JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 115, B04318, doi:10.1029/2009JB006369, 2010

Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data

M. Moorkamp,[1,2] A. G. Jones,[1] and S. Fishwick[3]

Received 9 February 2009; revised 19 October 2009; accepted 16 December 2009; published 30 April 2010.

[1] We present joint inversion of magnetotelluric, receiver function, and Raleigh wave dispersion data for a one-dimensional Earth using a multiobjective genetic algorithm (GA). The chosen GA produces not only a family of models that fit the data sets but also the trade-off between fitting the different data sets. The analysis of this trade-off gives insight into the compatibility between the seismic data sets and the magnetotelluric data and also the appropriate noise level to assume for the seismic data. This additional information helps to assess the validity of the joint model, and we demonstrate the use of our approach with synthetic data under realistic conditions. We apply our method to one site from the Slave Craton and one site from the Kaapvaal Craton. For the Slave Craton we obtain similar results to our previously published models from joint inversion of receiver functions and magnetotelluric data but with improved resolution and control on absolute velocities. We find a conductive layer at the bottom of the crust, just above the Moho; a low-velocity, low-resistivity zone in the lithospheric mantle, previously termed the Central Slave Mantle Conductor; and indications of the lithosphere-asthenosphere boundary in terms of a decrease in seismic velocity and resistivity. For the Kaapvaal Craton both the seismic and the MT data are of lesser quality, which prevents as detailed and robust an interpretation; nevertheless, we find an indication of a low-velocity low-resistivity zone in the mantle lithosphere. These two examples demonstrate the potential of joint inversion, particularly in combination with nonlinear optimization methods.

Citation: Moorkamp, M., A. G. Jones, and S. Fishwick (2010), Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data, *J. Geophys. Res.*, *115*, B04318, doi:10.1029/2009JB006369.

1. Introduction

[2] With increasing computational power, methods to jointly invert several data sets are gaining in popularity [e.g., *Julia et al.*, 2000; *Gallardo and Meju*, 2003; *Linde et al.*, 2006]. The rationale for this approach is to reduce the influence of noise and to increase resolution. Furthermore, if different types of data are inverted together the improved constraints on various physical parameters can give a better understanding of the underlying geology. However, combining different data in a single inversion bears the risk of forcing two or more inherently incompatible data sets into a common model that is not suitable for either of them.

[3] The factors that govern the distribution of seismic velocities and electrical conductivities within the Earth can be very different. Seismic velocities, particularly P wave velocities, are generally determined by bulk properties of the rock. With wavelengths of several kilometers for teleseismic data, the effective elastic moduli are an average over a large volume of material. Long-period magnetotelluric data also average over large volumes, but in a fundamentally different way. Here the question is often that of large-scale connectivity of conductive material; even small fractions of conductive material can determine the bulk conductivity, as long as it is connected [*Bahr and Simpson*, 2002]. These two different characteristics might lead to the conclusion that we should not expect any correlation between seismic velocities and electrical conductivities. This is generally true for the magnitude and its spatial derivatives; higher velocities are not usually associated with higher resistivities or vice versa, although for the mantle lithosphere of the Kaapvaal Craton [*Jones et al.*, 2009] recently showed a linear relationship between shear wave velocity and the logarithm of resistivity. In sedimentary environments porosity and fluids in the rock matrix provide a link that causes a correlation of seismic velocities and electrical conductivities [*Carcione et al.*, 2007]. In the mantle, temperature, iron content and composition all affect both the bulk and shear moduli and electrical conductivity [e.g., *Jones et al.*, 2009].

[4] More conservatively though, a geometrical and/or a structural link between the two parameters is quite likely.

[1]School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin, Ireland.
[2]Now at Leibniz-Institut für Meereswissenschaften an der Universität Kiel (IFM-GEOMAR), Kiel, Germany.
[3]Department of Geology, University of Leicester, Leicester, UK.

0148-0227/10/2009JB006369

Many discontinuities of seismic velocity or electrical properties are interpreted as boundaries between distinct geological units [e.g., *Marquis et al.*, 1995; *Lahti et al.*, 2005; *Tournerie and Chouteau*, 2005], which in some cases even originated in different regions and then accreted [e.g., *Jones et al.*, 2002; *Snyder and Bruneton*, 2007]. Under these circumstances changes in various petrophysical parameters will occur simultaneously. Also, the two major discontinuities in the lithosphere, the crust-mantle boundary (Moho) and the lithosphere-asthenosphere boundary (LAB) are both likely correlated with changes in conductivity [*Jones*, 1999; *Jones and Ferguson*, 1997; *Gatzemeier and Moorkamp*, 2005]. Therefore we can expect at least some structural correlation between these parameters. Still, we have to account for the possibility of disparate interfaces that do not allow a common parametrization. We will show to what extent our genetic algorithm approach helps to identify these situations.

[5] In this study we jointly invert magnetotelluric data, receiver functions and fundamental mode Rayleigh wave dispersion data. Previously [*Moorkamp et al.*, 2007], we showed how we can use the combination of magnetotellurics and receiver functions to construct joint electric and seismic models for two sites on the Slave Craton. This approach was motivated by the observation of similar interface locations in separate models from that area [*Davis et al.*, 2003; *Jones et al.*, 2003; *Snyder et al.*, 2004] and we obtained encouraging results. However, the limited sensitivity of receiver functions to absolute S wave velocities required us to prescribe crustal velocities taken from a regional seismic model [*Perry et al.*, 2002]. By adding fundamental mode Rayleigh wave dispersion data into the inversion process, we obtain a reference for absolute S wave velocities that should remove the need for tight constraints on the seismic parameters.

2. Data Sets

2.1. Magnetotelluric Data

[6] Magnetotellurics (MT) is the main geophysical method to derive the conductivity distribution of the Earth's crust and mantle. From simultaneous measurements of the horizontal components of the electric fields **E** and magnetic fields **H** at the surface, we can estimate the complex impedance tensor **Z** in the frequency domain, namely,

$$\mathbf{E} = \mathbf{ZH}. \tag{1}$$

For a uniform source field, the impedance tensor solely depends on the conductivity distribution in the subsurface. The impedances at high frequencies correspond to nearby structures, whereas the impedances at low frequencies correspond to more distant (deeper for a 1-D Earth) structures. In the most general case all entries of the impedance tensor are nonzero and we need 3-D modeling and inversion methods to fit all its components. Due to the difficulties associated with a full 3-D inversion and only recent availability of inversion codes [*Siripunvaraporn et al.*, 2005], a number of methods have been developed to identify data that can be adequately modeled with a 2-D, or even a 1-D model. We apply dimensionality analysis using the phase tensor [*Caldwell et al.*, 2004] to our data to identify suitable stations for our 1-D modeling approach.

