



**UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO**

**Geostatistical inversion of geophysical data for near-surface
modelling and characterization**

João Miguel Figueiredo Narciso

Supervisor: Doctor Leonardo Azevedo Guerra Raposo Pereira

Co-Supervisor: Doctor Ellen Van De Vijver

Thesis approved in public session to obtain the PhD Degree in

Earth Resources

Jury final classification: **Pass with Distinction**

2023



UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

**Geostatistical inversion of geophysical data for near-surface modelling and
characterization**

João Miguel Figueiredo Narciso

Supervisor: Doctor Leonardo Azevedo Guerra Raposo Pereira

Co-Supervisor: Doctor Ellen Van De Vijver

Thesis approved in public session to obtain the PhD Degree in
Earth Resources

Jury final classification: **Pass with Distinction**

Jury

Chairperson: Doctor Maria João Correia Colunas Pereira, Instituto Superior Técnico,
Universidade de Lisboa

Members of the Committee:

Doctor Thomas Hermans, Faculty of Sciences, University of Ghent, Belgium

Doctor Amílcar de Oliveira Soares, Instituto Superior Técnico, Universidade de Lisboa

Doctor Maria João Correia Colunas Pereira, Instituto Superior Técnico, Universidade de
Lisboa

Doctor Giulio Vignoli, Università Degli Studi di Cagliari, Italy

Doctor Ellen Van De Vijver, Faculty of Bioscience Engineering, University of Ghent, Belgium

Funding Institution

FCT – Fundação para a Ciência e a Tecnologia

2023

To A. and S.,
for giving me future in the present.

SUMMARY

The near-surface beneath our feet is the portion of the Earth that affects and is more impacted by human activities and yields important mineral and energy resources. It is, therefore, of the utmost interest to characterize the first meters of the subsurface and to accurately quantify its physical, structural, chemical, and biological properties. To overcome the limitations of direct measurements obtained from invasive methods, non-invasive geophysical methods have been applied in the modelling and characterization of complex and heterogeneous near-subsurface environments. Particularly frequency-domain electromagnetic (FDEM) induction methods have become one of the most widely used geophysical methods in near-surface applications due to their versatility, cost-effectiveness, and data sensitivity to subsurface changes of two physical properties: electrical conductivity (EC) and magnetic susceptibility (MS).

However, mapping geophysical data into numerical subsurface models concerns solving an ill-posed and nonlinear geophysical inverse problem with multiple solutions. While deterministic geophysical inverse solutions allow predicting smooth representations of the subsurface, they do not account for uncertainties and are unable to directly integrate direct observations, a probabilistic framework allows overcoming these limitations.

This thesis combines the advantages of FDEM induction measurements with the potential of probabilistic inversion and introduces a geostatistical FDEM inversion method to simultaneously model the spatial distribution of the subsurface EC and MS and assess the uncertainty of the predicted results. The proposed method is benchmarked with an alternative statistical-based FDEM inversion method. Since probabilistic inversion methods are computationally demanding when solving for large-scale three-dimensional inverse problems, the iterative geostatistical FDEM inversion is coupled with random tensor decomposition to alleviate the computational burden. From a multi-geophysical inversion approach, this thesis also presents a joint inversion method of electrical and electromagnetic data to reduce the uncertainty of the predicted subsurface models in near-surface applications. The methods are illustrated in both realistic synthetic and real application examples.

KEYWORDS: Near-surface modelling, Geostatistical FDEM inversion, FDEM data, Electrical data, Joint geophysical inversion.

RESUMO

A subsuperfície debaixo dos nossos pés é a parte da Terra que mais interage com actividades humanas e que armazena uma parte importante das águas subterrâneas e recursos minerais. É, portanto, fundamental caracterizar os primeiros metros da subsuperfície e quantificar com precisão as suas propriedades físicas, estruturais, químicas e biológicas. Para ultrapassar as limitações espaciais de medições directas por métodos invasivos, métodos geofísicos não invasivos têm sido aplicados na modelação e caracterização de ambientes subsuperficiais complexos e heterogéneos. Em particular, os métodos de indução electromagnética no domínio da frequência (FDEM) tornaram-se um dos métodos geofísicos mais utilizados nestes depósitos heterogéneos, devido à sensibilidade dos seus dados às alterações de duas propriedades físicas: a condutividade eléctrica e a susceptibilidade magnética. No entanto, transformar dados geofísicos em modelos espaciais da subsuperfície implica a resolução de um problema geofísico inverso, que é um problema não linear com múltiplas soluções. A resolução deste problema através de uma abordagem geoestatística permite combinar medições directas invasivas com dados geofísicos, para assim melhorar os modelos invertidos.

Combinando as vantagens dos dados FDEM com o potencial das metodologias de inversão geoestatística, esta tese apresenta um método de inversão geoestatística FDEM para modelar a distribuição espacial das propriedades na subsuperfície e avaliar a incerteza dos resultados previstos. A metodologia proposta é comparada com outro método probabilístico de inversão FDEM. Como os métodos de inversão geoestatísticos são computacionalmente exigentes quando se trata de resolver problemas inversos tridimensionais de grande escala, é apresentado um método de inversão probabilístico com um algoritmo de aprendizagem automática para melhorar a performance computacional. A partir de uma abordagem de inversão multi-geofísica, esta tese apresenta também um método de inversão conjunta de dados eléctricos e electromagnéticos que visa reduzir a incerteza dos modelos de subsuperfície previstos. Os métodos apresentados são aplicados em casos sintéticos e reais.

PALAVRAS-CHAVE: Modelação da subsuperfície, Inversão electromagnética geoestatística, Dados electromagnéticos, Dados Eléctricos, Inversão geofísica conjunta.

ACKNOWLEDGMENTS

“Mas quem vencer esta meta
Que diga se a linha é recta”,

in “De não saber o que me espera”, José Afonso

“Ensinas-me a fazer tantas perguntas
Na volta das respostas que eu trazia”, in “Inquietação”, José Mário Branco

To my supervisors / to my friends, Leonardo and Ellen, this was a ride only possible with you. Thank you for challenging/guiding/twisting/appeasing/revolting/growing me. Thank you for opening doors and for giving me ground. Thank you for your support in my darkest moments and your shade in my brightest. Thank you for being the model of scientific and ethical excellence to follow and for making the future now.

“Take these broken wings and learn to fly.
All your life, ”, in “Black Bird”, Beatles

Thanks to Fundação para a Ciência e Tecnologia (FCT) for funding and supporting this project and to CERENA research centre for providing all the conditions and opportunities to carry out my research. Thanks also to the ORBit research group (Ghent University) for hosting me on such a fruitful research stay.

“I’m pickin’ up good vibrations”, in “Good Vibrations”, Beach Boys

My deepest thanks to all the researchers and colleagues with whom I collaborated, carried out field work, provided me with their data and with whom I published. In particular to Christin, Mingliang, Dario, Daan, Guillaume, Jeroen, Philippe and Samuel.

“I’ll be around
To pick up your thoughts”, in “I’ll be around”, Yo La Tengo

An essential aspect for a positive experience in pursuing a PhD is to have a friendly working environment and collaborative, engaged, and supportive colleagues. I am fortunate to have had all of these within the CERENA’ research family. I am grateful to everyone, especially Filipa, João Lino, Minhoca, Neves, Eduardo, Helga, Roberto, Claudia, Camila, Lena, Susana, Gustavo, and Prof. Maria João.

“É tão bom uma amizade assim

Ai, faz tão bem saber com quem contar”, in “É tão bom”, Sérgio Godinho

Having friends is worth millions, as Sérgio says. And mine are worth billions. Always there, with a hug at the right moment and the perfect encouragement. To Luís, João Vasco and Rui and all the Magníficos, to Ivan, Julião and Marília, to H, Nocas and Margarida, to Cabé and Sales, to Joana, Rita and Mike, to Pedro, Barros, Tiago, and many others, you guys are awesome!

“It's all about the confidence”,

in “Mr. Motivator”, Idles

To my family, the bedrock of my being, I couldn't express greater gratitude. To my Parents, for giving true meaning to the word. To my brother, my best friend, for so much and for everything. And for making me an uncle. To Vasco, lovely Vasco, and his smiles. To Susana, for all the support and encouragement. To my uncle. To my Rita. To my André. And to everyone else. Thank you for all the love and demonstrations of pride. And to my aunt, always in my thoughts.

“Terra firme só não vale

Porto seguro é não ter medo”,

in “Terra Firme”, Benjamim

A. and S., the most rewarding outcome of this journey. Because of you, my initial self appears unrecognizable to me. Everything is heightened in proportion to your smile, your hunger for life. I don't know how I will be shaped by the weight of time, but I know I have only one enduring goal: to be your Terra Firme and have the ability to provide you Porto Seguro.

“I'm in love with the world

Through the eyes of a girl”,

in “Say Yes”, Elliott Smith

Tanto. Tanto. Tanto. Foi uma imensidão. Um mundo agarrado pelas entranhas do mais genuíno instinto. E nós aqui, onde tu nos trouxeste e nos reconstruíste. E continuarás a nortear, numa ode ao poder feminino. Tudo isto é mais teu que meu. E agradeço por isso. Porque és, e sempre serás, o meu *Verão enquanto caminho para o mar*.

TABLE OF CONTENTS

SUMMARY	v
RESUMO	vii
ACKNOWLEDGMENTS.....	ix
TABLE OF CONTENTS.....	xi
LIST OF FIGURES	xv
LIST OF TABLES	xxiii
LIST OF ACRONYMS	xxv
LIST OF SYMBOLS.....	xxvii
CHAPTER 1 Introduction	1
1.1 Background.....	2
1.2 Research objectives.....	5
1.3 Structure of the thesis.....	6
1.4 Research outcomes	8
CHAPTER 2 Geostatistical inversion of FDEM for near-surface modelling	11
2.1 Introduction.....	12
2.2 Methodology	15
2.2.1 EC and MS model generation	16
2.2.2 Forward response and sensitivity modelling.....	17
2.2.3 Stochastic model optimization	19
2.3 Synthetic case application	22
2.3.1 Data set description	22
2.3.2 Results	24
2.4 Real case application	28
2.4.1 Data set description	28
2.4.2 Results	29
2.5 Discussion	35
2.6 Conclusion	37
CHAPTER 3 Comparison between probabilistic inversion methods of FDEM	39
3.1 Introduction.....	40
3.2 Methodologies	42

3.2.1	Forward response and sensitivity modelling.....	42
3.2.2	The Kalman ensemble generator	43
3.2.3	GEMI inversion	45
3.3	Application.....	46
3.3.1	Data set description	46
3.3.2	Inversion parametrization	48
3.3.3	Results	50
3.4	Discussion	59
3.5	Conclusion	64
CHAPTER 4	Geostatistical Inversion of FDEM with Randomized Tensor Decomposition.....	65
4.1	Introduction.....	66
4.2	Methodology	69
4.2.1	Forward response and sensitivity modelling.....	69
4.2.2	Inverse method	70
4.2.3	Model re-parameterization	72
4.3	Synthetic case application	76
4.4	Real case application	80
4.5	Discussion	85
4.6	Conclusion	86
4.6.1	Conclusion data and materials availability	87
CHAPTER 5	Geostatistical joint inversion of FDEM and DC resistivity data	89
5.1	Introduction.....	90
5.2	Methodology	93
5.2.1	EC and MS model generation	93
5.2.2	FDEM forward model and sensitivity analysis.....	94
5.2.3	ERT forward model.....	95
5.2.4	Comparison and stochastic model optimization	96
5.3	Synthetic case application	99
5.3.1	Data set description	99
5.3.2	Results	100
5.4	Real case application	106
5.4.1	Data set description	106
5.4.2	Results	108
5.5	Discussion	112
5.6	Conclusion	116

CHAPTER 6 Conclusions & Future Perspectives	119
6.1 Conclusions.....	120
6.2 Future Perspectives	121
REFERENCES	123

LIST OF FIGURES

Figure 1.1 Structure of the thesis.	7
Figure 2.1 Schematic representation of the iterative geostatistical FDEM inversion workflow.	15
Figure 2.2 Relative normalized sensitivity analysis of IP and QP at depth for each coil configuration used in the synthetic case application, at position 10 m on the profile shown in Figure 2.3.	19
Figure 2.3 True electrical conductivity (left) and magnetic susceptibility (right) and location of the 4 boreholes providing conditioning data of the iterative geostatistical FDEM inversion. ..	24
Figure 2.4 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration.	24
Figure 2.5 a) Pointwise mean of all the MS models computed in the first iteration; b) pointwise mean of all the MS models computed in the last iteration; c) pointwise variance of all the MS models computed in the first iteration; d) Pointwise variance of all the MS models computed in the last iteration.	25
Figure 2.6 a) Histograms of the true EC and one EC model computed in the last iteration; b) Histogram of the residuals between one EC model computed in the first iteration and one EC model computed in the last iteration; c) Horizontal variogram models for EC; d) Vertical variogram models for EC.	26
Figure 2.7 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for four coil configurations (HCP orientation with 1 m and 2 m offset, PRP orientation with 1.1 m and 2.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical dashed lines indicate the location of the borehole data.	27
Figure 2.8 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for four coil configurations (HCP orientation with 1 m and 2 m offset, PRP orientation with 1.1 m and 2.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical dashed lines indicate the location of the borehole data.	28
Figure 2.9 IP maps (ppm) for PRP coil configuration with 1.1 m offset, using a 2D median filter of 0.3 x 0.3 m and: a) the observed FDEM data; b) the synthetic FDEM data computed from	

the pointwise mean models of EC and MS of the 6th iteration. The yellow lines represent the location of the vertical sections (A-B and C-D) of Figures 2.9 to 2.14. White points represent the locations of the available borehole data (blind well in red). Coordinate system in WGS84/UTM zone 30N, EPSG:32630.....30

Figure 2.10 Vertical sections extracted from: a) and b) pointwise mean model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise mean model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.31

Figure 2.11 Vertical sections extracted from: a) and b) pointwise variance model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise variance model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.32

Figure 2.12 Vertical sections extracted from: a) and b) pointwise mean model calculated from all models of MS generated during the 1st iteration; c) and d) pointwise mean model calculated from all models of MS generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.32

Figure 2.13 Vertical sections extracted from: a) and b) pointwise variance model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise variance model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.33

Figure 2.14 Comparison between observed (red stars) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration. The blue lines represent the minimum and maximum FDEM values predicted at a given iteration. The left column represents the predictions at the end of the first iteration and the right column at the end of the last iteration. Vertical dashed lines represent the location of the borehole data.34

Figure 2.15 Comparison between observed (red stars) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration. The blue lines represent the minimum and maximum FDEM values predicted at a given iteration. The left column represents the predictions at the end of the first iteration and the right column at the end of the last iteration. Vertical dashed lines represent the location of the borehole data.34

Figure 2.16 True and predicted values of EC and MS along the blind test for: a) EC during the 1st iteration; b) EC during the 6th iteration; c) MS during the 1st iteration; d) MS during the 6th iteration.36

Figure 2.17 LIN ECa maps (mS/m) of the HCP coil configuration with 2 m offset using a median filter of 0.3 x 0.3 m from a) the observed data b) the synthetic data predicted in the iteration 6th.36

Figure 3.1 Schematic representation of the noniterative constrained KEG method.....45

Figure 3.2 Simplified schematic representation of the iterative GEMI method.....46

Figure 3.3 Cross sections extracted from the reference models of (a) EC and (b) MS. The black dots represent the nine wells. (c) Location of the nine wells in the EC model map.48

Figure 3.4 Cross sections extracted from the pointwise average model of the prior ensemble for (a) EC and (b) MS.49

Figure 3.5 Cross sections from the variance of the prior ensemble for (a) EC and (b) MS. ...	50
Figure 3.6 Prior synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively), compared with reference IP and QP data (red).	50
Figure 3.7 Synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively) calculated for the KEG, compared with reference IP and QP data (red).	51
Figure 3.8 Synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively) calculated for the GEMI method, compared with reference IP and QP data (red).	51
Figure 3.9 Biplots between the reference IP and QP data and the synthetic IP and QP data for the 2 m HCP coil configuration, calculated from the mean of the model realizations, for each method.	52
Figure 3.10 Comparison between the histograms of the reference EC and a single realization generated by (a) the KEG and (c) the GEMI. Comparison between the histograms of the reference MS and a single realization generated by (b) the KEG and (d) the GEMI.	53
Figure 3.11 Vertical 2D section extracted from a realization of the posterior distribution obtained by the KEG for (a) EC and (b) MS.	53
Figure 3.12 Vertical 2D section extracted from a realization of the posterior distribution obtained by the GEMI for (a) EC and (b) MS.	54
Figure 3.13 Two-dimensional mean model of (a) EC and (b) MS using the KEG method. ...	54
Figure 3.14 Two-dimensional mean model of (a) EC and (b) MS using the GEMI method. ...	55
Figure 3.15 Two-dimensional difference between one model computed from the posterior realizations and the reference model of EC using (a) KEG and (b) GEMI.	55
Figure 3.16 Two-dimensional difference between the mean model computed from the posterior realizations and the reference model of EC using (a) KEG and (b) GEMI.	56
Figure 3.17 Two-dimensional difference between one model computed from the posterior realizations and the reference model of MS using (a) KEG and (b) GEMI.	56
Figure 3.18 Two-dimensional difference between the mean model computed from the posterior realizations and the reference model of MS using (a) KEG and (b) GEMI.	57
Figure 3.19 Comparison between the histograms of the residuals between one model computed from the posterior realizations and the reference model of (a) EC and (b) MS, and the residuals between the mean model computed from the posterior realizations and the reference model of (c) EC and (d) MS, using the KEG and GEMI.	57
Figure 3.20 Biplots between the reference model of EC and (a) one model and (b) the mean model computed from the posterior realizations, from each method.	58
Figure 3.21 Biplots between the reference model of MS and (a) one model and (b) the mean model computed from the posterior realizations, from each method.	58
Figure 3.22 (a) Horizontal and (b) vertical variogram models for EC calculated from the true models, the borehole data and one realization from the posterior ensemble predicted with KEG and with GEMI.	60

Figure 3.23 (a) Horizontal and (b) vertical variogram models for MS calculated from the true models, the borehole data and one realization from the posterior ensemble predicted with KEG and with GEMI.....	61
Figure 3.24 Pointwise variance model of EC obtained from the predicted models with (a) KEG and (b) GEMI.....	62
Figure 3.25 Pointwise variance model of MS obtained from the predicted models with (a) KEG and (b) GEMI.....	62
Figure 4.1 Tucker decomposition of a third-order tensor.....	74
Figure 4.2 Workflow of ES-MDA with RTD.....	76
Figure 4.3 True and prior mean of EC and MS of the synthetic case: (a) true EC model; (b) true MS model; (c) prior mean of EC; and (d) prior mean of MS. It is clear the influence of the borehole data in the prior mean of EC and MS.	77
Figure 4.4 Posterior mean, standard deviation (std.) and residual of EC and MS of the synthetic case: (a) posterior mean of EC models; (b) posterior mean of MS models; (c) posterior std. of EC; (d) posterior std. of MS; (e) absolute error between the posterior mean and the true EC; (f) absolute error between the posterior mean and the true MS.	78
Figure 4.5 Predicted IP data from the prior and posterior EC and MS models of the synthetic case. The red dots are the true measurements with noise; the black lines represent the true data without noise; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.	79
Figure 4.6 Predicted QP data from the prior and posterior EC and MS models of the synthetic case. The red dots are the true measurements with noise; the black lines represent the true data without noise; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.	79
Figure 4.7 Prior mean of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices, and the black dots represent the well locations.....	81
Figure 4.8 Posterior mean of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices.....	82
Figure 4.9 Posterior standard deviation of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices.	83
Figure 4.10 Predicted IP data from the prior and posterior EC and MS models of the real case. The red dots are the true measurements; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.	84
Figure 4.11 Predicted QP data from the prior and posterior EC and MS models of the real case. The red dots are the true measurements; the intervals in gray and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.	84

Figure 4.12 Comparison of the predicted and measured EC (a) and MS (b) at the blind well of the real case. The red dots are the true measurements; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior realizations, respectively; the blue lines represent the posterior mean.	86
Figure 5.1 Schematic representation of the iterative geostatistical joint inversion method using FDEM and ERT data.	94
Figure 5.2 True electrical conductivity (left) and magnetic susceptibility (right) and location of the 9 boreholes providing conditioning data of the iterative geostatistical joint inversion method.	100
Figure 5.3 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration. Vertical dot and blue lines indicate the location of the borehole data.	101
Figure 5.4 a) Pointwise mean of all the MS models computed in the first iteration; b) pointwise mean of all the MS models computed in the last iteration; c) pointwise variance of all the MS models computed in the first iteration; d) Pointwise variance of all the MS models computed in the last iteration. Vertical dot and blue lines indicate the location of the borehole data.	101
Figure 5.5 a) Histograms of the true Eca and the pointwise mean of all the Eca models computed in the first iteration; b) Histograms of the true Eca and the pointwise mean of all the Eca models computed in the last iteration; c) Histograms of the true Eca and the residuals of the pointwise mean of all the Eca models computed in the first iteration; d) Histograms of the true Eca and the residuals of the pointwise mean of all the Eca models computed in the last iteration.	102
Figure 5.6 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.	103
Figure 5.7 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for perpendicular coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.	103
Figure 5.8 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the	

predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.104

Figure 5.9 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for perpendicular coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.105

Figure 5.10 Comparison between observed and predicted apparent resistivity data; a) Observed apparent resistivity data; b) Predicted apparent resistivity data computed in the first iteration; c) Predicted apparent resistivity data computed in the last iteration; d) Histograms of the observed apparent resistivity data, one predicted apparent resistivity data computed in the first iteration and the corresponding residuals between both; e) Histograms of the observed apparent resistivity data, one predicted apparent resistivity data computed in the last iteration and the corresponding residuals between both.106

Figure 5.11 Map of the FDEM survey with QP data (ppm) for the HCP coil configuration with 4 m offset. The black line represents the location of the transect of Figures 5.12 to 5.15. Green points represent the locations of the available borehole data (blind well in red). Coordinates in Belgian Lambert'72 coordinate system.107

Figure 5.12 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration. Vertical magenta lines indicate the location of the borehole data. The vertical red dashed line represents the location of the blind well.109

Figure 5.13 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical grey lines indicate the location of the borehole data. The vertical dashed line represents the location of the blind well.109

Figure 5.14 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical grey lines indicate the location of the borehole data. The vertical dashed line represents the location of the blind well.110

Figure 5.15 Comparison between observed and predicted apparent resistivity data; a) Observed apparent resistivity data; b) Predicted apparent resistivity data computed in the first iteration; c) Predicted apparent resistivity data computed in the last iteration; d) Biplot between the observed apparent resistivity and the predicted apparent resistivity data computed in the first iteration; e) Biplot between the observed apparent resistivity and the predicted apparent resistivity data computed in the last iteration. Vertical grey lines indicate the location of the borehole data. The vertical dashed lines represent the location of the blind well.	111
Figure 5.16 a) Global correlation coefficient of the apparent resistivity data computed in all iterations; b) Global root-mean-square error of the QP data computed in all iterations.	112
Figure 5.17 a) Pointwise mean of all the EC models computed in the first iteration using GEMI; b) pointwise mean of all the EC models computed in the last iteration using GEMI; c) pointwise variance of all the EC models computed in the first iteration using GEMI; d) Pointwise variance of all the EC models computed in the first iteration using GEMI. Vertical magenta lines indicate the location of the borehole data. Vertical dot and blue lines indicate the location of the borehole data.	113
Figure 5.18 Two-dimensional difference between the pointwise mean model computed in the last iteration and the reference model of EC (Figure 5.2), using: a) iterative joint geostatistical inversion method; b) GEMI method. c) Histograms of the true electrical conductivity, the pointwise mean of all the EC models computed in the last iteration and the corresponding residuals between both, using the iterative joint geostatistical inversion method; d) Histograms of the true electrical conductivity, the pointwise mean of all the EC models computed in the last iteration and the corresponding residuals between both, using GEMI method.	114
Figure 5.19 Inversion realizations of Doelpolder data set obtained by GEMI method: a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration. Vertical magenta lines indicate the location of the borehole data. The vertical red dashed line represents the location of the blind well.	115
Figure 5.20 True and predicted values of EC along the blind well test for: a) EC during the first iteration obtained by the iterative joint geostatistical inversion method; b) EC during the last iteration by the iterative joint geostatistical inversion method; c) EC during the last iteration by the GEMI method.	115
Figure 5.21 Biplot along the blind well test between the observed EC and: a) the predicted EC obtained by GEMI method for the first iteration; b) the predicted EC obtained by GEMI method for the last iteration; c) the predicted EC obtained by the iterative geostatistical joint inversion method for the first iteration; d) the predicted EC obtained by the iterative geostatistical joint inversion method for the last iteration.	116

LIST OF TABLES

Table 2.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.....	23
Table 3.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.....	48
Table 4.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.....	77

LIST OF ACRONYMS

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
CC	Correlation coefficient
CP	Canonical polyadic
CPT-C	Conductivity cone penetration test
DC	Direct current
DOI	Depth of investigation
DSS	Direct sequential simulation
EC	Electrical conductivity
ECa	Apparent electrical conductivity
EM	Electromagnetic
EMI	Electromagnetic induction
EnKf	Ensemble Kalman filter
ERT	Electrical resistivity tomography
ES-MDA	Ensemble smoother with multiple data assimilation
FDEM	Frequency-domain electromagnetics
GEMI	Geostatistical frequency-domain electromagnetic inversion
HCP	Horizontal co-planar configuration
HOSVD	Higher-Order Singular Value Decomposition
ICA	Independent component analysis
IP	In-phase
KEG	Kalman ensemble generator
KF	Kalman filter
LIN	Low-induction number approximation
MCMC	Markov chain Monte Carlo
MS	Magnetic susceptibility
NMF	Non-negative matrix factorization
PDF	Probability density function
PRP	Perpendicular coil configuration
QP	Quadrature-phase
RMSE	Root-mean-square error

RTD	Randomized tensor decomposition
SGS	Sequential Gaussian simulation
SVD	Single value decomposition
UXO	Unexploded ordnance

LIST OF SYMBOLS

$\mathbf{m}$	Model of the subsurface
$\mathbf{d}_{\text{obs}}$	Observed data
$\mathbf{F}$	Forward model
h	Height
ϵ	Error
N_s	One model/realization of a property
H	Total magnetic field
H^P	Primary magnetic field
H^S	Secondary magnetic field
H^0	Free-space magnetic field
H^N	Normalized total magnetic field
M	Transmitter moment
J_0	Bessel functions of zeroth order
J_1	Bessel functions of first order
r	Transmitter-receiver offset
u_0	Wave-number of zeroth layer
λ	Hankel transformation
r_{TE}	Reflection coefficient
Y_0	Admittance of the air half-space
$\hat{Y}_1$	Surface admittance
n	Number of layers
$\Delta \mathbf{m}_n$	Perturbation of the physical property at a given layer
N	Number of layers
t_{coils}	Coil configuration
$sens_{EC}$	Normalized sensitivity of EC
$sens_{MS}$	Normalized sensitivity of MS
$\mathbf{S}$	Similarity coefficient
S_{EC}	Similarity weighted in depth by the normalized sensitivity of EC
S_{MS}	Similarity weighted in depth by the normalized sensitivity of MS
R_t	Bulk electrical resistivity
ϕ	Porosity

a	Tortuosity constant
n	Saturation exponent
R_w	Electrical resistivity of the pore fluid
S_w	Water saturation
k	Cementation exponent
x	Coordinate of the receiver coil
z	Coordinate of the receiver coil
$\mathbf{x}$	Observed FDEM data
$\mathbf{y}$	Synthetic FDEM data
ϵ_{err}	Measurement error
n_{ens}	Size of the ensemble
$\mathbf{m}^i$	Gaussian prior PDF
$\mathbf{d}^i$	Gaussian observed data PDF
$\mathbf{C}_D^e$	Covariance matrix of random observation errors
$\mathbf{C}_{mg}^e$	Ensemble covariance matrix between prior parameter vector and forward model response vector
$\mathbf{C}_{gg}^e$	Ensemble covariance matrix of the forward model response vector
G	Forward response ensemble matrix
n_{obs}	Number of observations
n_{par}	Number of model parameters
A	Prior ensemble matrix
G	Forward response ensemble matrix
$\mathbf{m}_j^p$	Prior model parameters
$\mathbf{m}_j^u$	Update model parameters
$\mathbf{d}_j^p$	Predicted FDEM data
$\tilde{\mathbf{d}}_j$	Observed FDEM data with random perturbation
$\mathbf{K}$	Kalman gain matrix
$\mathbf{C}_{md}^p$	Cross-variance matrix between the prior model parameters and the corresponding predicted data
$\mathbf{C}_{dd}^p$	Covariance of the predicted data
$\mathbf{C}_d$	Covariance matrix of the measurement error
$\bar{\mathbf{m}}^p$	Empirical mean of the ensemble of model variables
$\bar{\mathbf{d}}^p$	Empirical mean of the predicted data
$\mathbf{G}$	Forward response in linear case
$\mathbf{C}_m$	Covariance of model parameters

N_e	Ensemble of prior realizations
α_k	Inflation coefficient
$\mathcal{X}$	Tensor from Tucker decomposition
$\times_n$	Tensor-matrix multiplication along mode- n
$\mathcal{Z}$	Tensor
$\mathcal{G}$	Nondiagonal core tensor
$\mathbf{B}^{(n)}$	Factor matrix
$\mathbf{\Omega}^{(n)}$	Random matrix
$\mathcal{G}_{(n)}$	n -unfolding matrix of tensor $\mathcal{G}$
$\mathbf{Q}^{(n)}$	Orthonormal basis
$\mathbf{z}_j^k$	Reduced model parameters
N_m	Number of model parameters
N_d	Number of observations
σ	Electrical conductivity
ϕ_p	Potential field
I	Input current
δ	Dirac delta function
r^+	Positive current electrodes
r^-	Negative current electrodes
$\mathbf{D}$	Divergence matrix
$\mathbf{S}(\sigma)$	Diagonal matrix containing the electrical conductivity values
$\mathbf{G}$	Gradient matrix
$\hat{\phi}$	Vector of electric potentials
$\mathbf{A}(\sigma)$	Combined ERT forward operator
q	Vector containing the current electrode pairs
ρ_{app}	Apparent resistivity
K	Geometric factor

CHAPTER 1

Introduction

1.1 Background

The uppermost tens of meters beneath our feet (i.e., the near surface) are the portion of the Earth that affects and is most impacted by human activities: from construction to agriculture to repository of urban and industrial wastes. Also, yields some of our mineral and energy resources, which are critical to our modern lives. Therefore, it is of the utmost importance to characterize the near surface and to accurately quantify spatially its physical, chemical, and biological properties (e.g., Butler, 2005; Everett, 2013). Throughout the past century, mainly invasive techniques (e.g., boreholes) have been used to investigate and characterize the near surface. Although useful, invasive techniques are generally expensive to acquire and operate, impractical to implement in some locations, and provide only sparse and discrete direct observations with considerable limitations for the spatial characterization of such a dynamic system as the near surface (Pyrz and Deutsch, 2014).

The application of non-invasive geophysical techniques in near-surface characterization has increased considerably during the last decades, as the result of the recent developments on instrumentation equipment and improved computational capabilities. Also, these non-invasive geophysical techniques have been proven to be powerful tools in the spatial characterization of the subsurface properties, since they can acquire indirect, and virtually continuous, measurements of the physical properties of the subsurface (Minsley et al., 2012). These factors led to the use of near-surface geophysical techniques in more complex and heterogenous subsurface environments (e.g., landfill deposits), enabling the measurement of physical soil properties for an entire area of interest and the characterization of the subsurface in a spatially comprehensive way.

Electromagnetic methods (EM) have the broadest application range for near-surface characterization, due to the wide spectrum of instrumental systems and their respective configurations (Reynolds, 2011). EM methods were developed during the 1920s, with the first application to imaging the earth through the induction of eddy currents in the subsurface performed by Karl Sundberg (Sundberg and Hedstroem, 1934).

The following years witnessed an increasing development of induction-based EM equipment, all featured by one or more transmitter and receiver coils. Due to the ability of acquiring large data sets with spatially continuous information in a reasonable time frame, along with the versatility of the acquisition setups, EM surveys have become widely popular (Telford et al., 1990). Based on the nature and distinct features of the EM signal transmitted, EM methods can be classified into two main groups: time-domain (TDEM) characterized by a transient source; and frequency-domain (FDEM) electromagnetic methods characterized by a

continuous source. With the recent development of small-loop FDEM setups, characterized by the small distance between the transmitter and the receiver coil(s) within a single equipment, increasing versatility and efficiency and resulting in faster and more reliable surveys, FDEM methods have become one of the most widely used geophysical methods in near-surface characterization. Application examples cover a wide range of areas such as: determination of soil salinity (e.g., Akramkhanov et al., 2014) and soil compaction (e.g., Al-Gaadi, 2012), groundwater surveys (e.g., Huang et al., 2017; Paepen et al., 2020) and contamination detection (e.g., Sainato et al., 2018), archaeological prospection (e.g., Bongiovanni et al., 2008; De Smedt et al., 2013; Saey, et al., 2016), landfill surveys (e.g., Van De Vijver and Van Meirvenne, 2016; Van De Vijver, 2017) and soil pollution (e.g., Guérin et al., 2004; Blaha et al., 2008), geotechnical characterization (e.g., Saey et al., 2015) and geological 3D imaging (e.g., Monteiro Santos et al., 2011), and unexploded ordnance (UXO) detection (e.g., O'Neill et al., 2005; Saey et al., 2011). All these FDEM applications examples are characterized by shallow depths of investigations ($\sim 4\text{-}20$ m) and the ability to track lateral and/or vertical variations of electromagnetic parameters.

FDEM induction measurements can be linked to subsurface electrical conductivity (EC), magnetic susceptibility (MS) and dielectric permittivity by assuming an infinite homogeneous halfspace below the FDEM sensor (Hanssens et al., 2019). The spatial distributions of these properties can be resolved by the variations of the recorded electromagnetic data. Since small-loop FDEM methods work in low-frequency regimes (far below 10^5 Hz), the influence of dielectric permittivity on the acquired FDEM data can be considered negligible as it is only relevant at high-frequency (Hanssens et al., 2019). As this thesis focuses on methods and applications that take advantage of small-loop FDEM data, the focus was only on the investigation of electrical conductivity and magnetic susceptibility.

Despite these recent developments in FDEM methods for near-surface characterization, most of the potential of these methods has not yet been widely addressed. Particularly, the quantitative prediction of the near-surface properties from recorded FDEM data (i.e., FDEM inversion), and how direct and indirect measurements can be integrated in a consistent way, given their distinct nature.

The standard approach of translating geophysical data into numerical subsurface models is to solve a geophysical inverse problem (Tarantola, 2005). Opposite to solving a forward problem (i.e., the theoretical mathematical calculation of synthetic geophysical data from a known set of model parameters) a geophysical inversion aims at predicting the unknown spatial distribution of the model parameters. Under the scope of this thesis, the model parameter space is defined by the electrical conductivity and magnetic susceptibility, through the changes

imposed by these properties on the observed geophysical data. Geophysical inverse problems are ill-posed and nonlinear inverse problems with no unique or stable solutions, which arises from the fact that small differences in the input parameters can produce large differences in the solutions (Tarantola, 2005). The non-uniqueness of the solution is a result of the multiple parameter configurations that can produce similar data to the one observed, resulting in uncertain predictions.

Geophysical inverse methods can be categorized as either deterministic or probabilistic. Since probabilistic approaches are more computationally expensive and require a larger degree of expertise, the majority of methods currently used to solve a FDEM inverse problem are deterministic approaches (e.g., Zhdanov, 2002). Deterministic inverse methods rely on regularization methods to transform an ill-posed into a well-posed problem. These regularization methods, such as Tikhonov regularization (Tikhonov and Arsenin, 1977), simplify the inverted solution, predicting a single smooth representation of the subsurface and ignoring uncertainties about the predictions.

Opposed to a single solution outcome, probabilistic geophysical inversion methods predict the subsurface model parameters as probability distribution functions. These methods allow obtaining, or approximating, a posterior probability distribution of the model parameter space. In these types of methodologies, the non-uniqueness of the geophysical inversion is captured in the final solution and the uncertainty of the modelling procedure can be assessed. The growth of computational resources in the last decades allowed the development of probabilistic geophysical inversion methods and the spread of their application (Grana et al., 2022). Among these probabilistic methods, iterative geostatistical inversion methods emerge as a powerful tool to quantify physical subsurface properties, as they provide a framework to data integration and detailed description of the subsurface spatial heterogeneities (e.g., Azevedo and Soares, 2017; Tylor-Jones and Azevedo, 2022).

The inversion of electromagnetic data using probabilistic methods for modelling the near-surface physical properties is still underdeveloped and far from being the standard modelling technique (Bobe, 2020). This is mainly due to the relatively large size of the electromagnetic data sets, which makes probabilistic approaches computationally expensive, and to the highly non-linear nature of the forward model and the corresponding computational cost. This thesis proposes alternative approaches to close the existing gap related to the lack of geostatistical FDEM inversion methods.

Along with electromagnetic data, electrical tomography near-surface techniques, such as direct current (DC) resistivity methods, have demonstrated their efficiency to characterize near-surface heterogeneous environments such as ground water contamination, archaeological

mapping, or mineral prospecting. These methods are sensitive to subsurface changes of electrical resistivity (or electrical conductivity) (Reynolds, 2011). These electrical tomography methods can image the subsurface by deriving the spatial distribution of EC from the combination of different measurements configurations. Electromagnetic and electrical data are often acquired jointly but interpreted and modelled separately due to the nonlinearity of the inverse problem. Nevertheless, inverting both geophysical data sets in a joint inversion methodology can improve the accuracy of the predicted results due to the differences in the spatial resolution of both methods and the complementary information about the subsurface retrieved from each data set (Moorkamp, 2017). However, handling the differences in the resolution and nature of both methods is not straightforward and prone to uncertainties. This can be overcome by solving the joint inversion problem in a geostatistical framework. This thesis presents a multi-geophysical inversion approach, combining FDEM data and direct current resistivity data, focusing on the advantages of each geophysical method and their potential to model the spatial heterogeneities of a near-surface environment.

1.2 Research objectives

The main objective of this thesis is the development and implementation of iterative geostatistical FDEM, and joint FDEM and electric resistivity tomography (ERT), inversion methods able to predict the spatial distribution of EC and MS at high spatial resolution. These methods should be able to cope with complex and heterogeneous environments, while simultaneously assessing the uncertainty of the predicted models. In more detail, the objectives of this thesis can be described in the following four sub-objectives:

Objective one of this thesis is the development of a synthetic data set based on direct and laboratory measurements obtained from samples acquired in a complex near-surface deposit corresponding to a mine tailing. This data set is used to validate the inversion methodologies presented in this thesis.

Objective two of this thesis is the development and implementation of an iterative geostatistical frequency-domain electromagnetic inversion methodology that allows to simultaneously predict EC and MS. The proposed method can integrate existing direct measurements, assess uncertainties related to the inverse models and account for a priori geological knowledge about the spatial distribution of the soil properties.

Objective three of this thesis is to tackle the computational cost that iterative FDEM inversion methodologies have when solving large-scale three-dimensional problems in near-surface applications. Machine learning is applied for dimensionality reduction and the inversion is

performed in the reduced space without compromising the exploration of the model parameter space.

Objective four of this thesis is the development and implementation of an iterative geostatistical geophysical joint inversion approach to increase the accuracy of the predicted subsurface models while reducing their uncertainty.

1.3 Structure of the thesis

This thesis follows a structure based on research that is published, or submitted for publication, in international peer-reviewed journals. The thesis is divided into six chapters and its overall structure is outlined in Figure 1.1.

Chapter one underlines the rationale behind this thesis, the research questions that it proposes to answer, the main objectives of this work and describes the structure of the thesis.

Chapter two addresses the first and second objectives of this thesis. It describes the development of a realistic synthetic data set used to validate the geostatistical inversion methodologies developed and implemented under the scope of this thesis. It also describes an iterative geostatistical FDEM inversion (GEMI), validated in the realistic synthetic data set and tested in a real data set application example. The work of this chapter is under review for publication in a peer-reviewed journal.

Chapter three extends objective two with a comparison between the iterative geostatistical inversion method of FDEM data presented in chapter two and a probabilistic inversion method of FDEM data based on the Kalman Ensemble Generator technique. Both FDEM inversion methods are validated and tested on the same synthetic data set, exploring the potential and limitations of each method. The work of this chapter is published in a peer-reviewed journal (Narciso et al., 2022).

Chapter four presents a FDEM inversion methodology that combines an ensemble smoother with multiple data assimilation (ES-MDA) and model re-parameterization via randomized tensor decomposition (RTD), to simultaneously predict electrical conductivity and magnetic susceptibility from measured FDEM data. This chapter focuses on objective three and tries to overcome the computational costs associated with solving large-scale three-dimensional problems in near-surface modelling techniques. The method is applied to synthetic and noisy real data sets. The work of this chapter is published in a peer-reviewed journal (Liu et al., 2023).

Chapter five presents a geostatistical joint inversion methodology of electromagnetic and electrical resistivity data. This multi-geophysical inversion approach was developed and implemented to improve the detection of near-surface heterogeneities and spatial structures of the properties of interest when compared to single geophysical inversion methods. The proposed method aims at objective four. The joint inversion method is validated with a synthetic data set and then tested on a real data set. The work of this chapter is described in a manuscript under review for publication in a peer-reviewed journal.

Chapter six summarizes the main conclusions of this thesis and links the content and conclusions of each chapter to the corresponding objectives. Also, it discusses the limitations of the proposed methods and potential future research perspectives.

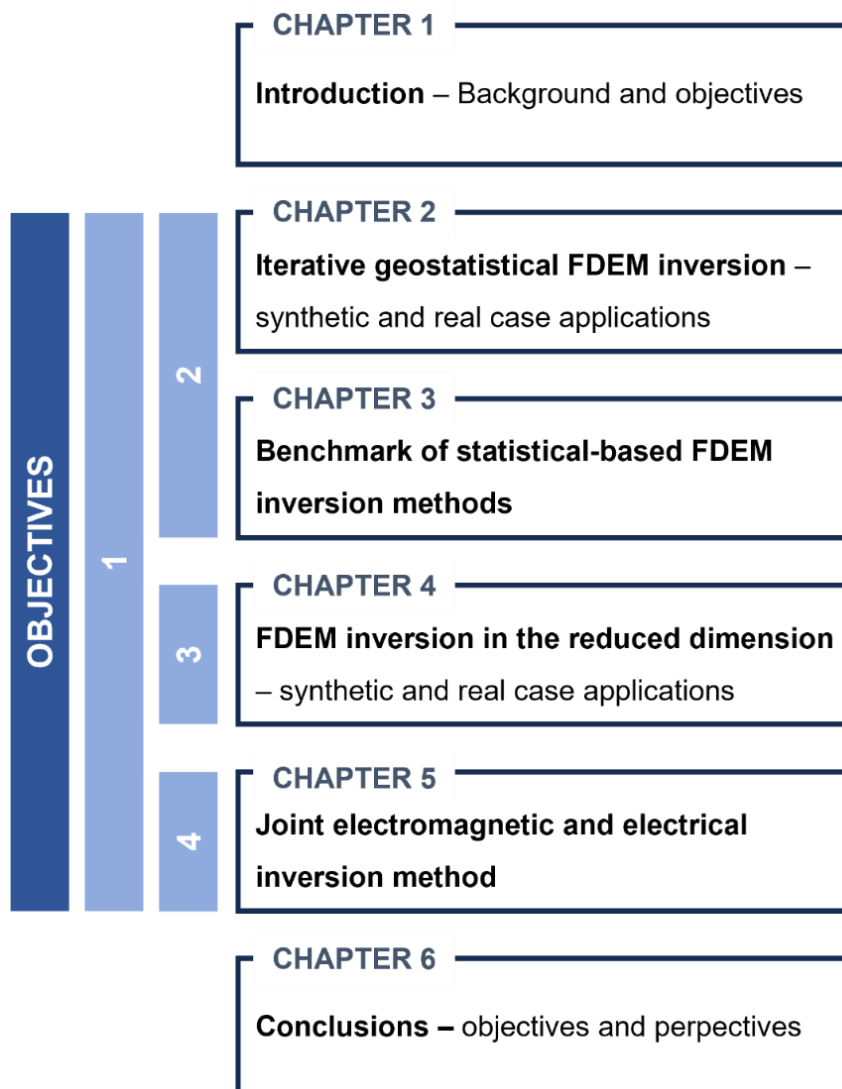


Figure 1.1 Structure of the thesis.

1.4 Research outcomes

In addition to the publications in international peer-reviewed journals on which the chapters of this thesis are based, the research developed under the scope of this thesis was presented at the following international scientific conferences, as poster or oral presentation:

Oral presentations

- Narciso, J., Azevedo, L., and E. Van De Vijver, 2023, Geostatistical joint inversion of FDEM and ERT data: a three-dimensional real case application. Near Surface Geoscience Conference 2023, EAGE, Edinburgh, UK, 3-7 September, doi.org/10.3997/2214-4609.202320177 (Extended Abstract).
- Narciso, J., Azevedo, L., and E. Van De Vijver, 2022, Geostatistical joint inversion of frequency-domain electromagnetic data and direct current resistivity data for modelling near-surface deposits. Near Surface Geoscience Conference 2022, EAGE, Belgrade, Serbia, 18-22 Sep. doi.org/10.3997/2214-4609.202220178 (Extended Abstract).
- Narciso, J., Van De Vijver, E., and L. Azevedo, 2022, Modelling the complexity beneath our feet: A joint inversion FDEM and ERT technique. GeoEnv 2022, Parma, Italy, 22-24 June.
- Narciso, J., Azevedo, L., Van De Vijver, E., and M. Van Meirvenne, 2021, Geostatistical inversion of electromagnetic induction data for modelling waste deposits. GeoEnv 2020, 13th International Conference on Geostatistics for Environmental Applications, Online, 18 June.
- Narciso, J., Liu, M., Van De Vijver, E., Azevedo, L., and D. Grana, 2021, Frequency-domain electromagnetic induction inversion with randomized tensor decomposition. EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-15471, doi.org/10.5194/egusphere-egu21-15471
- Azevedo, L., Narciso, J., and E. Van De Vijver, 2021, Geostatistical FDEM inversion: a three-dimensional real case application. EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-16130, doi.org/10.5194/egusphere-egu21-16130
- Narciso, J., Azevedo, L., Van De Vijver, E., and M. Van Meirvenne, 2021, Geostatistical Electromagnetic Inversion for Landfill Characterization. Symposium for the Application of Geophysics to Engineering and Environmental Problems (SAGEEP) 2021, Online, 14-19 March.

- Narciso, J., Azevedo, L., Van De Vijver, E., and M. Van Meirvenne, 2020, Geostatistical Electromagnetic Inversion for Landfill Characterization. Near Surface Geoscience Conference 2020, EAGE, Online, 7-8 Dec, doi.org/10.3997/2214-4609.202020154 (Extended Abstract). **Best oral communication of NSG2020.**
- Narciso, J., Azevedo, L., Van Meirvenne, M., and E. Van De Vijver, 2020, Geostatistical inversion of electromagnetic induction data for landfill modelling. EGU General Assembly 2020, Online, 04-08 May, EGU2020-20650, 10.5194/egusphere-egu2020-20650

Poster presentations

- Narciso, J., Bobe, C., Azevedo, L., and E. Van De Vijver, 2020, Comparing stochastic FDEM inversion methods for near-surface modelling. Abstract NS014-0011 presented at 2020 Fall Meeting, AGU, Online, 1-17 Dec.
- Narciso, J., Azevedo, L., and E. Van De Vijver, 2020, Geostatistical electromagnetic inversion for landfill modelling and characterization. In: Focaccia S., Kral U., Cormio C., Hengl I. (Eds.). Book of Abstracts: Poster session, Conference on Mining the European Anthroposphere, 20-21 Feb 2020, Bologna, 10.34726/4nat-yb46. **Best poster of the conference.**

CHAPTER 2

Geostatistical inversion of FDEM for near-surface modelling

The research presented in this chapter was submitted to:

Narciso, J.*, Van De Vijver, E., and L. Azevedo, (under review), Geostatistical Inversion of Frequency-Domain Electromagnetic Data for Near Surface Modelling: *Geophysics*.

The detailed characterization of near-surface deposits is important for both environmental and economic reasons. These shallow subsurface systems can be very complex and heterogenous due to natural dynamics and anthropogenic interferences. Modelling techniques based exclusively on direct sampling generate limited informed three-dimensional models of the near-surface. Geophysical methods provide valuable and additional information to model the spatial distribution of the near-surface for locations where direct observations are not available. From this set of methods, frequency-domain electromagnetic induction (FDEM) has been successfully applied to image complex near-surface deposits. Yet, predicting the spatial distribution of relevant subsurface properties from geophysical data, and the integration of direct observations, is not straightforward. It requires solving a challenging geophysical inversion problem. Geostatistical modelling tools have been effectively applied to couple direct observations with geophysical data such as seismic reflection. This chapter presents an iterative geostatistical FDEM inversion method able to integrate data from direct measurements of the near-surface with surface loop-loop FDEM measurements to simultaneously predict high-resolution models of electrical conductivity and magnetic susceptibility, and their associated uncertainty. The iterative geostatistical inversion method is based on stochastic sequential simulation and co-simulation as model perturbation and update techniques. The iterative optimization is based on the local data misfit between observed and simulated FDEM data, weighted by the sensitivity of the acquisition equipment. The proposed method is first demonstrated for a synthetic landfill data set created based on real data collected at a mine tailing disposal site in Portugal, and on a real data set collected at a site with archaeological features in Knowlton, UK. The results show the ability of the proposed method to accurately predict and characterize the spatial distribution of electrical conductivity and magnetic susceptibility down to the depth of interest while reproducing the recorded FDEM data.

2.1 Introduction

Subsurface environments can be highly complex and heterogeneous due to interacting processes of both natural and anthropogenic origins (e.g., Everett, 2013; Lehmann and Stahr, 2007; Morel et al., 2015; Reynolds, 2011). The effective sustainable management and use of the subsurface natural resources, especially in urban environments, including the sustainable remediation and redevelopment of contaminated or otherwise degraded natural or anthropogenic deposits, as well as climate-smart urban land-use planning, relies on the

availability of high-quality subsurface information. Methods for near-surface modelling and characterization based on discrete observations from conventional invasive sampling techniques, such as drilling and core sampling, have limitations to capture the spatial variability of heterogeneous near-surface systems (e.g., Van De Vijver, 2017). The detailed characterization and monitoring of these heterogeneous subsurface environments require novel and efficient data modelling methodologies that allow for the prediction of accurate high-resolution and spatially comprehensive subsurface models, while also providing a quantitative measure for the associated uncertainties. The latter can be critical information when subsurface models are used as decision-making support in land (re)development projects with important environmental and economic impacts.

Near-surface geophysical surveys have emerged as complementary and powerful data sources that, combined with direct observations for calibration and validation of the results, enable predicting the spatial distribution of subsurface physical properties of interest to characterize heterogeneous near-surface systems (e.g., Butler, 2005; Di Maio et al., 2018; Dumont et al., 2017; Narciso et al., 2022; Persico et al., 2018). From the non-invasive geophysical techniques applied in near-surface characterization, electromagnetic surveys have been one of the most widespread and applied techniques (e.g., Delefortrie et al., 2014a; Doolittle and Brevik, 2014; Moghadas et al., 2017; Triantafilis and Monteiro Santos, 2013). Particularly loop-loop frequency-domain electromagnetic induction (FDEM) methods have demonstrated their potential in various application domains, based on their compatibility with a wide variety of subsurface conditions, and capacity to collect high-resolution data in a time-efficient way, as direct ground contact between the instrument and the ground is not required. Besides, FDEM measurements are simultaneously sensitive to two key subsurface properties, electrical conductivity (EC) and magnetic susceptibility (MS). EC relates, amongst others, to soil salinity, texture, organic matter and moisture content, and bulk density. The MS has been proven useful for mapping natural variations in soil mineralogy as well as mapping traces of anthropogenic soil interference (Viscarra Rossel et al., 2011). Knowledge of the spatial variability of these subsurface properties can often be linked to subsurface structures and processes of interest in various application domains, such as agriculture (e.g., Pedrera-Parrilla et al., 2016; von Hebel et al., 2021), landscape archaeology (e.g., De Smedt et al., 2013), and environmental assessment (e.g., Van De Vijver et al., 2015).

