a new scheme which is a mixed concept of both DCG and Occam and a modification of the DCG method. Mathematics of the new scheme is in fact identical to the DCG method. Thus, it maintains the memory advantage of the DCG method over the Occam’s style. However, we intentionally design so that the new scheme spends computational time less than both DCG and Occam. This would make the mixed scheme as the efficient inversion.

Assuming that the goal of the inversion is the same for both DCG and Occam that is to bring the misfit down to the desired level. One distinct feature between both methods is at the λ value. In Occam’s inversion (Constable et al., 1987; Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005), in every iteration, λ in equation (3) is varied in order to search for the model producing the “least” RMS misfit (see Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005). With the Occam concept, λ is posed as both the step length and the regularization parameters. For the DCG method, λ is pre-selected and fixed in every iteration as shown in previous section in WSDCG3DMT. In DCG, λ therefore acts like a regularization or damping parameter.

In our mixed scheme, the algorithm is based mainly on the DCG method. However, λ is not fixed but varied as both step length and regularization parameter similar to the idea of the Occam's inversion. The difference from the Occam's method is we do not choose λ that minimize the RMS misfit, but we select λ that can both lower the misfit down and at the same time require small number of CG iterations per an outer loop iteration. The “optimal” λ is selected and varied based on our knowledge and experience gained from the studies in previous section 3. It is therefore not exactly the same philosophy as in the Occam's inversion, nor the DCG, but a mixed of both. This is why we refer to this method as a mixed DCG and Occam or in short WSMIX3DMT.

Based on earlier 3-D studies in section 3 and 2-D studies in Siripunvaraporn and Egbert (2007), r_{tol} for the inner CG loop is fixed at 10^{-2} as the optimal tolerance level for number of CG iterations. For early iterations, larger λ requires significantly smaller number of CG iterations than smaller λ and at the same time can lower the misfit down. We therefore choose to start our mixed scheme with large λ_{ini} (e.g., $\lambda_{ini} = 100$ or larger). To further decrease the misfit down, λ is automatically reduced by a factor of ε (e.g., $\varepsilon = 10$) in the next iteration. This automatic reduction is to avoid redundant computations as occurred when large λ is fixed (Figure 2). A reduction in λ was used before in Kelbert et al. (2008) but only when the misfit is not decreased in their non-linear conjugate gradient (NLCG) method. The automatic reduction in λ is continued successively for the next iterations until reaching λ_{min} (e.g., $\lambda_{min} = 0.1$). When λ below λ_{min} , it will set back to λ_{min} .

For example, $\lambda_{ini} = 100$, $\lambda_{min} = 0.1$ and $\varepsilon = 10$ is input in the first iteration. Values of λ for the 2nd, 3rd and 4th iterations would be 10, 1 and 0.1, respectively. If the inversion continues, 5th iteration and so on will have $\lambda = 0.1$. In addition, we also add a scheme to detect the divergence. Within N_{div} CG iterations (e.g., $N_{div} = 15$), if the divergence occurs, there is a high possibility that the inversion will fail to converge. If that happens, λ is automatically increased by a factor of ε and re-start the process again. This “extra” step may cause redundant computations but can help preventing the divergence inside the main inversion loop.

4.1 Numerical Experiments of WSMIX3DMT and Comparisons with WSINV3DMT and WSDCG3DMT

To check the efficiency of the WSMIX3DMT code, we apply it to the same synthetic data set generated from model in Figure 1. Four values of λ_{ini} are used ($\lambda_{\text{ini}} = 10000, 1000, 100$ and 10) with $\varepsilon = 10$. Figure 4 shows convergence rates from the WSMIX3DMT program with various initial λ_{ini} , in comparisons to those of WSINV3DMT (black) and WSDCG3DMT with $\lambda = 1$ (red). Figure 4 shows that all runs can converge to the desired level within 3-4 iterations. Most importantly, all WSMIX3DMT runs spend computational time less than both WSINV3DMT and fixed λ WSDCG3DMT. Inverted models from all runs with 1 RMS are similar to the inverted model plotted in Figure 3.

When λ_{ini} is too large (i.e. at 10000), redundant computation is occurred in the first iteration. Although the first iteration with $\lambda_{\text{ini}} = 10000$ runs very quick, it does not greatly reduce the misfit. When λ is decreased to 1000 in the next iteration. The misfit in this case is almost the same as starting the run with $\lambda_{\text{ini}} = 1000$. The first iteration of $\lambda_{\text{ini}} = 10000$ is therefore redundant and unnecessary. Starting the mixed inversion with $\lambda_{\text{ini}} \leq 10$ requires large computational time due to large number of CG iterations used in the first iteration. In addition, λ is decreased quickly to 1 and 0.1 in the next few iterations and would demand large number of CG iterations. In this case, we do not gain advantage of small number of CG iterations used from larger λ . It therefore become less effective as in WSDCG3DMT. Thus, we should avoid to start WSMIX3DMT with smaller λ or very large λ .

From the experiments, the “optimal” λ to start with would be around 100 to 1000 (Figure 4). Both cases spends computational time at about 100 minutes compared to 300 minutes of WSINV3DMT and 400 minutes of WSDCG3DMT. In addition, WSMIX3DMT requires memory the same as WSDCG3DMT, i.e. less than 0.4 Gbytes for this dataset, which is several factors less than WSINV3DMT. WSMIX3DMT which is a combination of DCG and Occam is the most efficient method compared to both WSINV3DMT and WSDCG3DMT.

Further studies show that ε around 10 is the optimal value. If ε too small, redundant computations can be occurred. If too large, WSMIX3DMT would not gain much advantage from smaller number of CG iterations when large λ used. This makes WSMIX3DMT less efficient.

5. Applications of WSMIX3DMT, WSDCG3DMT and WSINV3DMT to EXTECH data

To show the efficiency of our mixed scheme WSMIX3DMT in comparisons to the WSDCG3DMT and WSINV3DMT codes, we applied all three codes to the EXTECH dataset (Tuncer et al., 2006) conducting around the McArthur River mine, Saskatchewan, Canada (Figure 2 of Tuncer et al., 2006). The data consists of both impedance tensor (Z_{xx} , Z_{xy} , Z_{yx} and Z_{yy}) and the vertical magnetic field transfer function (VTF; T_{zx} and T_{zy}) for 131 stations and 16 periods (from 8000 Hz to 5 Hz). The data parameter N is therefore equal to 25,152. In all runs, minimum error bars for VTF is set at 15% of $(|T_{zx}|^2 + |T_{zy}|^2)^{1/2}$ and 5% of $|Z_{xy}Z_{yx}|^{1/2}$ for off-diagonal and 50% for diagonal terms. A 1000 Ω m half-space is used as an initial model and a prior model ($\mathbf{m}_0$) and is discretized at $56 \times 56 \times 33$ (+7 air layers). The model parameter M is therefore at 103,488.

To show the efficiency of the parallel codes, all runs are performed on a cluster computer which consists of 8 processor nodes with 8 GBytes in memory each. With 16 period data, two periods are distributed to compute on each processor node. In terms of memory, WSINV3DMT requires about 5 GBytes to store its two period sensitivities and the cross-product matrices. It also requires about 1 GBytes additional to store other necessary components. In contrast to WSINV3DMT, both WSDCG3DMT and WSMIX3DMT require less than 1 GBytes of RAM to perform the inversion of this EXTECH dataset. The EXTECH dataset and the model mesh used above are already at a maximum limitation of the cluster for WSINV3DMT. Because WSDCG3DMT and WSMIX3DMT use significantly less memory, they can therefore be applied on a bigger dataset and a bigger mesh on this cluster. However, here, same parameters are used for comparisons.

Convergence behaviors of the three methods are plotted in Figure 5 as a function of time in minutes. From Figure 5, WSINV3DMT requires about 870 minutes in 3 iterations to converge to its minimum at 1.52 RMS. After the 3rd iteration, the misfit is fluctuated above the minimum RMS. WSDCG3DMT with $\lambda = 1$ also requires 3 iterations to converge to 1.50 RMS but uses longer CPU time at about 1040 minutes. After the 3rd iteration, WSDCG3DMT increases its RMS to 1.57 in the 4th iteration and is terminated because of the divergence. With $\lambda < 0.5$, the WSDCG3DMT code diverges and fails after its first iteration.

For our mixed scheme, WSMIX3DMT with $\lambda_{ini} = 100$ can converge to 1.47 RMS slightly below the level of both WSINV3DMT and WSDCG3DMT in 3 iterations. Most importantly, the computational time is only about 450 minutes, about half of WSINV3DMT and WSDCG3DMT. At the 4th iteration when λ is reduced to 0.1, the scheme detected the divergence occurring inside the CG loop. The code is then re-started with a bigger $\lambda = 1$ on the 4th iteration. The process of increasing λ will cost some extra computational time. With the divergence detection scheme, the code can continue to run for several iterations.

After continuing the run, WSMIX3DMT can further reduce the misfit below the level that both WSINV3DMT and WSDCG3DMT can attain. At 5th iteration with $\lambda = 1$, the misfit is at the lowest RMS of 1.34. However, these 0.13 RMS difference from 3rd to 5th iteration require computational time almost 14 hours; about twice longer than the CPU time at the 3rd iteration. One can therefore stop at the 3rd iteration because the inverted models at the 3rd and 5th iteration are slightly different.

Convergence behavior from starting WSMIX3DMT with $\lambda_{ini} = 1000$ is redundant in early iterations similar to starting with $\lambda_{ini} = 100$, as shown in Figure 5. It therefore spends “extra” CPU time longer. Overall, it can still converge below 1.5 RMS within 500 minutes faster than both WSINV3DMT and WSDCGMT methods.

Inverted model from the 5th iteration of WSMIX3DMT starting with $\lambda_{ini} = 100$ is shown in Figure 6. It is similar to the inverted model from WSINV3DMT (Figure 11 of Siripunvaraporn and Egbert, 2009). Major differences are at the two conductors. Here, conductor on the eastern part of the profiles oriented in the NE-SW direction can be seen as shallow as 500 m depth. Northern conductor seems to be continuous from 800 m to 1.3 km depth. The difference of the two inverted models (Figure 6 here and Figure 11 of Siripunvaraporn and Egbert, 2009) and detail interpretation is beyond our scopes in this paper. For detail discussion of the EXTECH data set can be found in Tuncer et al. (2006) and Farquharson and Craven (2008).

6. Conclusions

In this paper, we implement and extend the data space conjugate gradient inversion for three-dimensional Magnetotelluric data (WSDCG3DMT). Numerical experiments on 3-D synthetic data show that WSDCG3DMT with some λ can converge to the desired level of misfit but often

spends longer computational time than the data space Occam's inversion (WSINV3DMT). However, because the whole sensitivity matrix is not explicitly formed and stored, its memory requirements are therefore minimal at a fraction of WSINV3DMT. This makes WSDCG3DMT practical for large to very large data set.

Based on the numerical experiments of WSDCG3DMT on synthetic data, number of CG iterations depends greatly on the λ values used. Larger λ usually requires smaller number of CG iterations per main inversion iteration but hardly converge to the "true" solution. Smaller λ requires larger number of CG iterations per main iteration but can converge to the desired level of misfit. However, if λ is too small, it can diverge. Computational time varies proportionally to the number of CG iterations. Thus, to use less CPU time, number of CG iterations per outer loop iteration must be minimized.

The information learned from the synthetic studies has inspired and led us to the creation of the mixed scheme of the Occam's and DCG methods or WSMIX3DMT. In DCG scheme, λ is fixed as a regularization parameter. In Occam's inversion, λ is varied as both step length and regularization parameters. In our mixed scheme, λ is varied but not in the same way as in the Occam's inversion. Instead of choosing λ that generates a model with smallest misfit as in Occam, we prefer λ that minimizes number of CG iterations but at the same time can reduce the misfit. With this strategy, λ should initially start from large value before reducing to smaller value for the next subsequent iterations. Our studies shows that λ between 100 to 1000 are the optimal λ to start with for the WSMIX3DMT code.

By applying all three algorithms (WSMIX3DMT, WSDCG3DMT and WSINV3DMT) on both synthetic and EXTECH field data, our mixed scheme (WSMIX3DMT) is significantly faster than both WSDCG3DMT and WSINV3DMT. Similar to WSDCG3DMT, it requires insignificant amount of memory. Because both computational time and memory performances are at minimum, we can conclude here that WSMIX3DMT is the most efficient inversion.

7. Acknowledgements

This research has been supported by the Thai Center of Excellence in Physics (ThEP) and by Thailand Research Fund (TRF: RMU5080025). The authors would like to thank XXX and XXX and the editor for their comments to help improve the manuscript. The authors would like to thank Jim Carven for allowing us to use EXTECH data sets

8. References

- Avdeev, D., and Avdeeva, A., 2009, 3D Magnetotelluric inversion using a limited-memory quasi-Newton optimization, *Geophysics*, 74, F45-F57.
- Árnason, K., Eysteinnsson, H., Hersir, G.P., 2010, Joint 1D inversion of TEM and MT data and 3D inversion of MT data in the Hengill area, SW Iceland, *Geothermics*, 39, 13-34
- Barret, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and Van der Vorst, H., 1994, Templates for the solution of linear systems: Building blocks for iterative methods: Soc. Ind. Appl. Math.
- Boonchaisuk, S., C. Vachiratiengchai and W. Siripunvaraporn, 2008, Two-dimensional direct current (DC) resistivity inversion: data space Occam's approach, *Phys. Earth. Planetary Interiors*, 168, 204-211.
- Constable, C. S., Parker, R. L., and Constable, C.G., 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, 52: 289-300.
- Goldak, D. and Kosteniuk, P., 2010, 3D Inversion of Transient Magnetotelluric Data: an Example from Pasfield Lake, Saskatchewan, EGM 2010 International Workshop, 11-14 April, 2010. Capri, Italy.
- Farquharson C.G., and Craven, J.A., 2008, Three-dimensional inversion of Magnetotelluric data for mineral exploration: An example from the McArthur River uranium deposit, Saskatchewan, Canada., *Jour. Applied. Geophysics.*, 68, 450-458.
- Han, N., Nam, M.J., Kim, H.J., Lee, T.J., Song, Y., Suh, J.H., 2008, Efficient three-dimensional inversion of Magnetotelluric data using approximate sensitivities, *Geophys. J. Inter*, 175, 477-485.

- Heise, W., Caldwell, T.G., Bibby, H.M., Bannister, S.C., 2008, Three-dimensional modelling of magnetotelluric data from the Rotokawa geothermal field, Taupo Volcanic Zone, New Zealand, *Geophys. J. Inter*, 173, 740-750.
- Hill, G.J., Caldwell, T.G., Heise, W., Chertkoff, D.G., Bibby, H.M., Burgess, M.K., Cull, J.P., Cas, R.A.F., 2009, distribution of melt beneath Mount St Helens and Mount Adams inferred from Magnetotelluric data, *Nature Geoscience*, 2, 785-789.
- Ingham, M.R., Bibby, H.M., Heise, W., Jones, K.A., Cairns, P., Dravitzki, S., Bennie, S.L., Caldwell, T.G., Ogawa, Y., 2009, A Magnetotelluric study of Mount Ruapehu volcano, New Zealand, 2009, *Geophys. J. Inter.*, 179, 887-904.
- Kelbert, A., Egbert, G.D., and Schultz, A., 2008, Non-linear conjugate gradient inversion for global EM induction: resolution studies, *Geophys. J. Int*, 173, 365-381.
- Lin, C., Tan, H., and Tong, T., 2009, Parallel rapid relaxation inversion of 3D Magnetotelluric data, *Applied Geophysics*, 6, 77-83.
- Lin, C., Tan, H., and Tong, T., 2008, Three-dimensional conjugate gradient inversion of Magnetotelluric sounding data, *Applied Geophys.*, 5, 314-321.
- Mackie, R. L., and Madden, T. R., 1993, Three-dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int*, 115: 215-229.
- Newman, G. A., and D. L. Alumbaugh, 1997, Three-dimensional massively parallel electromagnetic inversion – I. Theory, *Geophys. J. Int*, 128, 345-354.
- Newman, G. A., and D. L. Alumbaugh, 2000, Three-dimensional magnetotelluric inversion using non-linear conjugate gradients, *Geophys. J. Int*, 140, 410-424.
- Parker, R. L., 1994, *Geophysical Inverse Theory*, Princeton University Press.
- Patro, P.K. and Egbert, G.D., 2008, Regional conductivity structure of Cascadia: Preliminary results from 3D inversion of USArray transportable array Magnetotelluric data, *Geophys. Res. Lett.*, 35, art. no. L20311.
- Rodi, W. L., 1976, A technique for improving the accuracy of finite element solutions for Magnetotelluric data, *Geophys. J. Roy. Astr. Soc.*, **44**, 483-506.
- Sasaki, Y., 2001, Full 3D inversion of electromagnetic data on PC., *J. Appl. Geophys.*, 46, 45-54.

- Sasaki, Y. and Meju, M.A., 2006, Three-dimensional joint inversion for Magnetotelluric resistivity and static shift distributions in complex media, *Jour. Geophys. Res. B: Solid Earth*, 111, art. no. B05101.
- Siripunvaraporn, W. and Egbert, G., 2000. An efficient data-subspace inversion method for 2D magnetotelluric data, *Geophysics*, **65**(3), 791-803.
- Siripunvaraporn, W., Egbert, G., & Lenbury, Y., 2002. Numerical accuracy of magnetotelluric modeling: A comparison of finite difference approximations, *Earth Planets Space*, **54**(6), 721-725.
- Siripunvaraporn, W., M. Uyeshima and G. Egbert, 2004, Three-dimensional inversion for Network-Magnetotelluric data, *Earth Planets Space*, 56, 893-902.
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., & Uyeshima, M., 2005, Three-dimensional Magnetotelluric inversion: data-space method, *Phys. Earth and Planetary Interiors*, **150**, 3-14.
- Siripunvaraporn, W., and Egbert, G., 2007, Data space conjugate gradient inversion for 2-D Magnetotelluric data, *Geophys. Jour. Inter.*, 170, 986-994.
- Siripunvaraporn, W. and Egbert, G., 2009, WSINV3DMT: Vertical magnetic field transfer function inversion and parallel implementation, *Phys. Earth. Planet. Interiors.*, 173, 317-329.
- Streich, R., 2009. 3D finite-difference frequency-domain modeling of controlled-source electromagnetic data: Direct solution and optimization for high accuracy, *Geophysics*, **74**(5), F95-F105.
- Tuncer, V., M. J. Unsworth, W. Siripunvaraporn, and J.A. Craven, 2006, Exploration for unconformity-type uranium deposits with audiomagnetotelluric data: A case study from the McArthur River mine, Saskatchewan, Canada, *Geophysics*, 71, B201-B209.
- Türkoğlu, E., Unsworth, M., Pana, D., 2009, Deep electrical structure of northern Alberta (Canada): Implications for diamond exploration, *Canadian Jour. Of Earth Sciences*, 46, 139-154.
- Uyeshima, M., 2007, EM monitoring of crustal processes including the use of the Network-MT observations, *Surveys in Geophys.*, 28, 199-237.
- Zhdanov, M.S., S. Fang, G. Hursan, 2000, Electromagnetic inversion using quasi-linear approximation, *Geophysics*, 65, 1501-1513.

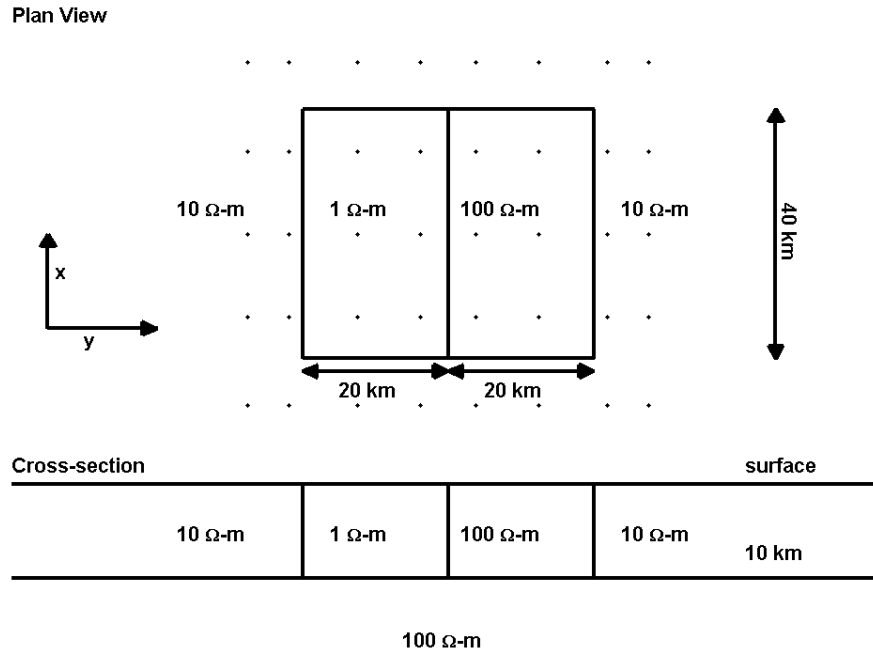


Figure 1. Two-block synthetic model used to test our inversions. The solid dots indicate the observational sites. A cross-section view in the lower panel is a profile cutting across the middle of the two anomalies in the upper panel, and is not to scale (after Siripunvaraporn et al., 2005; and Siripunvaraporn and Egbert, 2009).

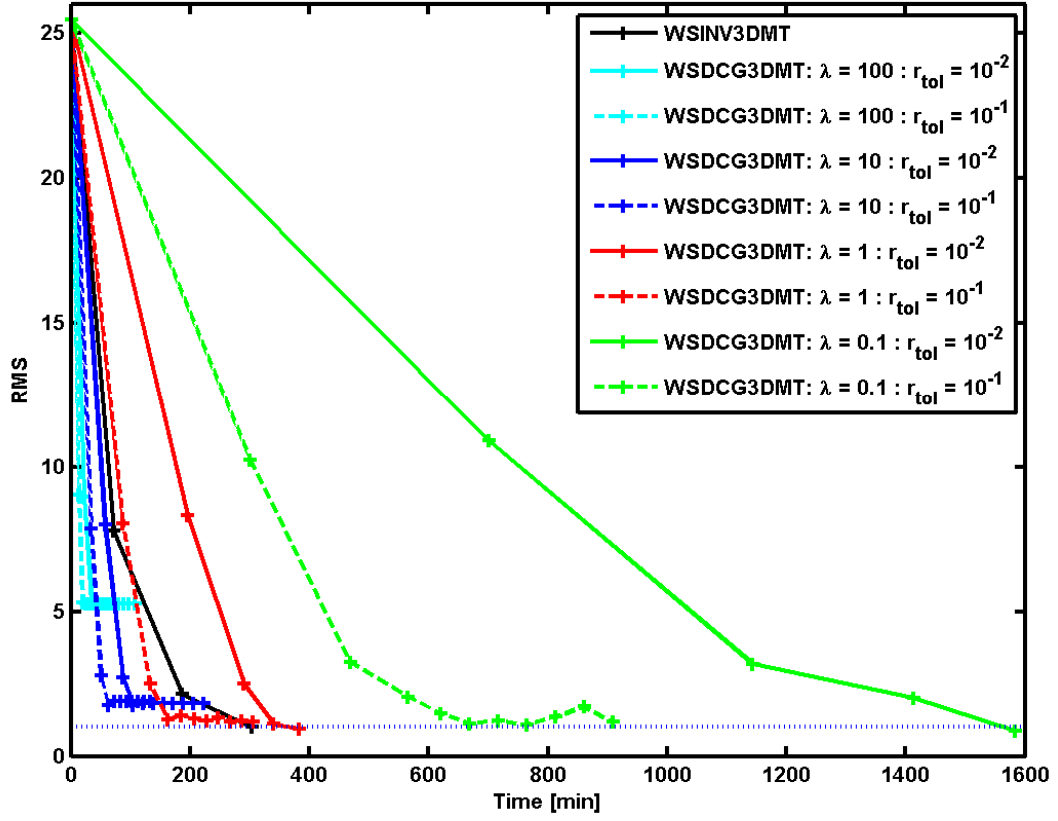


Figure 2. Convergence rates of WSINV3DMT (black) and WSDCG3DMT from various λ s and r_{tol} to the synthetic dataset generated from a model in Figure 1. Dash line for $r_{tol} = 10^{-1}$. Solid line for $r_{tol} = 10^{-2}$. Each plus symbol indicates one iteration.

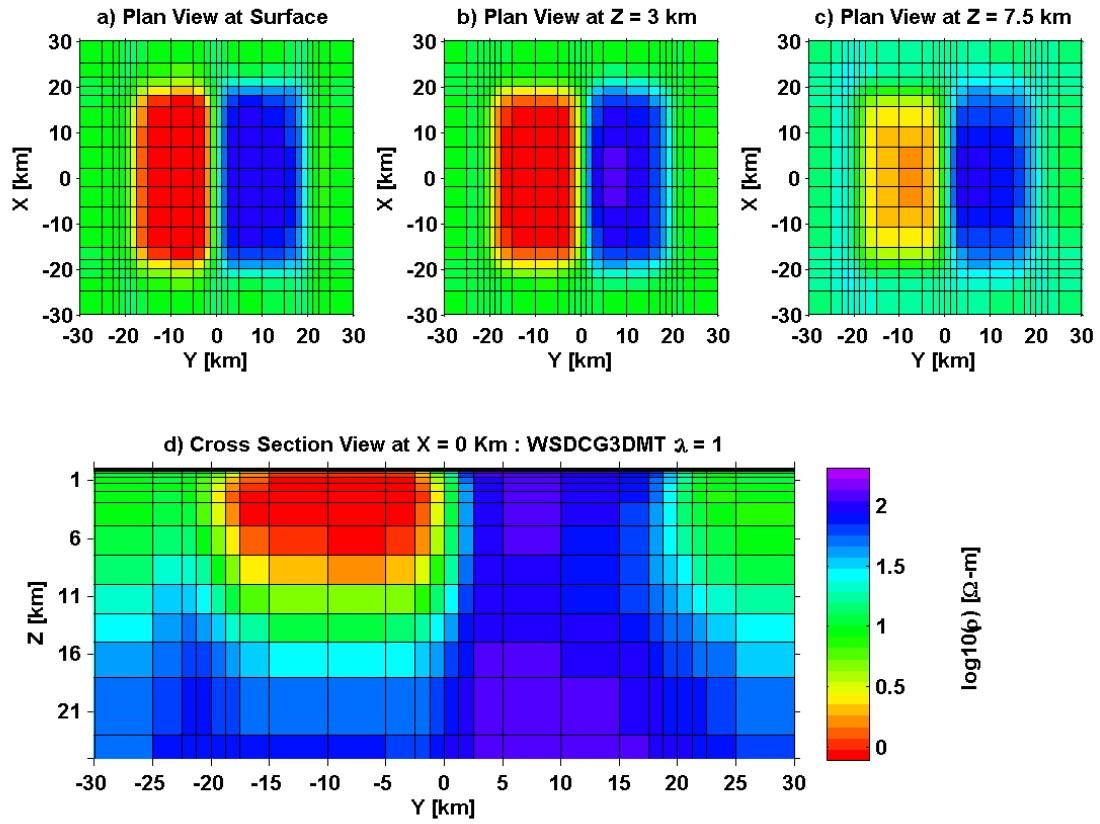


Figure 3. An inverted model from WSDCG3DMT with $\lambda = 1$. The synthetic data is generated from the model in Figure 1. The top panels (a)–(c) is a plan view at the surface, at 3 km and at 7.5 km depth, and the bottom panel (d) is a cross-section view cutting across the two anomalies at $X = 0$ km. The solution is shown only in the central area around the anomalies, not for the full model domain.

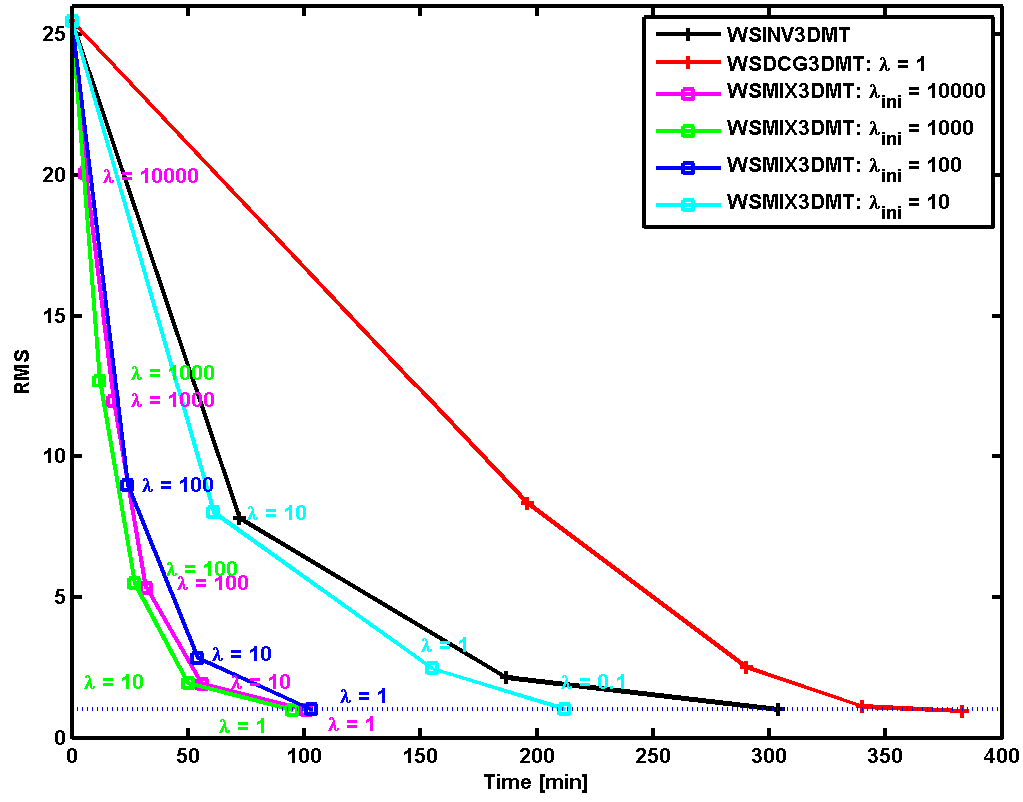


Figure 4. Convergence rates from WSINV3DMT (black), WSDCG3DMT with $\lambda = 1$ (red) and WSMIX3DMT with different initial λ_{ini} to the synthetic data generated from a model in Figure 1. Each square or plus symbol indicates one iteration. λ used in each iteration for WSMIX3DMT is printed next to its square symbols.

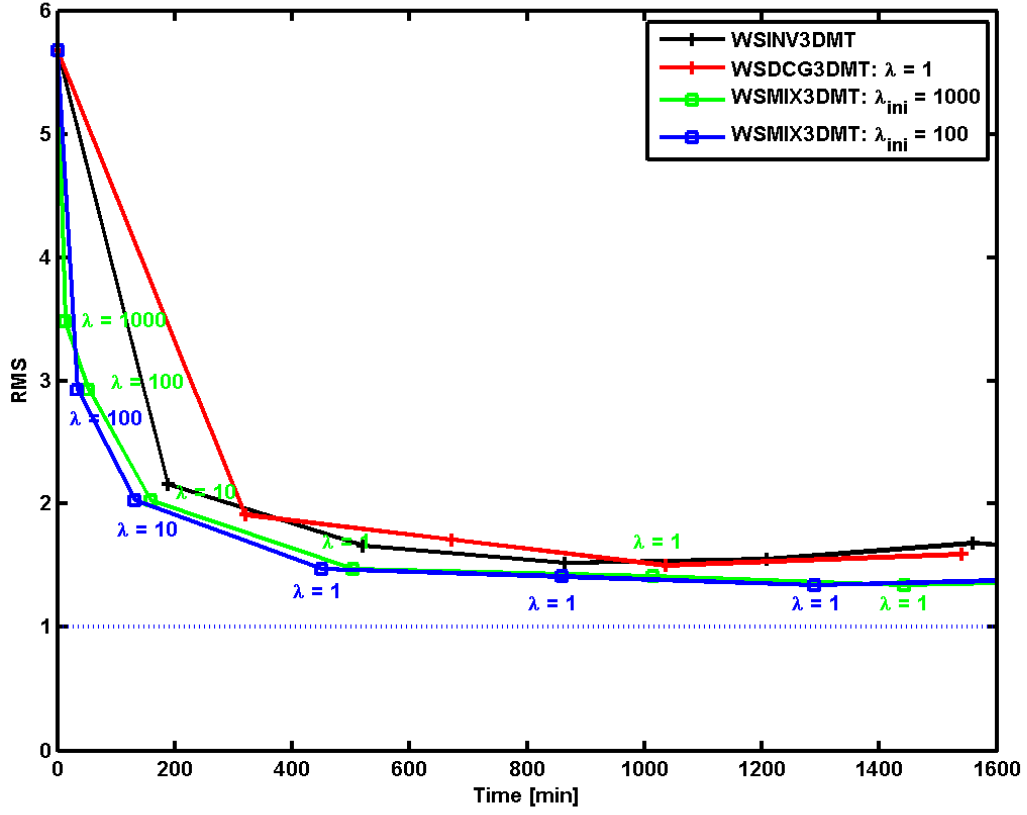


Figure 5. Convergence rates from WSINV3DMT (black), WSDCG3DMT with $\lambda = 1$ (red) and WSMIX3DMT with initial $\lambda_{ini} = 1000$ (green) and $\lambda_{ini} = 100$ (blue) to the EXTECH field dataset. Each square or plus symbol indicates one iteration. λ used in each iteration for WSMIX3DMT is printed next to its square symbols.

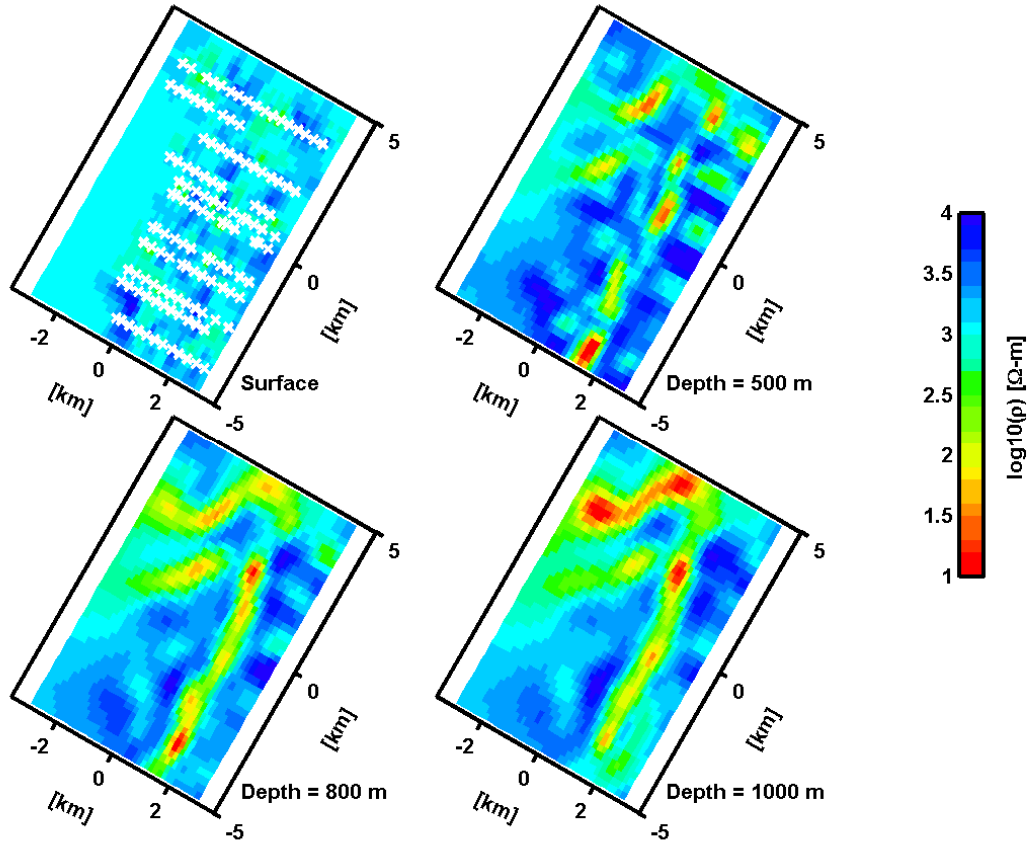


Figure 6. The inverse solution at various depths from the 5th iteration of the WSMIX3DMT method with initial $\lambda_{\text{ini}} = 100$. The EXTECH data used here consists of both vertical magnetic transfer function and full impedance tensor at 131 sites and 16 periods. The cross-symbols indicate the locations of the stations.

ภาคผนวก ข. Manuscript

Rung-Arunwan T. and W. Siripunvaraporn, 2010, An efficient modified hierarchical domain decomposition for 2-D Magnetotelluric forward modeling, submitted *Geophysical Journal International*, moderate revision.

An Efficient Modified Hierarchical Domain Decomposition for 2-D Magnetotelluric Forward Modeling

Tawat Rung-Arunwan^{1,2} and Weerachai Siripunvaraporn^{1,2}

¹Department of Physics, Faculty of Science, Mahidol University, Rama 6 Rd.,
Rachatawee, Bangkok, 10400, THAILAND

²ThEP, Commission of Higher Education, 328 Si Ayuthaya Road, Bangkok 10400,
THAILAND

Abstract

We use 2-D Magnetotelluric (MT) problems as a feasibility study to demonstrate that the 3-D MT modeling can be solved with a direct solver, even on a standard single processor PC. The scheme used is the hierarchical domain decomposition (HDD) method in which a global computational domain is uniformly split into many smaller non-overlapping subdomains. However, to make it more efficient, two modifications are made to the standard HDD method. Instead of three levels as in the standard HDD method, we classify the unknowns into four classes: the interiors, the horizontal and vertical interfaces and the intersections taking advantages of the finite-difference approximation. Four sets of smaller systems of equations are successively solved with a direct method (an LU factorization). The separation helps overcoming the memory overburden of a direct solver while remain computationally effective. To further enhance the speed of the code, a red-black ordering is applied to solve the horizontal and vertical interface reduced systems.

Numerical experiments on 2-D MT problem running on a single processor machine shows that CPU time and memory used are almost constant for any resistivity models, frequencies and modes as long as the model size remain the same. This is a clear advantage of our algorithm. Number of subdomains is a major factor controlling computational efficiency. Here, we also introduce a “memory map”, a tool we can use to pre-select “optimized” subdomains. Our 2-D experiments also shows that by splitting a domain with the optimized subdomains, this modified scheme can outperform the standard FD method in both CPU time and memory even running on a serial machine.

1. Introduction

To obtain magnetotelluric (MT) responses, the second order Maxwell's equation in either electric field or magnetic field is solved via three commonly used approaches: finite difference (FD) method (e.g. Mackie et al., 1994; Smith, 1996; Siripunvaraporn et al., 2002; Siripunvaraporn et al., 2005), finite element (FE) method (e.g. Wannamaker et al., 1987; Zyserman et al., 1999; Zyserman and Santos, 2000; Mitsuhata and Uchida, 2004;), and integral equation (IE) technique (e.g. Wannamaker, 1991; Xiong, 1992; Avdeed and Avdeeva, 2009). For complicated and geologically realistic two-dimensional (2D) and three-dimensional (3D) model, FD or FE methods are generally more efficient and robust than IE technique. In the past decades, FD method has gained more popularity due to its simplicity in technique and also its accuracy in solution.

In many problems, when model domain becomes very large, particularly in 3-D problems, solving the system of equations with the direct method is impractical in term of memory requirement (see Ben-Hadj-Ali et al., 2008 for 3-D frequency-domain full-waveform tomography; Streich, 2009 for 3-D MT;). The system is then alternatively solved with the iterative solvers (e.g. Bi-Conjugate Gradient (BiCG) method in Smith, 1996 and Xiong, 1999; Quasi Minimum Residual (QMR) in Siripunvaraporn et al., 2002; Preconditioned Conjugate Gradient (PCG) in Siripunvaraporn and Egbert, 2000; Minimum Residual Method (MRM) in Mackie et al., 1994). In many practical MT cases, the electrical resistivity model can be geologically complicated resulting in large conditioned number and therefore long computational time (see Patro and Egbert, 2009). Occasionally, the iterative solvers may become stagnant after many thousand of iterations and sometimes fail to converge. The calculated solution will therefore not be accurate and could mislead an interpretation if applied inside an inversion.

In high conditioned number case, being able to solve a problem with a direct solver is very crucial, if applicable. With direct method, accuracy is guarantee. Computational time is also controllable, because theoretically it is almost constant for any frequencies, modes or polarizations and resistivity models as long as the model domain remains the same size. In addition, the factorization used when solving the system can be re-used many times when

computing the sensitivity or Jacobian matrix (see Siripunvaraporn and Egbert, 2000) inside the inversion algorithm. In 3-D MT cases, the direct solver is still not practical with recent computer technology (see Streich, 2009). However, here we use the 2-D study to demonstrate that the 3-D problem can be efficiently solved with a direct solver even on a serial machine if the modified hierarchical domain decomposition developed in this paper is applied to.

Instead of computing on a large domain, a global domain can be splitting into several smaller local domains or subdomains. The solution on the global domain is then solved through the smaller systems of each subdomain. This technique is generally known as the domain decomposition (DD) technique. It is considered as a powerful tool in many large scale engineering problems (e.g. Lu and Shen, 1997; Bitzarakis et al., 1997; Larsson, 1999; Yin et al., 2002; Basermann et al., 2005; Lu et al., 2008; Wang et al., 2008;) and also in various multidimensional geophysical problems (e.g. Xiong, 1999; Zyserman et al., 1999; Zyserman and Santos, 2000; Xie et al., 2000; Pain et al., 2002; Ben-Hadj-Ali et al., 2008; Sourbier et al., 2008; Takei et al., 2010).

The domain decomposition method can be mainly classified into two categories: the overlapping technique where some region of the subdomain overlapping with the others (e.g. Xiong, 1999; Peng et al., 2009) and the non-overlapping method where neighboring subdomains share the same sub-boundaries (e.g. Lu and Shen, 1997; Zyserman et al., 1999; Zyserman and Santos, 2000; Lu et al., 2008; Wang et al., 2008). Comparison of the overlapping and the non-overlapping methods is mentioned in Chan and Goovaerts (1992) and Rice et al. (2000). Various schemes are used to solve the domain decomposition problems, such as the Schwartz algorithms (see Cai et al., 1998), Schur complement approach (see Smith et al., 1996; Saad, 2003; Zhang, 2005), the hierarchical domain decomposition approach (Smith et al., 1996; Takei et al., 2010), balancing domain decomposition method (Mandel, 1993), the interface relaxation methods (see Rice et al., 2000) among many other techniques.

In electromagnetic induction of the Earth, there are only a few papers demonstrating the use of domain decomposition method to solve MT forward problems. Zyserman et al. (1999) and Zyserman and Santos (2000) applied non-overlapping domain decomposition technique to 2-D and 3-D cases, respectively. In their techniques, sub-problems are iteratively solved via the interfaces enforced by the equivalent Robin-type transmission conditions. The memory requirement is significantly diminished due to no appearance of a large global matrix. Computational time is also greatly reduced when solving in the parallel computation (Zyserman and Santos, 2000). Although, the technique has proven to be numerically superior in the parallel system, the technique may not be suitable for serial computation. Xiong (1999) applied adaptive Schwarz overlapping domain decomposition technique for 3-D controlled source electromagnetic forward problems. In his method, all subdomains share overlapping regions. Each subdomain is independently solved and then updated from neighboring subdomains until the solution converges. The memory is significantly reduced. However, its total computational run time becomes larger than solving the whole system on single node processor (Xiong, 1999). Both schemes (Xiong, 1999; Zyserman et al., 1999; and Zyserman and Santos, 2000) show that efficiency in terms of computational time of the domain decomposition method can only be gained if running on parallel system. They are inferior if running on a serial machine.

In this paper, we investigated another method based on the hierarchical domain decomposition (HDD). Similar to other domain decomposition methods, the global domain is subdivided into many smaller subdomains. System of equations for each subdomain is separately formed and linked to the other via the interfaces. The hierarchical domain decomposition method can be directly applied to the MT problems both parallel and serial computations. Application of HDD on a parallel system is straightforward. Similar to others, calculation of each subdomain is performed separately on each processor node. A single interface system is then distributed to all processors for calculation. Theoretically, efficiency can be expected from applying the code to the parallel system. However, in practice, this parallel scheme requires substantial amount of communication time to exchange data among processors, particularly when solving the interface system. Efficiency is therefore platform-dependent. In this paper, we only illustrate the parallel algorithm but prefer not to demonstrate it numerically because our 2-D domain problem is “too” small for current

computer technology. The parallel algorithm will be later demonstrated on a bigger 3-D problem as a future research. In addition, this parallelization is not our main challenge. Our major challenge is the efficiency enhancement of HDD on a serial machine, not through a multi-processor machine.

Similar to other domain decomposition methods for MT problems (Xiong, 1999; Zyserman et al., 1999; and Zyserman and Santos, 2000), efficiency of HDD on a serial computation is low. However, in this paper, two modifications are developed and applied to the hierarchical domain decomposition method to increase its efficiency. First modification is the separation of interfaces into vertical and horizontal interfaces. This is natural for the finite-difference approximation scheme. Second modification is the application of red-black ordering to the re-ordered interface systems. With the two modifications, we will show that the modified HDD code for 2-D MT problems performs better than the conventional method even on a serial machine. Because we use a direct solver to solve system of equations, this 2-D experiment is also a feasibility study for future 3-D problems to demonstrate that the direct solver can be used to solve 3-D system of equations even with a serial calculation. These are therefore our main objectives for this paper.

Efficient modified HDD on a serial computation can also be applied to the parallel system. However, instead of parallelizing over subdomains, we parallelize over frequency. Calculation of MT responses of each frequency is performed serially on one processor. Thus, all frequencies are solved simultaneously but separately on multi-processor machines. This is used frequently in 3-D inversion algorithms (see Siripunvaraporn et al., 2004; 2005; Siripunvaraporn and Egbert, 2009; Siripunvaraporn and Sarakorn, 2010). In addition, this scheme does not require substantial amount of communication time between processors. It is therefore perfectly fit with the PC cluster platform which can be easily and cheaply built.

In addition, a major decisive factor that controls the efficiency of the modified HDD method is the number of subdomains. Selecting subdomains can be a trial and error processes. To

avoid wasting time to this process, here we introduce a “memory map” to help choosing “optimized” subdomains that yields the “best” computational performance. Memory map is pre-generated from several combinations of subdomains. Number of subdomains can be selected from the region of low memory in the memory map. This strategy often guarantees a faster CPU time than the standard method. The concept of memory map is new and first introduced here.

In the following, we first review the standard FD approach to solve a global domain problem. We then describes the basic idea of the hierarchical domain decomposition (HDD) and its parallel implementation. Then we describe the two modifications which help speeding up the HDD method on a serial calculation. Validations and numerical examples are given next along with the discussion. Conclusion are given at the end. Hereafter, we will refer to the standard finite difference for a global domain as FD2D, and to our modified hierarchical domain decomposition as MHDD2D.

2. Magnetotelluric forward modeling : Finite difference approach

Given an electrical conductivity (σ) or resistivity (ρ) model, to yield MT responses at the surface, the electric fields ($\mathbf{E}$) are computed from the second order Maxwell's equation,

$$\nabla \times \nabla \times \mathbf{E} = i\omega\mu\sigma\mathbf{E}, \quad (1a)$$

for the transverse electric field (TE) mode, while the magnetic fields ($\mathbf{H}$) are solved from,

$$\nabla \times \rho \nabla \times \mathbf{H} = i\omega\mu\mathbf{H}, \quad (1b)$$

for the transverse magnetic field (TM) mode, where ω is an angular frequency and μ the magnetic permeability. With finite difference approach, the conductivity or resistivity model is first discretized into many rectangular grids. An example of non-uniform grid discretization is shown in Figure 1. The unknown electric fields or magnetic fields are defined on the nodes (black dots) inside the domain, while the fields on the boundaries (left, right, top and bottom) are obtained from 1-D calculations. After applying finite difference to (1a) or (1b) and rearranging equation, both modes yield similar system of equations,

$$\mathbf{Ax} = \mathbf{b} , \quad (2)$$

where $\mathbf{x}$ represents the unknown internal electric or magnetic fields; $\mathbf{b}$ a vector containing the term associated with the boundary fields; and $\mathbf{A}$ a coefficient matrix which is large sparse five-banded symmetric and complex only on the diagonal (Siripunvaraporn and Egbert, 2000). Equation (2) for 2-D problem can be solved either directly or iteratively such as preconditioned conjugate gradient (PCG) method (Siripunvaraporn and Egbert, 2000). One of our aims is to demonstrate the use a direct solver for 3-D problem. An LU-factorization is therefore applied here to solve all systems of equations from FD2D and MHDD2D.