[7] For distortion by local structures of only the electric fields, the magnetotelluric phase tensor Φ represents the part of the impedance tensor that corresponds to the regional structure of the Earth and can be extracted without simplifying assumptions about dimensionality and galvanic distortion [*Bibby et al.*, 2005]. It is defined as

$$\Phi = \mathbf{X}^{-1}\mathbf{Y} \text{ where } \mathbf{X} = \mathrm{Re}(\mathbf{Z}), \mathbf{Y} = \mathrm{Im}(\mathbf{Z}). \tag{2}$$

[8] We can use the structure of this tensor to assess electromagnetic dimensionality. For one-dimensional structures the necessary conditions for the phase tensor are vanishing off-diagonal elements Φ_{21} and Φ_{21}, equal diagonal elements, and vanishing skew β [*Bibby et al.*, 2005], namely,

$$\beta = \frac{1}{2}\tan^{-1}\left(\frac{\Phi_{12} - \Phi_{21}}{\Phi_{11} + \Phi_{22}}\right). \tag{3}$$

[9] This condition can be alternatively expressed in terms of the ellipticity λ, as

$$\lambda = \frac{\Phi_{\max} - \Phi_{\min}}{\Phi_{\max} + \Phi_{\min}} = 0 \text{ and } \beta = 0. \tag{4}$$

Here $\Phi_{\max}$ and $\Phi_{\min}$ are the maximum and minimum principal values of Φ, respectively. For measured data that are affected by noise, this condition means that both values should not be different from zero in a statistical sense.

[10] For real data it is rare to find a site where the dimensionality condition is met at all frequencies and even when it is met the off-diagonal elements of the impedance tensor will be different from each other due to measurement noise. We can reduce the influence of noise and violations of the 1-D assumptions by inverting the Berdichevskiy invariant [*Berdichevskiy and Dmitriev*, 1976], the arithmetic average of the off-diagonal impedance values. As long as we do not have strong lateral interfaces we can expect good results with this approach [e.g., *Park and Livelybrooks*, 1989]. Another issue for the inversion of real data is the effect of small-scale inhomogeneities. The effect of these inhomogeneities can be described as a static distortion of the impedance tensor [*Bahr*, 1988; *Groom and Bailey*, 1989] and is commonly encountered. Various methods have been proposed to remove the effect of static distortion, the simplest manifestation of which is static shifts of the apparent resistivity curves [*Jones*, 1988], however these can only be applied effectively with a dense site coverage. One alternative is to model phase data alone, as the phases are not affected by static shifts, the most common form of static distortion. As the apparent resistivities define the depth scale in 1-D inversion [e.g., *Jones*, 1988], this requires prescribing the resistivity of some part of the model. The data we examine however do not appear to be affected significantly by static shift.

[11] We calculate synthetic impedances for a 1-D Earth of stacked isotropic layers using an implementation of the Wait algorithm [*Wait*, 1954]. This simple algorithm is a fast and accurate way to calculate the impedance for a layered sub-

surface. The inversion algorithm typically spends less than 5% of its time calculating synthetic MT responses.

2.2. Receiver Functions

[12] Receiver functions (RF) are commonly used as a tool to identify discontinuities in seismic velocities, particularly the Moho [e.g., *Langston*, 1979; *Kind and Vinnik*, 1988; *Langston and Hammer*, 2001]. Traditionally, analysis of P to S conversions, the so-called P receiver functions, have been used for this purpose, but recently also analysis of S to P conversion, also known as S receiver functions [*Yuan et al.*, 2006; *Vinnik et al.*, 2009], have gained popularity, particularly for identifying the LAB. In this study we focus solely on P receiver functions.

[13] There are various ways to calculate receiver functions from teleseismic data, the spectral water level deconvolution technique [*Langston*, 1979] probably being the most popular. In our experience the time domain iterative deconvolution technique [*Ligorria and Ammon*, 1999] gives superior results in the presence of noise. Regardless of the estimation technique, the result is a time series where time is a proxy for depth and significant positive or negative amplitudes correspond to an increase or decrease in seismic velocity, respectively. The visual interpretation of receiver functions is complicated though by multiple reverberations that occur particularly at later times, and the dependence of the time-depth mapping on the unknown velocity distribution in the subsurface.

[14] For this reason it is beneficial to model receiver functions for velocity structure, optimally through an inversion algorithm. In this case multiple reverberations are a benefit and help to constrain the position of interfaces. When RFs are used as an imaging tool rather than for modeling a common solution is to stack moveout-corrected receiver functions from different events with a wide range of ray parameters [e.g., *Nair et al.*, 2006]. While this reduces the influence of random noise and brings out primary P to S conversions, it has another effect that is beneficial for visual interpretation but detrimental for inversion. Multiple reverberations have a different moveout characteristic to primary conversions and therefore the amplitude of multiples is reduced by stacking. For visual interpretation this is beneficial as the multiples could otherwise be confused with primary conversions. When we model receiver functions, however, these multiples contain information about the Earth's structure. Furthermore if we remove them by using a stacked receiver function, we have to exclude the time windows of those multiples from the inversion as they will still be present in the synthetic receiver functions.

[15] For these reasons we stack the receiver functions sorted by ray parameters in bins with a width of 0.01 s/km. With this variation in ray parameter, the moveout of primary conversions and multiples from structures in the crust and mantle is small. While this increases the computational load compared to a single stack, it has the advantage that we preserve the information contained in the multiple reverberations. In theory the moveout characteristics also carry information about the structure of the Earth, however in practice the sensitivity is only small [*Ammon et al.*, 1990]. In any case we reduce the influence of noise on the inversion.