Multi-receiver FDEM data, considering both the in- and quadrature-phase (IP and QP, respectively) components of the signal, represent indirect measurement of the subsurface that can be used to predict the subsurface spatial distributions of EC and MS by solving a geophysical inversion problem (i.e., the FDEM inversion problem). Due to insufficient data, the bandlimited nature of the FDEM data, measurement noise and random and systematic errors

in data acquisition, the FDEM inversion problem is ill-posed, nonlinear and has multiple solutions (Tarantola, 2005).

FDEM inversion methods might be categorized into two main groups: deterministic and probabilistic methods. Most examples available in the literature to predict subsurface properties from recorded FDEM data use deterministic methods, which predict a single, smooth, best-fit subsurface model (e.g., Farquharson et al., 2003; Monteiro Santos, 2004; Sasaki et al., 2010; Dafflon et al., 2013; Huang et al., 2016). In addition, the uncertainty assessment of deterministic FDEM inversion methods is limited.

Subsurface models predicted from probabilistic FDEM inversion techniques have demonstrated to be more suitable to model heterogeneous subsurface systems and assess uncertainties about the predictions, thereby providing support for more informed decision-making (e.g., Jadoon et al., 2017; Moghadas and Vrugt, 2019). Besides, the increase in available computational resources led to a growth in the number of publications about this type of FDEM inversion methods in recent years (e.g., Minsley, 2011; Guillemoteau et al., 2016; Bobe et al., 2019; Narciso et al., 2022).

Amongst the probabilistic geophysical inversion methods, iterative geostatistical geophysical inversion methods (e.g., Hansen et al., 2006; Azevedo and Soares, 2017; Grana et al., 2022) have proven their value to predict high resolution rock and/or elastic properties models of the deep subsurface from seismic reflection data. These methods allow integrating different types of data (e.g., borehole data and a priori geological knowledge), provide a detailed description of the spatial distribution of the properties of interest (i.e., subsurface models with higher variability than the observed data), and assess the spatial uncertainty of the predicted model, but their application to near-surface geophysical techniques – other than seismic methods – is still very limited.

The application of this kind of methods to FDEM is particularly relevant considering FDEM inversion results are intrinsically sensitive to three-dimensional (3D) heterogeneity (e.g., Delefortrie et al., 2019; Moghadas et al., 2012) while, due to the computational costs involved in the computation of 3D forward models (i.e., an electromagnetic numerical simulator), the majority of FDEM inversion studies presented in the literature uses one-dimensional (1D) forward models (Guillemoteau et al., 2017), mostly assuming a horizontally layered earth and ignoring lateral variability.

In this chapter is presented an iterative geostatistical FDEM inversion (GEMI) method to simultaneously predict the spatial distribution of EC and MS, by integrating surface geophysical data and direct, in-situ, measurements in the same workflow. Stochastic sequential simulation

and co-simulation are used for model generation and update. The iterative procedure is driven by the misfit between predicted and observed FDEM data and depends on the sensitivity of the acquisition equipment in depth. The results of the proposed FDEM inversion method are a set of high-resolution subsurface EC and MS models that fit equally well the observed FDEM data for all the configurations considered. This ensemble of models can be used for example to assess regions of higher and lower uncertainty.

The proposed inversion method is illustrated for a synthetic and a real data set. The synthetic example shows the reliability of the method in reproducing the true EC and MS spatial distribution and the real case example allows assessing the performance of the method under real noise conditions. The results of the real case application are compared against borehole data that were not considered during the inversion (i.e., blind test).

The next section describes in detail each step of the proposed methodology. Then, are presented the results of its application to a 3D synthetic landfill mining data set created based on real data collected at a mine tailing disposal site in Portugal, and to a 3D real data set from a study area located in Knowlton, UK (Delefortrie et al., 2018), where the subsurface consists of Quaternary deposits overlying Cretaceous deposits. Finally, the results are discussed, and the main conclusions summarized.

2.2 Methodology

The GEMI method simultaneously predicts the spatial distribution of EC and MS from FDEM data (Figure 2.1).

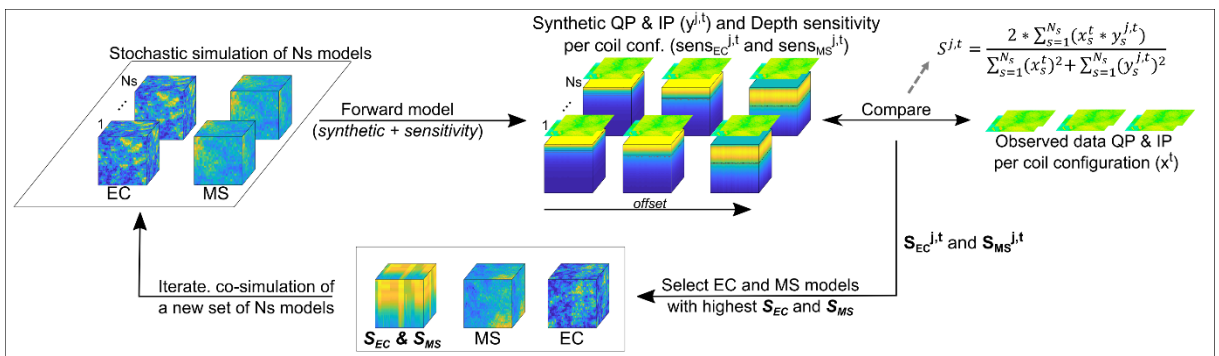


Figure 2.1 Schematic representation of the iterative geostatistical FDEM inversion workflow.

The relationship between the model parameters and the FDEM data can be mathematically summarized by:

$$\mathbf{m} = \mathbf{F}^{-1}(\mathbf{d}_{\text{obs}}) + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\mathbf{m}$ are the model parameters (i.e., EC and MS), $\mathbf{d}_{\text{obs}}$ are the observed FDEM data, $\mathbf{F}$ represents the nonlinear forward model that maps the model ($\mathbf{m}$) into the data ($\mathbf{d}_{\text{obs}}$) domain, and ϵ represents measurement errors and approximations to the physics governing the system under investigation (Tarantola, 2005).

The FDEM inversion problem is solved with an iterative geostatistical inversion technique that can be summarized in three main stages: i) perturbation of the model parameter space with stochastic sequential simulation and co-simulation (Deutsch and Journel, 1998); ii) computation of the synthetic FDEM responses and the sensitivity of these responses towards changes in EC and/or MS at a certain depth layer n , using the 1D FDEM forward model proposed by Hanssens et al. (2019) and assuming a layered half-space model; and iii) stochastic optimization with stochastic sequential co-simulation (Deutsch and Journel, 1998), driven simultaneously by the misfit between true and predicted FDEM responses taking into account the predicted depth of investigation (DOI) as retrieved from the sensitivity provided by the forward model and the different coil configurations that might exist in the acquisition system (as defined in ii)). These steps are described in detail below.

2.2.1 EC and MS model generation

The GEMI methodology starts with the generation of a set of N_s models of EC and MS with stochastic sequential simulation (Soares, 2001). Each model is simulated – in the first iteration – or co-simulated – in the subsequent iterations – for the entire inversion grid at once resulting in 3D or two-dimensional (2D) models depending on the recorded FDEM data. Available direct measurements of EC and MS from in-situ measurements, or borehole data, are used as conditioning experimental data in the geostatistical simulation. The spatial continuity pattern of the simulated models is defined by a variogram model, in 2D or 3D, fitted to experimental variograms computed from the available direct measurements (or borrowed from analogue areas or expert knowledge). In the iterative geostatistical FDEM inversion method proposed here, we use direct sequential simulation (DSS, Soares, 2001) and co-simulation with joint probability distributions (Horta and Soares, 2010) as model perturbation technique of EC and MS. Unlike sequential Gaussian simulation (SGS) (Deutsch and Journel, 1998), these stochastic sequential simulation techniques do not impose any condition on the data distribution (i.e., Gaussian) of the properties to be simulated, thereby avoiding the intermediate step of a data transformation of the distribution of the properties to be simulated. Instead, the marginal and joint distributions as inferred from the experimental data are used in the simulation and co-simulation procedures. For complex and highly nonlinear relationships between primary geophysical and secondary petrophysical properties related to electromagnetic induction measurements, the use of non-Gaussian stochastic sequential co-

simulation techniques allow a better reproduction of the relationship between variables as retrieved from the available direct measurements.

As the iterative inversion rely on stochastic sequential simulation and co-simulation as model perturbation and update technique, all the subsurface models generated during the iterative procedure reproduce the exact values of the borehole data at their locations, the global marginal and joint distributions of each property, and the imposed spatial continuity patterns expressed by variogram models.

2.2.2 Forward response and sensitivity modelling

The forward model (Eq. 2.1) is necessary to compute the theoretical FDEM instrument response, calculated from the differences between the primary electromagnetic field generated from the transmitter coil and the secondary electromagnetic field generated from conducting material, of a loop-loop system that is characterized by one transmitter coil and one or multiple receiver coils (Hanssens et al., 2019). The forward model can be formulated in 1D, 2D or 3D (Auken and Christiansen, 2004, Cox and Zhdanov, 2008, Farquharson et al., 2003). Here, we address the 1D vertical variations of both EC and MS, yet forward models can also address only EC, or even EC, MS and dielectric permittivity simultaneously.

The GEMI method uses a forward model that calculates the theoretical 1D normalized electromagnetic (EM) response according to Maxwell's equations and expressed in in-phase (IP) and quadrature-phase (QP) components, for a horizontal n -layered half-space model (Hanssens et al., 2019). Since it is a low-frequency application, the IP and QP components are mainly influenced by EC and MS, in a quasi-static approximation, neglecting the dielectric permittivity. We calculate N_s synthetic IP and QP responses per coil configuration used in the acquisition of the field data, for each pair of EC and MS models generated in the previous iteration, with N_s representing one model of each property.

This forward model considers a FDEM system positioned at the surface, or at a certain height above the surface (h), of an n -layered subsurface model. It uses Hankel functions, numerically calculated by means of a Guptasarma and Singh digital filter (Guptasarma and Singh, 1997), to determine a superposition of Bessel functions of the zeroth and/or first order that model the EM responses. The total magnetic field H (A/m; primary field, H^P , plus secondary field, H^S) is calculated for the two types of coil configurations used in the synthetic and real case applications, the horizontal co-planar (HCP) and the perpendicular (PRP) coil configurations, for a Z-directed magnetic dipole source located at $(0,0,-h)$:

$$H_{ZZ} = \frac{M}{4\pi} \int_0^\infty [e^{-u_0(z+h)} - r_{TE} e^{u_0(z-h)}] \lambda^2 J_0(\lambda r) d\lambda, \quad (2.2)$$

$$H_{ZX} = -\frac{M}{4\pi r} \int_0^\infty [e^{-u_0(z+h)} - r_{TE} e^{u_0(z-h)}] \lambda^2 J_1(\lambda r) d\lambda, \quad (2.3)$$

where M is the transmitter moment (A/m^2), x and z are the coordinates (m) of the receiver coil, h is the height (m) of the transmitter coil, r is the transmitter-receiver offset (m), J_0 and J_1 are, respectively, the Bessel functions of zeroth and first order, u_0 is the wave-number of zeroth layer, λ the Hankel transformation, and r_{TE} the reflection coefficient. In the methodology proposed herein the reflection coefficient approach is used, rather than the propagation matrix approach, because it is more computationally efficient. The reflection coefficient is defined by:

$$r_{TE} = \frac{Y_0 - \hat{Y}_1}{Y_0 + \hat{Y}_1}, \quad (2.4)$$

where Y_0 is the intrinsic admittance of the air half-space, and $\hat{Y}_1$ is the surface admittance (at $z = 0$), which can be determined recursively by starting at the basement half-space ($n = N$) toward the surface ($n = 1$), in a horizontally n -layered half-space model.

The free-space magnetic fields H^0 (A/m) used in normalization are given for a Z-directed magnetic dipole source located at $(0,0,-h)$:

$$H_z^0 = \frac{M}{4\pi} \int_0^\infty [e^{-u_0(z+h)}] \lambda^2 J_0(\lambda r) d\lambda. \quad (2.5)$$

The normalized total magnetic field H^N (in parts-per-million, ppm) is then given by:

$$H^N = \frac{H - H^P}{H^0} \cdot 10^6 = \frac{H^S}{H^0} \cdot 10^6. \quad (2.6)$$

The recent FDEM instruments use a phase-sensitive measurement between primary and secondary field, i.e., an IP (or real) and QP (or imaginary) measurement (in ppm):

$$IP = \text{Re}(H^N), \quad (2.7)$$

$$QP = \text{Im}(H^N). \quad (2.8)$$

Consequently, FDEM data $\mathbf{d}_{obs}$ (Eq. 2.1) are expressed in parts-per-million (ppm) of the total magnetic field H (A/m), related to the magnetic field of the zeroth layer (free space), H^0 (A/m).

This implementation provides a sensitivity modelling, which represents how sensitive the forward model is toward changes of a physical property $\mathbf{m}$ (i.e., EC and MS) at a specific layer n of the layered half-space. The sensitivity modelling thus calculates the vertical sensitivity distribution related to each physical property within the considered layered model, through a brute-force method, or perturbation method (McGillivray and Oldenburg, 1990), that uses a finite-difference formula based on the Taylor series for $H^N(\mathbf{m})$ and $H^N(\mathbf{m} + \Delta\mathbf{m}_n)$ of the

magnetic fields, where $\mathbf{m}$ is EC or MS, and $\Delta\mathbf{m}_n$ is the perturbation of the physical property (Figure 2.2). The forward model must be resolved N times, with N representing the number of layers, each time with a perturbation $\Delta\mathbf{m}_n$ of the governing n th layer model parameter $\mathbf{m}_n$, with $\mathbf{m}$ representing one of the physical properties. The sensitivity modelling (derived from Eqs. 2.2 and 2.3) can be used for estimating the depth of investigation (DOI).

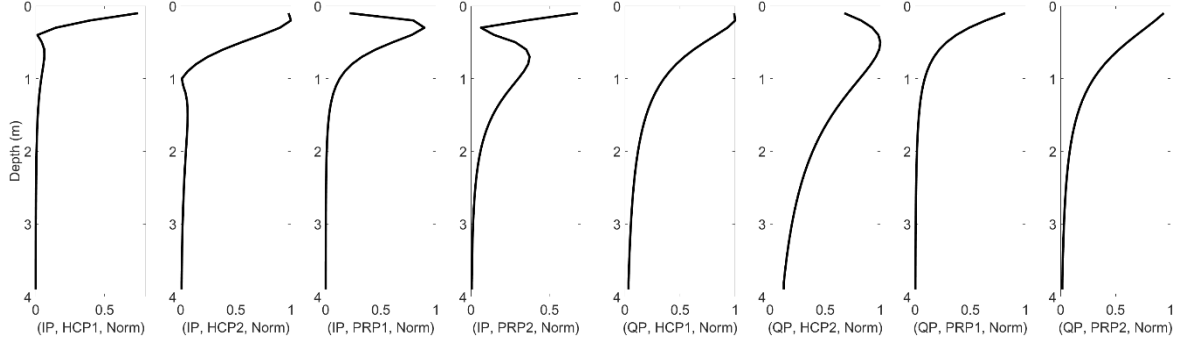


Figure 2.2 Relative normalized sensitivity analysis of IP and QP at depth for each coil configuration used in the synthetic case application, at position 10 m on the profile shown in Figure 2.3.

2.2.3 Stochastic model optimization

The model optimization is achieved by the maximization of the similarity coefficient (S) computed between predicted and observed FDEM data for all the coil configurations (t_{coils}). Each coil configuration corresponds to an offset (i.e., the distance between transmitter(s) and receivers) and an orientation (i.e., HCP or PRP). S is calculated per coil configuration and for each geostatistical realization (N_s) of EC and MS generated at a given iteration, following:

$$S^{j,t} = \frac{2 * \sum_{s=1}^N (x_s^t * y_s^{j,t})}{\sum_{s=1}^N (x_s^t)^2 + \sum_{s=1}^N (y_s^{j,t})^2}, \quad j = 1, \dots, N_s \quad \text{and} \quad t = 1, \dots, t_{\text{coils}}, \quad (2.9)$$

where x and y are the observed and synthetic QP (or IP) data with N samples, respectively. By construction S is bounded between -1 and 1, but negative values are truncated at zero so it can be used as secondary variable for the stochastic sequential co-simulation of a new set of N_s models in the subsequent iteration, constraining the realizations of that iteration. S is sensitive simultaneously to the shape and magnitude of the recorded IP and QP signal. We opted for this metric to avoid an objective function with two terms (i.e., both components of the data), dependent on user-defined parameters to weigh each term of the objective function. S is not computed for the entire series of the FDEM data at once, but for a set of non-overlapping 2D horizontal windows that visit the entire data series. These windows are randomly created, with different sizes, at the beginning of each iteration. Using multiple windows for the entire data grid (i.e., the FDEM measurement locations, with t_{coils} configurations per FDEM

measurement location), with each window having a smaller number of samples than the entire data set, this approach allows reaching higher S values at early stages of the iterative procedure. This procedure is similar to the one proposed in Azevedo and Soares (2017) for geostatistical seismic inversion.

Each S computed for each grid location is then weighted in depth by the normalized sensitivity curves of each coil configuration (Figure 2.2) obtained for the corresponding EC and MS resulting from the FDEM forward model. In the application examples shown below, assumption have been made that EC is directly dependent on QP and MS on IP (alternative assumption might be considered):

$$S_{EC}^{j,t} = sens_{EC}(z)^{j,t} * S^{j,t}, \quad j = 1, \dots, N_s \text{ and } t = 1, \dots, t_{coils}, \quad (2.10)$$

$$S_{MS}^{j,t} = sens_{MS}(z)^{j,t} * S^{j,t}, \quad j = 1, \dots, N_s \text{ and } t = 1, \dots, t_{coils}, \quad (2.11)$$

where $sens_{EC}$ and $sens_{MS}$ are the normalized sensitivities, in depth, at each FDEM data location within the inversion grid. This approach allows to weigh the assimilation of each coil configuration to the predicted EC and MS models in the subsequent iteration. The similarity obtained for a given coil configuration/offset influences in depth the stochastic update of EC and MS.

At the end of each iteration, auxiliary volumes of EC and MS are built with the parts of the N_s simulated EC and MS models (subvolumes of the inversion grid) that generate predicted QP and IP data with the highest S for a given location, considering simultaneously all the t_{coils} (i.e., coil configurations). The volumes of EC and MS are stored as auxiliary volumes along with the corresponding S .

In the subsequent iteration, these auxiliary variables (i.e., the selected S and corresponding EC and MS volumes) are used to co-simulate a new set of EC and MS models. For locations associated with $S \sim 1$ the new ensemble of co-simulated models of EC and MS will be similar to the auxiliary volumes. On the other hand, locations with $S < 0.5$ will exhibit larger variability within the ensemble of new models. This model update approach ensures the convergence of the geostatistical FDEM data inversion along the inversion procedure, with the minimization of the misfit between observed and predicted FDEM data.

The proposed iterative geostatistical inversion methodology for FDEM data may be summarized in the following sequence of steps (Figure 1):

- i) Generation of two ensembles of N_s models of EC and MS given borehole data and variogram models retrieved from these borehole data, with stochastic sequential

simulation (Soares, 2001) – in the first iteration – and co-simulation – in the next iterations – with joint probability distributions (Horta and Soares, 2010), respectively;

- ii) Calculation of the N_s synthetic FDEM data for each pair of models simulated in i) using a FDEM forward model. In the application examples shown below we use the 1D FDEM forward model proposed by Hanssens et al. (2019);
- iii) Compute S between true and predicted FDEM data per coil configuration, within each of the randomly created window at the beginning of the iteration;
- iv) Weight the similarity coefficient in depth by the normalized sensitivity analysis of the FDEM data for all coil configurations (Eqs. 2.10 and 2.11);
- v) Build three auxiliary volumes by selecting the EC and MS local value that ensure the highest S at a given location from all the N_s models of each iteration. Store the corresponding EC, MS and S values;
- vi) Generate a new ensemble of EC and MS models using co-DSS and the auxiliary volumes resulting from v) as secondary variables;
- vii) Iterate and repeat steps ii)-vii), until the global convergence of the method reaches a pre-defined threshold of global S computed between all offsets for IP and QP simultaneously.

All models of EC and MS generated during the iterative geostatistical FDEM inversion are conditioned locally by existing borehole data for EC and MS. They reproduce the global marginal and joint distributions between EC and MC as inferred from the borehole data and a pre-defined spatial continuity pattern as imposed by a variogram model.

The GEMI method is flexible and can be parameterized for all possible coil configurations (i.e., transmitter-receiver orientations and distances between the transmitter and the receiver) that are included in the FDEM survey data set and alternative forward models as long they provide a measure of sensitivity in depth.

2.3 Synthetic case application

2.3.1 Data set description

The synthetic data set used in this section comprises a three-dimensional grid of 30 by 40 by 4 meters (i.e., length, width, depth dimension, respectively) and a cell size of 0.1 m by 0.1 m by 0.1 m in each dimension, respectively. The data set was created based on real data collected at a mine tailing disposal site in Portugal (Panasqueira), in which the main mining production is copper and wolfram (Narciso et al., 2020). Laboratory measurements of porosity and particle density on two different geological materials, fine-shaly sands and quartz-schist gravels, were available. This information was used as conditioning data to generate realistic 3D numerical models of porosity, particle density and water content using geostatistical simulation and co-simulation (Deutsch and Journel, 1998). The variogram model for porosity was fitted to experimental variograms computed from the available samples. A spherical model was fitted, in the horizontal direction isotropy was assumed and a range of 8 m was used, in the vertical direction a range of 2 m was used. We assumed a nugget effect of 5% of the total variance of the available samples.

Particle density and water saturation models were generated with stochastic sequential co-simulation conditioned to the porosity model. Particle density was modelled with an omnidirectional spherical variogram model with a horizontal range of 8 m, a vertical range of 2 m and a nugget effect of 5% of the total variance of the available data. The spatial continuity pattern of water saturation was modelled with an omnidirectional spherical variogram model with a horizontal range of 12 m, a vertical range of 4 m and a nugget effect of 0% of the total variance of the collected samples. The sequential approach intends to ensure geological plausibility between properties.

EC was derived from the previously simulated models by applying Archie's law, relating the bulk electrical resistivity (R_t) of a porous medium to the porosity and the water content (Archie, 1942):

$$R_t = a \phi^{-k} R_w S_w^{-n}, \quad (2.12)$$

where a is the tortuosity constant, assumed as 0.88, ϕ is the porosity, R_w the electrical resistivity of the pore fluid, assumed as 0.25 ($\Omega \cdot m$), S_w is the water saturation, k is the cementation exponent that varies between 1.3 and 2.5 for most sedimentary rocks and was assumed as 1.37, and n is the saturation exponent, which depends on the type of the pore fluid and was set to 2. From the lithology and range of porosity values present in the synthetic

model, the values of a , k and n were defined from Keller (1987). The EC was then calculated based on Archie's second law (Archie, 1942):

$$EC = 1/R_t. \quad (2.13)$$

MS was simulated independently as it does not depend on EC. We used unconditional stochastic sequential simulation algorithms (Deutsch and Journel, 1998) and imposed an omnidirectional spherical variogram model with a horizontal range of 12 m, a vertical range of 4 m and a nugget effect of 0% of the total variance of the experimental data. These variogram ranges agree with common ranges for unconsolidated sediments (Hudson et al., 1999).

After the generation of the true 3D model for EC and MS, four boreholes were extracted equally spaced along one 2D transect. The data from the four boreholes are considered as experimental data to condition the iterative geostatistical inversion. In this way we include uncertainty in the spatial continuity model used within the inversion. Table 2.1 summarizes the main parameters of the variogram models imposed for the geostatistical FDEM inversion. The resulting variogram models have a higher nugget effect than the true ones as these were fitted to experimental variograms calculated from the four-borehole data. The nugget effect represents the lack of knowledge about EC and MS at the small-scale.

Table 2.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.

Omnidirectional spherical variogram	EC models	MS models
Horizontal range	8 m	10 m
Vertical range	4 m	4 m
Nugget effect	5 %	5 %

To generate the synthetic observed geophysical data, the coil configurations of a multi-receiver FDEM sensor was mimicked, namely a DUALEM-21S (DUALEM Inc., Milton, Canada). This equipment includes pairs of two different transmitter-receiver (loop-loop) orientations, the horizontal coplanar (HCP) and the perpendicular (PRP) configurations, and two offsets per coil orientation, 1 and 2 meters for HCP, and 1.1 and 2.1 meters for PRP. The true geophysical data was computed by applying the forward model described in the methodology section, which was also applied in the inversion (Hanssens et al., 2019). This approach assumes there is no uncertainty in the forward modelling, which is a hard assumption in complex and highly variable near-surface environments and neglects three-dimensional effects of the propagated field.

We ran the proposed iterative geostatistical FDEM inversion technique with six iterations and thirty-two realizations (i.e., $N_s = 32$) of EC and MS models per iteration spatially constrained to borehole data (Figure 2.3) and the imposed variogram models (Table 2.1).

2.3.2 Results

The results obtained with the proposed iterative geostatistical FDEM inversion method are illustrated for the 2D transect of EC and MS that intersects the four boreholes (Figure 2.3).

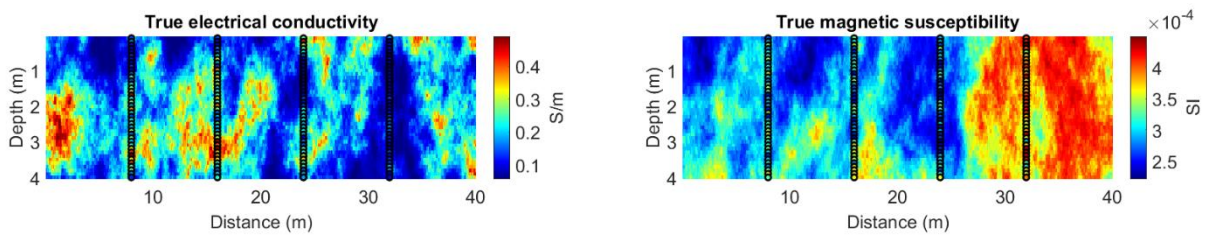


Figure 2.3 True electrical conductivity (left) and magnetic susceptibility (right) and location of the 4 boreholes providing conditioning data of the iterative geostatistical FDEM inversion.

The convergence of the iterative geostatistical FDEM inversion method with respect to parameter model reproduction is assessed by calculating the pointwise mean models computed from all the realizations generated at each iteration. The pointwise mean model is equivalent to the maximum a posteriori model from a Bayesian inversion. The predicted and the true EC and MS models show similar large-scale spatial patterns, but present small-scale differences, which are mainly located in the deeper part of the models where the sensitivity of the forward model is smaller (Figure 2.4b and 2.5b).

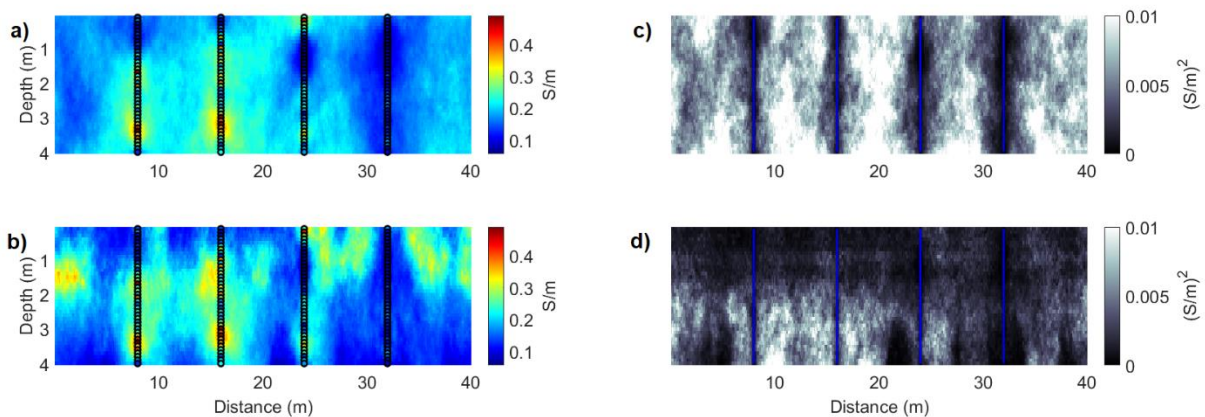


Figure 2.4 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the last iteration.

The iterative geostatistical FDEM inversion method reproduces the true model of EC and is sensitive to local transitions between high and low values of EC (Figures 2.3 and 2.4b). As showed in the pointwise average of an ensemble of realizations, the reproduction of the true small-scale heterogeneities cannot be evaluated based on the pointwise average of the ensemble of realizations, as they are – at least partly – cancelled out by averaging. Nevertheless, the global values and the areas of extreme values are properly matched (Figure 2.4b and 2.4d).

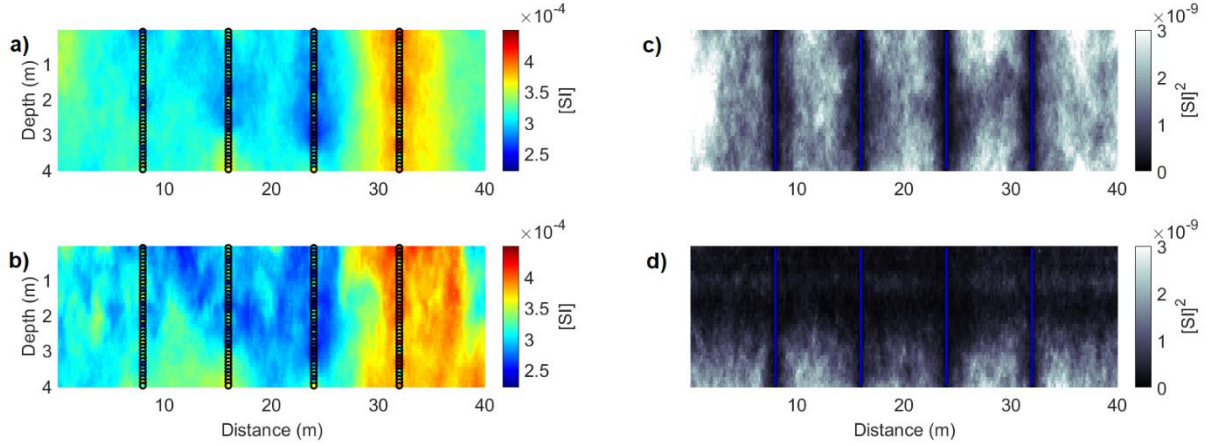


Figure 2.5 a) Pointwise mean of all the MS models computed in the first iteration; b) pointwise mean of all the MS models computed in the last iteration; c) pointwise variance of all the MS models computed in the first iteration; d) Pointwise variance of all the MS models computed in the last iteration.

The small-scale differences and the relationship with depth and sensitivity loss of the predicted solutions are also observed by calculating the pointwise variance models from the ensemble of EC and MS models generated at each iteration (Figures 2.4c, 2.4d, 2.5c and 2.5d). As expected, in the first iteration the spatial distribution of the variance is only dependent on the distance to the locations of the borehole data as the observed FDEM is not assimilated yet. The pointwise variance models of EC and MS computed from models predicted during the last iteration of the geostatistical inversion, shows the influence of the sensitivity provided by the forward model, as the higher variance values are mainly located in the deeper part of the model (Figure 2.4d and 2.5d). The spatial distribution of the pointwise EC variance model presents a reduction in in-depth sensitivity, directly dependent on the coil configurations used, although it shows robustness and accuracy in detecting the lowest local values of EC (Figure 2.4d). The predictions about MS are less sensitive at depth, with the spatial pattern of the pointwise variance model presenting a higher dependence on the coil configurations used (Figure 2.5d).

The proposed iterative inversion technique is able to reproduce the true models (e.g., EC, Figure 2.6a) and converge to the true solution, with the residuals between the true EC model

and one EC realization reducing significantly from the first to the last iteration (Figure 2.6b). The predicted models reproduce the spatial continuity patterns of the true models (Figure 2.6c and 2.6d). As expected, the sill of the predicted models is slightly smaller than in the variograms of the EC true models as result of the increased local variability of the ensemble coming from the stochastic simulation.

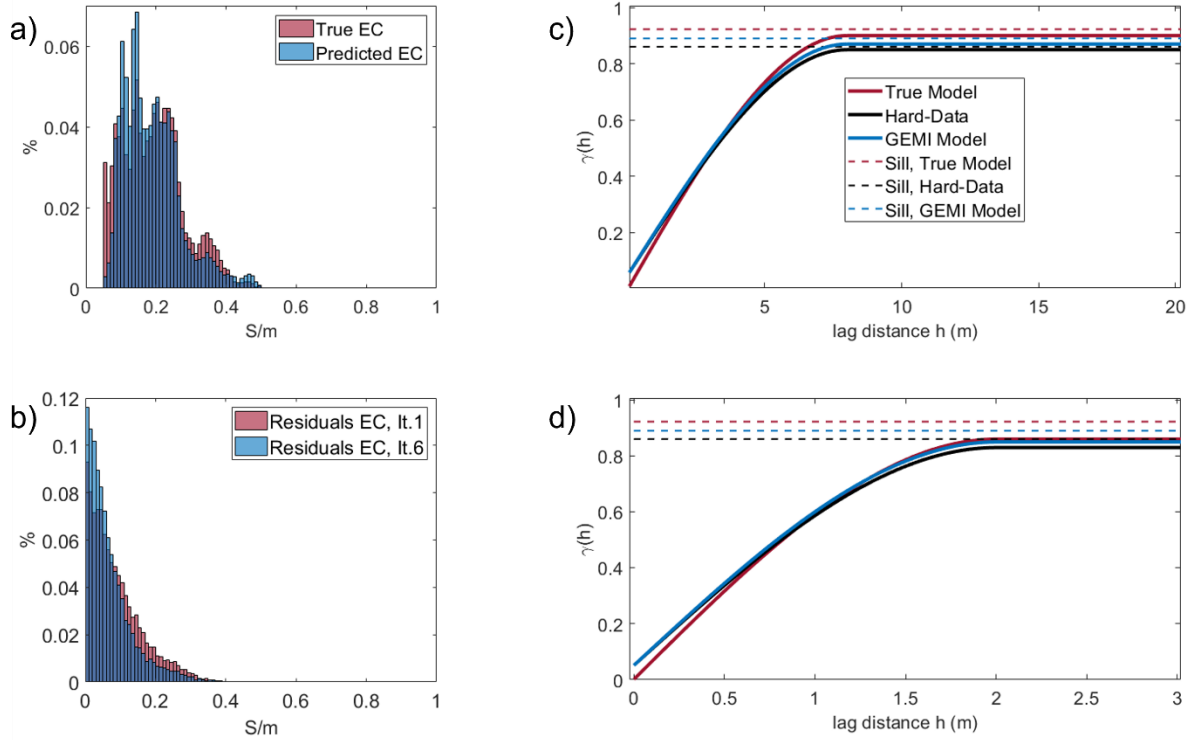


Figure 2.6 a) Histograms of the true EC and one EC model computed in the last iteration; b) Histogram of the residuals between one EC model computed in the first iteration and one EC model computed in the last iteration; c) Horizontal variogram models for EC; d) Vertical variogram models for EC.

The misfit between observed and predicted IP and QP data can be assessed (Figures 2.7 and 2.8). The figures show, for all coil configurations considered, the match between observed and predicted IP and QP responses increases from the first to the last iteration. The uncertainty envelope, as represented by the synthetic response of the ensemble of models in each iteration, narrows and encloses the observed IP and QP data as the iterative procedure advances. Although the uncertainty envelope of all coil configurations in the last iteration well encloses the true FDEM data, a better match is reached in QP responses and in smaller coil distances. This is due to a more stable signal in QP responses and a higher sensitivity to small-scale heterogeneities at shallow depths when the coils are closest to each other, although also losing sensitivity at greater depths, as shown in Figures 2.7 and 2.8. Also, due to an increasing and partly overlapping DOI of the different coil configurations, the shallowest depths are covered by all the FDEM measurement signals, while the largest depths are only covered by

one measurement signal. As expected, the predicted QP and IP responses at the borehole locations are exactly reproduced as the predicted EC and MS models are locally conditioned by the borehole data.

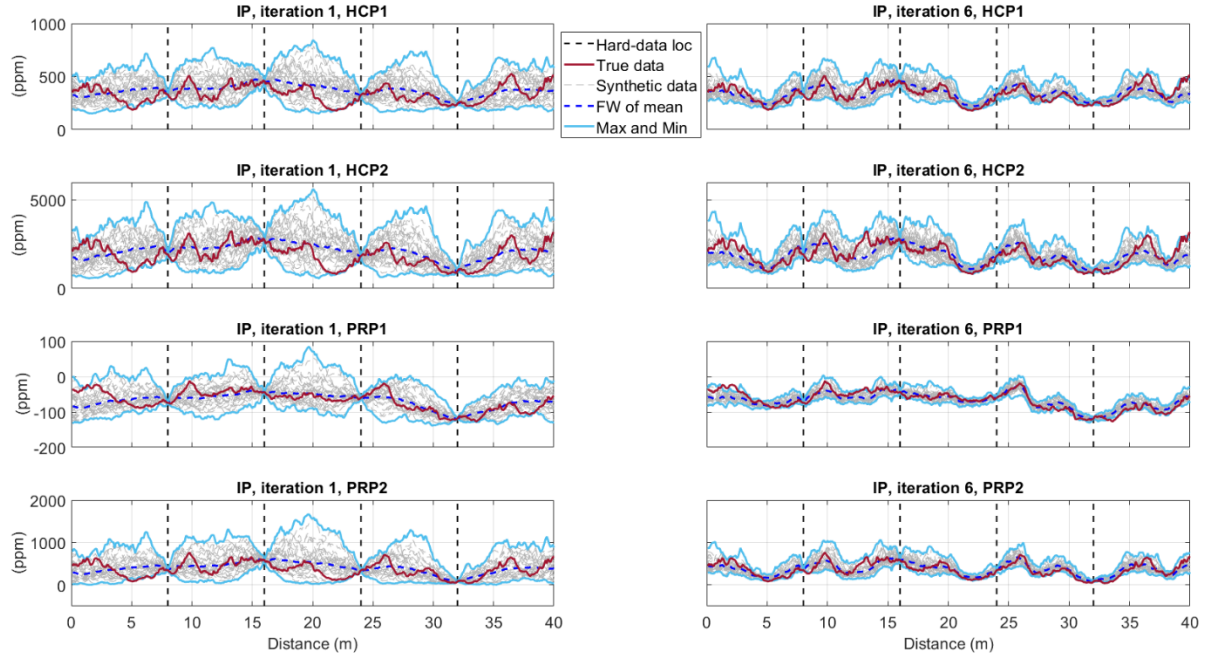


Figure 2.7 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for four coil configurations (HCP orientation with 1 m and 2 m offset, PRP orientation with 1.1 m and 2.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical dashed lines indicate the location of the borehole data.

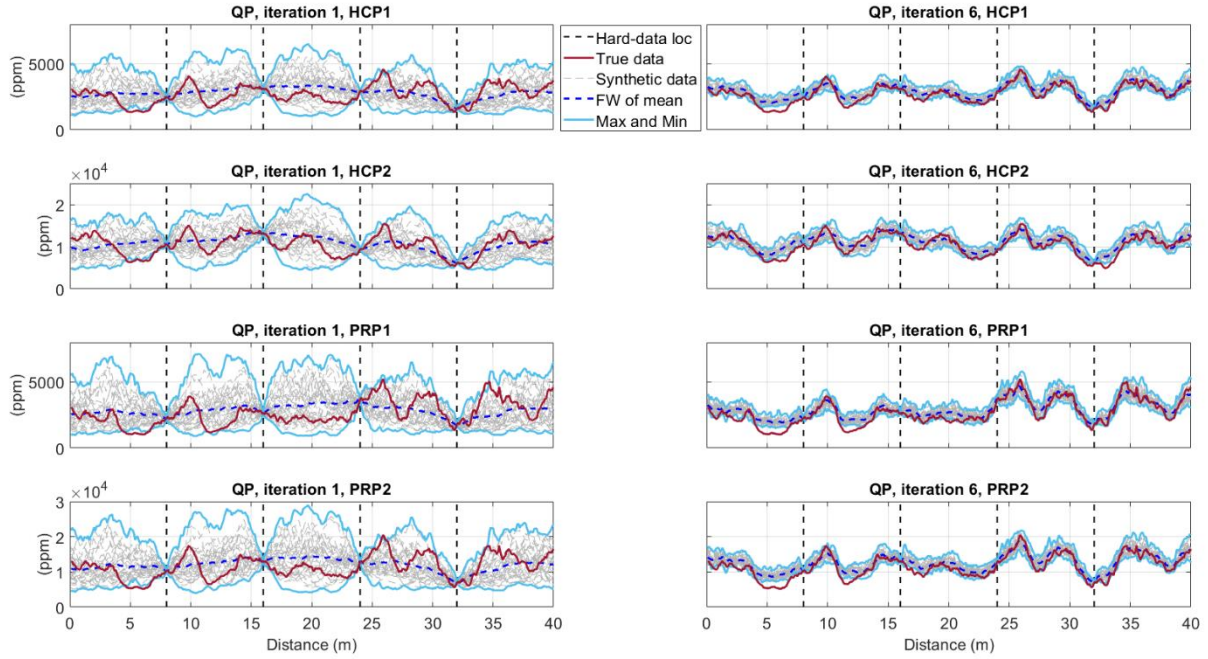


Figure 2.8 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for four coil configurations (HCP orientation with 1 m and 2 m offset, PRP orientation with 1.1 m and 2.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical dashed lines indicate the location of the borehole data.

2.4 Real case application

2.4.1 Data set description

The GEMI method was applied to a real data set obtained from an FDEM survey located at a site near Knowlton (Dorset, UK). The site is an arable land with a 20 cm thick rendzina soil cover, gentle slope, and geologically characterized by Cretaceous bedrock chalk overlain by Quaternary siliciclastic sand deposits. The bedrock chalk exhibits an overall background susceptibility of zero, while the topsoil is strongly magnetic ($MS \approx 1 \times 10^{-3}$). The subsurface has a low EC (~ 7 m/Sm), with the topsoil being slightly more conductive. The subsurface is also known to contain several archaeological features from the Stone Age that produce strong local IP anomalies. The FDEM data was collected during 2016 using a DUALEM-21HS instrument, with an operating frequency of 9000 Hz and pairing one transmitter with three horizontal coplanar receiver coils, at 0.5, 1 and 2 m spacing (HCPH, HCP1 and HCP2, respectively), and

three perpendicularly oriented receivers at 0.6, 1.1 and 2.1 m spacing (PRPH, PRP1 and PRP2, respectively). Yet, due to poor signal-to-noise ratio, the smallest offsets (0.5 and 0.6 m) were not used in this study. The FDEM survey was performed using a motorized survey configuration with the instrument elevated at 0.16 m above the surface, along parallel lines 1 m apart at a speed of ~8 km/h, and a sampling frequency of 8 Hz. A detailed description of the data set can be found in Delefortrie et al. (2018).

The pre-processing of the FDEM data included: i) the correction for the spatial offsets between the position and sensor data, following the procedure described in Delefortrie et al. (2016); ii) the correction for signal drift – a relative calibration, following the procedure in Delefortrie et al. (2014b); and iii) an absolute calibration per coil configuration to eliminate the presence of signal offsets, comparing the forward modelled responses at locations where in-situ measurements of EC and MS were available with the measured FDEM responses. Strong correlations between the theoretical and measured response were found, except for the PRP2 data, which may be related to low signal-to-noise ratio and/or high sensitivity to surface conditions.

Electrical conductivity and magnetic susceptibility data were collected at twelve calibration data locations (boreholes), with intervals measurements in depth of 5-10 cm, some reaching depths of 1.2 m. The magnetic susceptibility data were collected with an MS2H downhole probe (Bartington instruments, England) in 2.5 cm diameter gouge borehole, reaching a minimum of 15 cm in the chalk bedrock and its expected background susceptibility (zero). The electrical conductivity data were collected using a UMP-1 BTim field probe (UGT) in a 5 cm diameter borehole. Also, at each calibration location, a lithological description was made and the depth of the boundary between the two formations (depth to chalk) was measured. Figure 2.9 shows the location of the inversion grid used in this application example. This data corresponds to part of the entire survey data presented in Delefortrie et al. (2018). The high in-phase anomalies are related to buried archaeology, and the location of the available borehole data. From the existing set of direct measurements, one borehole was kept out of the conditioning data and used exclusively as blind test to evaluate the performance of the proposed FDEM inversion method.

2.4.2 Results

Figure 2.9b shows the interpolated map of the predicted IP data, for PRP coil configuration with 1.1 m offset, from the pointwise mean of EC and MS models generated during the last iteration of the inversion procedure. The main archaeological features observed in the field map do match the ones predicted by the most likely model for the same coil configuration (Figure 2.9a), with strong local in-phase anomalies. However, this prediction is smoother due

to the fact of being computed from the pointwise mean models of EC and MS generated during the last iteration of the inversion procedure.

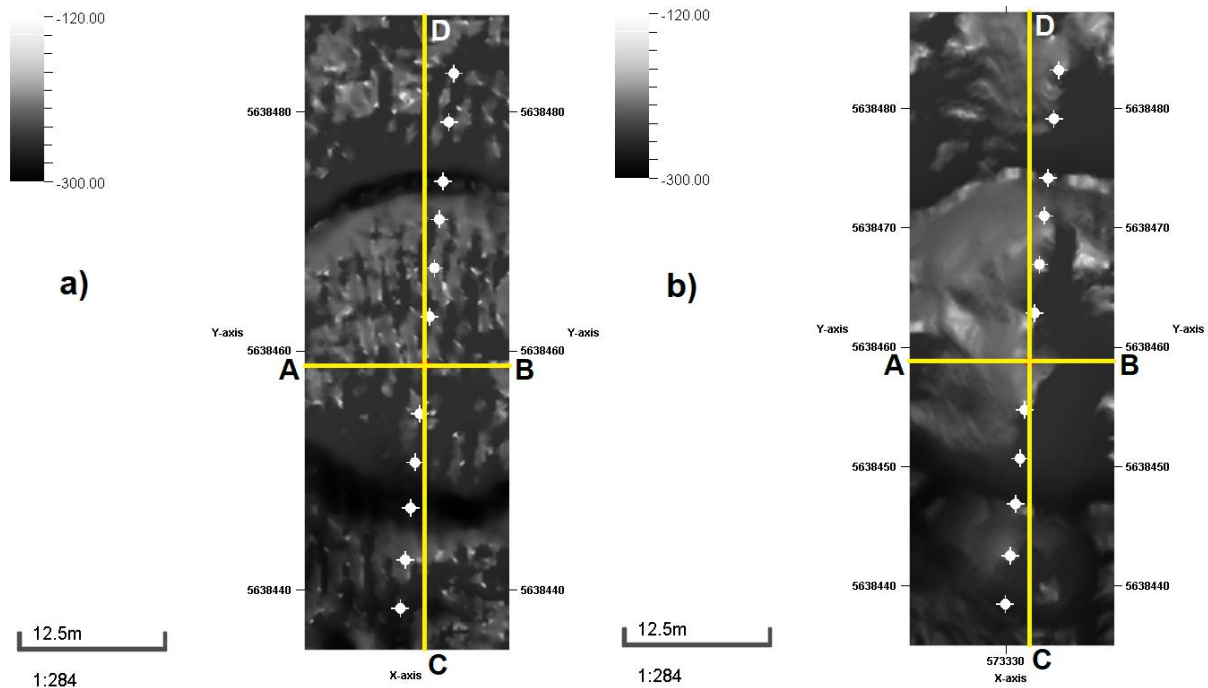


Figure 2.9 IP maps (ppm) for PRP coil configuration with 1.1 m offset, using a 2D median filter of 0.3 x 0.3 m and: a) the observed FDEM data; b) the synthetic FDEM data computed from the pointwise mean models of EC and MS of the 6th iteration. The yellow lines represent the location of the vertical sections (A-B and C-D) of Figures 2.9 to 2.14. White points represent the locations of the available borehole data (blind well in red). Coordinate system in WGS84/ UTM zone 30N, EPSG:32630

To assess the performance of the proposed iterative geostatistical FDEM inversion technique, Figures 2.10 to 2.13 show the pointwise mean and variance models of the ensemble of EC and MS models predicted at the first and the last iterations of the proposed iterative geostatistical FDEM inversion. The pointwise variance models of both properties clearly show the influence of including the sensitivity of the forward model to the model parameters, increasing in depth as the sensitivity of the FDEM decreases, in line with the results achieved in the synthetic case application. Two distinct regions can be clearly observed, a shallower one with lower variance and a deeper one with higher variance where the influence of the recorded FDEM data is small. With the coil configurations used, the sensitivity of the inversion procedure is limited to approximately 2 meters depth in EC models and 1 meter depth in MS models.

