

Improving Uncertainty Estimation in Geophysical Inversion Modelling

A dissertation presented
by
Sebastian Schnaidt

In fulfilment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Geophysics



THE UNIVERSITY

of ADELAIDE

Submitted to the
Department of Earth Sciences,
School of Physical Sciences, Faculty of Sciences

Adelaide, May 2015



THE UNIVERSITY
of ADELAIDE

© 2015
Sebastian Schnaidt
All Rights Reserved



**DEEP EXPLORATION
TECHNOLOGIES CRC**
Uncovering the future

I can live with doubt, and uncertainty, and not knowing. I think it's much more interesting to live not knowing than to have answers which might be wrong. I have approximate answers, and possible beliefs, and different degrees of certainty about different things, but I'm not absolutely sure of anything.

– Richard P. Feynman

CONTENTS

List of Tables	ix
List of Figures	xi
Abstract	xvii
Statement of Originality	xix
Acknowledgements	xxi
1 Introduction	1
1.1 Aims and Objectives	1
1.2 Introductory Background and Literature Review	2
1.2.1 Direct search methods	3
1.2.2 Joint inversion	5
1.3 Model Errors and Uncertainty	6
1.3.1 Data noise and solution non-uniqueness	6
1.3.2 User introduced model errors	8
1.4 Overview	9

2	Bootstrap Resampling as a Tool for Uncertainty Analysis in 2-D Magnetotelluric Inversion Modelling	13
	Summary	15
2.1	Introduction	16
2.2	Method	20
	2.2.1 Bootstrap resampling	20
	2.2.2 Occam's inversion	24
2.3	Ensemble Appraisal	25
	2.3.1 Variability in absolute value	25
	2.3.2 Variability in cell gradient	26
2.4	Results	28
	2.4.1 Synthetic test case	28
	2.4.2 Delamerian MT Transect	35
2.5	Discussion	42
2.6	Conclusion	46
	Acknowledgments	47
3	A Novel Approach to Pareto-optimal Joint Inversion Modelling and Data Set Compatibility Analysis of Geophysical Data	49
	Summary	51
3.1	Introduction	52
3.2	Theory	54
	3.2.1 Definition of multi-dimensional Pareto-optimality . .	54
	3.2.2 The Multi-objective Joint Optimisation Algorithm .	55
	3.2.3 Solution ensemble appraisal	61

3.2.4	Data set compatibility	62
3.3	Feasibility Study: Example Data	65
3.3.1	Data set properties	66
3.3.2	Resulting model distribution	68
3.4	Feasibility Study: Analysis	69
3.4.1	Two objectives - compatible data sets	70
3.4.2	Two objectives - incompatible data sets	78
3.4.3	Multiple objectives	82
3.4.4	Influence of the numbers of objectives on the stability of the compatibility measure	86
3.5	Discussion	88
3.6	Conclusion	93
	Acknowledgments	94
4	Parameter Studies of Geophysical Joint Inversion Modelling - Using Multi-objective Optimisation	95
	Summary	97
4.1	Introduction	98
4.2	Parameter Analysis	100
4.2.1	Redefinition of the misfit δ	100
4.2.2	The box size parameter ϵ	100
4.2.3	Model distribution analysis	107
4.3	Application to Real Data	110
4.3.1	Cooper Basin, South Australia	111
4.3.2	Isa Extension MT survey	119
4.4	Discussion	124

4.4.1	Case studies	126
4.5	Conclusion	128
	Acknowledgments	128
5	Summary and Conclusion	129
5.1	Bootstrap Resampling	129
5.2	Multi-objective Joint Optimisation	131
5.3	Concluding Comments	133
A	Appendix	137
A.1	Supporting Material for Chapter 4	137
A.1.1	Figures	137
A.1.2	Tables	139
	Bibliography	141

LIST OF TABLES

3.1	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the noise-free compatible 2-objective case.	73
3.2	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the compatible 2-objective case (with 3% noise).	75
3.3	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the 5/5-case and the 10/5-case.	77
3.4	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the noise-free incompatible 2-objective case	80
3.5	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the incompatible data 2-objective case with 3% white noise on the data.	82
3.6	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1,1) for the multi-objective case inversion runs.	84
3.7	Analysis of the deviation from the ideal line and median distance from the ideal point (1,1) for the 2-objective case without regularisation.	87

4.1 Analysis of the deviation from the ideal line for MOJO runs with varying ε -values for the 10/5-case. 105

4.2 Analysis of the deviation from the ideal line for the objective combinations MN102-MN103, MN102-LLS, and MN103-LLS. 114

4.3 Analysis of the deviation from the ideal line for the objective combinations MN102-MN103, MN102-LLD, and MN103-LLD. 114

4.4 Analysis of the deviation from the ideal line and for the Isa Extension case study. 120

A.1 Analysis of the deviation from the ideal line for the MN102-MN103 MOJO run. 139