[16] We use a two-step approach to calculate synthetic receiver functions for the inversion. First, we calculate synthetic seismograms using the code of *Randall* [1989]. Then we use the same iterative deconvolution routine that we use for real data to calculate receiver functions from these synthetic seismograms. In order to ensure that all reverberations are captured correctly, we calculate 200 s of synthetic seismograms even though the primary conversions and crustal multiples are all contained in the first 40 s of the receiver function. Together with the fact that we have to calculate separate seismograms for each ray parameter, this makes it the most time consuming part of the inversion. The inversion spends about 70% calculating the synthetic receiver functions.

2.3. Rayleigh Wave Dispersion Data

[17] Surface waves are one of the ideal tools for studying the structure of the crust and upper mantle, and for many events the surface wave train is the largest feature of the seismogram. A good vertical resolution of the variation in shear wave speed can be obtained as the depth sensitivity of surface waves are dependent on the period of the waveform that is exploited. At periods shorter than approximately 40 s there is strong sensitivity to crustal structure, whereas at longer periods the waveform becomes increasingly sensitive to variations in shear wave speed within the upper mantle. The horizontal propagation of the surface waves means that a wide distribution of events and recorders is required to produce a reliable tomographic image, but with a good path coverage it is possible to have reasonably good lateral resolution even in areas with very few seismic stations.

[18] Within the surface wave train it is possible to make use of both Rayleigh and Love waves, and also both the fundamental and higher modes. As the level of noise is generally much higher on the horizontal than vertical component, it is the Rayleigh wave that is most frequently analyzed in surface wave studies. For this reason in this study we work with Rayleigh wave dispersion data. A variety of methods exist to extract information from the component of the seismogram produced by the surface waves; for example, group velocities can be estimated using frequency time analysis [e.g., *Ritzwoller and Levshin*, 1998], two plane wave solutions for the wave field can be found for events measured within an array [e.g., *Forsyth and Li*, 2005], and multimode waveform inversion techniques can be used to estimate the path average velocity structure required to produce the observed surface wave dispersion [e.g., *Debayle*, 1999; *Lebedev et al.*, 2005]. In the last method, if dispersion information is required the path average velocity model can be used as an estimator of the average dispersion characteristics between the source and receiver [*Kennett and Yoshizawa*, 2002].

[19] Tomographic inversions are used to combine the data from multiple events measured across the array, or for multiple source-receiver pairs, and thus provide regional maps of the dispersion characteristics at a particular period. The local dispersion characteristics at any point within the region can then be obtained by extracting the information from a series of tomographic maps across the period range of interest.

[20] We use the surface wave dispersion code *sdisp*96 that is part of the Computer Programs in Seismology (R. B. Herrmann, 2002, available at http://www.eas.slu.edu/People/RBHerrmann/ComputerPrograms.html) to calculate

the synthetic dispersion curves. One difficulty with GA based inversions is that in the early stages the models are purely random and can contain strongly alternating low- and high-velocity layers. These types of structures can be numerically unstable and present a problem to some codes. With *sdisp*96 we have not encountered such problems. Also the forward calculation is relatively fast, only 20% of the time of the inversion is spent calculating the dispersion curves.

3. Inversion Method

[21] We chose the multiobjective genetic algorithm (GA) NSGA-II [*Deb et al.*, 2002] as the optimization method for our joint inversion approach. There are two main reasons for this choice: (1) genetic algorithms in general do not depend on any linearized approximation and (2) NSGA-II in particular yields a set of final models that demonstrate to what extent fitting one data set trades off against fitting the other data sets. When jointly inverting data that are sensitive to different parameters, we cannot be sure that these data sense the same structures and can be described by a joint model. We use the trade-off as an indicator of the compatibility of the data sets. These advantages come at the cost of a much higher computational load, as the lack of gradient information and the stochastic nature of the algorithm require a large number of forward calculations. In the following we will describe the basic concept of genetic algorithm inversion with particular focus on NSGA-II and our joint inversion problem.

[22] Genetic Algorithms (GA) are a class of stochastic optimization methods similar to Monte Carlo and Simulated Annealing approaches [e.g., *Sambridge and Mosegaard*, 2002]. The main difference of GAs to these two is that usually the algorithm does not directly work with the model parameters, but an encoded binary form that mimics RNA in biological evolution [*Goldberg*, 1989]. We therefore have to specify three parameters for each model parameter m_i: The minimum possible value $m_i^{\min}$, the discretization step Δm_i, and the number of bits used for the encoding N_i. Consequently genetic algorithms that use this form of representation are inherently constrained optimization methods. For geophysical applications this presents no problem, as reasonable and physically realistic values for the model parameters are usually known. For example, we search an S velocity range between 2.5 km/s and 5.6 km/s which covers realistic values in the crust and upper mantle [*Dziewonski and Anderson*, 1981]. For the MT data the apparent resistivity curve gives an indication of the order of magnitude of the expected resistivities and we search for resistivities between 1 Ωm and 10^6 Ωm. More important is the impact of the discretization step Δm_i. In the inversion procedure we have to assure that this step size is small enough to avoid aliasing effects; again for geophysical applications there are reasonable estimates of what minimum variation we can consider useful and the misfit surface is not that highly nonlinear that very small changes can have a major impact on the data fit. We use a step size of 0.1 km/s for the seismic model and 0.1 for logarithmic resistivity for the electric model. Comparison with inversions with finer discretization shows that we obtain essentially the same results but with less forward evaluations.

[23] After we have chosen encoding parameters for each model parameter, the genetic algorithm creates a random population of fixed size P. Each member of this population represents one possible model. For each population member $\mathbf{m}$ we transform the encoded bit string back into physical parameters and calculate the objective function value $O_f(\mathbf{m})$ for each objective function, e.g., the misfit of each data set and possible regularization functionals. NSGA-II preserves the multiobjective nature of the problem by storing the objective function values in a vector $\mathbf{O}(\mathbf{m})$ for each population member. This is a major difference to linearized schemes and some GAs that minimize a weighted sum of the objective functions values [e.g., *Julia et al.*, 2000; *Gallardo and Meju*, 2003; *Linde et al.*, 2006]: in our case we do not need to specify the nature of the weighting of the different data. This is a significant advantage and is discussed in greater detail below.