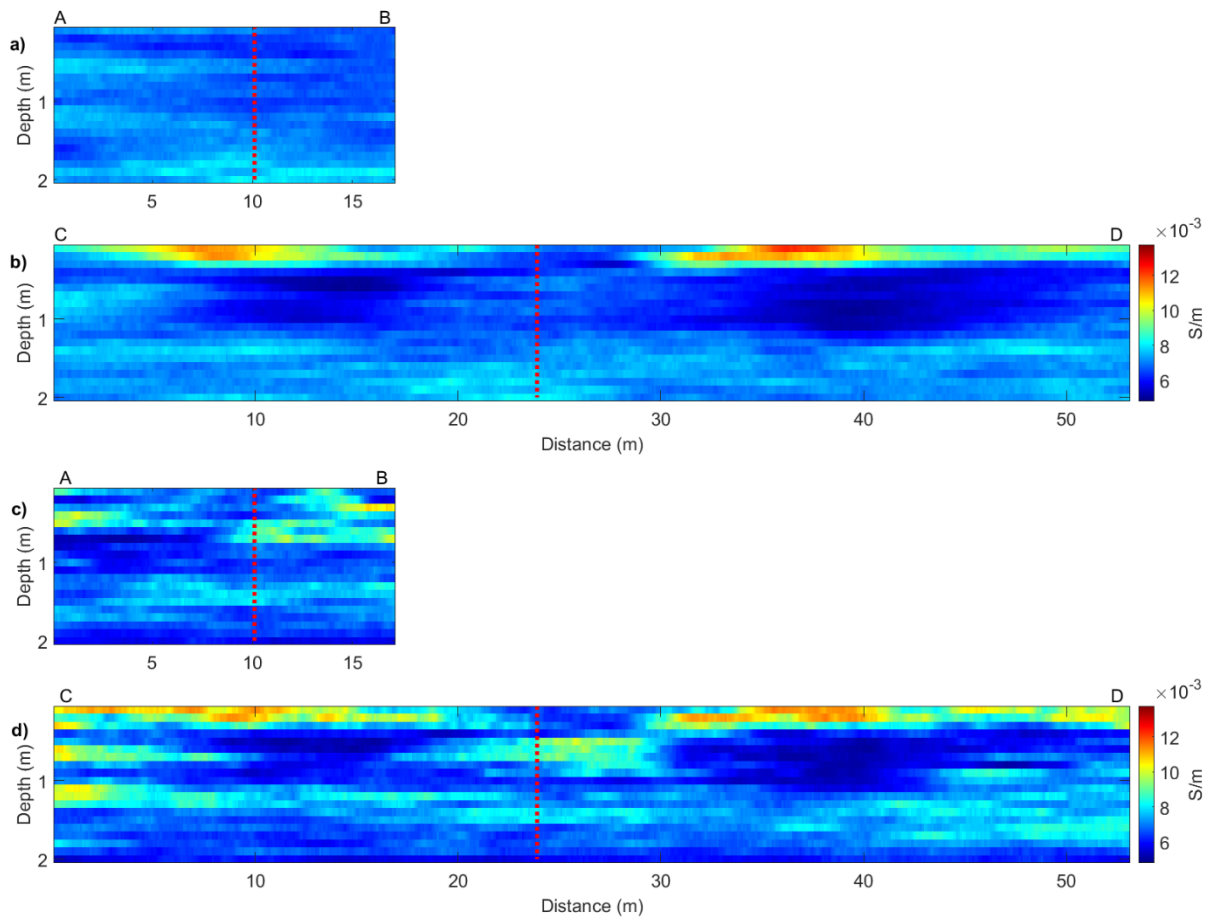
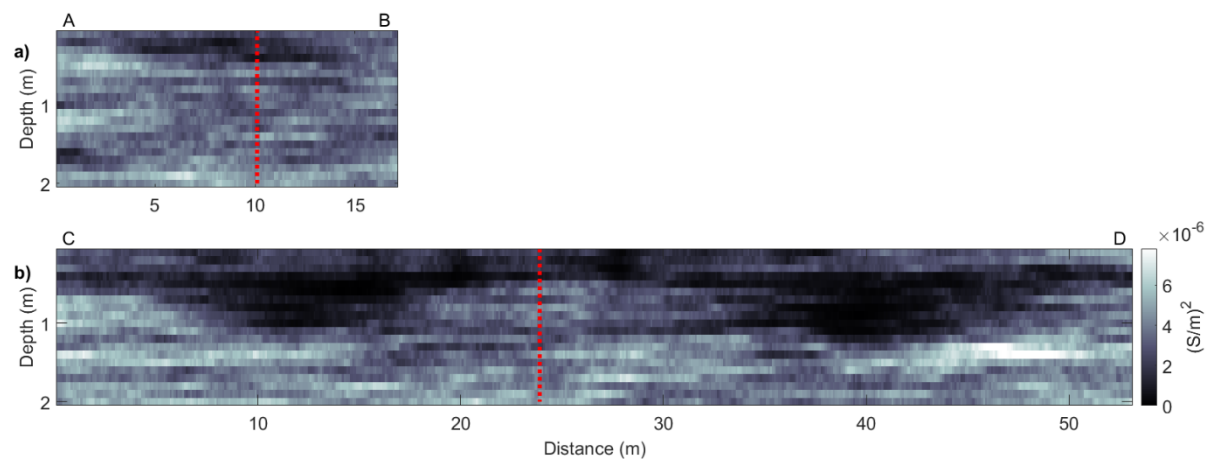


Figure 2.10 Vertical sections extracted from: a) and b) pointwise mean model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise mean model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.



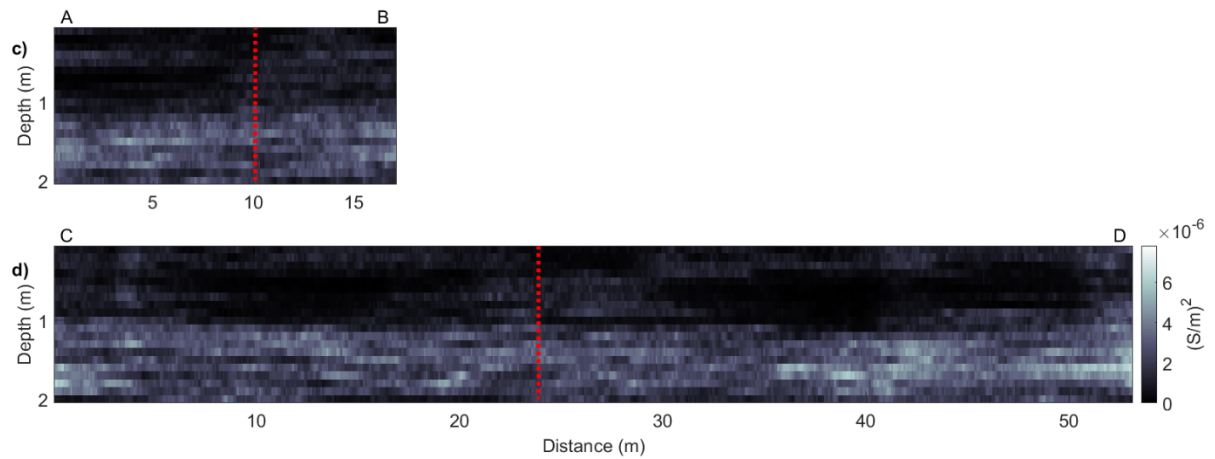


Figure 2.11 Vertical sections extracted from: a) and b) pointwise variance model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise variance model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.

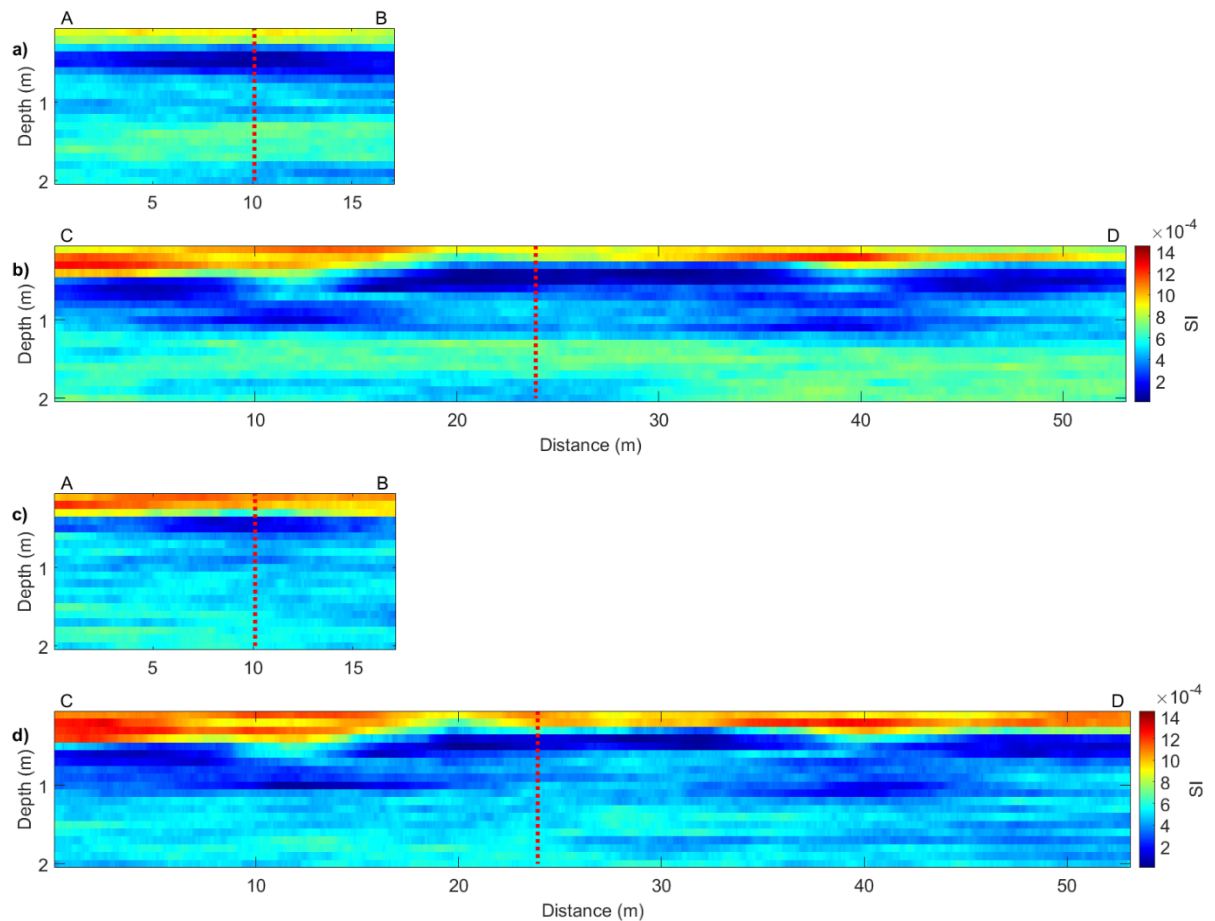


Figure 2.12 Vertical sections extracted from: a) and b) pointwise mean model calculated from all models of MS generated during the 1st iteration; c) and d) pointwise mean model calculated from all models of MS generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.

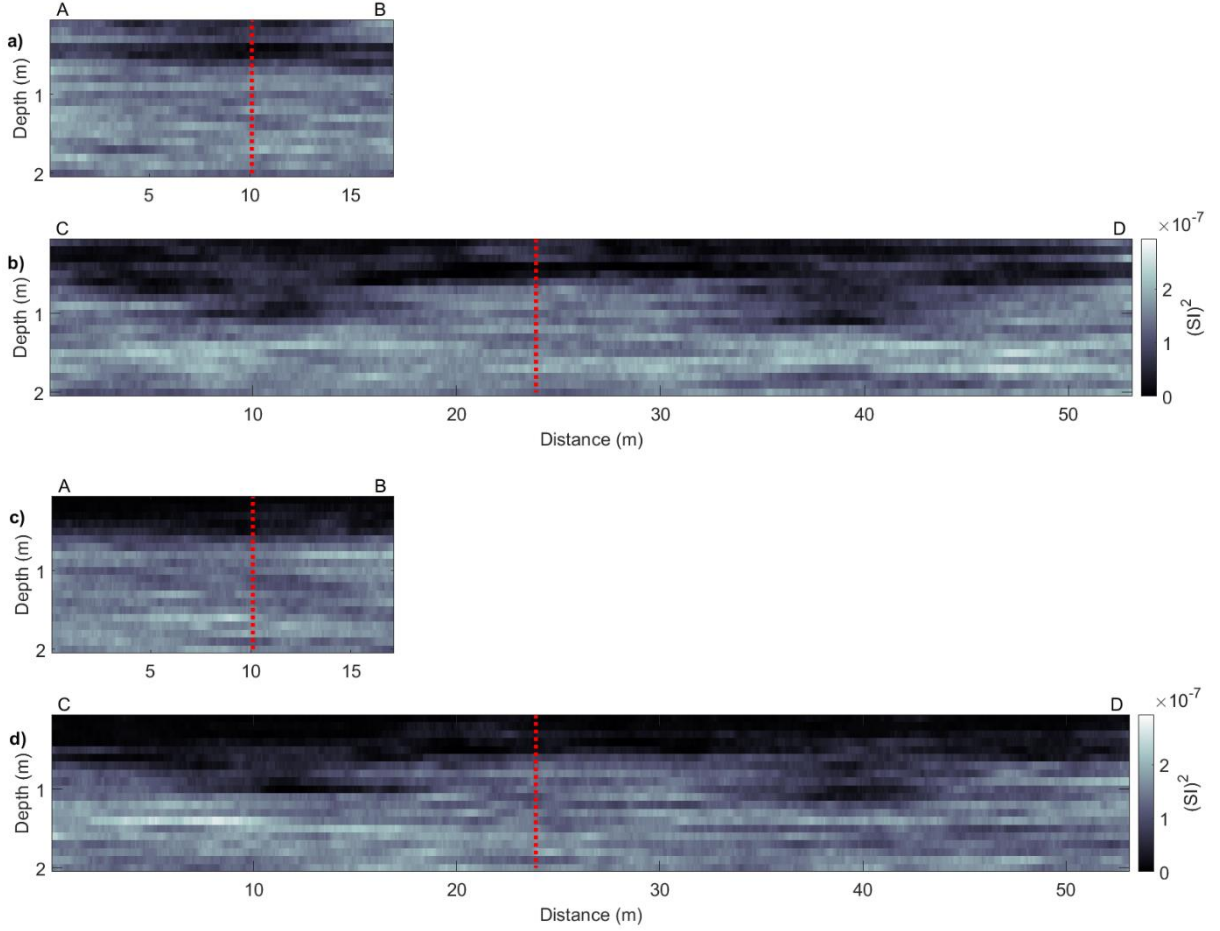


Figure 2.13 Vertical sections extracted from: a) and b) pointwise variance model calculated from all models of EC generated during the 1st iteration; c) and d) pointwise variance model calculated from all models of EC generated during the 6th iteration. The vertical dashed red line represents the location of the blind well.

The performance of the proposed methodology can also be assessed by the misfit between observed and predicted FDEM data (Figures 2.14 and 2.15) along the direction of the same two-dimensional profile C-D shown in Figures 9 to 13. The predicted FDEM responses were calculated from the ensemble of all models generated during the first and last iterations, for all coil configurations. The increasing convergence from iteration-to-iteration is illustrated by the envelope of the synthetic FDEM responses that gets narrower and closer to the observed data as the iterative procedure moves forward.

Contrary to the IP data, the observed QP responses are better enclosed by the predictions generated during the last iteration. The better reproduction of QP is due to the higher signal-to-noise ratio of this component of the data. This effect is a consequence of the type of FDEM equipment used during the data acquisition. In general, the predicted FDEM data for both the IP and QP signal components, and all coil configurations and offsets, do match the recorded field data.

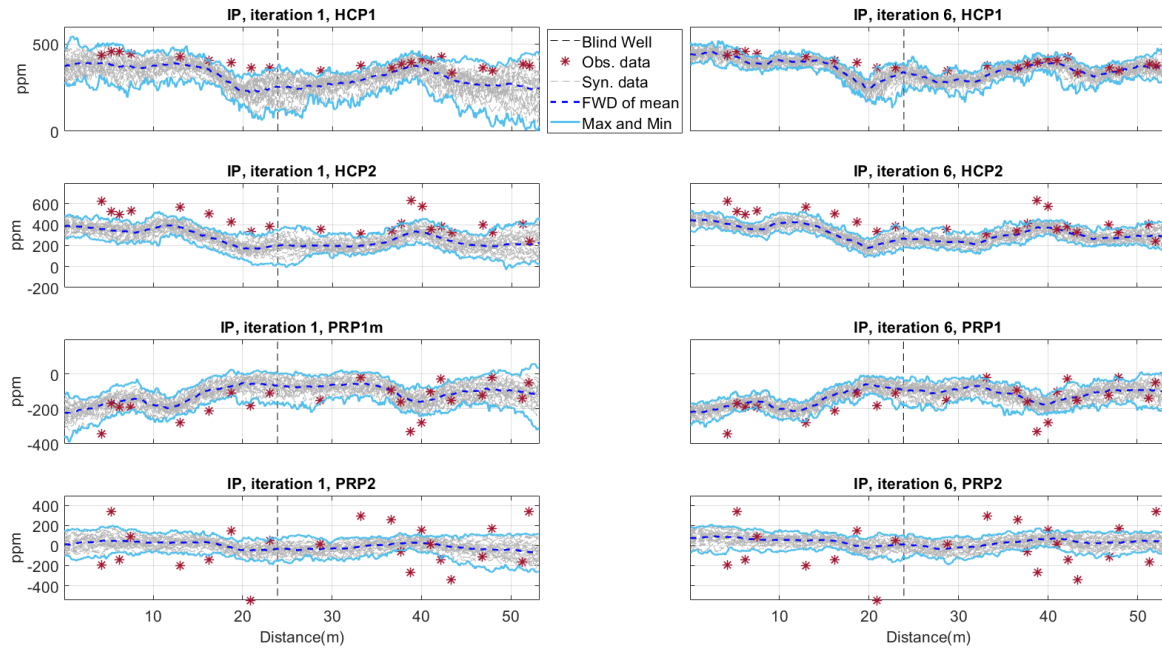


Figure 2.14 Comparison between observed (red stars) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration. The blue lines represent the minimum and maximum FDEM values predicted at a given iteration. The left column represents the predictions at the end of the first iteration and the right column at the end of the last iteration. Vertical dashed lines represent the location of the borehole data.

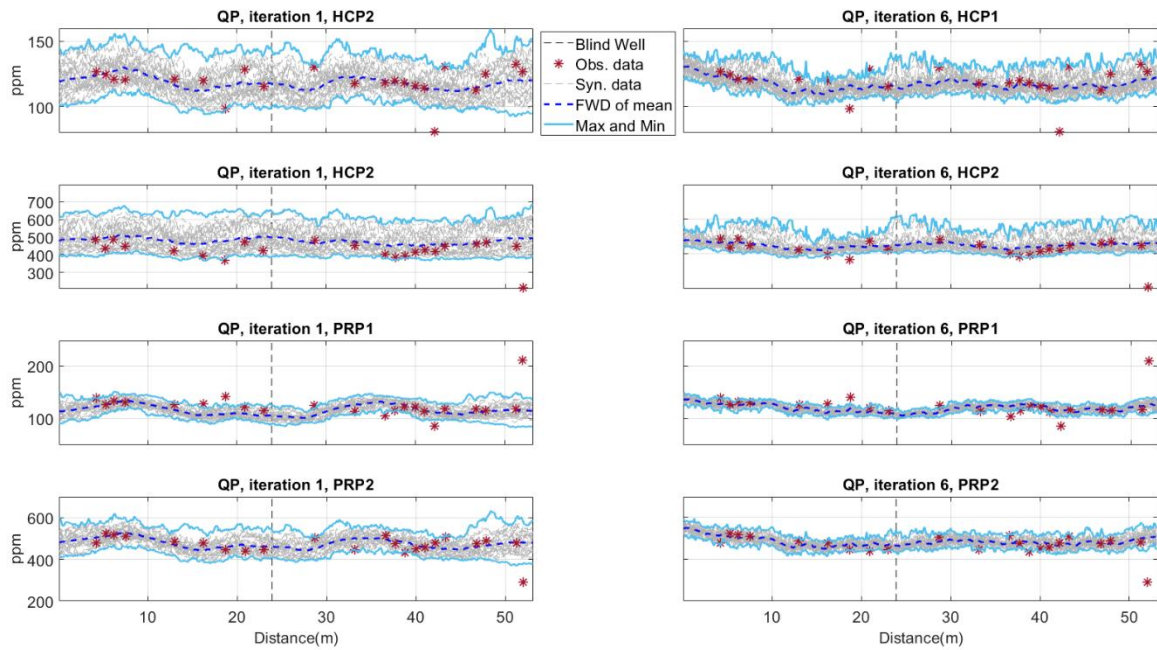


Figure 2.15 Comparison between observed (red stars) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration. The blue lines represent the minimum and maximum FDEM values predicted at a given iteration. The left column represents the predictions at the end of the first iteration and the right column at the end of the last iteration. Vertical dashed lines represent the location of the borehole data.

2.5 Discussion

The proposed iterative geostatistical FDEM inversion method predicts near-surface EC and MS models from survey FDEM data (i.e., simultaneously from the IP and QP components of the electromagnetic field). The inversion method is based on geostatistical simulation and co-simulation as model perturbation and update techniques. Therefore, the predicted models can be conditioned locally to existing borehole data and a spatial continuity pattern as described by a variogram model. The perturbation of the model parameters at each iteration leverages the sensitivity analysis provided by the forward model (i.e., the assimilation of the recorded FDEM data accounts for the sensitivity in depth per property as provided by the forward model). The proposed FDEM inversion method is based on a 1D forward model. This is a limitation as the propagated electromagnetic field is 3D in nature. Alternative 2D or 3D forward models can be used if they provide a sensitivity analysis to the model parameters. However, these forward models would increase the computational costs of the inversion. This hard assumption is somehow alleviated in the proposed methodology as the model perturbation is global for the entire grid at once (i.e., in 2D or 3D depending on the data availability).

The synthetic application example illustrates the potential of the proposed FDEM inversion method to predict a reliable near-surface model. However, this is a relatively simple example as the same forward model used to create the observed data was used in the inversion workflow. This limitation is surpassed in the real case application as the field data is three-dimensional in nature while we use the same 1D forward model approximation in the inversion procedure. This computational limitation introduces uncertainty in the model prediction.

We evaluate the predicted EC and MS models locally at the location of the “blind well” (Figure 2.16). Despite the much shallower borehole information compared to the depth of the inversion model, the predicted near-surface properties do match the observed one at the last iteration.

The match is better for MS when compared to EC, predicting properly the abrupt change of MS around 0.25 m depth (Figures 2.12 and 2.13), consistent with the expected values at depth of magnetic susceptibility for rendzina soil cover and the bedrock chalk. Overall, EC models confirm the low conductivity of the subsurface and the topsoil, with slight differences between both (Figures 2.10 and 2.11). Figure 2.16 also clearly shows the effect of the DOI for both properties. Below 0.5 m, for MS, and 1 m, for EC, the predictions at iteration 6 are more variable than above these depths. Finally, Figure 2.17 shows a comparison of the results of the sensor’s default output for apparent EC (ECa), between the interpolation of the observed QP component using the low-induction-number (LIN) approximation (Delefortrie et al., 2018), with the same LIN ECa interpolation using the best-fit synthetic QP component computed during

the inversion procedure. Both models are similar concerning the range of predicted values and large-scale spatial features. However, the predicted LIN ECa model from the synthetic data has more small-scale variability, which is originated by the stochastic nature of the proposed iterative geostatistical inversion method, resulting in an increased sensitivity to the spatial structure variations.

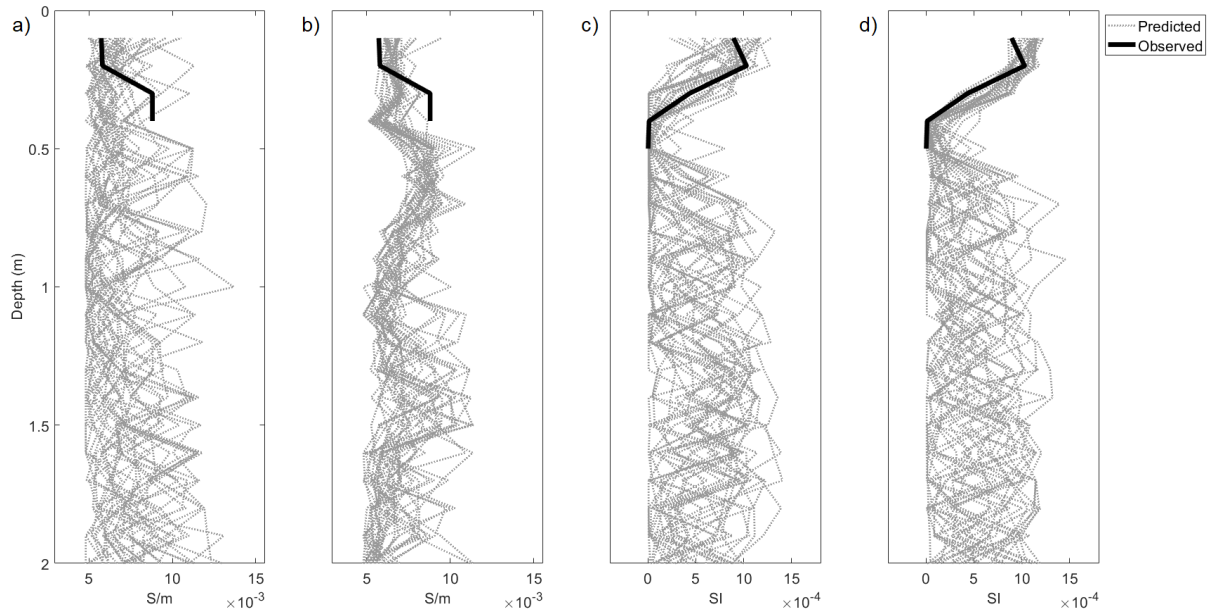


Figure 2.16 True and predicted values of EC and MS along the blind test for: a) EC during the 1st iteration; b) EC during the 6th iteration; c) MS during the 1st iteration; d) MS during the 6th iteration.

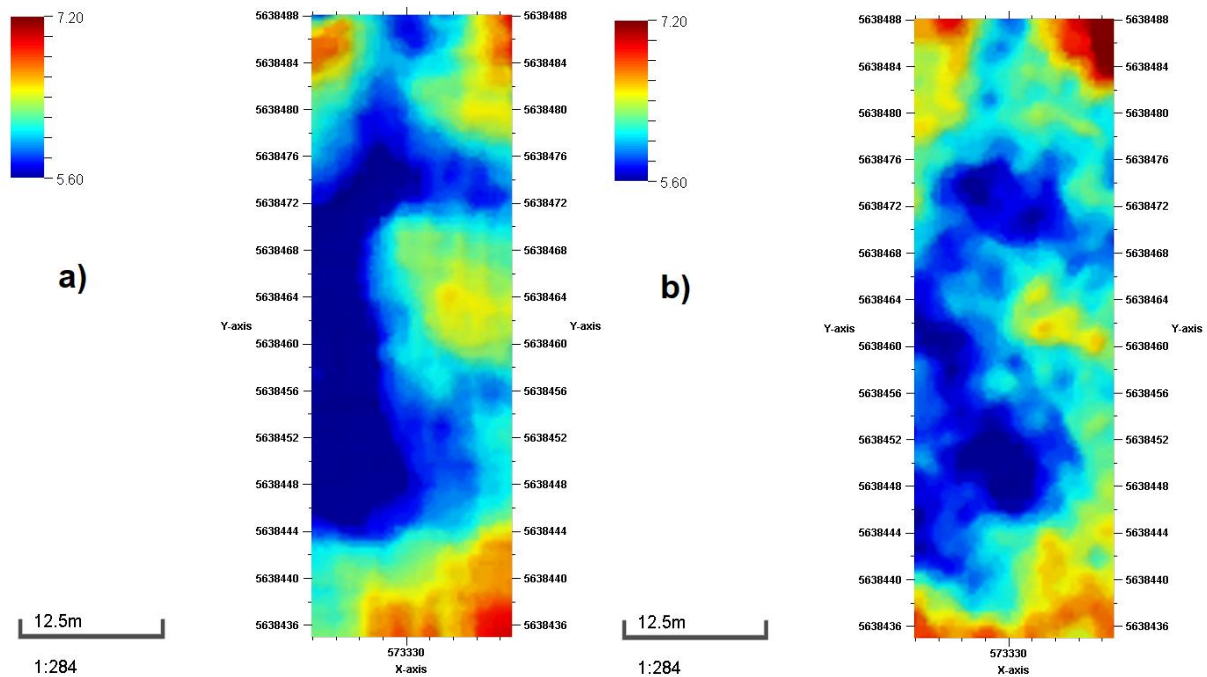


Figure 2.17 LIN ECa maps (mS/m) of the HCP coil configuration with 2 m offset using a median filter of 0.3×0.3 m from a) the observed data b) the synthetic data predicted in the iteration 6th.

2.6 Conclusion

This Chapter introduces an iterative geostatistical FDEM inversion method able to predict the spatial distribution of EC and MS simultaneously in complex and heterogeneous subsurface environments. It represents an alternative method in quantitative near-surface modelling using FDEM survey data and can be universally applied to characterize near-surface deposits of different types and nature, which is relevant to a wide variety of applications.

The FDEM inversion methodology was first validated using a developed three-dimensional synthetic data set rendering realistic spatial distributions of EC and MS and then applied to a real data set from a FDEM survey at a site containing several archaeological features that produce strong local in-phase anomalies. In both application examples, the pointwise mean and variance models were computed from geostatistical realizations generated during the first and last iterations, demonstrating that the predicted models reproduce the measured EC and MS. Also, the proposed methodology reproduces exactly the histograms retrieved from the borehole data and tends to reproduce the variogram models imposed during the stochastic sequential simulation and co-simulation of EC and MS.

The FDEM measurement responses derived from the ensemble of EC and MS models predicted by the iterative FDEM inversion method were able to enclose the true FDEM data in the synthetic case application, and in the real case application the methodology well reproduces the FDEM observed data from the coil configurations with higher signal-to-noise ratio. This methodology can assess the uncertainty of the FDEM responses as well the uncertainty of the posterior distributions of EC and MS.

The proposed inversion relies on a one-dimensional forward approximation but could be extended to more complex physical models from the imposed spatial structure. A critical aspect for the success of the proposed inversion method is the availability of calibration data, in the form of borehole observations and/or in-situ measurements, and their spatial distribution within the area of interest. The accuracy of the inversion results might be affected if only a limited number of calibration data are available, correspondingly leading to large uncertainties in the predictions. In real applications, this sometimes can be overcome by accounting for direct measurements from nearby areas close to the area under investigation.

CHAPTER 3

Comparison between probabilistic inversion methods of FDEM

The research presented in this chapter was published on:

Narciso, J.*, Bobe, C., Azevedo, L., and E. Van De Vijver, 2022, A comparison between Kalman ensemble generator and geostatistical frequency-domain electromagnetic inversion: the impacts on near-surface characterization. *Geophysics*, 27(5), 1-12.

The spatial distribution of the physical properties of the first meters beneath the earth's surface is often complex due to its highly dynamic nature and small-scale heterogeneities resulting from natural and anthropogenic processes. Therefore, obtaining numerical 3D models that accurately describe the spatial distribution of these properties is often challenging yet essential for different fields such as environmental assessment and remediation, geoarchaeological conservation, and precision agriculture. Frequency-domain electromagnetic (FDEM) induction methods have proven their potential to image these properties in high (spatial) detail because FDEM measurements are sensitive to two key soil properties: electrical conductivity and magnetic susceptibility. Predicting subsurface properties from FDEM data requires solving an ill-posed and nonlinear inverse problem with multiple solutions. Recently, there has been a rapid growth of FDEM inversion methods, which may be broadly divided into probabilistic and deterministic methods. This chapter presents the comparison between two stochastic FDEM inversion approaches: the Kalman ensemble generator (KEG) and the iterative geostatistical FDEM inversion presented in Chapter 2. Both methods are applied to a synthetic data set with spatially heterogeneous physical properties of interest, mimicking a real landfill mining site. The predicted models are compared with the reference models in terms of histogram and variogram models' reproduction and in their ability to quantify spatial uncertainty. The results indicate the ability of both methods to predict the reference values. Although the KEG is computationally efficient, it struggles to reproduce the extreme values. In contrast, the geostatistical inversion approach ensures the reproduction of the imposed histograms and variogram models in the predicted models. As the prior information is included in both inversion methods in different ways, the pointwise variance models computed from all of the posterior models have different information. The synthetic data set is available to the community, so it can be used as a benchmark for other FDEM inversion methods.

3.1 Introduction

The near surface is a complex and highly dynamic region of the subsurface due to intense natural and anthropogenic activities. These dynamics result in complex systems, which often are characterized by physical properties with small-scale heterogeneity. This complexity can make sparse and discrete direct observations (e.g., boreholes) insufficient to provide sufficient information about the spatial distribution of these properties in the horizontal direction, causing

a simple interpolation of the borehole data to be unsuitable to reproduce the natural spatial variability of these systems.

Contrary to direct observations, which are always scarce, several geophysical survey methods can be used to acquire indirect, and virtually continuous, measurements of the physical properties of the near surface. Particularly, frequency-domain electromagnetic (FDEM) induction methods have been found to be some of the most efficient methods to reach this objective due to their relatively low cost, operational flexibility, and sensitivity to two key subsurface properties: electrical conductivity (EC) and magnetic susceptibility (MS). Due to these features, FDEM induction methods have been used for, for example, environmental assessment and remediation (Van De Vijver et al., 2015; Dumont et al., 2017), soil characterization (e.g., Haber et al., 2004; Saey et al., 2015), groundwater characterization (Huang et al., 2017; Rejiba et al., 2018), archaeological prospection (e.g., Bongiovanni et al., 2008; De Smedt et al., 2011), and agricultural application (e.g., Pedrera-Parrilla et al., 2016; Badewa et al., 2018).

Due to the indirect and band-limited nature of the FDEM data, predicting the spatial distribution of the near-surface EC and MS from FDEM data, considering components of the signal — in phase (IP) and quadrature phase (QP) — is an ill-posed and non-linear inverse problem with multiple solutions. Stochastic frameworks partially address the previously mentioned challenges of this inversion problem as such settings enable accounting for non-uniqueness and nonlinearity. For this reason, and the increase in available computational resources, the number of publications concerning statistical-based inversion methods of FDEM data has been growing in recent years (e.g., Minsley, 2011; Bobe et al., 2019). The FDEM inversion problem can be mathematically summarized by Eq. 2.1.

This Chapter presents the results obtained by applying two distinct stochastic FDEM inversion methods to a highly spatially variable 3D synthetic data set composed of the reference EC and MS models, a set of synthetic borehole data extracted from the 3D models, and the theoretical FDEM response — IP and QP components — calculated from the EC and MS models. The FDEM responses calculated from the EC and MS models represent the reference FDEM measurement data mimicking a ground-based survey. The two FDEM inversion methods applied are based on (1) the Kalman ensemble generator (KEG; Nowak, 2009; Bobe et al., 2019) and (2) the GEMI method presented in Chapter 2.

The KEG is a Monte Carlo implementation of a Bayesian parameter estimation problem for Gaussian probability distributions, where the covariance matrices are replaced by sample covariance, avoiding a linearization of the forward equation (Zhou et al., 2011). The KEG uses the update step of the ensemble Kalman filter (EnKf) (Evensen, 2003), where the KEG update

is limited to the parameter space. For nonlinear inverse problems, the KEG gives a fast, first-order (Gaussian) approximation to the FDEM inverse solution (Bobe et al., 2019).

The GEMI method, as expose in Chapter 2, is an iterative FDEM inversion procedure based on the principles of global iterative geostatistical seismic inversion methods (Azevedo and Soares, 2017), which allows simultaneous prediction of EC and MS from FDEM data. This method has two underlying key ideas: (1) stochastic sequential simulation and co-simulation are used for global model parameter space perturbation and (2) the convergence is ensured by a global stochastic optimizer driven simultaneously by the misfit between reference and synthetic FDEM data and the predicted sensitivity of EC and MS. Both inversion methods use the same 1D forward model (Hanssens et al., 2019) as part of the inversion procedure to calculate the synthetic FDEM response for the model realizations.

This Chapter evaluate and compare quantitatively the results of both methods in terms of deviations from the reference model, reproduction of the global statistics and the variogram model computed from the reference EC and MS models in the predicted EC and MS models, and uncertainty assessment as represented by the pointwise variance of the posterior ensemble.

3.2 Methodologies

This section describes both stochastic FDEM inversion methods, providing a detailed description of the main principles of the KEG (Bobe et al., 2019). A detailed description of the GEMI method can be found in section 2.2. A detailed description of the forward model used by both methods (Hanssens et al., 2019) can be found in section 2.2.2.

3.2.1 Forward response and sensitivity modelling

The synthetic FDEM responses – IP and QP – are calculated in each method using a 1D nonlinear forward model that allows mapping the near-surface petrophysical properties (i.e., EC and MS) into the data domain, since IP and QP responses pertain to both MS and EC (Hanssens et al., 2019). This forward model replicates the components of the electromagnetic field as acquired by a loop-loop system characterized by one transmitter coil and one or more receiver coils. The primary field is computed by assuming an alternating current in the transmitter coil, while the secondary field is calculated based on induction currents in relation to the primary field.

The theoretical IP and QP responses are calculated per transmitter-receiver coil configuration located above a l -layered model by using Hankel functions, which are numerically calculated

by means of a Guptasarma and Singh digital filter (Guptasarma and Singh, 1997), to determine a superposition of Bessel functions of the zeroth and/or first order that model the EM responses. For low-frequency applications, a quasi-static approximation can be applied, so dielectric permittivity is negligible. This assumption results in a signal mostly depending on the spatial distribution of the subsurface EC and MS and on a lesser extent the characteristics of the acquisition equipment.

Along with the FDEM response, the GEMI method computes the sensitivity of the forward model with respect to changes of EC and MS at a specific layer l of the layered half-space. In contrast to the KEG, the GEMI methodology explores this result when doing the stochastic model update at the end of each iteration of the inversion procedure. However, this is a time-consuming calculation with an impact on the performance of the geostatistical inversion. This bottleneck can be mitigated by parallelizing the application of the forward model as this is a 1D approximation of the true three-dimensional field propagation or by making use of methods computing approximate sensitivities for the FDEM forward problem (Farquharson and Oldenburg, 1996; da Cruz Luz et al., 2013). A 1D forward model approximates the true subsurface field propagation that might not be suitable for geological settings highly affected by three-dimensional structure. A detailed mathematical description of this forward model is available in Hanssens et al. (2019) and in section 2.2.2.

3.2.2 The Kalman ensemble generator

The KEG is a variant of the widely applied EnKf (Evensen, 2003). The EnKf was introduced as an efficient Monte Carlo implementation of the Kalman filter (Kalman, 1960), where the covariance equations of the Kalman filter update are replaced by sample covariance derived from the Monte Carlo ensemble. The EnKF is based on the assumption that all probability density functions (PDFs) involved in the update are Gaussian but allow the application to nonlinear inversion problems (Zhou et al., 2011). Initially used for state estimation, the EnKf was soon applied to the joint estimation of states and time-invariant physical parameter updates (e.g., Hendricks Franssen and Kinzelbach, 2008). The EnKf update equation applied to pure parameter estimation problems was introduced by Nowak (2009) and called KEG because the filter function of the EnKf was ignored.

Using the KEG, a measurement response is simulated as follows:

$$\mathbf{d}_{sim}^i = g(\mathbf{m}_{prior}^i) + \epsilon_{err}, \quad i \in \{1, \dots, n_{ens}\}, \quad (3.1)$$

where ϵ_{err} refers to the measurement error and approximations of the true physical phenomena made during data processing and modelling, n_{ens} is the size of the ensemble, and

$\mathbf{m}^i$ and $\mathbf{d}^i$ are the realizations of the Gaussian prior PDF and Gaussian observed data PDF, respectively. The KEG update is given by (Evensen, 2003):

$$\tilde{\mathbf{m}}^i = \mathbf{m}_{prior}^i + \mathbf{C}_{mg}^e (\mathbf{C}_{gg}^e + \mathbf{C}_D^e)^{-1} \cdot (\mathbf{d}_i - g(\mathbf{m}_{prior}^i)), \quad (3.2)$$

for $i \in \{1, \dots, n_{ens}\}$; where $\mathbf{C}_D^e$ is the covariance matrix of random observation errors. The covariance matrices $\mathbf{C}_{mg}^e \in \mathbb{R}^{n_{par} \times n_{obs}}$ and $\mathbf{C}_{gg}^e \in \mathbb{R}^{n_{obs} \times n_{obs}}$ are derived from the prior ensemble captured by a matrix $A \in \mathbb{R}^{n_{par} \times n_{ens}}$ and the forward response ensemble captured by a matrix $G \in \mathbb{R}^{n_{obs} \times n_{ens}}$, with n_{obs} being the number of observations and n_{par} being the number of model parameters. Furthermore, the matrices A' and C' are defined as the mean-corrected versions of the matrices A and G , where A' is derived using the matrix of ensemble means:

$$\bar{A} = A \mathbf{1}_{n_{ens}}, \quad (3.3)$$

where $\mathbf{1}_{n_{ens}} \in \mathbb{R}^{(n_{ens} \times n_{ens})}$ is the matrix where all elements equal $1/n_{ens}$ (Evensen, 2003). Using the matrix of ensemble means, the mean-corrected matrix is derived by computing:

$$A' = A - \bar{A}. \quad (3.4)$$

Analogously, matrix G' is derived from matrix G . Using the previously defined matrices, the computation of the covariance matrices is given by:

$$\mathbf{C}_{mg}^e = A' G'^T \frac{1}{n_{ens}-1} \quad \text{and} \quad \mathbf{C}_{gg}^e = G' G'^T \frac{1}{n_{ens}-1}. \quad (3.5)$$

An efficient numerical analysis scheme for the computation of the update is given by Nowak (2009) and outlined in detail for the inversion of FDEM data in Bobe et al. (2019). In contrast to the strict 1D formulation given in Bobe et al. (2019), in this work (Figure 3.1) the prior covariance matrix is defined by 2D correlation functions (i.e., in the vertical and horizontal directions) as defined by a variogram model manually fitted to experimental variograms calculated from the borehole data. The borehole data are considered to be hard data without uncertainty. This is an approach similar to geostatistical simulation (Deutsch and Journel, 1998). The variogram models (i.e., spatial covariance matrices) used in this work are shown in section 3.3.

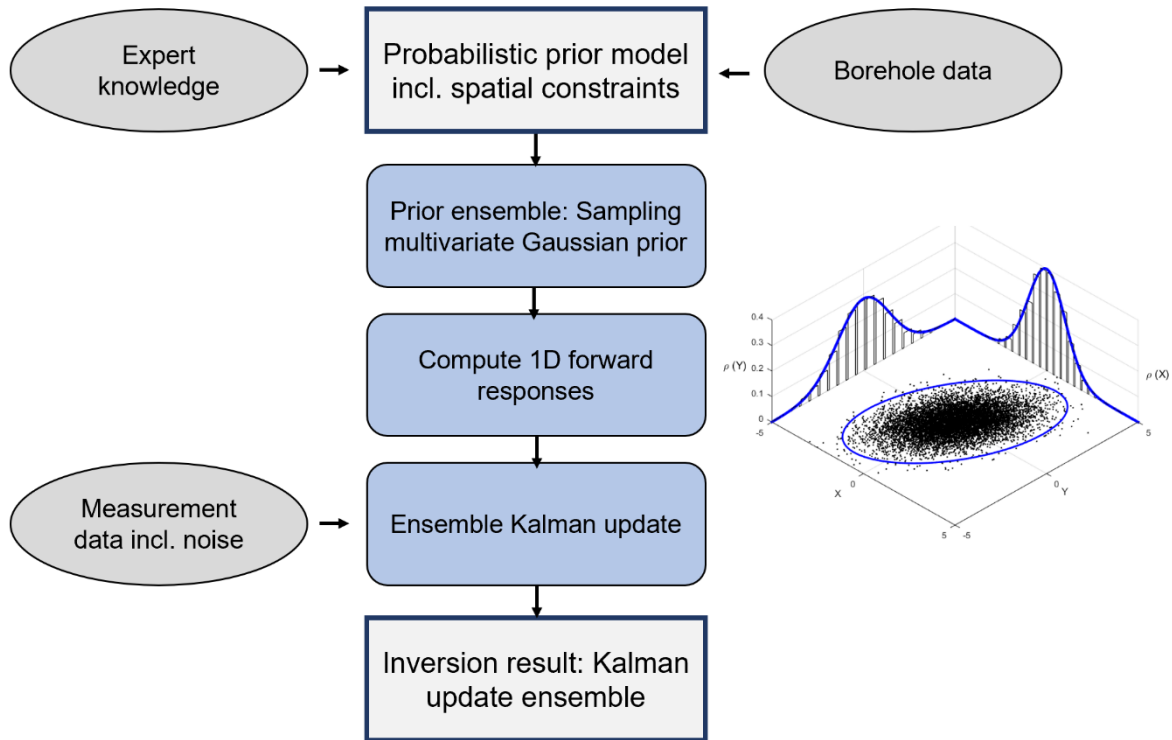


Figure 3.1 Schematic representation of the noniterative constrained KEG method.

3.2.3 GEMI inversion

The GEMI method (Figure 3.2) is an iterative inversion approach, which uses geostatistical simulation, namely, direct sequential simulation and co-simulation (Soares, 2001), to perturb the model parameter space, and uses the existing direct in situ measurements (e.g., data from borehole logs) as hard data without uncertainty. A variogram model is imposed during the geostatistical simulation to describe the expected spatial continuity pattern of EC and MS in three dimensions (Shamsipour et al., 2012). This variogram model is fitted to the experimental variogram calculated from the hard data but might be adjusted according to the expert knowledge of the expected geologic background. In all realizations generated during the iterative procedure, the hard data are reproduced exactly at their location (i.e., no uncertainty is considered at the locations of the direct observations) as well as their distributions (i.e., the histogram) and the spatial continuity pattern as defined by the variogram model. A detailed description of the GEMI method can be found in section 2.2.

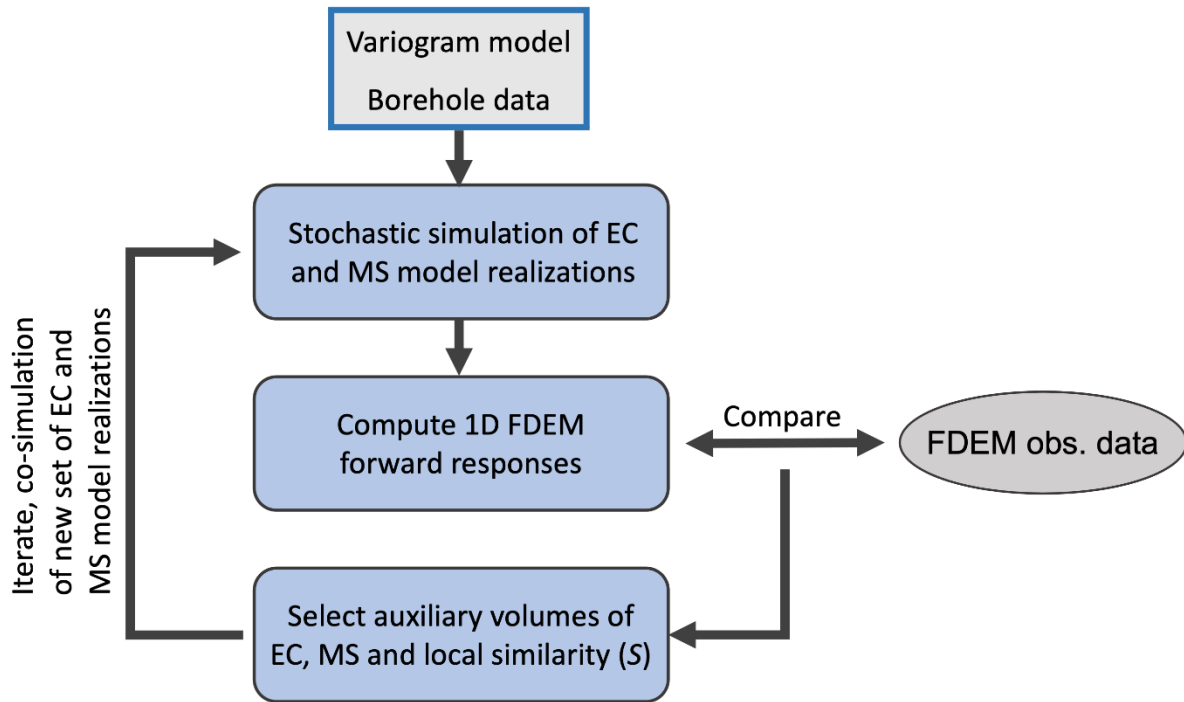


Figure 3.2 Simplified schematic representation of the iterative GEMI method.

3.3 Application

3.3.1 Data set description

A realistic 3D synthetic data set based on the synthetic data set described in section 2.3.1, was used as a benchmark between alternative FDEM inversion methods. Although this data set was modeled with some of the geologic samples used to develop the synthetic data set from section 2.3.1, and collected at the same mine tailing, the data set presented in this Chapter was created with the goal of mimicking a three-dimensional larger mine tailing, with more spatial continuity of the properties of interest. This approach enabled to test the robustness of the GEMI method, described in Chapter 2, when applied to different subsurface environments, in scale and heterogeneity.

Direct measurements of porosity and particle density were used to generate a 3D synthetic porosity subsurface model using stochastic sequential simulation (Deutsch and Journel, 1998) and imposing an omnidirectional spherical variogram model in the horizontal direction with a range of 10 m and a range of 1 m in the vertical direction and a nugget effect of 20% of the total variance. We modeled these variogram models using exclusively the location of the samples collected at the mine tailing. The model has a dimension of 150 m × 200 m × 4 m (i.e., length, width, and depth) with a cell size of 0.5 m × 0.5 m × 0.1 m, respectively. To ensure

a realistic relationship between properties, particle density and water content models were generated using stochastic sequential co-simulation (Deutsch and Journel, 1998) conditioned to the porosity model and imposing an omnidirectional spherical variogram model with a horizontal range of 10 m and a vertical range of 1 m and a nugget effect of 20% of the total variance for the particle density and an omnidirectional spherical variogram model with a horizontal range of 16 m and a vertical range of 2 m and a nugget effect of 10% of the total variance for the water content.

EC and MS models were then derived from these 3D physical property models and the Archie equation (Archie, 1942) and imposing an omnidirectional spherical variogram model with a horizontal range of 20 m and a vertical range of 4 m and a nugget effect of 10% of the total variance for the MS models.

From the resulting 3D models, were selected nine locations to represent synthetic boreholes that were used as in situ data conditioning the inversion. The corresponding observed FDEM data were obtained using the same forward model (Hanssens et al., 2019) as in the inversion procedures described in section 2.2.2, where the 1D forward model responses were stitched for forming the FDEM measurement transect following the borehole locations (Figure 3.3). The synthetic FDEM values were generated replicating one of the most common sensors for these types of near-surface surveys, namely, the DUALEM-421S (DUALEM Inc., Milton, Canada), considering two loop-loop coil orientations, a horizontal coplanar (HCP) and a perpendicular one (PRP), with the normal three offsets per coil orientation for this equipment, 1, 2, and 4 m for HCP and 1.1, 2.1, and 4.1 m for PRP, plus an extra offset per coil orientation, 10 m for HCP and 10.1 m for PRP, ensuring a theoretically larger depth of investigation (DOI).

This data set is considered as the reference to assess the performance of the GEMI and KEG inversion methods. The data set is available freely (<http://doi.org/10.5281/zenodo.5116420>) and can be used to assess the advantages and disadvantages of different methods.

For illustration purposes, the comparison between the two probabilistic inversion methods is shown for a 2D profile extracted from the true 3D model. The selected 2D profile was aligned with the nine boreholes where EC and MS logs were extracted (Figure 3.3). This geometry represents an ideal scenario in which the direct subsurface measurements are regularly spaced along the geophysical profile.

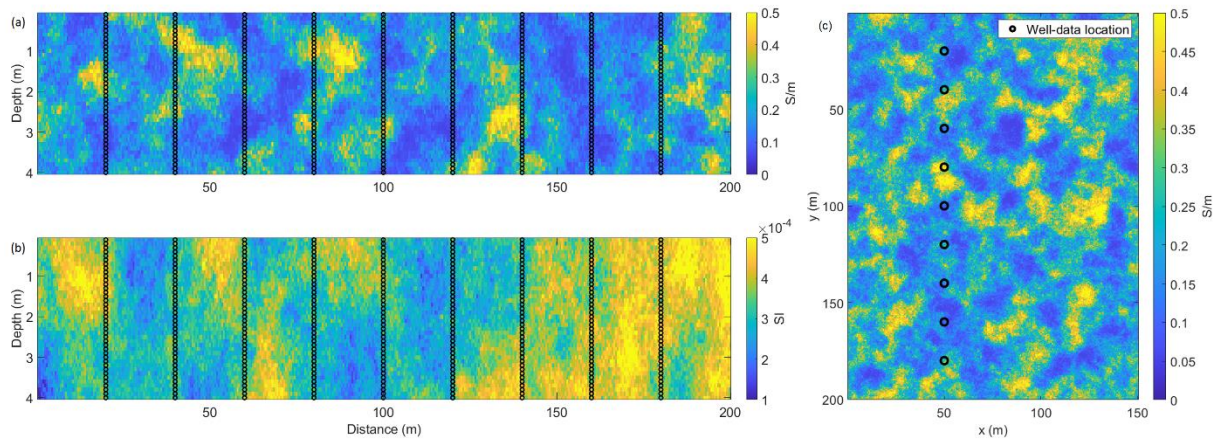


Figure 3.3 Cross sections extracted from the reference models of (a) EC and (b) MS. The black dots represent the nine wells. (c) Location of the nine wells in the EC model map.

3.3.2 Inversion parametrization

The GEMI ran with six iterations and generated sets of 32 realizations of EC and MS per iteration. EC models were simulated and cosimulated imposing an omnidirectional spherical variogram model with a horizontal range of 10 m and a vertical range of 4 m and a nugget effect of 20% of the total variance of the data. MS models were simulated and cosimulated imposing an omnidirectional spherical variogram model in the horizontal direction with a range of 40 m, a vertical range of 30 m, and a nugget effect of 10% of the total variance of the data. The variogram models were calculated relying exclusively on the borehole data set to mimic a real application scenario (Table 3.1). Consequently, these variogram models are not the same as those used in the true model generation. In addition to the borehole data, no other spatial constraint was considered in the geostatistical inversion.