After calculating the electric fields, the magnetic fields can be calculated from solving the first order Maxwell's equation, the Faraday's law. MT responses are then computed from the ratio of electric to magnetic fields at the surface.

3. Hierarchical Domain Decomposition method

An alternative method to solve (2) is via the domain decomposition method. There are many different domain decomposition techniques. Here, we applied the hierarchical domain decomposition (HDD) method which is a non-overlapping technique to our 2-D MT problems. We start this section by describing the basic idea of the HDD method.

In every domain decomposition techniques, the model domain is split into several smaller subdomains. For simplicity, example mesh in Figure 1 is redrawn as in Figure 2 with uniform space, and is uniformly partitioned into 3×4 subdomains only as an illustration. The unknown electric or magnetic fields located at the nodes can be classified into three “hierarchical” types: (1) the interiors ($\bullet$), (2) the interfaces ($\blacksquare$ and $\blacktriangle$) and (3) the intersections ($\times$) from lowest to highest level, as shown in Figure 2. The intersections are defined as the highest level because they separate the interfaces. Similarly, the interfaces separate the interiors, so they are defined the next lower level. The interiors are therefore the lowest. With this configuration, the intersections must be solved first. Once the intersections are obtained, the interfaces can be successively calculated from the intersections. Similarly, the interiors can be successively computed from the interfaces. This hierarchical classification is slightly different from the “classic” Schur complement method (see Smith et al., 1996; Saad, 2003; Zhang, 2005;). In Schur complement method, the unknown fields are classified only the interiors and the interfaces.

For 2-D MT problem, assuming that the model domain is equally divided into $p \times q (= r)$ subdomains where p and q are number of subdomains in z - and y - directions, respectively, and r is the total number of subdomains. These partitions will yield a total of l interiors (or l/r for each subdomain), total of m interfaces and n intersections. Specifically, an inner subdomain i which has $l_{zi} \times l_{yi}$ ($= l/r$) interiors would have $2l_{zi} + 2l_{yi}$ interfaces, and 4 intersections, while outer or boundary subdomains would have less depending on their locations. By using Figure 1 and Figure 2 as an example, the model in Figure 1 is discretized into 12×20 grids, which is later decomposed into $3 \times 4 (=12)$ subdomains. In this example, there would be a total of 209 unknowns inside a global domain. When partitioning into 3×4 subdomains, an inner subdomain would then have 12 interiors, 14 interfaces and 4 intersections. The total numbers of interiors, interfaces and intersections are 144, 59 and 6, respectively.

By organizing the unknowns into three levels, the system of equations (2) can be reordered according to this configuration as follows,

$$\begin{pmatrix} \mathbf{F} & \mathbf{D} & \mathbf{0} \\ \mathbf{D}^T & \mathbf{G} & \mathbf{E} \\ \mathbf{0} & \mathbf{E}^T & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \\ \mathbf{h} \end{pmatrix}, \quad (3)$$

where $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ are $l \times l$ global interior coefficient matrix, $m \times m$ global interface coefficient matrix, and $n \times n$ intersection coefficient matrices, respectively. Global interior matrix $\mathbf{F}$ composes of many smaller $l/r \times l/r$ local interior sub-matrix $\mathbf{F}_i$ where $i = 1$ to r . Each $\mathbf{F}_i$ corresponds to a coupling within the interior elements inside the i subdomain. Global interface matrix $\mathbf{G}$ gathers all coefficients corresponding to an interaction between the interface elements, while $\mathbf{H}$ is diagonal matrix associating with the intersection elements. The inter-coupling coefficients between the interiors and interfaces are given in $\mathbf{D}$ with a dimension of $l \times m$, and between the interfaces and intersections are given in $\mathbf{E}$ with a dimension of $m \times n$. There is no coupling between the interiors and the intersections in our 2-D MT case as shown in Figure 2. Vectors $\mathbf{f}$, $\mathbf{g}$ and $\mathbf{h}$ are domain boundary fields associated with the interiors ($\mathbf{u}$), interfaces ($\mathbf{v}$) and intersections ($\mathbf{w}$), respectively. Figure 2 shows that there are no boundary fields that belong to the intersections. Therefore, $\mathbf{h} = \mathbf{0}$ in our 2-D problems.

According to the hierarchical domain decomposition technique, equation (3) can be decomposed into two reduced systems: the interior-interface reduced system and the interface-intersection reduced system. The interior-interface reduced system is derived from the coupling between the interiors and interfaces,

$$\begin{pmatrix} \mathbf{F} & \mathbf{D} \\ \mathbf{D}^T & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} - \mathbf{E}\mathbf{w} \end{pmatrix}, \quad (4)$$

while the interface-intersection reduced system is from the coupling between the interfaces and intersections,

$$\begin{pmatrix} \mathbf{S} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{g}' \\ \mathbf{h} \end{pmatrix}, \quad (5)$$

where the interface Schur complement matrix $\mathbf{S} = \mathbf{G} - \mathbf{D}^T \mathbf{F}^{-1} \mathbf{D}$ and $\mathbf{g}' = \mathbf{g} - \mathbf{D}^T \mathbf{F}^{-1} \mathbf{f}$. The unknowns are then successively solved from the highest to the lowest level. The intersections $\mathbf{w}$ are solved first from

$$\mathbf{H}'\mathbf{w} = \mathbf{h}', \quad (6)$$

where the intersection Schur complement matrix $\mathbf{H}' = \mathbf{H} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{E}$, and its right-hand side $\mathbf{h}' = \mathbf{h} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{g}'$. Once solving the intersections, the interfaces $\mathbf{v}$ and the interiors $\mathbf{u}$ can then be consecutively solved from

$$\mathbf{S}\mathbf{v} = \mathbf{g}' - \mathbf{E}\mathbf{w}, \quad (7)$$

and

$$\mathbf{F}_i \mathbf{u}_i = \mathbf{f}_i - \mathbf{D}_i \mathbf{v}. \quad (8)$$

Algorithm of the standard HDD method can be summarized below after decomposing the global domain into several subdomains.

1. Form $\mathbf{F}_i$, $\mathbf{f}_i$, $\mathbf{D}_i$ and factorize $\mathbf{F}_i$ of each subdomain.
2. Compute $\mathbf{D}_i^T \mathbf{F}_i^{-1} \mathbf{D}_i$ and $\mathbf{D}_i^T \mathbf{F}_i^{-1} \mathbf{f}_i$ of each subdomain.
3. Form $\mathbf{G}$, $\mathbf{g}$, $\mathbf{H}$, $\mathbf{h}$ and $\mathbf{E}$.
4. Construct $\mathbf{S} = \mathbf{G} - \sum \mathbf{D}_i^T \mathbf{F}_i^{-1} \mathbf{D}_i$ and $\mathbf{g}' = \mathbf{g} - \sum \mathbf{D}_i^T \mathbf{F}_i^{-1} \mathbf{f}_i$.
5. Factorize $\mathbf{S}$.
6. Build $\mathbf{H}' = \mathbf{H} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{E}$ and $\mathbf{h}' = \mathbf{h} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{g}'$.
7. Solve $\mathbf{H}'\mathbf{w} = \mathbf{h}'$.
8. Solve $\mathbf{S}\mathbf{v} = \mathbf{g}' - \mathbf{E}\mathbf{w}$.
9. Solve $\mathbf{F}_i \mathbf{u}_i = \mathbf{f}_i - \mathbf{D}_i \mathbf{v}$.
10. Merge $\mathbf{u}_i$, $\mathbf{v}$ and $\mathbf{w}$ as a solution for the system of equations (2).

The intersection Schur complement matrix $\mathbf{H}'$ (step 7) is dense, but its dimension, $n \times n$, is relatively small and therefore would not require a lot of computations. Similarly, the classical Schur method has a similar dense matrix but with a dimension equal to numbers of interfaces and intersections, i.e. $m+n \times m+n$. Thus, the hierarchical domain decomposition method yields a significant smaller dense matrix. The interface Schur complement matrix $\mathbf{S}$, in the hierarchical case, is not dense but sparse matrix. Example of its sparse pattern is shown in Figure 3a) from subdomains of Figure 2.

All equations including equation (6), (7) and (8) are solved with a direct method (here, an LU-factorization). To construct $\mathbf{S} = \mathbf{G} - \mathbf{D}^T \mathbf{F}^{-1} \mathbf{D}$ and $\mathbf{H}' = \mathbf{H} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{E}$ in step 4 and 6, after factorizations, $\mathbf{F}$ and $\mathbf{S}$ systems are solved with a series of different right hand sides: $\mathbf{D}^T$ and $\mathbf{E}^T$ for m times and n times, respectively. Solving each system just one time requires relatively small amount of computational resources, both memory and CPU time. However, as showing in the algorithm above, both systems are solved several times. Computational time for numerous solving (step 2, 4 and 6) plus factorizations (step 1, 5 and 7) can be more than just solving one large global system (equation 2) on a serial machine. This statement is correctly confirmed in Xiong (1999) and also in our MT numerical experiments in the next section. Once all main matrices are obtained; equation (6) and (7) is solved just one to obtain $\mathbf{w}$ and $\mathbf{v}$ in step 7 and 8, respectively. Equation (8) is then consecutively solved to obtain the interiors $\mathbf{u}$ within each subdomain in step 9. If each subdomain is equally discretized, this is equivalent as solving equation (8) r times.

Because domain decomposition is not highly efficient on a serial machine, another way of using domain decomposition on a serial computation is to modify the hierarchical matrix (3) and used it as a preconditioner when solving the system with the iterative solvers (e.g., Bitzarakis et al., 1997; Larsson, 1999; Benedetti et al., 2009; Grasedyck et al., 2009).

3.1 Parallel Implementation of HDD

Most parallel domain decomposition algorithms distribute computations of each subdomain to each processor (see examples in Xiong, 1999; Zyserman et al., 1999; and Zyserman and Santos, 2000). In this parallel scheme, step 1, 2 and 9 of each subdomain are performed

separately on each processor. After calculations, all results are sent to the master node. The bottleneck of this parallelization occurs from step 3 to 8. The most difficult parts for parallelization are to factorize $\mathbf{S}$ in step 5, to construct $\mathbf{H}' = \mathbf{H} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{E}$ and $\mathbf{h}' = \mathbf{h} - \mathbf{E}^T \mathbf{S}^{-1} \mathbf{g}'$ in step 6 and to solve $\mathbf{S} \mathbf{v} = \mathbf{g}' - \mathbf{E} \mathbf{w}$ in step 8. Once distributing $\mathbf{S}$ to all processors, this process requires a lot of communication time among processors when factorizing and solving system of equations. Efficiency of this parallel scheme would depend significantly on the parallel algorithms which also depend on computer architectures (see Lu and Shen, 1997; Kocak and Akay, 2001). Many massive parallel manufacturers have provided their own efficient parallel algorithms to solve system of equations. These algorithms show best performance only on their own platforms.

However, this conventional parallel scheme could be a problem for PC cluster platform or distributed memory systems. Efficiency would be relatively low if switch or hub used to communicate among processors is slow regardless of how efficient the algorithm is. Parallel implementation is not the purposes of our paper as previously described. We therefore opt not to show the numerical experiments of HDD on parallel systems. Experiments with 3-D MT problems would be an interesting research to pursue which is beyond our scope here.

4. Modified hierarchical domain decomposition method

Earlier numerical experiments on single processor machine show that a straightforward application of the HDD method to the 2-D MT problems requires less memory storage than standard method. However, its computational time becomes longer. In order to make the hierarchical domain decomposition method more efficient on a serial machine for our 2-D MT problem, two modifications are necessary. First, the separation of the interfaces into vertical and horizontal interfaces will break the larger interface system into two smaller vertical and horizontal interface systems which would lead to a memory reduction. Second, the red-black ordering technique is applied inside the horizontal and vertical interface systems to further help decreasing the computational time.

Taking advantage of the rectangular discretization of the FD approximation, the interfaces can be further classified into two types: the horizontal interfaces (■ in Figure 2) and the

vertical interfaces ($\blacktriangle$ in Figure 2). Number of interfaces (m) is then divided into number of horizontal interfaces (m_h) and number of vertical interfaces (m_v) where $m = m_h + m_v$. The system of equations (3) can then be reassembled as follows,

$$\begin{pmatrix} \mathbf{F} & \mathbf{D}_H & \mathbf{D}_V & \mathbf{0} \\ \mathbf{D}_H^T & \mathbf{G}_H & \mathbf{0} & \mathbf{E}_H \\ \mathbf{D}_V^T & \mathbf{0} & \mathbf{G}_V & \mathbf{E}_V \\ \mathbf{0} & \mathbf{E}_H^T & \mathbf{E}_V^T & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v}_H \\ \mathbf{v}_V \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g}_H \\ \mathbf{g}_V \\ \mathbf{h} \end{pmatrix}, \quad (9)$$

where $_H$ and $_V$ represent horizontal and vertical interfaces, respectively. The main difference from the original hierarchical domain decomposition would be at the separation of $\mathbf{G}$ matrix into $\mathbf{G}_H$ and $\mathbf{G}_V$, where $\mathbf{G}_H$ gathers all coefficients corresponding to a coupling between the horizontal interfaces, and similarly for $\mathbf{G}_V$ corresponding to a coupling between the vertical interfaces. With new classification, both vertical interfaces ($\mathbf{v}_V$) and horizontal interfaces ($\mathbf{v}_H$) are situated in the middle level between the intersection ($\mathbf{w}$) and the interior ($\mathbf{u}$) which are the highest and lowest, respectively. The interior-interface and interface-intersection reduced systems in equation (4) and (5) become

$$\begin{pmatrix} \mathbf{F} & \mathbf{D}_H & \mathbf{D}_V \\ \mathbf{D}_H^T & \mathbf{G}_H & \mathbf{0} \\ \mathbf{D}_V^T & \mathbf{0} & \mathbf{G}_V \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v}_H \\ \mathbf{v}_V \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g}_H - \mathbf{E}_H \mathbf{w} \\ \mathbf{g}_V - \mathbf{E}_V \mathbf{w} \end{pmatrix}, \quad (10)$$

and

$$\begin{pmatrix} \mathbf{S}_{HH} & \mathbf{S}_{HV} & \mathbf{E}_H \\ \mathbf{S}_{VH} & \mathbf{S}_{VV} & \mathbf{E}_V \\ \mathbf{E}_H^T & \mathbf{E}_V^T & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{v}_H \\ \mathbf{v}_V \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{g}'_H \\ \mathbf{g}'_V \\ \mathbf{h} \end{pmatrix}, \quad (11)$$

respectively. Here, the interface Schur complement matrix $\mathbf{S}$ is decomposed into $\mathbf{S}_{HH}$, $\mathbf{S}_{HV}$, $\mathbf{S}_{VH}$ and $\mathbf{S}_{VV}$ as follow,

$$\begin{pmatrix} \mathbf{S}_{HH} & \mathbf{S}_{HV} \\ \mathbf{S}_{VH} & \mathbf{S}_{VV} \end{pmatrix} = \begin{pmatrix} \mathbf{G}_H & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_V \end{pmatrix} - \begin{pmatrix} \mathbf{D}_H^T \\ \mathbf{D}_V^T \end{pmatrix} \mathbf{F}^{-1} (\mathbf{D}_H \quad \mathbf{D}_V), \quad (12)$$

and

$$\begin{pmatrix} \mathbf{g}'_H \\ \mathbf{g}'_V \end{pmatrix} = \begin{pmatrix} \mathbf{g}_H \\ \mathbf{g}_V \end{pmatrix} - \begin{pmatrix} \mathbf{D}_H^T \\ \mathbf{D}_V^T \end{pmatrix} \mathbf{F}^{-1} \mathbf{f}. \quad (13)$$

Example of the sparsity pattern of the modified Schur interface (12) is shown in Figure 3b) to be compared with the original Schur interface matrix $\mathbf{S}$ (Figure 3a). Similar to the original hierarchical domain decomposition, the unknown fields are successively solved from the highest level to the lowest level. The intersections $\mathbf{w}$ will be solved first from

$$\mathbf{H}' \mathbf{w} = \mathbf{h}', \quad (14)$$

where, $\mathbf{H}' = \mathbf{H} - \begin{pmatrix} \mathbf{E}_H^T & \mathbf{E}_V^T \end{pmatrix} \begin{pmatrix} \mathbf{S}_{HH} & \mathbf{S}_{HV} \\ \mathbf{S}_{VH} & \mathbf{S}_{VV} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{E}_H \\ \mathbf{E}_V \end{pmatrix}$, and its right-hand side

$\mathbf{h}' = \mathbf{h} - \begin{pmatrix} \mathbf{E}_H^T & \mathbf{E}_V^T \end{pmatrix} \begin{pmatrix} \mathbf{S}_{HH} & \mathbf{S}_{HV} \\ \mathbf{S}_{VH} & \mathbf{S}_{VV} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{g}'_H \\ \mathbf{g}'_V \end{pmatrix}$. After solving the intersections $\mathbf{w}$, the vertical

interfaces $\mathbf{v}_V$ and the horizontal interfaces $\mathbf{v}_H$ can be split and solved separately as,

$$(\mathbf{S}_{VV} - \mathbf{S}_{VH} \mathbf{S}_{HH}^{-1} \mathbf{S}_{HV}) \mathbf{v}_V = \mathbf{g}'_V - \mathbf{E}_V \mathbf{w} - \mathbf{S}_{VH} \mathbf{S}_{HH}^{-1} (\mathbf{g}'_H - \mathbf{E}_H \mathbf{w}), \quad (15)$$

and,

$$\mathbf{S}_{HH} \mathbf{v}_H = \mathbf{g}'_H - \mathbf{E}_H \mathbf{w} - \mathbf{S}_{HV} \mathbf{v}_V. \quad (16)$$

Dimension of $\mathbf{S}_{HH}$ and $\mathbf{S}_{VV}$ from (15) and (16) are $m_h \times m_h$ and $m_v \times m_v$, respectively, which are smaller than $m \times m$ $\mathbf{S}$ matrix of (7). They are therefore faster to solve and less memory storage. This is one clear advantage of classifying the interfaces into the horizontal and

vertical interfaces. After obtaining $\mathbf{w}$ and $\mathbf{v}$, the interiors $\mathbf{u}$ can then be consecutively solved from

$$\mathbf{F}_i \mathbf{u}_i = \mathbf{f}_i - \mathbf{D}_{Hi} \mathbf{v}_H - \mathbf{D}_{Vi} \mathbf{v}_V. \quad (17)$$

To further increase the efficiency of our modified scheme, red-black coloring technique (See Press et al., 1992 and Saad, 2003) is applied to (15) and (16) to help reducing the computational time. Under the red-black ordering, the unknowns inside of $\mathbf{S}_{VV}$ and $\mathbf{S}_{HH}$ are classified into red and black unknowns. The idea of Schur complement is again applied to this coloring system of the interfaces. The reduced systems are then derived and recursively solved to the red and to the black systems. This modification demonstrates the application of Schur domain decomposition inside the hierarchical domain decomposition (see Rung-Arunwan, 2010 for further detail).

With both modifications, the modified hierarchical domain decomposition (MHDD2D) can outperform the FD2D code even running on a serial computational machine as showing in the next section.

5. Numerical Experiments

In this section, we first validate that the responses from our modified hierarchical domain decomposition method (MHDD2D) are as accurate as those from FD2D. Next, computational costs on a single processor are measured with different combinations of subdomains. A memory map is then introduced as a strategy to select an “optimized” number of subdomain where computational costs are minimized (i.e., relatively faster or at least equivalently to FD2D, but with a fraction of memory).

5.1 Validation Tests

To validate the MHDD2D approach, we show the apparent resistivities and phases of both TM and TE modes at three frequencies calculated from the model shown in Figure 1. The calculated responses from our MHDD2D approach are directly compared to those obtained from FD2D as in section 2. In this test, the model and air of Figure 1 is non-uniformly discretized into 80×240 grids in z - and y -direction, respectively. For FD2D method, the unknown to be solved is 18,881. For MHDD2D, the model domain is uniformly split into 4×8 (z - and y - direction, respectively) subdomains. With this 4×8 subdomains, the 18,881 unknowns will be divided into 551 interiors for each subdomain (or a total of 17,632 interiors), 696 horizontal interfaces and 532 vertical interfaces, and 21 intersections. Total memory requirement of MHDD2D is about 21.7 Mbytes, which is approximately one-third of FD2D (about 71.09 Mbytes). Memory estimation will be discussed in subsection 5.2.1.

Figure 4 shows that the calculated responses from both FD2D and MHDD2D are perfectly identical on both modes. Their difference is in the round-off level which is insignificant. This is expected since both methods solve the same system of equation, except that the MHDD2D method splits the computational domain into many smaller subdomains, and then solves smaller systems. In addition, we have performed validation tests on various synthetic models and real model (see inverted model from real data in Siripunvaraporn and Egbert, 2000) with several combinations of subdomains. All validation tests show that there is no difference from both methods (Rung-Arunwan, 2010). These have validated our MHDD2D method for both TM and TE modes.

5.2 Comparisons of Computational Efficiency

Next, to prove the efficiency of our modified domain decomposition scheme, we ran the code on several synthetic 2-D models and also real “inverted” model (from Siripunvaraporn and Egbert, 2000) for both TM and TE modes. Because a direct method (LU-factorization) is used to solve all systems of equations, computational time and memory requirements are no difference among different models, modes (TM or TE) and frequency if domain size is the same. Model of Figure 1 is therefore used as a representative to demonstrate the effectiveness of our code.

Model and air of Figure 1 is discretized into three size meshes: 40×120 (small), 80×240 (medium) and 120×360 (large). These three meshes are then uniformly subdivided into $p \times q$ subdomains, where p and q are numbers of subdomains in z-dir and y-dir, respectively, starting from 2. Estimated memory usage and actual calculation time for each combination of subdomains for each mesh are compared with those from FD2D. Comparison results are plotted and shown in Figure 5 for 40×120 mesh, Figure 6 for 80×240 mesh and Figure 7 for 120×360 mesh. Relative CPU time and memory (both in percents) are calculated from $(\text{time}_{\text{MHDD2D}} - \text{time}_{\text{FD2D}}) * 100 / \text{time}_{\text{FD2D}}$ and $(\text{mem}_{\text{MHDD2D}} - \text{mem}_{\text{FD2D}}) * 100 / \text{mem}_{\text{FD2D}}$, respectively. Positive relative time and relative memory indicate that MHDD2D is less efficiency than FD2D and therefore spend more calculation time and require more memory, while negative reflects the opposite, i.e. MHDD2D is more efficient. Actual memory usage of FD2D are 8.77 Mbytes, 71.09 Mbytes and 240.97 Mbytes for small, medium and large, respectively, while actual CPU time on an Intel Core Two Duo 6400, 2.13 GHz machine are 0.08 second, 1.12 second and 4.16 second, respectively. Actual CPU time and memory used of MHDD2D can thus be inferred from these actual values of FD2D and the maps shown in Figure 5, 6 and 7, respectively.

5.2.1 “Memory Map” and Memory Comparison

Total memory usage of MHDD2D can be calculated from numbers of subdomains in z-dir (p) and y-dir (q), number of interiors (l/r) for each subdomain, numbers of horizontal interfaces (m_h) and vertical interfaces (m_v) and number of intersections (n). However, it is quite complicated to express in a simple formula. It is therefore pre-estimated from the allocated variables inside the code to produce the “memory map” before running the actual code. Memory map displays minimum memory used for different combinations of subdomains as shown in Figure 5a, 6a and 7a. The concept of memory map is very useful and will be demonstrated in later subsection.

In contrast to MHDD2D, total memory usage for FD2D can be easily estimated from $(N_y - 1)(N_z - 1)(3N_z + 1) * 16$ where N_y and N_z is grid discretization in y-dir and z-dir, respectively.

Multiplication with 16 is required because complex double precision is used. Because a large global matrix (equation 2) of FD2D is broken into many smaller sub-matrices (equation 9) for MHDD2D, memory requirement for different combinations of subdomains should therefore be less than that of FD2D. This is evidently shown in Figure 5a, 6a and 7a, where negative percentage is all over the map indicating less memory requirement of MHDD2D. However, total memory usage varies according to numbers of subdomains used in both directions.

From all three figures, there are two cases where memory usage is relatively large (but still less than FD2D). First case is when the domain is divided into “large” numbers of subdomains. When number of subdomains become large (e.g., 20×30 subdomains in Figure 7a), number of interiors per subdomain is small (see Table 1), but total number of interfaces are high (Table 1). More memory is therefore required to store and solve those interface coefficient matrices ($\mathbf{G}_H$, $\mathbf{G}_V$, $\mathbf{S}_{HH}$, $\mathbf{S}_{HV}$, $\mathbf{S}_{VH}$ and $\mathbf{S}_{VV}$ in 10 and 11). Although intersections ($\mathbf{H}$) also increase, it would not significantly affect. In contrast, when small number of subdomains used (e.g., 3×3 subdomains in Figure 7a), total numbers of interfaces in both directions are small (see Table 1), but number of interiors per subdomain becomes very high (Table 1). Large number of interiors causes matrix $\mathbf{F}_i$ (equation 10) of each subdomain to require more memory to store and solve the system of equations (equation 13 and 17). Note that we use LU decomposition to solve all systems of equations. Some “extra” memory is therefore required to fill the empty band of the sparse matrix. This extra memory has already been accounted for in Figure 5a, 6a and 7a.

5.2.2 Comparisons of CPU time

Calculation time cannot be pre-estimated as the memory usage, it can only be obtained from running the actual code on the computer. Relative CPU time from small, middle and large meshes are shown in Figure 5b, 6b and 7b, respectively, from different combinations of subdomains. They are obtained from running on a single processor machine; here, an Intel Core Two Duo 6400, 2.13 GHz machine. Different machines or architectures may result differently. However, patterns of relative CPU time should remain approximately the same.

For small 40×120 mesh, relative CPU time of MHDD2D is at least 30% more than that of FD2D in every combination of subdomains (Figure 5b). Although a larger system of equations (equation 2) is broken into many smaller systems (equation 9), successively solving a series of these smaller systems (see equation 4-6, and 10-17) can outperform solving a global system of FD2D. This reflects in larger CPU time as shown with all positive in Figure 5b. Although there is no benefit of MHDD2D for smaller 40×120 meshes in term of CPU time, better efficiency can be gained up to 20% from larger meshes as shown with negative zones in Figure 6b for 80×240 mesh and in Figure 7b for 120×360 mesh. This shows that when grid discretization becomes large, MHDD2D will become more effective, even with a serial computation. This conclusion is significant, especially for future implementing the idea of MHDD2D to 3-D cases. In 3-D, the discretization mesh would be clearly a lot larger than what we used in 2-D case.

5.3 Optimized Number of Subdomains : Pre-Selection

Figure 5a, 6a and 7a show that there are regions where memory requirement is “minimum”. The minimized memory zones have the centers at 5×6 subdomains for 40×120 mesh (Figure 5a), at 8×8 subdomains for 80×240 mesh (Figure 6a) and at 10×9 subdomains for 120×360 mesh (Figure 7a). The interiors, horizontal interfaces, vertical interfaces and intersections for these three subdomains are given in Table 2.

By matching Figure 5a, 6a and 7a to Figure 5b, 6b and 7b, respectively, we found that the minimized memory zones are coincidentally occurred almost the same regions as the minimized CPU time zone. Both areas will be referred to as the “optimized” regions, because both memory and CPU time are least used. In this “optimized” regions, numbers of interiors, horizontal interfaces, vertical interfaces and intersections are properly justified or balancing (as shown in Table 2), so that solving and storing $\mathbf{F}_i$, $\mathbf{G}_H$, $\mathbf{G}_V$, $\mathbf{S}_{HH}$ and $\mathbf{S}_{VV}$ and $\mathbf{H}$ matrices are relatively fast and less memory requirement. Larger or smaller number of subdomains would cause an unbalance to these numbers. Larger number of subdomains would increase the interface sizes, while smaller number of subdomains would increase the interior size. Both cases would produce a large matrix, which would dominate both calculation time and memory usage.

The agreement between the optimized CPU time and memory usage has lead to the idea of subdomain selection. Usually, choosing number of subdomains that yields least CPU time and smallest memory requirement would be a trial and error strategy. Here, we propose to select the “optimized” subdomains from the memory map, shown in Figure 5a, 6a and 7a. Because memory usage can be pre-estimated from the variable allocations inside the code, this number can be printed out and plotted in a map from different combinations of subdomains. The optimized subdomains can therefore be chosen from the region of “least” memory requirement. There would be a higher chance that CPU time performance of MHDD2D would be better than FD2D if choosing subdomains from this region. When implementing MHDD2D to 3-D case, similar technique can be used to avoid trial and error selections.

5.4 Comparison of modified and non-modified hierarchical domain decomposition methods

For the original hierarchical domain decomposition technique, memory requirements for $\mathbf{F}$ and $\mathbf{H}$ matrices in (4) and (5) are identical to those in (10) and (11) for our modified hierarchical domain decomposition. However, interface matrices, $\mathbf{G}$ and $\mathbf{S}$ in (4) and (5) (Figure 3a), depends on the sum of horizontal interfaces and vertical interfaces ($m = m_h + m_v$). These matrices are therefore larger than $\mathbf{G}_H$, $\mathbf{G}_V$, $\mathbf{S}_{HH}$ and $\mathbf{S}_{VV}$ in (10) and (11) (Figure 3b) for the modified scheme around 20-50% depending on the number of subdomains (r). Memory requirement for non-modified hierarchical domain decomposition would therefore up to 50% more than the modified case from our 2-D study, but it is still less than FD2D.

In term of computational time, the standard hierarchical domain decomposition would require about the same CPU time to solve $\mathbf{F}_i$ and $\mathbf{H}$ systems of equations. However, our 2-D study reveal that for the interface parts, larger $\mathbf{G}$ and $\mathbf{S}$ in (4) and (5) of the non-modified code requires solving time slightly more or less than solving smaller $\mathbf{G}_H$, $\mathbf{G}_V$, $\mathbf{S}_{HH}$ and $\mathbf{S}_{VV}$ in (10) and (11) of the modified code. Not much can be gained in terms of CPU time in this part, but a lot more in terms of memory. However, by reducing the larger $\mathbf{G}$ into $\mathbf{G}_H$ and $\mathbf{G}_V$ (from Figure 3a to 3b), red-black ordering can be easily applied for solving $\mathbf{G}_H$ and $\mathbf{G}_V$, but not

directly to $\mathbf{G}$ in (4). With the red-black ordering, about 10-50% depending on a combination of subdomains can be gained comparing to the original HDD method for the 2-D case. Red-black ordering can be easily implementing in 3-D case as well, this would help further decreasing the computational time.

6. Conclusions

We have demonstrated the efficiency of the MHDD2D code for 2-D MT forward modeling. MHDD2D is a modified version of the hierarchical domain decomposition method. The original scheme begins by dividing a global computational domain into several subdomains. Then, the unknown nodes are classified into three different kinds: interiors, interfaces and intersections. A global system of equations is re-organized according to these configurations producing three sets of smaller systems of equations. The intersection reduced system of equations is solved first to obtain the intersections. The calculated intersections are then used in the right hand-side of the interface systems of equations to compute the interfaces. Similarly, the calculated interfaces are input in the interior systems of equations to compute the interiors inside each subdomain.

Normally, HDD is applied on a parallel system. Efficiency of the HDD method on a serial machine is very low comparing to the conventional method. To enhance the efficiency of the hierarchical method on single processor computer, the interfaces of the standard hierarchical domain decomposition method is further separated into horizontal interfaces and vertical interfaces by taking an advantage of the rectangular discretization of the finite difference. Our modified version will then have four sets of smaller systems of equations, instead of three as in the original version. The division of the interfaces into horizontal and vertical interfaces helps substantially decreasing the size of memory usage. However, it does little help in computing time. Red-black coloring is then applied to substantially reduce the computational time of the code.

By running MHDD2D with several combinations of subdomains on single processor machine, the optimized subdomains can be selected from the memory map generated prior

the run. Dividing the global domain with the optimized subdomains, MHDD2D can run up to 20-30% faster and require up to 70% less memory than FD2D on single processor machine. This conclusion is very crucial. It indicates that the same hierarchical domain decomposition algorithm can be extended and applied to 3-D problem. By applying modified HDD method to 3-D case, 3-D forward problem can now be solved with a direct method, even on standard single processor PC. With the direct solver, its factorized matrices can be re-used several times with different right-hand sides. This will help speeding up the sensitivity calculation in the 3-D inversion process. Most importantly for a direct solver, computational time is controllable and independent of frequencies, modes and resistivities, as long as the domain size remains the same.

7. References

- Avdeev, D. & Avdeeva, A., 2009, 3D Magnetotelluric inversion using a limited-memory quasi-Newton optimization, *Geophysics*, 74 (3), F45-F57.
- Basermann, A., Jaekel, U., Nordhausen, M., & Hachiya, K., 2005. Parallel iterative solvers for sparse linear systems in circuit simulation, *Future Generation Computer Systems*, 21(8), 1275-1284.
- Ben-Hadj-Ali, H., Operto, S., Virieux, J., & Sourbier, F., 2008. 3D frequency-domain full-waveform tomography based on a domain decomposition forward problem, *SEG Technical Program Expanded Abstracts*, 27(1), 1945-1949.
- Benedetti, I., Milazzo, A., & Aliabadi, M.H., 2009, A fast dual boundary element method for 3D anisotropic crack problems, *Inter. Jour. Numer. Methods. Eng.*, 80, 1356-1378.
- Bitzarakis, S., Papadarakakis, M., & Kotsopoulos, A., 1997. Parallel solution techniques in computational structural mechanics, *Computer Methods in Applied Mechanics and Engineering*, 148(1-2), 75-104.
- Cai, X.C., Casarin, M.A., Elliot Jr., F.W., & Widlund, O.B., 1998, Overlapping Schwarz Algorithms for solving Helmholtz equation, *Contemporary Math*, 218, 437-445.
- Chan, T.F. & Goovaerts, D., 1992, On the relationship between overlapping and nonoverlapping domain decomposition methods, *SIAM J. Matrix Anal. Appl.*, 13, 663-670.
- Grasedyck, L., Kriemann, R., & Le Borne, S, 2009, Domain decomposition based H-LU preconditioning, *Numerische Mathematik*, 112 (4), 565-600.

- Kocak, S., & Akay, H.U., 2001, Parallel Schur complement method for large-scale systems on distributed memory computers, *Applied Mathematical Modelling*, 25, 873-886.
- Larsson, E., 1999. A Domain Decomposition Method for the Helmholtz Equation in a Multilayer Domain, *SIAM Journal on Scientific Computing*, **20**(5), 1713-1731.
- Lu, Y. & Shen, C., 1997. A domain decomposition finite-difference method for parallel numerical implementation of time-dependent Maxwell's equations, *IEEE Transactions on Antennas and Propagation*, **45**(3), 556-562.
- Lu, Z., An, X., & Hong, W., 2008. A fast domain decomposition method for solving three-dimensional large-scale electromagnetic problems, *IEEE Transactions on Antennas and Propagation*, **56**(8 I), 2200-2210.
- Mackie, R., Smith, J., & Madden, T., 1994. Three-dimensional electromagnetic modeling using finite difference equations: The magne-totelluric example, *Radio Science*, **29**(4), 923-935.
- Mandel, J., 1993, Balancing domain decomposition, *Comm. Numer. Methods Engrg*, 9, 233-241.
- Mitsuhashi, Y. & Uchida, T., 2004, 3D Magnetotelluric modeling using the T- Ω finite-element method, *Geophysics*, 69 (1), 108-119.
- Pain, C., Herwanger, J., Worthington, M., & De Oliveira, C., 2002. Effective multidimensional resistivity inversion using finite-element techniques, *Geophys. J Int.*, **151**(3), 710-728.
- Patro, P.K., & Egbert, G.D., 2008, Regional conductivity structure of Cascadia: Preliminary results from 3D inversion of USArray transportable array Magnetotelluric data, *Geophys. Res. Lett.*, 35 (20), art. no. L20311.
- Peng, T., Sertel, K., & Volakis, J.L., 2009, Fully overlapping domain-decomposition for fast optimization of small antennas in large-scale composite media, *2009 Computational Electromagnetics International Workshop*, CEM 2009, art. no. 5228103, 77-81.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P., 1992. *Numerical Recipes in FORTRAN: The art of scientific computing*, Cambridge University Press, 2nd edn.
- Rice, J.R., Tsompanopoulou, P., & Vavalis, E., 2000, Interface relaxation methods for elliptic differential equations, *Applied. Numer. Math.*, 32, 219-245.

- Rung-Arunwan, T., 2010, An efficient modified hierarchical domain decomposition for 2D Magnetotelluric forward modeling, M.Sc. Thesis. Mahidol University.
- Saad, Y., *Iterative Methods for Sparse Linear Systems*, 2nd ed, SIAM, Philadelphia, 2003.
- Siripunvaraporn, W. & Egbert, G., 2000. An efficient data-subspace inversion method for 2D magnetotelluric data, *Geophysics*, **65**(3), 791-803.
- Siripunvaraporn, W., Egbert, G., & Lenbury, Y., 2002. Numerical accuracy of magnetotelluric modeling: A comparison of finite difference approximations, *Earth Planets Space*, **54**(6), 721-725.
- Siripunvaraporn, W., Uyeshima, M., & Egbert, G., 2004, Three-dimensional inversion for Network-Magnetotelluric data, *Earth Planets Space*, **56**, 893-902.
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., & Uyeshima, M., 2005, Three-dimensional Magnetotelluric inversion: data-space method, *Phys. Earth and Planetary Interiors*, **150**, 3-14.
- Siripunvaraporn, W & Egbert, G., 2009, WSINV3DMT: Vertical magnetic field transfer function inversion and parallel implementation, *Phys. Earth and Planetary Interiors*, **173**, 317-329.
- Siripunvaraporn, W & Sarakorn, W., 2010, An efficient three-dimensional Magnetotelluric inversion: a mixed of the data space conjugate gradient method and the data space Occam's method, *submitted to GJI*.
- Smith, B., Bjørstad, P., & Gropp, W., 1996. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press.
- Smith, J. T., 1996. Conservative modeling of 3D electromagnetic fields, Part I: Properties and error analysis, *Geophysics*, **61**(5), 1308-1318.
- Sourbier, F., Haidar, A., Giraud, L., Operto, S., & Virieux, J., 2008. Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D full-waveform inversion?, *SEG Technical Program Expanded Abstracts*, **27**(1), 2147-2151.
- Streich, R., 2009. 3D finite-difference frequency-domain modeling of controlled-source electromagnetic data: Direct solution and optimization for high accuracy, *Geophysics*, **74**(5), F95-F105.
- Takei, A., Sugimoto, S.-I., Ogino, M., Yoshimura, S., & Kanayama, H., 2010. Large-scale analysis of high frequency electromagnetic field by hierarchical domain decomposition

method with direct method in subdomains, *IEEE Transactions on Fundamentals and Materials*, **130**(3), 239-246.

- Wang, B., Mittra, R., & Shao, W., 2008. A domain decomposition finite-difference method utilizing characteristic basis functions for solving electrostatic problems, *IEEE Transactions on Electromagnetic Compatibility*, **50**(4), 946-952.
- Wannamaker, P., 1991. Advances in three-dimensional magnetotelluric modeling using integral equations, *Geophysics*, **56**(11), 1716-1728.
- Wannamaker, P., Stodt, J., & Rijo, L., 1987. A stable finite element solution for two-dimensional magnetotelluric modelling, *Geophys. J. R. astr. Soc.*, **88**(1), 277-296.
- Xie, G., Li, J., Majer, E., Zuo, D., & Oristaglio, M., 2000. 3D electromagnetic modeling and nonlinear inversion, *Geophysics*, **65**(3), 804-822.
- Xiong, Z., 1992. Electromagnetic modeling of 3D structures by the method of system iteration using integral equations, *Geophysics*, **57**(12), 1556-1561.
- Xiong, Z., 1999. Domain decomposition for 3D electromagnetic modeling, *Earth Planets Space*, **51**(10), 1013-1018.
- Yin, L., Wang, J., & Hong, W., 2002. A novel algorithm based on the domain-decomposition method for the full-wave analysis of 3D electromagnetic problems, *IEEE Transactions on Microwave Theory and Techniques*, **50**(8), 2011-2017.
- Zhang, F., 2005, The Schur complement and its applications, *Springer*, ISBN 0387242716.
- Zyserman, F. & Santos, J., 2000. Parallel finite element algorithm with domain decomposition for three-dimensional magnetotelluric modelling, *Journal of Applied Geophysics*, **44**(4), 337-351.
- Zyserman, F., Guarracino, L., & Santos, J., 1999. A hybridized mixed finite element domain decomposed method for two dimensional magnetotelluric modelling, *Earth Planets Space*, **51**(4), 297-306.

7. Acknowledgements

This research has been supported by Thai Center of Excellence in Physics (ThEP) and by Thailand Research Fund (TRF: RMU5080025). The authors would like to thank Colin G. Farquharson, anonymous reviewer and the editor, Oliver Ritter, for their comments to help improve the manuscript

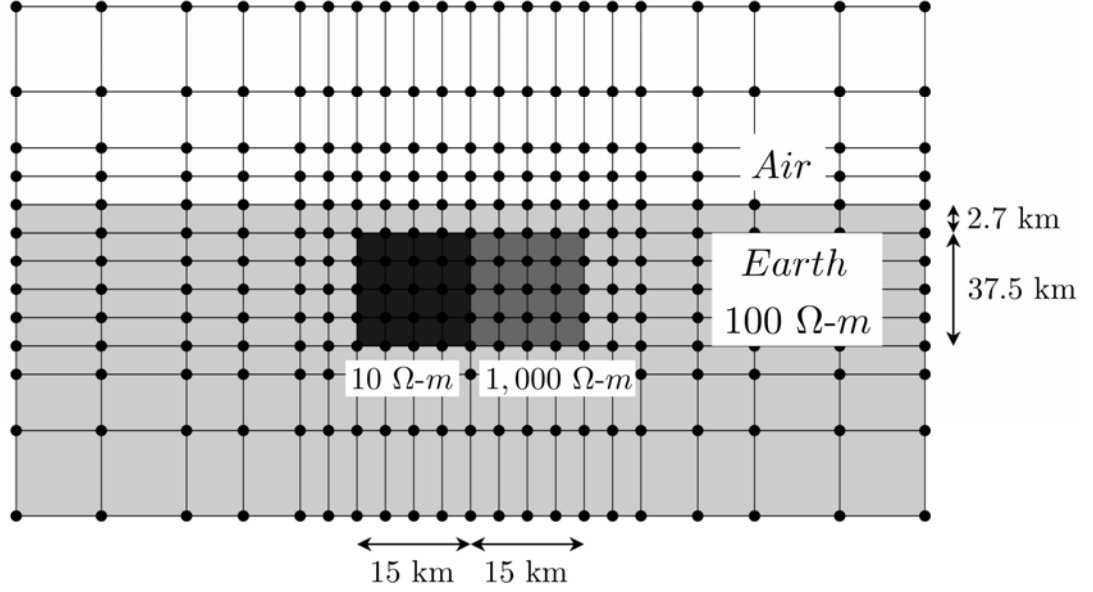


Figure 1. Model used to test the efficiency and accuracy of the modified hierarchical domain decomposition method. The model consists of two resistivity contrast blocks buried in a $100 \Omega-m$ half-space. The left and right blocks are $10 \Omega-m$ and $1,000 \Omega-m$, respectively. This model is discretized into three finite difference meshes: 40×120 , 80×240 and 120×360 and are used in the numerical experiment section. Discretization shown in this figure is merely an example to illustrate that the unknown fields are defined on the nodes (black dots).

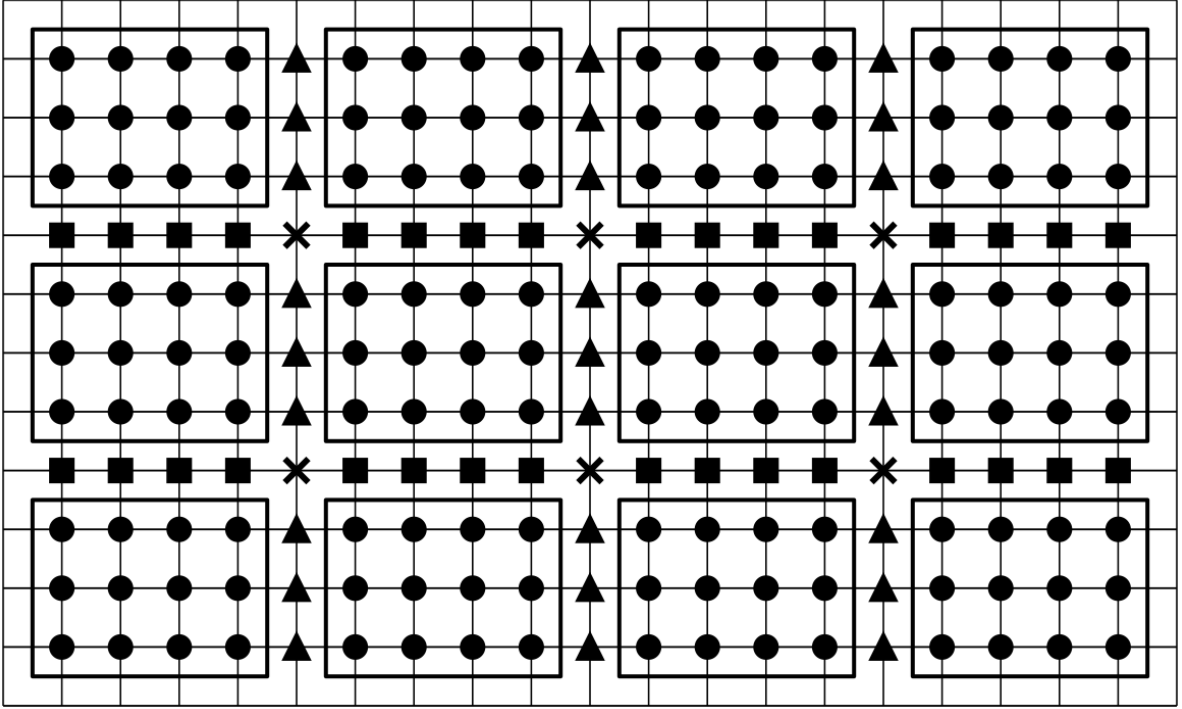


Figure 2. Example mesh of Figure 1 is uniformly redrawn, and subdivided into 3×4 subdomains as an illustration here. The interiors inside each subdomain are drawn with solid circle ($\bullet$). The horizontal and vertical interfaces between subdomains are shown with solid rectangle ($\blacksquare$) and solid triangle ($\blacktriangle$), respectively. The intersections from four subdomains are plotted with solid cross ($\times$).

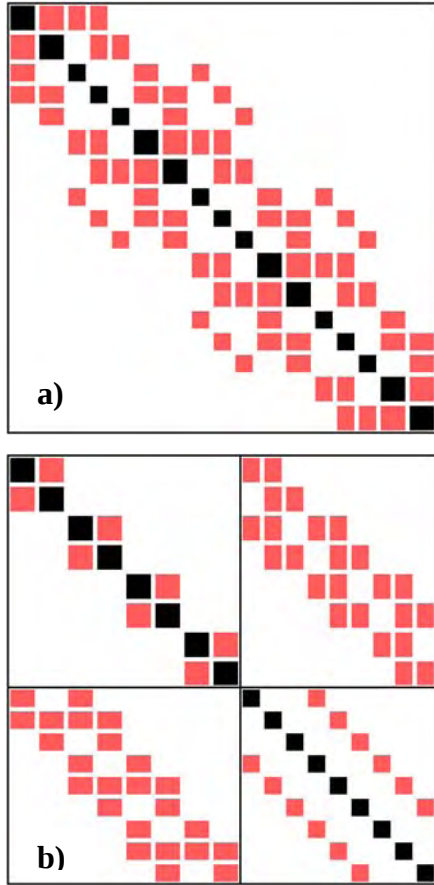


Figure 3. (a) Sparsity pattern of the Schur complement matrix S (equation 5) of the non-modified hierarchical domain decomposition. (b) Sparsity pattern of the Schur complement interface systems (S_{HH} , S_{HV} , S_{VH} and S_{VV} in equation 12) of the modified hierarchical domain decomposition.