[24] The next step in the algorithm is the selection of models for the next generation. Each population member is assigned a rank based on the concept of Pareto optimality. A population member $\mathbf{m}_i$ is said to be partially less than the member $\mathbf{m}_j$ if all objective function values are less or equal and at least one is less for $\mathbf{m}_i$,

$$\mathbf{m}_i \prec_p \mathbf{m}_j \Leftrightarrow \begin{array}{ll} \forall k: & O_k(\mathbf{m}_i) \le O_k(\mathbf{m}_j) \text{ and} \\ \exists k: & O_k(\mathbf{m}_i) < O_k(\mathbf{m}_j). \end{array} \quad (5)$$

As we can see from equation (5), the algorithm only compares objective function values for each data set or regularization functional. If there is no member in the population that is partially less then a given member $\mathbf{m}$, $\mathbf{m}$ is called locally Pareto optimal and assigned a rank of one. Conversely population members that are partially less than another member of the population are called dominated by that member. Note that there can be more than one locally Pareto optimal member in the population, the set of Pareto optimal members forms the so-called Pareto front; we will discuss the meaning and the implications of this below. After all locally Pareto optimal members have been identified, they are removed from the ranking procedure. Members of the population that are Pareto optimal after the removal of the first front are assigned a rank of two and the process is repeated with the remaining members with an increased rank until the whole population is ranked. We show a graphical representation of Pareto optimality and the ranking process in Figure 1.

[25] After the population has been ranked, candidates for the new population are selected by binary tournament selection. Two random members are drawn from the population and the one with lower rank is selected. If both have the same rank, the member with the larger crowding distance, i.e., the average distance to neighboring models in objective function space [*Deb et al.*, 2002], is used as a secondary criterion. If both criteria are equal one member is chosen randomly.

[26] So far the GA has not introduced any variations of the models. This is done by using crossover between models and mutation of individual models. With a probability p_c two models are selected for crossover and exchange their binary representation after a randomly chosen location. This process is very effective at distributing successful segments of the model across the population [*Goldberg*, 1989]. Muta-

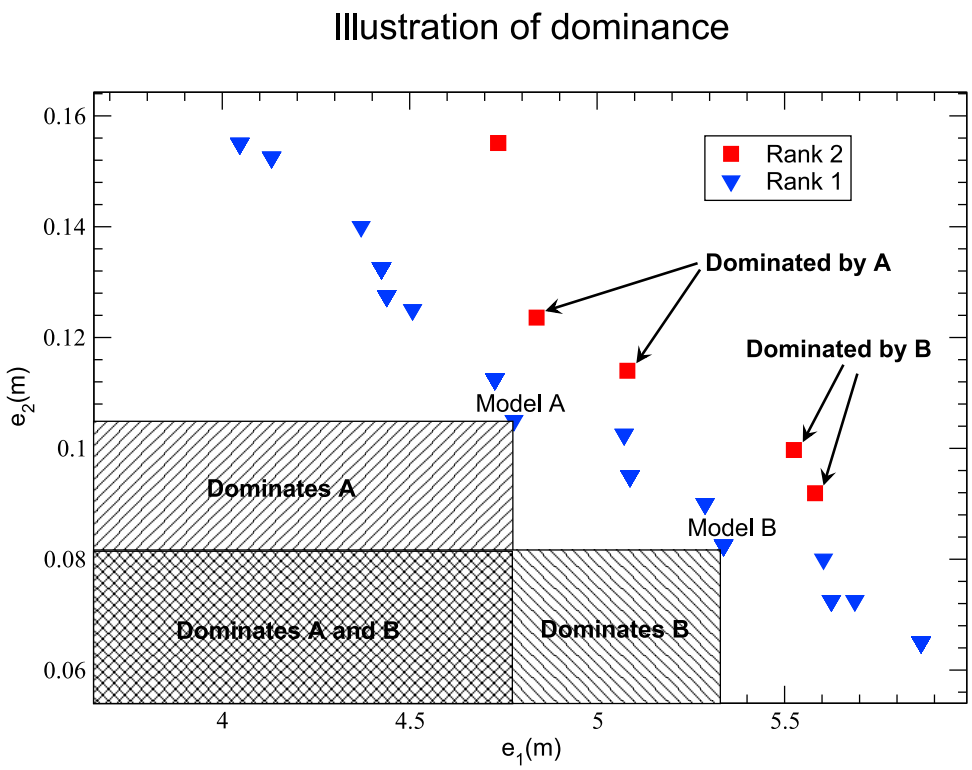

Figure 1. The concept of Pareto optimality. For two selected models from the Pareto optimal front, labeled model A and model B, we show the region in objective function space that these models would be dominated by and the members of the population these two model dominate. We also show the rank the population members would be assigned in the genetic algorithm.

tion on the other hand is a purely unstructured process; with a probability p_m one bit changes its value. This helps to explore completely new regions of model space.

[27] Before we start the new iteration, we ensure that the best models are always preserved by merging the new population with the old one, ranking the joint population and taking the P best models based on rank and crowding distance. With this new generation we repeat the above steps for a fixed number of iterations.

[28] The advantage of the concept of Pareto optimality over a weighted sum approach for joint inversion might not be intuitively clear. However, we consider it as one of the key advantages of optimization with NSGA-II and will therefore illustrate its merits here. For this purpose we use a simple regularized one-dimensional inversion problem for MT data. We can regard this as a joint inversion problem where we want to minimize data misfit and model complexity simultaneously and use it as a simple example to explain some basic ideas before we discuss joint inversion of several data sets. We generate synthetic MT data from a one-dimensional model, add random noise to the data and invert the two objective functions MT data misfit and resistivity variation using our joint inversion algorithm.

[29] The set of locally Pareto optimal models for selected generations in the inversion is shown in Figure 2a. We can readily identify the competitive nature of the two objective functions, data misfit and model complexity, from this plot. Within each front, models with a higher misfit have a lower roughness and vice versa. With successive generations the front on average moves closer to the origin, but at the same time maintains a diversity of models with a large span of misfit and roughness values. When the algorithm has completed, we have obtained an approximation to the L curve [*Hansen*, 1992] and can use criteria like maximum curvature or generalized cross validation [*Farquharson and Oldenburg*, 2004] to identify an appropriate model. Figure 2b shows three models from the final Pareto front, their position within the front is marked in Figure 2a.