Table 3.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.

Omnidirectional spherical variogram	EC models	MS models
Horizontal range	10 m	40 m
Vertical range	4 m	30 m
Nugget effect	20 %	10 %

The KEG used a prior ensemble with 500 models of EC and MS. This ensemble was generated using direct sequential simulation as the model perturbation technique of GEMI. The same parameterization in terms of the number of experimental data and variogram models used to constrain the GEMI was considered to create this ensemble of realizations. These models are equivalent to those generated in the first iteration of the GEMI. They reproduce the borehole

data at the corresponding locations and the variogram models imposed during the model simulation. In this way, the prior information of both inversion methods is the same. However, although the GEMI is an iterative procedure, the KEG infers the posterior distribution in a single-step update. A cross section extracted from the pointwise average EC and MS models computed from the prior ensemble is shown in Figure 3.4. All realizations match the true borehole data. The influence of these data in the horizontal direction is dependent on the horizontal variogram model. For distances greater than the horizontal range, the simulated values tend to be the mean value of the observed data.

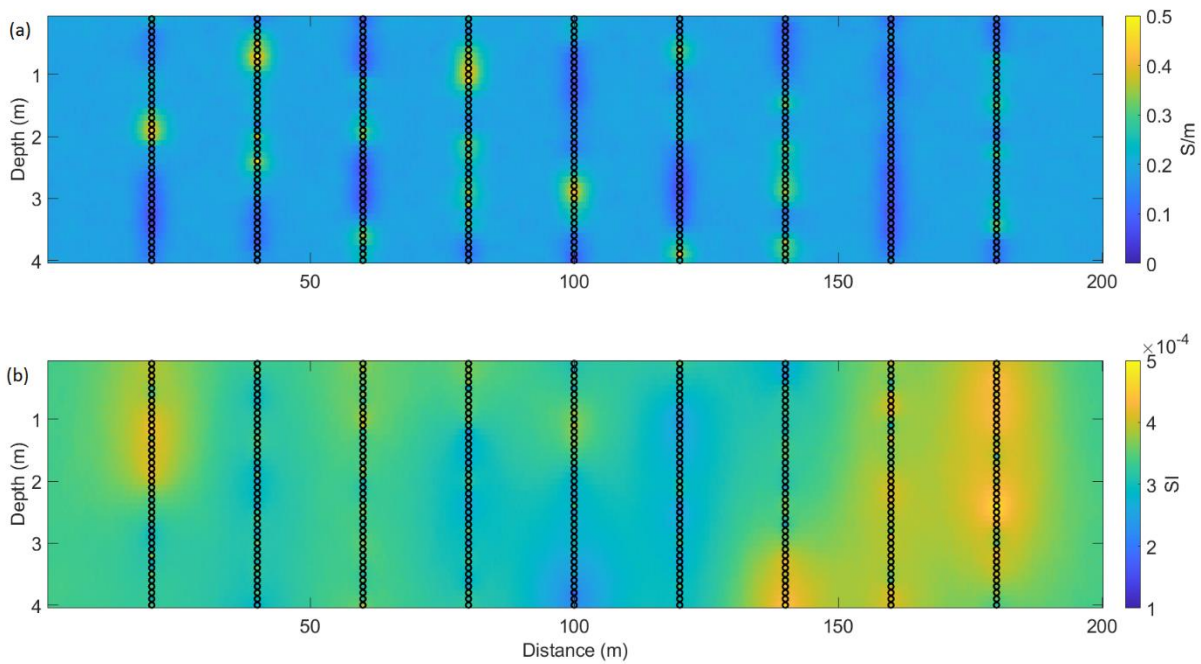


Figure 3.4 Cross sections extracted from the pointwise average model of the prior ensemble for (a) EC and (b) MS.

The same effect is observed from the pointwise variance models computed from this ensemble of models. As the stochastic sequential simulations are locally constrained by the borehole data, the variance is null at these locations (Figure 3.5). The synthetic response of the prior ensemble used in both inversion methods is shown in Figure 3.6. As the prior realizations were generated with geostatistical simulation, and we do not consider noise in this synthetic application, the synthetic responses calculated from the prior models do exactly reproduce the true measurement responses for the borehole locations. For other locations, the synthetic response of the pointwise average EC and MS models fails to reproduce the observed data.

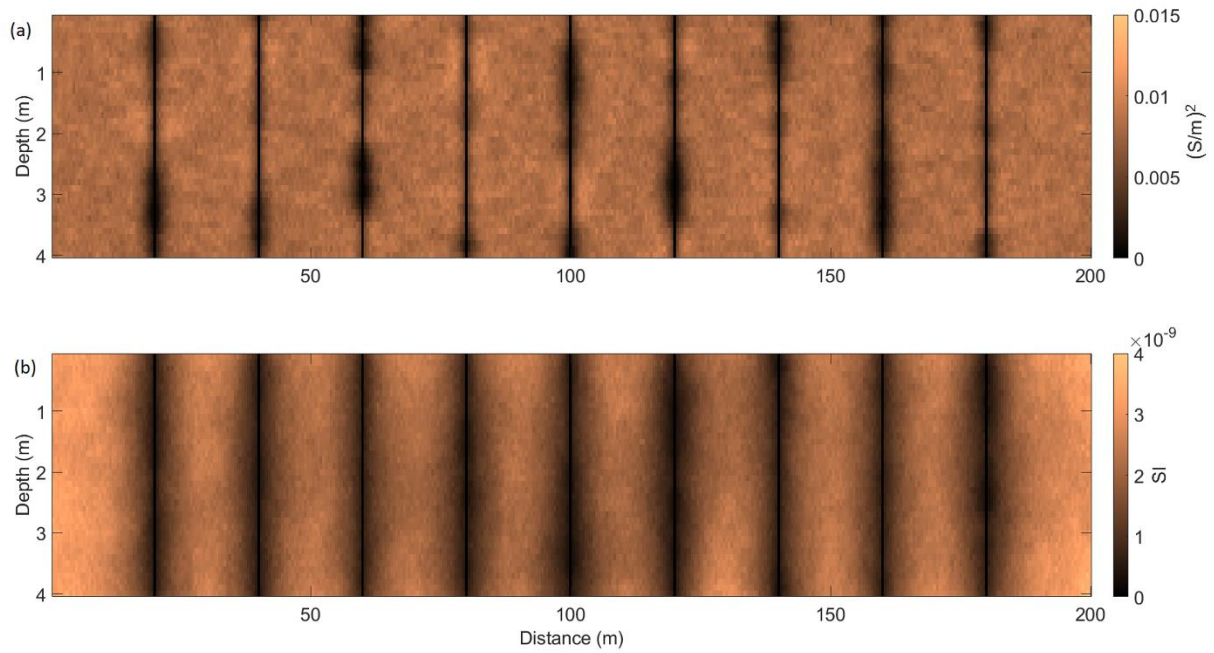


Figure 3.5 Cross sections from the variance of the prior ensemble for (a) EC and (b) MS.

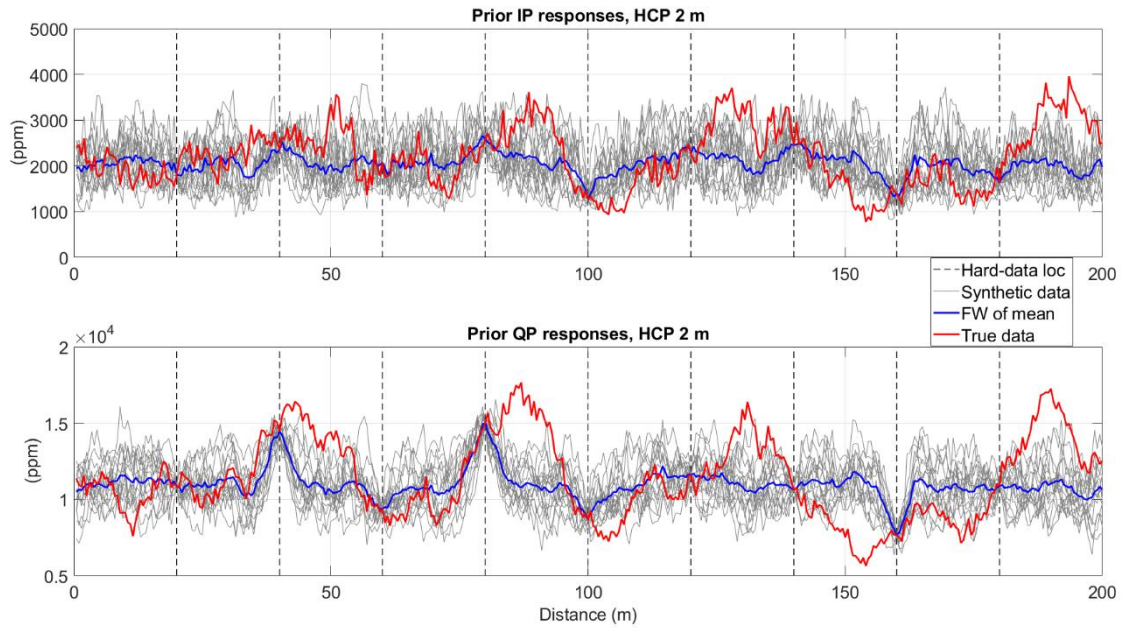


Figure 3.6 Prior synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively), compared with reference IP and QP data (red).

3.3.3 Results

The mismatch between the observed and predicted IP and QP data can be assessed in Figures 3.7 and 3.8, along the same 2D profile shown in Figure 3.3. For the sake of simplicity, is shown the results obtained for the 2 m HCP coil configuration. For the other coil configurations, similar results regarding the convergency of the data were obtained. Figures

3.7 and 3.8 show the synthetic response obtained from all the models of the posterior distribution predicted with the KEG and the ensemble of models cosimulated in the last iteration of the GEMI.

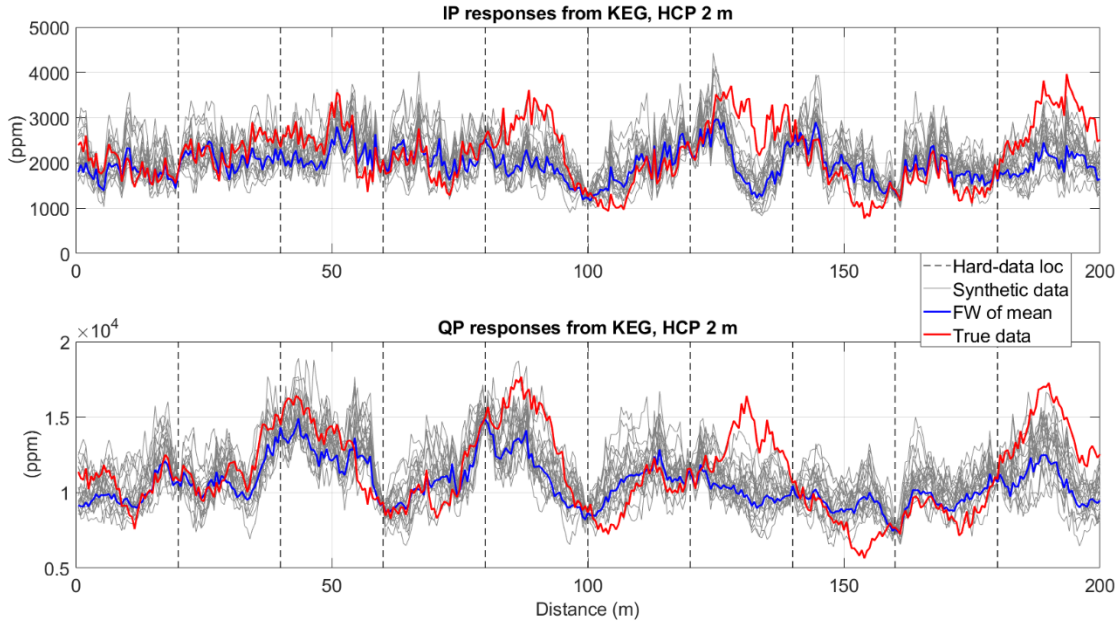


Figure 3.7 Synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively) calculated for the KEG, compared with reference IP and QP data (red).

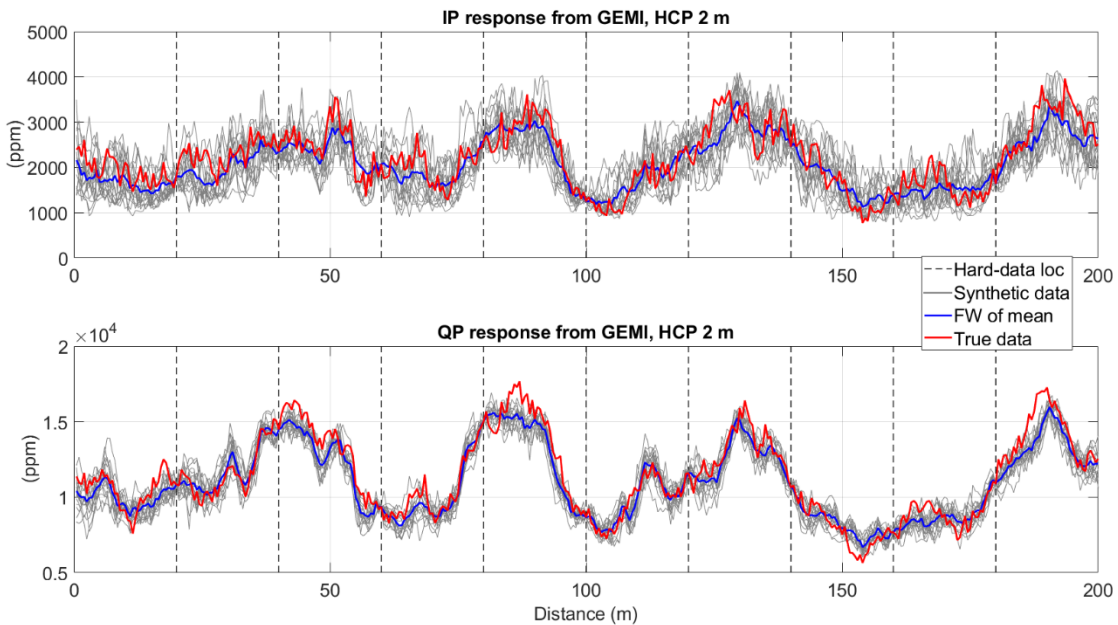


Figure 3.8 Synthetic IP and QP data for the 2 m HCP coil configuration, per model realization and the mean of the model realizations (grey and blue, respectively) calculated for the GEMI method, compared with reference IP and QP data (red).

The IP and QP responses predicted by the GEMI have a closer match than those predicted with the KEG concerning two aspects: (1) the predicted IP and QP calculated from the pointwise average model of EC and MS and (2) the uncertainty envelope, as represented by the synthetic response of all the realizations sampled from the posterior distributions (i.e., the grey lines in Figures 3.7 and 3.8), of the GEMI is tighter and better encapsulates the observed IP and QP data. Figure 3.9 summarizes the match between the true and predicted data, which supports the interpretation provided in (1).

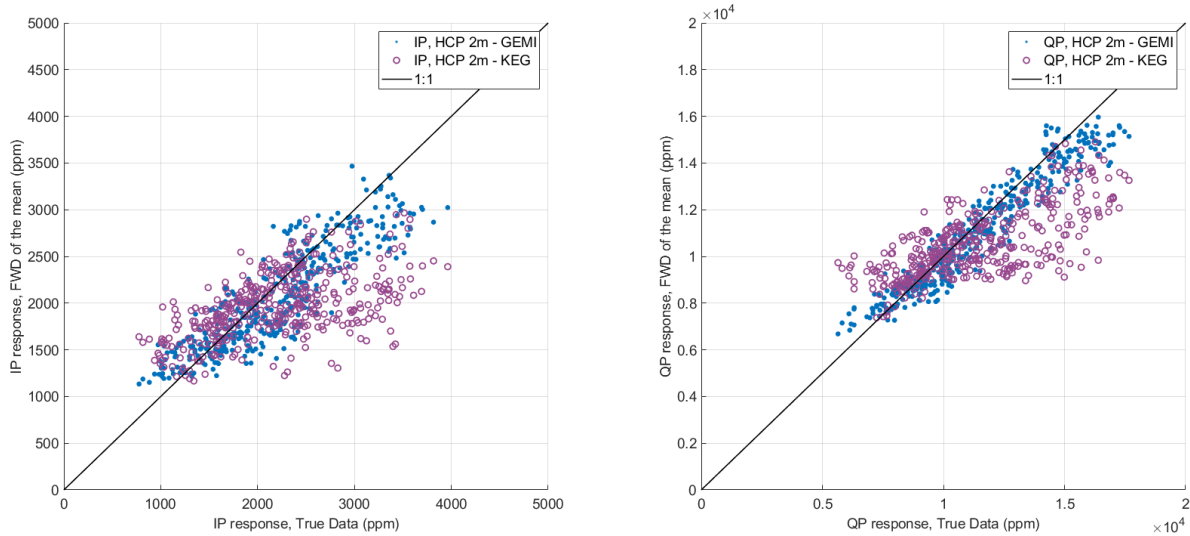


Figure 3.9 Biplots between the reference IP and QP data and the synthetic IP and QP data for the 2 m HCP coil configuration, calculated from the mean of the model realizations, for each method.

To compare the performance of the two methods with respect to parameter model reproduction, Figure 3.10 shows the posterior histograms of EC and MS of a single realization for the two methods. Both methods sufficiently reproduce the overall shape of the borehole data histograms. However, although the GEMI reproduces exactly the minimum and maximum values of the true distribution, the KEG predicts values not observed in the direct measurements and is characterized by longer tails in the predicted distributions. This difference is expected as by definition the stochastic sequential simulation method applied by GEMI uses a model perturbation technique which ensures the exact reproduction of extreme values. In addition, due to the Gaussian assumption of the KEG, the retrieved histograms from the posterior models are more Gaussian than those obtained from the GEMI.

Figures 3.11 and 3.12 present the comparison between the predicted models of both inversion methods using the posterior distribution of EC and MS from one single realization of each inversion method, which remains the same throughout the assessing analysis. Both models predict similar spatial patterns at the large scale for EC and MS but exhibit differences at the small scale and in the deeper part of the model. Consistently with what was observed for the

comparison of the histograms (Figure 3.10), the KEG predicts more pronounced local extremes and the GEMI predicts more accurately local distributions and spatial pattern.

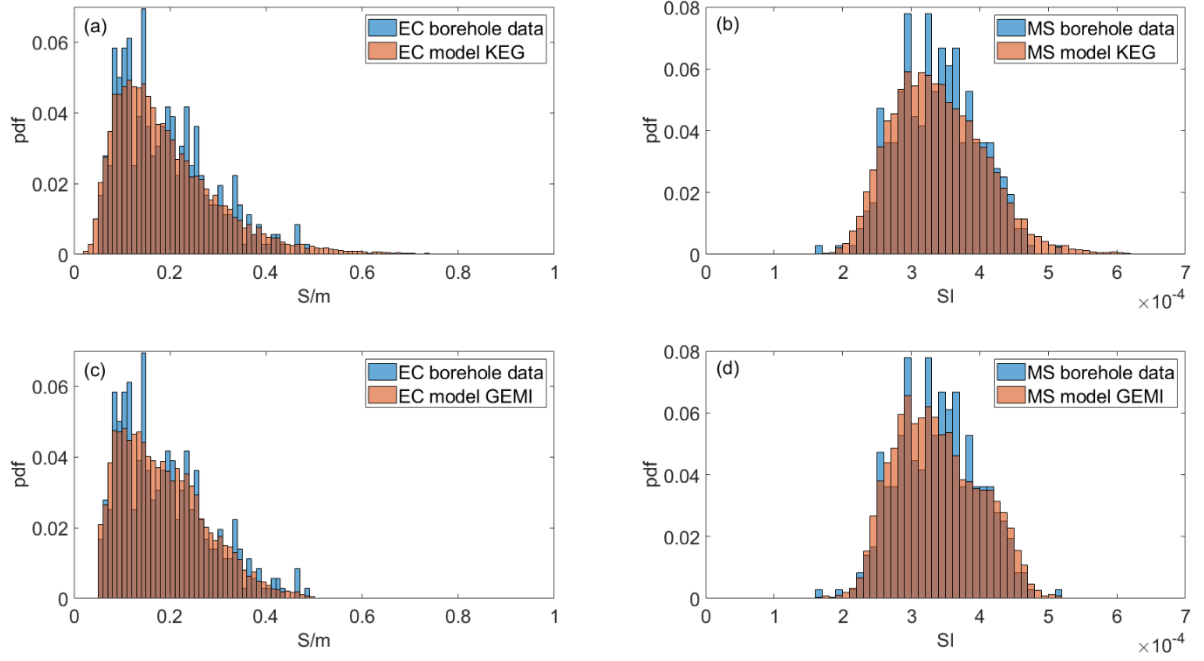


Figure 3.10 Comparison between the histograms of the reference EC and a single realization generated by (a) the KEG and (c) the GEMI. Comparison between the histograms of the reference MS and a single realization generated by (b) the KEG and (d) the GEMI.

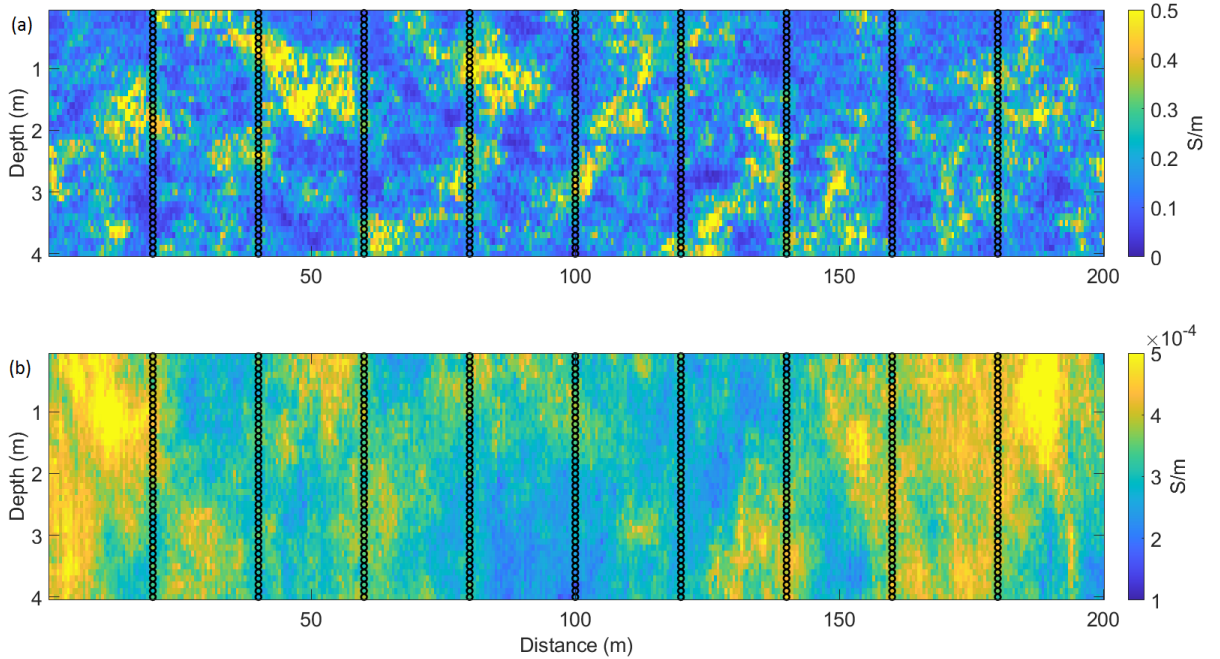


Figure 3.11 Vertical 2D section extracted from a realization of the posterior distribution obtained by the KEG for (a) EC and (b) MS.

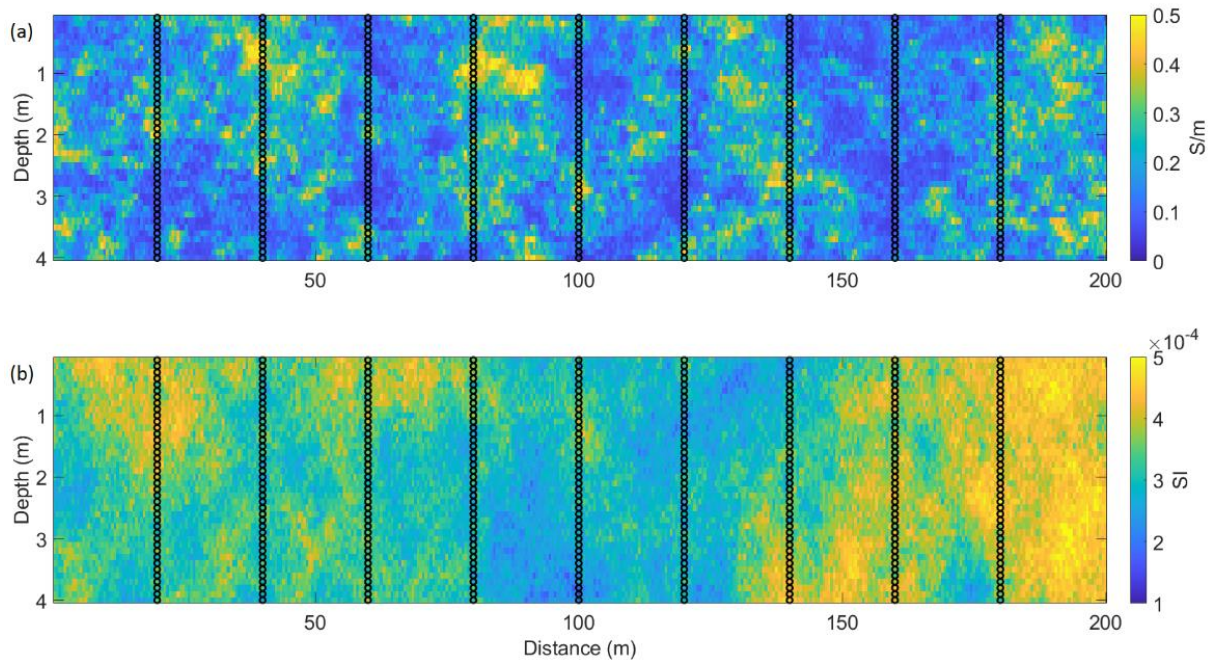


Figure 3.12 Vertical 2D section extracted from a realization of the posterior distribution obtained by the GEMI for (a) EC and (b) MS.

As an absolute comparison between single realizations is difficult and there is no equivalency between realizations sampled from the GEMI and the KEG, Figures 3.13 and 3.14 presents the pointwise average models computed from all the realizations from the posterior distribution predicted by both methods, with relatively large differences between the pointwise average models.

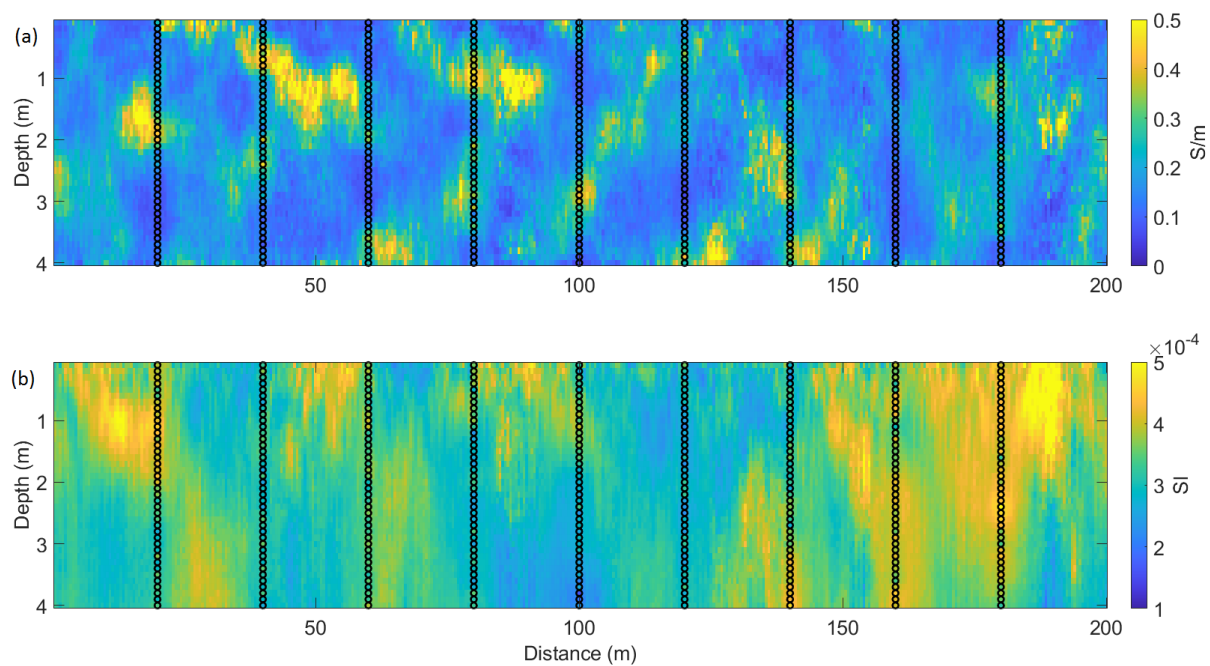


Figure 3.13 Two-dimensional mean model of (a) EC and (b) MS using the KEG method.

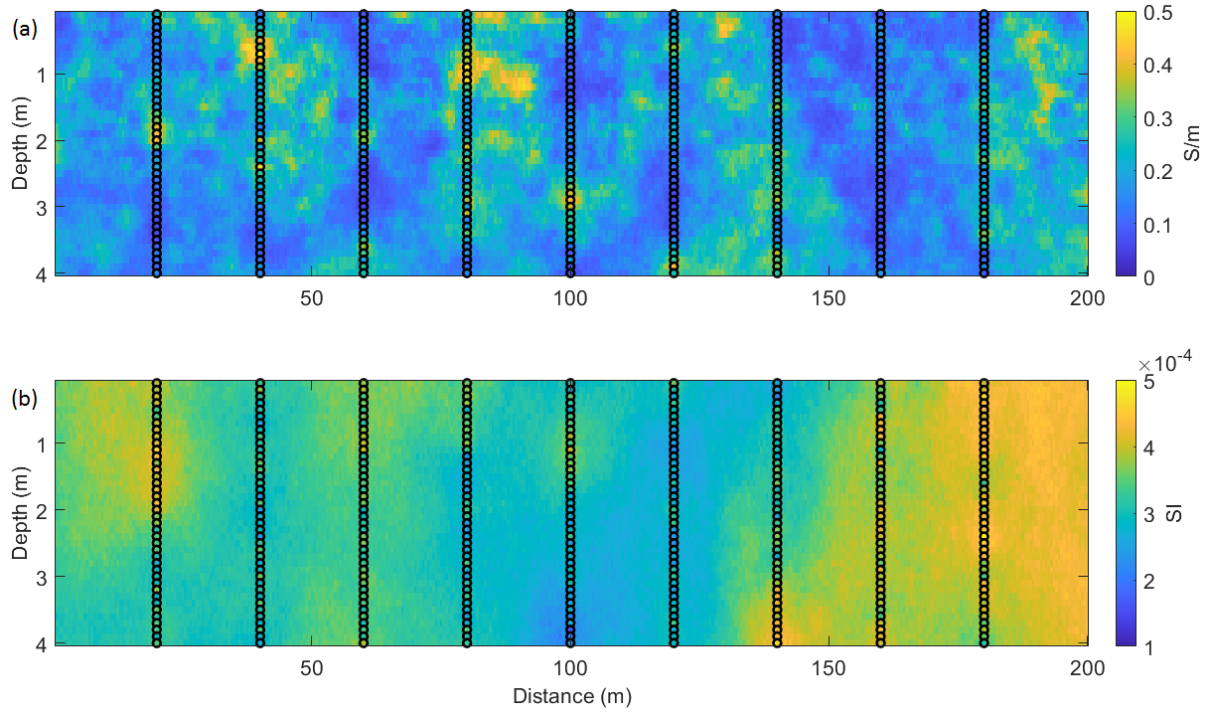


Figure 3.14 Two-dimensional mean model of (a) EC and (b) MS using the GEMI method.

The differences are further explored by comparing the residuals (i.e., the absolute differences between a single realization or the pointwise mean model of the ensemble realizations on the one hand and the reference EC and MS models on the other) in terms of their spatial distribution over the selected profile (Figures 3.15, 3.16, 3.17, and 3.18) and their global distribution based on the histogram (Figure 3.19).

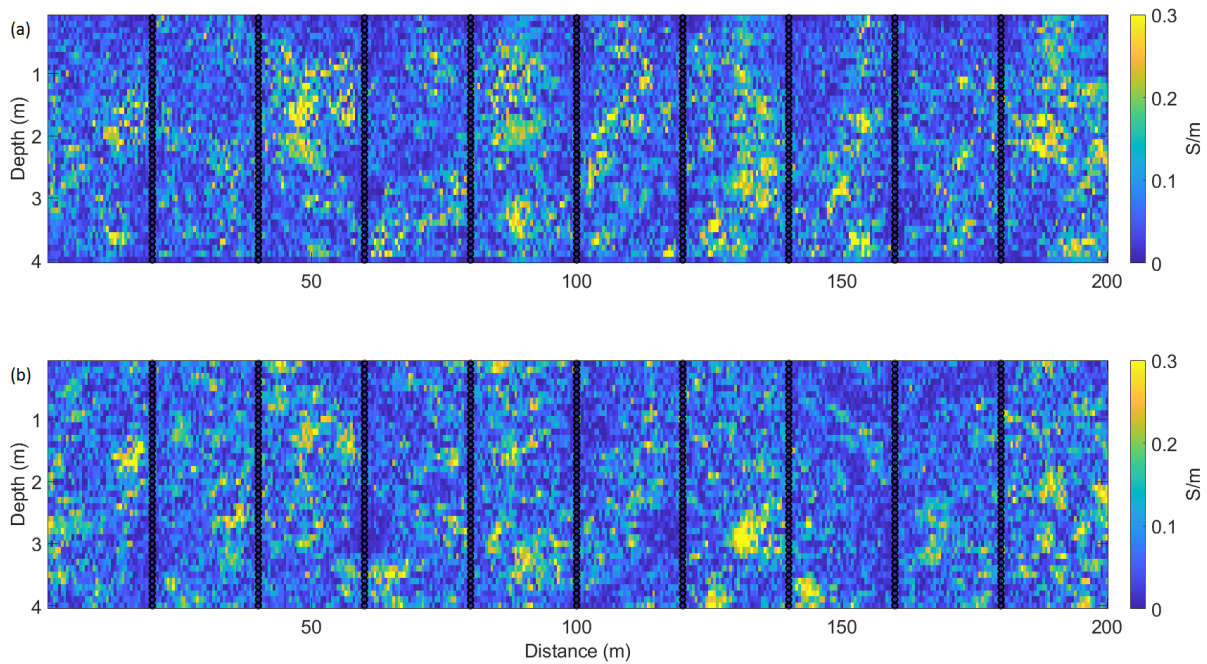


Figure 3.15 Two-dimensional difference between one model computed from the posterior realizations and the reference model of EC using (a) KEG and (b) GEMI.

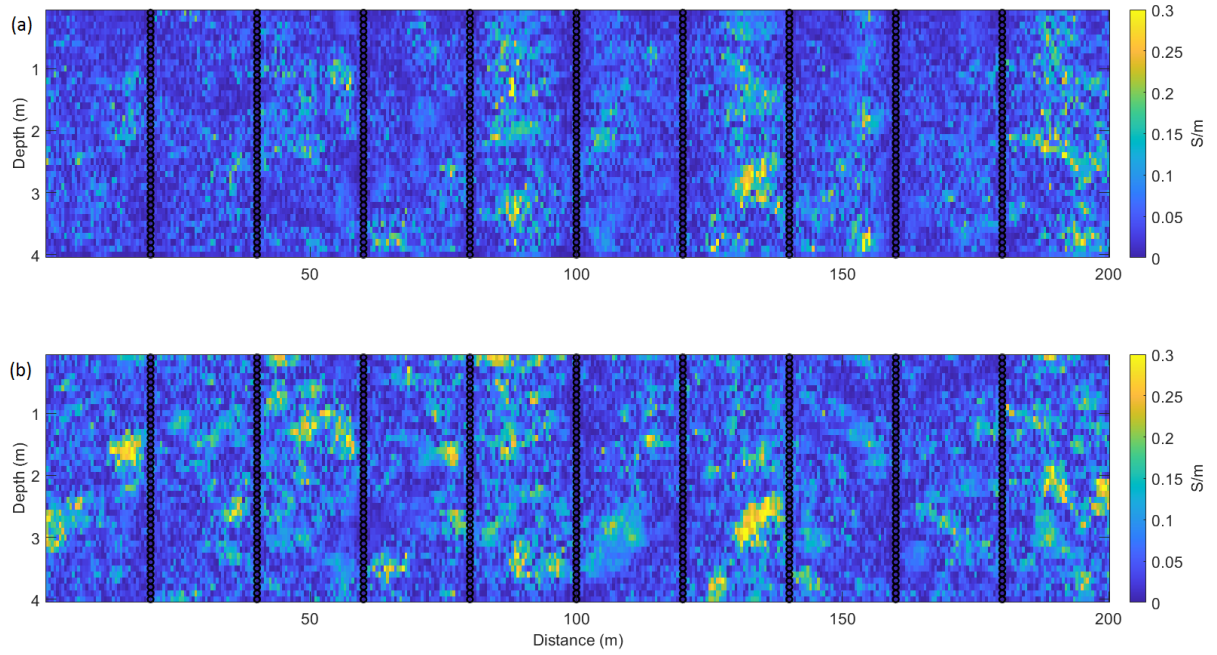


Figure 3.16 Two-dimensional difference between the mean model computed from the posterior realizations and the reference model of EC using (a) KEG and (b) GEMI.

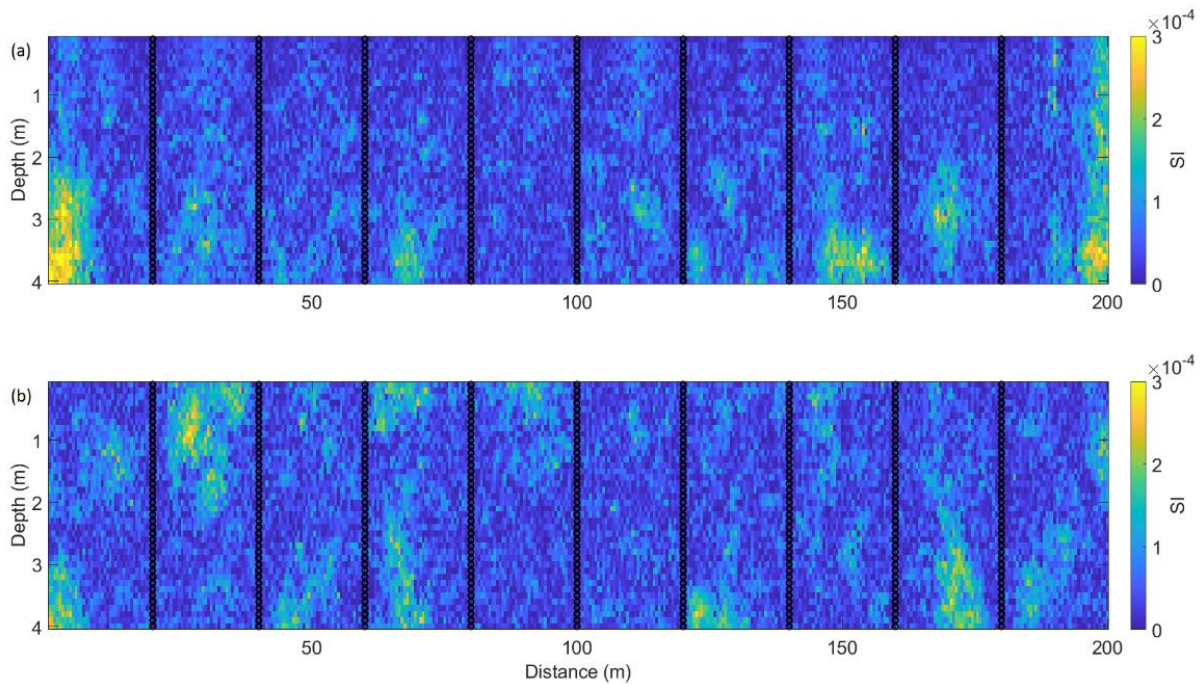


Figure 3.17 Two-dimensional difference between one model computed from the posterior realizations and the reference model of MS using (a) KEG and (b) GEMI.

In addition, Figures 3.20 and 3.21 show the biplots between the reference EC and MS models and the simulated EC and MS models. The profiles of the residuals of EC show a spatial structure in both methods, with low residuals where the reference model has low values and higher residuals where high reference values of EC occur (Figures 3.15 and 3.16). In line with the above, the histograms of the residuals EC for a single model realization are similar for both

methods but with the KEG predicting larger extremes (which can be larger than the maxima in the reference models). This effect is illustrated in the biplots between the reference EC model and the predicted EC models (Figure 3.20), and a tail of high residual values (Figure 3.19).

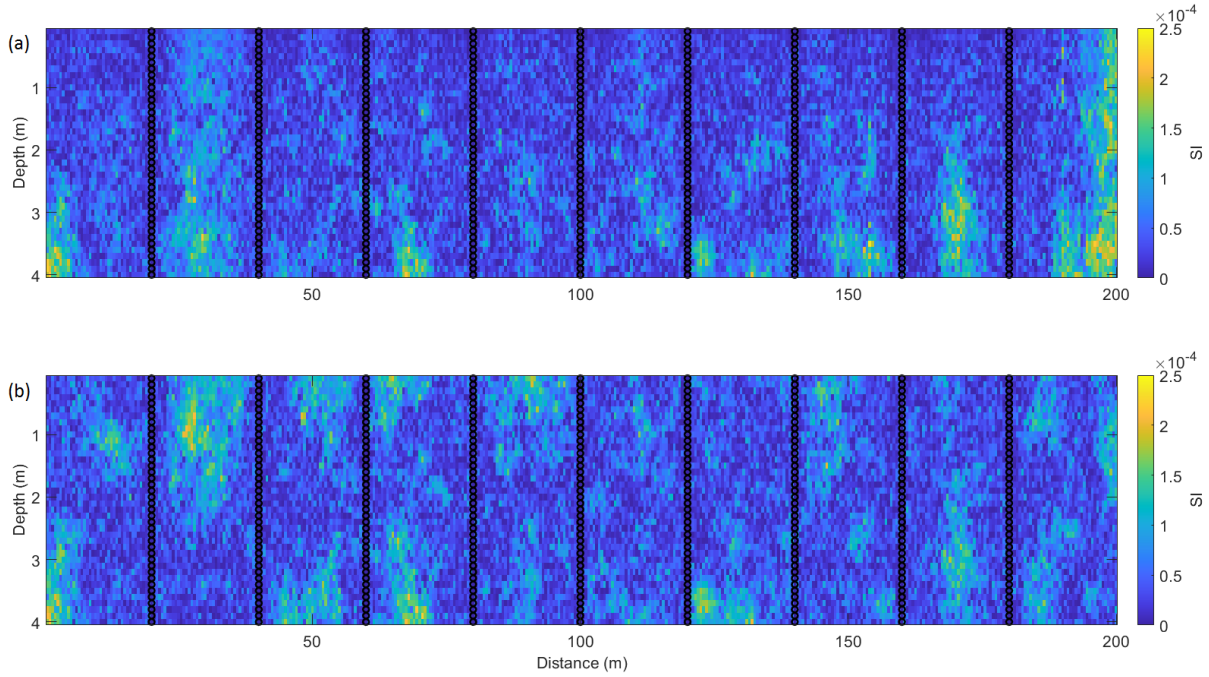


Figure 3.18 Two-dimensional difference between the mean model computed from the posterior realizations and the reference model of MS using (a) KEG and (b) GEMI.

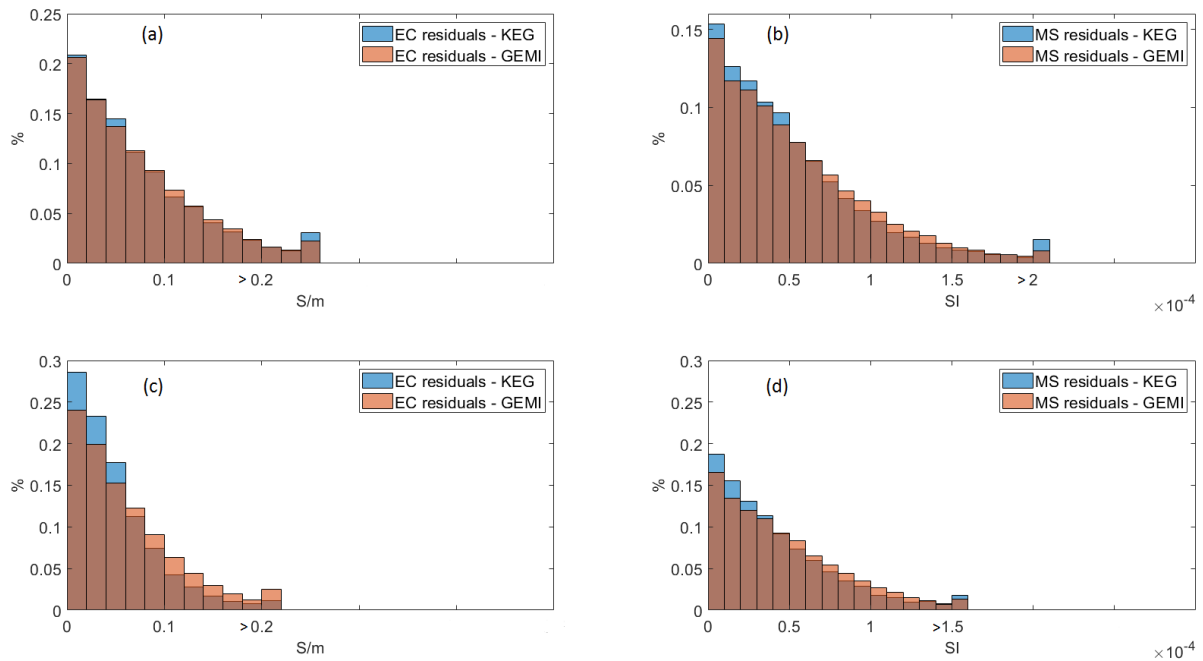


Figure 3.19 Comparison between the histograms of the residuals between one model computed from the posterior realizations and the reference model of (a) EC and (b) MS, and the residuals between the mean model computed from the posterior realizations and the reference model of (c) EC and (d) MS, using the KEG and GEMI.

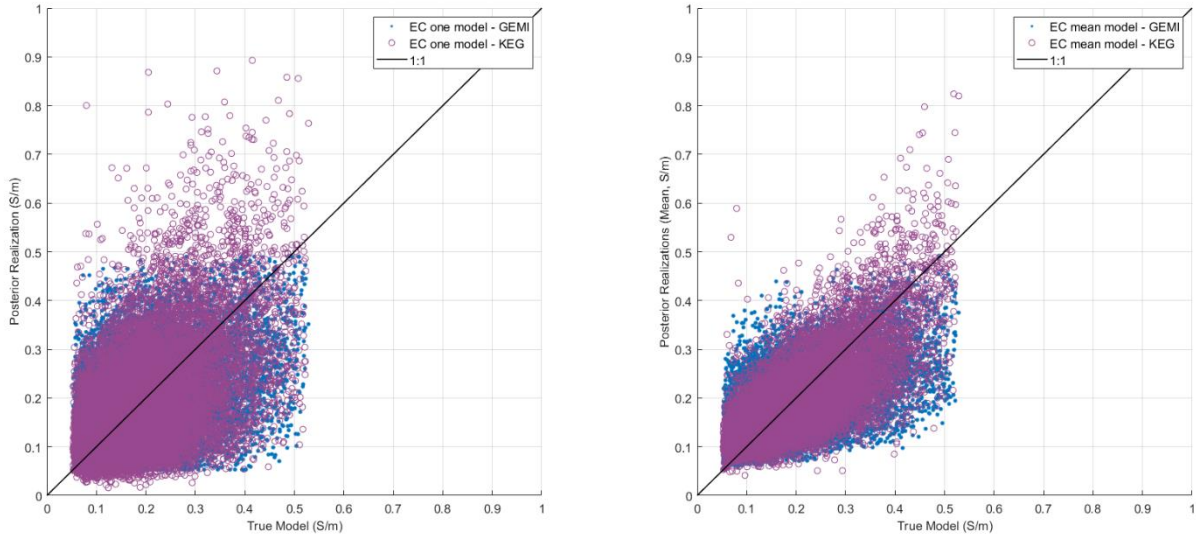


Figure 3.20 Biplots between the reference model of EC and (a) one model and (b) the mean model computed from the posterior realizations, from each method.

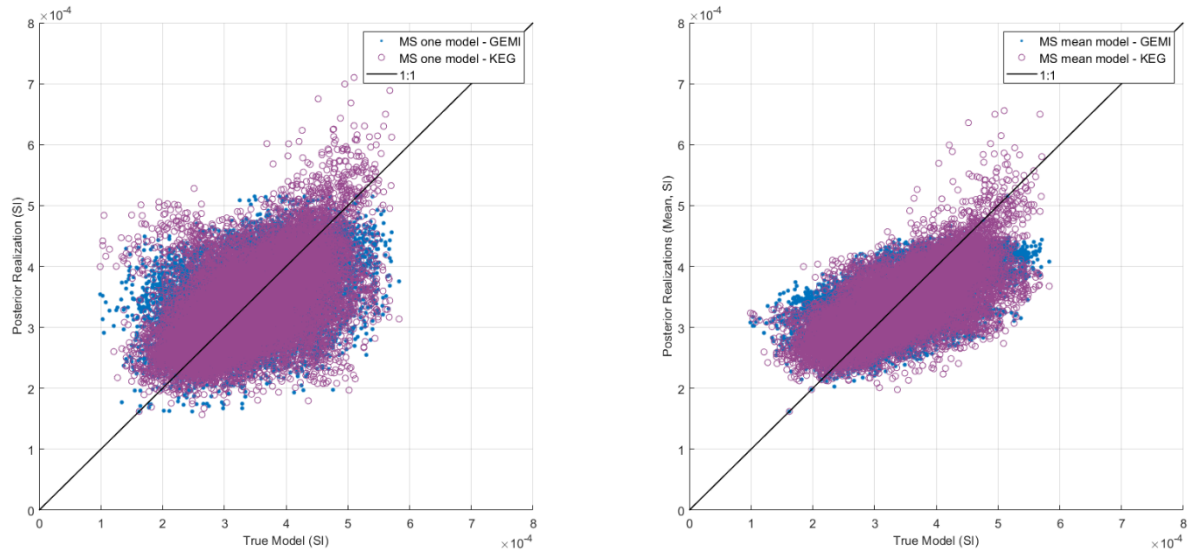


Figure 3.21 Biplots between the reference model of MS and (a) one model and (b) the mean model computed from the posterior realizations, from each method.

The profiles of the residuals of MS realizations of both methods show a low spatial structure correlation when comparing with the reference MS model (Figures 3.17 and 3.18), and the histograms and the biplots show similarities between KEG and GEMI, with the KEG predicting larger extremes (Figures 3.19 and 3.21), yet the distribution has a narrower peak for KEG, so small residuals are more frequently occurring than in the distribution obtained for the GEMI, but KEG again exhibits a tail of high residual values visible in the histograms and the biplots (Figures 3.19 and 3.21). These differences are due to the nature of the stochastic update of both inversion methods; because the KEG uses Gaussian statistics, the probability for EC and MS values beyond the prior ensemble values derived from the boreholes data is never zero. When comparing the mean of the ensemble residuals (Figures 3.16 and 3.18), i.e., absolute

differences between the mean models (Figures 3.13 and 3.14) and the reference EC and MS models (Figure 3.3), the same spatial structure in both methods is shown but the mean of the 500 realizations models of KEG reproduce better the reference model of EC and exhibit large residuals less frequently when compared with GEMI, which can derive from further exploration by GEMI of the model parameter space. However, it should be mentioned here that for the GEMI method the mean residuals are computed based on 32 model realizations only. Regarding the residuals of the average MS models, the same conclusion arises as that reached when only one MS realization of both methods is used, as described previously.