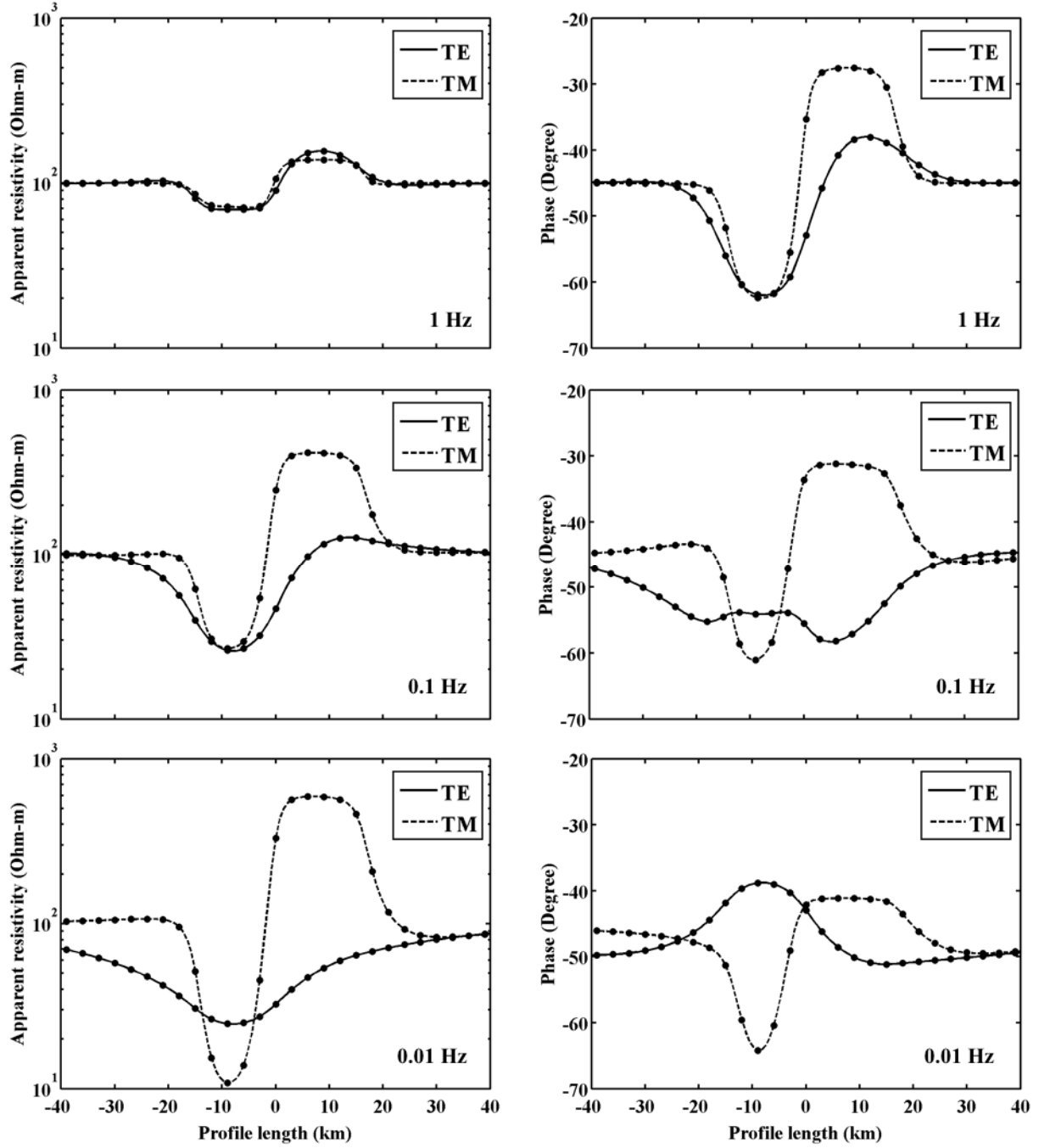


Figure 4. Apparent resistivities (Ohm-m) and phases (degree) of TM and TE modes from three different frequencies (1 Hz, 0.1 Hz and 0.01 Hz) across the model in Figure 1. Dots are from MHDD2D. Solid and dash lines are from TM and TE of FD2D, respectively. The differences of both responses from both methods are in the round-off level. This validates our MHDD2D code.

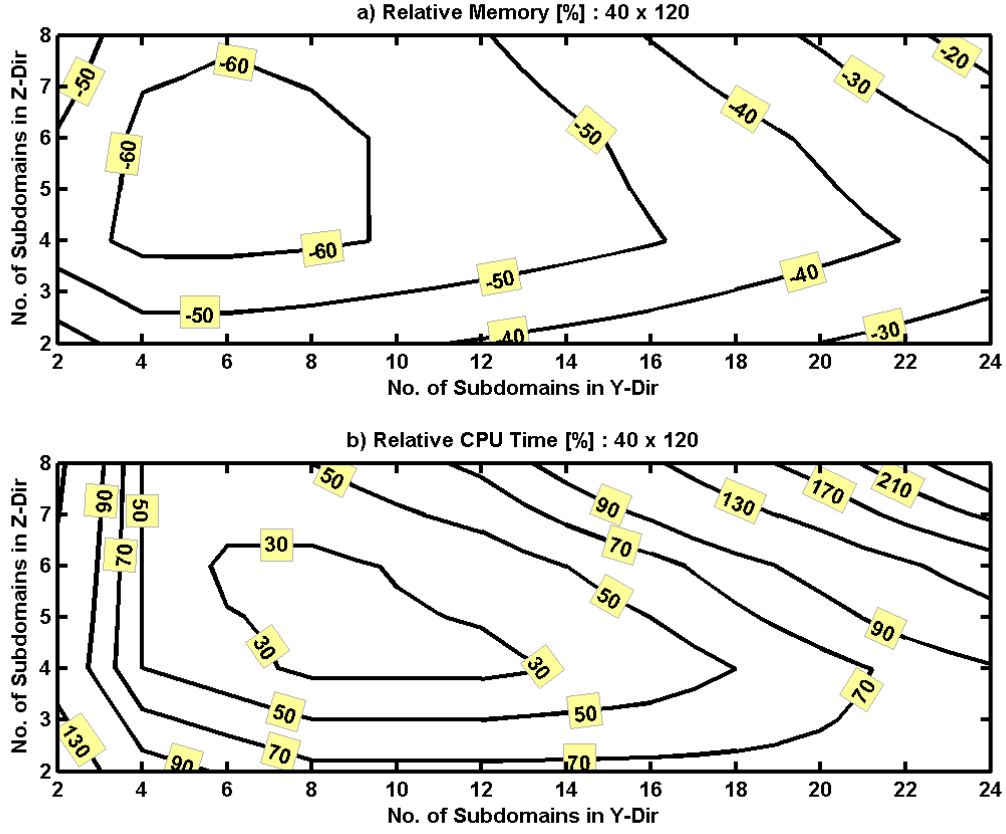


Figure 5. (a) Relative memory usage (in percent) and (b) relative CPU time (in percent) of MHDD2D to FD2D from several combinations of subdomains running on a 40×120 mesh. MHDD2D is more efficient than FD2D where larger negative percentage is presented, and less efficient where larger positive percentage.

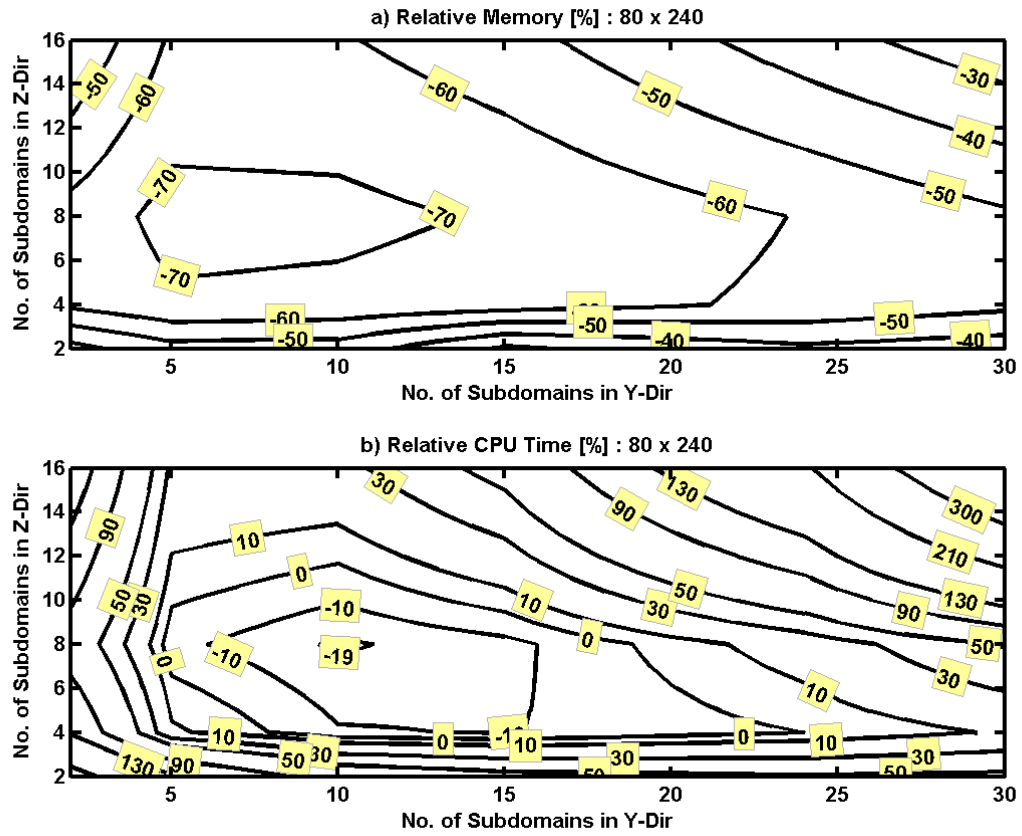


Figure 6. Same captions as in Figure 5 but for 80×240 mesh.

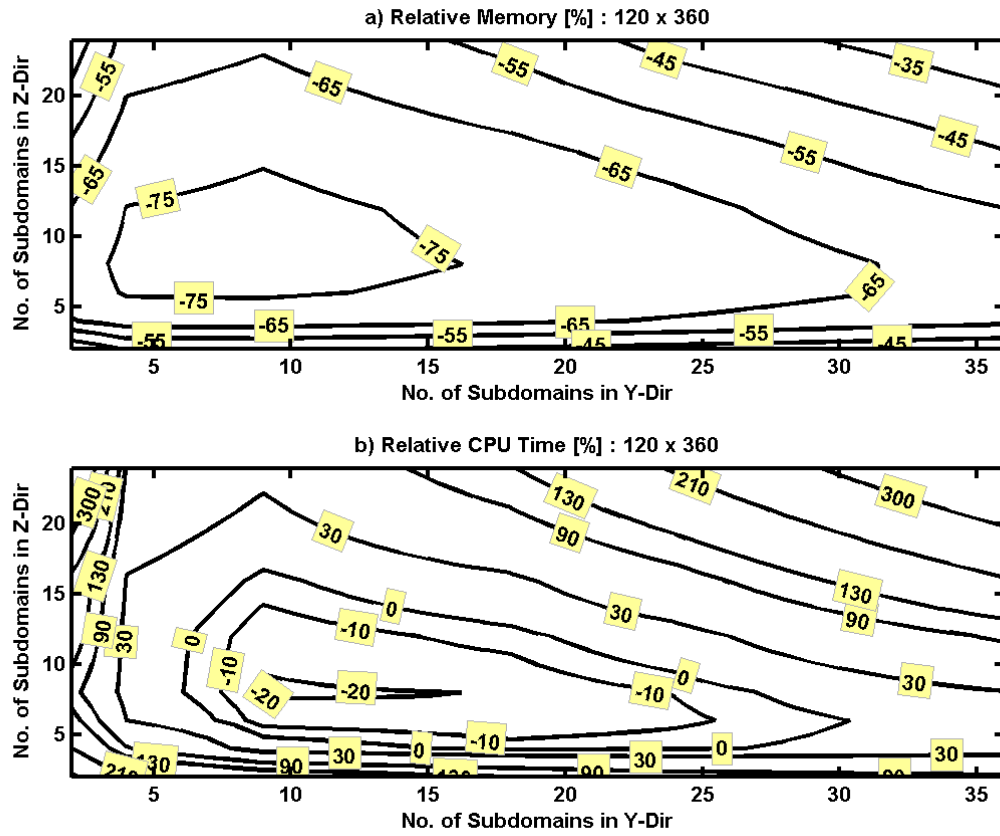


Figure 7. Same captions as in Figure 5 but for 120×360 mesh.

$p \times q$ subdomains	l/r (l)	n	m_h	m_v	m
3×3	4641 (41769)	4	714	234	948
10×9	429 (38610)	72	3159	880	4039
20×30	55 (33000)	551	6270	2900	9170

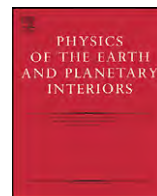
Table 1. Numbers of interiors per subdomain (l/r where l is total of interiors and $r = p \times q$), intersections (n), horizontal interfaces (m_h), vertical interfaces (m_v) and all interfaces (m) for three different numbers of subdomains running on a 120×360 mesh (Figure 7).

Center of optimized region	l/r (l)	n	m_h	m_v	m
5×6 subdomains on 40×120 mesh	133 (3990)	20	456	175	631
8×8 subdomains on 80×240 mesh	261 (16704)	49	1624	504	2128
10×9 subdomains on 120×360 mesh	429 (38610)	72	3159	880	4039

Table 2. Numbers of interiors per subdomain (l/r where l is total of interiors and $r = p \times q$), intersections (n), horizontal interfaces (m_h), vertical interfaces (m_v) and all interfaces (m) for 5×6 subdomains on 40×120 mesh (Figure 5), 8×8 subdomains on 80×240 mesh (Figure 6), and 10×9 subdomains on 120×360 mesh (Figure 7), respectively. These subdomains represent the center of optimized regions.

ภาคผนวก ค. Reprint

Siripunvaraporn W. and G. Egbert, 2009, WSINV3DMT: Vertical Magnetic Field Transfer Function Inversion and Parallel Implementation, *Physics of the Earth and Planetary Interiors* 173 (3-4), pp. 317-329



WSINV3DMT: Vertical magnetic field transfer function inversion and parallel implementation

Weerachai Siripunvaraporn^{a,*}, Gary Egbert^b

^a Department of Physics, Faculty of Science, Mahidol University, Rama VI Rd., Rachatawee, Bangkok 10400, Thailand

^b College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:

Received 15 September 2008

Received in revised form 21 January 2009

Accepted 25 January 2009

Keywords:

Magnetotellurics

Vertical magnetic transfer function

Data space method

3-D inversion

Occam's inversion

ABSTRACT

We describe two extensions to the three-dimensional magnetotelluric inversion program WSINV3DMT (Siripunvaraporn, W., Egbert, G., Lenbury, Y., Uyeshima, M., 2005, Three-dimensional magnetotelluric inversion: data-space method. *Phys. Earth Planet. Interiors* 150, 3–14), including modifications to allow inversion of the vertical magnetic transfer functions (VTFs), and parallelization of the code. The parallel implementation, which is most appropriate for small clusters, uses MPI to distribute forward solutions for different frequencies, as well as some linear algebraic computations, over multiple processors. In addition to reducing run times, the parallelization reduces memory requirements by distributing storage of the sensitivity matrix. Both new features are tested on synthetic and real datasets, revealing nearly linear speedup for a small number of processors (up to 8). Experiments on synthetic examples show that the horizontal position and lateral conductivity contrasts of anomalies can be recovered by inverting VTFs alone. However, vertical positions and absolute amplitudes are not well constrained unless an accurate host resistivity is imposed *a priori*. On very simple synthetic models including VTFs in a joint inversion had little impact on the inverse solution computed with impedances alone. However, in experiments with real data, inverse solutions obtained from joint inversion of VTF and impedances, and from impedances alone, differed in important ways, suggesting that for structures with more realistic levels of complexity the VTFs will in general provide useful additional constraints.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

WSINV3DMT (Siripunvaraporn et al., 2005) has been developed to invert Magnetotelluric (MT) impedance tensor components for three-dimensional (3-D) Earth conductivity. It was made freely available to the MT research community in 2006 and has since become one of the standard tools for 3-D inversion and interpretation (e.g., Tuncer et al., 2006; Heise et al., 2008; among others). The inversion algorithm used closely follows the two-dimensional (2-D) data space Occam's inversion of Siripunvaraporn and Egbert (2000) which has also been widely used for 2-D interpretation (e.g., Pous et al., 2002; Oskooi and and Pedersen, 2005; Toh et al., 2006; among others). Here we describe extensions to this code, which we illustrate with tests on synthetic and real data.

We first briefly summarize WSINV3DMT; see Siripunvaraporn et al. (2005) for more technical details. The algorithm used is based on the classic Occam's inversion introduced by Constable et al. (1987) for the one-dimensional (1-D) MT and DC resistivity sounding problems. The Occam inversion seeks a minimum structure

model (as defined by some model norm which penalizes roughness) subject to an appropriate fit to the data. The minimization is accomplished with a modified Gauss–Newton algorithm, in which the regularization parameter (which controls the tradeoff between model roughness and data fit) is also used for step length control (Parker, 1994). The main advantages of the Occam approach are its stability and robustness, and the fact that the scheme often converges to the desired misfit in a relatively small number of iterations (e.g., Siripunvaraporn and Egbert, 2000). Occam was extended to treat two-dimensional MT data by deGroot-Hedlin and Constable (1990), but for multi-dimensional inversion the originally proposed scheme can be computationally impractical, as the system of normal equations is explicitly formed and solved in the model space.

Siripunvaraporn and Egbert (2000) transformed the inverse problem into the data space (e.g., Parker, 1994). If the number of data (N) is small compared to the number of model parameters (M), as will typically be the case in 3-D, the data space variant requires a fraction of the CPU time and memory compared to a model space scheme. This data space Occam scheme forms the basis for the WSINV3DMT algorithm, which is summarized in Fig. 1.

The initial version of WSINV3DMT was only capable of inverting the impedance tensor $\mathbf{Z}$, the 2×2 complex frequency dependent

* Corresponding author. Tel.: +662 201 5770; fax: +662 354 7159.

E-mail address: scwsp@mahidol.ac.th (W. Siripunvaraporn).

Nomenclature

d	observed data
C_d	data error
m₀	initial and prior model
C_m	model covariance
m_k	model at <i>k</i> iteration
J_k	<i>N</i> × <i>M</i> sensitivity matrix forming from m_k
F[m_k]	forward responses of m_k
Γ_k	data space cross product matrix
R_k	representer for <i>k</i> iteration
λ	Lagrange multiplier
N_s	number of stations
N_m	number of modes
N_p	number of periods
N	number of data = <i>N_s</i> × <i>N_m</i> × <i>N_p</i>
M	number of model parameters

transfer function relating electric to magnetic fields

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix}. \quad (1)$$

The impedance tensor is frequently used by itself for 3-D conductivity imaging (e.g., Tuncer et al., 2006; Heise et al., 2008; Patro and Egbert, 2008). However, modern MT field practice typically includes measurement of vertical magnetic fields (particularly at long periods, where a tri-axial magnetometer is used), and thence computation of vertical field transfer functions (VTFs)

$$H_z = \begin{bmatrix} T_{zx} & T_{zy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix}. \quad (2)$$

The vertical magnetic field is only produced when there are lateral or horizontal variations of conductivity. Researchers have often used VTFs in the form of induction vectors (Parkinson, 1959) to indicate or point to the source of conductivity anomalies and to establish or verify geoelectric strike directions (e.g., Bedrosian et al., 2004; Uyeshima et al., 2005; Tuncer et al., 2006). A number of 2-D inversion codes (e.g., REBOCC of Siripunvaraporn and Egbert, 2000; and NLCG of Rodi and Mackie, 2001) allow inversion of VTFs (or “Tipper”), and these are often included along with TE and TM impedances in 2-D interpretations of MT profile data (e.g., Wannamaker et al., 1989; Wannamaker et al., 2008). Berdichevsky et al. (2003) studied VTFs using analytical and modeling studies, and concluded that inclusion of these additional induction transfer functions can substantially improve the reliability of geoelectrical models, because they are not affected as strongly by local distortion as the impedance tensor is.

Here, we describe the implementation of VTF inversion for the WSINV3DMT inversion code, and apply this to inversion of real and synthetic datasets. In addition, we describe implementation of a parallel version of WSINV3DMT, using MPI and parallelizing over frequencies to help reduce program execution times, which can be quite long for realistic modern datasets (e.g., Patro and Egbert, 2008).

The paper is organized as follows. First, we summarize the modifications to WSINV3D, for the most part omitting technical details. Next, we illustrate and test the new features on the same synthetic datasets previously used in Siripunvaraporn et al. (2005). Here we illustrate the speedup obtained with the parallelization, and explore the effectiveness of VTF data for recovering conductivity structures, alone, and in conjunction with impedance data. We then test the VTF inversion on the EXTECH dataset (Tuncer et al., 2006), comparing inverted models from only VTF data, from

only impedance data, and from a joint inversion of both data types.

2. Implementation of WSINV3DMT to include the vertical magnetic transfer function

There are only two major modifications to the WSINV3DMT codes required to allow inversion of VTFs: adding the VTF computation to the forward modeling routine, and the corresponding modifications for the sensitivities of the real and imaginary parts of the VTFs.

In WSINV3DMT, the electric fields are calculated by solving the second order Maxwell's equation using a staggered grid finite difference numerical scheme (Siripunvaraporn et al., 2002). Magnetic field components can then be computed (on grid cell faces) from Faraday's law $\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}$, and interpolated to the observation locations, which in the modified version of WSINV3D can be at any location on the surface. In order to compute the impedance tensor **Z** the forward equations are solved for two polarizations, and **Z** is calculated from the combination of horizontal electric and magnetic fields from both polarizations, as described in Siripunvaraporn et al. (2005).

The only modification required for the VTF is that the vertical magnetic field must also be computed at the observation location. As for the horizontal magnetic components, this is accomplished using Faraday's law, taking the curl of the horizontal **E** components on the model air–Earth interface, and interpolating the result (defined at cell centers) to the observation locations. Then, similarly to the impedance tensor, the vertical and horizontal magnetic fields computed from the solutions for both polarizations are combined to form the vertical magnetic field transfer function **T**,

$$\begin{bmatrix} H_z^1 & H_z^2 \end{bmatrix} = \begin{bmatrix} T_{zx} & T_{zy} \end{bmatrix} \begin{bmatrix} H_x^1 & H_x^2 \\ H_y^1 & H_y^2 \end{bmatrix} \quad (3)$$

Here **H_z¹** and **H_z²** are the *z*-component of magnetic fields for the **E_x–H_y** and **E_y–H_x** polarizations, respectively, and similarly for other field components. For a joint inversion with impedance tensor, computing the vertical magnetic transfer function does not require any extra forward modeling calls, as all transfer functions are computed from the same solutions.

The sensitivity calculation for VTFs is essentially identical to that used for impedances, which is based on the reciprocity approach described in Rodi (1976), Newman and Alumbaugh (2000), and Siripunvaraporn et al. (2005). Briefly, the linearized data functional, which is represented by linear combinations of electric field solution components on cell edges surrounding the observation point, is used to force the adjoint equation, and the result is mapped to perturbations in the model parameter, as described in Siripunvaraporn et al. (2005). Only the first step requires modification, with the coefficients for the linearized functionals for **T_{zx}** and **T_{zy}** replacing those for **Z_{xx}** and **Z_{xy}**. Details of this modification are straightforward, and are omitted here.

3. Parallel implementation with MPI

A major challenge in using WSINV3DMT, or for that matter, any 3-D MT inversion code, is that the program is very time consuming, especially when run with the sort of large dataset (and model domain) that justifies a 3-D interpretation. Run times exceeding a full month (on a single processor desktop computer, for the full inversion process, including multiple iterations of the outer loop of Fig. 1) have been reported when WSINV3D has been applied to even modest 3-D MT datasets (e.g., Patro and Egbert, 2008). These long run times primarily reflect the need for many forward modeling calls, each of which requires iterative

Serial WSINV3DMT algorithm:

- 1) Solve forward problem and compute misfit from model $\mathbf{m}_0$
- 2) Start WSINV3DMT outer loop iteration k :
 - 2.1) For $i = 1$ to $N_s * N_m * N_p$
 - Call forward solver to form $\mathbf{J}_{ki}$ sensitivity for data i
 - End
 - 2.2) Compute $\hat{\mathbf{d}}_k = \mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_k - \mathbf{m}_0)$
 - 2.3) Compute $\mathbf{\Gamma}_k = \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}$
 - 2.4) For various values of λ s
 - 2.4.1) Compute representer matrix $\mathbf{R}_k = [\lambda \mathbf{I} + \mathbf{\Gamma}_k]$
 - 2.4.2) Use Cholesky decomposition to solve

$$\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k \mathbf{C}_d^{-1/2} \mathbf{R}_k^{-1} \mathbf{C}_d^{-1/2} \hat{\mathbf{d}}_k$$
 - 2.4.3) Solve forward problem and Compute misfit from model $\mathbf{m}_{k+1}$
 - 2.4.4) Phase I :
 - Compare misfit from different λ s to seek for minimum misfit
 - Phase II:
 - Compare norm from different λ s to seek minimum norm
 - End
 - 2.5) Exit when misfit less than desired level with minimum norm

Fig. 1. Pseudo-code for serial WSINV3DMT (after Siripunvaraporn and Egbert, 2007).

solution of the large sparse linear system arising from discretization of Maxwell's equations. WSINV3D was developed as a serial code, to run on a single processor. An obvious way to speed up execution is to parallelize the code, and make use of the multiple processors which are increasingly common even in desktop computers.

There are several ways to redesign the codes to run on parallel system, and the most appropriate approach will depend on system architecture. For supercomputers or large clusters to make effective use of hundreds of processors it would be necessary to rewrite parts of the forward solver—e.g., parallelizing the iterative solver and preconditioner (e.g., Newman and Alumbaugh, 2000), or domain decomposition. Here, we consider a parallelization approach appropriate to small systems with a few to several tens of processors. Such small clusters and multi-processor workstations are now readily affordable and more widely available than supercomputers. To adapt WSINV3DMT for this class of systems, we parallelize over frequencies, adding calls to MPI (Message Passing Interface) library routines to the existing codes. In this way, we do not have to alter the core forward modeling and sensitivity calculation routines in any way. The parallel algorithm is summarized in Fig. 2.

Forward modeling and sensitivity calculations for each period are sent to one processor (Steps 2.1 and 2.2 in Fig. 2). If there are fewer processors than periods, each processor performs calculations for more than one period. With this simple parallelization, which requires minimal inter-processor communication, the computational time should be theoretically reduced by a factor P , the number of processors available. This parallel implementation also distributes storage of the sensitivity matrix over the available nodes. The $N \times M$ sensitivity matrix $\mathbf{J}$ requires $8NM$ bytes (in double precision), and the need to store this in RAM limits the size of datasets and model grids that can be practically treated. With the parallelization, memory required on each node is reduced to about two times $8NM/P$ (including temporary storage

for cross product computations), allowing WSINV3D to be run for larger models grids and datasets.

With the sensitivities distributed over processors, formation of the cross product matrix $\mathbf{\Gamma} = \mathbf{J} \mathbf{C}_m^{-1} \mathbf{J}^T$ also requires MPI calls. We have implemented this in a fairly simple way, breaking $\mathbf{\Gamma}$ into P^2 blocks to be computed on the P processors (Step 2.3 in Fig. 2). Diagonal blocks $\mathbf{\Gamma}_{ii}$ are computed on the local processor where the corresponding block $\mathbf{J}_i$ of the sensitivity matrix (corresponding to one or more frequencies) is computed and stored. The off-diagonal blocks ($\mathbf{\Gamma}_{ij}$) can only be formed by sharing blocks of $\mathbf{J}$ between nodes. Since $\mathbf{\Gamma}$ is symmetric, only upper off-diagonal blocks ($j > i$) need be formed. On step k block $\mathbf{J}_j$, where $j = \text{mod}(i + k, P)$ is sent to node i to compute $\mathbf{\Gamma}_{ij}$ where this block is stored. With this simple scheme the load is balanced and the number of steps required is approximately $(N_p + 1)/2$. Although computing the cross products requires significant communication time to share sensitivities between nodes, it can still significantly reduce the total computing time required to form $\mathbf{\Gamma}$ compared to single node processing.

In the data space Occam scheme used by WSINV3D the system of normal equations (Eq. (6) in Siripunvaraporn et al., 2005) must be solved for a series of trial values of the regularization parameter (about 3–7 from our experience) to find the optimal (in terms of data misfit and model norm) model parameter update. In the serial version of WSINV3D these dense systems are solved by Cholesky decomposition (Step 2.4.2 in Fig. 1). Parallel Cholesky decomposition subroutines are available (e.g., Choi and Moon, 1997), but these are generally tailored to a specific parallel architecture and are not easily adapted. To develop code that will be portable, and reasonably efficient on a generic multi-processor system, we have thus pursued a different strategy, using the easily parallelized preconditioned conjugate gradient (PCG) algorithm to solve the normal equations (Step 2.4.1.2 in Fig. 2). The major computation in the

Parallel WSINV3DMT algorithm for P -cluster PCs nodes:

- 0) Parent node distributes data to other nodes. Each node would then takes care a computational load of N_p/P data.
- 1) Each node separately solve forward problem and compute misfit from model $\mathbf{m}_0$
- 2) Start parallel WSINV3DMT outer loop iteration k :
 - 2.1) In each node,
 - For $i = 1$ to $N_s * N_m$; Call forward solver to form local $\mathbf{J}_{ki}$ sensitivity for data i
 - 2.2) Each node compute $\mathbf{d}_k = \mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_k - \mathbf{m}_0)$ separately.
 - 2.3) To compute $\mathbf{\Gamma}_k = \mathbf{C}_d^{-1/2} \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T \mathbf{C}_d^{-1/2}$,
 - 2.3.1) each node first computing local or diagonal $\mathbf{\Gamma}_{ii}$ from their local $\mathbf{J}_{ki}$
 - 2.3.2) each node cyclically sending their local $\mathbf{J}_{ki}$ to others nodes to compute the off-diagonal $\mathbf{\Gamma}_{ij}$.
 - 2.4) For various values of λ s
 - 2.4.1) If N is small (single node process)
 - 2.4.1.1) Sending $\mathbf{\Gamma}_k$ from local nodes to parent nodes to compute global representer matrix $\mathbf{R}_k = [\lambda \mathbf{I} + \mathbf{\Gamma}_k]$
 - 2.4.1.2) On parent node, applying Cholesky decomposition to solve $\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k \mathbf{C}_d^{-1/2} \mathbf{R}_k^{-1} \mathbf{C}_d^{-1/2} \mathbf{d}_k$
 - else if N is large (parallel process)
 - 2.4.1.3) Local $\mathbf{R}_k = [\lambda \mathbf{I} + \mathbf{\Gamma}_k]$ is formed in each node
 - 2.4.1.4) Parallel iterative solver (PCG) is applied to solve $\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k \mathbf{C}_d^{-1/2} \mathbf{R}_k^{-1} \mathbf{C}_d^{-1/2} \mathbf{d}_k$
 - 2.4.2) Each node separately solve forward problem and compute misfit from model $\mathbf{m}_{k+1}$
 - 2.4.3) On parent node
 - Phase I: Compare misfit from different λ s to seek for minimum misfit
 - Phase II: Compare norm from different λ s to seek minimum norm
- 2.5) Exit when misfit less than desired level with minimum norm

End parallel WSINV3DMT outer loop iteration

Fig. 2. Pseudo-code for parallel WSINV3DMT for cluster PCs system.

PCG algorithm is matrix–vector multiplication. This is readily parallelized by dividing the vectors and matrix into blocks, spreading computations for individual blocks over processors, and then gathering the results back to the master node. To simplify the algorithm we have distributed the full matrix to all computational nodes.

The preconditioner, based on the diagonals of the coefficient matrix, is also trivially parallelized. Because the coefficient matrices are dense, the parallel PCG scheme may not be efficient when N is small, since communication and other overhead may exceed the serial computational time. For smaller N , we therefore retain the option of solving the normal equations with a serial Cholesky decomposition, after all blocks $\mathbf{\Gamma}_{ij}$ are sent back to the parent node. The optimal choice of solution scheme (parallel or serial) for a specific value of N will depend on the cluster architecture. We give examples below where each approach is more efficient.

Once the new model $\mathbf{m}_{k+1}$ is obtained, the parallelized forward solver is called to compute the responses of each period, with the results gathered to the parent node to compute misfits (Step 2.4.2 in Fig. 2). All steps are repeated until an acceptable misfit and norm are achieved

4. Synthetic data examples

To illustrate the efficiency of the parallelized WSINV3D, and the effectiveness of the VTF inversion, we first consider inversion of synthetic datasets, revisiting the two synthetic examples previously used by Siripunvaraporn et al. (2005), reproduced in Fig. 3. The results of these tests are consistent with those obtained for other synthetic examples. Our basic test configuration is the two-block model (Fig. 3a) consisting of two anomalies, 1 Ω m and 100 Ω m located next to each other within a 10 Ω m host. The spatially homogeneous layer, which extends from the surface to 10 km depth, is underlain by a 100 Ω m half space. To test the efficiency of our parallel codes, and the VTF inversion, we generated VTF and impedance data at 16 periods (from 0.1 to 1000 s) for a total of 40 sites in a regular grid, as illustrated in Fig. 3a. Gaussian noise (5% of the data magnitude) was added to the generated data. The inversions for this case are performed on a $21 \times 28 \times 21$ (+7 air layers) mesh. The second model consists of a single conductive block (1 Ω m) buried in a 100 Ω m half-space (Fig. 3b), and responses were computed at 16 periods for 36 sites (Fig. 3b). The inversions

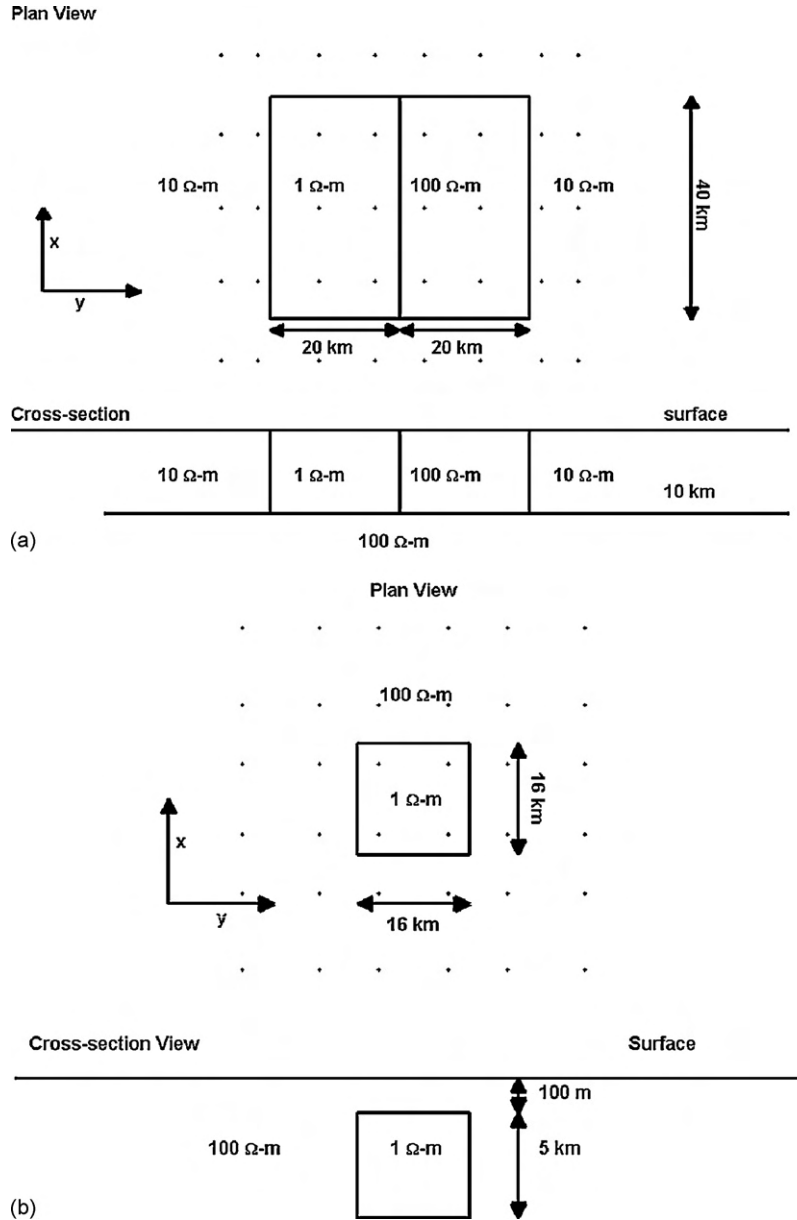


Fig. 3. Two synthetic models used to test our inversion. (a) Two-block synthetic model and (b) a single conductive block model. The solid dots indicate the observation sites. The cross-section view in the lower panel is a profile cutting across the middle of the model in the upper panel, and is not to scale (after Siripunvaraporn et al., 2005).

for the second case are performed on a $28 \times 28 \times 21$ (+7 air layers) mesh.

We first demonstrate the efficiency of the parallel version of WSINV3D, using both VTF and joint VTF/impedance datasets for tests. We then consider the effectiveness of VTF data for recovering conductivity variations, both alone, and in conjunction with impedances.

4.1. Parallel efficiency

We tested WSINV3DMT by running on 1, 4, 8 and 16 nodes for the first synthetic test case (Fig. 3a), with the 16 periods divided evenly among nodes (e.g., with 4 nodes, each solves for 4 periods). Tests were conducted on a small PC-clusters and a supercomputer (SGI Altix 4700) at the Earthquake Research Institute, University of Tokyo. To quantify efficiency of the parallel code, we display the speedup, defined as $S = T_1/T_P$, where T_1 is the execution time of the sequential WSINV3DMT algorithm and T_P is the execution time

of the parallel version, run on P processors. The idealized maximum speedup is P . However, due to computational overhead, the need for some computations to be performed only on the master node, and the time required to exchange information between nodes, S will always be less than P . Fig. 4 displays speedup versus the number of nodes. Inversions of all data (i.e., VTF + impedance, $N = 40 \times 12 \times 16 = 7680$) are plotted with solid lines. Inversions of the VTF only dataset ($N = 40 \times 4 \times 16 = 2560$, or one third the size of the joint inversion dataset) are plotted as dashed lines. We also compare speedups achieved with the two approaches for solving the normal equations: speedups obtained with the single processor Cholesky decomposition are plotted as solid symbols, while those obtained with the parallel PCG algorithm are plotted as open symbols.

For the inversion of the VTF dataset for this very small test problem, actual (wall clock) run times were about 186 min on a single node machine, 82 min on 4 nodes, 46 min on 8 nodes and 34 min on 16 nodes, resulting in speedups of about 2.2 for 4 nodes, 4 for 8

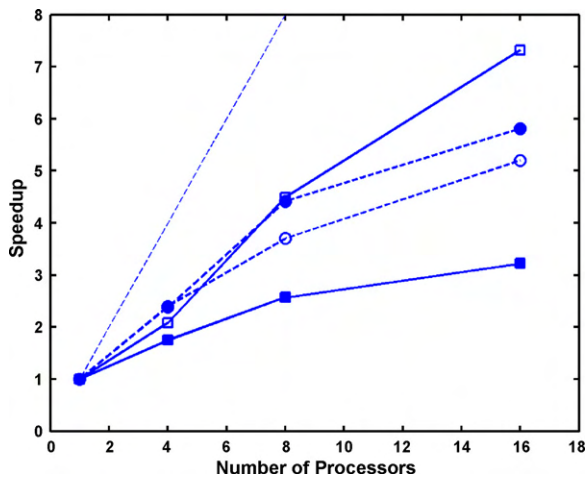


Fig. 4. Speedup versus the number of processors or nodes. Solid lines are the speedups from inversion using both VTF and impedance data ($N=7680$). Dashed lines are the speedups from inversion using only VTF data ($N=2560$). Results for the scheme which solves the normal equations by Cholesky decomposition on a single node (step 2.4.1.2 of Fig. 2) are plotted with solid symbols. The corresponding results with the parallel PCG solver (step 2.4.1.4 of Fig. 2) are plotted with open symbols. The thin-dashed line of slope one gives the ideal perfect speedup.

nodes and 5.4 for 16 nodes. Thus, as the number of nodes increases, the relative efficiency of additional nodes decreases. One reason for this is that the run time of the iterative forward modeling routine depends on the period of the data. Shorter periods typically require a larger number of iterations for convergence, and hence longer run times. Thus, some nodes are usually idle waiting for modeling computations to complete on other nodes, before moving on to the next step in the inversion. With fewer nodes some of the frequency-to-frequency variations average out, resulting in better balance.

Efficiencies are somewhat lower for the larger joint VTF/impedance dataset, when the serial Cholesky decomposition solver is used (solid line with solid square symbols of Fig. 4). Now the speedups are about 1.8, 2.6 and 3.2 for 4, 8 and 16 nodes, respectively, almost 50% below those achieved for the VTF only inversion. However, solving the normal equations with the parallel PCG solver (solid line with open square symbols in Fig. 4) significantly improves performance, increasing S to approximately 2, 4.5 and 7.3 for the three cases considered. In the VTF only case, where N is significantly smaller, both methods for solving the normal equations have similar performance (dashed lines in Fig. 4), and indeed the speedup is slightly greater when the single node Cholesky decomposition is used.

The difference between the two cases is readily understood. Operation counts for Cholesky decomposition scale as N^3 so computation times for the serial Cholesky decomposition in the all data case ($N=7680$) are expected to be about 27 times greater than for the VTF only case ($N=2560$). Other computational steps scale better with increasing N . For fixed model parameter size, total operation counts for the sensitivity calculations increase linearly in N , and formation of the cross product matrices increases as N^2 . Thus, as the size of the dataset increases, run times required for the serial Cholesky decomposition step become increasingly significant, and at large enough N this step will control the overall runtime. Operation counts for a single iteration in the parallel PCG scheme scale as N^2 , but overall runtimes will also depend on the number of iterations required. Although this should increase with N also, the dependence is weak, and so PCG becomes increasingly advantageous as N increases, particularly since computations for the PCG scheme can be distributed over the P processors.

The number of iterations for PCG also depends on the relative tolerance for the residual ($=||Ax - b||/||b||$) used to define convergence. We find that a tolerance of 10^{-4} results in models that are essentially identical to those obtained with the Cholesky decomposition technique. The number of iterations, and hence the run time of the parallel PCG scheme also depends on the condition number of the normal equations. For large values of the Lagrange multiplier (corresponding to a smoother model) the condition number is smaller, and the parallel solver thus converges in a small number of iterations. In contrast, when the Lagrange multiplier is very small (rough model) the parallel solver can require considerably more iterations, and solution times can exceed those for the serial Cholesky decomposition scheme. This occurred occasionally in our tests with the larger VTF/impedance dataset, but overall performance using the parallel PCG solver was much better when N is large enough.

We will not attempt to quantify more precisely how large N must be before the parallel approach to normal equation solution would be preferred. This will depend on the cluster architecture, especially on the sort of inter-processor communication used, since the parallel PCG solver requires substantial sharing of data.

In addition to reducing computational times, the parallel version also reduces the need for a large amount of memory on a single computer. Even for the small joint VTF/impedance inversion test example, about 1.5 GBytes are required for the representer and sensitivity matrices. In the parallel implementation, the required memory may be distributed over all of the nodes used. For example, with 16 nodes, each would require only 0.090 GBytes for storing the sensitivity matrix and forming cross products, almost 13 times less than required by the serial code. If the whole representer matrix is stored on a single processor (for the Cholesky decomposition, or to reduce the communication time between nodes for PCG) about 0.4 Gb are required on each node, still only a quarter required for a serial version.

The exact time speedup and per-node memory reduction factors will depend to some extent on the problem size, both in terms of model grid dimensions, and number of data. For larger problems, such as the real data EXTECH example considered below, similar performance gains were attained. For these larger problems, however, a speedup by a factor of roughly 7 means a run time that was perhaps 2–3 weeks on a single node is now reduced to 2–3 days, making inversion of realistic datasets considerably more practical. The practical impact of distributing memory is even greater. Total storage required by WSINV3D for the EXTECH example described below (joint inversion of the full impedance and VTFs) is at least 30 Gb, making this impractical on almost any shared memory machine.

4.2. Vertical magnetic transfer function inversion

We next consider the effectiveness of WSINV3DMT at correctly recovering resistivity when only VTF data are available. Because in practice one would not know *a priori* the correct background resistivity, we run the inversion using several prior (and starting) models. Inversion results for the synthetic VTF data from the test case of Fig. 3a are summarized in Figs. 5 and 6. Using a 50 Ω m half-space as a prior (this is intermediate between the true 10 Ω m upper layer background, and the 100 Ω m basement), inversion of VTF data reveals both the conductive body and the adjacent resistor, extending from near the surface to approximately 20 km depth. The calculated responses generated from the inverse solution of Fig. 5 fit the observed responses within 15% of the typical VTF amplitude (recall that 5% random noise was added to the synthetic data).

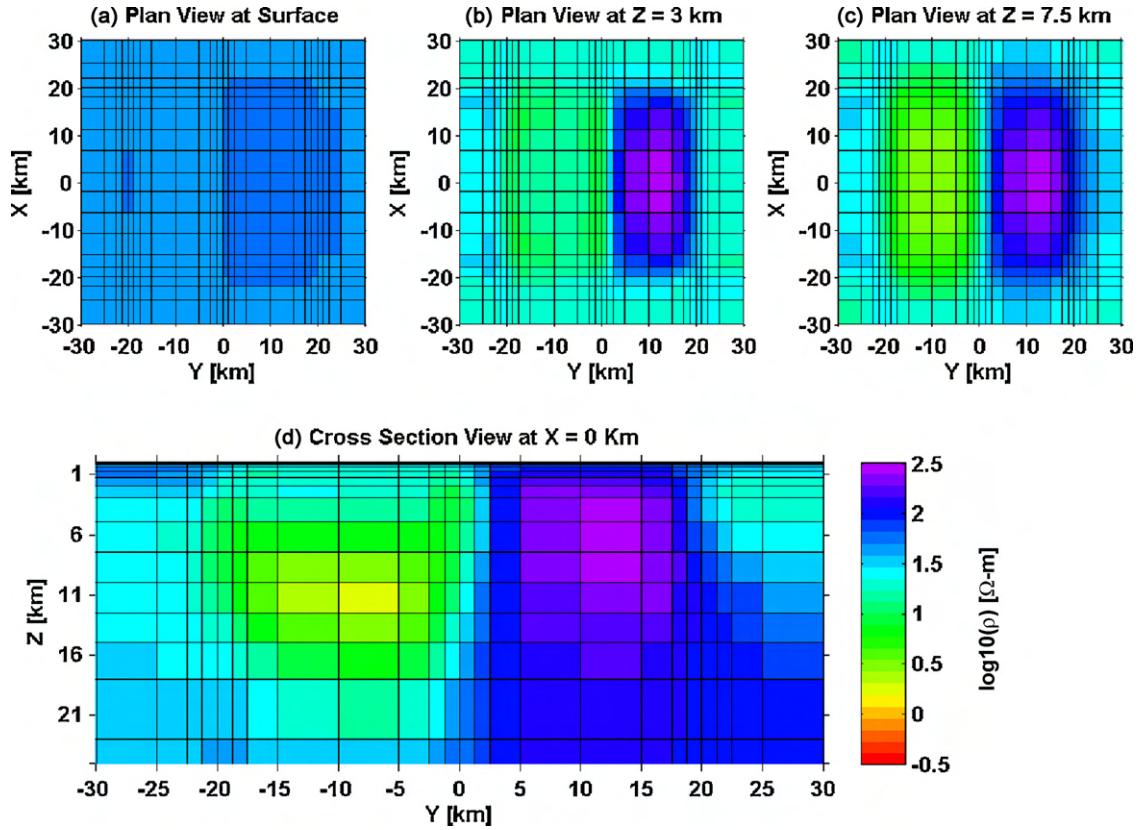


Fig. 5. An inverse solution from the VTF data alone after the 9th iterations with an RMS value of 1, fitting synthetic data generated from the model in Fig. 3a. The top panels (a)–(c) is a plan view at the surface, at 3 km and at 7.5 km depth, and the bottom panel (d) is a cross-section view cutting across the conductive block at $X = 0$ km. The solution is shown only in the central area around the anomalies, not for the full model domain.