[30] The results are similar to the Occam approach [*Constable et al.*, 1987], with the difference that now we obtain a range of models from which we can choose one that we regard as optimal in some sense. When we invert several data sets the algorithm will also yield a trade-off curve, or a multidimensional trade-off surface for three or more objective functions. In the ideal case where both data sets uniquely sense compatible structures and the genetic algorithm has fully converged, this curve will consist of only one point. In practice this will not be the case even when the data sets sense identical structures as each data set is affected differently by noise. Also, we are solving the problem in a discrete rather than a continuous manner. We will therefore always obtain some sort of trade-off curve. In the synthetic examples below we will show how we can use this information to test our assumptions about the level of noise in the data.

[31] When the data sets are incompatible, i.e., sense different structures that cannot be accommodated simultaneously in the joint model, we obtain a trade-off curve similar to the trade-off for the regularized inversion problem shown before. We will not be able to find a single model

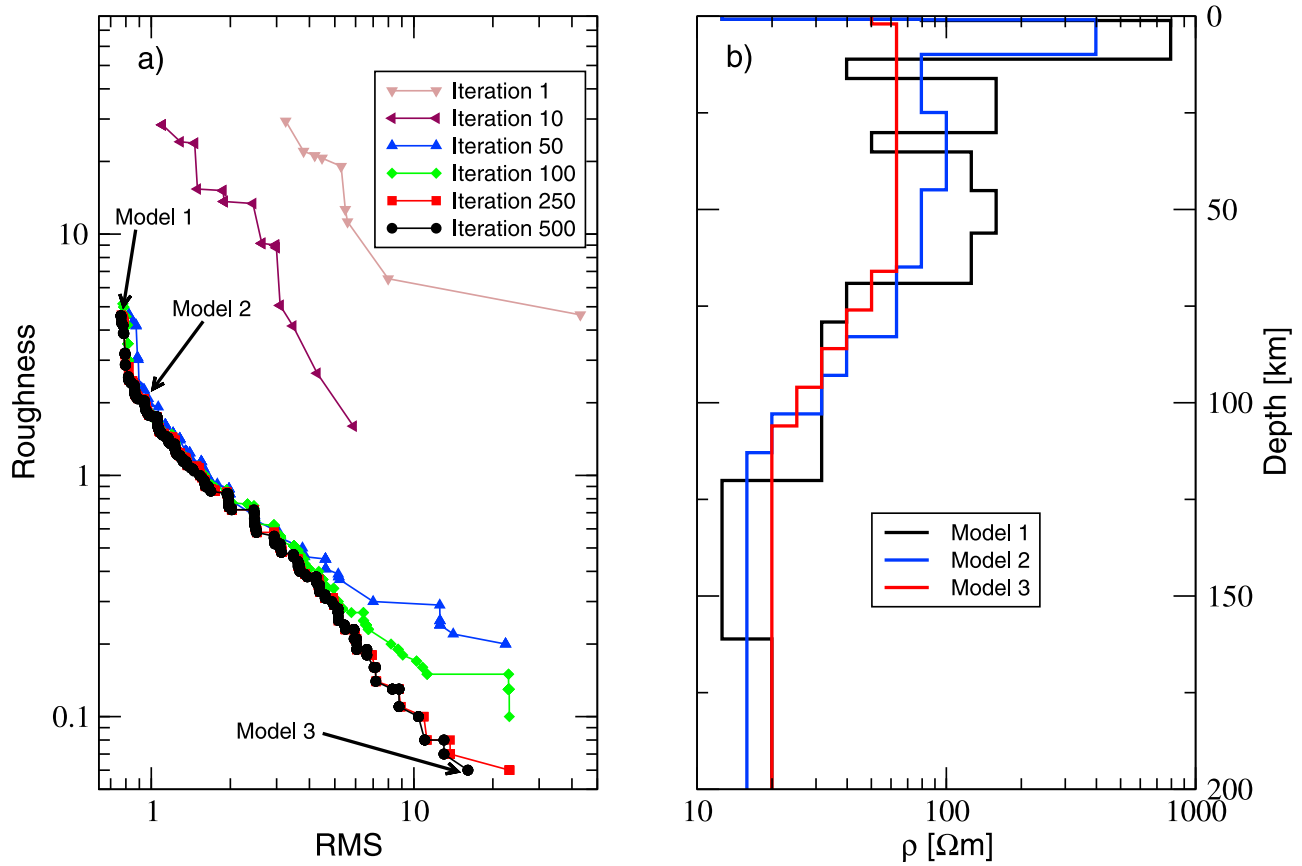

Figure 2. Evolution of the Pareto front for the regularized inversion example. (a) The evolution of the Pareto front through selected iterations and the position of the models in Figure 2b within the final Pareto optimal front. (b) Three models from the final iteration, the two extremal models and one with intermediate roughness and misfit.

that fits both data sets simultaneously. For more than two data sets the situation becomes more difficult to analyze, as we have to deal with a trade-off surface that becomes difficult to visualize. We will present a simple but reliable scheme to identify compatible and incompatible situations below.

[32] For the joint inversion we use a similar setup as in the previous study [*Moorkamp et al.*, 2007]. We invert for isotropic S velocity, isotropic resistivity and layer thickness for a fixed number of inversion layers. P velocities and densities for each layer are calculated by multiplying by a factor of $\sqrt{3}$ and using the empirical relationship of *Berteussen* [1977], respectively. We use a normalized RMS misfit Φ for each data set,

$$\Phi = \frac{1}{N}\sqrt{\sum_{j=1}^{N}\left(\frac{d_{obs} - d_{synth}}{\Delta d_{obs}}\right)^2}.$$

Dividing each datum by the estimated observation error Δd_{obs} makes the misfit dimensionless and it reaches a value of unity when the model on average fits the data within the observational error.