3.4 Discussion

Both stochastic inversion methods use computed FDEM signals to simultaneously predict the spatial distribution of the subsurface EC and MS models. The GEMI method does not assume any parametric distribution of the properties of interest and uses stochastic sequential simulation and co-simulation in a convergent and iterative procedure. However, it requires the existence of borehole log data of EC and MS, which might not always be available. The KEG method uses the update step of the EnKf to update a prior ensemble assimilating the geophysical measurement data.

The synthetic FDEM data generated from the inversion results encapsulate the reference observed FDEM data for both methods, but the GEMI method more closely resembles the reference observed QP and IP data along the cross section when comparing with the synthetic FDEM data calculated from the model realizations. For the GEMI method, 6.25% of the reference IP data and 21.75% of the QP data along the cross section were outside the range of the computed synthetic data, whereas for the KEG method we found 27.5% of the IP data and 24.75% of the QP data to be outside (Figures 3.7 and 3.8). Both methods predict EC and MS models that match the main spatial structures of the reference ones (Figures 3.11 and 3.12). However, as already mentioned previously, there are differences between the spatial continuity pattern of the predicted models with both methods. These differences can be assessed by modelling the horizontal and vertical variograms from a single realization sampled from the predicted posterior distribution and comparing them with the true variogram models and those obtained from the borehole data exclusively (Figures 3.22 and 3.23).

The set of borehole data for EC captures the spatial continuity pattern of the reference EC model in the horizontal and vertical directions (Figure 3.22). The best-fit inverted model predicted with GEMI exhibits a similar spatial behavior in terms of variogram range, sill, and nugget effect when compared with the reference models for both directions. This reproduction is expected as by definition stochastic sequential simulation and co-simulation methods, which

are used as model perturbation and stochastic update, do ensure the reproduction of the variogram models imposed. However, to ensure the match with the observed FDEM data, the iterative inversion procedure does need to perturb the imposed variogram range in the vertical direction. This results in the range of the vertical variogram obtained from the best-fit inverted model being smaller than the true one. On the other hand, the variogram models obtained from the inverted models with the KEG have higher sill (i.e., variance), shorter ranges, and larger variability at the small scale as represented by the nugget effect. The higher sill and nugget effect are consistent with the occurrence of larger extremes as discussed previously (Figures 3.10 and 3.19). This behavior is expected because the KEG update at a given location is independent of the updated values at the neighboring locations. The larger nugget effect and smaller variogram ranges result from this update procedure, which does not account for spatial correlation. In addition, unlike the more traditional inversion Bayesian update methodologies, which are done on an average model (i.e., a priori model), this KEG update is done on simulated realizations, which further enhances the nugget effect of resulting models.

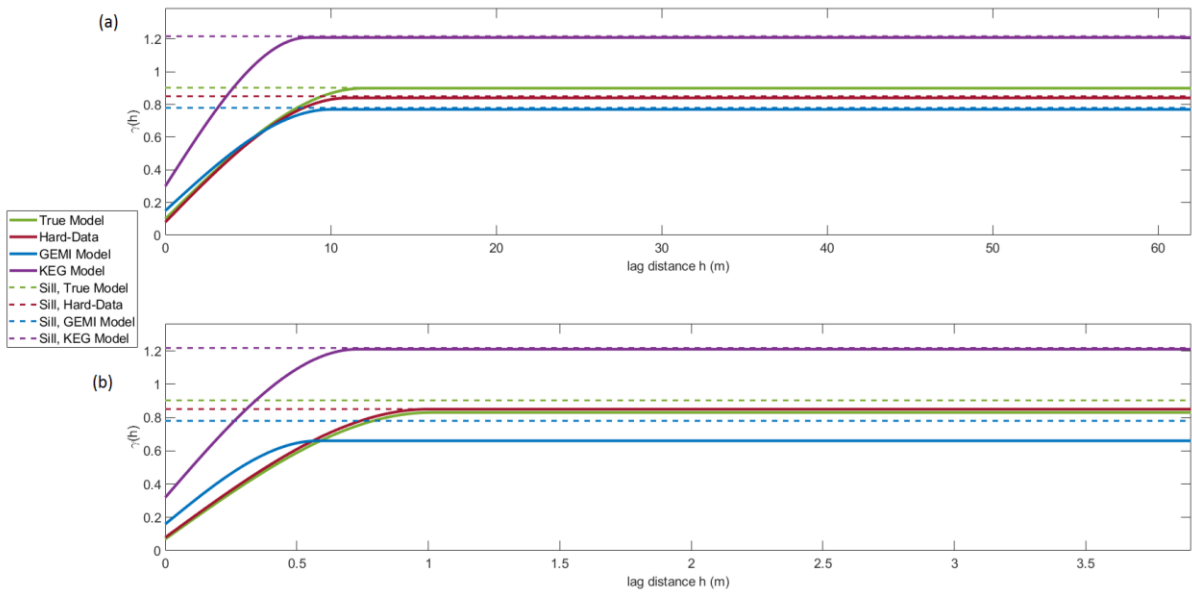


Figure 3.22 (a) Horizontal and (b) vertical variogram models for EC calculated from the true models, the borehole data and one realization from the posterior ensemble predicted with KEG and with GEMI.

Contrary to EC, for MS the sparse borehole data set is not able to reproduce the true variogram model of the reference 3D models of EC and MS in the horizontal and vertical directions (Figure 3.23). This behavior represents an obstacle for the GEMI because we use the variogram model retrieved from the borehole data in the stochastic sequential simulation and co-simulation of MS. If the a priori data are not similar to the reference model, this inversion method will struggle to predict the true spatial pattern. Azevedo and Demyanov (2019) propose a stochastic optimization approach to account for the uncertainty in variogram models. Figures 3.19, 3.20,

and 3.21 illustrate the potential of the KEG to move away from the spatial continuity pattern of the a priori ensemble of models toward the true one, which indicates that the KEG appears to be more flexible in the presence of non-exact prior information.

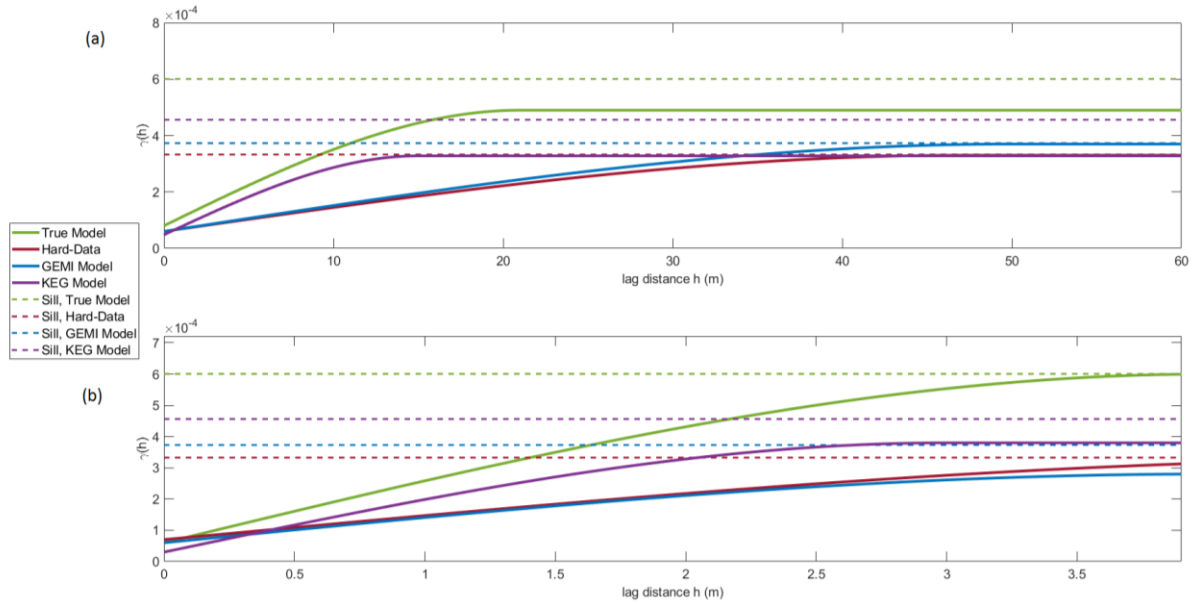


Figure 3.23 (a) Horizontal and (b) vertical variogram models for MS calculated from the true models, the borehole data and one realization from the posterior ensemble predicted with KEG and with GEMI.

The synthetic application example shown in this Chapter uses a large number of boreholes with a regular and small spacing relative to the range of the underlying variogram. This setup makes the simulated models (i.e., the a priori ensemble of models) very dependent on the conditioning borehole data. This is a very appropriate situation for updating by KEG because the simulated models are already close to the reference values. Reducing the number of boreholes might result in considerable differences by both methods considered herein.

In addition, the same forward model was used to generate the reference FDEM data and within the inversion procedures. This assumption implies that at the borehole locations there is no uncertainty. In a real case study, when one applies the KEG update of a value at the borehole location, the IP and QP values, obtained by the forward model, are not equal to the real values at borehole locations. Hence, in real applications, the updated EC and MS values of the final models will not honor the experimental values.

Finally, is presented the comparison of the spatial uncertainty assessed with both methods by computing the pointwise variance of the ensemble of models of EC and MS predicted with the KEG and the GEMI (Figures 3.24 and 3.25). Both methods honor exactly the borehole data, illustrated with null variance at these locations. However, the spatial behavior of both methods is quite different. For EC, in the models obtained with the GEMI, there is a clear influence of

the depth of sensitivity ($sens_{EC}$), whereas the predictions with the KEG have smaller variance in the first half-meter (Figure 3.24).

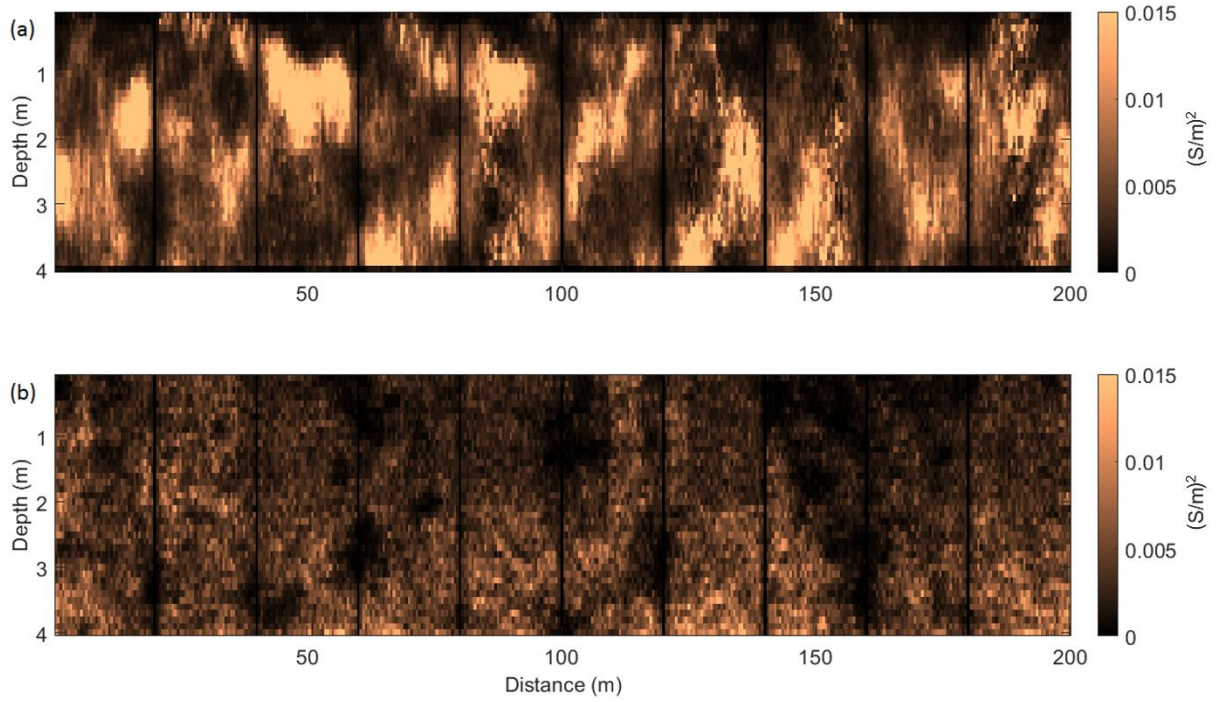


Figure 3.24 Pointwise variance model of EC obtained from the predicted models with (a) KEG and (b) GEMI.

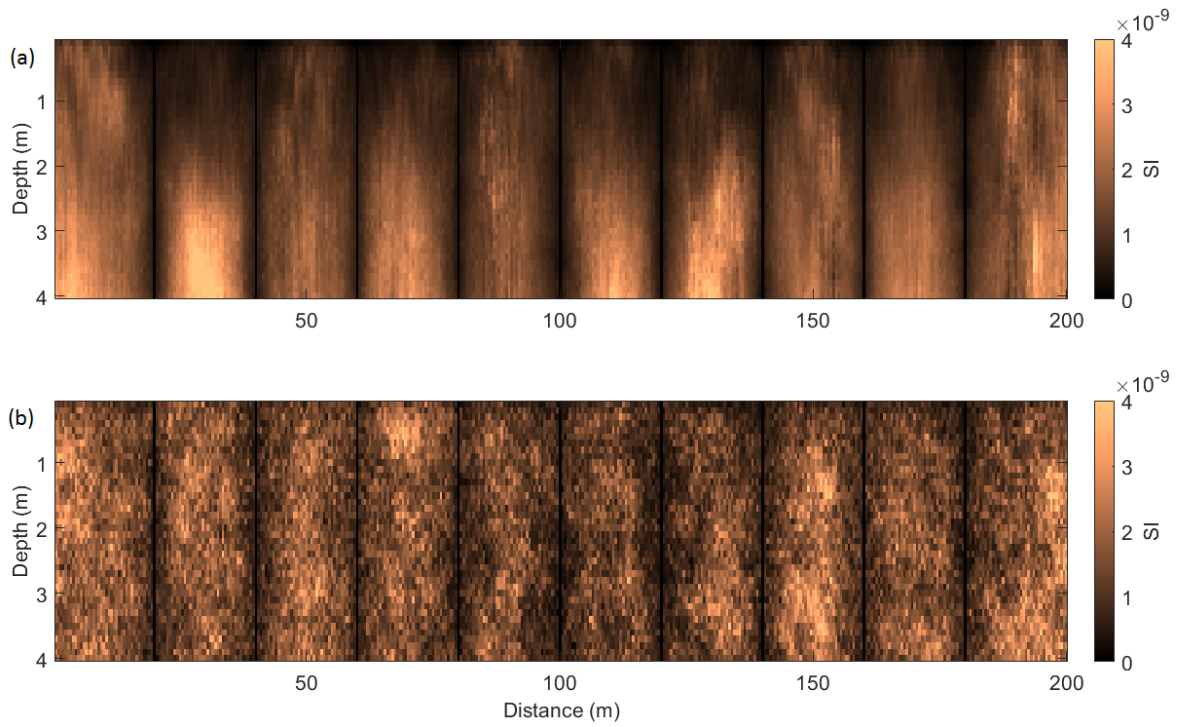


Figure 3.25 Pointwise variance model of MS obtained from the predicted models with (a) KEG and (b) GEMI.

For KEG, there seems to be a more pronounced proportionality effect, i.e., higher variance associated with higher predicted values of EC. This behavior can be explained by visual inspection of the prior model. Except in very close vicinity of the boreholes, the prior model's mean and variance EC are largely uniform. For MS, the prior model's mean is less uniform because larger correlation lengths are defined a priori and extreme values from the virtual boreholes have a larger influence in the prior model mean. For the uncertainty associated with the predicted MS, the KEG has a clear depth dependency, whereas the spatial uncertainty assessed with the GEMI appears to be unstructured and uninformative (Figure 3.25).

Regarding computational time performance, the KEG inversion method is more efficient because it is a single-step update method, and there is no need to compute the sensitivity for any of the ensemble's subsurface prior model realizations. The KEG typically works with larger ensemble sizes sampled from a prior model (500 realizations) as compared with the relatively low number of model realizations (32 realizations) considered in the individual iterations of the GEMI. However, the GEMI updates do take sensitivity into account, involving additional forward model computations necessary for the sensitivity derivation.

The GEMI inversion was performed on a workstation with Intel Core i7 3.40 GHz CPU and 16 GB RAM. The total computational time using the GEMI inversion procedure (32×6 realization models of EC and MS) was 21 h. The KEG inversion was performed on a personal computer with Intel Core i5 1.9 GHz and 8 GB RAM. The KEG inversion took 152 min, plus one additional hour for the prior ensemble simulation.

Both stochastic inversion methods considered are based on a 1D forward model. The same approximation was used to generate the synthetic data used in the application example shown herein. A 1D forward model represents a simplification of the complex 3D subsurface field propagation. In addition, by using the same forward model to build the synthetic and as part of the inversion methods, was not accounted the uncertainty related to numerical approximations. Three-dimensional forward models are required for complex subsurface geology (i.e., geology with significant 3D structure) because 1D approaches are not able to properly capture the complex subsurface field propagation and predict reliable inverse models, increasing the computational costs of both inversion methods. However, the main conclusions draw from the application examples shown in this Chapter would hold as the a priori information related to the spatial continuity patterns of EC and MS is not changed.

3.5 Conclusion

In this Chapter is presented a comparison between two different statistical-based FDEM inversion methods: the KEG and GEMI method. These inversion methods are applied to a freely available 3D synthetic data set rendering realistic spatial distributions of EC and MS.

The assessment of both methods was developed by comparing the predictions of EC and MS using a single realization selected from the posterior distribution and the pointwise mean and variance models computed from the posterior. The KEG frequently predicts values of EC and MS beyond the minimum and maximum values observed in the prior ensemble derived from the borehole data, whereas the GEMI reproduces exactly the histograms retrieved from the borehole data. This effect also is observed locally when computing the residuals between the reference and the predicted EC and MS models. Similarly, the GEMI tends to reproduce the variogram models imposed during the stochastic sequential simulation and co-simulation of EC and MS, whereas the KEG has larger degree of freedom to perturb the spatial continuity pattern. The latter might be important when the knowledge about the spatial distribution of the phenomena to be modeled is largely unknown.

Overall, both methods succeed in simultaneously reproducing subsurface EC and MS from FDEM data, but KEG results in more accurate reconstruction of the MS, considering that smaller residuals are more frequent. Overall, by comparing the histograms of the residuals of the pointwise average models, the KEG shows more smaller residuals more frequently. However, for the MS predictions, the GEMI method delivers very large residuals less frequently than the KEG. Another advantage of the GEMI method shows when an FDEM measurement response is computed from the resulting EC and MS subsurface models. The FDEM response derived from the GEMI subsurface model is much closer to the reference data than its equivalent for the KEG method.

The comparison of the inversion's uncertainty assessment is somewhat difficult because KEG and GEMI rely on different concepts regarding the interpretation of inverse model uncertainty. Because the uncertainty in the measurement data is uniform for the entire data set, the KEG shows larger uncertainty where the posterior model is more different from the prior model. For the GEMI, a link between the shape of measurement sensitivity and inverse model variance is partially observed.

Data associated with this research are available and can be accessed via the following URL: <http://doi.org/10.5281/zenodo.5116420> .

CHAPTER 4

Geostatistical Inversion of FDEM

with Randomized Tensor

Decomposition

The research presented in this chapter was published on:

Liu, M., Narciso, J.*, Grana, D., Van De Vijver, E., and L. Azevedo, (2023), Frequency-domain electromagnetic induction for the prediction of electrical conductivity and magnetic susceptibility using geostatistical inversion and randomized tensor decomposition: *Geophysics*, 88(6), 1-13.

High-resolution characterization of near-surface systems is crucial for a variety of subsurface applications. Frequency-domain electromagnetic induction (FDEM) has been widely used in near-surface characterization when compared with other geophysical methods due to its flexibility in acquisition and the ability to survey large areas with high-resolution but with relatively low costs. FDEM measurements are sensitive to subsurface electrical conductivity (EC) and magnetic susceptibility (MS). However, the prediction of these properties requires solving a geophysical inverse problem. This work combines ensemble smoother with multiple data assimilation (ES-MDA) and model re-parameterization via randomized tensor decomposition (RTD) to simultaneously predict electrical conductivity and magnetic susceptibility from measured FDEM data. ES-MDA is an iterative data assimilation method, which can be applied to nonlinear forward operators and provides multiple posterior realizations conditioned on the geophysical measurements to evaluate the model uncertainty. However, its application is usually computationally prohibitive for large-scale three-dimensional problems. To overcome this limitation, the model parameters are reduced using RTD to perform the inversion in the low-dimensional model space. The method is applied to synthetic and noisy real data sets. In the synthetic application example, the predicted posterior realizations illustrate the ability of the proposed method to recover the true models of EC and MS accurately. The real case application comprises FDEM data acquired over an arable land characterized by quaternary siliciclastic deposits with geoarchaeological features. The performance of the inversion method is assessed at a borehole location not used to constrain the inversion. The inverted models do capture the available log data, illustrating the applicability of the inversion method to noisy real data.

4.1 Introduction

Detailed modelling and characterization of near-surface is key to several applications, such as sustainable development of soil studies, archaeology, and groundwater management (De Smedt et al., 2013; Delefortrie et al., 2014; Simon et al., 2015). This is a challenging task as the near-surface is often characterized by strongly heterogeneous geological properties as the result of complex interacting processes of both natural and anthropogenic origins, which act at different spatiotemporal scales (Morel and Heinrich, 2008).

Due to the complex nature and dynamics of these systems, its characterization using traditional interpolation methods of sparse and discrete direct observations (e.g., borehole

data) is not suitable to capture the full spatial variability of the system. Recently, there has been an increased interest in using geophysical data to characterize the near-surface. This is mainly due to the ability to acquire high spatial resolution data over large areas at relatively low-cost, the usability of the existing equipment in different types of terrain, and the ability to image subsurface properties (Everett, 2013) that can be interpreted in terms of geological and physical processes.

Among the most common geophysical methods, frequency-domain electromagnetic induction (FDEM) allows collecting high-resolution data sets timely and efficiently (Hanssens et al., 2020), by providing indirect measurements of two key near-surface properties: electrical conductivity (EC); and magnetic susceptibility (MS). From a simplistic perspective, EC relates mainly to soil salinity, texture, organic matter, moisture content, and bulk density (Doolittle and Brevik, 2014; Everett, 2013; Islam et al., 2014a; Islam et al., 2014b, Reynolds, 2011), while MS tends to be related to the mineralogy of the near-surface rocks, and anthropogenic features (Van De Vijver, 2017). Nevertheless, all these geological properties affect jointly EC and MS.

However, predicting the spatial distribution of EC and MS from the observed FDEM data requires solving a non-linear, ill-conditioned inverse problem with multiple solutions due to measurement errors and uncertainties in the model and observations (Tarantola, 2005), the band-limited nature and resolution of the FDEM data, noise and physical assumptions associated with the forward operators (Qiu et al., 2020). The recorded electromagnetic fields, the in-phase (IP) and quadrature-phase (QP) signal components, are related to EC and MS through a forward operator $\mathbf{F}$ that is mathematically described in Equation 2.1. The operator $\mathbf{F}$ is commonly approximated using 1D or 2D numerical models, due to prohibitive computational costs of three-dimensional forward models (Li et al., 2019).

Deterministic algorithms as well as stochastic sampling and optimization methods have been proposed to solve geophysical inverse problems (Tarantola, 2005). Among stochastic approaches the most commonly used methods are Markov chain Monte Carlo (MCMC) and ensemble-based methods. FDEM inversion methods for near-surface characterization are generally based on deterministic approaches. These methods have been used successfully applied to model the spatial distribution EC and MS in the near-surface (Deidda et al., 2017; Farquharson et al., 2003; Guillemoteau et al., 2016). However, deterministic inversion methods predict a single best-fit model and have limited capabilities for uncertainty assessment. Due to the non-uniqueness of the solution of the inversion problem, stochastic inversion methods are generally preferable. In the stochastic approach, the solution can be expressed as an ensemble of models that fit the data within a tolerance and whose variability represents the uncertainty of the solution, which can be used to make informed decisions and quantify risks.

Stochastic geophysical inverse methods under a Gaussian assumption comprise the so-called pilot points method that use sequential geostatistical resampling techniques (e.g., Mariethoz et al., 2010; Alcolea et al., 2010; Fu and Gomez-Hernandez, 2009; Hansen et al., 2012; Zahner et al., 2016; Jäggli et al., 2017), principal component geostatistical approach (Lee and Kitanidis, 2014), methods based on circulant embedding of the covariance matrix (e.g., Hansen et al., 2012; Laloy et al., 2015); and methods that allow for jointly inferring the spatial correlation model (i.e., mean and variogram) together with the two- and three-dimensional spatial distribution of the property field values of interest (Laloy et al., 2015; Hunziker et al., 2017; Wang et al., 2022).

The available literature includes statistical approaches to FDEM inversion, but these are limited to the prediction of EC (Moghadas and Vrugt, 2019) or rely on Gaussian assumptions for the distribution of EC and MS. Trans-dimensional Bayesian inversion of electromagnetic data and Markov chain Monte Carlo methods have been proposed in (Blatter et al., 2018; Minsley, 2011; Ray and Key, 2012). These approaches generally allow an accurate quantification of the posterior distribution; however, the computational cost of the sampling and optimization is generally unfeasible for large geophysical datasets. Ensemble-based methods, such as ensemble smoother and ensemble Kalman filter (Evensen, 2009), provide a reliable alternative to MCMC methods, by finding a compromise between model accuracy and computational cost. For example, the Kalman ensemble generator (KEG) method (Bobe et al., 2019) detailed in section 3.2.1, provides such statistical framework for FDEM inversion. Most publications on the application of machine learning in geophysical inverse problem adopt deep learning algorithms to approximate the forward model and reduce the problem dimension and the computational cost (e.g., Manassero et al., 2020; Puzyrev and Swidinsky, 2021; Qi et al., 2019) or use them directly to approximate the inverse function and replace deterministic inversion methods (Hashemian et al., 2021; Li et al., 2021). For example, (Manassero et al., 2020) propose a reduced-order approach for the inversion of electromagnetic data.

The techniques to reduce the computational time in high-dimensional probabilistic inverse problems, can be generally divided in three categories: (1) approximation of the forward operator (i.e., surrogate modelling), (2) dimensionality reduction of the model and/or data spaces by re-parameterization, and (3) approximating the posterior distribution by making certain assumptions about its probability distribution. The method presented in this Chapter explores the points (2) and (3) by combining stochastic inversion with dimensionality reduction techniques to perform the inversion in a lower dimensional space. We propose a stochastic nonlinear method based on the ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013) to invert the FDEM data for EC and MS. ES-MDA is a derivative-free optimization method that proves useful when the code of forward simulators is

inaccessible, or the sensitivity matrix is challenging to derive. Unlike linear Bayesian methods, ES-MDA does not require a linear approximation of the forward operator, making it advantageous in improving inversion results for non-linear cases. However, ES-MDA poses prohibitive computational costs in large geological models. To address this challenge, is proposed the use of a randomized tensor decomposition (RTD) to sparsely re-parameterize the models and update the model parameters in the low-dimensional space. RTD is a high-order linear reduction method that can recover spatial structures between multiple dimensions of geological models and track uncertainty propagation during model order reduction. Compared with deep-learning-based methods, RTD is easier to integrate into the inversion workflow and is not limited by computing devices.

The method presented in this Chapter is first applied in a synthetic two-dimensional data set to validate the results obtained and then in a three-dimensional real case application to assess its performance in data contaminated with noise. The next sections describe in detail the modelling steps of the proposed methodology and the results of its application to the synthetic and real data sets.

4.2 Methodology

The proposed FDEM inversion method includes the integration of the forward model in section 4.2.1, the inverse method in section 4.2.2, and the model reparameterization in section 4.2.3.

4.2.1 Forward response and sensitivity modelling

The FDEM data comprise both the in-phase (IP) and quadrature-phase (QP) components of the electromagnetic field, generally acquired by a loop-loop system. To link the unknown near-surface properties (i.e., EC and MS) to the measured data, we use a one-dimensional nonlinear approximation of the propagated electromagnetic field (Hanssens et al., 2019). This forward operator calculates the IP and QP responses per transmitter-receiver coil offset located above a model with n layers. In addition to the IP and QP responses, the forward operator also calculates the sensitivity analysis through changes in the properties of interest at depth. A detailed description of the forward model (Hanssens et al., 2019) used in this method can be found in section 2.2.2.

In this work, geostatistical simulations are used to generate high-resolution subsurface models in 3D, then the forward geophysical model described in section 2.2.2 and based on a 1D approximation is applied to compute the IP and QP predictions, and the simulated models are then updated in 3D using the ES-MDA. The simulation and update of the models is done in

3D. To reduce the computational cost, a dimensionality reduction approach is proposed to perform the updating in a lower dimensional space and increase the computational efficiency (Liu et al., 2022a). The use of the 1D forward model is one of the main limitations of this approach as it is unable to capture the propagation of the EM field in the three directions of space; however, replacing the 1D approximation would dramatically increase the computational cost of the method proposed in this Chapter.

4.2.2 Inverse method

For the inversion, numerical approximate methods are adopted for the solution of the associated inverse problem. Stochastic method is applied, namely the ES-MDA (Emerick and Reynolds, 2013; Grana et al., 2021), for its computing efficiency and the ability to quantify uncertainty. ES-MDA is derived from Kalman Filter (Evensen, 2009) to overcome the limitation of the operator linearization in non-linear inverse problems and to improve the computational efficiency in large-scale optimization and inverse problems. Like the standard Kalman Filter (KF), the ES-MDA is based on a Bayesian updating approach and the estimation of model parameters from measurements includes two steps: prediction by the forward model from the prior realizations, and correction by the measurement according to the likelihood function. In the ES-MDA, the Kalman gain is empirically estimated from the ensemble of prior models. The ES-MDA updating equation of model parameters $\mathbf{m}$ of ES-MDA can be written as:

$$\mathbf{m}_j^u = \mathbf{m}_j^p + \mathbf{K}(\tilde{\mathbf{d}}_j - \mathbf{d}_j^p), \quad (4.1)$$

for $j = 1, \dots, N_e$ with N_e being the ensemble size, where $\mathbf{m}_j^p$ represents the prior model parameters, $\mathbf{m}_j^u$ represents the updated model parameters obtained by assimilating the measurements, $\mathbf{d}_j^p$ is the predicted data obtained from $\mathbf{m}_j^p$ through the forward operator $\mathbf{F}$, $\tilde{\mathbf{d}}_j$ is the observed data with random perturbation according to the distribution of the noise $\mathbf{e}$, and $\mathbf{K} \in \mathbb{R}^{N_m} \times \mathbb{R}^{N_d}$ is the Kalman gain matrix. In the ES-MDA, the Kalman gain matrix is empirically estimated from the prior models as

$$\mathbf{K} = \mathbf{C}_{\mathbf{md}}^p (\mathbf{C}_{\mathbf{dd}}^p + \mathbf{C}_{\mathbf{d}})^{-1}, \quad (4.2)$$

$$\mathbf{C}_{\mathbf{md}}^p = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{m}_j^p - \bar{\mathbf{m}}^p)(\mathbf{d}_j^p - \bar{\mathbf{d}}^p)^T, \quad (4.3)$$

$$\mathbf{C}_{\mathbf{dd}}^p = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{d}_j^p - \bar{\mathbf{d}}^p)(\mathbf{d}_j^p - \bar{\mathbf{d}}^p)^T, \quad (4.4)$$

where $\mathbf{C}_{\mathbf{md}}^p \in \mathbb{R}^{N_m} \times \mathbb{R}^{N_d}$ represents the cross-covariance matrix between the prior model parameters $\mathbf{m}^p$ and the corresponding predicted data $\mathbf{d}^p$, $\mathbf{C}_{\mathbf{dd}}^p \in \mathbb{R}^{N_d} \times \mathbb{R}^{N_d}$ is the covariance

of the predicted data $\mathbf{d}^p$, $\mathbf{C}_d \in \mathbb{R}^{N_d} \times \mathbb{R}^{N_d}$ is the covariance matrix of the measurement error $\mathbf{e}$, $\bar{\mathbf{m}}^p$ and $\bar{\mathbf{d}}^p$ are the empirical mean of the ensemble of model variables and predicted data, respectively. In the linear case, where the forward operator $\mathbf{F}$ can be expressed as matrix $\mathbf{G}$, the covariance matrices $\mathbf{C}_{md}^p$ and $\mathbf{C}_{dd}^p$ are corresponding to $\mathbf{C}_m \mathbf{G}^T$ and $\mathbf{G} \mathbf{C}_m \mathbf{G}^T$, respectively, where $\mathbf{C}_m$ is the covariance of model parameters. Equations 4.1 and 4.2 show that the Kalman Gain matrix controls the trade-off between the prior predictions and updated correction driven by measurements according to their uncertainties. The weights of measurements are large if the measurement errors are small and vice versa. For nonlinear inverse problems, it is necessary to iteratively update the model variables to achieve a satisfactory match between prediction and measurements. One common strategy is to sequentially assimilate observations at each time step (e.g., ensemble Kalman Filter), but this procedure requires to perform forward simulations every time step and thus it is computationally inefficient. Alternatively, in ES, all data available are simultaneously used for model updating. To guarantee the convergence between the model predictions and measurements in nonlinear cases, the simultaneous data assimilation is performed multiple times. This method is referred to as ES-MDA (Emerick and Reynold, 2013).

ES-MDA is an iterative method. An ensemble of prior models is first sampled from a prior distribution and iteratively updated until the models are consistent with the measured data. Each data assimilation step can be interpreted as a Bayesian updating process, where the models updated in the previous iteration are used as the prior at the current step and then corrected by assimilating the observations. The algorithm of ES-MDA can be summarized as follows:

- i) Define the ensemble size N_e , the number of data assimilations N and the inflation coefficients $\{\alpha_k\}_{k=1,\dots,N}$ with the constraint $\sum_{k=1}^N \alpha_k^{-1} = 1$.
- ii) Generate an ensemble of N_e prior realizations $\{m_j\}_{j=1,\dots,N_e}$ of the EC and MS models conditioned on the available borehole data using geostatistical simulation algorithms.
- iii) **For** $k = 1$ **to** N
 - Compute the geophysical response of each prior realization $\{\mathbf{d}_j^p\}_{j=1,\dots,N_e}$ using the forward model described in section 2.2.2.
 - Perturb the observations $\{\tilde{\mathbf{d}}_j\}_{j=1,\dots,N_e}$ for each ensemble member as

$$\tilde{\mathbf{d}}_j = \mathbf{d}_{\text{obs}} + \sqrt{\alpha_i} \mathbf{C}_d^{1/2} \boldsymbol{\epsilon}_j, \quad (4.5)$$

where $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_d})$.

- Update model ensemble $\{\mathbf{m}_j\}_{j=1,\dots,N_e}$ using Equations 4.1 to 4.4.

End

The solution of the inverse problem is a linear combination of the updated ensemble models. The ensemble models, at each iteration, are updated according to the residuals between predicted and observed data and the cross-covariance matrix of the residuals and model variables. The initial ensemble must be large enough to represent the prior variability. If the variability of the prior is too small, the uncertainty could be severely underestimated. The number of iterations is established through a trial-and-error approach. Publications on data assimilation in dynamic reservoir modelling show that a number of iterations between 4 and 8 is generally sufficient (Emerick and Reynolds, 2013). In geophysical inverse problems, the large amount of data makes the problem less underdetermined than fluid flow modelling problems, hence 4 iterations are generally sufficient (Grana et al., 2021). However, due to the large number of measurements, a large ensemble is necessary to avoid uncertainty underestimation or ensemble collapse. The prior model includes the prior distribution of the model variables and the spatial correlation model of the realizations. For datasets with large errors, the prior distribution has a strong impact on the posterior realizations, especially the spatial correlation model. In these cases, alternative methods that predict jointly the model parameters and the spatial correlation model can be used (Laloy et al., 2015; Hunziker et al., 2017; Wang et al., 2022). The vertical correlation can be estimated from well log data, whereas the lateral correlation must be assumed based on prior geological information. For simplicity, the data errors are assumed to be spatially uncorrelated with diagonal covariance matrix; however, if geophysical data are pre-processed for quality control and denoising, the error model could be correlated, and the covariance matrix of the data is banded. The assumption of the banded covariance matrix is generally challenging in practical applications. Large variances of the errors tend to make the prior dominant on the data-driven likelihood function and might lead to a poor data match, whereas small variances of the errors tend to make the likelihood function predominant and might lead to unphysical values of the model variables.

4.2.3 Model re-parameterization

Due to the large dimension of the model grids in real applications, the ES-MDA method is often computationally and memory prohibitive. Therefore, the method presented in this Chapter proposes to reduce the model parameters using the RTD method and then perform the data

assimilation in the reduced model space. After each data assimilation, the reduced model parameters can be back transformed to the full model space using the factor matrices obtained by RTD.

A tensor is a multi-index numerical array, which can be used to represent high-dimensional data. Conventional multivariate data analysis approaches based on standard flat-view matrix models requires reshaping the data tensor into a matrix or vector and applying classical matrix factorization methods, such as singular value decomposition (SVD) non-negative matrix factorization (NMF), or independent component analysis (ICA) (Cichocki et al., 2015). These methods can be efficiently implemented but they might struggle to capture spatial correlations in multiple dimensions, which limits their performance in high-dimensional data analysis. Tensor decomposition methods are based on multilinear algebra and can exploit the intrinsic multi-dimensional patterns in the model space, as the RTD used in this work. In recent years, many deep-learning-based reduction methods have been proposed to overcome the limitation in geoscience problems (e.g., Laloy et al., 2017; Laloy et al., 2018; Canchumuni et al., 2019; Liu and Grana, 2020; Lopez-Alvis et al., 2021; Mo et al., 2019). However, those methods based on deep neural networks usually require thousands of prior models for training and are relatively difficult to integrate with ES-MDA.

The canonical polyadic (CP) decomposition and the Tucker decomposition are the two most popular tensor decomposition algorithms (Rabanser et al., 2017). The CP decomposition represents a tensor as a linear combination of vectors, whereas the Tucker decomposition decomposes a tensor into a small dense core tensor and a set of factor matrices. In this work is adopted the Tucker decomposition, because it is more suitable for dimensionality reduction in which the core tensor can be regarded as the sparse features extracted from the original tensor data and the factor matrices can be used for back-transformation.

The Tucker decomposition of an N^{th} order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be expressed as:

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{B}^{(1)} \times_2 \mathbf{B}^{(2)} \dots \times_N \mathbf{B}^{(N)}, \quad (4.6)$$

where the symbol $\times_n$ represents the tensor-matrix multiplication along mode- n , $\mathcal{G}$ is the nondiagonal core tensor that includes the information for the extension of the tensor components, and $\{\mathbf{B}^{(n)}\}_{n=1, \dots, N}$ are factor matrices that represent the principal components in the respective tensor modes.

A graphical view of the Tucker decomposition is presented in Figure 4.1. Tensor decomposition is also non-unique. A decomposition where core tensor and all factor matrices are orthonormal

is referred to the Higher-Order Singular Value Decomposition (HOSVD) (De Lathauwer et al., 2000a; De Lathauwer et al., 2000b).

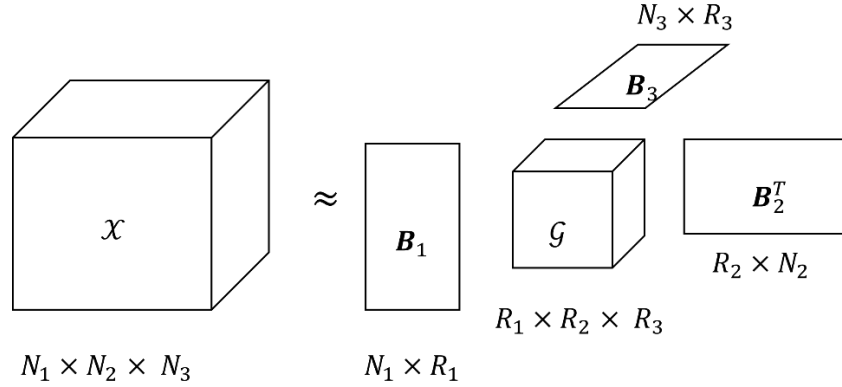


Figure 4.1 Tucker decomposition of a third-order tensor.

In practice, the 3D geological models might consist of millions of grid cells. Due to the limited memory and high computational complexity, the conventional tensor decomposition methods are usually not applicable. Randomized algorithms are then used for large-scale tensors. The randomized approach aims to find the low-rank approximation of the unfolding matrices of large-scale tensors via the probabilistic strategy (i.e., random sketching) and then perform matrix factorization on the small matrices. One popular method is the random projection in which a large-scale matrix is reduced using the transformation of a random matrix with given probability distribution. The detailed steps of the randomized HOSVD of an N^{th} order tensor data $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ are summarized as follows:

- i) Define the number of iterations N_{it} and a multilinear rank $(R_1, R_2, \dots, R_N)$.
- ii) Initialize the factor matrices $\{\mathbf{B}^{(n)} \in \mathbb{R}^{I_n \times \mathbb{R}^{R_n}}\}_{n=1, \dots, N}$ as random Gaussian matrices.
- iii) **For** $i = 1$ to N_{it} :
 - For** $n = 1$ to N :
 - $\mathcal{Z} = \mathcal{X} \times_{p \neq n} \{\mathbf{B}^{(n)T}\}$.
 - Generate a random matrix $\mathbf{\Omega}^{(n)} \in \mathbb{R}^{\prod_{p \neq n} R_p \times \mathbb{R}^{R_n}}$ drawn from Gaussian distribution.
 - Compute $\mathbf{W}^{(n)} = \mathcal{Z}_{(n)} \mathbf{\Omega}^{(n)}$ where $\mathcal{Z}_{(n)} \in \mathbb{R}^{I_n \times \mathbb{R}^{\prod_{p \neq n} I_p}}$ is the n -unfolding matrix of tensor $\mathcal{G}$.

- Compute the orthonormal basis $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n} \times \mathbb{R}^{R_n}$ of $\mathbf{W}^{(n)}$ by QR decomposition.

End

- Compute the core tensor $\mathcal{G} = \mathcal{Z} \times_1 \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)T} \dots \times_N \mathbf{Q}^{(N)T}$

End

Herein, is proposed the RTD algorithm as a dimensionality reduction method to reduce the dimension of the model and update the variables in a low dimensional model space. This approach allows improving the computational efficiency of the inversion. In this work, the uncertainty in the RTD transformation and the uncertainty in the inversion are not differentiated. The larger is the number of ensemble members, the smaller is the underestimation of the uncertainty. Similarly, the larger is the reduction of the model space, the larger is the overestimation of the uncertainty. By adopting a trial-and-error approach, can be determined, case by case, the optimal dimension of the model space and of the model ensemble. A detailed analysis of the uncertainty quantification in geophysical inverse problems with model and data reduction is presented in (Grana et al., 2019).

The presented FDEM inversion method predicts a set of model realizations that represent the posterior distribution of the inverse solution. Figure 4.2 illustrates the inversion workflow of ES-MDA with RTD. It starts with a set of prior realizations of EC and MS $\{\mathbf{m}_j^{k=0}\}_1^{N_e}$ simulated by geostatistics algorithms. Then, their EM responses $\{\mathbf{d}_j^k\}_1^{N_e}$ are predicted by the forward model and the reduced model parameters $\{\mathbf{z}_j^k\}_1^{N_e}$ are obtained by the RTD. The reduced model parameters are then updated $\{\mathbf{z}_j^{k+1}\}_1^{N_e}$ by assimilating the observations with ES-MDA. It is an iterative procedure in which the prior models in the next iteration $\{\mathbf{m}_j^{k+1}\}_1^{N_e}$ are back transformed from $\{\mathbf{z}_j^{k+1}\}_1^{N_e}$ by the inverse RTD. In the FDEM inversion method presented in this Chapter, both the ES-MDA and the model reduction with RTD affect the uncertainty assessment of the posterior solution. The performance of the ES-MDA depends on the number of models in the initial ensemble while the performance on the RTD depends on the dimensionality of the lower-dimensional space. Other stochastic inversion methods with robust uncertainty assessment, such as Markov chain Monte Carlo method (Blatter et al., 2018) could also be adopted.

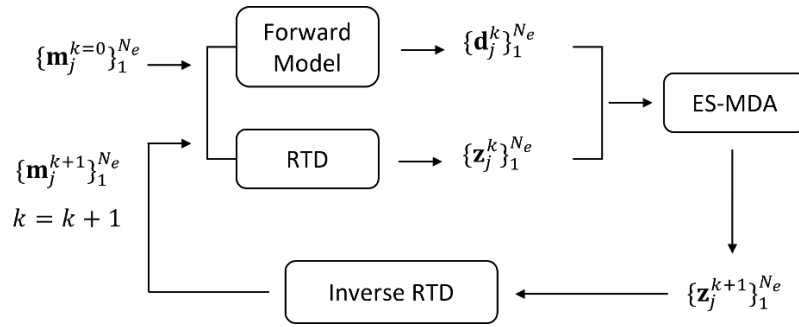


Figure 4.2 Workflow of ES-MDA with RTD.

4.3 Synthetic case application

The inversion method presented in this Chapter was first applied to a vertical section of the 3D synthetic data set presented in section 2.3.1 and based on real data collected at a mine tailing in Portugal (Panasqueira). This dataset comprises laboratory measurements of porosity and particle density obtained from samples collected from two main rock types of the site: fine-shaly sands (which constitutes the predominant rock type), and quartz-schist gravels. The models of the main physical properties (e.g., porosity, particle density) were first generated using direct sequential simulation (Soares, 2001) based on a variogram model that represents the expected spatial correlation of each property. The true EC model is then generated from these main physical properties using Archie's equation (Archie, 1942). The true MS model is generated using geostatistical simulations (Soares, 2001) based on a variogram model that describes the expected spatial distribution pattern of MS. From the resulting EC and MS models (Figures 4.3a and 4.3b), four pseudo-boreholes were extracted equally spaced along the vertical section. A detailed description of this synthetic data set can be found in section 2.3.1.

The prior ensembles of EC and MS include 500 geostatistical realizations generated using direct sequential simulation (Soares, 2001). This set of models represents the model parameter space and the histograms of both properties as retrieved from the borehole data. Therefore, this geostatistical simulation algorithm does not assume any parametric distribution for the property of interest. The EC and MS data extracted at the borehole locations are used as conditioning data for the geostatistical simulation so that all model realization reproduce the borehole data at the borehole locations. Based on the spatial continuity retrieved from the borehole data, the prior ensemble of realizations is simulated by imposing omnidirectional horizontal exponential variograms for EC and MS with a range of 6 m and 8 m, respectively. The vertical direction is modelled with exponential variograms with a range of 4 m for EC and 6 for MS (Table 4.1).

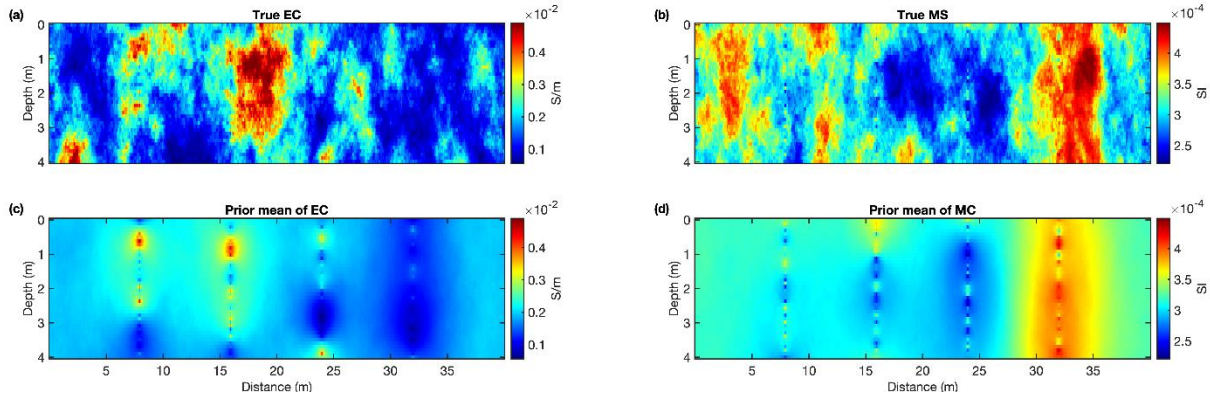


Figure 4.3 True and prior mean of EC and MS of the synthetic case: (a) true EC model; (b) true MS model; (c) prior mean of EC; and (d) prior mean of MS. It is clear the influence of the borehole data in the prior mean of EC and MS.

Table 4.1 Horizontal and vertical ranges and the nugget effects expressed as percentage of the total variance of the data, for the variogram models used to simulate and co-simulate EC and MS.

Omnidirectional spherical variogram	EC models	MS models
Horizontal range	6 m	8 m
Vertical range	4 m	6 m
Nugget effect	5 %	5 %

The true FDEM data were generated using the forward model (Hanssens et al., 2019) described in section 2.2.2, and mimicking the coil configurations of a multi-receiver FDEM sensor, namely a DUALEM-21S (DUALEM Inc., Milton, Canada). Therefore, was considered a loop-loop system setup, characterized by one transmitter coil and multiple receiver coils with two spatial configurations and two offsets per coil configuration, namely the horizontal coplanar (HCP) configuration with 1 and 2 m offset, and the perpendicular (PRP) configuration with 1.1 and 2.2 m offset. The data are contaminated by Gaussian noise and the noise level is 10% of the observations.

The model grid includes 400×40 cell in the i - and k - directions, respectively. The pointwise mean models of the prior EC and MS ensembles (Figures 4.3c and 4.3d) reproduce the true EC and MS measurements at the borehole locations. Far from the location of the boreholes, and for distances larger than the variogram range, these models tend to the average value of the distribution. For this 2D example, both EC and MS model are a third order tensor with a size of $400 \times 40 \times 500$ (corresponding to the numbers of model grids in the i - and k - directions, and ensemble size, respectively). The tensors of EC and MS model are reduced to $40 \times 5 \times 500$ by the RTD algorithm with four iterations before data assimilation. The ES-MDA is then applied in the reduced model space. The number of iterations of the ES-MDA is 4 with the inflation

coefficients of 9.33, 7.0, 4.0 and 2.0, which are recommended by Emerick and Reynolds (2013). The posterior mean of EC and MS is shown in Figures 4.4a and 4.4b, and the posterior standard deviation is shown in Figures 4.4c and 4.4d. The absolute residuals between the predicted posterior mean and the true models are shown in Figures 4.4e and 4.4f. The results capture small- and large-scale features of the true EC and MS models up to the depth of investigation provided directly by the forward operator used in the inversion (Hanssens et al., 2019) which is estimated to be approximately 3 m with the coil configurations used in this application. The inversion, by construction, reproduces exactly the measurements at the borehole locations. Hence, the posterior standard deviation is zero at the borehole location and it increases with the distance from the borehole locations.