LIST OF FIGURES

2.1	Workflow of the bootstrap resampling approach.	22
2.2	Definition of nearest neighbours for cell gradient calculation.	27
2.3	True synthetic model and the inversion result based on the data forward calculated from the true model.	30
2.4	Monte Carlo bootstrap results for the synthetic test data. .	32
2.5	Relative residuals of the low-noise comparison run minus the master-result.	35
2.6	Map of the station locations of the Delamerian MT transect.	36
2.7	Master-inversion (original data set) of Delamerian MT Transect data set.	37
2.8	Monte Carlo bootstrap results for the Delamerian MT transect data.	39
2.9	Reinterpretation of the master-model for the Delamerian MT transect data, taking into consideration the bootstrap results.	41
2.10	Differences between the residuals of the bootstrap average and the true model and the residuals of the master-model with the true model.	42
3.1	Illustration of Pareto-dominance using the example of a two objective minimisation problem.	55
3.2	Flowchart of the MOJO algorithms functionality.	58

3.3	Illustration of ε -dominance and ε -progress for a hypothetical two objective case.	60
3.4	Conceptual misfit visualisation of two objectives for a hypothetical archive of two compatible data sets.	63
3.5	Conceptual misfit visualisations for two hypothetical pairs of data sets: one pair of compatible data sets and one pair of incompatible data sets.	64
3.6	The synthetic 9-layer model Model I and the synthetic 7-layer model Model II.	67
3.7	Model results from the noise-free 2-objective case.	71
3.8	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the compatible noise-free 2-objective case.	72
3.9	Model results from the 2-objective case with 3% noise.	74
3.10	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the compatible 2-objective case with 3% noise.	75
3.11	Pareto-front for the AMT-objective and the BB-objective and the corresponding Theil-Sen regression, for the compatible 2-objective case with 3% noise and assigned errors for the BB data set and 5% noise and 5% assigned errors for the AMT data set.	76
3.12	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the compatible 2-objective case with 3% noise and assigned errors for the BB data set and 10% noise and 5% assigned errors for the AMT data set.	77
3.13	Model results from the incompatible noise-free 2-objective case.	79

3.14	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the incompatible noise-free 2-objective case.	80
3.15	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the incompatible 2-objective case with 3% noise.	81
3.16	Model results from the multi-objective case.	83
3.17	Pareto-front for the AMT-BB-objectives, AMT-WELL-objectives and BB-WELL-objectives with the corresponding Theil-Sen regressions, for the multi-objective case.	85
3.18	Pareto-front for the AMT-objective and the BB-objective, and the corresponding Theil-Sen regression, for the 2-objective case without regularisation.	87
4.1	Illustration of the effects of the ε -values on the solution distributions in objective space.	101
4.2	The true synthetic model for the 10/5-case.	103
4.3	Comparison of the solution distributions and corresponding Theil-Sen regressions for inversion runs of the 10/5-case for ε -values of $\varepsilon = 10^{-1}$, $\varepsilon = 10^{-2}$, $\varepsilon = 10^{-3}$, $\varepsilon = 10^{-4}$, and $\varepsilon = 10^{-5}$	104
4.4	Model distributions of the archive solutions for inversion runs of the 10/5 case for ε -values of $\varepsilon = 10^{-1}$, $\varepsilon = 10^{-3}$, and $\varepsilon = 10^{-5}$, expressed as resistivity value probabilities and interface probabilities.	106
4.5	Model distributions of different subsets of the final archive of the $\varepsilon = 10^{-5}$ run of the 10/5-case, expressed as resistivity value probabilities and interface probabilities.	108
4.6	The synthetic 3-layer models Model- α and Model- ω used to simulate incompatible data.	109

4.7	Model distributions of different subsets of the final archive of the $\varepsilon = 10^{-5}$ run of the synthetic incompatible test case, expressed as resistivity value probabilities and interface probabilities.	109
4.8	Locations of the GEL 185 MT survey lines, including the sites MN102 and MN103 used in this study, and the Moomba-73 well.	111
4.9	Solution distributions for the MN102-MN103-objective, MN102-LLS-objective, and MN103-LLS-objective and the corresponding Theil-Sen regressions.	113
4.10	Model distributions of different subsets of the final archive generated from the joint inversion of the data sets MN102, MN103, and LLS, expressed as resistivity value probabilities and interface probabilities.	115
4.11	Model distributions for the complete archive and the top 10% solution with respect to the best solution generated from the joint inversion of the data sets MN102, MN103, and LLD, expressed as resistivity value probabilities and interface probabilities.	116
4.12	Model responses for the optimal solution generated from the joint inversion of the data sets MN102, MN103, and LLS; MN102, MN103, and LLD; and MN102 and MN103 compared to the measured responses.	117
4.13	Model distributions for the complete archive and the top 10% solution with respect to the best solution generated from the joint inversion of the data sets MN102 and MN103, expressed as resistivity value probabilities and interface probabilities.	118
4.14	Location of the Isa Extension MT survey and the sites IEA00184 and IEB0046 used in this study.	119