[33] We achieve the coupling between the electrical and seismic structure by using a single set of thickness values for the inversion layers. Therefore a layer boundary in the seismic model is automatically a layer boundary in the electrical model, although we do not require a change in resistivity across that boundary. This type of setup is similar to the cross-gradient approach [*Gallardo and Meju*, 2003], and presents the weakest possible coupling between the data sets. Considering that the relationship between seismic velocities and resistivities at the scale of the lithosphere is largely unknown, this type of setup imposes the least assumptions on the joint inversion problem. Stronger assumptions, such as empirical relationships between conductivity and resistivity suggested by *Jones et al.* [2009], would increase the coupling between the data sets but also the possibility of totally misleading and erroneous results.

4. Synthetic Tests

[34] We perform a number of tests with synthetic data to demonstrate that we can identify situations where a joint modeling approach is sensible and where it is not, and that our approach yields meaningful results in cases where a joint model is appropriate. We construct two data sets for these tests. The receiver function and surface wave data are identical for both data sets. From the model shown in Figure 3c we calculate synthetic seismograms with ray parameters of 0.05, 0.06 and 0.07 s/km. We add random Gaussian noise with an amplitude of 2% of the direct

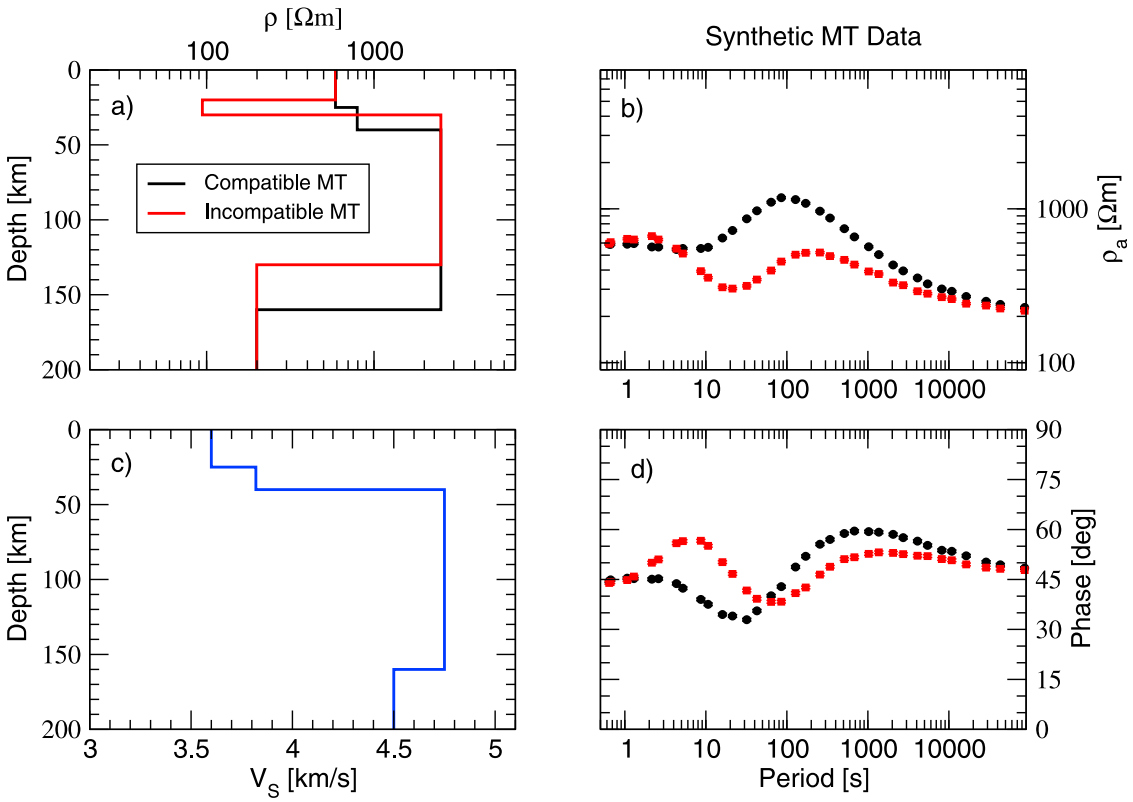

Figure 3. The synthetic models and data used to test our inversion algorithm. (a) The two MT models we combine with (c) the seismic model to construct the "compatible" and "incompatible" data sets. (b) Apparent resistivity and (d) phase for the two MT models.

P wave amplitude to the radial and vertical components and calculate synthetic receiver functions using the iterative deconvolution technique [*Ligorria and Ammon*, 1999]. From the same model we calculate synthetic seismograms for surface wave analysis, add 2% random Gaussian noise to each seismogram and estimate phase velocities using the tools and procedures described by R. B. Herrmann (Computer Programs in Seismology, 2002, available at http://www.eas.slu.edu/People/RBHerrmann/ComputerPrograms.html).

[35] The first data set is calculated from the compatible MT model in Figure 3a. The resistivities in this model are simply a scaled version of the seismic velocities, and therefore all interfaces are at exactly the same depths. We add 5% random Gaussian noise to the synthetic data calculated from this model and we will call the combination of this data with the seismic data the "compatible joint inversion data set."

[36] The second MT data set is calculated from the incompatible MT model in Figure 3a. Here we have added a conductive layer in the lower crust, a common observation in magnetotelluric data sets [*Jones*, 1992], and resistivity does not change across the Moho. Furthermore the conductive layer that simulates the lithosphere-asthenosphere boundary is shallower than the corresponding low-velocity layer in the seismic model. As above we add 5% random Gaussian noise. We will call the combination of this data with the synthetic seismic data the "incompatible joint inversion data set."

[37] Our first experiment demonstrates how we can identify the noise level in the seismic data set and distinguish the compatible and incompatible data set in the presence of noise. We run the inversion with 4 layers on both data sets and additionally on the compatible data set without any noise. This is the smallest number of layers that can reproduce the compatible data set and therefore provides the strongest coupling. Figure 4 shows the trade-off plots for the three data sets. As we are trying to fit MT data, receiver functions and surface wave dispersion data simultaneously, the resulting trade-off is a three-dimensional convex cloud of points. We project the trade-off onto the three planes spanned by the coordinate axes and show the third dimension by the color of the points.