Although both anomalies are detected in approximately the correct location, the true resistivities of Fig. 3a are not correctly estimated. However, calculating the average resistivity over the anomalous volumes we find for the inverse model of Fig. 5 an average resistivity of about $6.3 \Omega \text{ m}$ for the conductive anomaly, and of

about $453 \Omega \text{ m}$ for the resistive body, while the background resistivity of the inverse model was changed only slightly from the $50 \Omega \text{ m}$ prior. Computing the volume average resistivity ratios from left to right in Fig. 5d, we obtain values of $7.9 (=50/6.3)$, $72 (=453/6.3)$ and $9 (=453/50)$, compared to the actual ratios (Fig. 3a) of $10 (=10/1)$, 100

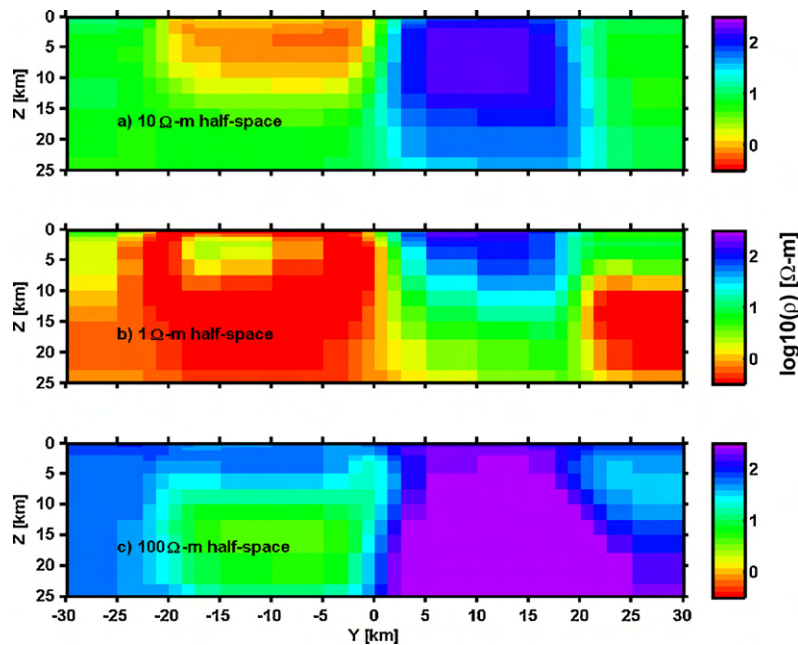


Fig. 6. Cross-sectional plots at $X = 0$ km (as in Fig. 5d) of the inverse solutions from VTF data alone, when the prior models are (a) $10 \Omega \text{ m}$ half-space, (b) $1 \Omega \text{ m}$ half-space and (c) $100 \Omega \text{ m}$ half-space.

(=100/1) and 10 (=100/10), respectively. The inversion thus results in roughly the correct structure, with approximately correct resistivity contrasts, but it does not recover the correct amplitude of either the background or the anomalies, or the actual depth extent of the anomalies.

To explore this issue further we ran the inversion on the same VTF dataset, using a range of values for the assumed half-space prior. Fig. 6 summarizes the results with cross-sectional plots of the inverse solutions at $X = 0$ km. When the prior model is the same as the correct background resistivity (i.e., a 10 Ω -m half-space in our example), the inversion quickly converges to the desired misfit within 4 iterations, even with error floors set to 5%. In this case, the inversion estimates the resistivity, and the depth extents, of the two anomalies quite well (Figs. 6a and 3a). However, the 100 Ω m basement resistivity (below 10 km depth in the synthetic test model of Fig. 3a) is not recovered—the prior resistivity of 10 Ω m remains unchanged at depth in the inverse solution. This again demonstrates that inversion of VTF data alone can only recover lateral resistivity contrasts, and is not effective at correcting resistivities, or their variations with depth.

Larger deviations of the prior model from the correct background result in even larger discrepancies in anomaly amplitudes and depths, but still generally allow the horizontal structure to be recovered. With a 1 Ω m half-space (Fig. 6b) data is fit to within 10%. Anomalies appear at very shallow depths (upper few km), with all features more conductive than their actual values. At greater depth, features with appropriate resistivity ratios are imaged, but the absolute levels are incorrectly estimated, and spurious structures appear. Using a 100 Ω m half-space as a prior, the VTF data can only be fit to within 20%. The basic structure is again recovered, but both anomalies are at greater depth (Fig. 6c) and have increased resistivity. The host resistivity is estimated to be slightly lower than the 100 Ω m starting value, but is still well above the correct value

of 10 Ω m. As in the other cases, the basement resistivity remains the same as the prior model.

All of these experiments suggest that when only VTF data are available, to achieve the target misfit and recover correct amplitudes and depths, the inversion must be started with a prior model that is close to the correct host resistivity. However, even starting far from the correct background model, anomalies are recovered with the correct horizontal location and dimensions. This result is not unexpected since the vertical magnetic fields are generated where there are lateral discontinuities, but are not inherently sensitive to the profile of vertical conductivity structure.

In addition, resistivities of anomalous bodies scale with the assumed prior background (Fig. 6), and resistivity contrasts (i.e., ratios) can be close to actual values, especially if the assumed background resistivity is not too far off. However, the VTFs provide little intrinsic constraint on contrasts in the vertical direction, including the location of the top or the bottom of the anomalies. The inversion only gets these properties of the anomalies correct if something close to the correct background is used (Fig. 6a).

Performing similar experiments to those summarized in Fig. 6, but using impedance tensor data shows that these inversions are much less sensitive to the assumed prior model. This is consistent with the basic physics, as the ratio of electric to magnetic fields is intrinsically related to the resistivity profile. In spite of the well-known uncertainties in depth and absolute resistivity level that may result from local static distortions, there is by now ample evidence (e.g., Tuncer et al., 2006; Unsworth et al., 2000) that, with proper care, MT impedances can yield reliable information about conductivity-depth profiles. The same does not appear to be true in practice with VTF data, although theoretical analysis of idealized models suggests otherwise (Berdichevsky et al., 2003).

The above results suggest that VTF data will be most useful as an adjunct to impedance data, which can provide the necessary con-

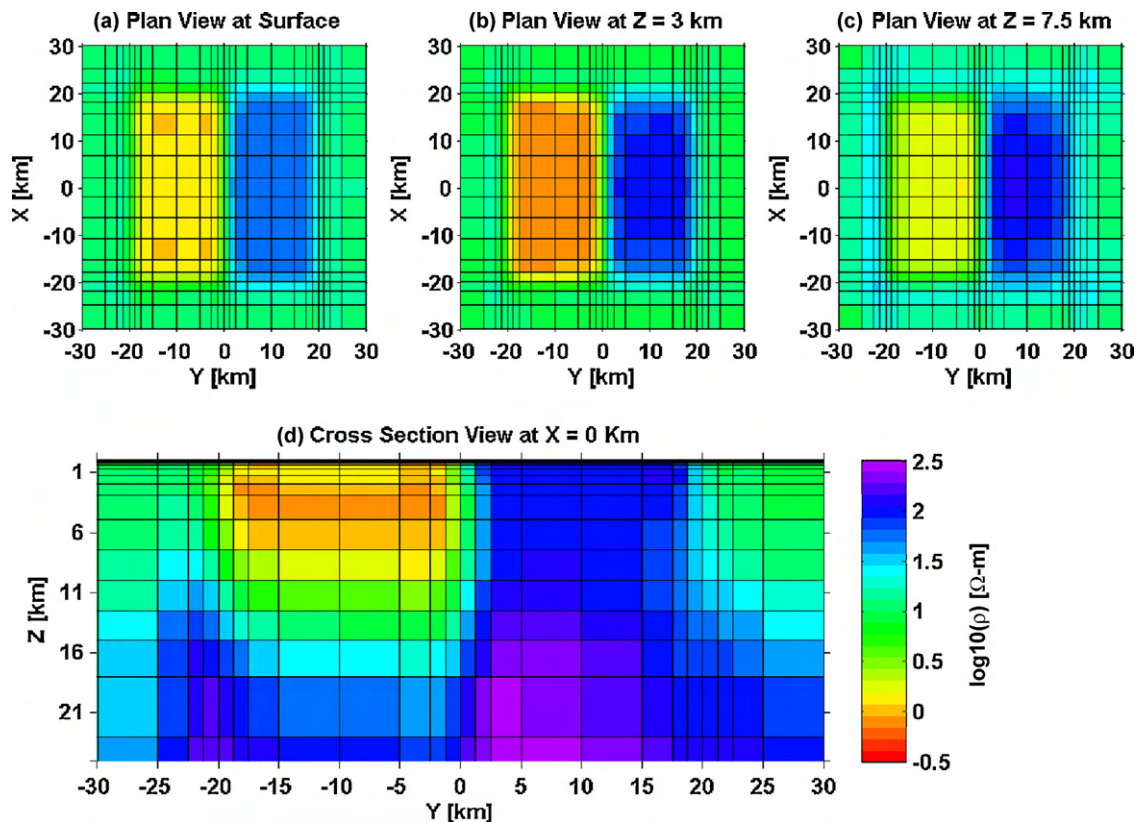


Fig. 7. Results from joint inversion of both VTF and impedance tensor data generated from the model in Fig. 3a. See caption of Fig. 4 for other details.

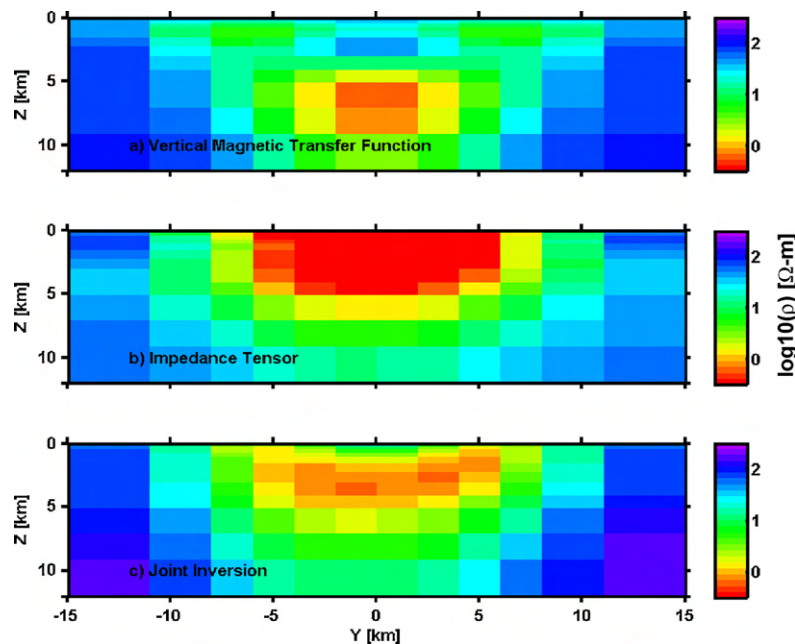


Fig. 8. Cross-sectional plots at $X=0$ km of the inverse solution from (a) fitting the vertical magnetic transfer function alone, (b) fitting the impedance tensor alone, (c) fitting both data types. The data is generated from the synthetic model in Fig. 3b.

straint on background resistivities. As a first example, we consider joint inversion of VTF and the impedance tensor data derived for the synthetic model of Fig. 3a. As above we again tried a range of priori/initial models. Although in general the impedance tensor data can adjust the resistivity background, we still had difficulties getting the joint inversion to converge to the desired 5% misfit level, especially with priori models that differ greatly from the correct background resistivities. In this and other examples, we found that to achieve the target misfit for both data types, it was necessary to first fit the impedances to a half-space model, to determine a prior model for the joint inversion. Even with this additional step, we typically found it necessary to use increased error floors for the VTF data (but not the impedances) to achieve a normalized RMS of one.

Not surprisingly, a $50\ \Omega\text{m}$ half-space (as in Figs. 5 and 6 of Siripunvaraporn et al., 2005) yields a good fit to the synthetic impedance data for case 1. With error floors set to 15% for VTF data and 5% for impedance tensor data, the joint inversion converged to the target misfit in 5 iterations. In the final iteration (Fig. 7) the two anomalies are recovered with essentially correct background resistivities. In fact, in comparison with the inverse model obtained from inversion of just the impedance data (Fig. 6 of Siripunvaraporn et al., 2005), there is little difference. Clearly, the relatively simple structures in this synthetic example are well enough constrained already by the array of 40 MT sites that addition of the VTF data can add little. In any event, this example demonstrates the consistency of the two datasets, as both can be fit simultaneously with the same inverse solution.

Other synthetic examples demonstrate the potential benefit of joint inversion a bit more clearly. We performed three inversion tests on the second test case, with data generated for the synthetic model of Fig. 3b, as described above. Error floors were set at 10% and 5% for the VTF and the impedance data, respectively. Initial models for all runs are $50\ \Omega\text{m}$ half space. The first inversion was performed using just the VTF data, the second with just the impedance tensor, and the last with both data types. All inversion reaches the target misfit of 1 RMS. Fig. 8 displays cross-sectional plots at $X=0$ km.

In all cases the conductor is recovered, although for the VTF case the burial depth is greater than what it should be (Fig. 8a). This again

shows that the VTF data can primarily constrain the location of the conductor in the horizontal, but not the vertical. Inversion of the impedance tensor alone recovers the anomalous volume quite well (Fig. 8b), but the conductivity is noticeably above the correct value of $1\ \Omega\text{m}$ (Fig. 8b). The best results are obtained by the joint inversion, where the resistivity, shape, size and depth of the conductor are close to correct. It is not clear why this example demonstrates a benefit of including VTF data, and the other does not; possibly different results would be obtained if the experiment was repeated with different realizations of random noise added to the data, or if the locations of the MT sites were perturbed, or different initial or prior models were used. Clearly the need to satisfy additional data constraints reduces the effects of noise in the data, and is likely to improve the fidelity of the inverse solution. For more complex structure the value of additional constraints provided by the VTF inversion are even clearer, as we show next by consideration of an example with real data.

5. Numerical experiments on real data

We applied the VTF inversion to the EXTECH dataset (Tuncer et al., 2006), consisting of tensor audio-magnetotelluric (AMT) soundings for 131 stations around the McArthur River mine, Saskatchewan, Canada. The goal in this survey was to use electromagnetic data to detect and map low resistivity graphite which is indicative of unconformity-type uranium deposits. A full description of the survey, and an interpretation of this dataset based on 2-D and 3-D analysis (including inversion with WSINV3D), is given in Tuncer et al. (2006). Further efforts at 3-D interpretation are given in and recently Farquharson and Craven (2008).

Here, we invert VTF and impedance data from 16 periods (from 8000 Hz to 5 Hz) at 131 sites (Fig. 2 of Tuncer et al., 2006), comparing results obtained with the two sorts of responses, separately and in combination. We use a $1000\ \Omega\text{m}$ half-space as an initial and prior model for all runs, as previous inversion of the impedance tensor suggests that this is a reasonable average background, and should thus produce sensible results when inverting the VTF alone. For inversion of the VTF (T_{zx} and T_{zy}) only, minimum error bars were set at 15% of $(|T_{zx}|^2 + |T_{zy}|^2)^{1/2}$. The inversion required about 8 iterations

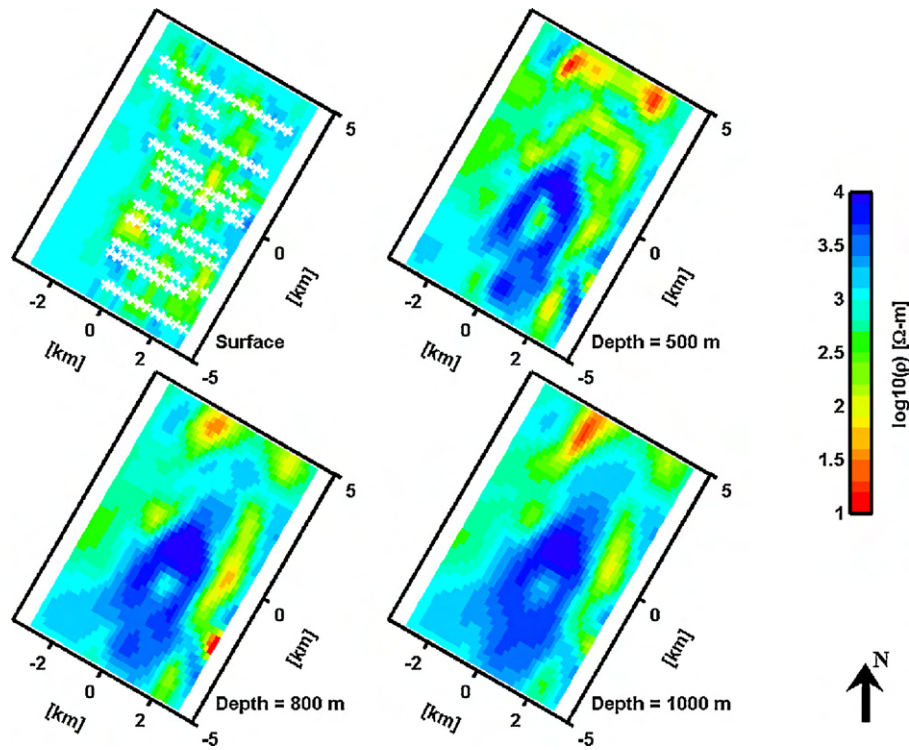


Fig. 9. The inverse solution at various depths from fitting the vertical magnetic transfer functions of the EXTECH dataset. The cross-symbols indicate the location of stations.

to converge to a minimum RMS of 1.2. Results for this inversion are given in Fig. 9.

For the second run we inverted the impedance tensor alone. In previous results using WSINV3D, reported in Tuncer et al. (2006) only the off-diagonal components (Z_{xy} and Z_{yx}) of the impedance were inverted. Here, we used all components including Z_{xx} and Z_{yy}

also. The minimum error bar for this run was set at 5% of $|Z_{xy}^{1/2} Z_{yx}^{1/2}|$ for off-diagonal and 50% for diagonal terms. When the same error floors were tried for off-diagonal and diagonal terms, the misfit could not be reduced below 3 RMS. With the modified error floors, the inversion required 4 iterations to converge to the target level of 1 RMS. The resulting model is shown in Fig. 10. The last run was a joint

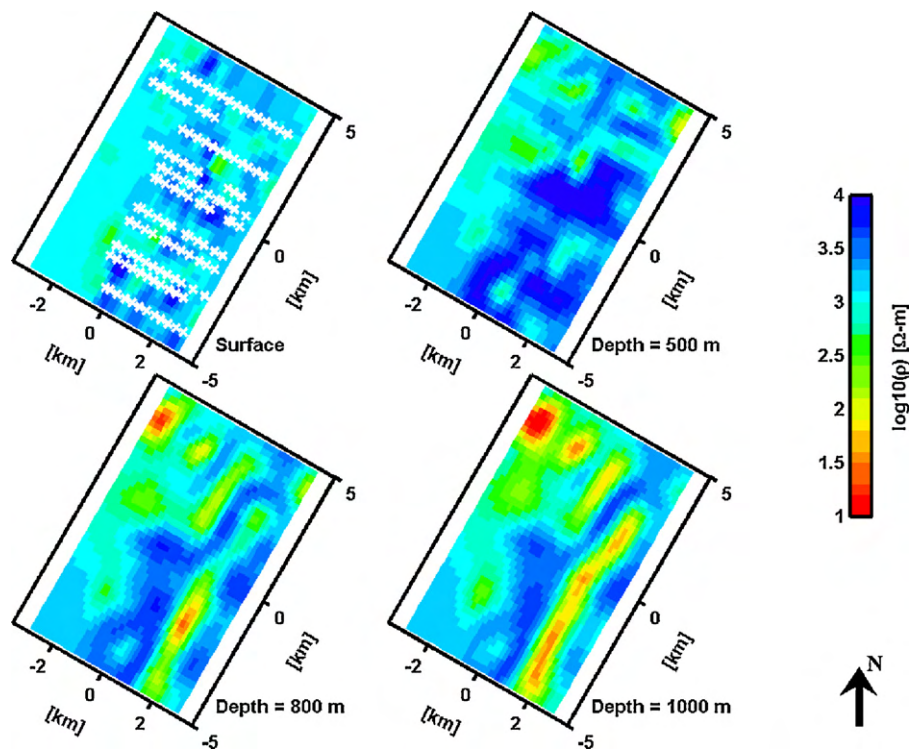


Fig. 10. The inverse solution at various depths from fitting all components of the impedance tensors of the EXTECH dataset.

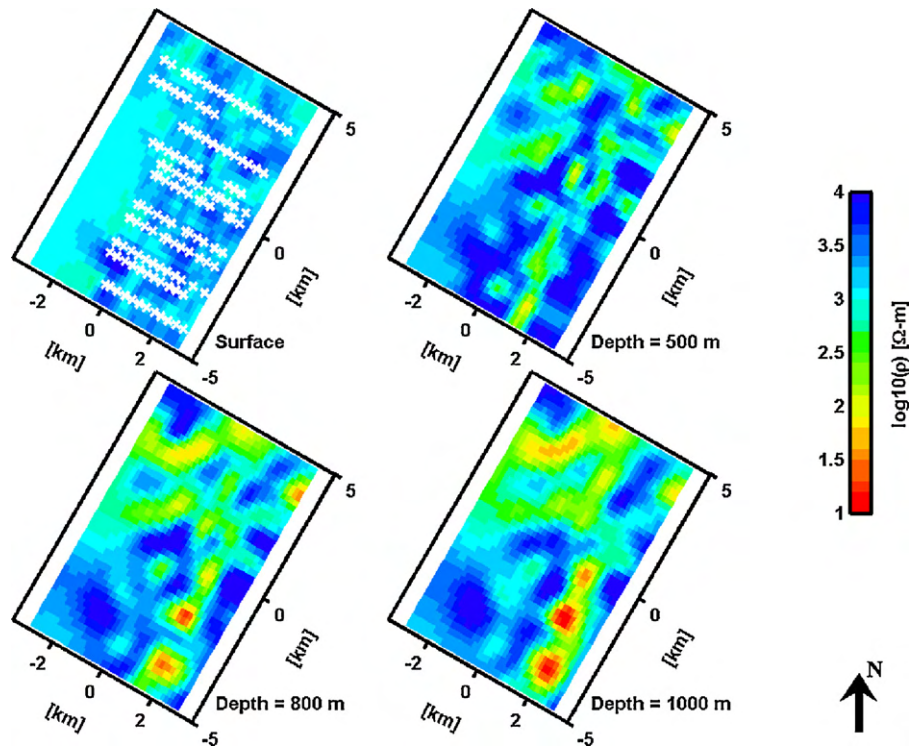


Fig. 11. The inverse solution at various depths from fitting both VTF and the impedance tensors of the EXTECH dataset.

inversion of the full impedance tensor and the vertical magnetic transfer function, with error floors set as in the first two runs. The inversion reduced the RMS misfit to 1.3 in 5 iterations. The model from the joint inversion is shown in Fig. 11.

Inverting just the impedance tensor (Fig. 10) reveals two main zones of high conductivity at 1000 m depth—an elongated feature of about $100 \Omega \text{ m}$ running perpendicular to the profiles on the east side of the model domain, and an area of variable (but

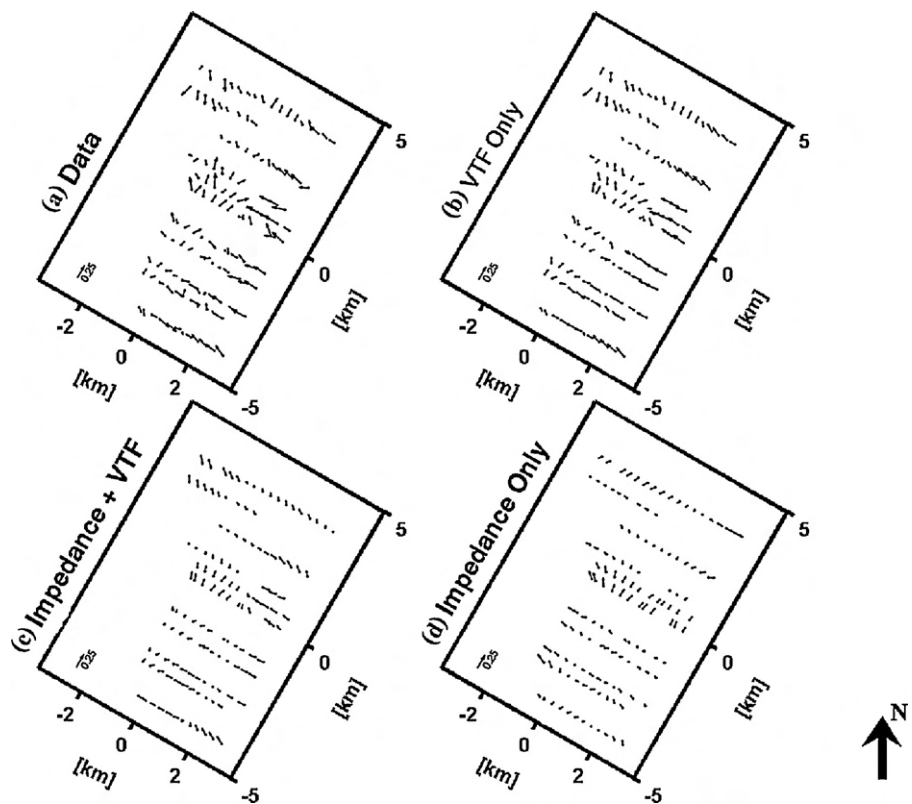


Fig. 12. The induction vectors at 100 Hz generated from (a) the observed VTF data, (b) the VTF inversion alone of Fig. 9, (c) the joint inversion of both impedance tensor and VTF data of Fig. 11, and (d) the impedance tensor inversion alone of Fig. 10. Notice that the calculated induction vectors in (d) fit the observed induction vectors more poorly.

generally higher) conductivity located in the northwest. The same features are evident, but somewhat weaker, in the 800 m layer. Similar features were obtained by inverting only the VTF data (Fig. 9). However, depth resolution appears poorer, as the inversion spreads the conductive features to shallower depths, particularly in the north, beyond the area covered by the MT profiles. The independent inversions of each data type confirm the lateral locations of the conductors. However, based on our experiments with synthetic data, the vertical position and extents of the conductive zones are almost certainly better constrained by the impedance tensor.

Results from joint inversion (Fig. 11) show increased conductivity in the same two general areas at 1000 m depth. However, the elongated conductor to the east now appears to be broken into segments, with patches of resistivity as low as $10 \Omega \text{ m}$, separated by areas with resistivities of several hundred $\Omega \text{ m}$. In contrast, inverting impedances alone results in a more uniform (approximately $100 \Omega \text{ m}$) continuous feature. Apparently, the VTFs cannot be fit by such a simple uniform conductor, but rather require significant along-strike variability (see Fig. 12). The feature to the north is also substantially modified by inclusion of both data types. Compared to the VTF only inversion, the depth of this feature is now clearly localized at around 1000 m, constrained by the impedance tensor. Inclusion of the VTF data also reduces peak conductivities in this area, and results in more linear conductive features which strike approximately east–west.

It is instructive to consider fits of the inverse solutions of Figs. 9–11 to the VTF data. Real induction vectors (with the Parkinson convention, so that arrows point toward conductors) are plotted in Fig. 12 for a frequency of 100 Hz, along with computed responses for the VTF only, impedance only, and joint inversions. The induction vectors are consistent with the presence of conductive features in the southeastern and northern parts of the array—e.g., note the clear reversal of vectors on most lines as they cross the elongated conductive feature at 1000 m depth (clearest in Fig. 10), and the reversal from South to North pointing vectors in the Northern corner of the study area. However, as noted by Tuncer et al. (2006) patterns in the observations are much more complex than can be reproduced by simple 3-D models. The VTF only inversion reproduces almost all of the complexity seen in the data (Figs. 12a and b). The joint inversion results in a smoother VTF response, and a slightly poorer fit to the data (Fig. 12; this is consistent with the larger error floor assumed in this case), but again, significant features in the data are reproduced in the fitted response. In contrast, the solution obtained from fitting the impedance tensor data alone (Fig. 12c) fits the VTF observations considerably less well, suggesting that the result from the joint inversion (Fig. 11) is more reliable than that from the impedance tensor alone (Fig. 10). A more detailed interpretation of this dataset is beyond the scope of this paper. See Tuncer et al. (2006) and Farquharson and Craven (2008) for further interpretation and discussion of the EXTECH data, and Craven et al. (2006) for comparison of inversion techniques using this data.

6. Conclusions

Experiments on both synthetic and real data show that inverting VTFs alone can recover anomalous structures, particularly if the prior model is close to the correct background or host value. In general, the qualities of the inverse solution obtained from VTF data alone are inferior to those obtained from inverting the impedance tensor alone. Vertical magnetic fields are generated whenever lateral conductivity gradients align with the normal inducing field. Thus, VTFs are sensitive to horizontal structures, and to some extent to resistivity contrasts, but not to depths or absolute values of resistivity. If some constraint on host resistivity can be provided, either *a priori*, or through inversion of impedances, the VTF data

can result in accurate 3-D imaging of the anomalous structures. Joint inversion of VTFs and the impedance tensor can help constrain subsurface structures, as shown in both synthetic and real data examples. In cases with very simple structures which are already well resolved by the impedance data VTFs add little to the inverse solution. However, with more realistic levels of complexity, as exemplified by the EXTECH data, inclusion of VTF data results in significant modifications to the inverse solution. Because the joint inversion model fits both datasets, it is likely to be more reliable.

One issue that deserves further investigation is the inability of the inversion to fit synthetic VTF data to within the tolerance implied by the noise level, which of course is well known in synthetic tests. We speculate that the VTF data can only be fit perfectly when the background resistivity is correct—implying at least a weak sensitivity of this sort of data to the background, as the analysis of Berdichevsky et al. (2003) in fact showed. In the case of using the wrong background resistivity (for which the data have little sensitivity) no nearby model parameters can provide a better fit, perhaps after adjusting conductivities of the anomalous bodies to roughly fit the VTFs, the Occam inversion is stuck in a local minimum of the penalty functional, and cannot escape from. It would be useful to compare other search algorithms (e.g., NLGG) to see if they suffered from similar problems.

A significant drawback with WSINV3DMT has been the large amount of memory required to store the sensitivity matrix, and the extensive computational time required for forward and sensitivity solutions. These drawbacks can be ameliorated by adapting the code to run with MPI to on parallel systems. We have parallelized the computations over frequencies, requiring no significant changes to our forward modeling routine. This approach is probably most appropriate for small cluster type machines. To make efficient use of a cluster or supercomputer with more than a few tens of processors would require different approaches, such as decomposing the modeling domain for the forward solver. We have also parallelized computation of cross products, sharing rows of the sensitivity computed on separate nodes to compute blocks of the coefficient matrix needed for the Gauss–Newton normal equations. The resulting dense system of normal equations can be solved on the master node, or using a parallel solver based on iterative methods. The optimal choice here depends on the size of the data space, with the iterative parallel solver only efficient for large datasets. The speedup of the code on a test dataset with 16 periods is nearly linear (with a coefficient of roughly 0.5) for up to 8 processors, but rolls over for a further increase to 16 processors. Even so, the parallelization should make use of the code on realistic 3-D datasets significantly more practical.

Acknowledgments

This research has been supported by a grant from the Thailand Research Fund (RMU5080025) to W.S., and by grant DE-FG02-06ER15819 from the U.S. Dept. of Energy to G.E. Other partial supports for this work to W.S. are from Thai National Grid Center, Thailand and EMPulse, Canada. The authors would like to thank Jim Carven for allowing us to use EXTECH data sets, and Collin Farquharson, George Helffrich, and an anonymous reviewer for comments that improved the manuscript.

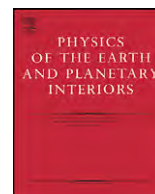
References

- Berdosian, P.A., Unsworth, M.J., Egbert, G.D., Thurber, C.H., 2004. Geophysical images of the creeping segment of the San Andreas fault: implications for the role of crustal fluids in the earthquake process. *Tectonophysics* 385, 137–158.
- Berdichevsky, M.N., Dmitriev, V.I., Golubtsova, N.S., Mershchikova, N.A., Pushkarev, P.Yu., 2003. Magnetovariational sounding: new possibilities, *Izvestiya. Physics of the Solid Earth* 39, 701–727.

- Choi, J., Moon, S., 1997. A parallel Cholesky factorization routine with a new version of PB-BLAS. In: *International Conference on Parallel and Distributed Systems (ICPADS'97)*, pp. 52–58.
- Constable, C.S., Parker, R.L., Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics* 52, 289–300.
- Craven, J.A., Farquharson, C., Mackie, R., Siripunvaraporn, W., Tuncer, V., Unsworth, M., 2006. A comparison of one-, two- and three-dimensional modeling of audiomagnetotellurics data collected at the world's richest uranium mine, Saskatchewan, Canada. In: *18th International Workshop on Electromagnetic Induction in the Earth*, El Vendrell, Spain, 17–23 September.
- deGroot-Hedlin, C., Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. *Geophysics* 55, 1613–1624.
- Farquharson, C.G., Craven, J.A., 2008. Three-dimensional inversion of Magnetotelluric data for mineral exploration: an example from the McArthur River uranium deposit, Saskatchewan, Canada. *J. Appl. Geophys.*, doi:10.1016/j.jappgeo.2008.02.002.
- Heise, W., Caldwell, T.G., Bibby, H.M., Bannister, S.C., 2008. Three-dimensional modeling of magnetotelluric data from the Rotokawa geothermal field, Taupo volcanic zone, New Zealand. *Geophys. J. Int.* 173, 740–750.
- Newman, G.A., Alumbaugh, D.L., 2000. Three-dimensional magnetotelluric inversion using non-linear conjugate gradients. *Geophys. J. Int.* 140, 410–424.
- Oskooi, B., Pedersen, L.B., 2005. Comparison between VLF and RMT methods. A combined tool for mapping conductivity changes in the sedimentary cover. *J. Appl. Geophys.* 57, 227–241.
- Parker, R.L., 1994. *Geophysical Inverse Theory*. Princeton University Press.
- Parkinson, W.D., 1959. Directions of rapid geomagnetic variations. *Geophys. J. Roy. Astronom. Soc.* 2, 1–14.
- Patro, Egbert, G., 2008. Regional conductivity structure of Cascadia: preliminary results from 3D inversion of USArray transportable array magnetotelluric data. *Geophys. Res. Lett.* 35, L20311, doi:10.1029/2008GL035326.
- Pous, J., Heise, W., Schnegg, P., Munoz, G., Marti, J., Soriano, C., 2002. Magnetotelluric study of the Las Canadas caldera (Tenerife, Canary Islands): structural and hydrogeological implications. *Earth Planet. Sci. Lett.* 204, 249–263.
- Rodi, W.L., 1976. A technique for improving the accuracy of finite element solutions for Magnetotelluric data. *Geophys. J. Roy. Astr. Soc.* 44, 483–506.
- Rodi, W.L., Mackie, R.L., 2001. Nonlinear Conjugate Gradients Algorithm for 2-D magnetotelluric inversion. *Geophysics* 66, 174–187.
- Siripunvaraporn, W., Egbert, G., 2000. An efficient data-subspace inversion method for 2-D magnetotelluric data. *Geophysics* 65, 791–803.
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., 2002. Numerical accuracy of magnetotelluric modeling: a comparison of finite difference approximations. *Earth Planets Space* 54, 721–725.
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., Uyeshima, M., 2005. Three-dimensional magnetotelluric inversion: data-space method. *Phys. Earth Planet. Interiors* 150, 3–14.
- Siripunvaraporn, W., Egbert, G., 2007. Data space conjugate gradient inversion for 2-D magnetotelluric data. *Geophys. J. Int.* 170, 986–994.
- Toh, H., Baba, K., Ichiki, M., Motobayashi, T., Ogawa, Y., Mishina, M., Takahashi, I., 2006. Two-dimensional electrical section beneath the eastern margin of Japan sea. *Geophys. Res. Lett.* 33, L22309.
- Tuncer, V., Unsworth, M.J., Siripunvaraporn, W., Craven, J.A., 2006. Exploration for unconformity-type uranium deposits with audiomagnetotelluric data: a case study from the McArthur River mine, Saskatchewan, Canada. *Geophysics* 71, B201–B209.
- Unsworth, M., Bedrosian, P., Eisel, M., Egbert, G., Siripunvaraporn, W., 2000. Along strike variations in the electrical structure of the San Andreas Fault at Parkfield, California. *Geophys. Res. Lett.* 27 (18), 3021–3024.
- Uyeshima, M., Ogawa, Y., Honkura, Y., Koyama, S., Ujihara, N., Mogi, T., Yamaya, Y., Harada, M., Yamaguchi, S., Shiozaki, I., Noguchi, T., Kuwaba, Y., Tanaka, Y., Mochido, Y., Manabe, N., Nishihara, M., Saka, M., Serizawa, M., 2005. Resistivity imaging across the source region of the 2004 Mid-Niigata Prefecture earthquake (M6.8), central Japan. *Earth Planets Space* 57, 441–446.
- Wannamaker, P.E., Booker, J.R., Jones, A.G., Chave, A.D., Filloux, J.H., Waff, H.S., Law, L.K., 1989. Resistivity cross section through the Juan de Fuca subduction system and its tectonic implications. *J. Geophys. Res.* 94 (B10), 14,114–14,127.
- Wannamaker, P.E., Hasterok, D.P., Johnston, J.M., Stodt, J.A., Hall, D.B., Sodergren, T.L., Pellerin, L., Maris, V., Doerner, W.M., Groenewold, K.A., Unsworth, M.J., 2008. Lithospheric dismemberment and magmatic processes of the Great Basin–Colorado Plateau transition, Utah, implied from magnetotellurics. *Geochem. Geophys. Geosyst.* 9 (5), doi:10.1029/2007GC001886, Q05019.

ภาคผนวก จ. Reprint

Boonchaisuk, S., Vachiratietchai, C., Siripunvaraporn, W., 2008, Two-dimensional direct current (DC) resistivity inversion: Data space Occam's approach, *Physics of the Earth and Planetary Interiors* 168 (3-4), pp. 204-211



Two-dimensional direct current (DC) resistivity inversion: Data space Occam's approach

Songkhun Boonchaisuk, Chatchai Vachiratienchai, Weerachai Siripunvaraporn*

Department of Physics, Faculty of Science, Mahidol University, Rama VI Road, Rachatawee, Bangkok 10400, Thailand

ARTICLE INFO

Article history:

Received 3 October 2007

Received in revised form 25 May 2008

Accepted 30 June 2008

Keywords:

Occam's inversion

DC resistivity

Data-space method

ABSTRACT

A data space Occam's inversion algorithm for 2D DC resistivity data has been developed to seek the smoothest structure subject to an appropriate fit to the data. For traditional model space Gauss–Newton (GN) type inversion, the system of equations has the dimensions of $M \times M$, where M is the number of model parameter, resulting in extensive computing time and memory storage. However, the system of equations can be mathematically transformed to the data space, resulting in a dramatic drop in its dimensions to $N \times N$, where N is the number of data parameter, which is usually less than M . The transformation has helped to significantly reduce both computing time and memory storage. Numerical experiments with synthetic data and field data show that applying the data space technique to 2D DC resistivity data for various configurations is robust and accurate when compared with the results from the model space method and the commercial software RES2DINV.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The direct current (DC) resistivity method has been used for various applications in hydrogeological, mining, and geotechnical investigations and environmental surveys (e.g., Ward, 1990; Daily et al., 1992, 1995; Ramirez et al., 1993, 1996; LaBrecque and Ward, 1990; among many others). The measured voltages caused by injected current bring out information on the earth's structure. The inversion program is then applied to interpret the measured voltages to obtain the Earth's resistivity structure.

The development of DC resistivity inversions has progressed successfully. Various techniques have been proposed for the two-dimensional (2D) and three-dimensional (3D) DC resistivity inversion (e.g., Pelton et al., 1978; Tripp et al., 1984; Nariida and Vozoff, 1984; Tong and Yang, 1990; Park and Van, 1991; Ellis and Oldenburg, 1994; Li and Oldenburg, 1994; Sasaki, 1994; Loke and Barker, 1995; Zhang et al., 1995; Loke and Dahlin, 1997, 2002; Tsourlos et al., 1998; Jackson et al., 2001; Pain et al., 2002; Loke et al., 2003; Günther et al., 2006; Pidlisecky et al., 2007; among many others). The most direct approach is the Gauss–Newton (GN) and its variant methods (e.g., Sasaki, 1994; Li and Oldenburg, 1994; Loke and Dahlin, 1997). Other limited memory optimization algorithms are the Quasi-Newton (QN) method (Loke and Barker, 1996; Loke and Dahlin, 1997, 2002; Tsourlos et al., 1998),

the conjugate gradient (CG) type inversion (Zhang et al., 1995) and the non-linear conjugate gradient (NLCG) (Ellis and Oldenburg, 1994). These are the schemes that require the gradient of the function. The derivative-free methods are neural networks (El-Qady and Ushijima, 2001) and genetic algorithms (Schwarzbach et al., 2005).

One of the main disadvantages of the GN-type inversion is that it requires solving a large and dense $M \times M$ system of equations, where M is the number of model parameters. Another disadvantage is the formation of the full $N \times M$ Jacobian or sensitivity matrix. Calculation of the full Jacobian requires a numerical solution of many forward problems. Both disadvantages, consequently, result in extensive computing time and memory storage. For example, in the 3D inversion, the synthetic model of a burial mound and data used by Günther et al. (2006) has 23,109 parameter cells ($M=23,109$) which is a lot more than the number of data parameters ($N=3439$). Inverting the $23,109 \times 23,109$ matrix and forming the Jacobian would require about 4–5 GBytes of RAM and many hours of CPU time.

The problem for the 3D DC resistivity inversion is quite similar to (though not as severe as) that for the 3D magnetotelluric (MT) survey, where the model parameter (M) is significantly greater than the data parameter (N). Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2005) could overcome this difficulty by transforming the model space inverse problem into the data space problem for their 2D and 3D Magnetotelluric data, respectively. With the transformation, the computational time and memory storage are greatly reduced by a factor of several (Siripunvaraporn and

* Corresponding author. Tel.: +66 2 201 5764; fax: +66 2 354 7159.
E-mail address: scwsp@mahidol.ac.th (W. Siripunvaraporn).

Egbert, 2000; Siripunvaraporn et al., 2005). However, formation of the Jacobian matrix is still a requirement.

For the limited memory optimization schemes such as QN, the full Jacobian or sensitivity matrix and the large and dense coefficient matrix of the system of equations are not necessarily constructed. Instead, a multiplication of the Jacobian with any vector can be calculated by solving the forward problem. These methods therefore never require a large memory storage as in GN-type inversions. Another advantage of QN-type inversions over the model space GN-type is their speed. However, their stability may be questionable (Loke and Dahlin, 2002). Though GN-type inversions may use extensive computing time, their main advantages are stability and robustness. They require fewer iterations to converge to the solution than limited memory methods (Loke and Dahlin, 2002; Siripunvaraporn and Egbert, 2007).

Because of their stability, we still have confidence in GN-type inversion techniques, especially Occam's method as first introduced by Constable et al. (1987). Siripunvaraporn and Egbert (2007) showed that for 2D MT data, the computing time of a GN-type inversion in the data space is actually comparable to that of the CG or NLCG inversion. For all of these reasons, here we propose to solve the multi-dimensional DC resistivity inverse problem using one variant GN-technique, Occam's inversion. However, instead of solving the problem in model space as others have (e.g., Constable et al., 1987; Sasaki, 1994), we propose to solve the DC resistivity inverse problem in data space as in Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2004, 2005). In order to test the feasibility and practicality of the data space approach for 3D DC resistivity data, we developed the 2D DC resistivity inversion based on the data space approach of Siripunvaraporn et al. (2005), which will be extended to 3D in the future.

We first start the paper by briefly reviewing the basic idea of Occam's inversion in the usual model space formulation, and then from a data space perspective. We then describe the implementation of the data space technique to a 2D DC resistivity data set. Numerical experiments of both synthetic and real field data in comparison with the commercial software RES2DINV version 3.55 (Loke and Barker, 1996) are shown at the end.

2. Occam's inversion: model space approach versus data space approach

Constable et al. (1987) introduced the Occam method for 1D MT and Schlumberger sounding data. Since then it has become one of the "classic" inversion techniques for various geophysical data (e.g., deGroot-Hedlin and Constable, 1990, 2004; deGroot-Hedlin, 1995; LaBrecque et al., 1996; Siripunvaraporn and Egbert, 2000; Huang et al., 2003; Siripunvaraporn et al., 2005; Greenhalgh et al., 2006; among others). For more general and detailed discussions of the Occam approach, see Constable et al. (1987), deGroot-Hedlin and Constable (1990), Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2004, 2005).

The philosophy of the Occam approach is to seek for the "smoothest" or "minimum" structure model subject to a constraint on the misfit (Constable et al., 1987), which can be mathematically translated into a problem of minimization of an unconstrained functional $U(\mathbf{m}, \lambda)$,

$$U(\mathbf{m}, \lambda) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \lambda^{-1} \{(\mathbf{d} - \mathbf{F}[\mathbf{m}])^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{F}[\mathbf{m}]) - X^{*2}\}. \quad (1)$$

Here $\mathbf{m}$ is a resistivity or conductivity model of dimension M , $\mathbf{m}_0$ a base or prior model, $\mathbf{C}_m$ a model covariance matrix which defines the model norm, $\mathbf{d}$ the observed data with dimension N , $\mathbf{F}[\mathbf{m}]$ the forward model response, $\mathbf{C}_d$ a data covariance matrix, X^* the desired

level of misfit, and λ^{-1} a Lagrange multiplier. In the 2D DC resistivity case, the data $\mathbf{d}$ are the apparent resistivities from different configurations. The model response $\mathbf{F}[\mathbf{m}]$ is computed by solving the DC resistivity forward problem, which we will describe later.

Instead of directly minimizing (1), Constable et al. (1987) consider the penalty functional $W_\lambda(\mathbf{m})$,

$$W_\lambda(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \lambda^{-1} \{(\mathbf{d} - \mathbf{F}[\mathbf{m}])^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{F}[\mathbf{m}])\}. \quad (2)$$

When λ is fixed, $\partial U / \partial \mathbf{m}$ and $\partial W_\lambda / \partial \mathbf{m}$ yield the same result. Therefore, minimizing W_λ with a series of λ values, and choosing λ for which the smallest minimum is achieved, is equivalent to minimizing the original functional U of (1).

Because of the non-linearity of the inverse problem, the Occam's inversion starts with the linearization of the response function $\mathbf{F}[\mathbf{m}]$ based on the Taylor series expansion, $\mathbf{F}[\mathbf{m}_{k+1}] = \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_{k+1} - \mathbf{m}_k)$. Inserting the series expansion in (2), and then solving for the stationary points, a series of iterative approximate solutions is then obtained,

$$\mathbf{m}_{k+1}(\lambda) - \mathbf{m}_0 = [\lambda \mathbf{C}_m^{-1} + \mathbf{J}_k^T \mathbf{C}_d^{-1} \mathbf{J}_k]^{-1} \mathbf{J}_k^T \mathbf{C}_d^{-1} \mathbf{d}_k, \quad (3)$$

where $\mathbf{d}_k = \mathbf{d} - \mathbf{F}[\mathbf{m}_k] + \mathbf{J}_k(\mathbf{m}_k - \mathbf{m}_0)$, the subscript k denotes the iteration number, and $\mathbf{J}_k = (\partial \mathbf{F} / \partial \mathbf{m})_k$ is the $N \times M$ sensitivity or Jacobian matrix calculated at $\mathbf{m}_k$. Note that the system of Eq. (3) has dimensions of $M \times M$. We therefore called this technique the "model space" Occam's inversion.

Parker (1994) showed that the solution to (3) for iteration k can be transformed to

$$\mathbf{m}_{k+1} - \mathbf{m}_0 = \mathbf{C}_m \mathbf{J}_k^T \boldsymbol{\beta}_{k+1}, \quad (4)$$

where $\boldsymbol{\beta}_{k+1}$ is an unknown expansion coefficient vector. The derivation of (4) from (3) is also given in Siripunvaraporn et al. (2005). Searching for the stationary points with the transformation (4), a series of iterative solutions is again obtained,

$$\boldsymbol{\beta}_{k+1} = [\lambda \mathbf{C}_d + \mathbf{J}_k \mathbf{C}_m \mathbf{J}_k^T]^{-1} \mathbf{d}_k. \quad (5)$$

Note that the system of Eq. (5) has dimensions $N \times N$, rather than $M \times M$ as in (3). Here is the main difference between (3) and (5). Because we transform the computation from model space to data space, we therefore called this technique after the transformation the "data space" Occam's inversion. If all the same parameters are used the solutions from both approaches will be identical (Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005). For MT data, the number of model parameters M is usually much larger than the number of data values N . Both the calculation time and memory are significantly decreased with the transformation to data space (Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2005). Here, we apply this method to DC resistivity data and we expect to gain the same benefits.

The beauty of Occam's inversion is here, which makes it different from other regularized inverse problems. In either the model space or data space approach, the goal is to search for the minimization of (1). This can be performed by two stages. The first stage (Phase I) is to bring the misfit down to the target level by varying λ values in (3) and (5) for each iteration. Once the target misfit is achieved, Phase II keeps the misfit at the desired level and searches for the minimum norm model by again varying λ values in each iteration. The addition of Phase II is to guarantee that the model structure does not contain unwanted or spurious structures (Siripunvaraporn et al., 2004, 2005).