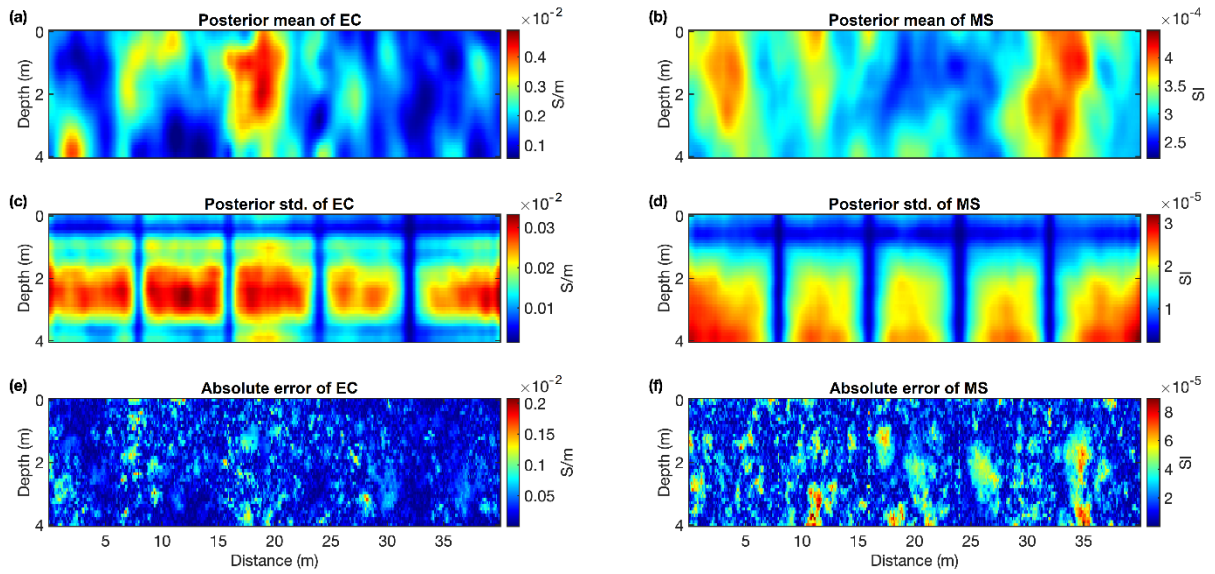


Figure 4.4 Posterior mean, standard deviation (std.) and residual of EC and MS of the synthetic case: (a) posterior mean of EC models; (b) posterior mean of MS models; (c) posterior std. of EC; (d) posterior std. of MS; (e) absolute error between the posterior mean and the true EC; (f) absolute error between the posterior mean and the true MS.

Despite the large variability in the values of the QP and IP responses predicted from the prior models for all coil offsets (Figures 4.5 and 4.6) the posterior model of QP and IP matches the observed data for most of the observations. At some locations the predicted models show mismatches with the observation. These results might be related simultaneously to the noise component within the data and the uncertainty originated due to the dimensionality reduction technique applied as part of the proposed method (i.e., RTD) (Grana et al., 2019; Liu et al., 2022a).

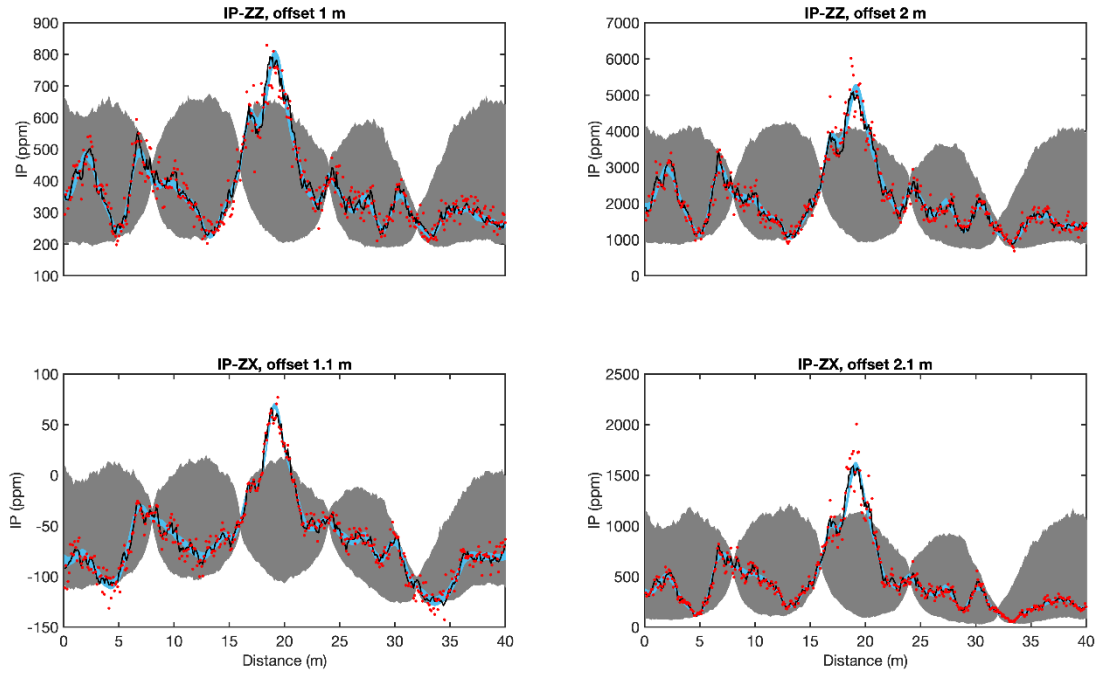


Figure 4.5 Predicted IP data from the prior and posterior EC and MS models of the synthetic case. The red dots are the true measurements with noise; the black lines represent the true data without noise; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.

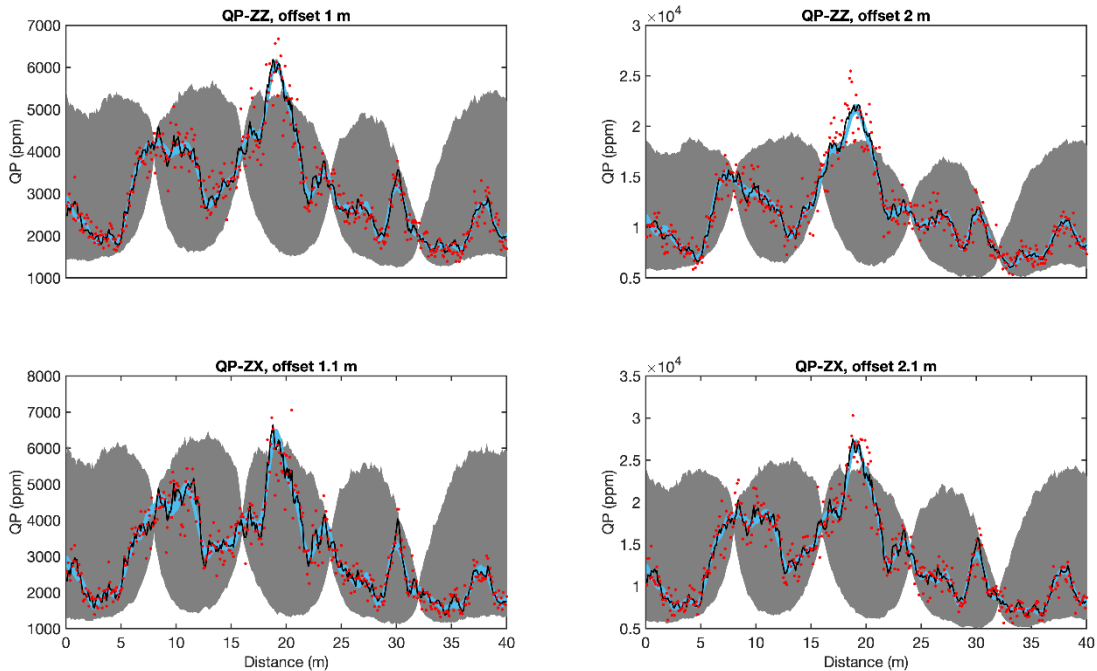


Figure 4.6 Predicted QP data from the prior and posterior EC and MS models of the synthetic case. The red dots are the true measurements with noise; the black lines represent the true data without noise; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.

4.4 Real case application

The proposed ES-MDA with RTD to the FDEM inversion was applied to the same a real data set from a FDEM survey described in section 2.4.1. This FDEM survey was acquired over arable land with slight slope and 20 cm of rendzina soil cover, located near Knowlton (Dorset, UK), and containing several archaeological features. The region of interest is characterized by Cretaceous chalk in the shallow subsurface, and calcareous ooze, overlain by Quaternary siliciclastic sand deposits. The Cretaceous formation is characterized by a background susceptibility of zero and a low EC (~ 7 m/Sm), while the sand deposits is strongly magnetic ($MS \approx 1 \times 10^{-3}$) and slightly more conductive than the bedrock. In this area IP anomalies are related to the buried archeology and the background geology provides a large range of QP values (Delefortrie et al., 2018).

The FDEM data are collected using a DUALEM 21HS instrument, with an operating frequency of 9000 Hz in a loop-loop setup, elevated at 0.16 m from the surface pulled by a quadbike. The data acquisition was performed along parallel lines 1 m apart at a speed of ~ 8 km/h, and a sampling frequency of 8 Hz. This application used the FDEM data collected from one transmitter paired with two coplanar receiver coils, in horizontal mode, at 1 and 2 m from the transmitter (HCP1 and HCP2, respectively), and two receivers in vertical model, 1.1 and 2.1 m from the transmitter (PRP1 and PRP2, respectively). The measured IP and QP data are noisy and with systematic errors; therefore, a calibration was performed before the inversion using the existing EC and MS borehole measurements, applying a drift correction consisting of tie-line levelling as described in (Delefortrie et al., 2018). However, was not tackled the striping effect present in the PRP IP signal and the point anomalies observed in HCP QP signal (Delefortrie et al., 2018). These characteristics of the observed signal do affect the quality of the inversion results and the match between predicted and observed data.

EC and MS data are collected in twelve boreholes at intervals of 5-10 cm, reaching a maximum depth of 1.2 m and a minimum of 0.8 m. Eleven boreholes were used to compute horizontal and vertical experimental variograms based on a spherical model (omnidirectional in the horizontal direction) for both EC and MS properties, with horizontal range of 16.8 m for EC and 28.2 m for MS, and vertical range of 0.8 m for EC and 0.7 m for MS.

The model grid includes $531 \times 171 \times 20$ cells in the i -, j - and k - directions. A set of 500 geostatistical realizations of EC and MS is then generated conditioned on the borehole data and assuming variogram models fitted to experimental variograms computed from the borehole data. The mean prior models of EC and MS (Figures 4.7a and 4.7b) match the

borehole measurements and tend to the mean of the direct measurements away from the borehole.

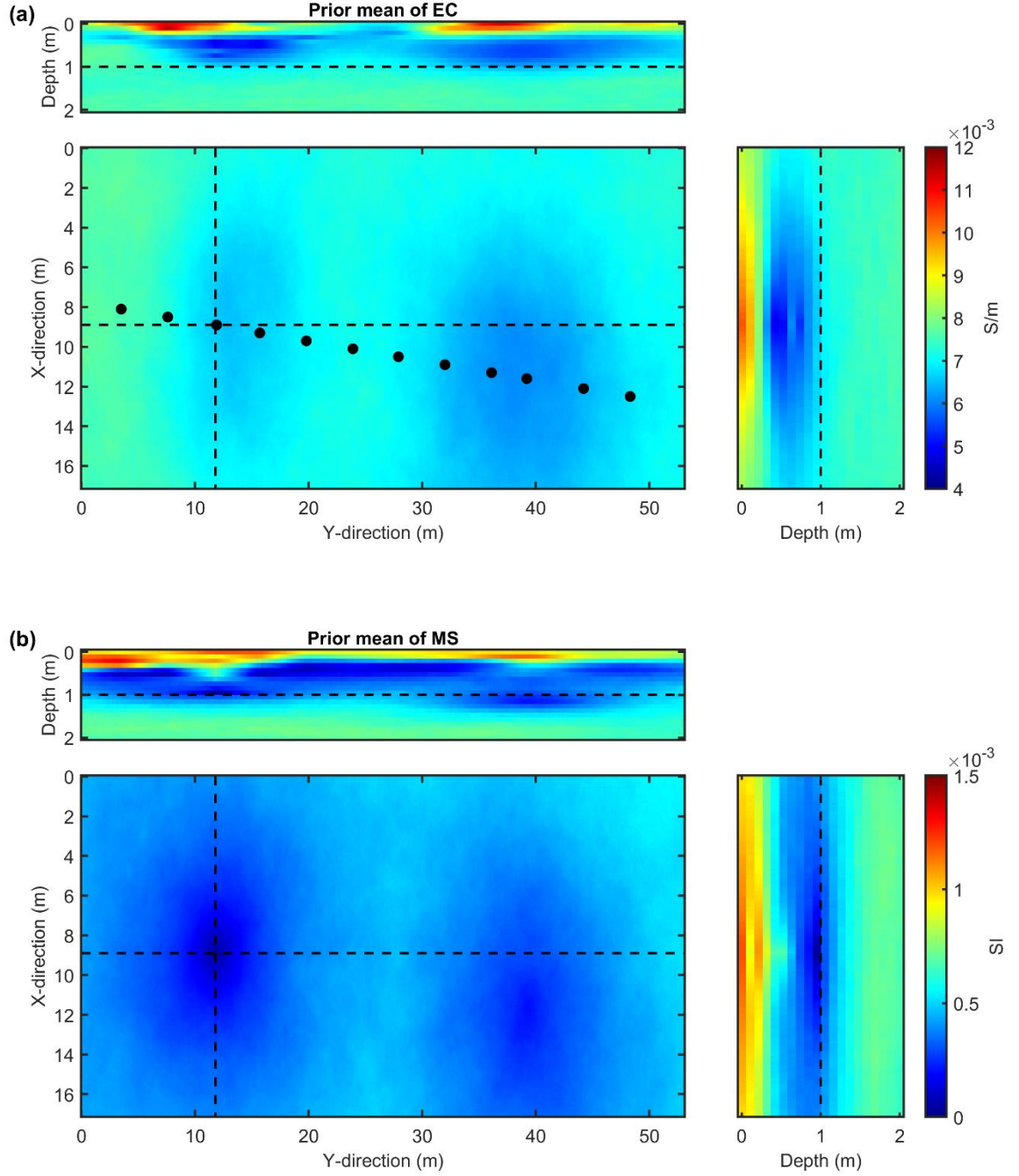


Figure 4.7 Prior mean of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices, and the black dots represent the well locations.

For this 3D case, both EC and MS model is a fourth order tensor with a size of $531 \times 171 \times 20 \times 500$ (corresponding to the numbers of model grids in the i -, j - and k - directions, and ensemble size, respectively). The tensors of EC and MS model are reduced to $20 \times 40 \times 5 \times 500$ by the RTD algorithm with four iterations before data assimilation. Then was applied the ES-MDA inversion with 4 iterations and inflation coefficients of 9.33, 7.0, 4.0 and

2.0 (Emerick and Reynolds, 2013). The posterior mean models (Figures 4.8a and 4.8b) show a detailed spatial distribution pattern with a layer of continuous high conductivity and susceptibility at around 1 m depth. This value is consistent with the observed depth of the top chalk as interpreted from the existing borehole data (Delefortrie et al., 2018). The posterior standard deviation of EC and MS is shown in Figure 4.9. As the prior ensemble of EC and MS was constructed with geostatistical simulation, the predicted EC and MS values at the borehole locations are exactly reproduced.

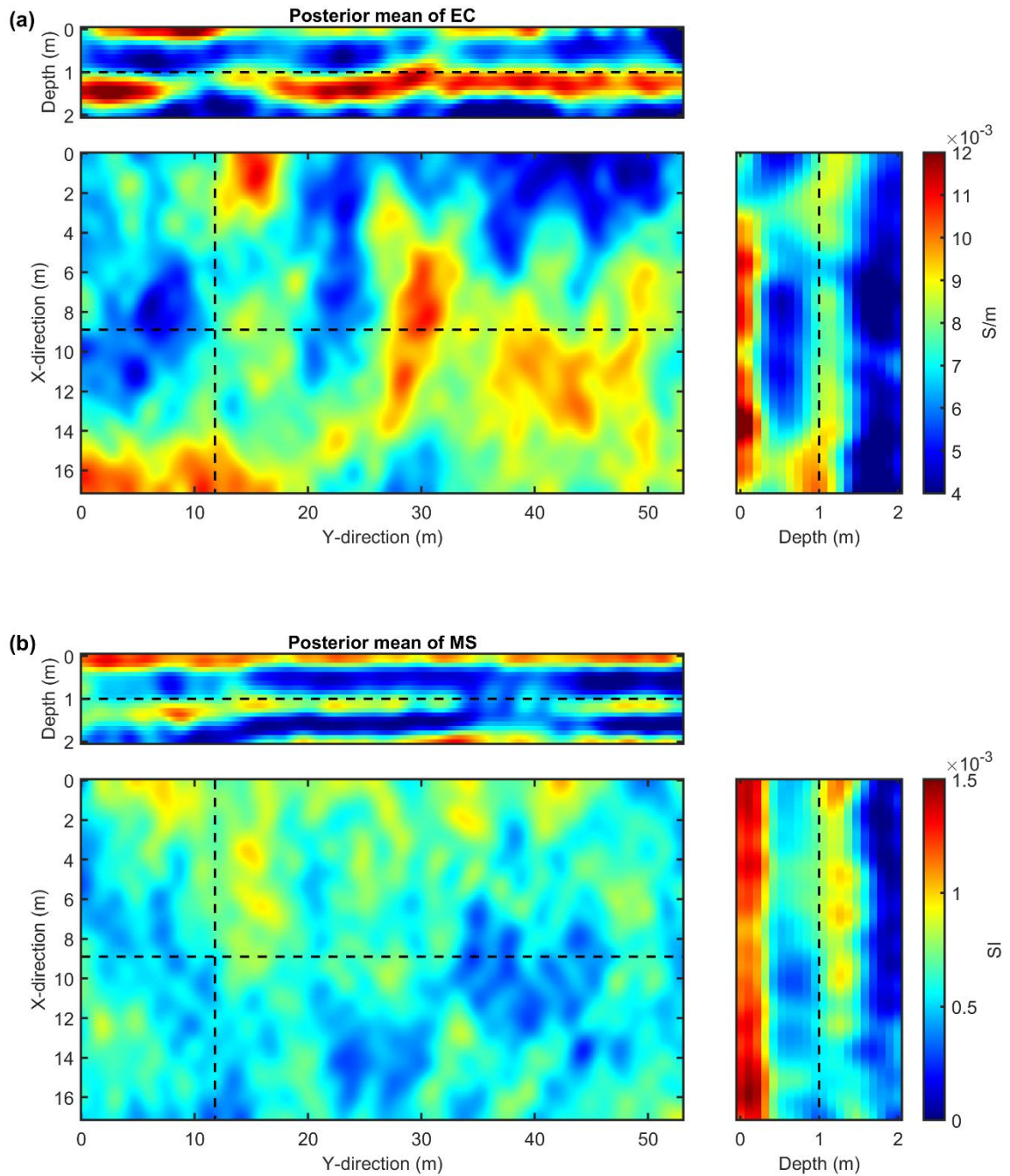


Figure 4.8 Posterior mean of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices.

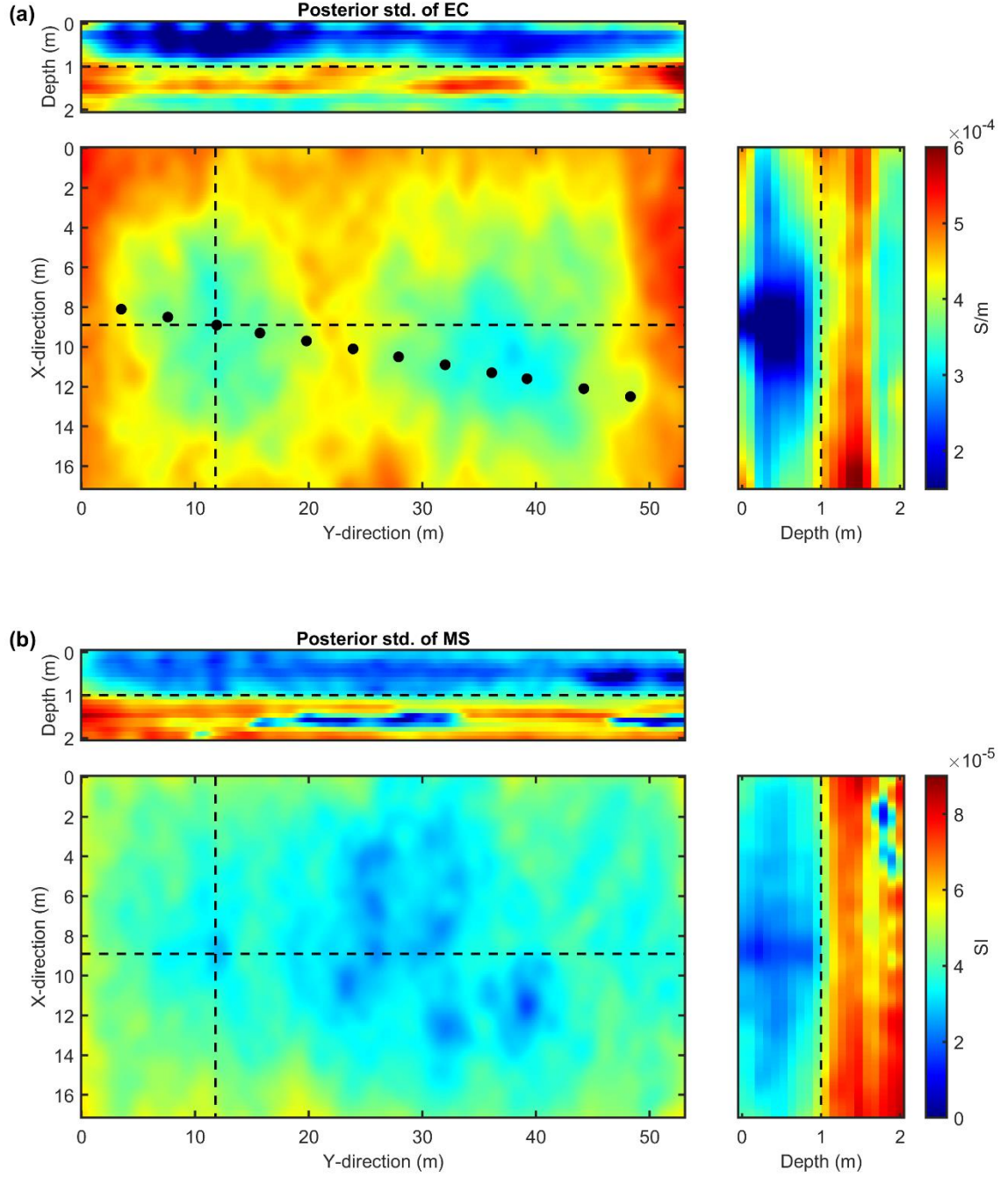


Figure 4.9 Posterior standard deviation of EC (a) and MS (b) models of the real case. The black dashed lines represent the locations of the X-, Y- and depth slices.

In Figures 4.10 and 4.11, is shown the comparison between predicted and measured IP and QP data. The predicted data match relatively well the observed data. The mismatch between the predicted and observed data might be due to the noisy nature of the data, as described above, and the one-dimensional approximations of the forward operator that cannot model complex and highly heterogenous lateral distributions of electrical properties. The parametrization of the RTD and the ES-MDA might be partly attributed to the misfit between predicted and observed data.

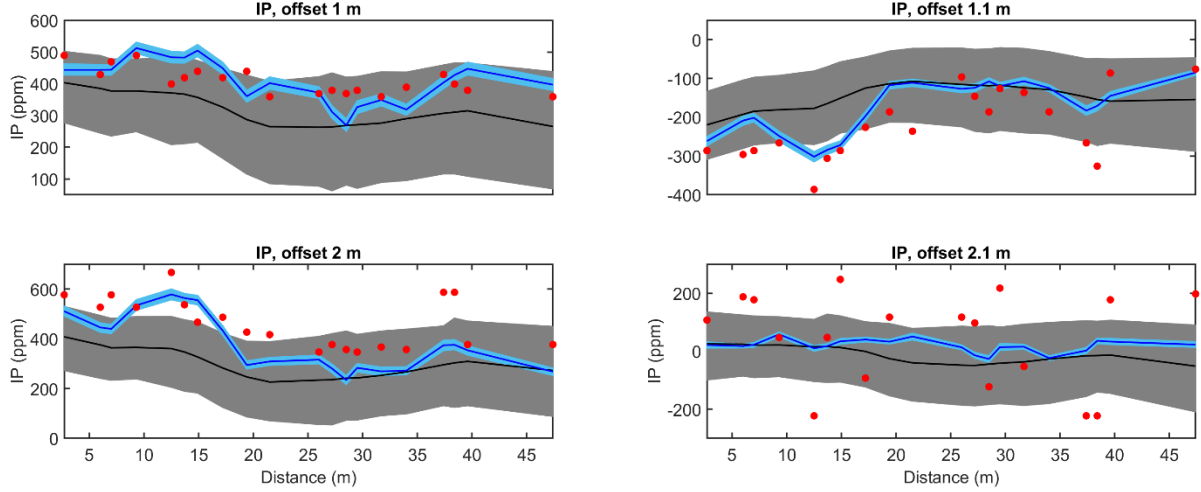


Figure 4.10 Predicted IP data from the prior and posterior EC and MS models of the real case. The red dots are the true measurements; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.

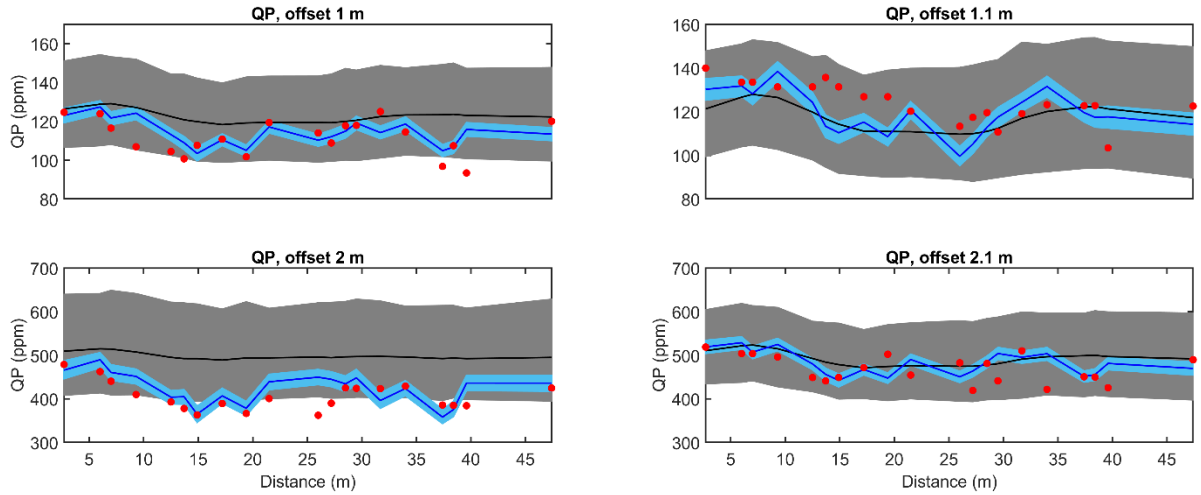


Figure 4.11 Predicted QP data from the prior and posterior EC and MS models of the real case. The red dots are the true measurements; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior prediction, respectively; the black and blue lines represent the prior mean and posterior mean, respectively.

The computational cost of one updating step for ES-MDA is $\mathcal{O}(N_e^2 N_m + N_e^2 N_d)$ where N_e , N_m and N_d are the ensemble size, the number of model parameters and observations, respectively (Nino Ruiz et al., 2015). In the real case, the ensemble size is 500; the number of observations is 26,640; the numbers of model parameters with and without reduction are 1,816,020 ($171 \times 531 \times 20$) and 4,000 ($20 \times 40 \times 5$), respectively. The speed-up ratio with model dimension reduction by RTD is roughly 60.14.

4.5 Discussion

This Chapter proposes a stochastic FDEM inversion method in a reduced space leveraging the benefits of RTD with respect to dimensionality reduction. The method is illustrated in two application examples: one synthetic and one real. First, we considered a 2D synthetic data set to evaluate the accuracy of the predictions. Then, the proposed method was applied in a real 3D data set to assess its performance under real noise conditions. In both application examples the data predictions do match the observed FDEM data (Figures 4.5, 4.6, 4.10 and 4.11). Besides, when observing the residuals between model predictions of EC and MS and the true model (Figure 4.4e and 4.4f) they do not exhibit any spatial continuity pattern that is consistent with the true EC and MS spatial continuity pattern. In both application examples the prior ensembles of EC and MS are constructed via geostatistical simulation. While alternative methods can be applied, this class of methods have the ability to reproduce direct observations (i.e., borehole data, histograms and spatial continuity patterns as revealed by variogram models). For this reason, when relying on geostatistical simulation to build the prior ensemble, a critical aspect for the success of the proposed inversion method is the availability of borehole data and its spatial distribution within the area of interest. Spatial sampling, including the spatial distributions of conditioning data, has been extensively studied in mining engineering (Journel and Huijbregts, 1978). A limited number of boreholes might affect the accuracy of the inversion and lead to large uncertainties in the predictions. In real applications, the prior distributions and variogram models assumed in the generation of the prior ensembles should account for prior geological information available for the area under investigation as well as direct measurements from nearby areas.

Figures 4.5, 4.6, 4.10 and 4.11 show that the P5-P95 interval of the predicted data do not encapsulates entirely the observed data. In other words, there is an underestimation of the predicted uncertainty. This fact might be originated by two complementary reasons: the reparameterization of the model parameter space with the RTD affects the uncertainty assessment (Grana et al., 2019 and Liu et al., 2022b); the ES-MDA has a better performance for non-Gaussian and non-linear inverse problems. Finally, to assess the performance of the inversion locally, one borehole from the conditioning data set was removed in the real case application. Removing a larger number of conditioning boreholes would decrease the accuracy of the predictions as the estimation of the EC and MS distributions would be poor. The comparison between the predicted properties and the borehole measurements at the location of the borehole not used to constrain the inversion is shown in Figure 4.12. Despite the limited length of the measured EC and MS, the estimated posterior distribution matches the true EC and MS. Due to the relatively small number of samples the predictions in the deeper part of the model are less reliable.

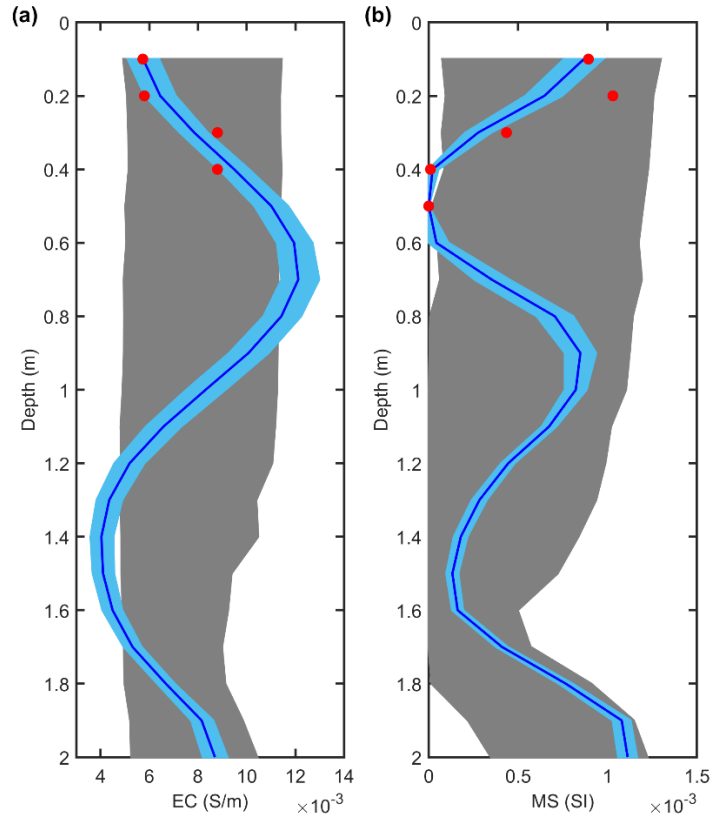


Figure 4.12 Comparison of the predicted and measured EC (a) and MS (b) at the blind well of the real case. The red dots are the true measurements; the intervals in grey and light blue correspond to the region between the percentiles P5 and P95 of the prior and posterior realizations, respectively; the blue lines represent the posterior mean.

The ES-MDA was used due to its relatively simplicity of implementation and its potential to efficiently assess the posterior distribution in geophysical inversion problems. However, the computational cost of ES-MDA might be prohibitive for large-dimensional inverse problems such as FDEM inversion. For this reason, was combined RTD, a dimensionality reduction technique of the model parameter space, with ES-MDA. The application examples shown herein, show that the coupling of both methodologies is an efficient solution to alleviate the computational burden of ES-MDA without compromising the model predictions and the uncertainty assessment despite assumptions about the prior distributions of the model parameters.

4.6 Conclusion

In this Chapter is proposed a FDEM inversion method that combines ES-MDA with RTD to predict the spatial distribution of EC and MS. The initial prior ensemble of models is generated using geostatistical simulation, to model the complex and heterogeneous subsurface distributions. Then, RTD coupled with ES-MDA makes the inversion method computationally

feasible and applicable to 3-dimensional grids with a large number of cells. This FDEM inversion method was validated on a two-dimensional synthetic data set and then applied to a 3-dimensional real data set. In both application examples, the predicted models reproduce the measured EC and MS data while allowing assessing the uncertainty of the predictions. The proposed inversion relies on a one-dimensional forward approximation but could be extended to more complex physical models.

4.6.1 Conclusion data and materials availability

The code and synthetic data are freely available on GitHub (<https://github.com/theanswer003/ES-RTD-FDEM>)

CHAPTER 5

Geostatistical joint inversion of FDEM and DC resistivity data

The research presented in this chapter was submitted to:

Narciso, J.*, Van De Vijver, E., and L. Azevedo, (under review), Geostatistical joint inversion of frequency-domain electromagnetic data and direct current resistivity data for modelling near-surface deposits: *Geophysics*.

Due to their sensitive to subsurface electrical conductivity (EC), direct-current resistivity and electromagnetic geophysical methods, particularly electrical resistivity tomography (ERT) and frequency-domain electromagnetic (FDEM) methods, have been widely applied in different near-surface activities, such as agriculture, urban development, or investigation of mineral and groundwater resources. Predicting the spatial distribution of EC from FDEM and ERT data requires solving a geophysical inversion problem. Due to the different spatial resolutions of both methods, and the nonlinearity of the inverse problem, individual inversions of each type of the geophysical data have been the standard to predict EC. However, the joint inversion of FDEM and ERT data has the potential to reduce the non-uniqueness of the inversion solution and to increase the ability to model the small-scale spatial heterogeneity which is characteristic of near-surface environments. We propose herein an iterative geostatistical joint inversion method of FDEM and ERT data. A geostatistical framework is used to couple both data domains in a consistent spatial model. The misfit between predicted and observed data simultaneously for both domains drives the convergence of the iterative procedure. The method is validated in a synthetic data set that illustrates a complex and highly heterogeneous near-surface environment. The proposed joint inversion method is also applied in a real case, characterized by high conductivity field data. The joint inversion results present improvements over the FDEM inversion results in both synthetic and real case applications, specifically in the modelling of the small-scale variability and the reduction of the spatial uncertainty at depth. In both application examples, the models predicted at the last iteration agree with the expected spatial distribution of the true EC field.

5.1 Introduction

The near-surface is a heterogeneous and highly dynamic region of the subsurface, particularly in urban environments, as the result of complex, interacting processes of both natural and anthropogenic origin (e.g., Lehmann and Stahr, 2007; Morel and Heinrich, 2008). Due to these reasons, an accurate characterization of the spatial distribution of the near-surface geological properties is often challenging, yet essential for different activities (e.g., groundwater contamination, geotechnical engineering, mineral resources prospecting, soil assessment, archaeological detection). The characterization of these systems based exclusively on discrete direct observations acquired through conventional invasive sampling techniques, such as drilling and core sampling, can capture the vertical spatial variability of these heterogeneous deposits at sparse location in space. These techniques are expensive, impractical to perform

in some sites and reveal limitations in capturing the lateral spatial variability of near-surface properties. Non-invasive geophysical surveys, particularly electrical and electromagnetic methods, have been proven powerful tools for the collection of virtually spatially continuous high-resolution datasets that can be translated in detailed images of the near-surface physical properties (Moorkamp, 2017). Within this context, frequency-domain electromagnetic (FDEM) induction and electrical resistivity tomography (ERT) methods have demonstrated their efficiency to characterize heterogeneous subsurface systems. Both methods are sensitive to electrical conductivity (EC), while frequency-domain electromagnetic data can additionally be linked to the subsurface magnetic susceptibility (MS) and dielectric permittivity (Everett, 2013). EC is directly related to porosity, water saturation and the conductivity of pore fluids, while MS is a function of the metal content in the subsurface.

The data acquired from both geophysical methods can be translated into numerical subsurface models of the physical subsurface properties of interest by solving a geophysical inversion problem. These geophysical inverse problems are ill-posed nonlinear problems and have a nonunique solution due to the larger number of model parameters when compared against the observed data. This is because geophysical measurements are band-limited and contaminated by noise and inconsistencies during the collection of the data set, resulting in uncertainties in the inverse models (Tarantola, 2005).

A joint inversion of multiple geophysical methods, sensitive to a common subsurface property, or properties, can potentially improve the predicted subsurface models, while reducing the uncertainty of the predictions (Moorkamp, 2017). The joint inversion leverages the benefits of each method individually resulting in better predictions about the geometry and spatial distribution of the subsurface properties when compared to using just a single method, hence mitigating the non-uniqueness of the inverse problem (Bobe et al, 2020). Although FDEM and ERT data are sensitive to the same subsurface physical property, both methods are often interpreted and modelled separately. Nevertheless, inverting both data sets in a joint inversion framework is generally a preferable approach due to the complementary information about the subsurface provided by each method due to differences in the spatial resolution. Regarding the complementary information from FDEM and ERT data sets, while FDEM inversion can detect thin conductive layers and fails to model thin resistive layers in conductive environments, ERT inversion have the opposite characteristics (Sharma and Kaikkonen, 2003). However, handling the differences in the resolution and nature of both methods is not straightforward.

A few approaches for joint inversion of direct current (DC) electrical resistivity and electromagnetic induction (EMI) data have been presented (e.g., Raiche et al., 1985; Sharma

and Kaikkonen, 2003; Yi and Sasaki, 2015). These methods apply deterministic gradient-based geophysical inversion approaches, which are not capable to assess uncertainty of the predicted subsurface model. Also, these methods are based on the use of arbitrary weights in the objective function to balance the importance, and assimilation of each data set, during the inversion procedure.

Probabilistic inversion approaches can quantify the uncertainty related to the prediction obtained from solving a geophysical inverse problem. The common approach in geophysical inversion is the use of Markov chain Monte Carlo (MCMC) (Tarantola, 2005). However, in joint inversion application with highly parametrized models, the forward model is computationally expensive, which can limit the applicability of the MCMC methods and their convergence within a reasonable number of forward model runs (Sambridge and Mosegaard, 2002). Rosas-Carbajal et al. (2013) jointly invert DC resistivity and EMI data using a joint inversion method based on MCMC and the plane waves approximation. Their results show a reduction of the uncertainty in the predicted models by comparison to the separate MCMC inversion of these geophysical data. Bobe et al. (2020) introduce a joint inversion of DC resistivity and small-loop EMI data based on the Kalman ensemble generator (KEG) as alternative to a MCMC inversion framework. While this KEG method is computationally less expensive than MCMC it assumes Gaussian distribution for the probability distribution of the model parameters and the errors present in the observed data.

In this work, we present an iterative geostatistical joint inversion technique of FDEM and ERT data for EC, based on a previously established iterative geostatistical FDEM inversion technique (Narciso et al., 2022) (Chapter 2). A geostatistical framework is used to couple both data domains in a consistent spatial model while assessing the uncertainty of the predicted models. For each geophysical data, a dedicated forward model is used to compute synthetic data. The misfit between predicted and observed data for each domain drives the convergence of the iterative procedure, conditioning the co-simulation of new EC models in the subsequent iterations. The method is validated in a synthetic data set that illustrates a complex and highly heterogeneous near-surface environment, developed from direct and laboratory measurements on geological samples collected at a mine tailing disposal site in Portugal (Panasqueira). The proposed iterative geostatistical joint inversion method is also applied in a real case study, characterized by high conductivity field data. The results obtained are discussed against the individual inversion of the FDEM data, and present improvements in both synthetic and real case applications, specifically in the modelling of the small-scale variability and the reduction of the spatial uncertainty at depth. In both applications, the models at the last iteration that predict FDEM and ERT data match better the observed data of each

geophysical method and reproduces better the true electrical conductivities, than the models obtain from the FDEM inversion.

5.2 Methodology

The proposed iterative geostatistical joint inversion method predicts the spatial distribution of EC from FDEM and ERT data and also predicts the spatial distribution of MS from FDEM data. The relationship between the model parameters ($\mathbf{m}$) (i.e., EC and MS) and both geophysical data ($\mathbf{d}_{\text{obs}}$), contaminated by noise (ϵ), can be mathematically summarized by

$$\mathbf{m} = \mathbf{F}^{-1}(\mathbf{d}_{\text{obs}} + \epsilon). \quad (5.1)$$

We approximate the inverse problem stated in Eq. 5.1 with an iterative geostatistical geophysical inversion method based on global approach (Azevedo and Soares, 2017, Narciso et al., 2020). It relies on two key main ideas: i) the perturbation of the model parameter space and update technique with stochastic sequential simulation and co-simulation (Soares, 2001); and ii) the convergence is ensured by a global stochastic optimizer driven simultaneously by the misfit between true and synthetic FDEM and ERT data. The proposed iterative geostatistical joint inversion methodology may be summarized in the following sequence of steps (Figure 5.1) and divided in four main steps, which are described in detail below.

5.2.1 EC and MS model generation

The proposed iterative geostatistical joint inversion methodology starts with the generation of sets of N_s models of EC and MS with stochastic sequential simulation and co-simulation (Deutsch and Journel, 1998). Each model is simulated, or co-simulated, for the entire inversion grid at once. Available direct measurements of EC and MS from borehole data are used as conditioning experimental data. The spatial continuity pattern of the simulated models is imposed through a variogram model, fitted to experimental variograms computed from the available direct measurements, or borrowed from a prior geological knowledge. In the proposed joint inversion methodology, we use direct sequential simulation (Soares, 2001) and co-simulation with joint probability distributions (Horta and Soares, 2010) as model perturbation technique. Unlike sequential Gaussian simulation (SGS) (Deutsch and Journel, 1998), these stochastic sequential simulation techniques do not impose any condition on the data distribution (i.e., Gaussian) of the properties to be simulated, thereby avoiding the intermediate step of a data transformation of the distribution of the properties to be simulated. Rather, the marginal and joint distribution as inferred from the experimental data are used in the simulation procedure. The use of non-Gaussian stochastic sequential co-simulation

techniques allow a better reproduction of the relationship between variables as retrieved from the available direct measurements, especially for complex and highly nonlinear relationships between primary geophysical and secondary petrophysical properties related to the geophysical measurements used in the inversion procedure, electromagnetic induction and direct-current resistivity measurements.

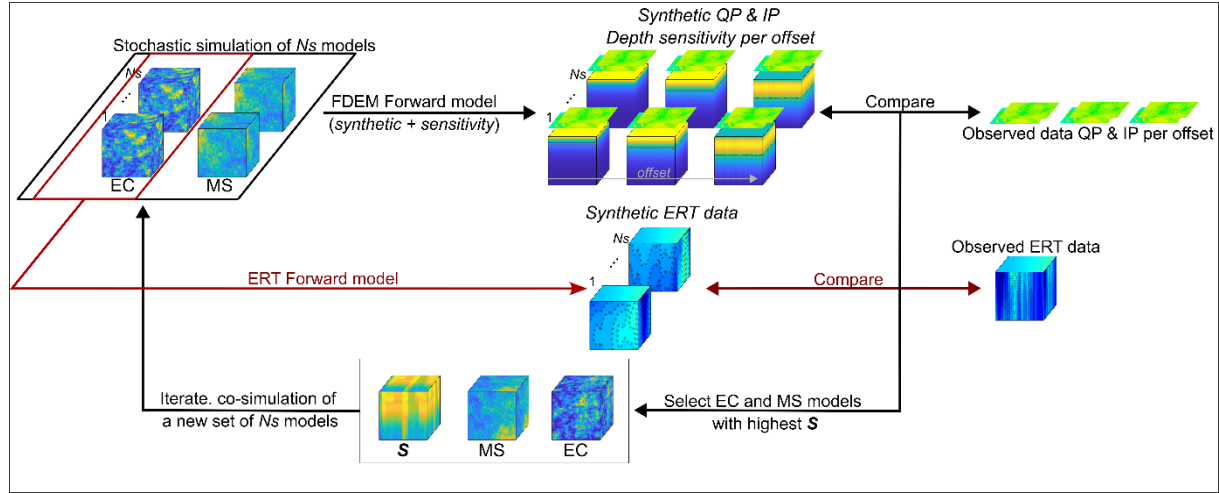


Figure 5.1 Schematic representation of the iterative geostatistical joint inversion method using FDEM and ERT data.

As we rely on stochastic sequential simulation and co-simulation, all the subsurface models generated during the iterative procedure, reproduce exactly the values of the borehole data at their locations, the global marginal and joint distribution of EC and MS, and the imposed spatial continuity pattern (i.e., the variogram models for each property individually).

5.2.2 FDEM forward model and sensitivity analysis

The forward model F maps the model ($\mathbf{m}$) into the data ($\mathbf{d}_{\text{obs}}$) domain. Thus, a forward model is necessary to calculate the theoretical FDEM instrument response, existing in two components – the in-phase (IP) and quadrature-phase (QP) – for a given subsurface distribution of EC and MS. The FDEM forward model can be formulated in 1D, 2D or 3D (e.g., Auken and Christiansen, 2004, Cox and Zhdanov, 2008, Farquharson et al., 2003), to simultaneously address EC, both EC and MS or even EC, MS and dielectric permittivity.

The proposed geostatistical joint inversion method uses a forward model that calculates the theoretical 1D normalized EM response, expressed in IP and QP components, of a small loop-loop system, that is characterized by one transmitter coil and one or multiple receiver coils (Hanssens et al., 2019). This forward model considers a FDEM system positioned at the surface of an n -layered subsurface model, that accounts for EC, MS and dielectric permittivity. This forward model uses Hankel functions, numerically calculated by means of a Guptasarma

and Singh digital filter (Guptasarma and Singh, 1997), to determine a superposition of Bessel functions of the zeroth and/or first order that model the EM responses.

The sensitivity modelling represents how sensitive the forward model is toward changes of a physical property $\mathbf{m}$ (i.e., EC) at a specific layer n of the layered half-space. The sensitivity modelling calculates, with the corresponding forward response, the vertical sensitivity distribution related to each physical property within the considered layered model.

After this step, we obtain a set of N_s responses of IP and QP per coil configuration, and the corresponding vertical sensitivity profiles. A detailed description of the FDEM forward model is shown in Section 2.2.2.

5.2.3 ERT forward model

The forward model used in the proposed iterative geostatistical joint inversion method to compute N_s synthetic apparent resistivity models from the previously generated electrical conductivity geostatistical realizations is a two and a half-dimensional forward model (Pidlisecky and Knight, 2008).

In ERT surveys, a series of known currents are injected in the ground using two current electrodes, then a series of voltage measurements are obtained in two other electrodes. Poisson's equation (Eq. 5.2) can be used to describe the electric potential field generated when a current passes across an electrode dipole:

$$-\nabla \cdot (\sigma \nabla \phi_p) = I(\delta(r - r^+) - \delta(r - r^-)), \quad (5.2)$$

where σ is the electrical conductivity [$M^{-1}L^{-3}T^3I^2$], ϕ_p is the potential field [$ML^2T^{-3}I^{-1}$], I is the input current [I], δ is the Dirac delta function, and r^+ and r^- are the locations of the positive and negative current electrodes, respectively. To solve numerically Eq. (5.2) for the electric potential, ϕ_p , numerical gradient, and divergence approximations are required. Once numerical finite difference operators have been derived for gradient and divergence, Eq. 5.2 can be written in matrix notation as:

$$(\mathbf{DS}(\sigma)\mathbf{G})\hat{\phi} = \mathbf{A}(\sigma)\hat{\phi} = q, \quad (5.3)$$

where $\mathbf{D}$ is the divergence matrix, $\mathbf{S}(\sigma)$ is a diagonal matrix containing the electrical conductivity values, $\mathbf{G}$ is the gradient matrix, $\hat{\phi}$ is a vector of electric potentials, $\mathbf{A}(\sigma)$ is the combined forward operator, and q is a vector containing the current electrode pairs (Pidlisecky and Knight, 2008). Equation (5.3) is solved to yield the potential field:

$$\hat{\phi} = \mathbf{A}^{-1}(\sigma)q. \quad (5.4)$$

A vector of electric potential values for the cells in the model is the result of Eq. (5.4). Potential differences can be calculated across each measurement pair, from the known locations of the survey potential electrodes. To calculate apparent resistivities (ρ_{app}), these measurements are divided by the input current I and then multiplied by a specific geometric factor K for each survey configuration:

$$\rho_{app} = \frac{\Delta\hat{\phi}}{I} K. \quad (5.5)$$

The geometric factor (K) depends on the arrangement of the four electrodes and corresponding acquisition geometry (i.e., depends on the distance between each electrode and the measurement). If there is no topography, the conventional formula of K can be used to calculate the apparent resistivity (Loke, 2018):

$$K = \frac{2\pi}{\frac{1}{r_{C1-P1}} + \frac{1}{r_{C1-P2}} + \frac{1}{r_{C2-P1}} + \frac{1}{r_{C2-P2}}}, \quad (5.6)$$

where r_{C1-P1} is the distance between current electrode C1 and potential electrode P1, r_{C1-P2} is the distance between current electrode C1 and potential electrode P2, r_{C2-P1} is the distance between current electrode C2 and potential electrode P1, and r_{C2-P2} is the distance between current electrode C2 and potential electrode P2.

5.2.4 Comparison and stochastic model optimization

The model optimization is achieved by the maximization of an objective function that measures the similarity coefficient between synthetic and observed FDEM and ERT data (Eq. 5.7 and 5.10, respectively). For the FDEM data, that is based on the GEMI method (Chapter 2). The similarity coefficient (S) is calculated per coil configuration, t_{coils} (i.e., the distance between transmitter and receivers and the orientation of the coils), between the N_s synthetic IP and QP responses obtained for each pair of EC and MS models and the corresponding observed IP and QP data:

$$S^{j,t} = \frac{2 * \sum_{s=1}^N (x_s^t * y_s^{j,t})}{\sum_{s=1}^N (x_s^t)^2 + \sum_{s=1}^N (y_s^{j,t})^2}, \quad j = 1, \dots, N_s \quad \text{and} \quad t = 1, \dots, t_{coils}, \quad (5.7)$$

where x and y are the real and synthetic QP (or IP) data with N samples. The negative values of S are truncated at zero. This metric avoids an objective function with two terms, and the need of user-defined parameters to weight each term. S is not computed for the entire series of data but along a set of non-overlapping windows, which are randomly created at the beginning of each iteration with different sizes and positions. Each S computed for each grid location is then weighted in depth by the normalized sensitivity curves of each coil configuration

resulting from the FDEM forward model (Eq. 5.8 and 5.9). In the synthetic and real case applications of the joint inversion method proposed, assumption have been made that EC is directly dependent on QP and MS on IP (alternative assumption might be considered)

$$S_{EC}^{j,t} = sens_{EC}(z)^{j,t} * \mathbf{S}^{j,t}, \quad j = 1, \dots, N_s \text{ and } t = 1, \dots, t_{coils}, \quad (5.8)$$

$$S_{MS}^{j,t} = sens_{MS}(z)^{j,t} * \mathbf{S}^{j,t}, \quad j = 1, \dots, N_s \text{ and } t = 1, \dots, t_{coils}, \quad (5.9)$$

where $sens_{EC}$ and $sens_{MS}$ are the sensitivity analysis of each FDEM data at each location within the inversion grid. A more detailed description of the model optimization using FDEM data can be found in Section 2.2.3.