4.15	Model distributions of different subsets of the final archive generated from the joint inversion of the IEA00184 and IEB0046, expressed as resistivity value probabilities and interface probabilities.	121
4.16	The modelled responses of the Queensland data sets compared to the measured responses.	123
4.17	Illustration of the different stages of the MOJO workflow . .	125
A.1	Solution distributions for the MN102-MN103-objective, MN102-LLD-objective, and MN103-LLD-objective and the corresponding Theil-Sen regressions.	137
A.2	Solution distribution for the MN102-MN103 MOJO run and the corresponding Theil-Sen regression.	138
A.3	Solution distribution for the Isa Extension Survey data and the corresponding Theil-Sen regression.	138

ABSTRACT

Improving Uncertainty Estimation in Geophysical Inversion Modelling

by Sebastian Schnaidt

Numerical inversion modelling is an integral part of geophysical data interpretation. Growing computational resources are used to invert ever-growing data sets and higher dimensional data. However, models without meaningful uncertainty estimates are difficult to interpret reliably and limited attention has been paid to the advancement of model quality estimation techniques to keep up with the more sophisticated inversion schemes. The employment of meaningful uncertainty estimation methods is often hindered by the complicated implementation of those methods, and inadequate model quality estimators are frequently used. This project was aimed at the advancement of model uncertainty estimation, to enable a more common use.

Two different approaches were developed, approaching the problem from different directions:

Firstly, a bootstrap resampling approach for the qualitative estimation of model uncertainties is presented. The algorithm is characterised by an easy implementation and the fact that it can provide model quality estimation capabilities to existing inversion algorithms without requiring access to the inversion algorithm's source code. A given data set is repeatedly resampled to create multiple realisations of the data set. Each realisation is individually inverted and the variations between the generated models are analysed and visualised to generate interpretable uncertainty maps. The capabilities

of the approach are demonstrated using the example of synthetic and real 2-D magnetotellurics data.

Secondly, the multi-objective joint optimisation algorithm MOJO is presented, which aims to remedy the common shortcomings of classical joint inversion approaches. Joint inversion modelling is a powerful tool to improve model results and reduce the effects of data noise and solution non-uniqueness. Nevertheless, the classic joint inversion approaches have a variety of shortcomings, such as a dependency on the choice of data weights, optimising only a single solution resulting in inadequate uncertainty estimates, and the risk of model artefacts being introduced by the accidental joint inversion of incompatible data. MOJO is based on the concept of Pareto-optimality and treats each data set as a separate objective, avoiding data-weighting. The algorithm generates solution ensembles, which are statistically analysed to provide model uncertainty estimates. The shapes and evolutions of the solutions ensemble's distribution in objective space is dependent on the level of compatibility between the data set. The solution distributions are compared against a theoretical solution distribution corresponding to perfectly compatible data to estimate the compatibility state of any given objective-pair, allowing to distinguish between compatible and incompatible data, as well as identify data sets that are neither mutually exclusive nor sensitive to common features. MOJO's effectiveness was demonstrated in extensive feasibility studies on synthetic data as well as real data. The algorithm is adaptive and can be expanded to incorporate a variety of different data types.

Additionally, ways were explored to make the communication of the modelling results and the model quality estimates as clear and concise as possible, to allow the user to make an informed decision and avoid misinterpretations.

Thesis Supervisors: Prof. Dr. Graham Heinson, Dr. Stephan Thiel

STATEMENT OF ORIGINALITY

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

Date

ACKNOWLEDGEMENTS

Thank you to my supervisors Graham Heinson and Stephan Thiel for giving me the opportunity to come to Adelaide and perform research on this interesting topic. Thanks for letting me work freely, be it day or night, and for always providing support if needed.

Also, thank you to Kelly Keats for giving me the chance to gather valuable experience in the industry.

A special thanks to all my colleagues and fellow Ph.D. students in the geophysics group for creating a friendly working atmosphere and taking the time to discuss and talk, sometimes even about science. Thanks to all the people who helped proof read my papers and also this thesis and making it sound a little less German. This goes especially for Lars who provided critical comments and helped to improve my manuscripts, and was always ready to discuss ideas and help wherever he could.

Thanks for the good times I had on fieldwork and other trips, be it setting up instruments in 45 °C heat or spending a 0 °C night in a van. A special thanks goes to Göran, it is always a pleasure working with you.

Finally, thanks to my family and friends, near and far, for supporting me in this great adventure, especially Benedikt and Lisa for always being there despite the bane of being nine time zones apart. Also, thanks you Kate for being a tireless teacher and a friend on far shores.

The work has been supported by the Deep Exploration Technologies Cooperative Research Centre whose activities are funded by the Australian Government's Cooperative Research Centre Programme. This is DET CRC Document 2015/699.