3. Implementations for 2D DC resistivity data

As stated earlier, the goal of this paper is to test the feasibility and practicality of the data space approach to fit general DC resistivity data by using 2D DC resistivity data as an example. To develop the codes, major implementations are the 2D DC resistivity forward modeling and the sensitivity calculation routines. Other parts closely followed Siripunvaraporn and Egbert (2000) and Siripunvaraporn et al. (2005).

3.1. Two-dimensional DC resistivity forward modeling

A first and important step is to develop a 2D DC resistivity forward modeling routine. A good and efficient forward modeling routine helps make the inversion effective because it is the basis of most processes of the inversion. There are many techniques for solving the 2D resistivity forward problem, such as the transmission surface method (Swift, 1971; Madden, 1971; Pelton et al., 1978), the finite difference method (Dey and Morrison, 1979; Mufti, 1976; Mundry, 1984; Lowry et al., 1989) and the finite element method (Coggon, 1971; Rijo, 1977; Pelton et al., 1978; Pridmore et al., 1980, 1981; Uchida and Murakami, 1990; Queralt et al., 1991).

One of the advantages of the finite difference and finite element methods over the other methods is their well-known ability to quickly approximate the solutions for any arbitrary and complex structure models. Finite difference method is relatively fast compared with finite element method. However, to include a general topography, the finite element method becomes a better selection. The DC resistivity survey is usually applied for shallow studies in which the topography must be accounted for. Here, we therefore choose the finite element method for our 2D DC resistivity problem. A brief review of our implementation is stated next. For specific details, readers are encouraged to consult Coggon (1971), Rijo (1977), Queralt et al. (1991), Xu et al. (2000) and Boonchaisuk (2007).

In the 2D DC resistivity forward problem, the governing equation for the electrical potential must be transformed into Fourier space (Pelton et al., 1978; Dey and Morrison, 1979; Uchida and Murakami, 1990; Queralt et al., 1991) to remove the strike direction variable, i.e., from $\phi_c(x, y, z)$ into $\varphi_f(x, k_y, z)$, where y is the strike direction, k_y is the wave number, and ϕ_c and φ_f are the electrical potential in Cartesian coordinates and in Fourier space, respectively. The finite element method, closely following Rijo (1977) and Queralt et al. (1991), is then applied by using triangular elements for the model discretization. However, the mixed boundary conditions of Queralt et al. (1991) and Dey and Morrison (1979) are imposed at all grid boundaries. These conditions help produce better solutions than the classical ones (Dirichlet or Neumann) (Queralt et al., 1991). For each k_y wave number, the global system of equations $\mathbf{K}\varphi_f = \mathbf{F}$ is then obtained, where $\mathbf{K}$ is the discretized differential operator, φ_f is the unknown potential vector in Fourier space and $\mathbf{F}$ is the imposed boundary condition. The Cholesky decomposition method is later applied to solve the system of equations to obtain φ_f .

The final step for the forward modeling routine is to convert the potential vector φ_f in Fourier space back to the potential vector ϕ_c in Cartesian coordinates. This process can be done by directly applying the inverse Fourier transform to the Fourier potential. However, in order to obtain an accurate result, a direct computation would require many Fourier potential solutions from various wave numbers k_y (more than 10; Boonchaisuk, 2007). Thus, the number of times required to solve the system of equations $\mathbf{K}\varphi_f = \mathbf{F}$ would be equal to the number of wave numbers n_k used. Xu et al. (2000) proposed an optimization technique that requires at least four wave numbers to generate an accurate solution. Hence we closely fol-

lowed the method of Xu et al. (2000) but slightly adapted it so that the number of wave numbers used is around 8–10.

Once the inverse Fourier transform is performed using the solutions φ_f from various k_y , the electrical potentials at the surface can be obtained and are then used to compute the apparent resistivities for any array configuration. Extensive tests on simple to complex models were conducted to guarantee the accuracy of the forward modeling routine. By comparing the results to the analytic solutions (where applicable) and to other existing codes such as the RES2DMOD program version 2.2 (Loke and Barker, 1996), the RMS misfit is about 1% or less. We therefore conclude that our finite element code is accurate and comparable to other existing codes (Boonchaisuk, 2007). Note that our codes are developed only with MATLAB scripts.

3.2. Sensitivity calculation

The sensitivity term $\mathbf{J}_k = (\partial \mathbf{F} / \partial \mathbf{m})_k$ arises inevitably in every non-linear inversion process. A single sensitivity value denotes the change of the forward response with respect to a change of the model parameter. In general, there are three ways to compute the sensitivity for the DC resistivity (Spitzer, 1998): the perturbation method, the sensitivity forward calculation and the potential approximation. Here, to form the sensitivity matrix $\mathbf{J}$, we used the adjoint Green's function technique described in McGillivray and Oldenburg (1990). This technique requires a number of forward problem calls of only $n_k \times N$, in contrast with $n_k \times M$ when not using the adjoint method. For a given k_y , since the matrix $\mathbf{K}$ has already been decomposed and stored, the system of equations can then be quickly solved.

3.3. Model covariance

For the data space approach, the model covariance $\mathbf{C}_m$ is the same as that used in Siripunvaraporn et al. (2005). For the model space approach, a roughening matrix similar to those of deGroot-Hedlin and Constable (1990) was used as the inverse of the model covariance.

4. Numerical experiments

In this section, we will investigate the efficiency of our inversion codes for the 2D resistivity data. The first test is performed on synthetic data generated from the two blocks with a contrast resistivity model. Next, we test our codes from the field data collected in our geophysical test area. Our codes are written with MATLAB and run on a personal computer (PC: Pentium IV-3.0 GHz with 1 GB of RAM).

4.1. Examples with synthetic data: two blocks with contrast resistivity model

The synthetic data are generated from the two blocks with the contrast resistivity model shown in Fig. 1. The two $12.5 \times 20 \text{ m}^2$ blocks located next to each other are buried in the $10 \Omega \text{ m}$ half-space at 2.5 m from the ground surface. The left block has $100 \Omega \text{ m}$ and the right block has $1 \Omega \text{ m}$ (Fig. 1). The apparent resistivity data

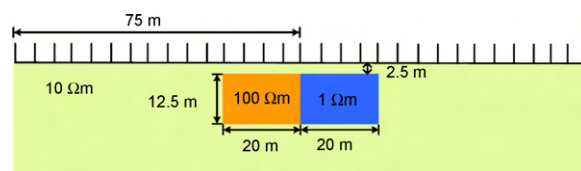


Fig. 1. Two blocks with contrast resistivity model. The solid marks above the surface indicate the measuring electrodes.

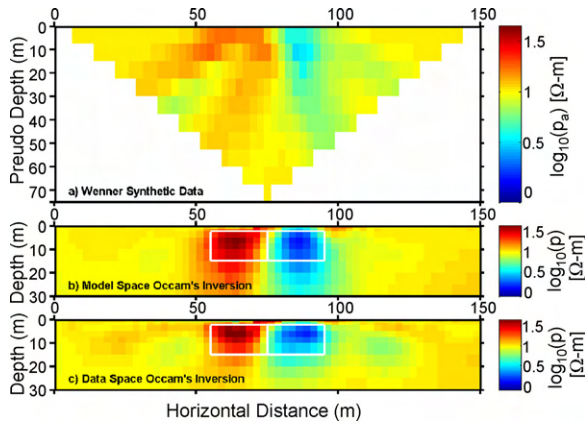


Fig. 2. The solution models from (b) the model space and (c) the data space Occam's program inverted from the synthetic Wenner array data shown in pseudosection (a). White lines indicate the two blocks of Fig. 1.

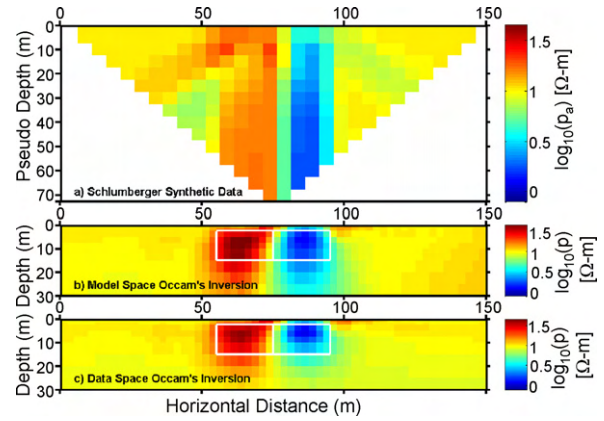


Fig. 4. The solution models from (b) the model space and (c) the data space Occam's program inverted from the synthetic Schlumberger array data shown in pseudosection (a). White lines indicate the two blocks of Fig. 1.

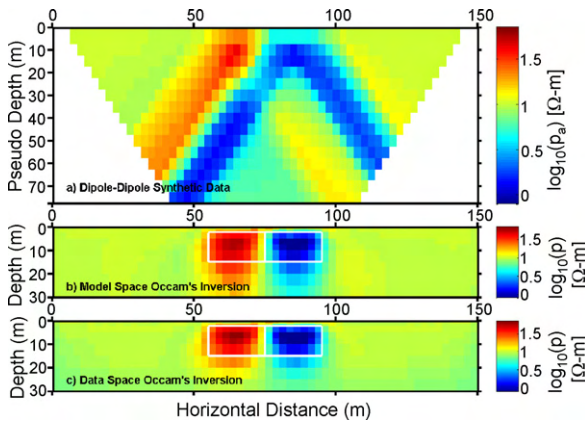


Fig. 3. The solution models from (b) the model space and (c) the data space Occam's program inverted from the synthetic Dipole–Dipole array data shown in pseudosection (a). White lines indicate the two blocks of Fig. 1.

sets for the Wenner, Dipole–Dipole and Schlumberger arrays are obtained from using 31 electrodes with an electrode distance of 5 m and a separation factor of $n = 1–10$ for Wenner (Fig. 2a), $n = 1–15$ for Dipole–Dipole (Fig. 3a) and $n = 1–14$ for Schlumberger arrays (Fig. 4a). These results have $N = 145$ for the Wenner array, 315 for the Dipole–Dipole array and 210 for the Schlumberger array. The finite element mesh used to generate the data is 150×32 in the horizontal and vertical directions, respectively. However, the 80×25 model mesh ($M = 2000$) is used for the inversion. Five percent Gaussian noises were added to the synthetic data. The data variance is set at 5%, calculating from the apparent resistivity to accommodate the discretization errors from using different meshes for the inversion and the forward modeling and the 5% Gaussian noises.

The starting model for the inversion is set to be the same as the base model ($\mathbf{m}_0$) which is a homogeneous half-space, with its resistivity value (ρ_{avg}) determined from the geometric mean of the apparent resistivity data (ρ_a), $\rho_{avg} = \sqrt[N]{\rho_a^1 \cdot \rho_a^2 \cdot \dots \cdot \rho_a^N}$. Thus, the geometric mean of apparent resistivity is about $9.22 \Omega m$ for the

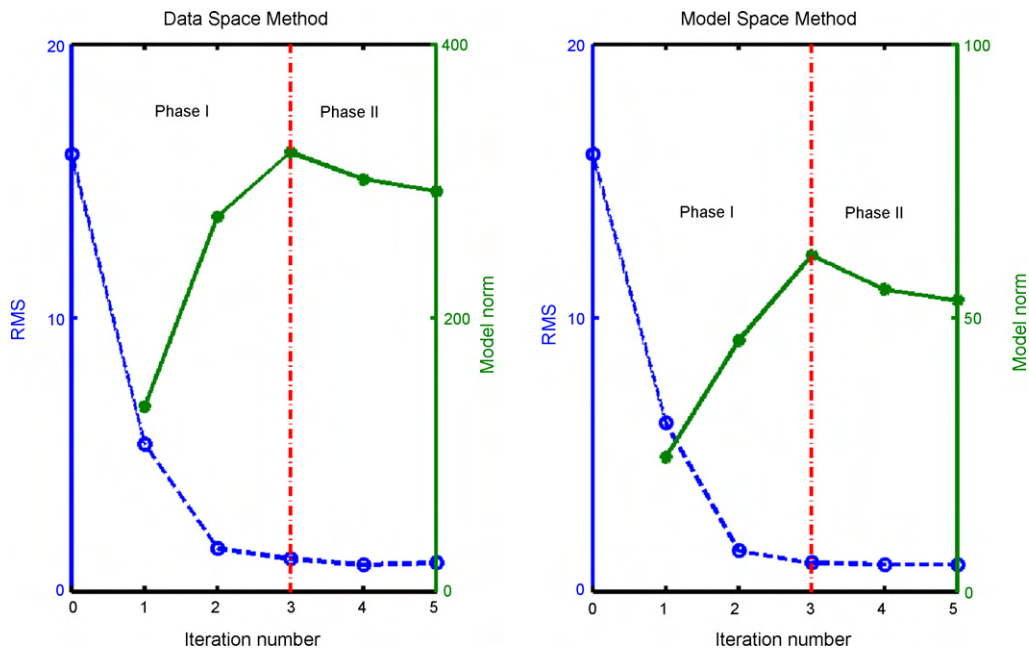


Fig. 5. The RMS misfit (dashed line; left y-axis) and model norm (solid line; right y-axis) versus the iteration number of the model space (right) and the data space (left) Occam's inversions for the synthetic Dipole–Dipole data (Fig. 3a) generated from two blocks with a contrast resistivity model (Fig. 1).

Wenner data (Fig. 2a), $7.03 \Omega \text{ m}$ for Dipole–Dipole data (Fig. 3a) and $8.18 \Omega \text{ m}$ for Schlumberger data (Fig. 4a). The initial RMS is about 6.13 for Wenner, 15.97 for Dipole–Dipole and 12.64 for Schlumberger. The target misfit set for the inversion is equal to 1, implying that the data are being fitted within their error levels of 5%. For Wenner and Schlumberger arrays, both approaches require 3 iterations to complete the minimization of U . For Dipole–Dipole data, the model space method requires 4 iterations, while the data-space method requires 5 iterations (Fig. 5).

Fig. 2b and c displays the final inverted models obtained at the target misfit with the minimum norm after 3 iterations for the Wenner configuration array with the model space and data space Occam's inversion, respectively. Similarly, Fig. 3b and c are for the Dipole–Dipole configuration, and Fig. 4b and c are for the Schlumberger configuration. In all figures, both model and data space approaches can recover both resistivity blocks quite well; however, the qualities may be different. The white line indicates the block boundaries. The top surfaces of the two blocks are clearly seen at a depth around 2.5 m for both methods. However, the bottom boundaries are resolved better with the data space approach. This is probably due to the different model covariance used for both methods.

Fig. 5a and b displays the convergence plot versus iteration number for the model space and data space algorithms, respectively, for Dipole–Dipole configurations. Both inversions requires about 3 iterations to reach the target misfit (Phase I), but require another one or two iterations to complete Phase II for the model or data-space method, respectively. Both programs are terminated in Phase II if the model norm of the next iteration increases or changes only slightly. The right y-axis of both figures shows that the model norm is being minimized. Note that the model covariance of the model space and data-space methods is different, resulting in different levels of model norms. Other convergence plots for the Wenner and Schlumberger arrays are similar, and are therefore not shown here.

Computational times for the model space and the data space approaches are not significantly different since the problem is 2D where the model domain is still small. However, in the process of solving the system of equations of the inversion, the model space method must spend more computing time than the data-space method. That is because the inverted matrix has a size of 2000×2000 compared with 315×315 for the data-space method. The time difference would be higher if the model domain were larger as in the case of the 3D problem. In addition, the memory required to store the system of equations is about a factor of 40 times greater in the model space than is required by the data-space method, as expected from the theory (Siripunvaraporn and Egbert, 2000; Siripunvaraporn et al., 2004, 2005).

We have also performed various tests on other synthetic data generated from different models, both simple and complex. We found that our inversion codes in both model space and data-space

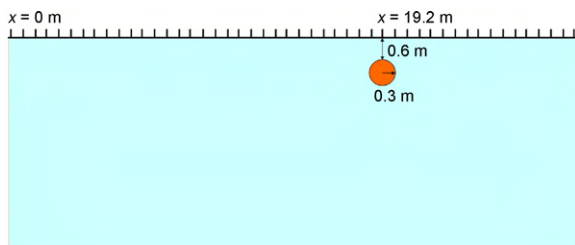


Fig. 6. Location of the drainpipe known a priori from the map of the Faculty of Science, Mahidol University. The solid marks above the surface indicate the electrodes with a spacing of 0.6 m.

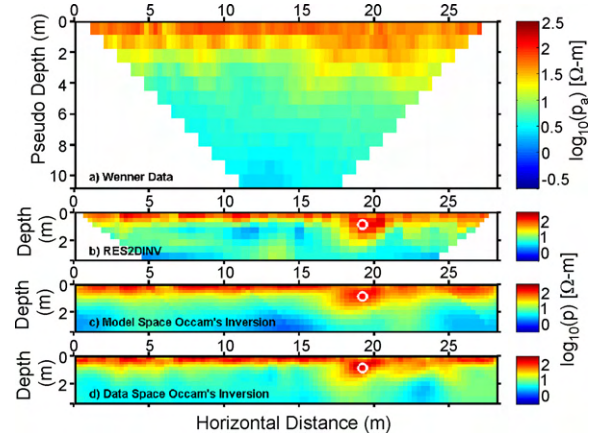


Fig. 7. The solution models from (b) RES2DINV, (c) our model space, and (d) our data space programs inverted from the observed Wenner array data shown in pseudosection (a). The white circle indicates the circular drainpipe in Fig. 6.

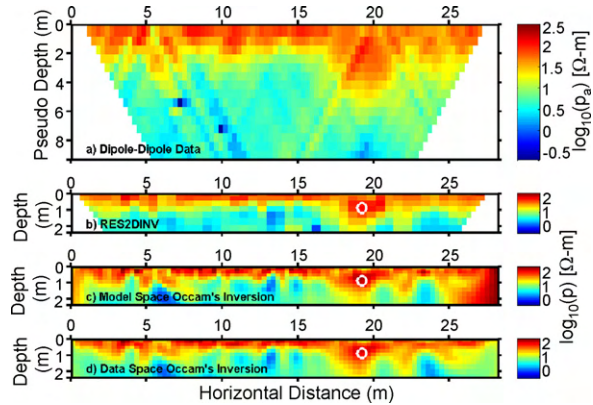


Fig. 8. The solution models from (b) RES2DINV, (c) our model space, and (d) our data space programs inverted from the observed Dipole–Dipole array data shown in pseudosection (a). The white circle indicates the circular drainpipe in Fig. 6.

methods are capable of recovering the anomalies buried inside and background structures of the model (Boonchaisuk, 2007).

4.2. Example with field data

In the previous section, we have shown that our inversion codes work reasonably well with synthetic data. In this section we show

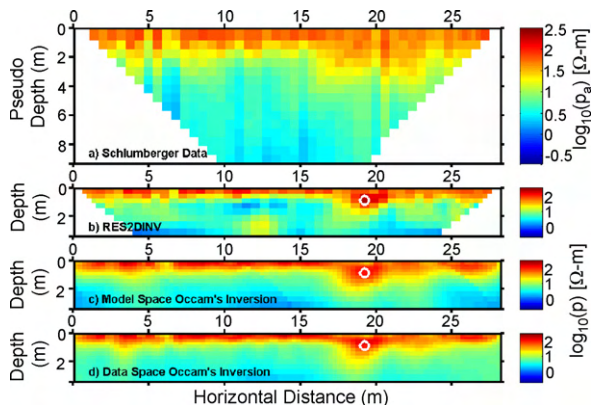


Fig. 9. The solution models from (b) RES2DINV, (c) our model space, and (d) our data space programs inverted from the observed Schlumberger array data shown in pseudosection (a). The white circle indicates the circular drainpipe in Fig. 6.

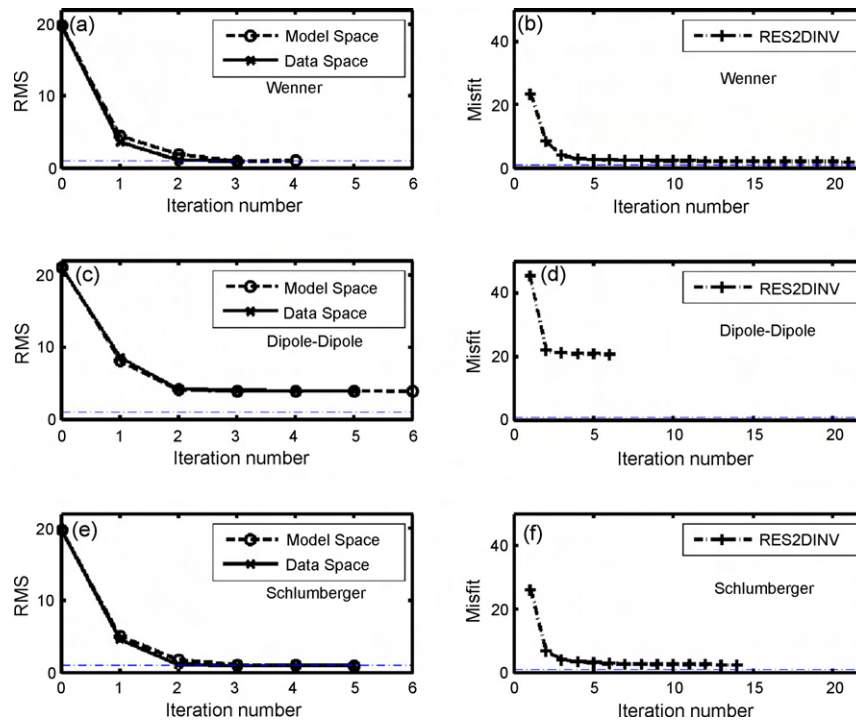


Fig. 10. (a, c, e) The RMS misfits versus the iteration numbers of the model (dashed line) and data space (solid line) methods. (b, d, f) The calculated misfits versus the iteration numbers from the RES2DINV program. Top row is for Wenner data, middle row is for Dipole–Dipole data and bottom row is for Schlumberger data. Note that the RES2DINV program uses a different formula from our RMS misfit to calculate its own misfit.

the flexibility, robustness and accuracy of our codes to the real observed data, which usually contains noise from many sources.

The DC resistivity data were collected at the geophysical test area inside the Faculty of Science, Mahidol University to detect the buried drainpipe; for which the position and depth was known a priori. The circular drainpipe is 0.3 m in radius and is buried at a depth of 0.9 m (from the surface to the center) as shown in Fig. 6. A SYSCAL R1 PLUS Switch-48 instrument was used to collect the observed data. A profile array of 48 electrodes with an electrode spacing of 0.6 m was applied perpendicular to the course of the drainpipe to obtain the 2D data.

Figs. 7a, 8a and 9a display the 335, 555 and 452 observed data values constituted from the Wenner, Dipole–Dipole and Schlumberger configurations, respectively, by using separation factors of 1–12, 1–15 and 1–15, respectively. The 114×21 model mesh is used for the inversions. The data variance is set to 5% of the apparent resistivity to account for the noise from measurement and other sources. The total number of model parameters is therefore 2394 (114×21), and the data parameters are 335, 555 and 452 for Wenner, Dipole–Dipole and Schlumberger configurations, respectively.

Both model space and data space Occam's inversion methods were applied to our field data set, along with the commercial software, RES2DINV version 3.55 (Loke and Barker, 1996), using default parameters of the program in order to provide a reference to our inverted models and to show the accuracy and robustness of our developed codes. The final inverse solutions for Wenner, Dipole–Dipole and Schlumberger data sets are shown in Figs. 7b, 8b and 9b for the RES2DINV program, Figs. 7c, 8c and 9c for our model space code, and Figs. 7d, 8d and 9d for our data space inversion, respectively. All inverted models from our codes and the commercial code show that there are two layers beneath the surface and a high-resistivity portion indicating the drainpipe. The high-resistivity ($100 \Omega \text{ m}$) top layer is about 1 m in depth lying on top

of the conductive ($10 \Omega \text{ m}$) layer. The high-resistivity layer is interpreted as the top soil filled on top of the clay conductive layer. The drainpipe is located directly in the high-resistivity zone matching the location plotted in the map.

Fig. 10 shows the misfit plots versus iteration numbers. The RES2DINV misfit is calculated differently from our RMS misfit, so we plot them on separate figures. The left column shows the RMS misfits calculated from our model and data space Occam's inversions (Fig. 10a, c and e). The right column shows the misfits generated from the RES2DINV code (Fig. 10b, d and f). For only the Dipole–Dipole configuration data, the RMS misfit did not converge to the desired misfit of 1 for the model and data-space methods. This also happened for the RES2DINV code, where a high RES2DINV misfit is still shown. In all cases, the RES2DINV code requires more iterations than both types of Occam's inversion.

5. Discussion and conclusions

The data-space method has been widely used in many fields including geophysics (Parker, 1994; Egbert et al., 1994; Chua and Bennett, 2001). It was recently used for MT data by Siripunvaraporn and Egbert (2000) for their 2D code and later applied to a 3D code (Siripunvaraporn et al., 2004, 2005). Here, we have shown that the same technique can be also used for 2D DC resistivity data. In the two examples presented here, we show that our developed 2D data space code is robust and accurate, and comparable to the commercial software RES2DINV program.

For any configuration, N is always less than M by a large factor. This fact helps enhance the benefit of using the data-space method, especially for the GN-type inversions. Since the size of the system of equations is significantly dropped to $N \times N$ from the original $M \times M$ in traditional model space, both the computing time and memory storage of the data-space method become a fraction of those in the model space method. Applying the data-space method to 3D

data would definitely yield an advantage, because in 3D, N/M is much smaller than in the 2D case. However, in the case where M is greater N (if occurred), the data space approach would not yield any advantages. In this case, we can select to run our code in the model space. Alternatively, a subset data space inversion, similar to REBOCC code (Siripunvaraporn and Egbert, 2000), can be applied in order to further reduce the computational costs.

Both the model and data space Occam's inversion methods require the full sensitivity matrix. This process consumes extensive computing time and large memory storage. Directly extending our 2D code to a 3D code would still yield an impractical 3D code. However, Siripunvaraporn and Egbert (2007) showed that we can avoid constructing the full sensitivity matrix by applying the conjugate gradient technique. This would help significantly in reducing the size of memory storage; however, the computing time may not necessarily decrease (Siripunvaraporn and Egbert, 2007).

In order to speed up the codes, an approximate sensitivity matrix is another option to consider. The Quasi-Newton method has been used to estimate the sensitivity (Loke and Barker, 1996). It was then used in combination with the GN method in order to maintain the accuracy of the GN method along with the speed of QN method (Loke and Dahlin, 2002). A hybrid method combining the advantages of the data-space method, GN, QN and other techniques is possible and should be explored for the 3D cases.

Acknowledgments

This research has been supported by a Development and Promotion of Science and Technology Talents Project (DPST) scholarship to SB, partially by Thai Center of Excellence for Physics (ThEP) funding to CV and by Thailand Research Fund (TRF: RMU5080025) funding to WS. The authors would like to thank the Faculty of Science, Mahidol University for their support for geophysics instruments. The authors would also like to thank two anonymous reviewers and the editor for their comments to help improve the manuscript.

References

- Boonchaisuk, S., 2007. Two-dimensional direct current resistivity inversion: finite element method. M.Sc. Thesis. Mahidol University.
- Chua, B.S., Bennett, A.F., 2001. An inverse ocean modeling system. *Ocean Model.* 3, 137–165.
- Coggon, J.H., 1971. Electromagnetic and electrical modeling by the finite element method. *Geophysics* 36, 132–155.
- Constable, C.S., Parker, R.L., Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics* 52, 289–300.
- Daily, W.D., Ramirez, A.L., LaBrecque, D.J., Nitao, J., 1992. Electrical resistivity tomography of vadose water movement. *Water Resources* 28, 1429–1442.
- Daily, W.D., Ramirez, A.L., LaBrecque, 1995. Electrical resistance tomography during in-situ trichloroethylene remediation at the Savannah River Site. *J. Appl. Geophys.* 33, 239–249.
- deGroot-Hedlin, C., 1995. Inversion for regional 2-D resistivity structure in the presence of galvanic scatterers. *Geophys. J. Int.* 122, 877–888.
- deGroot-Hedlin, C., Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. *Geophysics* 55, 1613–1624.
- deGroot-Hedlin, C., Constable, S., 2004. Inversion of magnetotelluric data for 2D structure with sharp resistivity contrasts. *Geophysics* 69, 78–86.
- Dey, A., Morrison, H.F., 1979. Resistivity modeling for arbitrarily shaped two-dimensional structures. *Geophys. Prosp.* 27, 106–136.
- Egbert, G.D., Bennett, A.F., Foreman, M.G., 1994. TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.* 99, 24821–24852.
- El-Qady, G., Ushijima, K., 2001. Inversion of DC resistivity data using neural networks. *Geophys. Prosp.* 49, 417–430.
- Ellis, R.G., Oldenburg, D.W., 1994. The pole-pole 3-D DC-resistivity inverse problem: a conjugate-gradient approach. *Geophys. J. Int.* 119, 187–194.
- Greenhalgh, S.A., Bing, Z., Green, A., 2006. Solutions, algorithms and inter-relations for local minimization search geophysical inversion. *J. Geophys. Eng.* 3, 101–113.
- Günther, T., Rucker, C., Spitzer, K., 2006. Three-dimensional modeling and inversion of dc resistivity data incorporating topography-II. Inversion. *Geophys. J. Int.* 166, 506–517.
- Huang, Z.X., Su, W., Peng, Y.J., Zheng, Y.J., Li, H.Y., 2003. Rayleigh wave tomography of China and adjacent regions. *J. Geophys. Res.* 108 (B2), ARTN 2073.
- Jackson, P.D., Earland, S.J., Reece, G.J., 2001. 3D resistivity inversion using 2D measurements of the electric field. *Geophys. Prosp.* 49, 26–39.
- LaBrecque, D.J., Ward, S.H., 1990. Two-dimensional cross-borehole resistivity model fitting. In: S. Ward (Ed.), *Geotechnical and Environmental Geophysics*, vol. 1, pp. 51–57.
- LaBrecque, D.J., Miletto, M., Daily, W., Ramirez, A., Owen, E., 1996. The effect of noise on Occam's inversion of resistivity tomography data. *Geophysics* 61, 538–548.
- Li, Y., Oldenburg, D.W., 1994. Inversion of 3-D DC resistivity data using an approximate inverse mapping. *Geophys. J. Int.* 116, 527–537.
- Loke, M.H., Barker, R.D., 1995. Least-squares deconvolution of apparent resistivity pseudosections. *Geophysics* 60, 1682–1690.
- Loke, M.H., Barker, R.D., 1996. Rapid least-squares inversion of apparent resistivity pseudosections by a Quasi-Newton method. *Geophys. Prosp.* 44, 131–152.
- Loke, M.H., Dahlin, T., 1997. A combined Gauss-Newton and Quasi-Newton inversion method for the interpretation of apparent resistivity pseudosections. In: *Proc. 3rd Meeting of the European Association for Environmental and Engineering Geophysics*, 8–11 Sept. 1997, Aarhus, Denmark, pp. 139–142.
- Loke, M.H., Dahlin, T., 2002. A comparison of Gauss-Newton and Quasi-Newton inversion method in resistivity imaging inversion. *J. Appl. Geophys.* 49, 149–162.
- Loke, M.H., Acworth, I., Dahlin, T., 2003. A comparison of smooth and blocky inversion methods in 2D electrical imaging surveys. *Explor. Geophys.* 34, 182–187.
- Lowry, T., Allen, M.B., Shive, P.N., 1989. Singularity removal: a refinement of modeling techniques. *Geophysics* 54, 766–774.
- Madden, T.R., 1971. The resolving power of geoelectric measurements for delineating resistive zones within the crust. In: J.G. Heacock (Ed.), *The Structure and Physical Properties of the Earth's Crust: AGU Geophys. Monogr. Ser.*, vol. 14, 95–105.
- McGillivray, Oldenburg, 1990. Methods for calculating Frechet derivatives and sensitivities for the non-linear inverse problem: a comparative study. *Geophys. Prosp.* 38, 499–524.
- Mufti, I.R., 1976. Finite-difference resistivity modeling for arbitrary shaped two-dimensional structures. *Geophysics* 41, 62–78.
- Mundry, E., 1984. Geoelectrical model calculations for two-dimensional resistivity distribution. *Geophys. Prosp.* 32, 124–131.
- Nariida, C.S., Vozoff, K., 1984. Two-dimensional DC resistivity inversion for dipole-dipole data. *IEEE Trans. Geosci. Remote Sens.* GE-22, 21–28.
- Pain, C.C., Herwanger, J.V., Worthington, M.H., de Oliveira, C.R.E., 2002. Effective multidimensional resistivity inversion using finite-element techniques. *Geophys. J. Int.* 151, 710–728.
- Park, S.K., Van, G.P., 1991. Inversion of pole-pole data for 3-D resistivity structure beneath arrays of electrodes. *Geophysics* 56, 951–960.
- Parker, R.L., 1994. *Geophysics Inversion Theory*. Princeton University Press.
- Pelton, W.H., Rijo, L., Swift, C.M.J.R., 1978. Inversion of two-dimensional resistivity and induced-polarization data. *Geophysics* 43, 788–803.
- Pidlisecky, A., Haber, E., Knight, R., 2007. RESINVM3D: a 3D resistivity inversion package. *Geophysics* 72, H1–H10.
- Pridmore, D.F., Hohmann, G.W., Ward, S.H., Sill, W.R., 1980. An investigation of finite-element modeling for electrical and electromagnetic data in three dimensions. *Geophysics* 46, 1009–1024.
- Pridmore, D.F., Hohmann, G.W., Ward, S.H., Sill, W.R., 1981. An investigation of finite-element modeling for electrical and electromagnetic data in three dimensions. *Geophysics* 46, 1009–1024.
- Queral, P., Pous, J., Marcuello, A., 1991. 2-D resistivity modeling: an approach to arrays parallel to the strike direction. *Geophysics* 56, 941–950.
- Ramirez, A.L., Daily, W.D., LaBrecque, D.J., Owen, E., Chesnut, D., 1993. Monitoring an underground steam injection process using electrical resistance tomography. *Water Resource Res.* 29, 73–87.
- Ramirez, A.L., Daily, W.D., Binley, A.M., LaBrecque, Roelant, D., 1996. Detection of leaks in underground storage tanks using electrical resistance methods. *J. Environ. Eng. Geophys.* 1, 189–203.
- Rijo, L., 1977. Modeling of electric and electromagnetic data. Ph.D. thesis. University of Utah.
- Sasaki, Y., 1994. 3-D resistivity inversion using the finite element method. *Geophysics* 55, 1839–1848.
- Schwarzbach, C., Borner, R., Spitzer, K., 2005. Two-dimensional inversion of direct current resistivity data using a parallel, multi-objective genetic algorithm. *Geophys. J. Int.* 162, 685–695.
- Siripunvaraporn, W., Uyeshima, M., Egbert, G., 2004. Three-dimensional inversion for Network-Magnetotelluric data. *Earth Planets Space* 56, 893–902.
- Siripunvaraporn, W., Egbert, G., 2000. An efficient data-subspace inversion method for 2-D magnetotelluric data. *Geophysics* 65, 791–803.
- Siripunvaraporn, W., Egbert, G., 2007. Data space conjugate gradient inversion for 2-D magnetotelluric data. *Geophys. J. Int.* 170, 986–994.
- Siripunvaraporn, W., Egbert, G., Lenbury, Y., Uyeshima, M., 2005. Three-dimensional magnetotelluric inversion: data-space method. *Phys. Earth Planetary Interiors* 150, 3–14.
- Spitzer, K., 1998. The three-dimensional dc sensitivity for surface and subsurface sources. *Geophys. J. Int.* 134, 736–746.
- Swift Jr., C.M., 1971. Theoretical magnetotelluric and turam response from two-dimensional inhomogeneities. *Geophysics* 36, 38–52.
- Tong, L., Yang, C., 1990. Incorporation of topography into two-dimensional resistivity inversion. *Geophysics* 55, 354–361.

- Tripp, A.C., Hohmann, G.W., Swift, C.M., 1984. Two-dimensional resistivity inversion. *Geophysics* 49, 1708–1717.
- Tsourlos, P.I., Szymanski, J.E., Tsokas, G.N., 1998. A smoothness constrained algorithm for the fast 2-D inversion of DC resistivity and induced polarization data. *J. Balkan Geophys. Soc.* 1, 3–13.
- Uchida, T., Murakami, Y., 1990. Development of Fortran code for the Two-Dimensional Schlumberger Inversion: Geological Survey of Japan Open-File Report No. 150, pp. 1–12.
- Ward, S.H., 1990. Geotechnical and environmental geophysics. *Investigat. Geophys.* 5, 147–189.
- Xu, S., Duan, B., Zhang, D., 2000. Selection of the wavenumber k using an optimization method for the inverse Fourier transform in 2.5D electrical modelling. *Geophys. Prosp.* 48, 789–796.
- Zhang, J., Mackie, R.L., Madden, T.R., 1995. 3-D resistivity forward modeling and inversion using conjugate gradients. *Geophysics* 60, 1313–1325.

ภาคผนวก ง. Reprint

- Kalscheuer, T., Pedersen, L.B., Siripunvaraporn, W., 2008, Radiomagnetotelluric two-dimensional forward and inverse modelling accounting for displacement currents, *Geophysical Journal International* 175 (2), pp. 486-514

Radiomagnetotelluric two-dimensional forward and inverse modelling accounting for displacement currents

Thomas Kalscheuer,¹ Laust B. Pedersen¹ and Weerachai Siripunvaraporn²

¹Department of Earth Sciences, Uppsala University, Villavägen 16, 75236 Uppsala, Sweden. E-mail: Thomas.Kalscheuer@geo.uu.se

²Department of Physics, Faculty of Science, Mahidol University, Rama VI Rd., Rachatawee, Bangkok 10400, Thailand

Accepted 2008 June 30. Received 2008 June 17; in original form 2007 October 30

SUMMARY

Electromagnetic surface measurements with the radiomagnetotelluric (RMT) method in the frequency range between 10 and 300 kHz are typically interpreted in the quasi-static approximation, that is, assuming displacement currents are negligible. In this paper, the dielectric effect of displacement currents on RMT responses over resistive subsurface models is studied with a 2-D forward and inverse scheme that can operate both in the quasi-static approximation and including displacement currents.

Forward computations of simple models exemplify how responses that allow for displacement currents deviate from responses computed in the quasi-static approximation. The differences become most obvious for highly resistive subsurface models of about 3000 Ω m and more and at high frequencies. For such cases, the apparent resistivities and phases of the transverse magnetic (TM) and transverse electric (TE) modes are significantly smaller than in the quasi-static approximation. Along profiles traversing 2-D subsurface models, sign reversals in the real part of the vertical magnetic transfer function (VMT) are often more pronounced than in the quasi-static approximation. On both sides of such sign reversals, the responses computed including displacement currents are larger than typical measurement errors.

The 2-D inversion of synthetic data computed including displacement currents demonstrates that serious misinterpretations in the form of artefacts in inverse models can be made if displacement currents are neglected during the inversion. Hence, the inclusion of the dielectric effect is a crucial improvement over existing quasi-static 2-D inverse schemes. Synthetic data from a 2-D model with constant dielectric permittivity and a conductive block buried in a highly resistive layer, which in turn is underlain by a conductive layer, are inverted. In the quasi-static inverse model, the depth to the conductive structures is overestimated, artefactual resistors appear on both sides of the conductive block and a spurious conductive layer is imaged at the surface.

High-frequency RMT field data from Ävrö, Sweden, are re-interpreted using the newly developed 2-D inversion scheme that includes displacement currents. In contrast to previous quasi-static modelling, the new inverse models have electrical resistivity values comparable to a normal-resistivity borehole log and boundaries between resistive and conductive structures, which correlate with the positions of seismic reflectors.

Key words: Numerical solutions; Inverse theory; Electrical properties; Electromagnetic theory; Magnetotelluric.

1 INTRODUCTION

Since many electromagnetic (EM) methods utilize frequencies below 10 kHz, the quasi-static assumption that displacement currents are much smaller than conduction currents (i.e. $\omega\epsilon \ll \sigma$ with angular frequency $\omega = 2\pi f$, dielectric permittivity $\epsilon = \epsilon_r\epsilon_0$, free air permittivity ϵ_0 and electrical conductivity σ) is stipulated and displacement currents are neglected during the data interpretation. For the radiomagnetotelluric (RMT) method, which uses EM fields in the VLF (3–30 kHz) and LF (30–300 kHz) frequency ranges, the validity of the quasi-static assumption is questionable. For a typical relative dielectric permittivity $\epsilon_r = 5$ (e.g. mildly fractured crystalline bedrock), displacement currents are equally strong as conduction currents for, for example, an electrical resistivity $\rho = 1/\sigma = 10\,000\ \Omega\text{m}$ and a frequency $f = 360\ \text{kHz}$. This means that the dielectric effect is non-negligible even at a combination of lower frequencies and/or resistivities. In fact, it can be argued that the

dielectric effect should be accounted for as soon as the perturbation it causes is roughly equal to the measurement errors of the data. For a typical error level of, say, 2 per cent on the impedance tensor elements, vertically incident plane waves and a relative dielectric permittivity $\epsilon_r = 5$, it is shown in Section 3.1 that the effect of displacement currents on the impedance phase is above the error level at, for example, $f = 15$ kHz and $\rho = 10\,000\ \Omega\text{ m}$ or $f = 170$ kHz and $\rho = 1000\ \Omega\text{ m}$.

In the following subsections, the existing knowledge of the dielectric effect on plane-wave and controlled-source frequency-domain electromagnetic (FDEM) responses is reviewed. With respect to the RMT method, the plane-wave FDEM responses are of special importance. After that, the resolvability of anomalous dielectric permittivities and previous attempts of quasi-static interpretation of high-frequency RMT data are discussed. In the last part of the introduction, we give an outlook at our 2-D inverse scheme for RMT data that allows for displacement currents and summarize the assumptions we make. Note that for the treatment of the FDEM theory, we choose an $\exp +i\omega t$ time dependence throughout this paper.

1.1 Dielectric effect on frequency-domain EM responses

Several publications describe the effect of displacement currents on plane-wave and controlled-source FDEM responses in the VLF and LF frequency ranges, based on analytic solutions by Wait (1953, 1970) and Wait & Nabulsi (1996) for a 1-D layered Earth.

In plane-wave FDEM methods like the RMT method, EM fields generated by powerful radio transmitters operating in the VLF and LF frequency ranges are used as primary signals. The aerials employed with the remote radio transmitters are vertical electric dipoles. At distances beyond several free-air wavelengths from the transmitter, that is, in the so-called far-field zone, the EM field essentially resembles that of a plane wave, which is obliquely incident on the Earth's surface (McNeill & Labson 1991). An excellent summary of the theory of plane-wave FDEM impedance, VMT and wave tilt measurements that covers both the quasi-static approximation and the general case with displacement currents, as well as the nature of the radio transmitter source field, is given by Crossley (1981).

For plane-wave EM fields, Sinha (1977) investigates the influence of displacement currents on the wave tilt, that is, the ratio of the horizontal to vertical electric field. On the surface of a homogeneous half-space, both amplitude and phase approach the values of the quasi-static approximation at low frequencies, although they become significantly smaller than the quasi-static responses with increasing frequency.

The dielectric effect on apparent resistivities and phases of radiomagnetotelluric surface impedances is deduced in Crossley (1981), Zacher (1992) and Persson & Pedersen (2002) from 1-D forward computations. On the surface of a homogeneous half-space, both apparent resistivity and phase are smaller than their constant counterparts in the quasi-static approximation. The differences become stronger with increasing frequency.

Wait (1953), Sinha (1977), Crossley (1981) and Song *et al.* (2002) emphasize the importance of the angle of incidence for wave tilt, surface impedance and VMT measurements conducted with plane-wave FDEM methods. The EM field is transmitted vertically into the Earth, independent of the angle of incidence, when the quasi-static approximation is valid. In the general case with displacement currents, however, the angles of incidence and transmission are related through Snell's law. As a consequence, the TM- and TE-mode impedances vary with the angle of incidence and differ at oblique incidence, even if measured on the surface of a layered half-space (Song *et al.* 2002).

For controlled source air-borne FDEM measurements, Fraser *et al.* (1990), Huang & Fraser (2002) and Yin & Hodges (2005) simulate responses due to a pair of horizontal coplanar transmitting–receiving coils, operating in the frequency range of 0.4 to 100 kHz. The ratio of secondary magnetic field intensity to primary magnetic field intensity is split into an in-phase component (real part) and a quadrature component (imaginary part). According to Fraser *et al.* (1990) and Huang & Fraser (2002), displacement currents in the Earth lead to a decrease of the in-phase component and an increase of the quadrature component, compared with the quasi-static case for which both components are positive. The influence of displacement currents in the air (an increase of both components) is rather small compared with that in the Earth (Yin & Hodges 2005).

1.2 Resolvability of permittivity anomalies

The resolvability of the relative dielectric permittivity from both plane-wave and controlled-source FDEM measurements is assessed by Nabulsi & Wait (1996), Stewart *et al.* (1994), Huang & Fraser (2002) and Persson & Pedersen (2002) with 1-D simulations.

Using obliquely incident plane waves in the VHF range (30–300 MHz), Nabulsi & Wait (1996) illustrate that a dielectric layer embedded in a highly resistive host is detectable if its thickness and relative permittivity are sufficiently high.

For a controlled source coil–coil FDEM method which operates in the MF (0.3–3 MHz) and HF (3–30 MHz) frequency ranges, Stewart *et al.* (1994) show that the anomalous response of both a resistive and conductive thin layer is significantly enlarged by the dielectric effect even if there is no contrast of dielectric permittivity between the layers of the model. Stewart *et al.* (1994) present two field examples, where tilt angle and ellipticity data of the magnetic field polarization ellipse have been successfully inverted for both electric resistivity and dielectric permittivity, with a 1-D inverse scheme.

At frequencies lower than those employed by Nabulsi & Wait (1996) and Stewart *et al.* (1994), displacement currents become weaker and the resolvability of permittivity anomalies within a limited range of possible relative permittivity values deteriorates. Huang & Fraser (2002) (see Section 1.1) estimate a single value of relative permittivity at their highest frequency of 100 kHz, as it is a badly resolved parameter at lower frequencies.

Persson & Pedersen (2002) invert RMT data with frequencies up to 250 kHz for dielectric permittivity, using 1-D models. The differences of inverse models are found to be negligible if the relative dielectric permittivities are limited to the range between 4 and 10, typical of bedrock, and if the resistivities are not larger than 20 000 Ω m (Persson & Pedersen 2002). Relative dielectric permittivities larger than 10 are typical of water bearing sedimentary rocks and soils (Reynolds 1997). Due to the high water content, such formations have relatively low resistivities (typically up to about 500 Ω m), which reduce the importance of displacement currents at VLF and LF frequencies. It is therefore sufficient, in many practical cases of RMT data interpretation, to account for displacement currents by selecting a dielectric permittivity representative of high-resistivity structures in the subsurface.

1.3 Quasi-static interpretation of high-frequency RMT data

The difficulties of the interpretation of high-frequency RMT field data in the quasi-static approximation are discussed by Persson & Pedersen (2002) and Linde & Pedersen (2004). For synthetic 1-D RMT impedance responses computed with displacement currents, Persson & Pedersen (2002) compare 1-D inversion results from inverse schemes that utilize both the quasi-static approximation and displacement currents. For a homogeneous half-space model, neglecting displacement currents during the inversion leads to an inverse model with a conductor close to the surface, followed by alternating layers of high and low resistivity at depth (Persson & Pedersen 2002). Similarly, Linde & Pedersen (2004) observe for quasi-static 1-D inversions of RMT field data from the island Åvrö, Sweden, that a conductive surface layer is modelled more conductive and the underlying unfractured bedrock is modelled more resistive than in the 1-D inversions with displacement currents. The models, due to inversion with displacement currents, are supported by logging data of Gentzschin *et al.* (1987).

In fact, the work presented by Linde & Pedersen (2004) is a typical example of the interpretation strategies chosen until now, in cases where the dielectric effect in RMT data is to be accounted for. In the absence of a 2-D inversion program that allows for displacement currents, the data interpretation has, so far, been restricted to 1-D inversions with modified analytic forward and Frechet derivative routines, the exclusion of the higher frequency data in 2-D inversions and 3-D forward modelling with the integral equation code X3D by Avdeev *et al.* (2002).