[38] For the inversion run without added noise we observe misfit values below an RMS of 1 for all data sets and all models. We assumed the same noise level as for the data sets with added noise for the misfit calculation so this result is not surprising. The trade-off between receiver functions and surface waves shows a single model with significantly smaller misfit values then all other models in the Pareto front. We mark this model with a yellow triangle in Figures 4a–4c. Theoretically we expect only a single solution when inverting noise free synthetic data with the same number of layers as the model the input data were generated from. Two factors contribute to the formation of an extended front. First, to guarantee searching all model space we would have to run the genetic algorithm for an infinite number of iterations. In practice the GA approaches the vicinity of the solution fast, but full convergence of the Pareto front to the single optimal solution takes a large number of iterations.

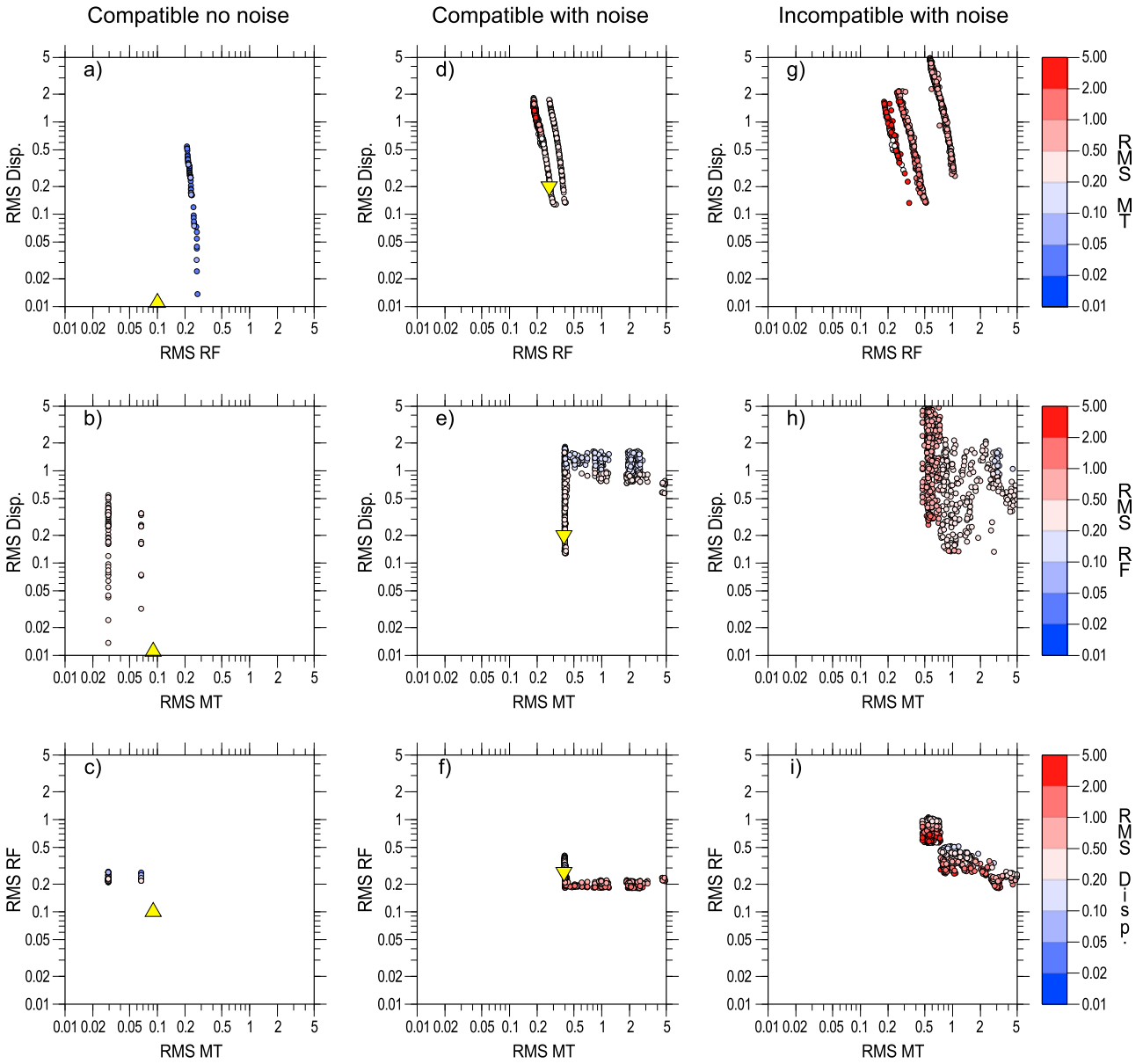

Figure 4. Trade-off plots for (a–c) the compatible problem without noise, (d–f) the compatible problem with noise, and (g–i) the incompatible problem. Within each plot we show two objective function values against each other and color each point by the third objective function value. The optimum model for the compatible case without noise is marked by a yellow triangle. We mark the compatible problem with noise with an inverted triangle.

Second, the exact solution might not be reachable for the GA due to the chosen parameter discretization. In this case parameter values above and below the true solution that can be represented by the GA appear in the Pareto front. This effect can be seen in Figure 4c where the front clusters in three areas. As we chose a discretization step of 0.01 km/s and 0.01 for the logarithmic resistivity for this test the differences between the three solutions are below the resolution of real data and comparison with the true model shows that all three are practically identical.