For the ERT data, the similarity coefficient ($\mathbf{S}_{ERT}$) is calculated between the N_s apparent resistivity computed from each EC model using the forward model described in Section 5.2.3, and the observed ERT data, using a non-overlapping moving window that visits all the inversion grid locations:

$$\mathbf{S}_{ERT}^j = \frac{2 * \sum_{s=1}^N (\mathbf{x}_s * \mathbf{y}_s^j)}{\sum_{s=1}^N (\mathbf{x}_s)^2 + \sum_{s=1}^N (\mathbf{y}_s^j)^2}, \quad j = 1, \dots, N_s, \quad (5.10)$$

where $\mathbf{x}$ and $\mathbf{y}$ are the observed and synthetic apparent resistivity, respectively. The width and height of the moving window is randomly generated at the beginning of each iteration to avoid biasing the results from iteration to iteration. $\mathbf{S}_{ERT}$ is bounded between -1 and 1, but negative values are truncated at zero.

The selection of the maximum similarity coefficient in both data domains is performed after computing a linear interpolation between $\mathbf{S}_{ERT}$ and all the S_{EC} , for each coil configurations used, at each location within the inversion grid. The samples of EC, corresponding to a given EC geostatistical realization, that originated the maximum similarity coefficient interpolated between the two methods, are stored together with the maximum similarity coefficient in two auxiliary arrays. These arrays are used as conditioning information in the subsequent iteration. The samples of MS corresponding to a given MS geostatistical realization that originated the highest $S_{MS}^{j,t}$ are also stored in an auxiliary volume.

The auxiliary volumes of EC and MS, and the corresponding similarity coefficients volumes, are used as secondary variable for the stochastic sequential co-simulation of a new set of EC and MS models in the subsequent iteration. The magnitude of the maximum similarity coefficient determines the variability of the new ensemble of EC and MS models. The higher the maximum similarity coefficient is, the less variable the new ensembles at each location within the inversion grid will be. For locations associated with $S \sim 1$ the new ensemble of co-

simulated models of EC (and MS) will be similar to the ones corresponding to the auxiliary volumes. This model update approach ensures the convergence of the geostatistical FDEM data inversion from iteration to iteration. The proposed iterative geostatistical joint inversion method for FDEM and ERT data may be summarized in the following sequence of steps (Figure 5.1):

- i) Simulation of two ensembles of N_s models of EC and MS given borehole data and a calibrated variogram model computed from these borehole data, with stochastic sequential simulation (Soares, 2001) and co-simulation with joint probability distributions (Horta and Soares, 2010);
- ii) Calculation of the N_s synthetic FDEM data for each pair of models simulated in i) using a FDEM forward model. In the application examples shown below we use the 1D approximation proposed by Hanssens et al. (2019);
- iii) Calculation of the N_s synthetic ERT data for each EC model simulated in i) using a ERT forward model. In the application examples shown below we use the 2D approximation proposed by Pidlisecky and Knight (2008);
- iv) Compute the local S_{EC} and S_{MS} between true and synthetic FDEM data weighted in depth by the normalized sensitivity curves of each coil configuration resulting from the FDEM forward model;
- v) Compute the local S_{ERT} between true and synthetic ERT data;
- vi) Compute the maximum similarity coefficient for EC by interpolating S_{ERT} and S_{EC} at each location within the inversion grid;
- vii) Build four auxiliary volumes by selecting the EC and MS traces (i.e., vertical column of grid cells) that ensure the highest S from each property at a given iteration. Store the corresponding S values;
- viii) Generate a new ensemble of EC and MS models using co-DSS and the auxiliary volumes resulting from vii) as secondary variables. All models of EC and MS generated during the iterative geostatistical joint inversion are conditioned locally by existing borehole data for EC and MS. They reproduce the global marginal and joint distributions between EC and MC as inferred from the borehole data and a pre-defined spatial continuity pattern as imposed by a variogram model;
- ix) Iterate and repeat steps ii-viii, while the global convergence of the method reaches a pre-defined threshold of maximum similarity.

The proposed iterative geostatistical joint inversion method of FDEM and ERT data is flexible and can be parameterized for all possible coil configurations that are considered in the FDEM survey and/or the configurations possible in a ERT survey. Alternative forward models can be parametrized.

5.3 Synthetic case application

5.3.1 Data set description

A realistic 3D synthetic data set (Narciso et al., 2022) was used as benchmark for the proposed iterative geostatistical joint inversion methodology. The data set was modelled based on geological samples collected at a mine tailing disposal site in Portugal (Panasqueira) and laboratory measurements of porosity and particle density. From these physical properties and using stochastic sequential simulation (Deutsch & Journel, 1998), we generated a three-dimensional synthetic porosity subsurface model, with a dimension of 150 by 200 by 4 meters with a cell size of 0.5 m by 0.5 m by 0.1 m. Water content was then generated with stochastic sequential co-simulation (Deutsch & Journel, 1998) conditioned to the porosity model. EC and MS models were then derived from these 3D physical property models and the Archie equation (Archie, 1942). From the EC and the MS models, nine locations were selected to represent synthetic boreholes, that were used to condition the inversion procedure and to model the spatial structure imposed via the variogram model.

The corresponding observed FDEM data were obtained using a 1D forward model (Hanssens et al., 2019) and replicating a commonly used sensor for this type of near-surface surveys. More particularly, data from a DUALEM-421S sensor were mimicked, including two loop-loop coil orientations, a horizontal coplanar (HCP) and a perpendicular one (PRP), and 3 offsets per coil orientation, 1, 2 and 4 meters for HCP and 1.1, 2.1 and 4.1 meters for PRP.

The observed apparent resistivity data along the transect randomly selected and used as true ERT data during the application of the joint inversion method, was calculated considering a Wenner-Schlumberger acquisition array (e.g., Loke 2018) composed by 412 electrodes spaced 0.5 m and solving the 2.5D forward model to yield the potential field, following Pidlisecky and Knight (2008). The same forward models used to calculate the observed FDEM and the true apparent resistivity field was used as part of the inversion. Therefore, in this application example we assume that no uncertainty is considered in the forward model, which might be a strong assumption in real case applications.

5.3.2 Results

The results obtained with the proposed iterative geostatistical joint inversion method for FDEM and ERT data are illustrated for the 2D transect of EC and MS that intersects nine boreholes (Figure 5.2).

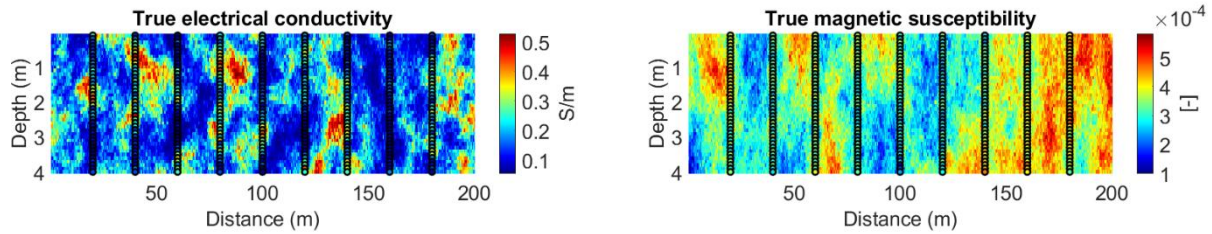


Figure 5.2 True electrical conductivity (left) and magnetic susceptibility (right) and location of the 9 boreholes providing conditioning data of the iterative geostatistical joint inversion method.

The convergence of the iterative geostatistical joint inversion method with respect to the reproduction of the model parameters is assessed by calculating the pointwise mean models of EC and MS computed from all the realizations generated at the last iteration, which is equivalent to the maximum a posteriori model from a Bayesian geophysical inversion (Bober, 2020). Although the reproduction of the true small-scale heterogeneities cannot be evaluated by these models due to the smoothing effect of the e-type mode, the predicted and true EC and MS models show similar large-scale spatial patterns and are sensitive to the high and low values of true EC and MS (Figures 5.3b and 5.4b).

The small-scale differences between the true and predicted EC and MS models and the relationship with depth and sensitivity loss of the predicted solutions are also observed by calculating the pointwise variance models from the ensemble of EC and MS models generated at each iteration (Figures 5.3c, 5.3d, 5.4c, 5.4d). As expected, in the first iteration the spatial distribution of the variance for both properties is only dependent on the distance to the locations of the borehole data. The pointwise variance models of EC and MS computed from models predicted during the last iteration of the geostatistical joint inversion, shows the influence of each geophysical data and the sensitivity provided by the FDEM forward model (Figures 5.3d, 5.4d). With the observed FDEM and ERT data assimilated in the models at the last iteration, the spatial distribution of the pointwise variance model of EC shows lower variance values along all the grid model and an influence of the sensitivity provided by the FDEM forward model together with the configuration of the ERT survey (Figure 5.3d). The predictions about MS are less sensitive at depth, with the lack of ERT data constraint in the MS predicted models, along with the shallow sensitivity provided by the FDEM forward model (higher dependence on the coil configurations used). This is demonstrated in spatial pattern of the pointwise variance MS

model as the higher variance values are mainly located in the deeper part of the model (Figure 5.4d).

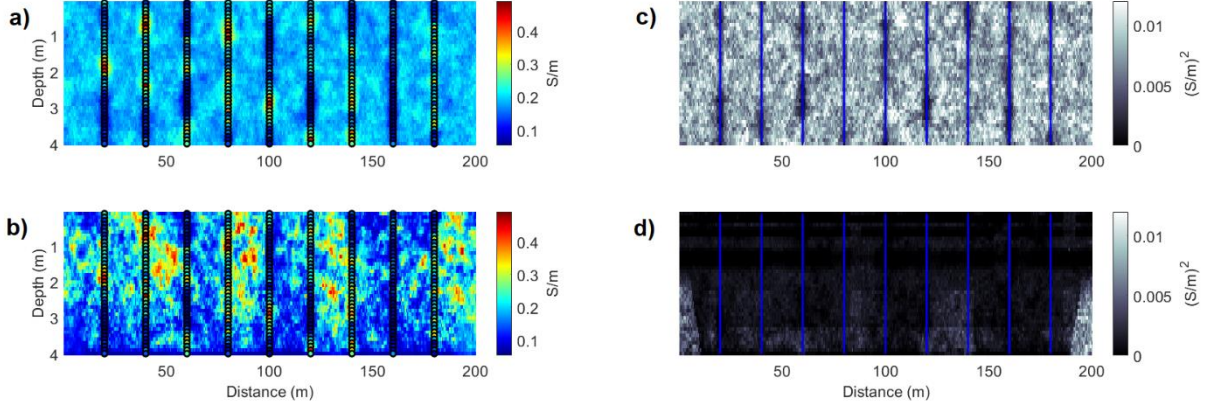


Figure 5.3 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration. Vertical dot and blue lines indicate the location of the borehole data.

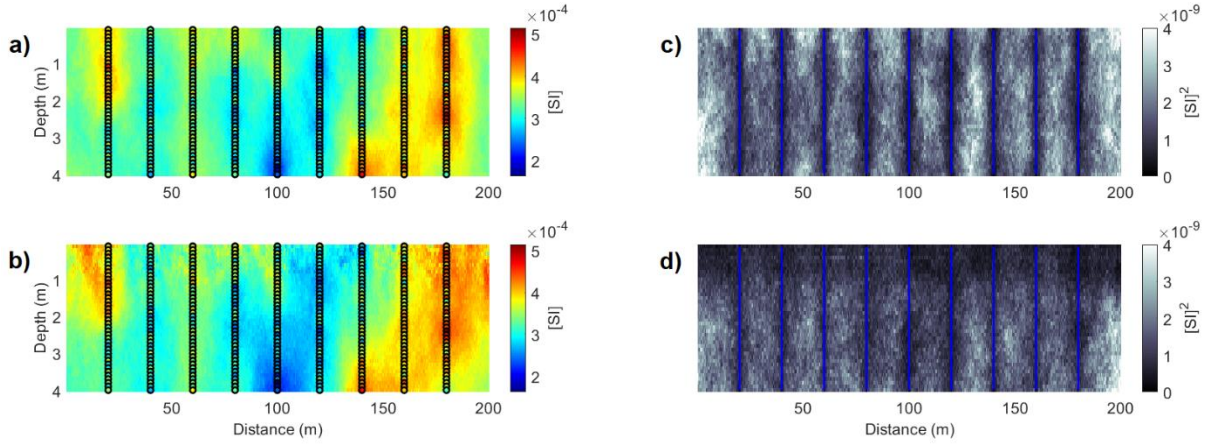


Figure 5.4 a) Pointwise mean of all the MS models computed in the first iteration; b) pointwise mean of all the MS models computed in the last iteration; c) pointwise variance of all the MS models computed in the first iteration; d) Pointwise variance of all the MS models computed in the last iteration. Vertical dot and blue lines indicate the location of the borehole data.

The proposed iterative geostatistical joint inversion method reproduces the true EC models and converge to the true solution from the first iteration (Figure 5.5a) to the last iteration (Figure 5.5b), with the residuals between the true EC model and one EC realization reducing along the iterations (Figures 5.5c and 5.5d).

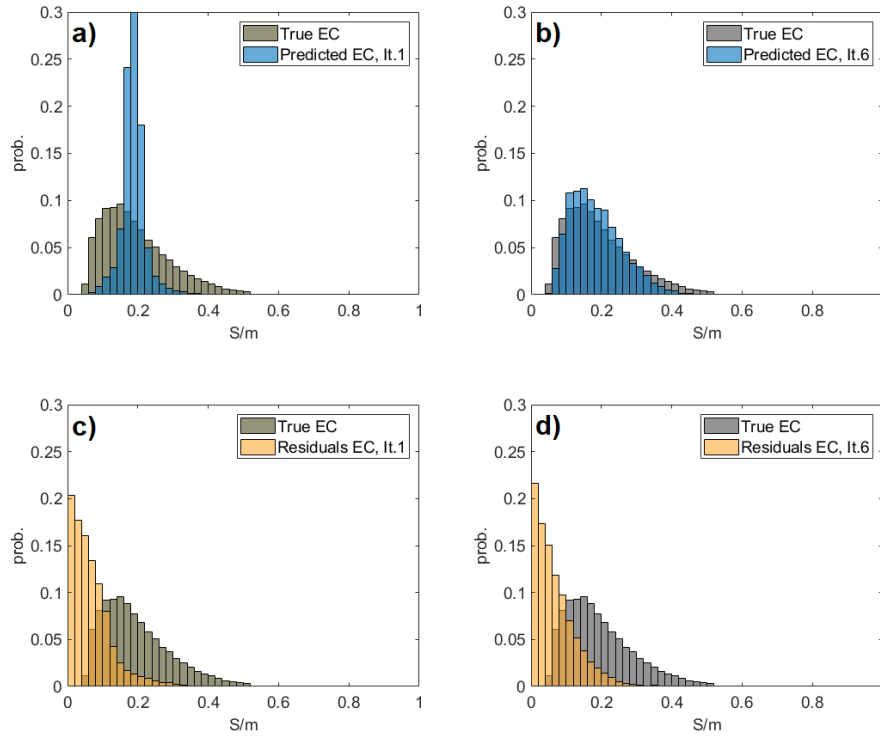


Figure 5.5 a) Histograms of the true EC and the pointwise mean of all the EC models computed in the first iteration; b) Histograms of the true EC and the pointwise mean of all the EC models computed in the last iteration; c) Histograms of the true EC and the residuals of the pointwise mean of all the EC models computed in the first iteration; d) Histograms of the true EC and the residuals of the pointwise mean of all the EC models computed in the last iteration.

The misfit between observed and predicted FDEM data, for both IP and QP components is shown in Figures 5.6 to 5.9. For all coil configurations considered, the match between observed and predicted IP and QP responses increases from the first to the last iteration. The uncertainty envelope, as represented by the synthetic response of the ensemble of models in each iteration, narrows and encloses the observed IP and QP data as the iterative procedure advances. Although the uncertainty envelope of all coil configurations in the last iteration well encloses the true FDEM data, a better match is reached in QP responses and in smaller coil distances. This is due to a more stable signal in QP responses and a higher sensitivity to small-scale heterogeneities at shallow depths when the coils are closest to each other, but also to the influence of ERT data in a faster convergence of the EC inversion to the true solution. EC is directly dependent on QP and MS on IP is shown here.

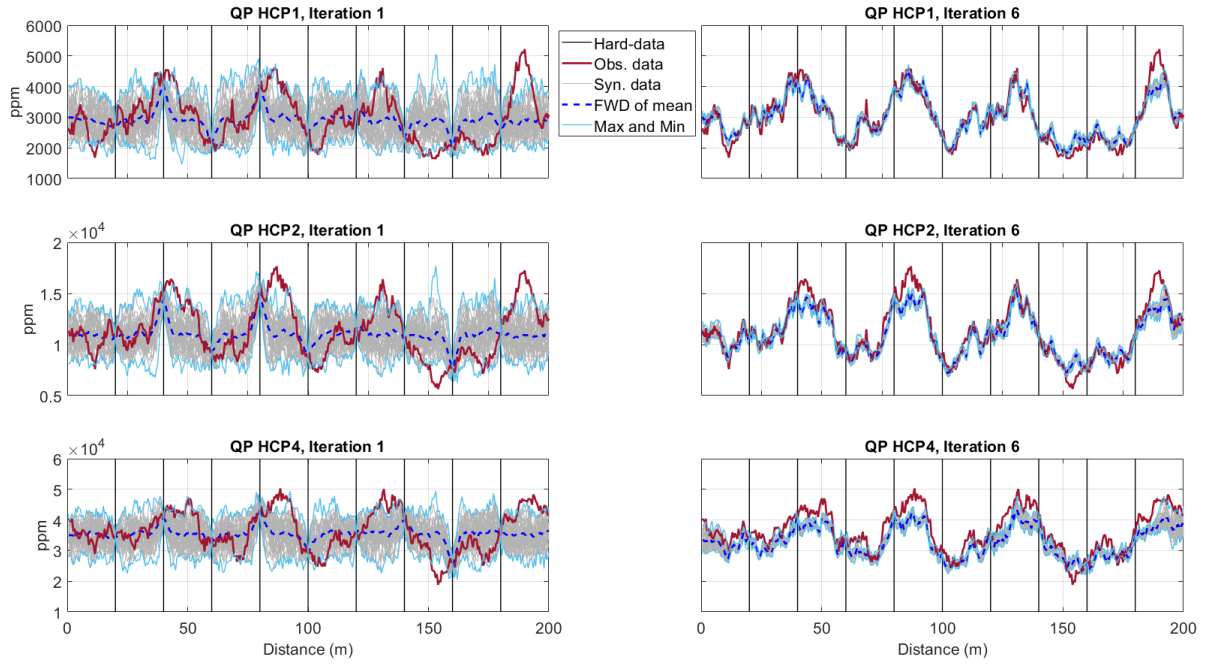


Figure 5.6 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.

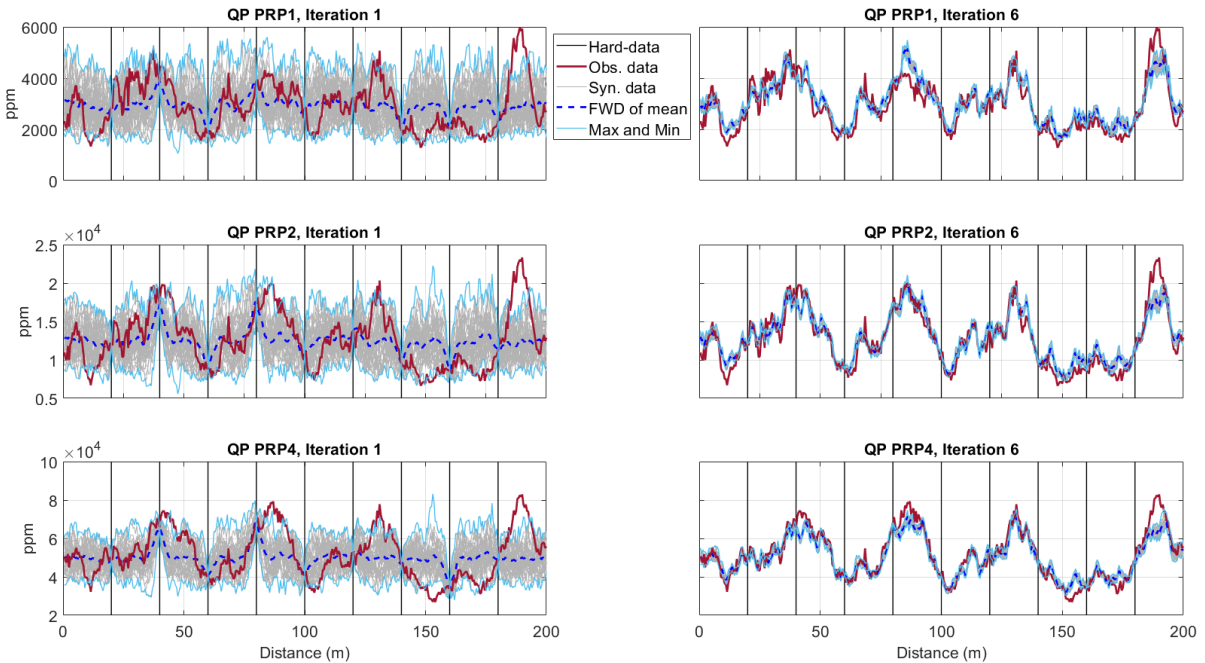


Figure 5.7 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for perpendicular coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and

maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.

As expected, the predicted QP and IP responses at the borehole locations are exactly reproduced as the predicted EC and MS models are locally conditioned by the borehole data.

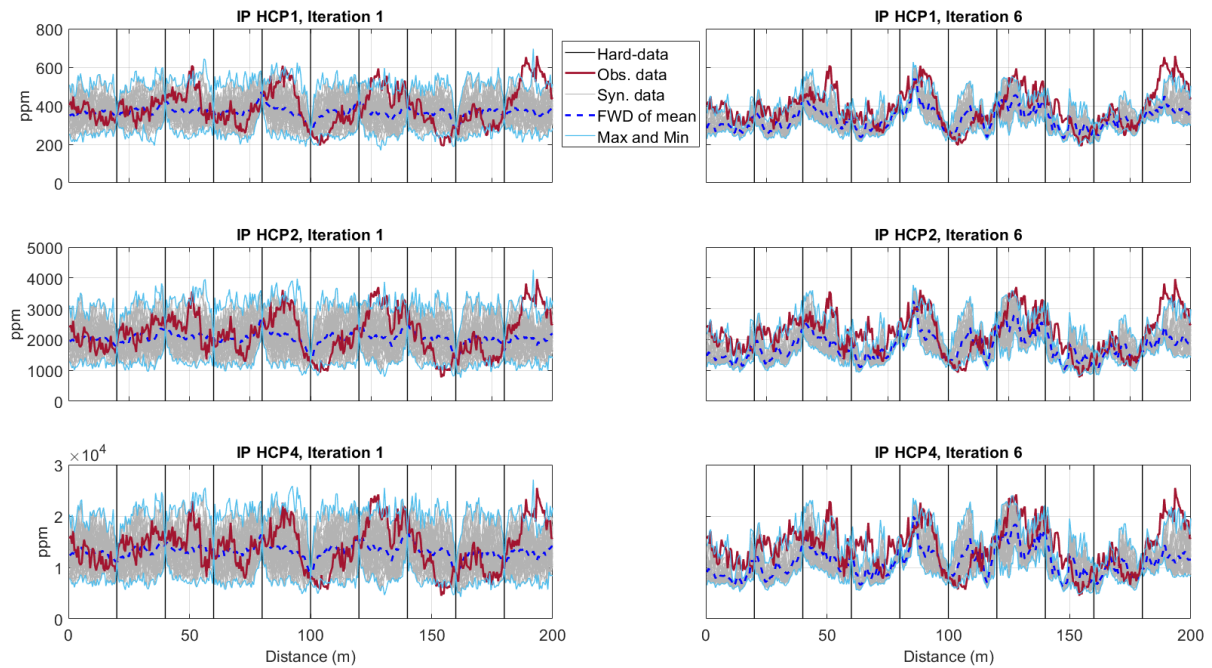


Figure 5.8 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.

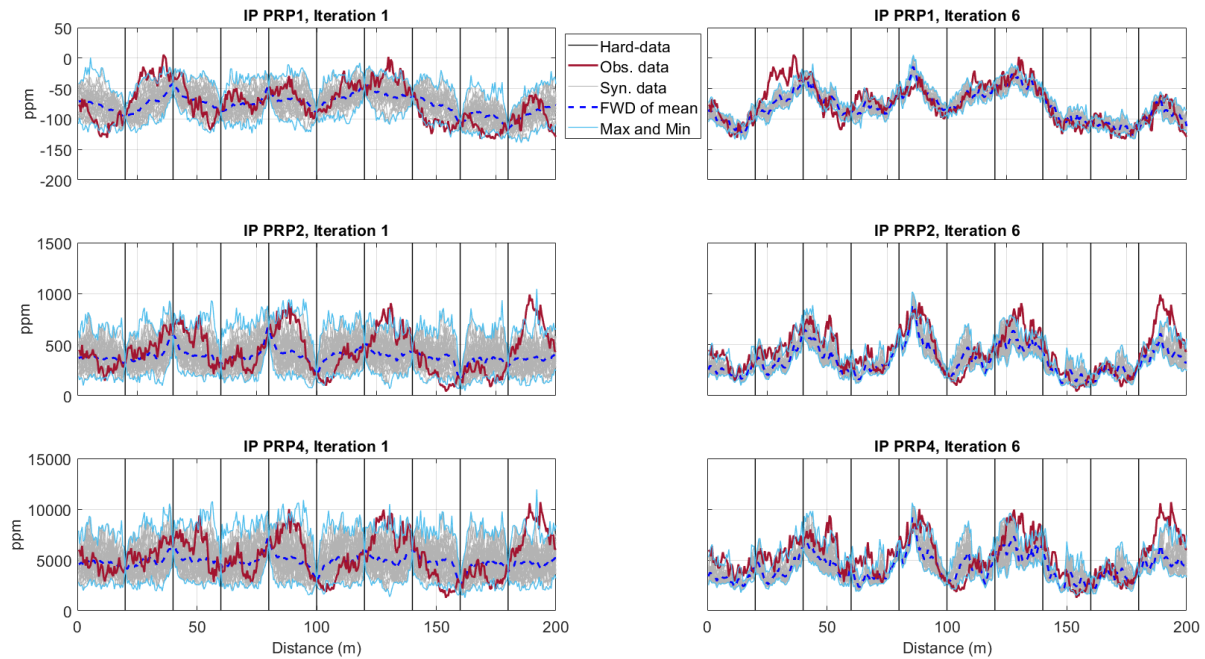


Figure 5.9 Comparison between observed (red line) and predicted IP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC and MS models generated at a given iteration (dashed dark blue line) for perpendicular coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical lines indicate the location of the borehole data.

The misfit between observed and predicted ERT data can be assessed in Figures 5.10. The figures show how the match between observed and predicted apparent resistivity increases from the first to the last iteration. The predicted and observed apparent resistivity show similar large-scale spatial patterns and match the high and low values (Figures 5.10a and 5.10c). However, it is observed an increase in small-scale differences between the true and predicted apparent resistivity, particularly below the 3 m depth, arising from sensitivity loss of ERT data at depth.

The reproduction of the observed apparent resistivity from the first to the last iteration of the proposed iterative geostatistical joint inversion method can also be assessed by the residuals computed between the observed data and the predicted data from a single geostatistical realization generated during the last iteration (Figures 5.10d and 5.10e).

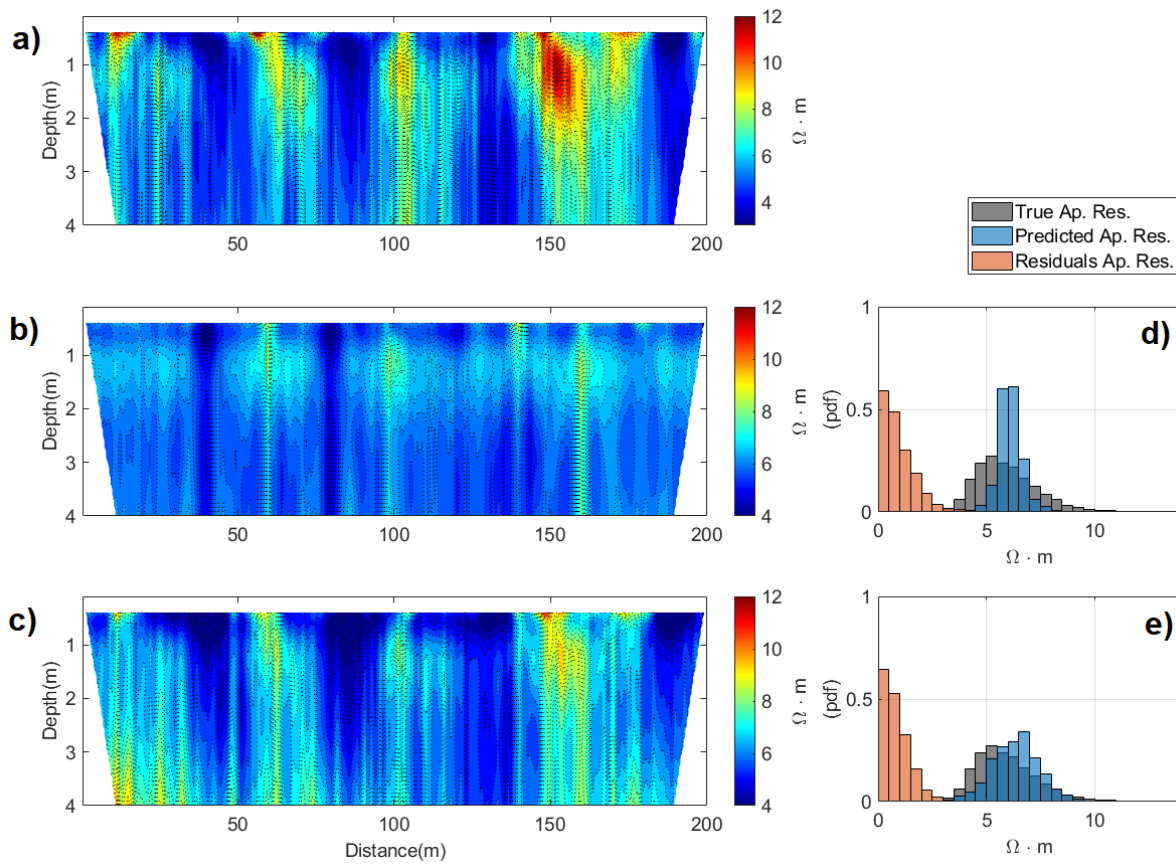


Figure 5.10 Comparison between observed and predicted apparent resistivity data; a) Observed apparent resistivity data; b) Predicted apparent resistivity data computed in the first iteration; c) Predicted apparent resistivity data computed in the last iteration; d) Histograms of the observed apparent resistivity data, one predicted apparent resistivity data computed in the first iteration and the corresponding residuals between both; e) Histograms of the observed apparent resistivity data, one predicted apparent resistivity data computed in the last iteration and the corresponding residuals between both.

5.4 Real case application

5.4.1 Data set description

The proposed iterative geostatistical joint inversion method of FDEM and ERT data was applied to a real data set obtained in the nature reserve of Doelpolder Noord, situated north of Antwerp harbour, on the left bank of the Scheldt River (Verhegge et al., 2016a). The site is characterized by topsoils that range from heavy clay to clay with moderately bad to bad drainage, covering Pleistocene sands with the top between 4- and 9-meter depth. Manual coring and geophysical data identified a river dune buried between 2 and 6 m deep, flanking a large depression with the base reaching up to 9 m below the surface. The dune is characterized

by low values of FDEM response and higher resistance and is less conductive compared to the surrounding subsurface (Figures 5.11 and 5.15).

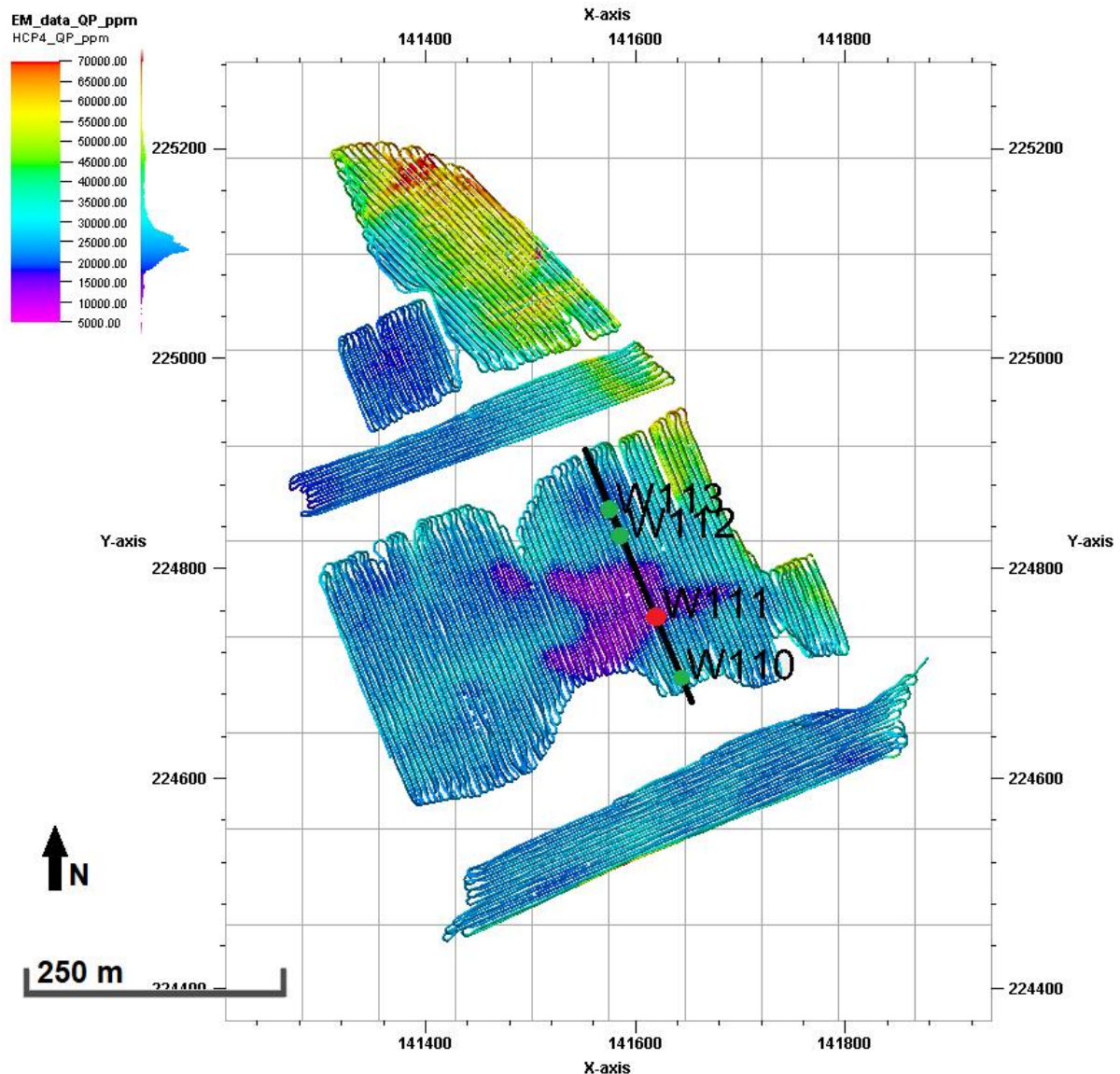


Figure 5.11 Map of the FDEM survey with QP data (ppm) for the HCP coil configuration with 4 m offset. The black line represents the location of the transect of Figures 5.12 to 5.15. Green points represent the locations of the available borehole data (blind well in red). Coordinates in Belgian Lambert'72 coordinate system.

FDEM data was collected on August 2013 using a DUALEM-421S (DUALEM Inc., Milton, Canada) low-frequency domain EMI sensor, recording both IP and QP response to an induced field with a frequency of 9 kHz (Verhegge et al., 2016a). The data were registered in HCP coil configuration with 1, 2 and 4 m separation between coils, and PRP configuration with 1.1, 2.1 and 4.1 m between coils. The survey was carried out using a quad-pulled sled, with the sensor elevated 16 cm from the surface, along parallel lines every 3 m, and the responses registered at 8 Hz while driving 7-8 km/h. The pre-processing of the FDEM data included: i) the correction

for the spatial offsets between the position and sensor data, following the procedure described in Delefortrie et al. (2016); ii) the correction for signal drift – a relative calibration, following the procedure in Delefortrie et al. (2014b); and iii) an absolute calibration per coil configuration to eliminate the presence of signal offsets, comparing the forward modelled responses at locations where in-situ measurements of EC and MS were available with the measured FDEM responses.

Electrical resistivity data were collected using an AGI Supersting R8 with an inverse Schlumberger configuration, allowing the use of multi-channel possibilities and increasing the survey speed (Verhegge et al., 2016b). The electrodes were positioned 2 m apart to obtain an estimated 1 m spatial resolution (Baines et al., 2018). The ERT data were despiked and concatenated using RES2Dinv software (Loke, 2018). No topographical correction was needed because the relatively flat topography.

A series of conductivity cone penetration tests (CPT-C), i.e., direct measurements of electrical conductivity, were carried out using a dielectric cone, a frequency-domain method at 20MHz (Hilhorst, 1998), reaching the 10-meter depth each one.

To test the iterative geostatistical joint inversion method, a 260 meter transect was selected containing maximal subsoil variability including the top of a Pleistocene micro-sand ridge (buried about 2 m deep), FDEM data, a 2D ERT data profile and 4 CPT-C (Figure 5.11). Although the iterative geostatistical joint inversion can predict EC and MS inversion models in the same inversion application, and since MS was not measured by direct measurements of borehole data, the inverse modelling of MS and the prediction of IP component will not be assessed in the application of the iterative joint inversion method to this real case data set.

5.4.2 Results

To assess the performance of the proposed iterative geostatistical joint inversion method of FDEM and ERT data, Figures 5.12 show the pointwise mean and variance models of the ensemble of EC models predicted at the first and the last iterations. The pointwise variance of EC models demonstrates the influence of including the sensitivity of the forward model to the model parameters, increasing in depth as the sensitivity of the FDEM decreases, and the constraint of the ERT data and survey configuration in the predicted EC models, decreasing the variance of the ensemble, in line with the results achieved in the synthetic case application. With the FDEM coil configurations used, the constraint of the FDEM data in the inversion results is limited to approximately 5 meters depth in the predicted EC models. Two distinct regions can be clearly observed, a shallower one with lower electrical conductivity values and a deeper one more conductive.

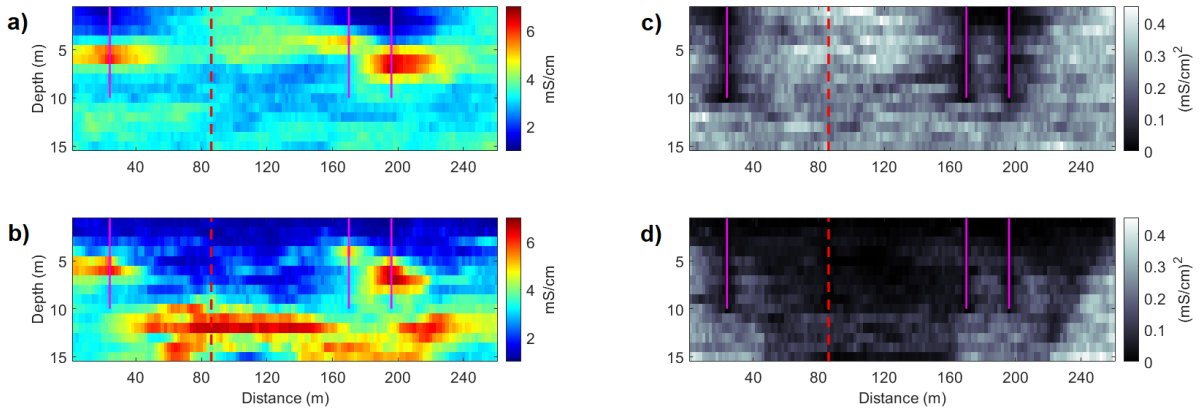


Figure 5.12 a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the last iteration. Vertical magenta lines indicate the location of the borehole data. The vertical red dashed line represents the location of the blind well.

The performance of the proposed methodology can also be assessed by the match between observed and predicted FDEM data (Figures 5.13 and 5.14).

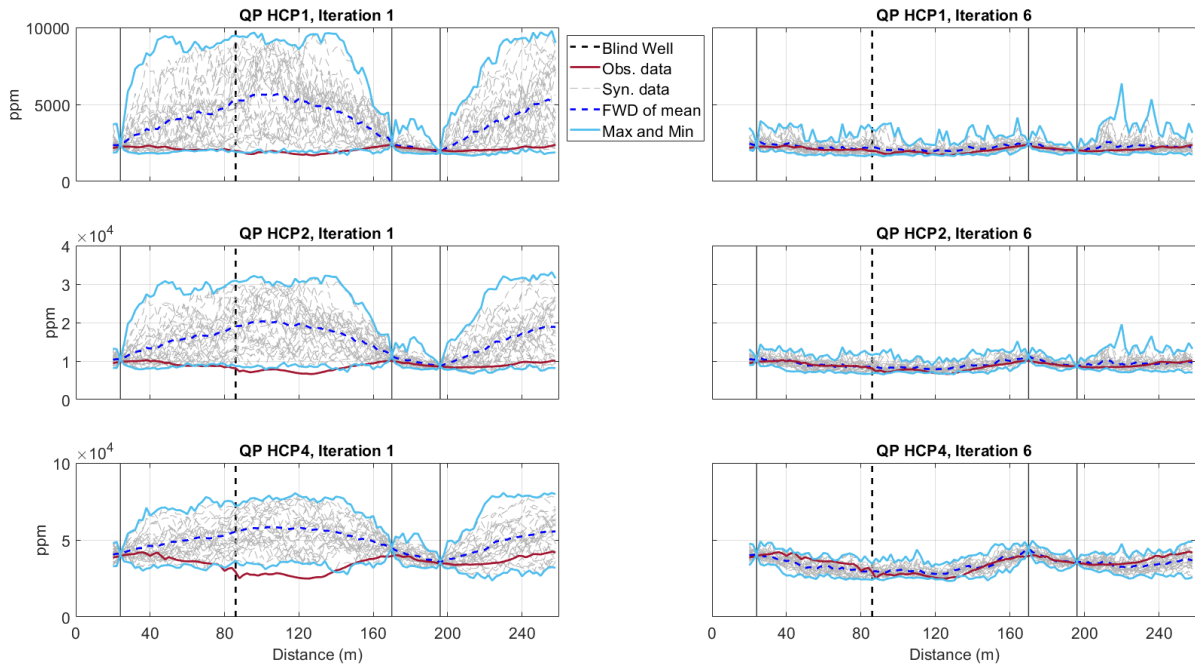


Figure 5.13 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (HCP orientation with 1, 2 and 4 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical grey lines indicate the location of the borehole data. The vertical dashed line represents the location of the blind well.

The predicted FDEM responses were calculated from the ensemble of all models generated during the first and last iterations, for all coil configurations. The increasing convergence from iteration-to-iteration is illustrated by the envelope of the synthetic FDEM responses that gets narrower and closer to the observed data as the iterative procedure moves forward. In general, the predicted QP signal component of the FDEM data for all coil configurations match the recorded field data.

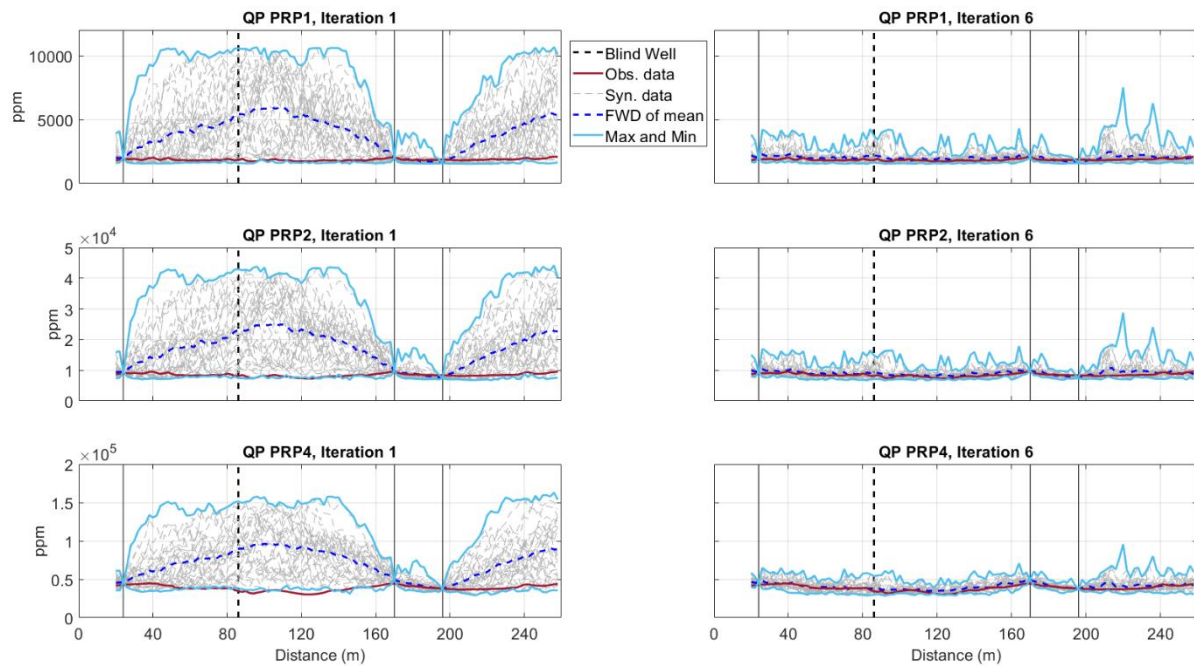


Figure 5.14 Comparison between observed (red line) and predicted QP data for all models generated at a given iteration (grey), the predicted FDEM data from the pointwise mean of the EC models generated at a given iteration (dashed dark blue line) for horizontal coil configurations (PRP orientation with 1.1, 2.1 and 4.1 m offset). The light blue lines represent the minimum and maximum FDEM values predicted at a given iteration. In the left column the predictions at the end of the first iteration are represented and in the right column the predictions at the end of the last iteration are represented. Vertical grey lines indicate the location of the borehole data. The vertical dashed line represents the location of the blind well.

The misfit between observed and predicted ERT data can be assessed in Figures 5.15. The match between predicted and observed apparent resistivity increases from iteration to iteration and present similar large-scale spatial patterns, reproducing the high and low values (Figures 5.15a to 5.15c). The spatial reproduction of the Late Glacial (river) dune, characterized by high values of apparent resistivity, is also detected in the predicted apparent resistivity.

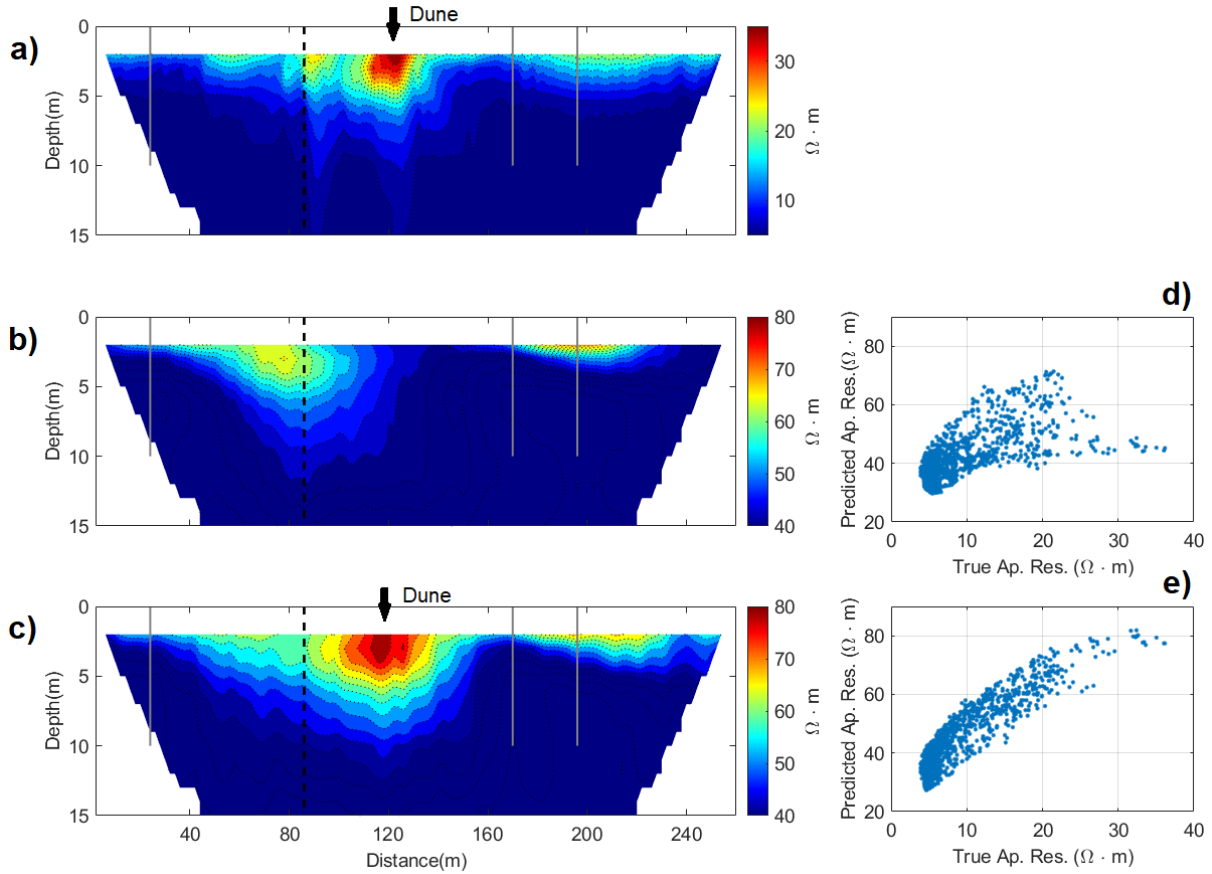


Figure 5.15 Comparison between observed and predicted apparent resistivity data; a) Observed apparent resistivity data; b) Predicted apparent resistivity data computed in the first iteration; c) Predicted apparent resistivity data computed in the last iteration; d) Biplot between the observed apparent resistivity and the predicted apparent resistivity data computed in the first iteration; e) Biplot between the observed apparent resistivity and the predicted apparent resistivity data computed in the last iteration. Vertical grey lines indicate the location of the borehole data. The vertical dashed lines represent the location of the blind well.

The quality of the inversion results and the convergence of the data in both domains can be assessed by computing same correlation metrics (e.g., the global Pearson correlation coefficient (CC)) between the observed ERT data and the predicted ERT data computed from each realization, and the root-mean-square errors (RMSE) between the observed FDEM data and the predicted QP component from each realization (Figure 5.16). In both data domains, the convergence to the observed data is achieved, with CC of 0.90 in ERT data in the last iteration, and lower RMSE values of FDEM data from iteration to iteration.

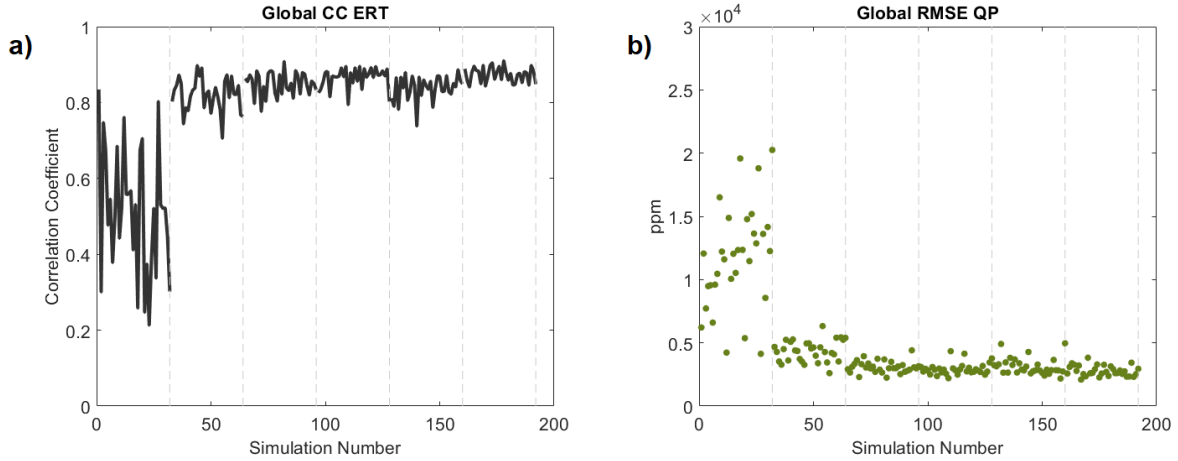


Figure 5.16 a) Global correlation coefficient of the apparent resistivity data computed in all iterations; b) Global root-mean-square error of the QP data computed in all iterations.