1.4 2-D inversion of RMT data allowing for displacement currents

For the first time, we take displacement currents in a 2-D forward and inverse modelling scheme for RMT data into account by selecting a value of dielectric permittivity that is typical of the subsurface and assuming vertically incident plane waves. As the EM field from remote VLF transmitters can be expected to be incident at an angle closer to 90° (grazing incidence), it is shown in Section 2.2 that the presence of a moderately resistive surface layer reduces the influence of the angle of incidence considerably. We investigate the effect of displacement currents on 2-D forward responses in the TM-mode, the TE-mode and the VMT and compare our results with the responses computed by the integral equation code X3D by Avdeev *et al.* (2002), which, at the time of writing, was the only forward code known to us that operates in two or three dimensions and includes displacement currents. Especially, the effect on VMT responses was not considered in the past (*cf.* Avdeev *et al.* 2002; Persson & Pedersen 2002). Possible misinterpretations, in the form of artefacts with excessively extreme resistivities in models from quasi-static inverse schemes, are highlighted. The RMT data from Åvrö (Linde & Pedersen 2004) are re-interpreted with the inverse scheme that allows for displacement currents. The resulting inverse models are compared with the borehole data of Gentzschin *et al.* (1987) and the seismic reflection model of Juhlin & Palm (1999).

We have added our forward and sensitivity routines, which allow for displacement currents, to the popular 2-D magnetotelluric inverse code REBOCC by Siripunvaraporn & Egbert (2000).

2 THEORY

2.1 Electromagnetic equations

Assuming a volume of conductivity σ , dielectric permittivity ϵ and vacuum permeability μ_0 , Maxwell's equations are written in the frequency domain as

$$\nabla \times \mathbf{E} = -(\mathrm{i}\omega\mu_0)\mathbf{H} = -\hat{\mathbf{z}}\mathbf{H} \quad \text{Faraday's law} \quad (1)$$

$$\nabla \times \mathbf{H} = (\sigma + \mathrm{i}\omega\epsilon)\mathbf{E} = \hat{\mathbf{y}}\mathbf{E} \quad \text{Ampere's law} \quad (2)$$

$$\nabla \cdot (\epsilon\mathbf{E}) = q \quad \text{Gauss' law} \quad (3)$$

$$\nabla \cdot \mathbf{H} = 0 \quad (4)$$

where $\mathbf{E}^{\mathrm{tot}}$ and $\mathbf{H}^{\mathrm{tot}}$ are the electric and magnetic field vectors, varying in time t , with angular frequency ω (e.g. Ward & Hohmann 1987) and q is the charge density. On the right-hand sides of eqs (1) and (2), the definitions of the impedivity $\hat{\mathbf{z}} = \mathrm{i}\omega\mu_0$ and admittivity $\hat{\mathbf{y}} = \sigma + \mathrm{i}\omega\epsilon$ are used. The quantities $\mathbf{j}_{\mathrm{cond}} = \sigma\mathbf{E}$, $\mathbf{j}_{\mathrm{disp}} = \mathrm{i}\omega\epsilon\mathbf{E}$, and $\mathbf{j} = \hat{\mathbf{y}}\mathbf{E}$ are the conduction, displacement and total current densities, respectively. The

displacement current density $i\omega\epsilon\mathbf{E}$ describes the dielectric effect due to electronic, atomic, molecular and space charge derived polarization of matter with dielectric permittivity ϵ in the presence of a time-varying electric field (Keller 1987). In the case that conduction currents dominate over displacement currents (i.e. $\sigma \gg \omega\epsilon$), displacement currents may be neglected in eq. (2). This simplification is known as the quasi-static approximation.

In the following, it is assumed that plane waves are obliquely incident on the Earth's surface in the y - z plane and that the x -direction is the geoelectrical strike direction. Therefore, the admittivity $\hat{y}$ and the EM field components vary only in y and z direction. This choice leads to the definition of the transverse electric (TE) and transverse magnetic (TM) modes for which the vertical electrical and vertical magnetic field components, respectively, vanish. The sets of equations for the TE- and TM-modes are

(1) TE-mode:

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \hat{y} E_x, \quad (5)$$

$$\frac{\partial E_x}{\partial z} = -\hat{z} H_y, \quad (6)$$

$$\frac{\partial E_x}{\partial y} = \hat{z} H_z. \quad (7)$$

(2) TM-mode:

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\hat{z} H_x, \quad (8)$$

$$\frac{\partial H_x}{\partial z} = \hat{y} E_y, \quad (9)$$

$$\frac{\partial H_x}{\partial y} = -\hat{y} E_z. \quad (10)$$

An illustration of the EM field components of the TM-mode and a 2-D subsurface with a cylindrical structure of anomalous electrical properties and infinite extension along the x -axis, that is, the strike direction, is given in Fig. 1. The EM field is obliquely incident at an angle θ_0 , thereby having a wavenumber vector $\mathbf{k}_0 = (0, k_{0y}, k_{0z})$. According to the definition of the TM-mode, the incident, reflected and transmitted magnetic fields $\mathbf{H}_i = (H_{ix}, 0, 0)$, $\mathbf{H}_r = (H_{rx}, 0, 0)$, and $\mathbf{H}_t = (H_{tx}, 0, 0)$, respectively, are all directed along the strike direction whereas the incident, reflected and transmitted electric fields $\mathbf{E}_i = (0, E_{iy}, E_{iz})$, $\mathbf{E}_r = (0, E_{ry}, E_{rz})$, and $\mathbf{E}_t = (0, E_{ty}, E_{tz})$, respectively, are all directed perpendicularly to the strike direction.

In the quasi-static approximation of the TM-mode, $\mathbf{j} = (0, \sigma E_y, \sigma E_z)$ vanishes in the air half-space (Brewitt-Taylor & Weaver 1976) where $\sigma_{\text{air}} = 0$ is assumed. As a consequence of eqs (9) and (10), H_x is then constant in the air half-space, and an inclusion of the air half-space in the modelling domain can be omitted. If displacement currents are accounted for, the magnetic field in the air is no longer independent of the resistivity distribution in the Earth, as the vertical component of the current density is continuous at the air-Earth interface

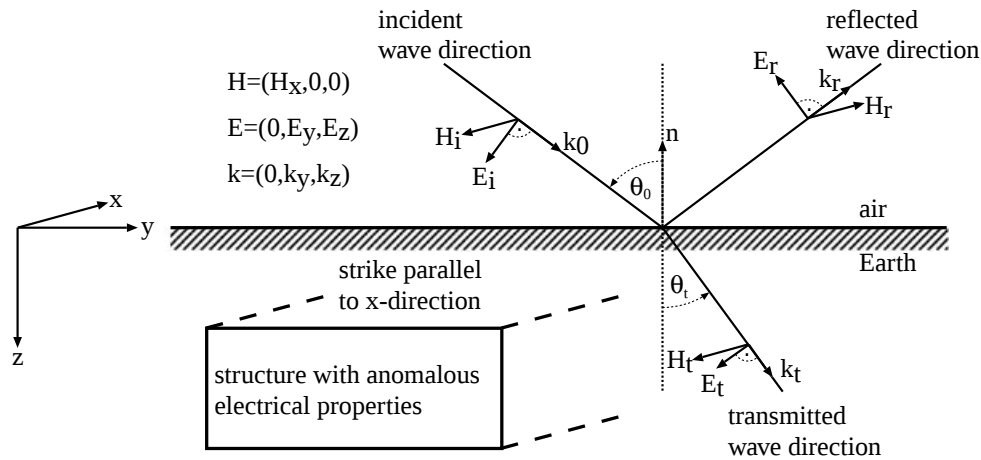


Figure 1. EM field components of the TM-mode on a 2-D earth model. The model consists of a structure with anomalous electrical properties that has its strike direction parallel to the x -axis. The EM field is obliquely incident at an angle θ_0 , thereby having a wavenumber vector $\mathbf{k}_0 = (0, k_{0y}, k_{0z}) = (0, k_0 \sin \theta_0, k_0 \cos \theta_0)$. The incident, reflected and transmitted magnetic fields $\mathbf{H}_i = (H_{ix}, 0, 0)$, $\mathbf{H}_r = (H_{rx}, 0, 0)$ and $\mathbf{H}_t = (H_{tx}, 0, 0)$, respectively, are all directed along the strike direction. The incident, reflected and transmitted electric fields $\mathbf{E}_i = (0, E_{iy}, E_{iz})$, $\mathbf{E}_r = (0, E_{ry}, E_{rz})$, and $\mathbf{E}_t = (0, E_{ty}, E_{tz})$, respectively, are all directed perpendicular to the strike direction. On top of a conductive subsurface, the electromagnetic field is refracted towards the normal, that is, $\theta_t < \theta_0$.

and in the air,

$$j_y = \hat{y} E_y = i\omega\epsilon_0 E_y = \frac{\partial H_x}{\partial z}, \quad (11)$$

$$j_z = \hat{y} E_z = i\omega\epsilon_0 E_z = -\frac{\partial H_x}{\partial y} \quad (12)$$

differ from zero (cmp. eqs 9 and 10). Hence, the air half-space must be included in the simulation of the TM-mode.

The electric and magnetic field components E_x and H_x of the TE- and TM-modes, respectively, fulfil the Helmholtz equations (cf. eqs 1 and 2)

$$\begin{aligned} -(\nabla \times \nabla \times \mathbf{E})_x &= -[\nabla(\nabla \cdot \mathbf{E})]_x + (\nabla^2 \mathbf{E})_x \\ &= \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} \\ &= \hat{z} \hat{y} E_x \end{aligned} \quad (13)$$

and

$$\begin{aligned} -\left(\nabla \times \frac{1}{\hat{y}} \nabla \times \mathbf{H}\right)_x &= -\frac{1}{\hat{y}} [\nabla(\nabla \cdot \mathbf{H})]_x + \frac{1}{\hat{y}} \nabla^2 H_x - \left[\left(\nabla \frac{1}{\hat{y}}\right) \times \nabla \times \mathbf{H}\right]_x \\ &= \frac{1}{\hat{y}} \left(\frac{\partial^2 H_x}{\partial y^2} + \frac{\partial^2 H_x}{\partial z^2}\right) + \frac{\partial}{\partial y} \left(\frac{1}{\hat{y}}\right) \cdot \frac{\partial H_x}{\partial y} + \frac{\partial}{\partial z} \left(\frac{1}{\hat{y}}\right) \cdot \frac{\partial H_x}{\partial z} \\ &= \hat{z} H_x, \end{aligned} \quad (14)$$

where the 2-D assumption $\partial/\partial x = 0$ and eq. (4) were used.

In a homogeneous volume, for instance, the general solution of the scalar Helmholtz equations (eqs 13 and 14) is given by

$$\begin{aligned} \{E, H\}_x &= (\{E, H\}_x^{++} e^{-ik_z z} + \{E, H\}_x^{--} e^{ik_z z}) e^{-ik_y y} \\ &\quad + (\{E, H\}_x^{+-} e^{-ik_z z} + \{E, H\}_x^{-+} e^{ik_z z}) e^{+ik_y y}. \end{aligned} \quad (15)$$

Here, k_y and k_z are the horizontal and vertical components of the wavenumber vector $\mathbf{k}$ (see above). The substitution of eq. (15) into eq. (13) yields

$$k^2 = k_y^2 + k_z^2 = -\hat{z} \hat{y} \quad (16)$$

where $k_y = k \sin \theta$ and $k_z = k \cos \theta$. The complex wavenumber can be split as $k = \alpha - i\beta$ where the real numbers α and β represent propagation and attenuation, respectively, and

$$\alpha = \omega \sqrt{\frac{\mu_0 \epsilon}{2} \left[\sqrt{1 + \frac{\sigma^2}{\epsilon^2 \omega^2}} + 1 \right]}, \quad (17)$$

$$\beta = \omega \sqrt{\frac{\mu_0 \epsilon}{2} \left[\sqrt{1 + \frac{\sigma^2}{\epsilon^2 \omega^2}} - 1 \right]}. \quad (18)$$

The inverse of the imaginary part gives the skin depth $\delta = \frac{1}{\beta}$ over which the amplitude of the EM field is reduced by a factor $1/e$. In the quasi-static approximation, the real and imaginary parts are equal, that is, $\alpha = \beta = \sqrt{\frac{\omega \mu_0 \sigma}{2}}$.

The reflection and refraction of plane EM waves at the Earth's surface are governed by Snell's law and the Fresnel equations (Ward & Hohmann 1987). Hence, the EM field measured on the Earth's surface depends on the angle of incidence (see Fig. 1). Three cases of the angle of incidence θ_0 are distinguished. The cases $\theta_0 = 0^\circ$ and $\theta_0 = \pm 90^\circ$ are known as normal (or vertical) incidence and grazing (or parallel) incidence, respectively. The cases $90^\circ > \theta_0 > 0^\circ$ and $0^\circ > \theta_0 > -90^\circ$ are called oblique incidence. The refraction of obliquely incident EM waves into the subsurface is conveniently demonstrated for a layered half-space. As a consequence of the boundary conditions for the EM field components at layer interfaces, the horizontal component of the wavenumber vector is constant (Ward & Hohmann 1987), that is,

$$k_{y,j} = k_0 \sin \theta_0 = k_j \sin \theta_j. \quad (19)$$

Here, $k_0 = \sqrt{\omega^2 \mu_0 \epsilon_0}$ is the wavenumber of the air and θ_0 is the angle of incidence. Similarly, k_j and θ_j are the wavenumber and angle of transmission of the j th layer, respectively. According to eqs (16) and (19) the vertical wavenumber of the j th layer has the form

$$\begin{aligned} k_{z,j} &= k_j \cos \theta_j = k_j \sqrt{1 - \sin^2 \theta_j} \\ &= k_j \sqrt{1 - \frac{k_0^2}{k_j^2} \sin^2 \theta_0}. \end{aligned} \quad (20)$$

At sufficiently low frequencies, that is, when the quasi-static approximation is valid, $k_0^2/k_j^2 = \omega^2 \mu_0 \epsilon_0 / (\omega^2 \mu_0 \epsilon_j - i\omega \mu_0 \sigma_j) \rightarrow 0$ as $\omega \rightarrow 0$ and $k_{z,j} \cong k_j$. Hence, it is only in the quasi-static approximation or at vertical incidence that the EM field is transmitted vertically into the Earth. At high frequencies and oblique incidence, the angle of transmission generally deviates from 0° .

After solving the Helmholtz equations (eqs 13 and 14) for E_x or H_x of a 2-D conductivity distribution, the auxiliary fields H_y and H_z or E_y and E_z can be computed with eqs (6) and (7) or eqs (9) and (10), respectively.

The off-diagonal elements of the complex 2-D impedance tensor relate the horizontal magnetic fields to the horizontal electric fields of the TE- and TM-mode as

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} 0 & Z_{xy} \\ Z_{yx} & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad (21)$$

and yield the responses commonly used in radiomagnetotellurics, that is, the apparent resistivities

$$\rho_a^{xy} = \frac{1}{\omega \mu_0} |Z_{xy}|^2 \quad \text{and} \quad \rho_a^{yx} = \frac{1}{\omega \mu_0} |Z_{yx}|^2 \quad (22)$$

and phases

$$\phi^{xy} = \arg(Z_{xy}) \quad \text{and} \quad \phi^{yx} = \arg(Z_{yx}). \quad (23)$$

Ward *et al.* (1968) establish a more direct link to the electrical properties of the subsurface, in the general case with displacement currents, by defining an apparent conductivity and an apparent dielectric permittivity.

Due to the dependence of the EM field on the angle of incidence, the amplitude and phase of the impedances of the TM- and TE-mode differ even if measured on the surface of a layered Earth. Only if the quasi-static approximation is valid or if the EM field is vertically incident, the TE- and TM-mode impedances of a layered half-space satisfy the relationship $Z_{xy} = -Z_{yx}$.

For plane waves vertically incident on the surface of a homogeneous half-space with impedivity $\hat{z}$ and admittivity $\hat{y}$, the TM-mode impedance has the form $Z_{yx} = \sqrt{\hat{z}/\hat{y}}$ (Wait 1970; Ward & Hohmann 1987). In the quasi-static approximation, the latter expression simplifies to $Z_{yx} = \sqrt{i\omega\mu_0/\sigma}$, and only in this case, the apparent resistivities and phases measured on a homogeneous half-space equal the resistivity of the half-space and 45° , respectively.

In the TE-mode, the vertical magnetic field H_z is related to the horizontal magnetic field H_y through the complex 2-D VMT B :

$$H_z = B H_y. \quad (24)$$

For plane waves obliquely incident on a layered Earth, the VMT generally differs from zero. However, for vertically incident plane waves or in the quasi-static approximation, a VMT that differs from zero is only observed if the admittivity $\hat{y}$ varies laterally (see eqs 5–7).

2.2 Normal and oblique incidence

In the case of grazing or oblique incidence, both the incident electric and the incident magnetic fields can have vertical components (see Fig. 1). Already for a 1-D earth model, the TE- and TM-mode are then defined, by demanding that either the electric or the magnetic field be perpendicular to the plane of incidence (Wait 1970; Ward & Hohmann 1987), and the impedance tensor and VMT measured on the Earth's surface depend on the angle of incidence (see Section 2.1). It is therefore important to appraise the error made by assuming vertical incidence during the modelling process. For a layered earth model, the deviations of the TE- and TM-mode impedance amplitudes and phases at an arbitrary angle of incidence from those at normal incidence can be estimated with well-known recurrence formulae (see e.g. Wait 1953, 1970; Crossley 1981; Ward & Hohmann 1987; Song *et al.* 2002).

For the half-space model shown in Fig. 2(a), consisting of two layers with resistivities of 600 and 30 000 Ω m and layer thicknesses of 25 and 75 m, a confining half-space with a resistivity of 600 Ω m and a constant relative permittivity $\epsilon_r = 6$, the deviations of the amplitude and phase of the TM- and TE-mode impedances from their respective values at normal incidence are shown in Fig. 2. The maximal deviations of 1.5 per cent and 1° for the amplitude and phase, respectively, occurring at parallel incidence, are of the order of typically expected error levels. A similar model that consists of the uppermost layer underlain by a confining half-space of 30 000 Ω m shows maximal deviations of 1.0 per cent and 0.25° , respectively, indicating that a considerable part of the distortion in the first case is due to the reflection of the EM energy on the top of the confining half-space.

The angle of incidence can be estimated with the scheme by Song *et al.* (2002), which requires that the horizontal EM field components are measured simultaneously at adjacent receiver sites. In a typical RMT field campaign, however, a single receiver is moved along the profile. The interpretation is further complicated, as the EM fields of different transmitters, with frequencies close to a nominal frequency, are used to estimate the TM- and TE-mode impedances (Bastani & Pedersen 2001). Generally, the transmitters are off the profile or strike direction and have different angles of incidence; but the angle of incidence, normally, is close to 90° (grazing incidence) at the site of investigation (Crossley 1981).

As the aeriels employed by the remote radio transmitters, typically, are vertical electric dipoles, the incident EM field is that of a TM-mode. Hence, the definitions of TE- and TM-mode based on the geoelectrical structure of the subsurface and on the nature of the incident field are conciliable only for the TM-mode, given that the direction to the remote radio transmitter coincides with the profile direction (as in Fig. 1 for instance). If the transmitter was located off the profile direction, the wavenumber vector $\mathbf{k}$ would have an x -component, which, in the general case, would persist within the Earth and invalidate the 2-D assumption $\partial/\partial x = 0$. However, even this problem is amended if a

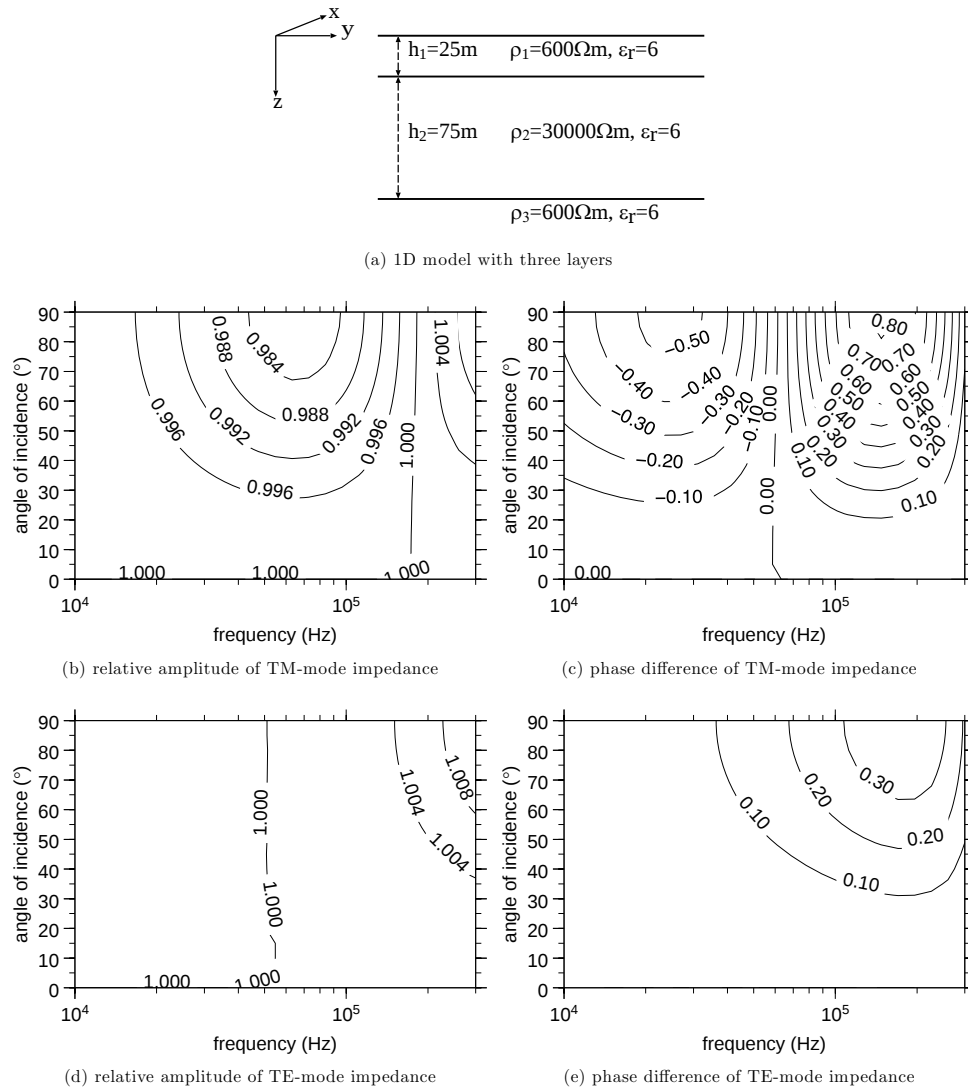


Figure 2. Relative amplitude and phase difference for the TM- and TE-mode impedances with respect to the case of normal incidence for angles of incidence between 0° (normal incidence) and 90° (grazing incidence) and at frequencies between 10 and 300 kHz (panels b–e). The earth is assumed to consist of two layers with resistivities of 600 and 30 000 Ω m and layer thicknesses of 25 and 75 m, respectively, a confining half-space with a resistivity of 600 Ω m and a constant relative permittivity $\epsilon_r = 6$ (panel a). The deviations from the impedance values at normal incidence are largest at grazing incidence.

moderately resistive or conductive surface layer is present, as the EM field is then transmitted almost vertically into the subsurface, and the definition of different modes can be based on the geoelectrical structure.

We consider only vertically incident plane-wave fields. As the above example shows, the presence of a moderately resistive or conductive near-surface layer reduces the importance of the angle of incidence, and deviations of the responses for different angles of incidence are then rather small.

2.3 Computation of forward responses and sensitivities

The forward problem, that is, the computation of responses for a given model, is solved by discretizing the modelling domain with the finite-difference approximation (FDA), following Hohmann (1987) and Aprea *et al.* (1997). The derivations of the FDAs for the TE- and TM-modes can be found in Appendix A. Both direct and iterative solvers for the system of linear equations, arising from the FDA of the forward problem, are discussed in Appendix B. As we have not yet managed to implement an appropriate iterative solver, we rely on the LU-decomposition (also known as Gaussian elimination) by Anderson *et al.* (1999).

The sensitivity matrix $\mathbf{J} \in \mathbb{R}^{N \times M}$ describes the perturbations ensuing for N forward responses $\mathbf{F}[\mathbf{m}] \in \mathbb{R}^N$ due to perturbations of M model parameters $\mathbf{m} \in \mathbb{R}^M$. The entry of the sensitivity matrix for the k th datum with respect to the l th model parameter is then calculated as a partial derivative:

$$J^{kl}(\mathbf{m}) = \frac{\partial F^k[\mathbf{m}]}{\partial m^l}. \quad (25)$$

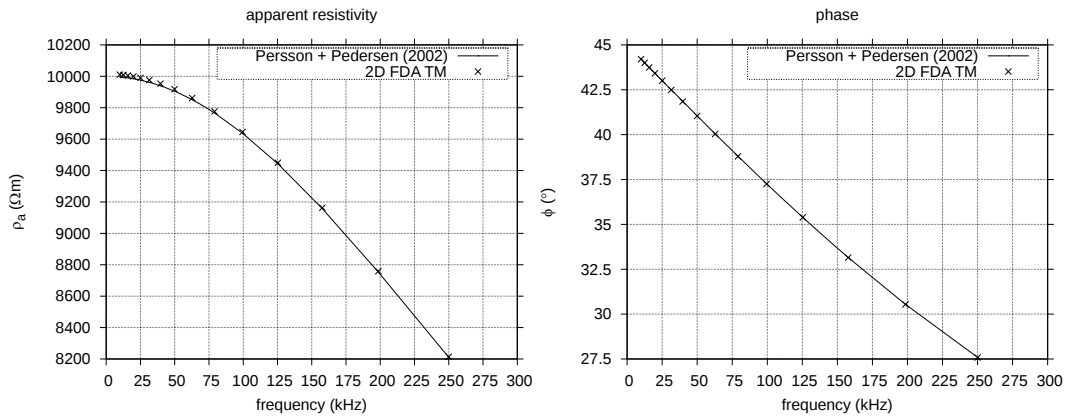


Figure 3. Analytic 1-D solutions by Persson & Pedersen (2002) and 2-D FDA solutions of apparent resistivity ρ_a and phase ϕ for the TM-mode on the surface of a homogeneous half-space with $\rho = 10\,000\ \Omega\text{ m}$ and $\epsilon_r = 5$. The responses were computed for frequencies between 10 and 250 kHz and under the assumption of normal incidence. At high frequencies, ρ_a and ϕ are both significantly smaller than their quasi-static values of $10\,000\ \Omega\text{ m}$ and 45° , respectively. With decreasing frequency, ρ_a and ϕ approach their quasi-static values asymptotically.

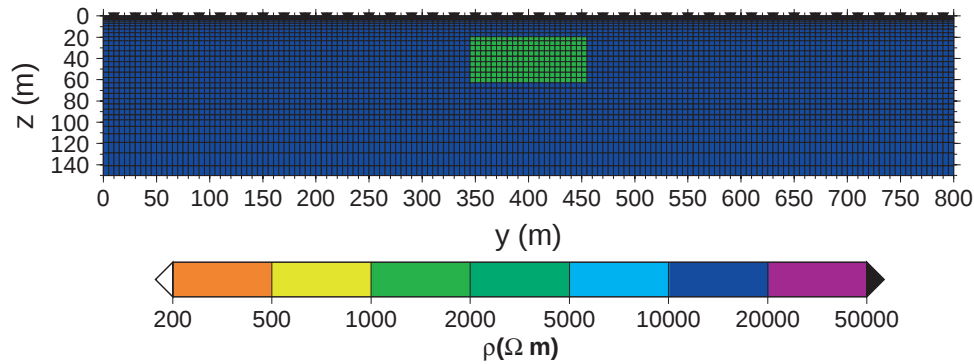


Figure 4. Simple 2-D model with a conductive block of $\rho = 1000\ \Omega\text{ m}$ in a half-space with a resistivity of $\rho = 10\,000\ \Omega\text{ m}$ and constant relative dielectric permittivity $\epsilon_r = 5$. Receiver positions are indicated by black triangles. The TM-mode, TE-mode and VMT responses of this model are shown in Fig. 5.

The entries of the sensitivity matrix are typically given for the logarithms of the apparent resistivities and the phases of the impedance tensor elements and the real and imaginary parts of the VMT with respect to the logarithms of the cell resistivities. The logarithms are typically chosen relative to the base 10. The sensitivity matrix is computed with the scheme by Rodi (1976) and depends on the FDA of the forward problem. Further information on this algorithm is given in Appendix C. An example of sensitivity matrix entries is given at the end of Section 3.2.

2.4 Mesh design

To obtain accurate modelling results, the total extent of the modelling domain (i.e. the finite-difference mesh) and the sizes of individual cells of the finite-difference mesh need to be well adapted to the settings of the experiment, that is, the length of the profile on which measurements were conducted, the lowest and highest frequencies of the measurements and the distributions of electrical conductivity and dielectric permittivity present in the model.

The horizontal and, below the air–Earth interface, the vertical extents of the finite difference mesh must be larger than those used in the quasi-static approximation, as the skin depth $\delta = \frac{1}{\beta}$ computed with displacement currents (see eq. 18) is larger than its quasi-static counterpart.

Furthermore, the node spacing must be small compared with the scale lengths across which the EM fields vary, that is, the inverse real and imaginary parts of the complex wavenumber k . In the quasi-static approximation, this leads to the well-known requirement that the node spacing must be small compared with the local skin depth (Aprea *et al.* 1997). In the general case, $1/\alpha < 1/\beta$ and the local node spacing must be considerably smaller than $1/\alpha$.

A small vertical node spacing is essentially important for the air half-space since the vertically incident plane wavefield propagates undamped (assuming $\sigma_{\text{air}} = 0\ \text{S m}^{-1}$). In the air, the largest vertical mesh cell dimension must be smaller than $1/\alpha$ of the highest frequency. This results in the following comparison. In the REBOCC inverse scheme (Siripunvaraporn & Egbert 2000), the conductivity of the air half-space is assumed to be $\sigma_{\text{air}} = 10^{-10}\ \text{S m}^{-1}$, and the quasi-static skin depth at a frequency of 300 kHz is 92 km. In the general case with displacement currents, $\sigma_{\text{air}} = 0\ \text{S m}^{-1}$ and the inverse real part of the wavenumber is $1/\alpha = 159\ \text{m}$ for $f = 300\ \text{kHz}$. In the former case, the

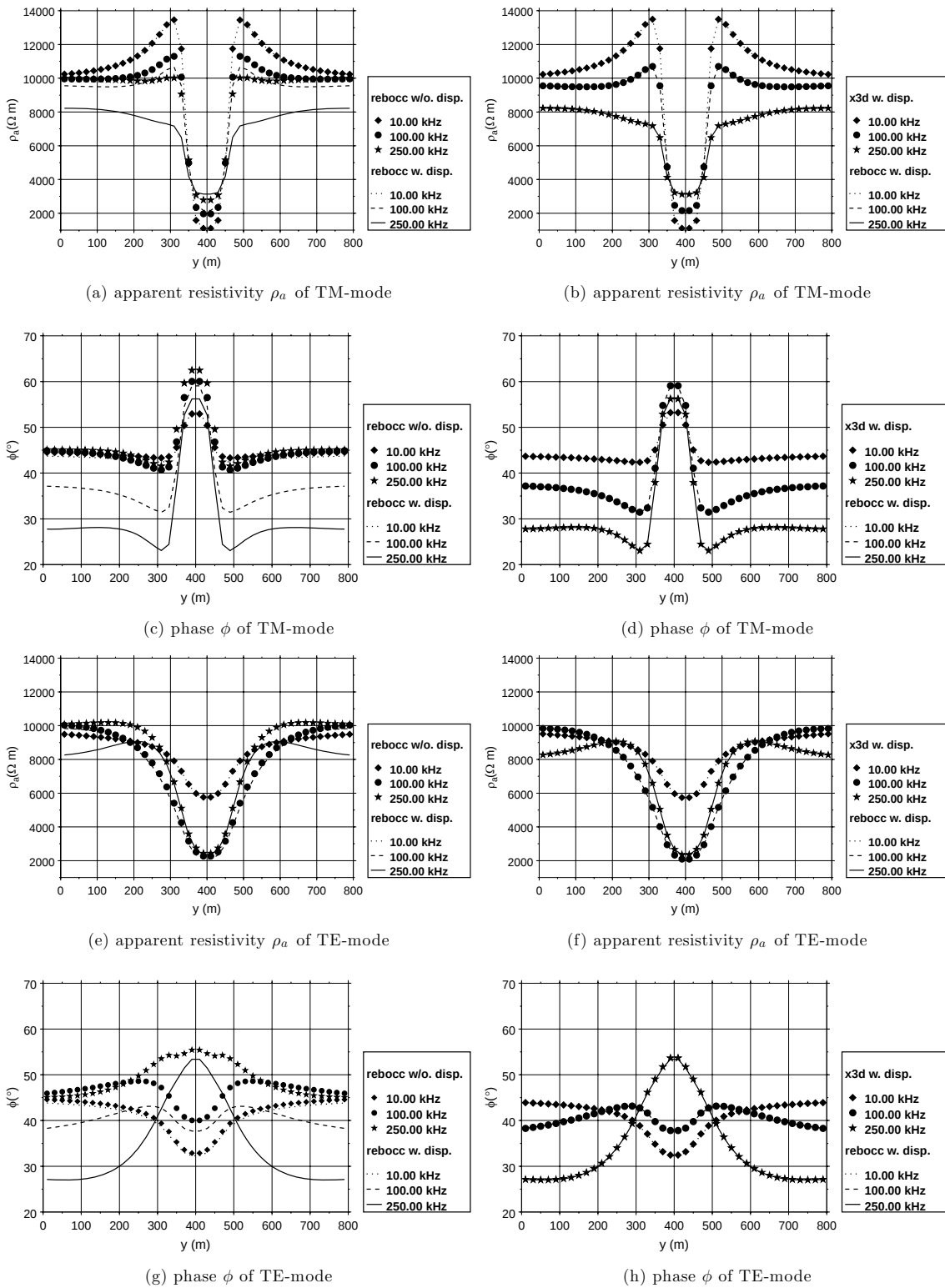


Figure 5. Comparison of 2-D FDA forward responses of the block model shown in Fig. 4 computed with displacement currents (shown as lines in both the left- and right-hand columns) with 2-D FDA solutions computed in the quasi-static approximation (shown as symbols in the left-hand column) and 3-D integral equation solutions computed with displacement currents (shown as symbols in the right-hand column). Panels (a)–(d) show the responses for the TM-mode apparent resistivity and phase, respectively. Panels (e)–(h) show the responses for the TE-mode apparent resistivity and phase, respectively. Panels (i)–(l) show the responses for the real and imaginary part of the VMT, respectively. The TM-mode and TE-mode responses computed with displacement currents are generally smaller than those computed in the quasi-static approximation, especially to the sides of the conductive block. The real part of the VMT response computed with displacement currents shows distinct sign reversals [marked by labels (2) and (6) in panel i] to the sides of the conductive block. The corresponding maximum and minimum are marked by labels (1) and (7), respectively, in panel (i). The 2-D FDA and integral equation solutions are in good agreement (right-hand column).

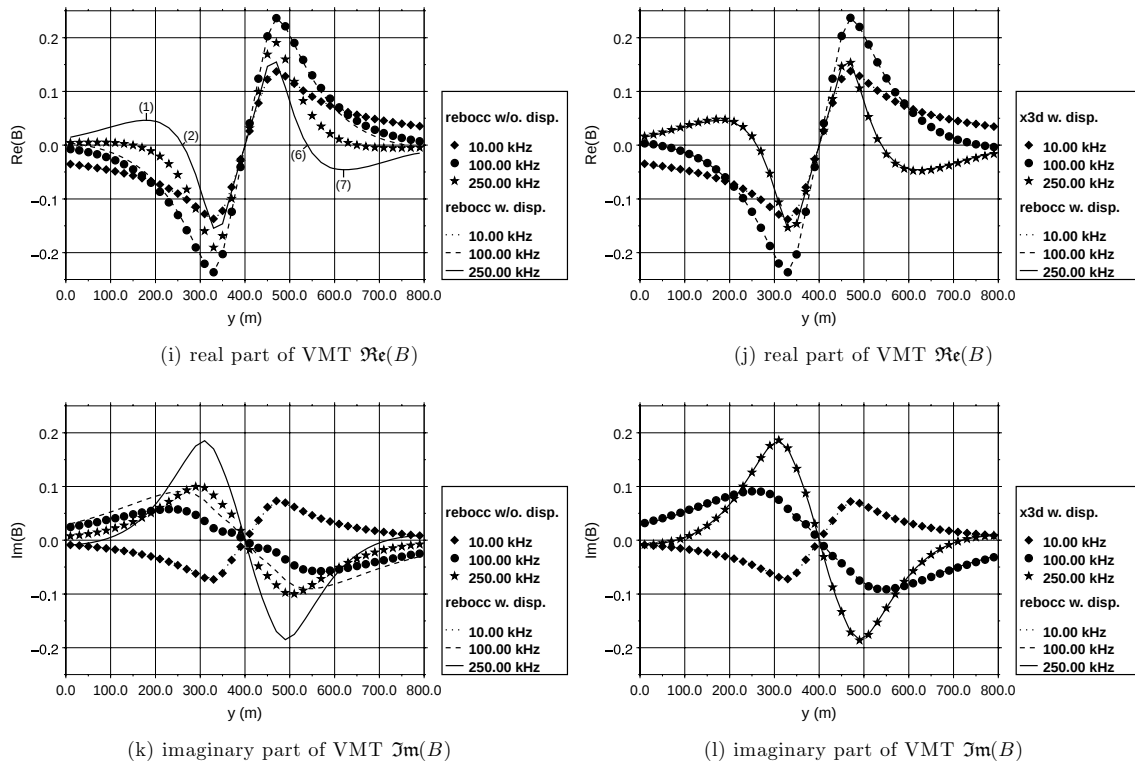


Figure 5. (Continued.)

skin depth exceeds the size of the modelling domain by far, and it is therefore appropriate to address the primary field of the quasi-static case as a uniform inducing field rather than a plane wave incident on the Earth's surface.

3 SYNTHETIC EXAMPLES

3.1 Forward modelling examples

We consider a forward modelling example for a homogeneous half-space, with a resistivity of $10\,000\,\Omega\text{m}$ and a relative permittivity $\epsilon_r = 5$. Assuming vertically incident plane waves, analytic 1-D solutions with the algorithm by Persson & Pedersen (2002) and 2-D FDA solutions were computed for the apparent resistivities and phases of the TM- (Fig. 3) and TE-mode (not shown) at frequencies between 10 and 250 kHz. The comparison of the analytic 1-D solution (marked by a solid line) and the 2-D FDA solution (marked by crosses) shows excellent agreement. At high frequencies, the effect of displacement currents is to decrease the apparent resistivity and phase below the apparent resistivity of $10\,000\,\Omega\text{m}$ and phase of 45° , respectively, typical of the quasi-static approximation. For a typical error level of 2 per cent on the impedance, the deviations from the quasi-static values are as large as the given errors at 105 kHz for the apparent resistivity and 15 kHz for the phase.

For the simple 2-D model with a block of $\rho = 1000\,\Omega\text{m}$ in a half-space, with a resistivity of $\rho = 10\,000\,\Omega\text{m}$ and $\epsilon_r = 5$ throughout, shown in Fig. 4, 2-D FDA forward responses with displacement currents are compared with both 2-D FDA forward responses for the quasi-static approximation and the 3-D integral equation solution by Avdeev *et al.* (2002). Responses were computed for the TM-mode impedance, the TE-mode impedance and the VMT. Fig. 5 shows the 2-D FDA forward responses, computed with and without displacement currents in the left-hand column, and the comparison of 2-D FDA forward responses, computed with displacement currents, and 3-D integral equation solutions, with displacement currents, in the right-hand column. The latter comparison indicates that the finite-difference forward scheme is rather accurate. For the given mesh discretization, the relative deviations between the impedance responses of the FDA and integral equation solutions are below 3.0 per cent. The absolute deviations between the VMT responses of the FDA and integral equation solutions are below 0.003. As errors in the computation of two field components might cancel when taking their ratio, a further comparison was done for the 2-D FDA and 3-D integral equation solutions of individual field components (not shown). After an appropriate normalization, the scaled complex field components of the TE-mode deviate by less than 0.7 per cent, whereas the field components of the TM-mode differ by as much as 3.0 per cent. As we do not have insight into the code by Avdeev *et al.* (2002), it is difficult to give an explanation for the discordance in the latter case.

For the lowest frequency of about 10 kHz, the responses computed with (dotted lines in left-hand column of Fig. 5) and without displacement currents (diamond symbols in left-hand column of Fig. 5) are very similar. At 100 kHz (dashed lines and filled circle symbols)

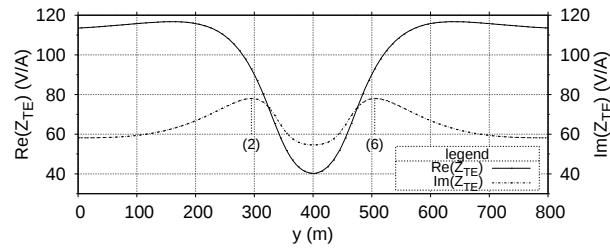
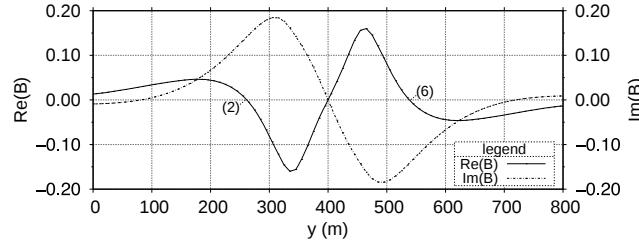
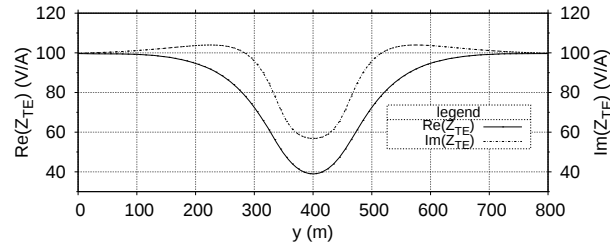
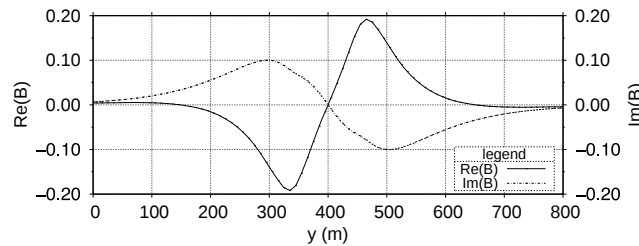
(a) $\Re(Z_{TE})$ and $\Im(Z_{TE})$ in the general case with displacement currents(b) $\Re(B)$ and $\Im(B)$ in the general case with displacement currents(c) $\Re(Z_{TE})$ and $\Im(Z_{TE})$ in the quasi-static case(d) $\Re(B)$ and $\Im(B)$ in the quasi-static case

Figure 6. The TE-mode impedance Z_{TE} and the VMT B for the model shown in Fig. 4 at $f = 250$ kHz, in the general case with displacement currents (panels a and b) and the quasi-static case (panels c and d). The VMT and the TE-mode impedance roughly follow the relations $\Re(\frac{H_z}{H_y}) \propto \Im(\frac{\partial Z_{xy}}{\partial y})$ and $\Im(\frac{H_z}{H_y}) \propto -\Re(\frac{\partial Z_{xy}}{\partial y})$. Hence, for instance, zero transitions of the real part of the VMT are observed at approximately the same positions where the imaginary part of the impedance has minima or maxima. The labels (2) and (6) mark two such pairs of zero transitions in the real part of the VMT and maxima of the imaginary part of the impedance.

and 250 kHz (solid lines and star symbols), the influence of displacement currents is considerable, given the chosen resistivity distribution and relative dielectric permittivity.

For stations located on the sides of the conductive block, the effect of displacement currents on TE- and TM-mode impedances is most obvious. Towards the left- and right-hand edges of the mesh, the apparent resistivities and phases approach those of the corresponding homogeneous half-space (see Fig. 3). Also at sites above the conductive block, apparent resistivity and phase are generally smaller than in the quasi-static approximation.

An important effect of displacement currents on the real and imaginary parts of the VMT at high frequencies is the occurrence of lateral sign reversals, located symmetrically around the conductive block. For $f = 250$ kHz, lateral sign reversals are shown at 260 and 540 m along the profile in the real part of the VMT [marked by labels (2) and (6), respectively, on the solid line in Fig. 5i] and at 75 and 725 m along

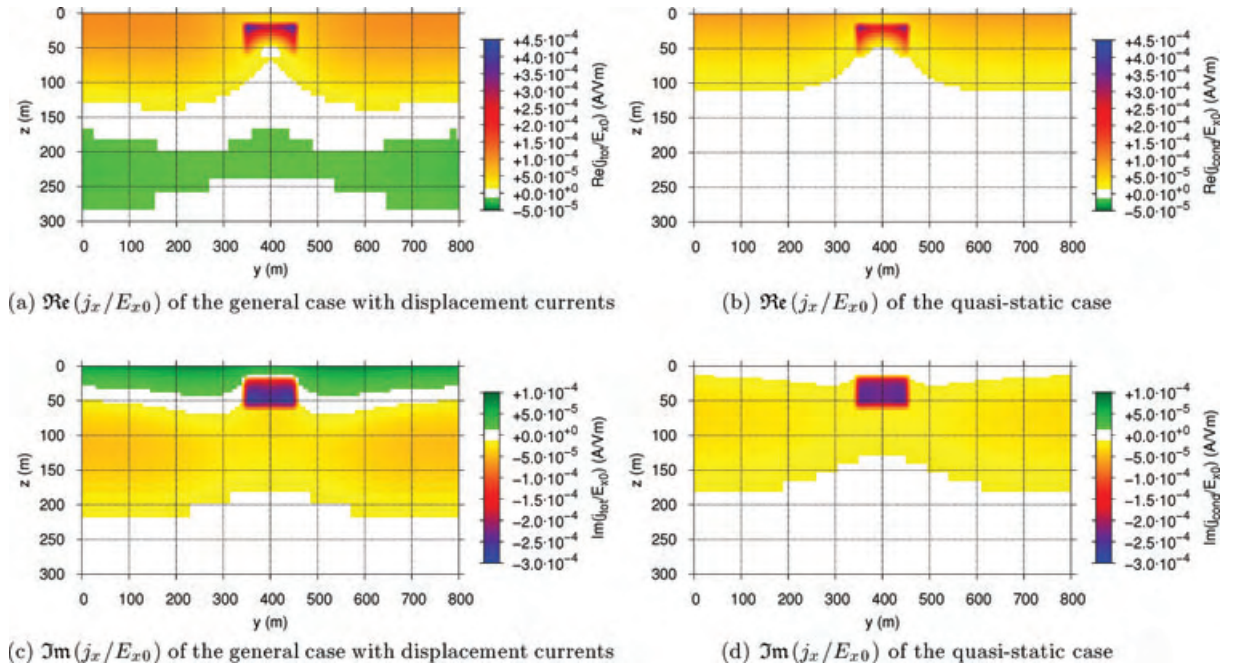


Figure 7. Real (upper row) and imaginary (lower row) parts of the normalized current density j_x/E_{x0} at $f = 250$ kHz of the TE-mode for the general case with displacement currents (left-hand column) and the quasi-static case (right-hand column) for the model shown in Fig. 4. Normalized current densities, which are close to $0 \text{ A V}^{-1} \text{ m}^{-1}$, are plotted in white. Different colourscales were used for the real and imaginary parts. In the general case with displacement currents, the current system penetrates deeper into the subsurface than in the quasi-static case.

the profile in the imaginary part of the VMT (solid line in Fig. 5k). In addition to the lateral sign reversals, the real part of the VMT has a maximum at $y = 180 \text{ m}$ [marked by label (1) in Fig. 5i] and a minimum at $y = 620 \text{ m}$ [marked by label (7) in Fig. 5i]. The responses at the maximum and minimum are $|\Re(B)| = 0.05$. Sign reversals to the sides of the conductive block can also be observed in the real part of the quasi-static response at $y = 150 \text{ m}$ and $y = 650 \text{ m}$ (star symbols in Fig. 5i). However, the quasi-static response is comparatively small at sites further away from the block (no larger than $|\Re(B)| = 0.006$) and would most likely be masked by noise effects (a typical absolute error is e.g. $\Delta \Re(B) \approx 0.01$) if measured in the field. In the general case with displacement currents, the deduction of the horizontal centre of conductive structures from the positions of zero transitions of the VMT B becomes intricate in more complex geological settings. Artefacts might be introduced to inverse models in a quasi-static interpretation.