5.5 Discussion

The proposed iterative geostatistical joint inversion method predicts near-surface EC and MS models from survey FDEM and ERT data. The inversion method is based on geostatistical simulation and co-simulation as model perturbation and stochastic update techniques. Therefore, the predicted models can be conditioned locally to existing borehole data and a spatial continuity pattern as described by a variogram model. The perturbation of the model parameters at each iteration leverages the sensitivity analysis provided by the FDEM forward model (i.e., the assimilation of the recorded FDEM data accounts for the sensitivity in depth per property as provided by the forward model) and the local predicted by the ERT data at the deeper depths.

The proposed joint inversion method is based on a 1D FDEM forward model and a 2D ERT forward model. Using a one- and two-dimensional forward models represents a limitation when computing the electromagnetic and direct-current resistivity response, particularly in highly complex geological settings, as the propagation of the electromagnetic field and the injected electrical current into the subsurface flows three-dimensionally through preferential paths that could bypass some structures, imposing artifacts in a two-dimensional representation. In these cases, alternative three-dimensional forward models could be used, but the computational costs of the proposed methodology would increase. This hard assumption is somehow alleviated in the proposed methodology as the model perturbation is global for the entire grid at once (i.e., in 2D or 3D depending on the data availability).

The synthetic application example illustrates the potential of the proposed joint inversion method to predict a reliable near-surface EC model. However, this is a relatively simple

example as the same forward model used to create the observed data was used in the inversion workflow. The potential of the joint inversion method compared to simple geophysical inversion methods can be assessed by testing the GEMI method (Chapter 2) in the same synthetic data set used with the joint inversion method, using the same inversion and modeling parameterization along with the same FDEM data. The EC realizations predicted from the GEMI method capture the spatial continuity of the true EC model (Figure 5.2) and reproduce the high and low local values of EC (Figure 5.17b). However, from the pointwise variance of the EC ensemble in the first and last iteration, a higher spatial uncertainty and small-scale variability of EC is predicted below the 2 m depth (Figure 5.17d), in comparison to the EC predicted models from the joint inversion method (Figure 5.3d). The increase in the spatial uncertainty in depth from the EC predicted models computed by the GEMI method is directly related to the sensitivity decrease of FDEM data in depth along with the lack of ERT data to converge the results to the true solution.

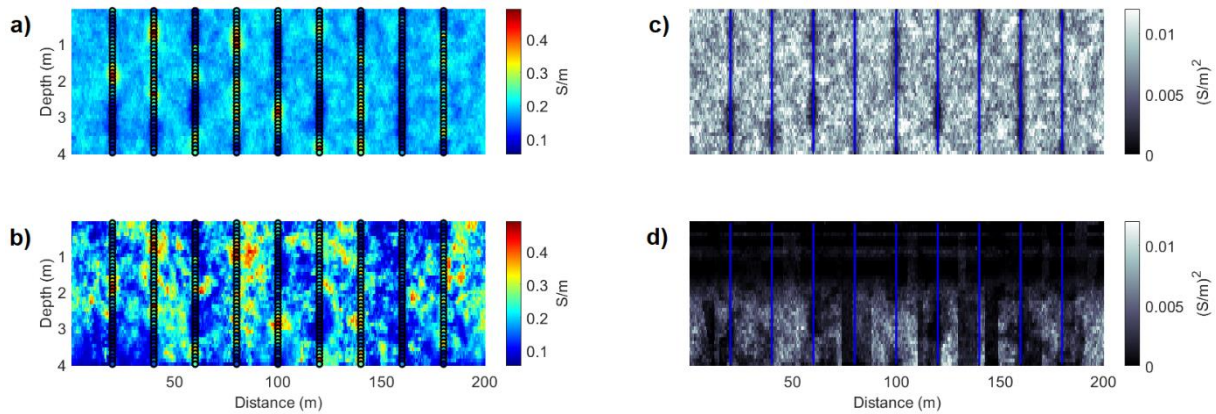


Figure 5.17 a) Pointwise mean of all the EC models computed in the first iteration using GEMI; b) pointwise mean of all the EC models computed in the last iteration using GEMI; c) pointwise variance of all the EC models computed in the first iteration using GEMI; d) Pointwise variance of all the EC models computed in the first iteration using GEMI. Vertical magenta lines indicate the location of the borehole data. Vertical dot and blue lines indicate the location of the borehole data.

The same conclusions arise from computing the residuals between the pointwise mean model computed in the last iteration and the reference model of EC (Figure 5.18). The GEMI method predicted EC models with more spatial uncertainty and difference to the true solution. The residuals from the pointwise mean computed from the EC ensemble models predicted from the joint inversion method are also lower than the ones predicted by the GEMI method (Figures 5.18c and 5.18d).

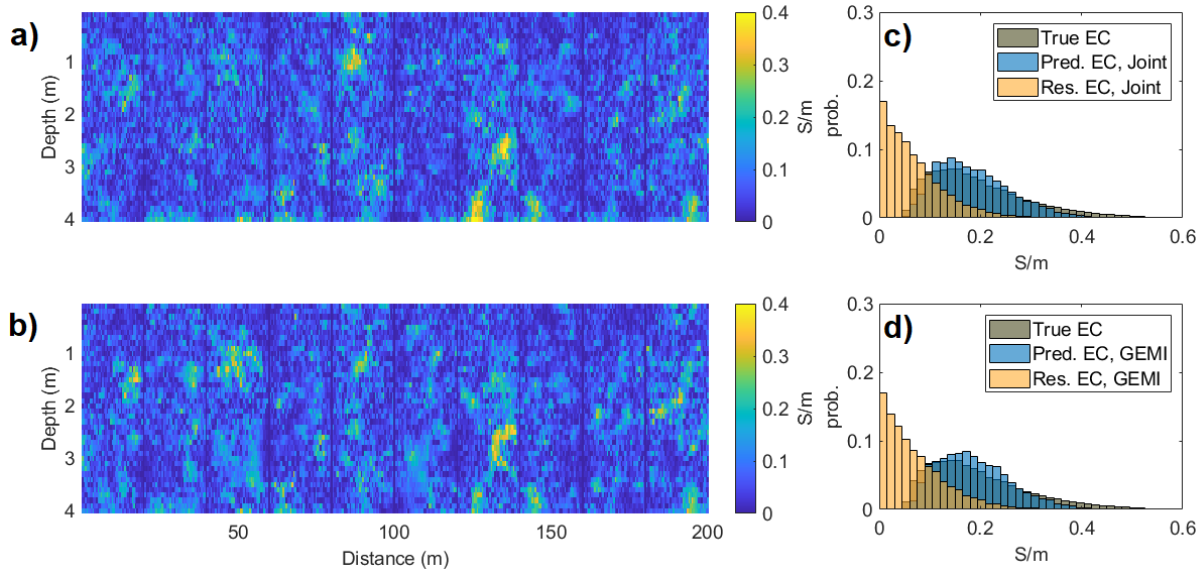


Figure 5.18 Two-dimensional difference between the pointwise mean model computed in the last iteration and the reference model of EC (Figure 5.2), using: a) iterative joint geostatistical inversion method; b) GEMI method. c) Histograms of the true electrical conductivity, the pointwise mean of all the EC models computed in the last iteration and the corresponding residuals between both, using the iterative joint geostatistical inversion method; d) Histograms of the true electrical conductivity, the pointwise mean of all the EC models computed in the last iteration and the corresponding residuals between both, using GEMI method.

The potential of the joint inversion method can also be assessed by applying the GEMI method in the same real data set of Doelpolder Noord, using the same inversion and modeling parameterization, and removing the same borehole information in the iterative modelling. The predicted EC models from this GEMI application are conditioned to the spatial continuity pattern imposed by a variogram model, but the perturbation of the model parameters is dependent on the sensitivity analysis of the FDEM data, which decreases significantly below the 5 m depth. This effect can be observed by computing the pointwise mean and variance of the EC models predicted by the GEMI method (Figure 5.19). The spatial uncertainty only decreases in the shallower grid locations of the EC models in the last iteration, here the FDEM data perturb the model parameters (Figure 5.19d).

By testing both inversion methods in the real case data set without using one of the boreholes information, we evaluate the predicted EC model locally at the location of the blind well (Figures 5.11 and 5.20). Despite the borehole information does not reaches the full depth of the inversion model, the predicted near-surface properties from the joint inversion method do match the observed one at the last iteration.

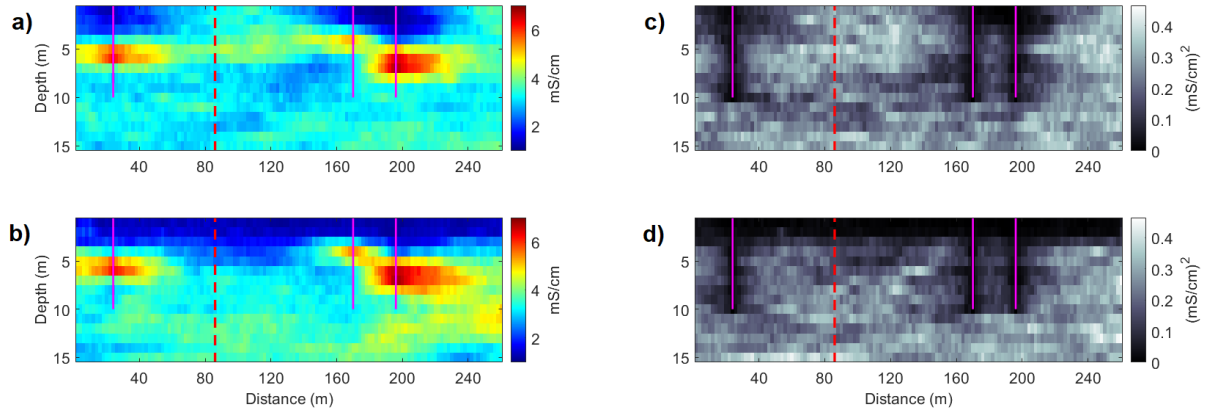


Figure 5.19 Inversion realizations of Doelpolder data set obtained by GEMI method: a) Pointwise mean of all the EC models computed in the first iteration; b) pointwise mean of all the EC models computed in the last iteration; c) pointwise variance of all the EC models computed in the first iteration; d) Pointwise variance of all the EC models computed in the first iteration. Vertical magenta lines indicate the location of the borehole data. The vertical red dashed line represents the location of the blind well.

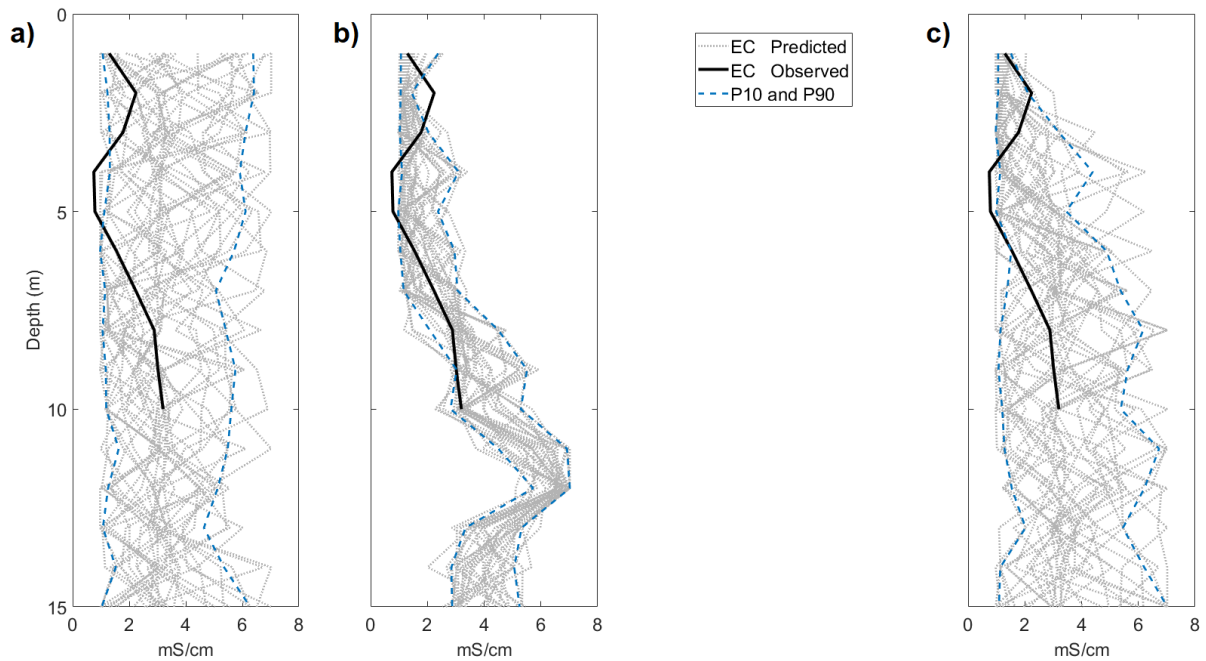


Figure 5.20 True and predicted values of EC along the blind well test for: a) EC during the first iteration obtained by the iterative joint geostatistical inversion method; b) EC during the last iteration by the iterative joint geostatistical inversion method; c) EC during the last iteration by the GEMI method.

A comparison between the predicted EC results in the blind well location computed by both inversion methods, demonstrated that the iterative geostatistical joint inversion method can predict the EC model decreasing the uncertainty at depth, at opposite to the GEMI method below the 5 m depth (Figure 5.20b and 5.20c). The predicted EC values from the joint inversion

method exhibit a linear correlation to the true EC values from the borehole data that was not used in the inversion procedure (Figure 5.21).

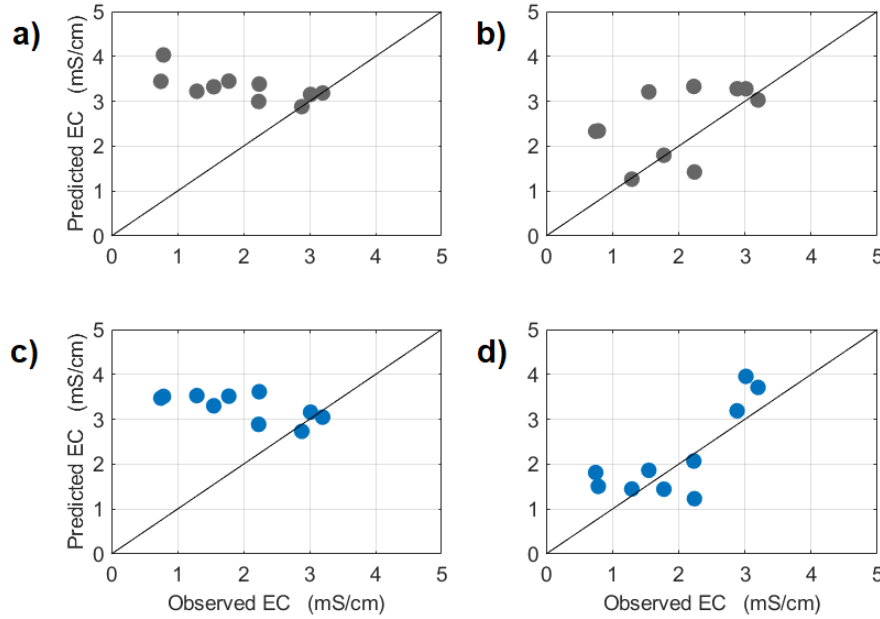


Figure 5.21 Biplot along the blind well test between the observed EC and: a) the predicted EC obtained by GEMI method for the first iteration; b) the predicted EC obtained by GEMI method for the last iteration; c) the predicted EC obtained by the iterative geostatistical joint inversion method for the first iteration; d) the predicted EC obtained by the iterative geostatistical joint inversion method for the last iteration.

5.6 Conclusion

This Chapter introduces an iterative geostatistical joint inversion method that represents a contribution to probabilistic joint inversion of DC resistivity data and FDEM data. The proposed iterative joint inversion method can predict the spatial distribution of EC and MS simultaneously (although the models of MS are only conditioned to the FDEM data).

The predicted FDEM and ERT data computed from each EC realization of the joint inversion match better the observed data of each geophysical method and reproduces better the true electrical conductivities, than the models obtain from the FDEM inversion.

The joint inversion methodology was first validated using a developed synthetic data set rendering realistic spatial distributions of EC and MS and then applied to a real data set containing FDEM data, ERT data and CPT-C (direct measurements of EC). The pointwise mean and variance models of predicted EC from both application examples demonstrated that, not only the joint inversion methodology reproduces exactly the histograms retrieved from the

borehole data, but also presented improvements in the spatial continuity reproduction and uncertainty at depth, when compared to the separated FDEM inversion method.

CHAPTER 6

Conclusions & Future Perspectives

6.1 Conclusions

The main goal of this thesis was to develop and implement an iterative geostatistical geophysical inversion framework able to predict the subsurface spatial distribution of electrical conductivity and magnetic susceptibility at high spatial resolution from FDEM and ERT data. All the methodologies proposed herein were validated in a realistic synthetic data set and applied in real case examples. As main conclusion from these application examples, we can highlight that the geostatistical framework can handle the different spatial resolution from the geophysical and borehole data, that the proposed methodologies could cope with heterogeneous subsurface environments, predicting local small-scale variability, while assessing simultaneously the uncertainty of the predicted models. The main conclusions of the four objectives of the thesis follow below.

Objective one: Realistic synthetic data set. We developed a realistic synthetic data set based on direct and laboratory measurements obtained from samples acquired in a mine tailing. These data can be used to benchmarking different geophysical inversion methods that have the potential to be applied in complex and heterogeneous near-surface environments. The work proposed in Bobe et al. (2019) (Chapter 3) is an illustrative example. This data set proved to be useful to test the sensibility of the proposed inversion methods to discontinuities in the physical properties and to capture their spatial continuity in highly heterogeneous environments. It also was useful to validate the proposed inversion methodologies throughout this thesis and to compare the corresponding predicted models. The data set is publicly available in <http://doi.org/10.5281/zenodo.5116420>

Objective two: Iterative geostatistical inversion of FDEM data. We developed and implemented an iterative geostatistical FDEM inversion methodology (Chapter 2) that allows to simultaneously predict EC and MS and can be applied to characterize complex and heterogeneous near-surface deposits of different types and nature. The proposed method was validated in the 3D synthetic data set developed under objective one, was tested in a real data set containing several archaeological features and strong local IP anomalies and was compared to a probabilistic KEG method and their predicted results. The results show the ability of the proposed method to reproduce the true EC and MS and the predicted FDEM measurements responses well enclosed the true FDEM. The uncertainty of the posterior distributions of EC and MS and the FDEM responses computed from the predicted models can also be assessed, presenting an advantage compared to deterministic FDEM inversion methods.

Objective three: Optimization of FDEM inversion. To improve the computational cost of the developed iterative geostatistical FDEM inversion method, we proposed a FDEM inversion method that performs the inversion in a reduced space without compromising the exploration of the model parameter space. We use a FDEM inversion scheme that combines ES-MDA with RTD and was able to predict the spatial distribution of EC and MS, in both synthetic and real case application examples. In both application examples, the predicted models reproduced the measured EC and MS data while allowing to assess the uncertainty of the predictions. The proposed methodology has the potential to solve large-scale three-dimensional problems in near-surface applications. The code of this method is available at: <https://github.com/theanswer003/ES-RTD-FDEM>

Objective four: Iterative geostatistical joint inversion of FDEM and ERT data. We developed and implemented an iterative geostatistical joint inversion method that couples data from different geophysical methods. The proposed method combines the benefits of the separate inversion methods of small-loop FDEM and direct current resistivity data in a joint inversion framework. Using a joint inversion approach, the perturbation of the joint parameter space represents improvements over the joint interpretation of the separate inversion. Though, most of the joint inversion methods that combine these two geophysical data use deterministic frameworks which require to explicitly weight the influence of the different data types. This work represents a milestone in the probabilistic joint inversion of FDEM and ERT data, as the proposed joint inversion method is, as far as our knowledge go, the first geostatistical joint inversion method of FDEM and ERT data, with the flexibility of application in a significant range of near-surface activities. From the application examples shown herein, we concluded that the proposed joint inversion method presents benefits over the separate inversion methods, increasing the accuracy of the predicted EC subsurface models with a better reproduction of the true EC models while reducing the uncertainty at the local small-scale, particularly at depth.

6.2 Future Perspectives

The iterative geostatistical geophysical inversion methods proposed in this thesis could predict three-dimensional models of electrical conductivity and magnetic susceptibility using a one-dimensional formulation of the FDEM physical forward equations and capturing the spatial structure by imposing auxiliary data-based variogram models with vertical and lateral spatial correlations. The one-dimensionality of the forward model represents one of the main limitations of the proposed methodologies. This could be overcome with numerical solutions to 3D forward modelling algorithms, which recently became available in open-source code repositories (e.g., Heagy et al., 2017; Werthmüller et al., 2019). Using 3D forward models may

allow the GEMI and the joint inversion methods to go beyond in the modelling of complex and heterogeneous subsurface environments. However, since the increase in the subsurface structure complexity would result in the use of millions of model parameters, this would increase the computational burden of the inversion procedure to a point that could become unfeasible. Also, resolving the inverse problem of complex subsurface structures using FDEM data could imply collecting large data sets of FDEM surveys. Following that demand, FDEM instrumentation and data collection has undergone recent advances, with the acquisition of large data sets sampled at a high resolution with the use of drones, with particular interest for advanced 3D inversion of that data. A line of future research could be the integration of 3D forward models in geostatistical FDEM inversion methods and the management of these large FDEM data sets using machine learning and deep learning algorithms, following the steps of FDEM inversion proposed in Chapter 4.

A complementary line of research would be to push the boundaries of joint inversion methods applied to near-surface heterogeneous environments, integrating more geophysical methods that would complementary improve the predicted subsurface physical models. Seismic methods have the potential to characterize the interfaces between some subsurface structures that, combined with FDEM and direct current resistivity methods, could retrieve more reliable numerical models of subsurface physical properties, while simultaneously assessing the uncertainty of each data domain. Following this line of research, geophysical data could be integrated with remote sensing images to create 3D models in areas where only 2D transects of geophysical data are available. This type of spatial data is increasingly abundant, available across large areas and easy to access, and can be a great addition to geophysical data, particularly FDEM data, contributing to better understanding the subsurface.

Adding another dimensionality to the subsurface characterization is also a way to go. Geophysical time-lapse studies, with repeated geophysical measurements, such as FDEM surveys, along the time in the same study area can be used for advanced monitoring and characterization of the subsurface. These could be particularly useful in near-surface applications of environmental and groundwater contamination, characterizing through time-lapse FDEM measurements the evolution of the subsurface contamination. The new FDEM data in each period could be used in the co-simulation of updated EC and MS predicted models.

REFERENCES

- Akramkhanov, A., Brus, D.J., and D.J.J., Walvoort, 2014, Geostatistical monitoring of soil salinity in Uzbekistan by repeated EMI surveys. *Geoderma*, 213, 600-607.
- Alcolea, A., and P. Renard, 2010, Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data. *Water Resources Research*, 46, W08511.
- Al-Gaadi, K.A., 2012, Employing Electromagnetic Induction Technique for the Assessment of Soil Compaction. *American Journal of Agricultural and Biological Sciences*, 7, 425-434.
- Archie, G. E., 1942, The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Trans.*, 146, 54–62.
- Auken, E., and A. V. Christiansen, 2004, Layered and laterally constrained 2D inversion of resistivity data. *Geophysics*, 69, 752–761.
- Azevedo, L., and V. Demyanov, 2019, Multiscale uncertainty assessment in geostatistical seismic inversion. *Geophysics*, 84, no. 3, R355–R369.
- Azevedo, L., Nunes, R., Soares, A., Neto, G., and T. Martins, 2018, Geostatistical seismic Amplitude-versus-angle inversion. *Geophysical Prospecting*, 66, 116-131.
- Azevedo, L., and A. Soares, 2017, Geostatistical Methods for Reservoir Geophysics. *Advances in Oil and Gas Exploration & Production*, Springer.
- Badewa, E., Unc, A., Cheema, M., Kavanagh, V., and L. Galagedara, 2018, Soil moisture mapping using multi-frequency and multi-coil electromagnetic induction sensors on managed podzols. *Agronomy*, 8, 224.
- Baines, D., Smith, D. G., Froese, D. G., Bauman, P., and G. Nimeck, 2002, Electrical resistivity ground imaging (ERGI): A new tool for mapping the lithology and geometry of channel-belts and valley-fills. *Sedimentology*, 49, 441-449.
- Blaha, U., Appel, E., and H. Stanjek, 2008, Determination of anthropogenic boundary depth in industrially polluted soil and semi-quantification of heavy metal loads using magnetic susceptibility. *Environmental Pollution*, 156, 278-289.

Blatter, D., Key, K., Ray, A., Foley, N., Tulaczyk, S., and E. Auken, 2018, Trans-dimensional Bayesian inversion of airborne transient EM data from Taylor Glacier, Antarctica. *Geophysical Journal International*, 214(3), 1919-1936.

Bobe, C., 2020, Efficient probabilistic processing of frequency-domain electromagnetic data for subsurface modelling. PhD thesis, Ghent University.

Bobe, C., Hanssens, D., Hermans, T., and E. Van De Vijver, 2020, Efficient probabilistic joint inversion of direct current resistivity and electromagnetic induction data. *Algorithms*, 13(144), 1-16.

Bobe, C., Van De Vijver, E., Keller, J., Hanssens, D., Van Meirvenne, M., and P. De Smedt, 2019, Probabilistic 1-D inversion of frequency-domain electromagnetic data using a Kalman ensemble generator. *IEEE Transactions on Geoscience and Remote Sensing*, 58, 3287–3297.

Bongiovanni, M. V., Bonomo, N., de la Vega, M., Martino, L., and A. Osella, 2008, Rapid evaluation of multifrequency EMI data to characterize buried structures at a historical Jesuit Mission in Argentina. *Journal of Applied Geophysics*, 64, 37–46.

Butler, D., 2005, Near-Surface Geophysics: Society of Exploration Geophysicists.

Canchumuni, S. W., Emerick, A. A., and M.A.C. Pacheco, 2019, Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, 128, 87–102.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa C., and H.A. Phan, 2015, Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2), 145-163.

Cox, L. H., and M. S. Zhdanov, 2008, Advanced computational methods of rapid and rigorous 3-D inversion of airborne electromagnetic data. *Communications in Computational Physics*, 3, 160–179.

da Cruz Luz, E., da Silva, V., and C. Régis, 2013, Fast sensitivities for 2D marine CSEM inversion. 83rd Annual International Meeting, SEG, Expanded Abstracts, 845–849.

Dafflon, B., Hubbard, S., Ulrich, C., and J. E. Peterson, 2013, Electrical conductivity imaging of active layer and permafrost in an arctic ecosystem, through advanced inversion of electromagnetic induction data. *Vadose Zone Journal*, 12(4), 1-19.

De Lathauwer, L., De Moor, B., and J. Vandewalle, 2000a, A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4), 1253-1278.

De Lathauwer, L., De Moor, B., and J. Vandewalle, 2000b, On the best rank-1 and rank-($r_1, r_2, \dots, r_n$) approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4), 1324-1342.

De Smedt, P., Saey, T., Lehouck, A., Stichelbaut, B., Meerschman, E., Islam, M. M., Van De Vijver, E., and M. Van Meirvenne, 2013, Exploring the potential of multi-receiver EMI survey for geoarchaeological prospection: A 90ha dataset. *Geoderma*, 199, 30–36.

De Smedt, P., Van Meirvenne, M., and D. Simpson, 2011, Multi-signal EMI and geoarchaeology: Evaluating integrated magnetic susceptibility measurements for archaeological prospection. 9th International Conference on Archaeological Prospections, Archaeology and Art Publications, 54–57.

Deidda, G. P., de Alba, P. D., and G. Rodriguez, 2017, Identifying the magnetic permeability in multi-frequency EM data inversion. *ETNA Electronic Transactions on Numerical Analysis*, 27, 1–17.

Delefortrie, S., Saey, T., Van De Vijver, E., De Smedt, P., Missiaen, T., Demerre, I., and M. Van Meirvenne, 2014a, Frequency domain electromagnetic induction survey in the intertidal zone: Limitations of low-induction-number and depth of exploration. *Journal of Applied Geophysics*, 100, 14–22.

Delefortrie, S., De Smedt, P., Saey, T., Van De Vijver, E., and M. Van Meirvenne, 2014b, An efficient calibration procedure for correction of drift in EMI survey data. *Journal of Applied Geophysics*, 110, 115–125.

Delefortrie, S., Saey, T., De Pue, J., Van De Vijver, E., De Smedt P., and M. Van Meirvenne, 2016, Evaluating corrections for a horizontal offset between sensor and position data for surveys on land. *Precis. Agric.*, 17, 349–364.

Delefortrie, S., Hanssens, D., and P. De Smedt, 2018, Low signal-to-noise FDEM in-phase data: Practical potential for magnetic susceptibility modelling. *Journal of Applied Geophysics*, 152, 17–25.

Delefortrie, S., Hanssens, D., Saey, T., Van De Vijver, E., Smetryns, M., Bobe, C., and P. De Smedt, 2019, Validating land-based FDEM data and derived conductivity maps: Assessment of signal calibration, signal attenuation and the impact of heterogeneity. *Journal of Applied Geophysics*, 164, 179–190.

Deutsch, C., and A. G. Journel, 1998, GSLIB: Geostatistical software library and users' guide. *Oxford University Press*.

- Di Maio, R., Fais, S., Ligas, P., Piegari, E., Raga, R., and R. Cossu, 2018, 3D geophysical imaging for site-specific characterization plan of an old landfill. *Waste Management*, 76, 629-642.
- Doolittle, J., and E. Brevik, 2014, The use of electromagnetic induction techniques in soils studies. *Geoderma*, 223-225, 33-45.
- Dumont, G., Robert, T., Marck, N., and F. Nguyen, 2017, Assessment of multiple geophysical techniques for the characterization of municipal waste deposit sites. *Journal of Applied Geophysics*, 145, 74-83.
- Emerick, A. A., and A. C. Reynolds, 2013, Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55, 3-15.
- Evensen, G., 2003, The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343–367.
- Evensen, G., 2009, Data Assimilation: the ensemble Kalman filter. Springer.
- Everett, Mark E., 2013. Near-Surface Applied Geophysics. Cambridge: Cambridge University Press.
- Farquharson, C. G., and D. W. Oldenburg, 1996, Approximate sensitivities for the electromagnetic inverse problem. *Geophysical Journal International*, 126(1), 235-252.
- Farquharson, C. G., Oldenburg, D. W., and P. S. Routh, 2003, Simultaneous 1D inversion of loop–loop electromagnetic data for magnetic susceptibility and electrical conductivity. *Geophysics*, 68, 1857–1869.
- Fu, J., and J. J. Gomez-Hernandez, 2009, A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling. *Mathematical Geoscience*, 41, 105–128.
- Grana, D., Azevedo, L., de Figueiredo, L., Connolly, P., and T. Mukerji, 2022, Probabilistic inversion of seismic data for reservoir petrophysical characterization: Review and examples. *Geophysics*, 87(5), M199–M216.
- Grana, D., de Figueiredo, L. P., and L. Azevedo, 2019, Uncertainty quantification in Bayesian inverse problems with model and data dimension reduction. *Geophysics*, 84, M15-M24.
- Grana, D., Liu, M., and M. Ayani, 2021, Prediction of CO₂ Saturation Spatial Distribution Using Geostatistical Inversion of Time-Lapse Geophysical Data. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5), 3846 – 3856.

Guérin, R., Munoz, M.L., Aran, C., Laperrelle, C., Hidra, M., Drouart, E., and S. Grellier, 2004, Leachate recirculation: moisture content assessment by means of a geophysical technique. *Waste Management*, 24, 785–794.

Guillemoteau, J., Simon, F.-X., Luck, E., and J. Tronicke, 2016, 1D sequential inversion of portable multi-configuration electromagnetic induction data. *Near Surface Geophysics*, 14(5), 423-432.

Guptasarma, D., and B. Singh, 1997, New digital linear filters for Hankel J0 and J1 transforms. *Geophysical Prospecting*, 45, 745–762.

Haber, E., Ascher, U. M., and D. W. Oldenburg, 2004, Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69, 1216–1228.

Hansen, T. M., Journal, A. G., Tarantola, A., and K. Mosegaard, 2006, Linear inversion Gaussian theory and geostatistics. *Geophysics*, 71(6), R101-R111.

Hansen, T. M., Cordua, K. C., and K. Mosegaard, 2012, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. *Computational Geosciences*, 16, 593–611.

Hanssens, D., Delefortrie, S., De Pue, J., Van Meirvenne, M., and P. De Smedt, 2019, Frequency-domain electromagnetic forward and sensitivity modeling: Practical aspects of modeling a magnetic dipole in a multilayered half-space. *IEEE Geoscience and Remote Sensing Magazine*, 7, 74–85.

Hanssens, D., Waegeman, W., Declercq, Y., Dierckx, H., Verschelde H., and P. De Smedt, 2020, High-resolution surveying with small-loop frequency-domain electromagnetic systems: Efficient survey design and adaptive processing. *IEEE Geoscience and Remote Sensing Magazine*, 9(1), 167-183.

Hashemian, A., Garcia, S., Rivera, J. A., and D. Pardo, 2021, Massive database generation for 2.5 D borehole electromagnetic measurements using refined isogeometric analysis. *Computers & Geosciences*, 155, 104808.

Heagy, L. J., Cockett, R., Kang, S., Rosenkjaer, G. K., and D. W. Oldenburg, 2017, A framework for simulation and inversion in electromagnetics. *Computers & Geosciences*, 107, 1-19.

Hendricks Franssen, H. J., and W. Kinzelbach, 2008, Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44, W09408.

Hilhorst, M. A., 1998, Dielectric Characterisation of Soil. Landbouwniversiteit Wageningen. Wageningen Agricultural University, Wageningen

Horta, A., and A. Soares, 2010, Direct Sequential Co-simulation with Joint Probability Distributions. *Mathematical Geoscience*, 42, 269–292.

Huang, J., Scudiero, E., Clary, W., Corwin, D.L., and J. Triantafilis, 2016, Time-lapse monitoring of soil water content using electromagnetic conductivity imaging, *Soil Use Manage*, 33, 191–204.

Huang, J., Koganti, T., Santos, F. A., and J. Triantafilis, 2017, Mapping soil salinity and a fresh-water intrusion in three-dimensions using a quasi-3D joint-inversion of DUALEM-421S and EM34 data. *Science of the Total Environment*, 577, 395–404.

Hudson, M. Mikolas, M., Geissman, J., and B. Allen, 1999, Paleomagnetic and rock magnetic properties of Santa Fe Group sediments in the 98th Street core hole and correlative surface exposures, Albuquerque basin, New Mexico. *New Mexico Geology Society Guidebook*, 50, 355–361.

Hunziker, J., Laloy, E., and N. Linde, 2017, Inference of multi-Gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data. *Geophysics*, 82(5), H25-H40.

Islam, M. M., Meerschman, E., Saey, T., De Smedt, P., Van De Vijver, E., Delefortrie, S., and M. Van Meirvenne, 2014a, Characterizing compaction variability with an electromagnetic induction sensor in a puddled paddy rice field. *Soil Science Society of America Journal*, 78(2), 579-588.

Islam, M. M., Saey, T., De Smedt, P., Van De Vijver, E., Delefortrie, E., and M. Van Meirvenne, 2014b, Modeling within field variation of the compaction layer in a paddy rice field using a proximal sensing system. *Soil Use and Management*, 30(1), 99-108.

Jadoon, K.Z., Altaf, M.U., McCabe, M.F., Hoteit, I., Muhammad, N., Moghadas, D., and L. Weihermuller, 2017, Inferring soil salinity in a drip irrigation system from multi-configuration EMI measurements using adaptive Markov chain Monte Carlo. *Hydrology and Earth System Sciences*, 21(10), 5375–5383.

Jäggli, C., Straubhaar, J., and P. Renard, 2017, Posterior population expansion for solving inverse problems. *Water Resources Research*, 53, 2902–2916.

Journel, A. G., and C.J. Huijbregts, 1978, Mining Geostatistics. Academic Press, New York.

Kalman, R. E., 1960, A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35–45.

Keller, G. V., 1987, Rock and mineral properties: in *Electromagnetic Methods in Applied Geophysics*, Investigations in Geophysics Series 1, Society of Exploration Geophysics

Laloy, E., Linde, N., Jacques D., and J.A. Vrugt, 2015, Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction. *Water Resources Research*, 51(6), 4224–4243.

Laloy, E., Herault, R., Lee, J., Jacques, D., and N. Linde, 2017, Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, 110, 387–405.

Laloy, E., Hérault, R., Jacques D., and N. Linde, 2018, Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 54(1), 381–406.

Lee, J., and P. K. Kitanidis, 2014, Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA). *Water Resources Research*, 50(7), 5410–5427.

Lehmann, A., and K. Stahr, 2007, Nature and significance of anthropogenic urban soils: *Journal of Soils and Sediments*, 7, 247–260.

Li, R., Zhang, H., Gao, S., Wu, Z., and C. Guo, 2021, An improved extreme learning machine algorithm for transient electromagnetic nonlinear inversion. *Computers & Geosciences*, 156, 104877.

Liu, M., and D. Grana, 2020, Petrophysical characterization of deep saline aquifers for CO₂ storage using ensemble smoother and deep convolutional autoencoder. *Advances in Water Resources*, 142, p.103634.

Liu, M., Grana, D., and T. Mukerji, 2022a, Randomized Tensor Decomposition for Large-Scale Data Assimilation Problems for Carbon Dioxide Sequestration. *Mathematical Geosciences*, 54, 1139–116.

Liu, M., Grana D., and L.P. de Figueiredo, 2022b, Uncertainty quantification in stochastic inversion with dimensionality reduction using variational autoencoder. *Geophysics*, 87(2), pp.M43-M58.

Liu, M., Narciso, J., Grana, D., Van De Vijver, E., and L. Azevedo, (2023), Frequency-domain electromagnetic induction for the prediction of electrical conductivity and magnetic susceptibility using geostatistical inversion and randomized tensor decomposition. *Geophysics*, 88(6), 1-13.

Loke, M.H., 2018, Tutorial 2-D and 3-D electrical imaging surveys. Geotomosoft Solutions, Malaysia.

Lopez-Alvis, J., Laloy, E., Nguyen, F., and T. Hermans, 2021, Deep generative models in inversion: the impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder. *Computational Geosciences*, 152, 104762.

Manassero, M. C., Afonso, J. C., Zyserman, F., Zlotnik, S., and I. Fomin, 2020, A reduced order approach for probabilistic inversions of 3-D magnetotelluric data I: general formulation. *Geophysical Journal International*, 223(3), 1837-1863.

Mariethoz, G., Renard, P., and J. Caers, 2010, Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research*, 46, W11530.

McGillivray, P. R., and D. W. Oldenburg, 1990, Methods for calculating Fréchet derivatives and sensitivities for the non-linear inverse problem: A comparative study. *Geophysical Prospecting*, vol. 38, 499–524.

Minsley, B. J., 2011, A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. *Geophysical Journal International*, 187, 252–272.

Minsley, B. J., Smith, B. D., Hammack, R., Sams, J. I., and G. Veloski, G., 2012, Calibration and filtering strategies for frequency domain electromagnetic data. *Journal of Applied Geophysics*, 80, 56–66.

Mo, S., Zabaras, N., Shi, X., and J. Wu, 2019, Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, 55, 3856–3881.

Moghadas, D., André, F., Bradford, J. H., van der Kruk, J., Vereecken, H., and S. Lambot, 2012, Electromagnetic induction antenna modelling using a linear system of complex antenna transfer functions. *Near Surface Geophysics*, 10, 237-247.

Moghadas, D., Jadoon, K. Z., and M. F. McCabe, 2017, Spatiotemporal monitoring of soil water content profiles in an irrigated field using probabilistic inversion of time-lapse EMI data. *Advances in Water Resources*, 110, 238-248.

Moghadas, D., and J. Vrugt, 2019, The Influence of Geostatistical Prior Modeling on the Solution of DCT-Based Bayesian Inversion: A Case Study from Chicken Creek Catchment. *Remote Sensing*, 11, 1549-1571.

Monteiro Santos, F.A., 2004, 1-D laterally constrained inversion of EM34 profiling data. *Journal of Applied Geophysics*, 56(2), 123–134.

Monteiro Santos, F. A., Triantafilis, J., and K. Bruzgulis, 2011, A spatially constrained 1D inversion algorithm for quasi-3D conductivity imaging: Application to DUALEM-421 data collected in a riverine plain. *Geophysics*, 76(2), B43–B53.

Moorkamp, M., 2017, Integrating electromagnetic data with other geophysical observations for enhanced imaging of the earth: a tutorial and review. *Surveys in Geophysics*, 38, 935–962.

Morel, J. L., and A. B. Heinrich, 2008, SUITMA—soils in urban, industrial, traffic, mining, and military areas. *Journal Soils Sediments*, 8, 206–207.

Morel, J. L., Chenu, C., and K. Lorenz, 2015, Ecosystem services provided by soils of urban, industrial, traffic, mining, and military areas (SUITMAs). *Journal of Soils and Sediments*, 15, 1659–1666.

Narciso, J., Azevedo, L., Van De Vijver, E., and M. Van Meirvenne, 2020, Geostatistical electromagnetic inversion for landfill characterization: Presented at the *NSG2020 26th European Meeting of Environmental and Engineering Geophysics*

Narciso, J., Bobe, C., Azevedo, L., and E. Van De Vijver, 2022, A comparison between Kalman ensemble generator and geostatistical frequency-domain electromagnetic inversion: the impacts on near-surface characterization. *Geophysics*, 27(5), 1-12.

Nino Ruiz, E.D., Sandu A., and J. Anderson, 2015, An efficient implementation of the ensemble Kalman filter based on an iterative Sherman–Morrison formula. *Statistics and Computing*, 25(3), pp.561-577.

Nowak, W., 2009, Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research*, 45, W04431.

O'Neill, K., Shubitidze, F., Sun, K., Shamatava, I., and K. Paulsen, 2005, EMI obscuration of buried UXO by geophysical magnetic permeability, anthropogenic clutter, and by magnitude disparities. in *Proceedings of SPIE - The International Society for Optical Engineering*, 5811, 87-99.

Paepen, M., Hanssens, D., Smedt, P. D., Walraevens, K., and T. Hermans, 2020, Combining resistivity and frequency domain electromagnetic methods to investigate submarine groundwater discharge (SGD) in the littoral zone. *Hydrology and Earth System Sciences Discussions*, 1–23.

Pedrerá-Parrilla, A., Van De Vijver, E., Van Meirvenne, M., Espejo-Pérez, A.J., Giraldez, J.V., and K. Vanderlinden, 2016, Apparent electrical conductivity measurements in an olive orchard under wet and dry soil conditions: significance for clay and soil water content mapping. *Precision Agriculture*, 17, 531-545.

Pidlisecky, A., and R. Knight, 2008, FW2 5D: a MATLAB 2.5-D electrical resistivity modelling code. *Computers & Geosciences*, 34(12), 1645-1654.

Pyrz, M., and C. Deutsch, 2014, *Geostatistical reservoir modeling*: Oxford University Press.

Persico, R., Piro, S., and N. Linford, 2018, *Innovation in near-surface geophysics: instrumentation, application, and data processing methods*: Elsevier.

Puzyrev, V., and A. Swidinsky, 2021, Inversion of 1D frequency- and time-domain electromagnetic data with convolutional neural network. *Computers & Geosciences*, 149, 104681.

Qi, Y., El-Kaliouby, H., Revil, A., Ahmed, A. S., Ghorbani, A., and J. Li, 2019, Three-dimensional modeling of frequency-and time-domain electromagnetic methods with induced polarization effects: *Computers & Geosciences*, 124, 85-92.

Qiu, C., Liang, B., Han, F., Li, J., Fang, G., and Q. H. Liu, 2020, 3-D Full-Wave Inversion of Helicopter Transient Electromagnetic Data in Frequency Domain. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), 4928-4938.

Rabanser, S., Shchur, O., and S. Günnemann, 2017, Introduction to tensor decompositions and their applications in machine learning, arXiv preprint, arXiv:1711.10781.

Raiche, A., Jupp, D., Rutter, H., and K. Vozoff, 1985, The joint use of coincident loop transient electromagnetic and Schlumberger sounding to resolve layered structures. *Geophysics*, 50, 1618-1627.

Ray, A., and K. Key, 2012, Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm. *Geophysical Journal International*, 191(3), 1135-1151.

Rejiba, F., Schamper, C., Chevalier, A., Deleplancque, B., Hovhannissian, G., Thiesson, J., and P. Weill, 2018, Multiconfiguration electromagnetic induction survey for paleochannel internal structure imaging: A case study in the alluvial plain of the River Seine, France. *Hydrology and Earth System Sciences*, 22, 159–170.

Reynolds, J. M., 2011, An Introduction to Applied and Environmental Geophysics (2nd ed.), Chichester, England: John Wiley & Sons, Ltd.

Rosas-Carbajal, M., Linde, N., Kalscheuer, T., and J. A. Vrugt, 2013, Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data. *Geophysical Journal International*, 196(3), 1508-1524.

Saey, T., De Smedt, P., Delefortrie, S., Van De Vijver, E., and M. Van Meirvenne, 2015, Comparing one- and two-dimensional EMI conductivity inverse modeling procedures for characterizing a two-layered soil. *Geoderma*, 241, 12–23.

Saey, T., Gheyle, W., Stichelbaut, B., Bourgeois, J., Verplaetse, S., Van Eetvelde, V., Note, N., and M. Van Meirvenne, 2016, The Characterization of a Former World War I Battlefield by Integrating Multiple Signals from a Multireceiver EMI Soil Sensor. *Geoarchaeology*, 31, 267-281.

Saey, T., Van Meirvenne, M., Dewilde, M., Wyffels, F., De Smedt, P., Meerschman, E., Islam, M.M., Meeuws, F., and L. Cockx, 2011, Combining multiple signals of an electromagnetic induction sensor to prospect land for metal objects. *Near Surface Geophysics*, 9.

Sainato, C.M., Losinno, B.N., Márquez Molina, J.J., and R.A., Espada, 2018, Electromagnetic soundings to detect groundwater contamination produced by intensive livestock farming. *Journal of Applied Geophysics*, 154, 159-166.

Sambridge, M. and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3), 3-1.

- Sasaki, Y., Kim, J.-H., and S.-J. Cho, 2010, Multidimensional inversion of loop-loop frequency-domain EM data for resistivity and magnetic susceptibility. *Geophysics*, 75(6), F213–F223.
- Shamsipour, P., Marcotte, D., and M. Chouteau, 2012, 3D stochastic joint inversion of gravity and magnetic data. *Journal of Applied Geophysics*, 79, 27–37.
- Sharma, S. P., and P. Kaikkonen, 2003, Appraisal of equivalence and suppression problems in 1D EM and DC measurements using global optimization and joint inversion. *Geophysical Prospecting*, 47, 219–249.
- Simon, F.-X., Sarris, A., Thiesson, J., and A. Tabbagh, 2015, Mapping of quadrature magnetic susceptibility/magnetic viscosity of soils by using multi-frequency EMI. *Journal of Applied Geophysics*, 120, 36–47.
- Soares, A., 2001, Direct Sequential Simulation and Cosimulation. *Mathematical Geology*, 33, 911–926.
- Sundberg, K., and F., Hedstroem, 1934, Structural investigations by electromagnetic methods. In Proceedings of the World Petroleum Congress. World Petroleum Congress.
- Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia.
- Telford, W. M., Geldart, L., Sheriff, R. E., and R. Sheriff, 1990, Applied Geophysics. Cambridge university press.
- Triantafyllis, J., and F. A. Monteiro Santos, 2013, Electromagnetic conductivity imaging (EMCI) of soil using a DUALEM-421 and inversion modelling software (EM4Soil). *Geoderma*, 211–212, 1, 28–38.
- Tikhonov, A. N., and V.Y. Arsenin, 1977, Solutions of ill-posed problems. WH Winston, Washington, DC, 330.
- Tylor-Jones, T., and L. Azevedo, 2022, A Practical Guide to Seismic Reservoir Characterization. Advances in oil and gas exploration & production. Springer, Berlin.
- Van De Vijver, E., 2017, Proximal soil sensing in the context of urban (re)development: an evaluation of multi-receiver electromagnetic induction and stepped-frequency ground penetrating radar at landfills and industrial sites. Ghent, Belgium, Ghent University. Faculty of Bioscience Engineering.

Van De Vijver, E., and M. Van Meirvenne, 2016, Delving into the potential of multi-receiver electromagnetic induction surveying for enhanced landfill exploration in view of ELFM. In 3rd International symposium on Enhanced Landfill Mining, 175–187.

Van De Vijver, E., Van Meirvenne, M., Saey, T., Delefortrie, S., De Smedt, P., De Pue, J., and P. Seuntjens, 2015, Combining multi-receiver electromagnetic induction and stepped frequency ground penetrating radar for industrial site investigation. *Eur J soil Sci*, 66, 688-698.

Verhegge, J., Saey, T., Van Meirvenne, M., Missiaen, T., and P. Crombé, 2016a, Reconstructing early neolithic paleogeography: EMI-based subsurface modeling and chronological modeling of Holocene peat below the lower Scheldt floodplain in NW Belgium. *Geoarchaeology: An International Journal*, 32, 159-176.

Verhegge, J., Missiaen, T., and P. Crombé, 2016b, Exploring integrated geophysics and geotechnics as a paleolandscape reconstruction tool: Archaeological prospection of (Prehistoric) sites buried deeply below the Scheldt Polders (NW Belgium). *Archaeological Prospection*, 23, 125-145.

Viscarra Rossel, R., Adamchuk, V., Sudduth, K., McKenzie, N., and C. Lobsey, 2011, Chapter Five - Proximal soil sensing: an effective approach for soil measurements in space and time. *Advances in Agronomy*, 113, 243-291.

von Hebel, C., Reynaert, S., Pauly, K., Janssens, P., Piccard, I., Vanderborght, J., van der Kruk, J., Vereecken, H., and S. Garré, 2021, Toward high-resolution agronomic soil information and management zones delineated by ground-based electromagnetic induction and aerial drone data. *Vadose Zone Journal*, 20(4), 1-18.

Wang, L., Kitanidis, P. K., and J. Caers, 2022, Hierarchical Bayesian inversion of global variables and large-scale spatial fields. *Water Resources Research*, 58.

Werthmüller, D., Mulder, W., and E. Slob, 2019, emg3d: A multigrid solver for 3D electromagnetic diffusion. *Journal of Open Source Software*, 4(39), 1463.

Yi, M-J, and Y. Sasaki, 2015, 2-D and 3-D joint inversion of loop–loop electromagnetic and electrical data for resistivity and magnetic susceptibility. *Geophysical Journal International*, 203, 1085-1095.

Zhdanov, M. S., 2010, Electromagnetic geophysics: Notes from the past and the road ahead. *Geophysics*, 75(5), 75A49–75A66.

Zahner, T., Lochbuhler, T., Mariethoz, G., and N. Linde, 2016, Image synthesis with graph cuts: A fast model proposal mechanism in probabilistic inversion. *Geophysical Journal International*, 204(2), 1179–1190.

Zhou, H., Gomez-Hernandez, J.J., Franssen, H.-J. H., and L. Li, 2011, An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, 34, 844–864.