It is instructive to relate the lateral sign reversals of the VMT to the gradient of the TE-mode impedance Z_{xy} by considering eqs (7) and (21):

$$\hat{z}H_z = \frac{\partial E_x}{\partial y} = \frac{\partial}{\partial y}(Z_{xy}H_y).$$

For small deviations of H_y from its normal field component H_y^n , that is, the H_y -component of the corresponding 1-D model without the conductive block, this yields

$$\frac{H_z}{H_y} \approx \frac{H_z}{H_y^n} = \frac{1}{\hat{z}} \frac{\partial Z_{xy}}{\partial y} = -\frac{i}{\omega\mu_0} \frac{\partial Z_{xy}}{\partial y}, \quad (26)$$

which corresponds to the following relationships for the real and imaginary parts of the VMT

$$\begin{aligned} \Re\left(\frac{H_z}{H_y}\right) &\propto -i \cdot i \cdot \Im\left(\frac{\partial Z_{xy}}{\partial y}\right) = \Im\left(\frac{\partial Z_{xy}}{\partial y}\right), \\ \Im\left(\frac{H_z}{H_y}\right) &\propto -\Re\left(\frac{\partial Z_{xy}}{\partial y}\right). \end{aligned}$$

In the synthetic example for $f = 250$ kHz, the variation of H_y away from its approximate 1-D values at the beginning and end of the profile is less than 22 per cent in the quasi-static case (not shown) and less than 12 per cent in the general case (not shown). The real and imaginary parts of the VMT are in good agreement with the expected variation with the lateral derivative of the TE-mode impedance Z_{xy} for the quasi-static (Figs 6c and d) and general cases (Figs 6a and b). For the general case, the positions of the lateral sign reversals in the real part of the VMT and the corresponding maxima in the imaginary part of the impedance are marked with the labels (2) and (6) in Figs 6(b) and (a), respectively. The zero transitions of the VMT are somewhat shifted from their predicted positions, where impedance maxima are less distinct. This disagreement is related to the fact that the assumption of small deviations of H_y from its normal component is slightly violated.

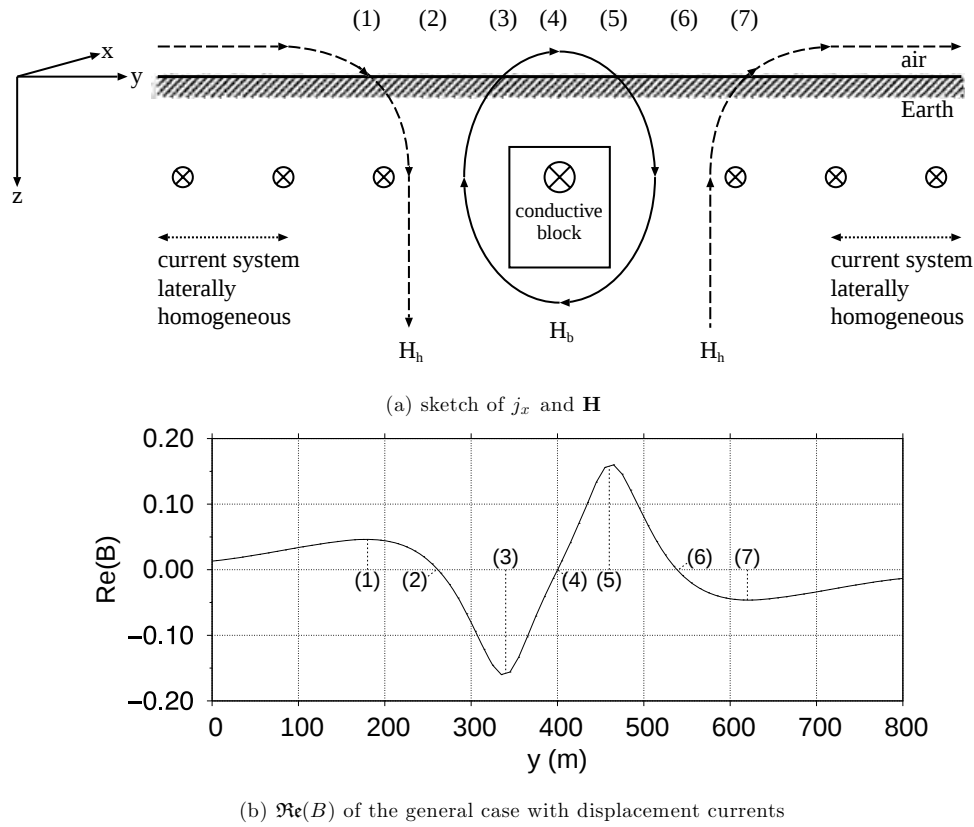


Figure 8. A sketch of the real part of the current density j_x in the subsurface with the emerging real part of the magnetic field $\mathbf{H}$ (panel a) and the real part of the VMT response for the model shown in Fig. 4, at $f = 250$ kHz in the general case with displacement currents (panel b). At positions (1) and (7), the magnetic field $\mathbf{H}_h$ due to currents in the resistive host is larger than the magnetic field $\mathbf{H}_b$ due to currents in the conductive block, leading to a maximum and a minimum of the real part of the VMT at (1) and (7), respectively. In the vicinity of the conductive block, the magnetic field is dominated by $\mathbf{H}_b$ resulting in a minimum and a maximum of the real part of the VMT at (3) and (5), respectively. At positions (2) and (6), the vertical components of $\mathbf{H}_h$ and $\mathbf{H}_b$ are equal in magnitude but opposite in direction and, hence, the VMT B is zero. The lateral position of the sign reversal at (4) coincides with the centre of the conductive block.

A more quantitative explanation for the lateral zero transitions can be arrived at by investigating eq. (5). As the curl operator treats the real and imaginary parts of $\mathbf{H}$ separately,

$$\frac{\partial \Re(H_z)}{\partial y} - \frac{\partial \Re(H_y)}{\partial z} = \Re(\hat{y}E_x) \quad \text{and} \quad \frac{\partial \Im(H_z)}{\partial y} - \frac{\partial \Im(H_y)}{\partial z} = \Im(\hat{y}E_x)$$

are directly related to the real and imaginary parts of the current density $j_x = \hat{y}E_x$ of the TE-mode. However, the current densities of the quasi-static and general cases in the subsurface are not directly comparable. As the propagation of the electric field in the air is modelled differently (i.e. through conduction currents in the quasi-static approximation with a conductivity $\sigma_{\text{air}} = 10^{-10} \text{ Sm}^{-1}$ and through displacement currents in the general case), there is a large difference in the scale lengths over which the electric field varies in the air (see Section 2.4). This leads to different phases and amplitudes of the electric fields of the two cases at the air–Earth interface, even if equal amplitudes and phases of the electric field are chosen as boundary conditions on the upper edge of the finite-difference mesh. In addition, different vertical node spacings were chosen in the air half-space for the quasi-static and general cases, according to the considerations in Section 2.4. To circumvent this problem, the electric field is scaled by its surface value at the left-hand edge of the mesh.

For the general case, the real and imaginary parts of the normalized current density at $f = 250$ kHz are shown in Figs 7(a) and (c), respectively. Similarly, for the quasi-static case, the real and imaginary parts of the normalized current density at $f = 250$ kHz are shown in Figs 7(b) and (d), respectively. The area of the highest normalized current density amplitude (up to $5.7 \times 10^{-4} \text{ A V}^{-1} \text{ m}^{-1}$) coincides with the conductive block. To the sides of the block at $y < 340$ m and $y > 460$ m, the normalized current density amplitude reaches $1.8 \times 10^{-4} \text{ A V}^{-1} \text{ m}^{-1}$, with only small lateral changes of the real and imaginary parts at the beginning and end of the profile.

An important simplification ensues for the real part of the VMT of the general case, as $\Re(H_y)$ exceeds $\Im(H_y)$ by at least a factor 4.4 at all positions along the profile (not shown). Hence, the real part of the VMT can be approximated as $\Re(B) \approx \Re(H_z)/\Re(H_y)$ and is mostly determined by $\Re(j_x)$ in Fig. 7(a). We illustrate the sign reversals in the real part of the VMT for the general case with a sketch (Fig. 8a) that describes the real part of the current system in the subsurface and the emerging magnetic field. The real part of the magnetic

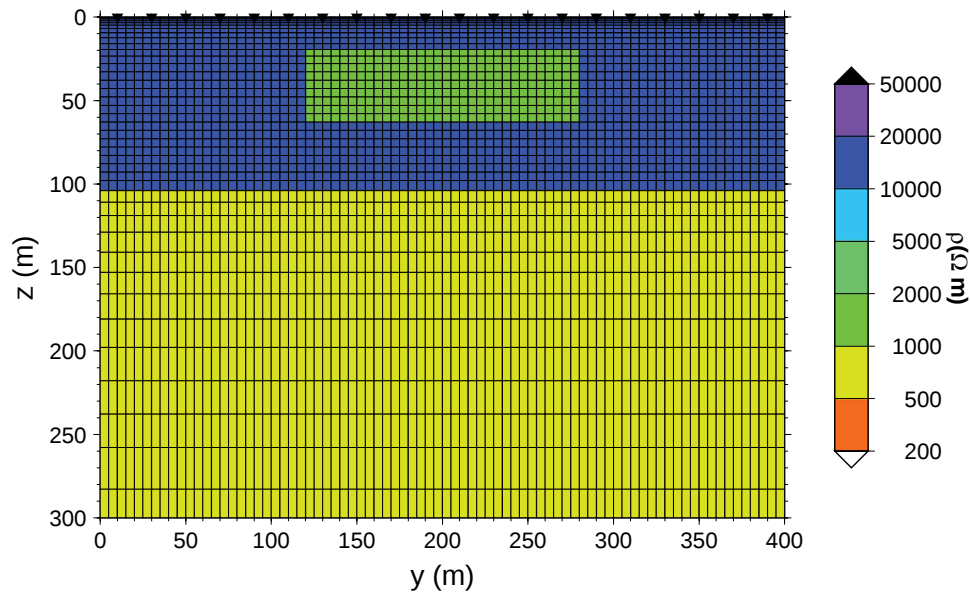


Figure 9. Simple 2-D model with a buried elongated block of a resistivity of 1000 Ω m in a resistive layer of 10 000 Ω m and underlain by a half-space of 500 Ω m. The relative dielectric permittivity is assumed to be $\epsilon_r = 5$. Receiver positions are indicated by black triangles.

field due to currents in the resistive host is designated as $\mathbf{H}_h$ and the part due to currents in the conductive block is designated as $\mathbf{H}_b$ (Fig. 8a). To facilitate a simpler comparison, the real part of the VMT for the general case (solid line in Fig. 5i) is plotted in Fig. 8(b). At the beginning and end of the profile [i.e. to the left-hand side of position (1) and to the right-hand side of position (7) in Fig. 8], the lateral homogeneity of the current system generates a magnetic field with a very small H_z -component. At positions (1) and (7), that is, at $y = 180$ and 620 m, the magnetic field $\mathbf{H}_h$ due to the resistive host is larger than the magnetic field $\mathbf{H}_b$ due to the conductive block, leading to a maximum and a minimum of the real part of the VMT at (1) and (7), respectively. At positions (2) and (6) to the sides of the block, that is, at $y = 260$ and 540 m, the H_z -components of $\mathbf{H}_h$ and $\mathbf{H}_b$ are equal in amplitude but point in opposite directions, leading to zero-transitions of the real part of the VMT. The minimum, zero transition and maximum of the real VMT response at positions (3), (4) and (5), respectively, are similar in both the quasi-static and general cases (see Fig. 5i). Though in magnitude smaller than the current system in the block, the lateral current system is strong enough to generate a commensurable maximum and minimum of the real part of the VMT at $y = 180$ and 620 m, respectively (Fig. 8b). Hence, the main effect of displacement currents on the real part of the VMT is to increase the response at the edges of the conductor. As noted before, there are no such distinct maxima or minima associated with the lateral sign reversals in the quasi-static VMT response at $f = 250$ kHz (star symbols in Fig. 5i). The reason is most likely that the vertical extent of the current systems and the total current strengths to the sides of the block (Figs 7b and d) are smaller than in the general case with displacement currents (Figs 7a and c), whereas the current within the conductive block has a comparable amplitude in both cases. It should also be noted that the imaginary part of the VMT increases quite strongly in amplitude if displacement currents are included (Fig. 5k). An explanation with regard to the imaginary part of the current density (shown in Fig. 7c) does not appear to be possible as the imaginary part of H_z and the real part of H_y are involved.

3.2 Inverse modelling examples

Synthetic responses of a simple 2-D model (Fig. 9), with constant relative dielectric permittivity $\epsilon_r = 5$ and an elongated block with a resistivity of 1000 Ω m that is buried in a resistive layer of 10 000 Ω m and underlain by a half-space of 500 Ω m, were computed for the TM- and the TE-mode. The responses were computed at 20 receiver sites for 15 frequencies, ranging from 10 to 250 kHz giving a total of 600 data points. Gaussian white noise, corresponding to 2.5 per cent of the modulus of the computed impedances, was added to the forward responses of both polarizations.

After that, two inversions of the synthetic data set were performed with the REBOCC inverse scheme (Siripunvaraporn & Egbert 2000). During the first inversion, displacement currents were allowed for, whereas they were neglected during the second inversion. In both inversions, the error floor was assumed to correspond to 2.5 per cent of the modulus of the impedances, and the starting model was a homogeneous half-space of 10 000 Ω m.

After six iterations with the inversion that allows for displacement currents, a model was obtained (Fig. 10), which fits the data to a rms misfit of 1.04. Additional iterations with REBOCC did not decrease the rms misfit further. The inverse model reproduces the edges of the block and the resistivities of the block and layered half-space rather accurately. The transition from the lower edge of the conductor into the resistive layer is, however, smeared out.

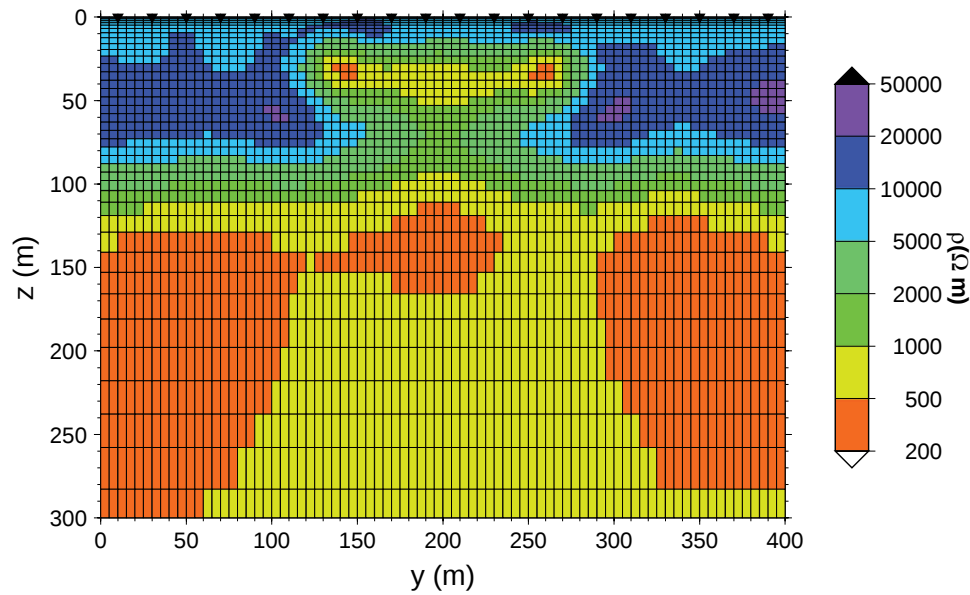


Figure 10. 2-D REBOCC inversion result of synthetic data from the 2-D block model in Fig. 9. Displacement currents were allowed for during the inversion. After six iterations, a rms misfit of 1.04 was reached. The inverse process has reconstructed the edges of the conductive block and the resistivities of the block and layered half-space rather accurately. The lower edge of the conductor is smeared out due to the damping of the electromagnetic field in the block.

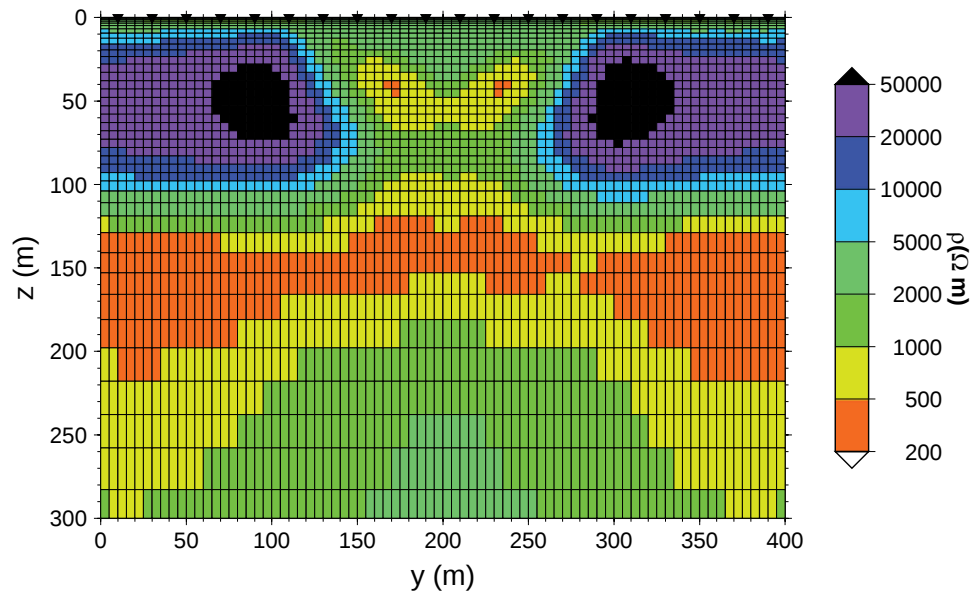


Figure 11. 2-D REBOCC inversion result of synthetic data from the 2-D block model in Fig. 9. Displacement currents were not allowed for during the inversion. After four iterations, the lowest rms misfit of 1.95 was reached. Artefactual structures in the form of a conductive near-surface layer, two resistors at the sides of the conductive block, a distorted shape of the block and a too large depth to the top of the confining half-space are consequences of the omission of displacement currents during the inversion.

Neglecting displacement currents results in convergence problems and an inverse model with many artefactual structures (Fig. 11). The lowest rms misfit of 1.95 was obtained after four iterations. Clearly, an artefactual thin conductive layer is visible at the surface (a similar conductive layer is also observed by Persson & Pedersen (2002) in 1-D inverse models, computed in the quasi-static approximation, for synthetic data of a homogeneous half-space). The lateral extent of the conductive block and the top of the central parts of the block are grossly in error. Two artefactual resistors with resistivities close to 100 000 Ω m appear to the left- and right-hand side of the block. The depth to the top of the underlying conductive layer is shifted from 105 to 130 m. If the synthetic data were generated from a model without the underlying conductive layer, the artefactual resistors would be observed, both to the sides of and below the conductive block (not shown).

A comparison of the relative errors, that is, the differences between the synthetic data and the forward responses scaled by the data errors, generated by the two inverse schemes, is shown in Fig. 12. The relative errors from the inversion that accounts for displacement currents

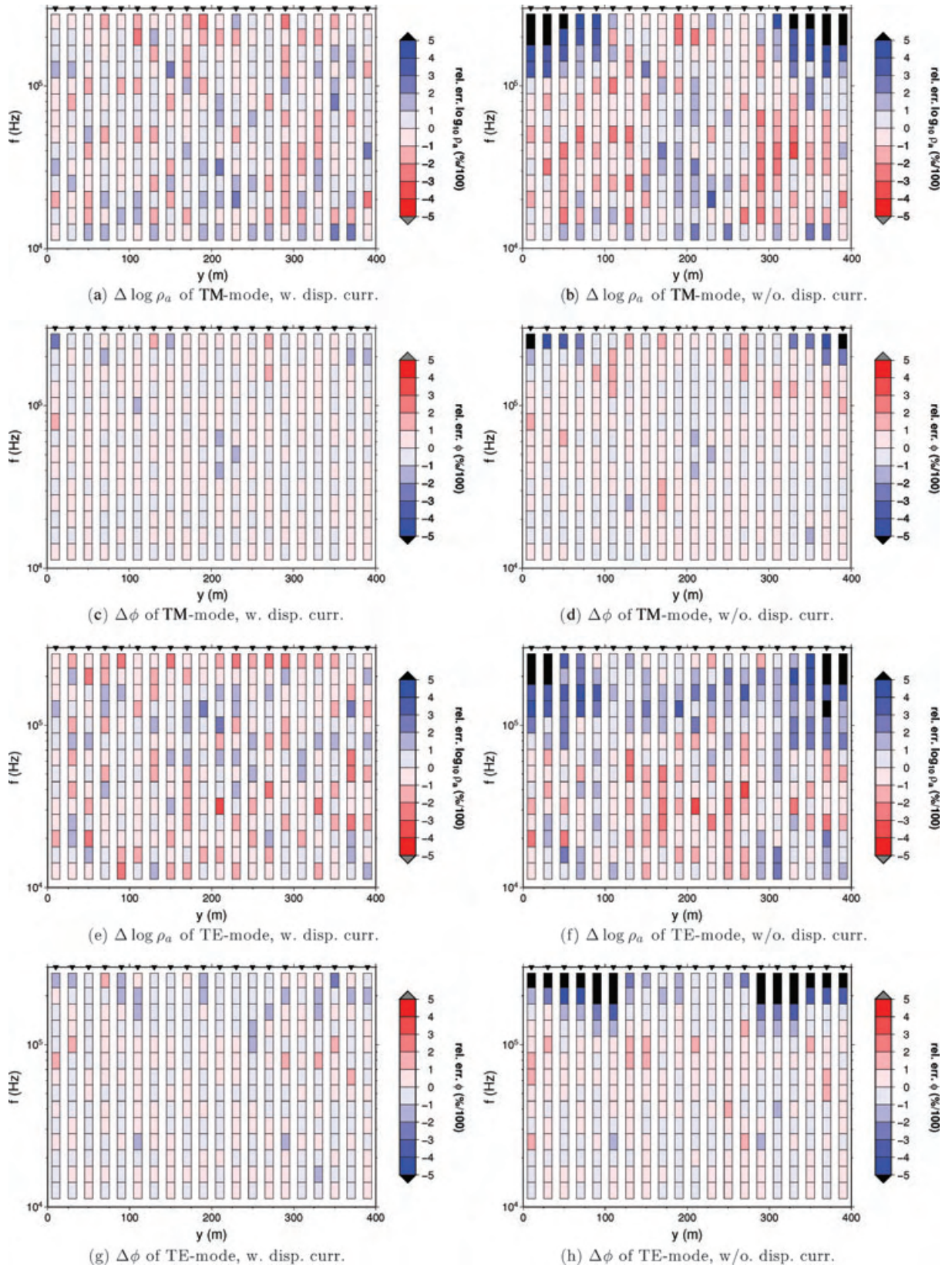


Figure 12. Relative errors obtained from the inversions of the synthetic data of the model shown in Fig. 9. The relative errors from inversions that account for displacement currents and that neglect displacement currents are shown in the left- and right-hand columns, respectively. The relative errors of $\log_{10} \rho_a$ and ϕ for the TM-mode are shown in panels (a)–(d), whereas the relative errors of $\log_{10} \rho_a$ and ϕ for the TE-mode are shown in panels (e)–(h). Systematic deviations from the synthetic data are mostly observed at high frequencies and stations to the sides of the conductive block for the model from the inverse scheme that does not allow for displacement currents (Fig. 11).

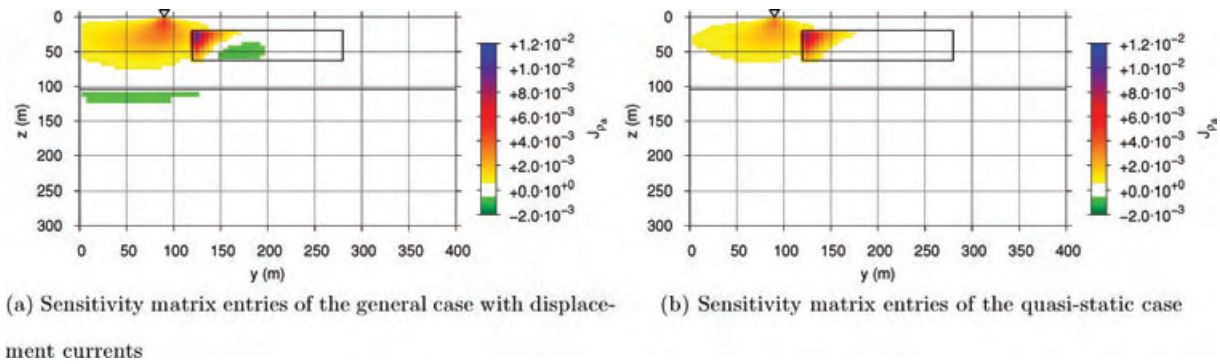


Figure 13. Sensitivity matrix entries for the block model in Fig. 9 and the TE-mode apparent resistivity at $f = 250$ kHz and a receiver site at $y = 90$ m (to the left-hand side of the conductive block). The receiver site is marked by a triangle. The edges of the conductive block and the interface between the upper layer and the confining half-space are depicted as solid black lines. Sensitivity values, which are close to 0, are plotted in white. As expected, sensitivities are largest at the left-hand upper edge of the conductive block and the sensitivity entries of the general case (panel a) encompass a larger volume with non-zero values (observe the negative sensitivity values marked in green) than those calculated in the quasi-static approximation (panel b).

(left-hand column of Fig. 12) show relatively random deviations of the forward data from the synthetic data. In contrast to this, the relative errors from the quasi-static inversion (right-hand column of Fig. 12) exhibit systematic deviations in the form of frequency ranges common to groups of neighbouring stations, with relative errors that have absolute values significantly larger than one and the same sign. The systematic deviations originate from the false assumption that displacement currents can be neglected during the inversion. As expected, the misfit is most severe at high frequencies and receiver sites to the sides of the conductive block.

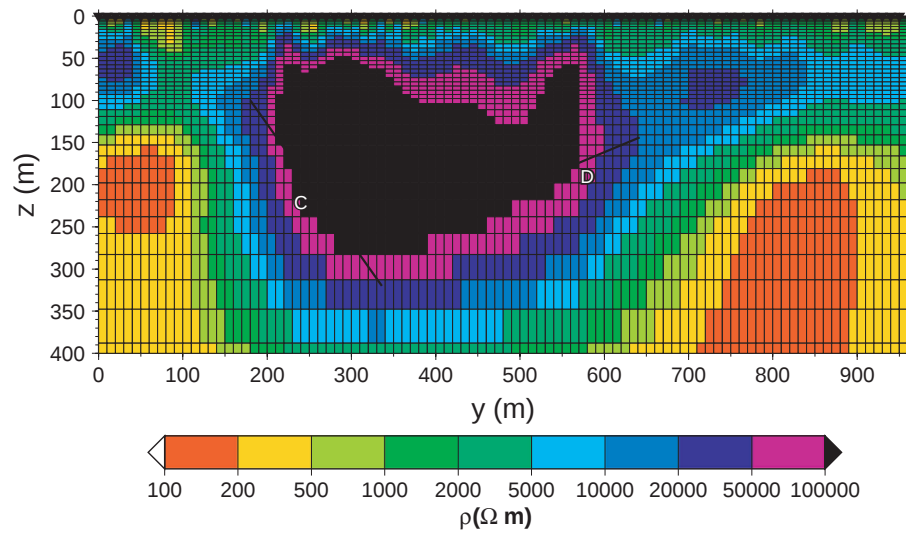
As an example, the row of the sensitivity matrix for the block model in Fig. 9 and the TE-mode apparent resistivity at $f = 250$ kHz and a receiver site at $y = 90$ m (to the left-hand side of the conductive block) is shown in Fig. 13. Model parameters with sensitivities close to zero (shown in white colours in Fig. 13) have little influence on the considered data item. As expected, sensitivities, which were computed for the general case (Fig. 13a), encompass a larger volume with non-zero values than those computed in the quasi-static approximation (Fig. 13b). Especially, the depth extend for the non-zero sensitivity values of the general case is larger. This larger depth range is equivalent to a larger depth of investigation for the general case as already indicated in Section 2.4.

4 A FIELD DATA EXAMPLE

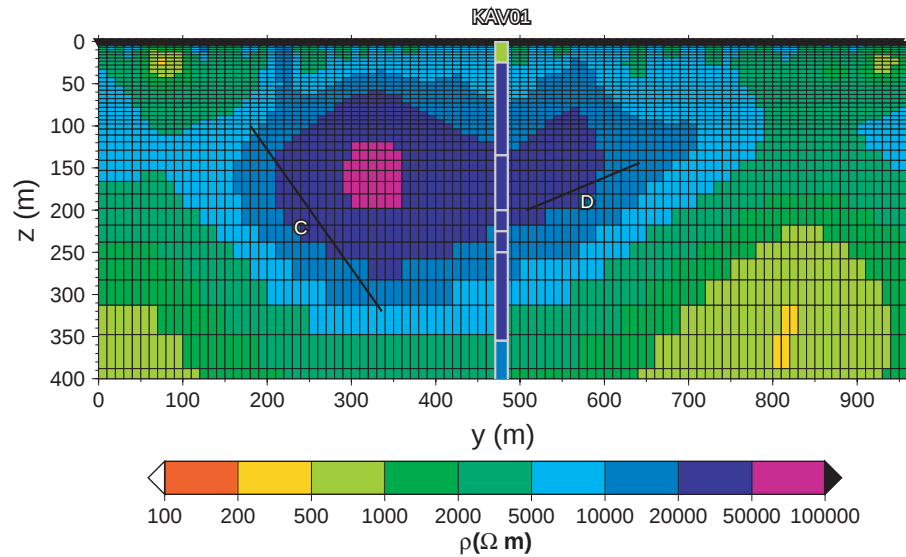
Linde & Pedersen (2004) investigate highly resistive granitic bedrock on the small island Ävrö, Sweden, with tensor RMT, in the frequency range of 14–226 kHz. RMT data were acquired on an east–west profile, with a total length of 960 m and a station spacing of 10 m. On Ävrö, the typical soil thickness is between 0 and 1 m. The bedrock consists mostly of granite. In some locations, aplitic and pegmatitic dykes are encountered (Gentzschein *et al.* 1987). Previous geophysical studies include borehole measurements by Gentzschein *et al.* (1987) and a seismic reflection study on the same profile by Juhlin & Palm (1999). A normal-resistivity log and a fracture frequency log of borehole KAV01, located in the central part of the profile, reveal an upper weathered layer, with a thickness of up to 30 m and a resistivity of about 600 Ω m, followed by almost intact and highly resistive bedrock down to a depth of 200 m and with a resistivity between 32 000 and 40 000 Ω m (Gentzschein *et al.* 1987). Between 200 and 400 m depth, the resistivity slowly decreases to 10 000 Ω m. At greater depth, the bedrock is more fractured and saline pore fluids decrease the electrical resistivity to a few thousand Ω m. Juhlin & Palm (1999) describe two major seismic reflectors (see Fig. 14d) for the depth range down to 400 m. Reflector C is located beneath the western part of the profile, at a depth between 100 and 320 m and dips approximately 60° to the east. Reflector D is located beneath the central part of the profile at a depth between 150 and 200 m and dips approximately 20° to the west.

To mitigate the effects of displacement currents, Linde & Pedersen (2004) restrict the data set used in quasi-static 2-D inversions with the REBOCC scheme (Siripunvaraporn & Egbert 2000) to frequencies up to 56 kHz. Linde & Pedersen (2004) perform inversions for the TE-mode, TM-mode, TE- and TM-modes together and the determinant of the impedance tensor. By computing synthetic TE-mode, TM-mode and determinant data for a 3-D model and comparing the corresponding 2-D inversions, Pedersen & Engels (2005) show that the inversion of determinant data is less prone to introducing artefacts from 3-D structures off the profile to 2-D inverse models. Furthermore, the inverse model of the determinant data, presented by Pedersen & Engels (2005) has a better data fit than their other models. For the inversion of the RMT data from Ävrö, this leads us to concentrate on the inversion of determinant data, as the data at both ends of the profile show a high degree of three-dimensionality (Linde & Pedersen 2004). At a few stations, the determinant data of the highest frequencies (160 and 226 kHz) have very small negative phases, which can be indicative of displacement currents (Song *et al.* 2002). As the rather irregular behaviour of the apparent resistivities at the same stations and frequencies hints at problems with measurement accuracy, we excluded such data points from the inversion.

In the following, we examine the effect of displacement currents, by first considering the inversion of the restricted set of frequencies and then for the full set of frequencies. For each data set, inversions were carried out in both the quasi-static approximation and with displacement



(a) model QL for the low-frequency data set and the quasi-static case



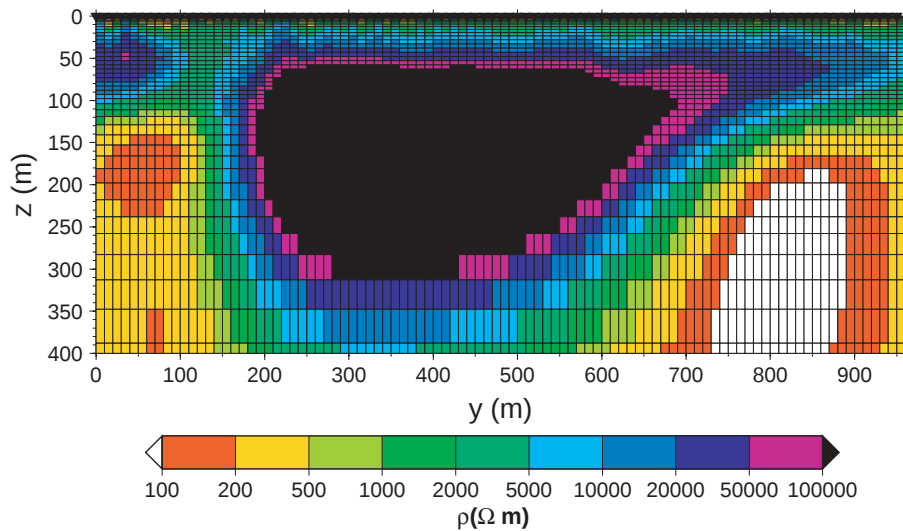
(b) model DL for the low-frequency data set and the general case with displacement currents

Figure 14. Models of the inversion of determinant data from Ävrö for (a) the low-frequency data set in the quasi-static approximation (model QL), (b) the low-frequency data set allowing for displacement currents (model DL), (c) the full set of frequencies in the quasi-static approximation (model QF) and (d) the full set of frequencies allowing for displacement currents (model DF). The lines marked by C and D indicate seismic reflectors from Juhlin & Palm (1999) (their fig. 8). The resistivity values of borehole KAV01 are taken from the normal-resistivity log presented in Gentzschlein *et al.* (1987). In contrast to models QL and QF, models DL and DF have a more realistic range of resistivities if compared in terms of the range observed in the normal-resistivity log. Furthermore, the resistivity–depth section of model DF at borehole KAV01 is in good agreement with the normal-resistivity log down to a depth of 230 m, and the positions of the seismic reflectors are in good agreement with resistivity contrasts in model DF.

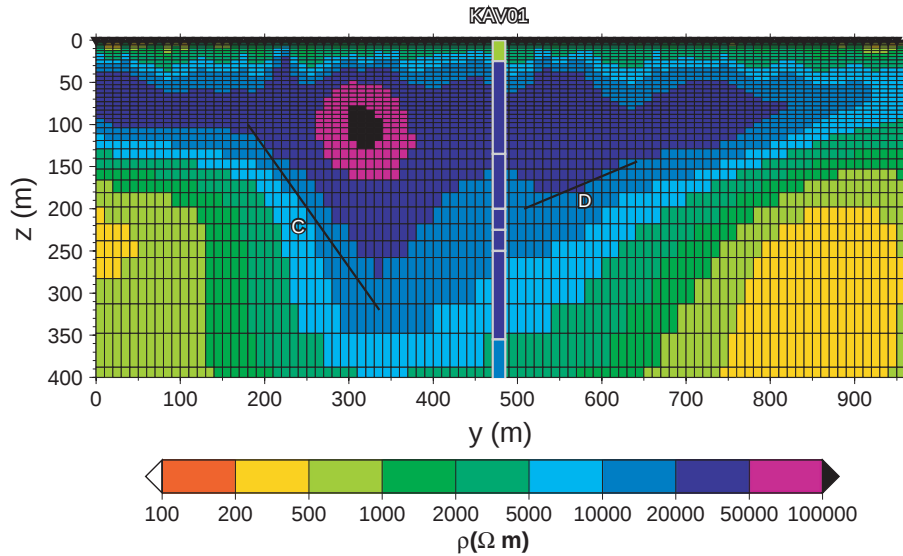
currents. We assumed the relative dielectric permittivity to be $\epsilon_r = 6$, which, for granite, is in the range between 5 and 8, given by Reynolds (1997). Variation of the permittivity in this range leads to only small differences of the resistivity models for the Ävrö data (not shown).

Our quasi-static determinant model for the lowest frequencies up to 56 kHz (model QL) in Fig. 14(a) resembles the corresponding model by Linde & Pedersen (2004) (their fig. 9d) strongly. We did not include the shallow sea (less than 10 m deep) to the east of Ävrö as *a priori* information, as this turned out to be of negligible importance. The central unfractured granite reaches resistivities up to 500 000 Ω m. The conductor at the western end of the profile is interpreted by Linde & Pedersen (2004) as a 150 m wide wet fracture zone, assumed to be related to seismic reflector C of Juhlin & Palm (1999), although the positions of the conductor and reflector are not in very good agreement. The subhorizontal seismic reflector D does not appear to be related to any structure in the resistivity model. The rms misfit of model QL is 1.56.

The inversion with displacement currents for the low-frequency data set gives a model (model DL in Fig. 14b) with a significantly reduced range of resistivities from 300 to 100 000 Ω m. The conductors at 50 and 850 m along the profile appear at greater depth and the



(c) model QF for the full set of frequencies and the quasi-static case



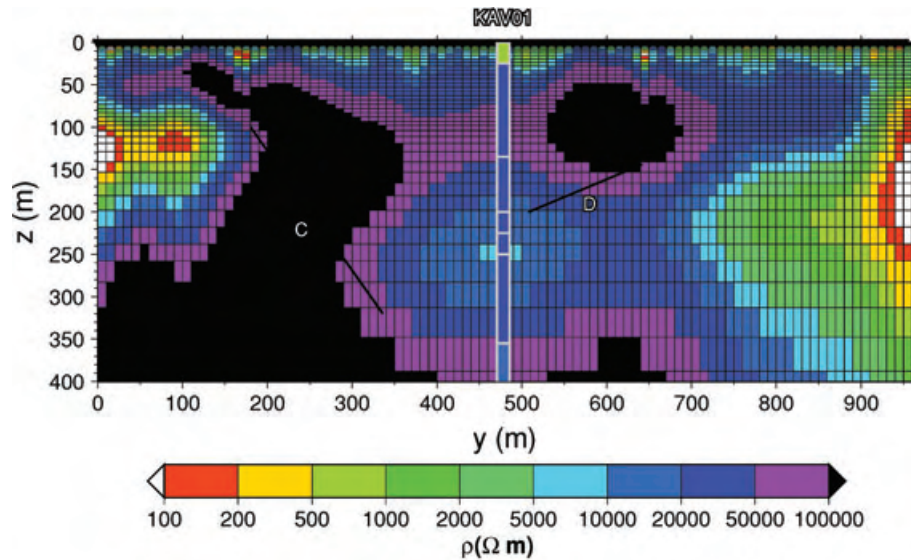
(d) model DF for the full set of frequencies and the general case with displacement currents

Figure 14. (Continued.)

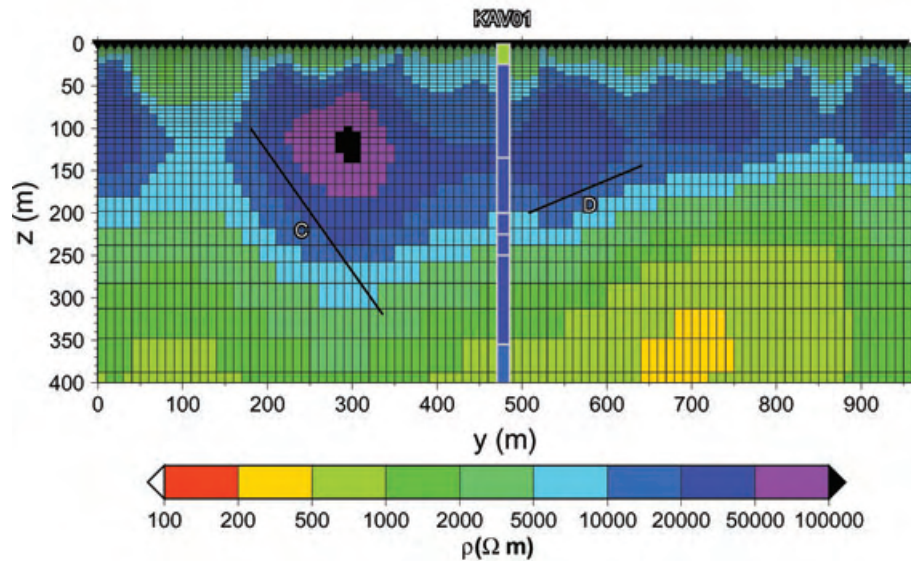
boundary between the central resistor and the western conductors is less steep than in the model QL (Fig. 14a). The rms misfit of model DL is 2.03.

The quasi-static inversion for the full set of frequencies (model QF in Fig. 14c) leads to a transition into the top of the central resistor that is sharper than in model QL (Fig. 14a). The resistivity of the central resistor is as high as $5 \times 10^6 \Omega \text{ m}$. The positions of the conductors at profile metres 50 and 850 is very similar to the positions in model QL. The rms misfit of model QF is 3.16.

In comparison to the quasi-static inversions, the inversion with displacement currents for all frequencies gives a model [model DF in Fig. 14(d) with an rms error of 2.60] that shows a less extreme range of resistivities, both at depth and close to the surface. The depth to the conductors at both ends of the profile is about 50 m larger than in the quasi-static models QL and QF in Figs 14(a) and (c), respectively. The model is also in better agreement with the positions of the seismic reflectors. The position of seismic reflector C conforms to an expected boundary between an unfractured resistive granite body and water saturated fractured bedrock. Therefore, we would expect reflector C to represent a boundary of rock units, with different grades of fracturing, rather than a 150 m wide fracture zone, as proposed by Linde & Pedersen (2004). Similarly, reflector D appears to coincide with a subhorizontal boundary of rock units. Furthermore, the model in Fig. 14(d) is in very good agreement with the resistivities of the normal-resistivity log of borehole KAV01 (Gentzschin *et al.* 1987), down to a depth of 230 m. At greater depth, the model might be more influenced by the smoothness constraint imposed during the inversion than the data. Compared with model DF, model DL (Fig. 14b) deviates from the normal-resistivity log at shallow depth down to 100 m and the positions of



(a) model TEDF for the full set of frequencies in the TE-mode and the general case with displacement currents



(b) model TMDF for the full set of frequencies in the TM-mode and the general case with displacement currents

Figure 15. Models of the inversion of (a) TE-mode data (model TEDF) and (b) TM-mode data (model TMDF) from Ävrö for the full set of frequencies allowing for displacement currents. The lines marked by C and D indicate seismic reflectors from Juhlin & Palm (1999). The resistivity values of borehole KAV01 are taken from the normal-resistivity log presented in Gentschein *et al.* (1987). The resistivities of the central resistor in model TEDF are as high as 300 000 Ω m. Compared with model DF (Fig. 14d), neither model TEDF nor model TMDF shows similarly good agreement with the positions of the seismic reflectors C and D or the normal-resistivity log.

the seismic reflectors are not as representative as bounds of different rock units. Hence, it appears that the inclusion of high-frequency data is of great importance during the modelling process.

As a verification that the 2-D inverse models of determinant data are less biased by 3-D structures off the profile, the inverse models of TE-mode data (model TEDF) and TM-mode data (model TMDF) are shown in Figs 15(a) and (b), respectively. In both inversions, the full set of frequencies was used and displacement currents were accounted for. The rms fits of 4.56 for the TE-mode model (reached after nine iterations) and 3.67 for the TM-mode model (reached after five iterations) are both significantly higher than that of model DF. The worst data fits of models TEDF and TMDF (not shown) are obtained at the western end of the profile, where strong 3-D effects in the VMT are observed by Linde & Pedersen (2004). In model TEDF, resistivities of the central resistor are as high as 300 000 Ω m. Compared with model

DF (Fig. 14d), neither model TEDF nor model TMDF shows similarly good agreement with the positions of the seismic reflectors C and D of Juhlin & Palm (1999) or the normal-resistivity log by Gentzschin *et al.* (1987).

5 DISCUSSION AND CONCLUSIONS

We demonstrated the effect of displacement currents on 2-D TM-mode, TE-mode and VMT data, measured with the RMT method at frequencies between 10 and 300 kHz. Forward modelling of subsurfaces with resistivities larger than $1000 \Omega \text{ m}$ confirms that responses computed in the quasi-static approximation, that is, when displacement currents are neglected, become increasingly inaccurate, with rising frequency. For a homogeneous half-space, both apparent resistivity and phase, computed with displacement currents, decrease from their constant values in the quasi-static approximation, with increasing frequency. At high frequencies, the dielectric effect leads to the occurrence of distinct sign reversals in the real part of the VMT, which are not observed in the quasi-static approximation and might lead to artefactual 2-D or 3-D structures in an interpretation, based on the quasi-static approximation.

The interpretation of high-frequency RMT data with an inverse scheme that operates in the quasi-static approximation will inevitably lead to an inverse model with artefactual structures. As can be seen from the quasi-static interpretation of our synthetic data example in Fig. 11, the resistivities found in this inverse model vary over a larger range than those of the true model (Fig. 9). Typical artefactual structures include conductive near-surface layers, regions of excessively high resistivities next to conductors, as well as conductors that deviate strongly from their true shapes and positions. As only the resistivity distribution is inverted in the scheme presented here, a value for the dielectric permittivity must be chosen before the inversion. The relative dielectric permittivity of bedrock is typically in the range of 5 to 9 (e.g. Reynolds 1997, table 12.3), and a variation in this range does not lead to any important differences in the obtained resistivity models.

Typically, the primary EM field from remote radio transmitters has an angle of incidence that is close to grazing incidence at the measurement site. The assumption of vertically incident plane waves in the modelling code is a limitation, which is of minor importance in many practical situations. Often, a conductive surface layer consisting of, for instance, weathered bedrock or glacial till is present in the area of interest and refracts the incident field towards the vertical due to its relatively low resistivity.

For the Ävrö field data, the inversion that allows for displacement currents and includes high-frequency data produces a model that is in very good agreement with the results of other geophysical methods. The seismic reflectors C and D by Juhlin & Palm (1999) coincide with the boundaries between structures of different conductivity (Fig. 14d). The resistivity depth section of the model at borehole KAV01 matches the normal-resistivity log by Gentzschin *et al.* (1987) very well, down to a depth of 230 m below which the model might be strongly influenced by the smoothness constraint applied during the inversion. The inverse models computed in the quasi-static approximation (Figs 14a and c) contain artefactual structures, with unrealistically large resistivities, even if only the low-frequency data set is inverted (Fig. 14a).

ACKNOWLEDGMENTS

The original 2-D MT code by WS at Mahidol University is supported by a research grant from the Thailand Research Fund (TRF: RMU5080025). Nazli Ismail of Uppsala University computed the 3-D forward responses with the X3D code. Dmitry Avdeev, IZMIRAN, Russian Academy of Sciences, spent a large amount of money on the telephone, from Japan, to correct our X3D input files. A discussion with John Weaver, University of Victoria, clarified a few theoretical aspects. We would also like to thank Ute Weckmann, GFZ Potsdam and an anonymous reviewer for their comments, which helped to improve the manuscript.

REFERENCES

- Anderson, E. *et al.*, 1999. *LAPACK Users' Guide*, 3rd edn, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- Apria, C., Booker, J.R. & Smith, J.T., 1997. The forward problem of electromagnetic induction: accurate finite-difference approximations for two-dimensional discrete boundaries with arbitrary geometry, *Geophys. J. Int.*, **129**, 29–40.
- Avdeev, D.B., Kuvshinov, A.V., Pankratov, O.G. & Newman, G.A., 2002. Three-dimensional induction logging problems, part 1: an integral equation solution and model comparisons, *Geophysics*, **67**(2), 413–426.
- Bastani, M. & Pedersen, L.B., 2001. Estimation of magnetotelluric transfer functions from radio transmitters, *Geophysics*, **66**(4), 1038–1051.
- Brewitt-Taylor, C.R. & Weaver, J.T., 1976. On the finite difference solution of two-dimensional induction problems, *Geophys. J. R. astr. Soc.*, **47**(2), 375–396.
- Crossley, D.J., 1981. The theory of EM surface wave impedance measurements, in *Geophysical Applications of Surface Wave Impedance Measurements*, Paper no. 81-15, pp. 1–17, eds Collet, L.S. & Jensen, O.G., Geological Survey of Canada.
- Fraser, D.C., Stodt, J.A. & Ward, S.H., 1990. The effect of displacement currents on the response of a high-frequency helicopter electromagnetic system, in *Geotechnical and Environmental Geophysics*, Vol. 2, pp. 89–95, ed. Ward, S.H., SEG, Tulsa.
- Freund, R.W., 1992. Conjugate gradient-type methods for linear systems with complex symmetric coefficient matrices, *SIAM J. Sci. Stat. Comput.*, **13**(1), 425–448.
- Gentzschin, B., Nilsson, G. & Steinberg, L., 1987. Preliminary investigations of fracture zones at Ävrö - results from investigations performed July 1986 - May 1987, SKB Progress Report 25–87-16, SKB.
- Golub, G. & van Loan, C., 1996. *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, 3rd edn, Johns Hopkins University Press, Baltimore and London.
- Hohmann, G.W., 1987. Numerical modeling for electromagnetic methods in geophysics, in *Electromagnetic Methods in Applied Geophysics*, Vol. 1 (Theory), pp. 313–363, ed. Nabighian, M.N., SEG, Tulsa.
- Huang, H.P. & Fraser, D.C., 2002. Dielectric permittivity and resistivity mapping using high-frequency, helicopter-borne EM data, *Geophysics*, **67**(3), 727–738.
- Juhlin, C. & Palm, H., 1999. 3-D structure below Ävrö island from high-resolution reflection seismic studies, southeastern Sweden, *Geophysics*, **64**, 662–667.
- Keller, G.V., 1987. Rock and mineral properties, in *Electromagnetic*

- Methods in Applied Geophysics*, Vol. 1 (Theory), pp. 12–51, ed. Nabighian, M.N., SEG, Tulsa.
- Linde, N. & Pedersen, L.B., 2004. Characterization of a fractured granite using radio magnetotelluric (RMT) data, *Geophysics*, **69**(5), 1155–1165.
- McNeill, J.D. & Labson, V., 1991. Geological mapping using VLF radio fields, in *Electromagnetic Methods in Applied Geophysics*, Vol. 2 (Application), Parts A. & B, pp. 521–640, ed. Nabighian, M.N., SEG, Tulsa.
- Nabulsi, K.A. & Wait, J.R., 1996. TM reflection from an insulating layer within a lossy half-space, *J. Appl. Geophys.*, **34**(4), 271–276.
- Pedersen, L.B. & Engels, M., 2005. Routine 2D inversion of magnetotelluric data using the determinant of the impedance tensor, *Geophysics*, **70**(2), G33–G41.
- Persson, L. & Pedersen, L.B., 2002. The importance of displacement currents in RMT measurements in high resistivity environments, *J. Appl. Geophys.*, **51**(1), 11–20.
- Reynolds, J.M., 1997. *An Introduction to Applied and Environmental Geophysics*, John Wiley & Sons Ltd, Chichester, UK.
- Rodi, W.L., 1976. A technique for improving the accuracy of finite element solutions for magnetotelluric data, *Geophys. J. R. astr. Soc.*, **44**, 483–506.
- Rodi, W. & Mackie, R.L., 2001. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, **66**(1), 174–187.
- Sinha, A.K., 1977. Influence of altitude and displacement currents on plane-wave EM fields, *Geophysics*, **42**(1), 77–91.
- Siripunvaraporn, W. & Egbert, G., 2000. An efficient data-subspace inversion method for 2-D magnetotelluric data, *Geophysics*, **65**(3), 791–803.
- Song, Y., Kim, H.J. & Lee, K.H., 2002. High-frequency electromagnetic impedance method for subsurface imaging, *Geophysics*, **67**(2), 501–510.
- Stewart, D.C., Anderson, W.L., Grover, T.P. & Labson, V.F., 1994. Shallow subsurface mapping by electromagnetic sounding in the 300 kHz to 30 MHz range: model studies and prototype system assessment, *Geophysics*, **59**(8), 1201–1210.
- Wait, J.R., 1953. Propagation of radio waves over a stratified ground, *Geophysics*, **18**(2), 416–422.
- Wait, J.R., 1970. *Electromagnetic Waves in Stratified Media*, Vol. 3: International series of monographs in electromagnetic waves, Pergamon Press, Oxford, New York.
- Wait, J.R. & Nabulsi, K.A., 1996. Wave tilt of radio waves propagating over a layered ground, *Geophysics*, **61**(6), 1647–1652.
- Ward, S.H. & Hohmann, D.W., 1987. Electromagnetic theory for geophysical applications, in *Electromagnetic Methods in Applied Geophysics*, Vol. 1 (Theory), pp. 131–311, ed. Nabighian, M.N., SEG, Tulsa.
- Ward, S.H., Jiracek, G.R. & Linlor, W.L., 1968. Electromagnetic reflection from a plane-layered lunar model, *J. Geophys. Res.*, **73**(4), 1355–1372.
- Weaver, J.T., LeQuang, B.V. & Fischer, G., 1985. A comparison of analytical and numerical results for a two-dimensional control model in electromagnetic induction, I: B-polarization calculations, *Geophys. J. R. astr. Soc.*, **82**, 263–277.
- Weaver, J.T., LeQuang, B.V. & Fischer, G., 1986. A comparison of analytical and numerical results for a two-dimensional control model in electromagnetic induction, II: E-polarization calculations, *Geophys. J. R. astr. Soc.*, **87**, 917–948.
- Yin, C. & Hodges, G., 2005. Influence of displacement currents on the response of helicopter electromagnetic systems, *Geophysics*, **70**(4), G95–G100.
- Zacher, G., 1992. Hochfrequente Elektromagnetik auf Altlasten, in *Protokoll über das 14. Kolloquium Elektromagnetische Tiefenforschung, Borkheide*, pp. 357–363, eds Haak, V. & Rodemann, H., Dt. Geophys. Gesellschaft, GFZ Potsdam.

APPENDIX A: FINITE-DIFFERENCE APPROXIMATION

The derivation of the finite-difference approximation (FDA) with the integration method following Hohmann (1987) and Aprea *et al.* (1997) is particularly instructive. In a finite-difference mesh, a node (i, j) is a corner point of four finite-difference cells, with admittivities $\hat{y}_{i-1/2, j-1/2}$, $\hat{y}_{i+1/2, j-1/2}$, $\hat{y}_{i-1/2, j+1/2}$, and $\hat{y}_{i+1/2, j+1/2}$ (Fig. A1). The cells have widths $\Delta y_{i-1/2}$ and $\Delta y_{i+1/2}$ and heights $\Delta z_{j-1/2}$ and $\Delta z_{j+1/2}$. The rectangle A has the centres of the cells as its corner points (Fig. A1). It is assumed that the horizontal electric field component of the TE-mode and the horizontal magnetic field component of the TM-mode at node (i, j) are $E_x^{i,j}$ and $H_x^{i,j}$, respectively, and that the magnetic permeability is equal to its vacuum value μ_0 . Nodes along the boundary of the finite-difference mesh are called boundary nodes. All other nodes are called inner nodes. Finite-difference equations for the TM- and TE-mode are obtained by integrating eqs 14 and 13, respectively, over the surface of A . This surface integral is then transformed to a contour integral around the perimeter ∂A of the surface A with Gauss' Theorem.

Hence, for the TE-mode,

$$\int_A (\hat{z} \hat{y} E_x) dA = \int_A (\nabla \cdot \nabla E_x) dA = \int_{\partial A} (\mathbf{n} \cdot \nabla E_x) dl, \quad (\text{A1})$$

where $\mathbf{n}$ is an outward unit normal vector on the edges of A .

The part of A , for instance, which is entirely situated to the upper left-hand side of node (i, j) , contributes to the surface integral in eq. (A1) with

$$\int_A (\hat{z} \hat{y} E_x) dA \approx \hat{z} \hat{y}_{i-1/2, j-1/2} E_x^{i,j} \frac{\Delta y_{i-1/2}}{2} \frac{\Delta z_{j-1/2}}{2}. \quad (\text{A2})$$

For the upper edge of the rectangle A , for instance, the integral around the perimeter ∂A can be approximated as

$$\int_{\partial A} (\mathbf{n} \cdot \nabla E_x) dl \approx \left[-\frac{E_x^{i,j} - E_x^{i,j-1}}{\Delta z_{j-1/2}} \right] \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{2}, \quad (\text{A3})$$

where $(\Delta y_{i+1/2} + \Delta y_{i-1/2})/2$ is the length of the perimeter ∂A on the upper edge of A and $\mathbf{n}$, which equals $(0, 0, -1)$ on the upper edge of A , collects the vertical component of ∇E_x , that is, $(\nabla E_x)_z \approx (E_x^{i,j} - E_x^{i,j-1})/\Delta z_{j-1/2}$, multiplied by -1 .

In total, this leads to the following approximations (cf. Aprea *et al.* 1997):

$$\int_A (\hat{z} \hat{y} E_x) dA \approx \frac{1}{4} \hat{z} \hat{y}_{i,j}^{\text{int}} E_x^{i,j} + O(\Delta^3), \quad (\text{A4})$$

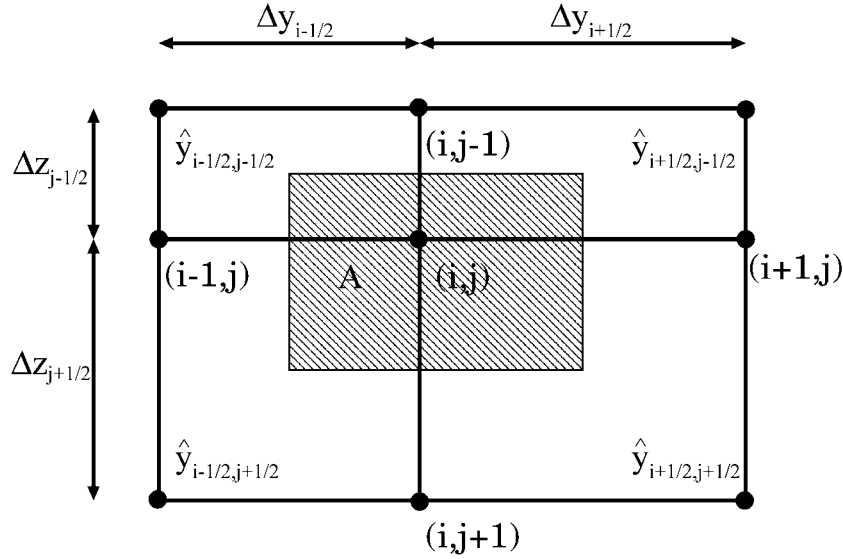


Figure A1. Finite-difference mesh around node (i, j) . The four surrounding cells have admittivities $\hat{y}_{i-1/2, j-1/2}$, $\hat{y}_{i+1/2, j-1/2}$, $\hat{y}_{i-1/2, j+1/2}$ and $\hat{y}_{i+1/2, j+1/2}$. The heights and widths of the surrounding cells are $\Delta z_{j-1/2}$ and $\Delta z_{j+1/2}$ and $\Delta y_{i-1/2}$ and $\Delta y_{i+1/2}$, respectively. The rectangle A has its corner points at the centres of the cells.

$$\begin{aligned} \int_{\partial A} (\mathbf{n} \cdot \nabla E_x) dl &\approx \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{2} \left[\frac{E_x^{i,j+1} - E_x^{i,j}}{\Delta z_{j+1/2}} - \frac{E_x^{i,j} - E_x^{i,j-1}}{\Delta z_{j-1/2}} \right] \\ &+ \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{2} \left[\frac{E_x^{i+1,j} - E_x^{i,j}}{\Delta y_{i+1/2}} - \frac{E_x^{i,j} - E_x^{i-1,j}}{\Delta y_{i-1/2}} \right] \\ &+ O(\Delta^2), \end{aligned} \quad (\text{A5})$$

where

$$\begin{aligned} \hat{y}_{i,j}^{\text{int}} &= \hat{y}_{i-1/2, j-1/2} \Delta y_{i-1/2} \Delta z_{j-1/2} + \hat{y}_{i+1/2, j-1/2} \Delta y_{i+1/2} \Delta z_{j-1/2} \\ &+ \hat{y}_{i-1/2, j+1/2} \Delta y_{i-1/2} \Delta z_{j+1/2} + \hat{y}_{i+1/2, j+1/2} \Delta y_{i+1/2} \Delta z_{j+1/2}, \end{aligned} \quad (\text{A6})$$

$$\hat{y}_{i,j}^{\text{avg}} = \frac{1}{\Delta y_{i-1/2} \Delta z_{j-1/2} + \Delta y_{i+1/2} \Delta z_{j-1/2} + \Delta y_{i-1/2} \Delta z_{j+1/2} + \Delta y_{i+1/2} \Delta z_{j+1/2}} \hat{y}_{i,j}^{\text{int}} \quad (\text{A7})$$

and $O(\Delta^2)$ are terms of second or higher order in $\Delta y_{i\pm 1/2}$ or $\Delta z_{j\pm 1/2}$.

After rearranging, one obtains

$$\begin{aligned} 0 &= 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j+1/2}} E_x^{i,j+1} + 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j-1/2}} E_x^{i,j-1} \\ &+ 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i+1/2}} E_x^{i+1,j} + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i-1/2}} E_x^{i-1,j} \\ &- \left\{ 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j+1/2}} + 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j-1/2}} \right. \\ &\quad \left. + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i+1/2}} + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i-1/2}} + \hat{y}_{i,j}^{\text{int}} \right\} E_x^{i,j}. \end{aligned} \quad (\text{A8})$$

Similarly, the FDA of the TM mode can be derived. Gauss' Theorem gives

$$\int_A (\hat{z} H_x) dA = \int_A \left(\nabla \cdot \frac{1}{\hat{y}} \nabla H_x \right) dA = \int_{\partial A} \left(\mathbf{n} \cdot \frac{1}{\hat{y}} \nabla H_x \right) dl \quad (\text{A9})$$

and the single terms can be approximated as

$$\int_A (\hat{z} H_x) dA \approx \hat{z} H_x^{i,j} \frac{(\Delta y_{i+1/2} + \Delta y_{i-1/2})(\Delta z_{j+1/2} + \Delta z_{j-1/2})}{4} + O(\Delta^3), \quad (\text{A10})$$

$$\begin{aligned}
\int_{\partial A} \left(\mathbf{n} \cdot \frac{1}{\hat{y}} \nabla H_x \right) dl &\approx \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{2} \left[\frac{1}{\hat{y}_{i,j}^d} \frac{H_x^{i,j+1} - H_x^{i,j}}{\Delta z_{j+1/2}} - \frac{1}{\hat{y}_{i,j}^u} \frac{H_x^{i,j} - H_x^{i,j-1}}{\Delta z_{j-1/2}} \right] \\
&+ \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{2} \left[\frac{1}{\hat{y}_{i,j}^r} \frac{H_x^{i+1,j} - H_x^{i,j}}{\Delta y_{i+1/2}} - \frac{1}{\hat{y}_{i,j}^l} \frac{H_x^{i,j} - H_x^{i-1,j}}{\Delta y_{i-1/2}} \right] \\
&+ O(\Delta^2),
\end{aligned} \tag{A11}$$

where the vertically and horizontally averaged inverse admittivities are given by

$$\frac{1}{\hat{y}_{i,j}^d} = \frac{\frac{1}{\hat{y}_{i+1/2,j+1/2}} \Delta y_{i+1/2} + \frac{1}{\hat{y}_{i-1/2,j+1/2}} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}}, \tag{A12}$$

$$\frac{1}{\hat{y}_{i,j}^u} = \frac{\frac{1}{\hat{y}_{i+1/2,j-1/2}} \Delta y_{i+1/2} + \frac{1}{\hat{y}_{i-1/2,j-1/2}} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}}, \tag{A13}$$

$$\frac{1}{\hat{y}_{i,j}^r} = \frac{\frac{1}{\hat{y}_{i+1/2,j+1/2}} \Delta z_{j+1/2} + \frac{1}{\hat{y}_{i+1/2,j-1/2}} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}}, \tag{A14}$$

$$\frac{1}{\hat{y}_{i,j}^l} = \frac{\frac{1}{\hat{y}_{i-1/2,j+1/2}} \Delta z_{j+1/2} + \frac{1}{\hat{y}_{i-1/2,j-1/2}} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}}. \tag{A15}$$

Again, rearranging gives

$$\begin{aligned}
0 = & 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j+1/2}} \frac{1}{\hat{y}_{i,j}^d} H_x^{i,j+1} + 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j-1/2}} \frac{1}{\hat{y}_{i,j}^u} H_x^{i,j-1} \\
& + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i+1/2}} \frac{1}{\hat{y}_{i,j}^r} H_x^{i+1,j} + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i-1/2}} \frac{1}{\hat{y}_{i,j}^l} H_x^{i-1,j} \\
& - \left\{ 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j+1/2}} \frac{1}{\hat{y}_{i,j}^d} + 2 \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\Delta z_{j-1/2}} \frac{1}{\hat{y}_{i,j}^u} \right. \\
& \quad \left. + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i+1/2}} \frac{1}{\hat{y}_{i,j}^r} + 2 \frac{\Delta z_{j+1/2} + \Delta z_{j-1/2}}{\Delta y_{i-1/2}} \frac{1}{\hat{y}_{i,j}^l} \right. \\
& \quad \left. + \hat{z} (\Delta y_{i+1/2} + \Delta y_{i-1/2}) (\Delta z_{j+1/2} + \Delta z_{j-1/2}) \right\} H_x^{i,j}.
\end{aligned} \tag{A16}$$

If considered at all inner mesh nodes (i, j) , eqs (A8) and (A16) form systems of linear equations in the unknown horizontal electric and magnetic field components of the TE-mode and the TM-mode, respectively. Assuming that there are N_{za} air cells, N_{zb} earth cells and a total of $N_z = N_{za} + N_{zb}$ cells in the vertical direction and N_y cells in the horizontal direction, the horizontal field components E_x or H_x are to be computed at $(N_z - 1) \cdot (N_y - 1)$ inner mesh nodes. Boundary values have to be supplied at the edges of the mesh. Along the upper edge of the air half-space, the incident plane wave is assumed to have unit amplitude and zero phase. At the lower edge of the earth half-space, the electromagnetic (EM) field is assumed to have totally decayed, and along the lateral edges, the horizontal field components E_x or H_x are assumed to be that of the corresponding 1-D admittivity section along the particular side. This results in a system of $(N_z - 1) \cdot (N_y - 1)$ linear equations $\mathbf{K}\mathbf{x} = \mathbf{s}$ (one equation for each interior node), with the coefficient matrix $\mathbf{K}$, the vector $\mathbf{x}$ of unknown horizontal field components of the TE- or TM-mode and a vector $\mathbf{s}$ of boundary values. If a central node is located next to one or two boundary nodes, the terms in eqs (A8) or (A16), which contain the electric or magnetic boundary field components, are placed in the corresponding row of the right-hand side vector $\mathbf{s}$. If the nodes are arranged such that the vertical index j varies fastest, the finite-difference eq. (A8) or (A16) of central node (i, j) is contained in row number $(i - 2)(N_z - 1) + (j - 1)$ of $\mathbf{K}\mathbf{x} = \mathbf{s}$.

The auxiliary field components (H_y, H_z) of the TE-mode are derived as partial derivatives of E_x at the air-Earth interface (nodes at $j = N_{za} + 1$), by expanding E_x in a Taylor series of second order, around the considered node (i, j) and substituting eq. (13) for the second-order term as proposed by Weaver *et al.* (1986). Vertical expansion, both upwards and downwards from the considered node yields a central

difference formula for the horizontal magnetic field component H_y as

$$\begin{aligned} -\hat{z}H_y^{i,j} &= \left(\frac{\partial E_x}{\partial z} \right)_{i,j} \\ &= \frac{\Delta z_{j+1/2} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \left\{ \frac{E_x^{i,j+1}}{\Delta z_{j+1/2}^2} - \frac{E_x^{i,j-1}}{\Delta z_{j-1/2}^2} \right. \\ &\quad \left. - \left[\frac{1}{\Delta z_{j+1/2}^2} - \frac{1}{\Delta z_{j-1/2}^2} + \frac{1}{2} \hat{z} (\hat{y}_{i,j}^d - \hat{y}_{i,j}^u) \right] E_x^{i,j} \right\}, \end{aligned} \quad (\text{A17})$$

where

$$\hat{y}_{i,j}^d = \frac{\hat{y}_{i+1/2,j+1/2} \Delta y_{i+1/2} + \hat{y}_{i-1/2,j+1/2} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}}, \quad (\text{A18})$$

$$\hat{y}_{i,j}^u = \frac{\hat{y}_{i+1/2,j-1/2} \Delta y_{i+1/2} + \hat{y}_{i-1/2,j-1/2} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}}. \quad (\text{A19})$$

Similarly, horizontal expansion both to the left- and right-hand side of the considered node yields a central difference formula for the vertical magnetic field component H_z of the form

$$\begin{aligned} \hat{z}H_z^{i,j} &= \left(\frac{\partial E_x}{\partial y} \right)_{i,j} \\ &= \frac{\Delta y_{i+1/2} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}} \left\{ \frac{E_x^{i+1,j}}{\Delta y_{i+1/2}^2} - \frac{E_x^{i-1,j}}{\Delta y_{i-1/2}^2} \right. \\ &\quad \left. - \left[\frac{1}{\Delta y_{i+1/2}^2} - \frac{1}{\Delta y_{i-1/2}^2} + \frac{1}{2} \hat{z} (\hat{y}_{i,j}^r - \hat{y}_{i,j}^l) \right] E_x^{i,j} \right\} \end{aligned} \quad (\text{A20})$$

where

$$\hat{y}_{i,j}^r = \frac{\hat{y}_{i+1/2,j+1/2} \Delta z_{j+1/2} + \hat{y}_{i+1/2,j-1/2} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \quad (\text{A21})$$

$$\hat{y}_{i,j}^l = \frac{\hat{y}_{i-1/2,j+1/2} \Delta z_{j+1/2} + \hat{y}_{i-1/2,j-1/2} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}}. \quad (\text{A22})$$

A corresponding derivation of the auxiliary electric field components (E_y , E_z) of the TM-mode follows Weaver *et al.* (1985). The horizontal electric field component E_y is computed by expanding H_x in a Taylor series, both upwards and downwards from the considered node (i,j). Hence,

$$j_y^{i,j} = \left(\frac{\partial H_x}{\partial z} \right)_{i,j} = \frac{N_{i,j}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}}, \quad (\text{A23})$$

where

$$\begin{aligned} N_{i,j} &= \frac{\Delta z_{j-1/2}}{\Delta z_{j+1/2}} H_x^{i,j+1} - \frac{\Delta z_{j+1/2}}{\Delta z_{j-1/2}} H_x^{i,j-1} - \frac{O_{i,j}}{2} \left(\frac{H_x^{i+1,j}}{\Delta y_{i+1/2}} - \frac{H_x^{i-1,j}}{\Delta y_{i-1/2}} \right) \\ &\quad + \left[\frac{\Delta z_{j+1/2}}{\Delta z_{j-1/2}} - \frac{\Delta z_{j-1/2}}{\Delta z_{j+1/2}} + \frac{1}{2} \frac{\Delta y_{i-1/2} - \Delta y_{i+1/2}}{\Delta y_{i-1/2} \Delta y_{i+1/2}} O_{i,j} \right. \\ &\quad \left. + \hat{z} \frac{\Delta z_{j-1/2} \Delta z_{j+1/2}}{2} (\hat{y}_{i,j}^u - \hat{y}_{i,j}^d) \right] H_x^{i,j}, \\ O_{i,j} &= -\frac{\Delta z_{j-1/2} \Delta z_{j+1/2}}{\Delta y_{i-1/2} + \Delta y_{i+1/2}} \left[\hat{y}_{i,j}^d \left(\frac{1}{\hat{y}_{i+1/2,j+1/2}} - \frac{1}{\hat{y}_{i-1/2,j+1/2}} \right) \right. \\ &\quad \left. - \hat{y}_{i,j}^u \left(\frac{1}{\hat{y}_{i+1/2,j-1/2}} - \frac{1}{\hat{y}_{i-1/2,j-1/2}} \right) \right]. \end{aligned}$$

After obtaining $j_y^{i,j}$ from the above equations, the two one-sided values $E_y^{i-,j}$ and $E_y^{i+,j}$ of the electric field component E_y can be computed. In contrast to eq. (24) in Weaver *et al.* (1985), the current density $j_y^{i,j}$ must be divided by the left- and right-hand sided vertically averaged admittivities (eqs A22 and A21, respectively) to obtain $E_y^{i-,j}$ and $E_y^{i+,j}$, respectively, that is,

$$\hat{y}_{i,j}^l E_y^{i-,j} = \hat{y}_{i,j}^r E_y^{i+,j} = j_y^{i,j}. \quad (\text{A24})$$

The use of averaged admittivities is motivated by considering the integrated current $I_y = \iint_{x-z\text{-plane}} j_y \, dx \, dz$ through any surface $y = \text{const.}$

To assign a unique value to $E_y^{i,j}$, the current density $j_y^{i,j}$ is typically divided by the average admittivity $\hat{y}_{i,j}^{\text{avg}}$ given in eq. (A7), that is,

$$\hat{y}_{i,j}^{\text{avg}} E_y^{i,j} = j_y^{i,j}. \quad (\text{A25})$$

APPENDIX B: SOLUTION METHODS FOR LINEAR SYSTEMS

As shown in Appendix A, the FDAs of Helmholtz eqs (13) and (14) result in system matrices that are sparse, complex and symmetric with two subdiagonals and two super-diagonals. For the solution of the FDAs, both direct and iterative solvers are desirable. The solution with a direct method is rewarding, as soon as the same system matrix is used for the solution with multiple right-hand side vectors (Rodi 1976; Siripunvaraporn & Egbert 2000), for example, for multiple pseudo-forward problems arising in the computation of the sensitivity matrix (see Appendix C). As the system matrix is non-Hermitian and, hence, not positive-definite, the LU-decomposition rather than the Cholesky-decomposition has to be used as a direct method (Golub & van Loan 1996). If the system of linear equations is to be solved for a single right-hand side vector, an iterative solver can provide significant computational savings over a direct method. The iterative bi-conjugate gradient method (BiCG), which is used by Siripunvaraporn & Egbert (2000) for quasi-static problems, breaks down if applied to the general forward problem with displacement currents. Freund (1992) gives two possible reasons for the breakdown. As the proper inner product for the base vectors of a Krylov subspace constructed during the Lanczos process of complex-symmetric matrices is $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$, quasi-null vectors $(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j) = \tilde{\mathbf{v}}_j^T \tilde{\mathbf{v}}_j = 0$ may occur and cause a division by zero, during the normalization of the new base vector $\mathbf{v}_j = \tilde{\mathbf{v}}_j / (\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)$. In computer arithmetic, the tridiagonal matrix, constructed during the Lanczos-process, might be nearly singular and the solution update, constructed from the Petrov-Galerkin conditions, might give a bad approximation to the true solution, leading to erratic convergence behaviour with wildly varying residual norms. It should be possible to circumvent these problems by using the quasi-minimal residual method (QMR) as proposed by Freund (1992), where the problem of quasi-null vectors is amended with look-ahead techniques and the Petrov-Galerkin conditions are replaced by a quasi-minimal residual property. We have not yet succeeded, however, in implementing the QMR solver for the general forward problem with displacement currents. Therefore, we use the direct LU-method as the sole solver at present.

APPENDIX C: COMPUTATION OF THE SENSITIVITY MATRIX

An efficient scheme for the computation of sensitivity matrices was proposed by Rodi (1976) and Rodi & Mackie (2001). The k th impedance or VMT datum for a given model $\mathbf{m}$ is expressed in terms of the horizontal electric or magnetic field component of the TE- or TM-mode, respectively, as

$$Z^k(\mathbf{m}) = \frac{\mathbf{a}_k(\mathbf{m})^T \mathbf{x}(\mathbf{m})}{\mathbf{b}_k(\mathbf{m})^T \mathbf{x}(\mathbf{m})}, \quad (\text{C1})$$

where

$$\mathbf{x}(\mathbf{m}) = \begin{cases} H_x & \text{at inner mesh nodes for TM-mode impedance} \\ E_x & \text{at inner mesh nodes for TE-mode impedance or VMT} \end{cases} \quad (\text{C2})$$

and $\mathbf{a}_k(\mathbf{m})$ and $\mathbf{b}_k(\mathbf{m})$ are coefficient vectors from the central difference computation of the auxiliary fields in the TM-mode impedance, TE-mode impedance and VMT.

The entry of the sensitivity matrix for the k th impedance or VMT datum with respect to (w.r.t.) the l th model parameter is then computed as

$$\begin{aligned} J^{kl}(\mathbf{m}) &= \frac{\partial Z^k(\mathbf{m})}{\partial m^l} \\ &= \frac{1}{\mathbf{b}_k^T \mathbf{x}} \frac{\partial (\mathbf{a}_k^T \mathbf{x})}{\partial m^l} - \frac{\mathbf{a}_k^T \mathbf{x}}{(\mathbf{b}_k^T \mathbf{x})^2} \frac{\partial (\mathbf{b}_k^T \mathbf{x})}{\partial m^l} \\ &= \left(\frac{1}{\mathbf{b}_k^T \mathbf{x}} \frac{\partial \mathbf{a}_k}{\partial m^l} - \frac{\mathbf{a}_k^T \mathbf{x}}{(\mathbf{b}_k^T \mathbf{x})^2} \frac{\partial \mathbf{b}_k}{\partial m^l} \right)^T \mathbf{x} + \left(\frac{1}{\mathbf{b}_k^T \mathbf{x}} \mathbf{a}_k - \frac{\mathbf{a}_k^T \mathbf{x}}{(\mathbf{b}_k^T \mathbf{x})^2} \mathbf{b}_k \right)^T \frac{\partial \mathbf{x}}{\partial m^l}. \end{aligned} \quad (\text{C3})$$

The definitions

$$\mathbf{c}_k = \frac{1}{\mathbf{b}_k^T \mathbf{x}} \mathbf{a}_k - \frac{\mathbf{a}_k^T \mathbf{x}}{(\mathbf{b}_k^T \mathbf{x})^2} \mathbf{b}_k, \quad (\text{C4})$$

$$\mathbf{d}_{kl} = \left(\frac{1}{\mathbf{b}_k^T \mathbf{x}} \frac{\partial \mathbf{a}_k}{\partial m^l} - \frac{\mathbf{a}_k^T \mathbf{x}}{(\mathbf{b}_k^T \mathbf{x})^2} \frac{\partial \mathbf{b}_k}{\partial m^l} \right) \quad (\text{C5})$$

and the relation from the forward problem,

$$\mathbf{K} \frac{\partial \mathbf{x}}{\partial m^l} = -\frac{\partial \mathbf{K}}{\partial m^l} \mathbf{x} + \frac{\partial \mathbf{s}}{\partial m^l}, \quad (\text{C6})$$

give

$$\begin{aligned} J^{kl}(\mathbf{m}) &= \mathbf{d}_{kl}^T \mathbf{x} + \mathbf{c}_k^T \mathbf{K}^{-1} \left(-\frac{\partial \mathbf{K}}{\partial m^l} \mathbf{x} + \frac{\partial \mathbf{s}}{\partial m^l} \right) \\ &= \mathbf{d}_{kl}^T \mathbf{x} + \mathbf{u}_k^T \left(-\frac{\partial \mathbf{K}}{\partial m^l} \mathbf{x} + \frac{\partial \mathbf{s}}{\partial m^l} \right), \end{aligned} \quad (\text{C7})$$

where the computation of $\mathbf{u}_k$ as a solution of the pseudo-forward problem $\mathbf{K}^T \mathbf{u}_k = \mathbf{c}_k$ exploits the reciprocity of the forward problem (i.e. the symmetry of the forward system matrix $\mathbf{K}$ and its inverse). The computation of $\mathbf{u}_k$ for $k = 1, \dots, N$ is significantly faster than the computation of $\mathbf{K}^{-1}(-\frac{\partial \mathbf{K}}{\partial m^l} \mathbf{x} + \frac{\partial \mathbf{s}}{\partial m^l})$ for $l = 1, \dots, M$ in the case of typical 2-D problems, where the number of model parameters exceeds the number of data, that is, where $M > N$.

A model parameter $m^l = \rho_{i+1/2, j+1/2}$ is connected to the cell to the lower right-hand side of a node (i, j) through $l = j - N_{za} + (i - 1) * N_{zb}$. Similarly, each data index k is connected to a single surface node $(i_s, j_s = N_{za} + 1)$ for a given frequency, where i_s and j_s indicate the node at which a certain receiver station is located.

C1 TM-mode

The computation of the sensitivities for the TM-mode turns out to be intriguingly complicated, as the derivatives of inverse admittivities w.r.t. resistivities are involved.

In the TM-mode, E_y is expressed through $\mathbf{a}_k$ and H_x according to eqs (A23) and (A25) and $\mathbf{b}_k$ is zero except for the k th entry, which is 1. Hence,

$$\mathbf{a}_k^T \mathbf{x} = E_y^k,$$

$$\mathbf{b}_k^T \mathbf{x} = H_x^k$$

and

$$\mathbf{c}_k = \frac{1}{H_x^k} \mathbf{a}_k - \frac{Z_{yx}^k}{H_x^k} \left(0, \dots, 0, \underbrace{1}_{k\text{th entry}}, 0, \dots, 0 \right), \quad (\text{C8})$$

$$\mathbf{d}_{kl} = \frac{1}{H_x^k} \frac{\partial \mathbf{a}_k}{\partial m^l}. \quad (\text{C9})$$

The computation of $\mathbf{d}_{kl}$ is simplified as $\mathbf{b}_k$ does not depend on any model parameter. Furthermore, as $\mathbf{a}_k$ is computed with a five-point-stencil FDA, each $\mathbf{a}_k$ depends only on the admittivities of the four cells surrounding a node (i_s, j_s) , with a receiver, and sensitivities are only computed for the resistivities of two such cells, that is, those immediately below the surface. Consequently, the indices of the involved model parameters are $i_s - 1/2, j_s + 1/2$ and $i_s + 1/2, j_s + 1/2$, respectively. For brevity, the notation $i = i_s$ and $j = j_s$ is used. Eqs (A23) and (A25) yield

$$\begin{aligned} \frac{\partial \mathbf{a}_k^T}{\partial \rho_{i \pm 1/2, j+1/2}} \mathbf{x} &= \mathbf{j}_y^k \frac{\partial \hat{y}_{i,j}^{\text{avg} \pm}}{\partial \rho_{i \pm 1/2, j+1/2}} + \frac{1}{\hat{y}_{i,j}^{\text{avg}}} \frac{1}{2} \frac{\Delta z_{j-1/2} \Delta z_{j+1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \\ &\quad \left\{ -\hat{z} \frac{\partial \hat{y}_{i,j}^d}{\partial \rho_{i \pm 1/2, j+1/2}} H_x^{i,j} + \frac{1}{\Delta y_{i-1/2} + \Delta y_{i+1/2}} \right. \\ &\quad \left. \frac{\partial}{\partial \rho_{i \pm 1/2, j+1/2}} \left[\hat{y}_{i,j}^d \left(\frac{1}{\hat{y}_{i+1/2, j+1/2}} - \frac{1}{\hat{y}_{i-1/2, j+1/2}} \right) \right] \right. \\ &\quad \left. \left(\frac{H_x^{i+1,j}}{\Delta y_{i+1/2}} + \frac{\Delta y_{i+1/2} - \Delta y_{i-1/2}}{\Delta y_{i-1/2} \Delta y_{i+1/2}} H_x^{i,j} - \frac{H_x^{i-1,j}}{\Delta y_{i-1/2}} \right) \right\} \\ &= \mathbf{j}_y^k \frac{\partial \hat{y}_{i,j}^{\text{avg} \pm}}{\partial \rho_{i \pm 1/2, j+1/2}} + \frac{1}{\hat{y}_{i,j}^{\text{avg}}} \frac{1}{2} \frac{\Delta z_{j-1/2} \Delta z_{j+1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \\ &\quad \left[\left(-\hat{z} \frac{\partial \hat{y}_{i,j}^{\pm}}{\partial \rho_{i \pm 1/2, j+1/2}} + \frac{\Delta y_{i+1/2} - \Delta y_{i-1/2}}{\Delta y_{i+1/2} \Delta y_{i-1/2}} P_{i,j}^{\pm} \right) H_x^{i,j} \right. \\ &\quad \left. + \frac{1}{\Delta y_{i+1/2}} P_{i,j}^{\pm} H_x^{i+1,j} - \frac{1}{\Delta y_{i-1/2}} P_{i,j}^{\pm} H_x^{i-1,j} \right], \quad (\text{C10}) \end{aligned}$$

where the definitions

$$\begin{aligned} \frac{\partial \hat{y}_{i,j}^{\text{avg} \pm}}{\partial \rho_{i \pm 1/2, j+1/2}} &= \frac{\partial}{\partial \rho_{i \pm 1/2, j+1/2}} \left(\frac{1}{\hat{y}_{i,j}^{\text{avg}}} \right) \\ &= \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{(\hat{y}_{i,j}^r \Delta y_{i+1/2} + \hat{y}_{i,j}^l \Delta y_{i-1/2})^2} \Delta y_{i \pm 1/2} \frac{\Delta z_{j+1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \sigma_{i \pm 1/2, j+1/2}^2 \\ &= \frac{1}{\hat{y}_{i,j}^{\text{avg}}} \frac{1}{\hat{y}_{i,j}^r \Delta y_{i+1/2} + \hat{y}_{i,j}^l \Delta y_{i-1/2}} \Delta y_{i \pm 1/2} \frac{\Delta z_{j+1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \sigma_{i \pm 1/2, j+1/2}^2, \end{aligned}$$

$$\begin{aligned}
\partial \hat{y}_{i,j}^{d\pm} &= \frac{\partial \hat{y}_{i,j}^d}{\partial \rho_{i\pm 1/2, j+1/2}} \\
&= - \frac{\Delta y_{i+1/2} + \Delta y_{i-1/2}}{\left(\frac{1}{\hat{y}_{i+1/2, j+1/2}} \Delta y_{i+1/2} + \frac{1}{\hat{y}_{i-1/2, j+1/2}} \Delta y_{i-1/2} \right)^2} \Delta y_{i\pm 1/2} \frac{\sigma_{i\pm 1/2, j+1/2}^2}{\hat{y}_{i\pm 1/2, j+1/2}^2} \\
&= - \frac{\hat{y}_{i,j}^d}{\frac{1}{\hat{y}_{i+1/2, j+1/2}} \Delta y_{i+1/2} + \frac{1}{\hat{y}_{i-1/2, j+1/2}} \Delta y_{i-1/2}} \Delta y_{i\pm 1/2} \frac{\sigma_{i\pm 1/2, j+1/2}^2}{\hat{y}_{i\pm 1/2, j+1/2}^2}, \\
P_{i,j}^{\pm} &= \frac{1}{\Delta y_{i+1/2} + \Delta y_{i-1/2}} \frac{\partial}{\partial \rho_{i\pm 1/2, j+1/2}} \left[\hat{y}_{i,j}^d \left(\frac{1}{\hat{y}_{i+1/2, j+1/2}} - \frac{1}{\hat{y}_{i-1/2, j+1/2}} \right) \right] \\
&= \frac{1}{\Delta y_{i+1/2} + \Delta y_{i-1/2}} \left[\partial \hat{y}_{i,j}^{d\pm} \left(\frac{1}{\hat{y}_{i+1/2, j+1/2}} - \frac{1}{\hat{y}_{i-1/2, j+1/2}} \right) \pm \hat{y}_{i,j}^d \frac{\sigma_{i\pm 1/2, j+1/2}^2}{\hat{y}_{i\pm 1/2, j+1/2}^2} \right]
\end{aligned}$$

were used. The quantity $\hat{y}_{i,j}^{\text{avg}}$ is given according to eq. (A7).

The derivative of the system matrix of the forward problem $\mathbf{K}$ w.r.t. a single model parameter m^l in eq. (C7) results in a matrix $\partial \mathbf{K} / \partial m^l$ that has only four rows with non-zero entries. The parameter $m^l = \rho_{i+1/2, j+1/2}$ enters into the rows of $\mathbf{K}$ that correspond to the central nodes (i, j) , $(i, j + 1)$, $(i + 1, j)$ and $(i + 1, j + 1)$ (cf. Fig. A1), that is, into rows number

$$iul = (i - 2)(N_z - 1) + (j - 1) \quad (\text{C11})$$

$$idl = (i - 2)(N_z - 1) + j \quad (\text{C12})$$

$$iur = (i - 1)(N_z - 1) + (j - 1) \quad (\text{C13})$$

$$idr = (i - 1)(N_z - 1) + j. \quad (\text{C14})$$

The computation of $\partial \mathbf{K} / \partial m^l$ is further simplified by the symmetry of $\mathbf{K}$. For a central node (i, j) the coefficient of the EM field component at its right-hand side node is the same as the coefficient of the EM field component at the left-hand side node of its neighbouring central node $(i + 1, j)$. Similarly, for a central node (i, j) the coefficient of the EM field component at its lower node is the same as the coefficient of the EM field component at the upper node of its neighbouring central node $(i, j + 1)$. Furthermore, as the left- and right-hand coefficients contain vertically averaged inverse admittivities and the lower and upper coefficients contain horizontally averaged inverse admittivities, the derivative of the coefficient of the right-hand node of the central node (i, j) w.r.t. $\rho_{i+1/2, j+1/2}$ equals the derivative of the coefficient of the right-hand node of its neighbouring central node $(i, j + 1)$ w.r.t. $\rho_{i+1/2, j+1/2}$. Similar rules are valid for the coefficients of left-hand, upper and lower nodes at correspondingly neighbouring nodes. Hence,

$$\begin{aligned}
\frac{\partial \mathbf{K}(iul, iul + (N_z - 1))}{\partial \rho_{i+1/2, j+1/2}} &= \frac{\partial \mathbf{K}(idl, idl + (N_z - 1))}{\partial \rho_{i+1/2, j+1/2}} = \frac{\partial \mathbf{K}(iur, iur - (N_z - 1))}{\partial \rho_{i+1/2, j+1/2}} \\
&= \frac{\partial \mathbf{K}(idr, idr - (N_z - 1))}{\partial \rho_{i+1/2, j+1/2}} = 2 \frac{\Delta z_{j+1/2}}{\Delta y_{i+1/2}} \frac{\sigma_{i+1/2, j+1/2}^2}{\hat{y}_{i+1/2, j+1/2}^2}, \quad (\text{C15})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbf{K}(iul, iul + 1)}{\partial \rho_{i+1/2, j+1/2}} &= \frac{\partial \mathbf{K}(iur, iur + 1)}{\partial \rho_{i+1/2, j+1/2}} = \frac{\partial \mathbf{K}(idl, idl - 1)}{\partial \rho_{i+1/2, j+1/2}} \\
&= \frac{\partial \mathbf{K}(idr, idr - 1)}{\partial \rho_{i+1/2, j+1/2}} = 2 \frac{\Delta y_{i+1/2}}{\Delta z_{j+1/2}} \frac{\sigma_{i+1/2, j+1/2}^2}{\hat{y}_{i+1/2, j+1/2}^2}, \quad (\text{C16})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbf{K}(iul, iul)}{\partial \rho_{i+1/2, j+1/2}} &= \frac{\partial \mathbf{K}(iur, iur)}{\partial \rho_{i+1/2, j+1/2}} = \frac{\partial \mathbf{K}(idl, idl)}{\partial \rho_{i+1/2, j+1/2}} \\
&= \frac{\partial \mathbf{K}(idr, idr)}{\partial \rho_{i+1/2, j+1/2}} \\
&= -2 \frac{\Delta z_{j+1/2}}{\Delta y_{i+1/2}} \frac{\sigma_{i+1/2, j+1/2}^2}{\hat{y}_{i+1/2, j+1/2}^2} - 2 \frac{\Delta y_{i+1/2}}{\Delta z_{j+1/2}} \frac{\sigma_{i+1/2, j+1/2}^2}{\hat{y}_{i+1/2, j+1/2}^2}. \quad (\text{C17})
\end{aligned}$$

C2 TE-mode

In the TE-mode, H_y is expressed through $\mathbf{b}_k$ and E_x according to eq. (A17) and $\mathbf{a}_k$ is zero except for the k th entry, which is 1. Hence,

$$\mathbf{a}_k^T \mathbf{x} = E_x^k,$$

$$\mathbf{b}_k^T \mathbf{x} = H_y^k$$

and

$$\mathbf{c}_k = \frac{1}{H_y^k} \left(0, \dots, 0, \underbrace{1}_{k\text{th entry}}, 0, \dots, 0 \right) - \frac{Z_{xy}^k}{H_y^k} \mathbf{b}_k, \quad (\text{C18})$$

$$\mathbf{d}_{kl} = -\frac{Z_{xy}^k}{H_x^k} \frac{\partial \mathbf{b}_k}{\partial m^l}, \quad (\text{C19})$$

as $\mathbf{a}_k$ does not depend on any model parameter.

The derivative in $\mathbf{d}_{kl}^T \mathbf{x}$ is computed from eq. (A17) as

$$\frac{\partial \mathbf{b}_k^T}{\partial \rho_{i\pm 1/2, j+1/2}} \mathbf{x} = \frac{1}{2} \frac{\Delta z_{j+1/2} \Delta z_{j-1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \left(-\sigma_{i\pm 1/2, j+1/2}^2 \Delta y_{i\pm 1/2} \frac{1}{\Delta y_{i+1/2} + \Delta y_{i-1/2}} \right) E_x^{i,j}. \quad (\text{C20})$$

As m^l only enters into the coefficient of the central node in eq. (A8), the matrix $\partial \mathbf{K} / \partial m^l$ contains only four non-zero entries, which are all on the diagonal. Hence,

$$\begin{aligned} \frac{\partial \mathbf{K}(iul, iul)}{\partial \rho_{i+1/2, j+1/2}} &= \frac{\partial \mathbf{K}(idl, idl)}{\partial \rho_{i+1/2, j+1/2}} = \frac{\partial \mathbf{K}(iur, iur)}{\partial \rho_{i+1/2, j+1/2}} = \frac{\partial \mathbf{K}(idr, idr)}{\partial \rho_{i+1/2, j+1/2}} \\ &= -\hat{z} \left(-\sigma_{i+1/2, j+1/2}^2 \Delta z_{j+1/2} \Delta y_{i+1/2} \right). \end{aligned} \quad (\text{C21})$$

C3 VMT mode

In the VMT mode, H_z is expressed through $\mathbf{a}_k$ and E_x according to eq. (A20) and H_y is expressed through $\mathbf{b}_k$ and E_x according to eq. (A17). Hence,

$$\mathbf{a}_k^T \mathbf{x} = H_z^k,$$

$$\mathbf{b}_k^T \mathbf{x} = H_y^k$$

and

$$\mathbf{c}_k = \frac{1}{H_y^k} \mathbf{a}_k - \frac{B_k}{H_y^k} \mathbf{b}_k, \quad (\text{C22})$$

$$\mathbf{d}_{kl} = \frac{1}{H_y^k} \frac{\partial \mathbf{a}_k}{\partial m^l} - \frac{B_k}{H_y^k} \frac{\partial \mathbf{b}_k}{\partial m^l}. \quad (\text{C23})$$

The derivative in the first term of $\mathbf{d}_{kl}^T \mathbf{x}$ is computed from eq. (A20) as

$$\frac{\partial \mathbf{a}_k^T}{\partial \rho_{i\pm 1/2, j+1/2}} \mathbf{x} = \frac{1}{2} \frac{\Delta y_{i+1/2} \Delta y_{i-1/2}}{\Delta y_{i+1/2} + \Delta y_{i-1/2}} \left(\pm \sigma_{i\pm 1/2, j+1/2}^2 \frac{\Delta z_{j+1/2}}{\Delta z_{j+1/2} + \Delta z_{j-1/2}} \right) E_x^{i,j}. \quad (\text{C24})$$

The derivative in the second term of $\mathbf{d}_{kl}^T \mathbf{x}$, that is, $(\partial \mathbf{b}_k / \partial m^l)^T \mathbf{x}$, is already given by eq. (C20).

As the linear system of equations that is solved in the forward problem of the VMT is the one solved in the TE-mode, the entries of the matrix $\partial \mathbf{K} / \partial m^l$ are given as in eq. (C21).