[39] Compared to the noise free case the trade-off plots for the compatible case with added noise (Figures 4d–4f) span a larger range of misfits for all three data sets, as expected we do not reach as low misfits as for the noise free case. There is also an indication of an L curve for the trade-off between the two seismic data sets that have been generated from the same model. The misfit for the surface wave dispersion data varies by more than an order of magnitude, while the misfit of the receiver functions varies by a factor of two. This is an effect of the noise added to the synthetic data. The model that fits the receiver function data best, also fits aspects of the noise that are not compatible with the dispersion data. We can therefore use the trade-off to check the accuracy of our noise estimate. Here we use the maximum curvature criterion [*Hansen*, 1992] that suggests an RMS of 0.27 for the receiver function data and 0.2 for the dispersion data, we mark this model with an inverted yellow triangle in Figures 4d–4f. In both cases the processing applied to the

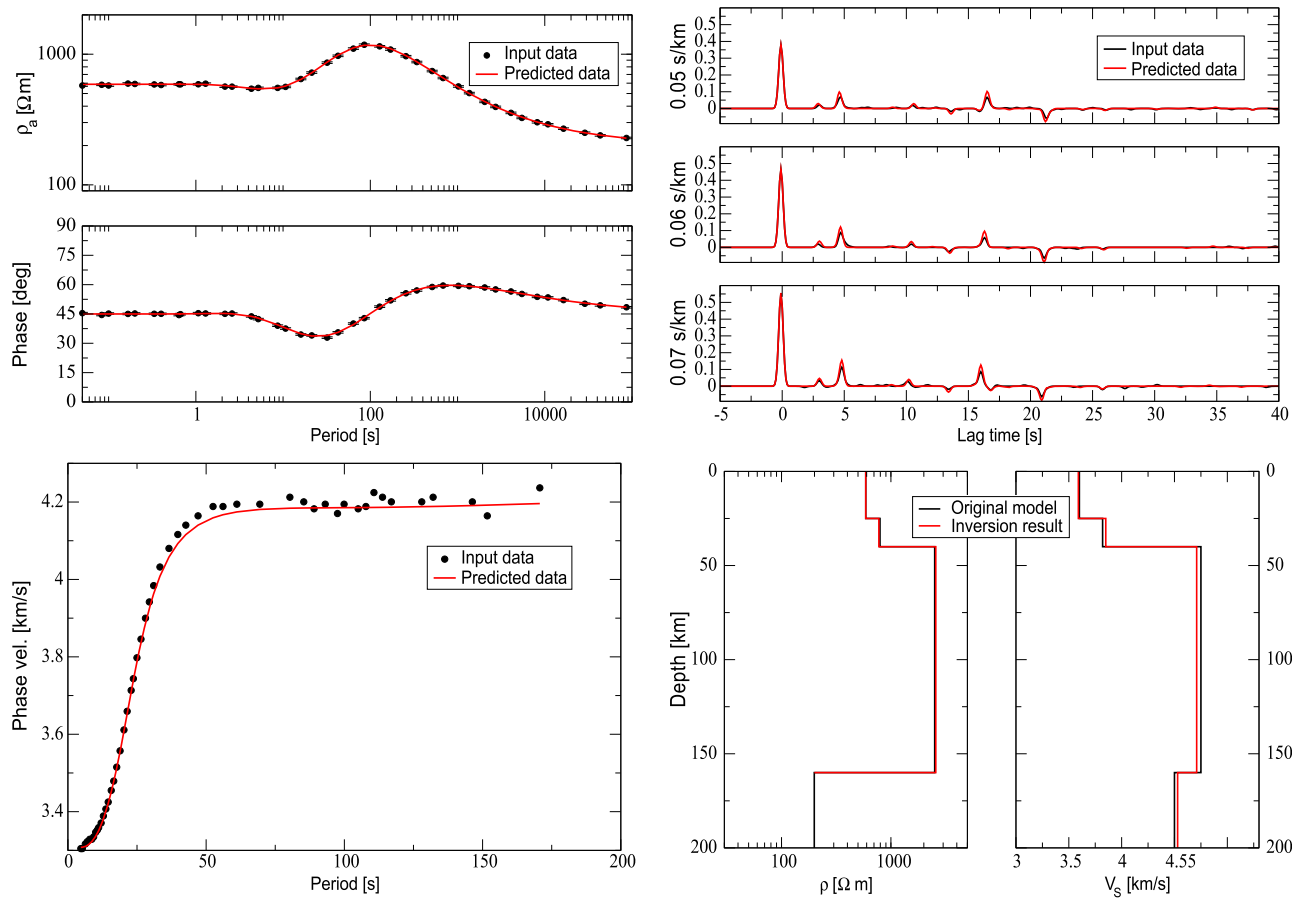

Figure 5. Comparison between inversion results and synthetic input data for the compatible case with added noise.

noise contaminated synthetic seismograms has reduced the effective noise on the data used within the inversion. This impression is confirmed by visual inspection of the computed receiver functions, for example, and demonstrates the robustness of the iterative deconvolution approach. The genetic algorithm has therefore helped to identify the appropriate level of noise to avoid overfitting the data.

[40] The trade-off front between surface wave dispersion data and MT (Figure 4e) for the compatible case is nearly a straight line, a large number of models with varying dispersion misfit has nearly identical MT misfit. This is a result of the coupling we use between the electrical and seismic parameters. All these models have identical layer thicknesses and resistivities, but varying seismic velocities. As we can see the model with the optimum trade-off for the seismic data is part of this front and also minimizes the MT data. The trade-off between the MT data and the receiver function data shows a sharply bent L curve with the optimum model in this case not at the point of maximum curvature. Again this is a result of the noise added to the data. The models in this area have a high dispersion misfit. The model determined this way for the compatible case matches the layer geometry of the synthetic model used to generate the data exactly, the seismic velocities within each layer differ by up to 0.02 km/s and 2% for the resistivities (see Figure 5).

[41] For the incompatible test case the trade-off plot for the two seismic data sets looks similar to the trade-off plot for the compatible case. Although the seismic data sets are identical in both cases we do not obtain identical trade-off curves for these data sets for two reasons. First, the genetic algorithm is a stochastic method that despite some mechanisms to enhance the diversity of the population has a tendency to find models that fit all data sets well, but is less efficient at finding extreme models that fit only one data set [*Goldberg*, 1989]. These extreme models might not appear in every run of the GA. Second, the interaction of the Pareto front with the incompatible MT data precludes some models from the front, its shape in all dimensions depends on the interaction of all objective functions and these effects cannot simply be separated. Still, the optimum misfit that we identify for the seismic data for this case is identical to the compatible case, as we expect for identical input data. This also demonstrates that the GA yields robust results.

[42] At first glance the trade-off plot between surface wave data and MT data for the incompatible case (Figure 4h) looks similar to the compatible case. In both cases we obtain an L-shaped cloud of points. It appears though that for the incompatible case the models with low MT and surface wave misfit all have high receiver function misfit; significantly higher than the optimum misfit we determined from the trade-off of the seismic data. However from such a cloud of points

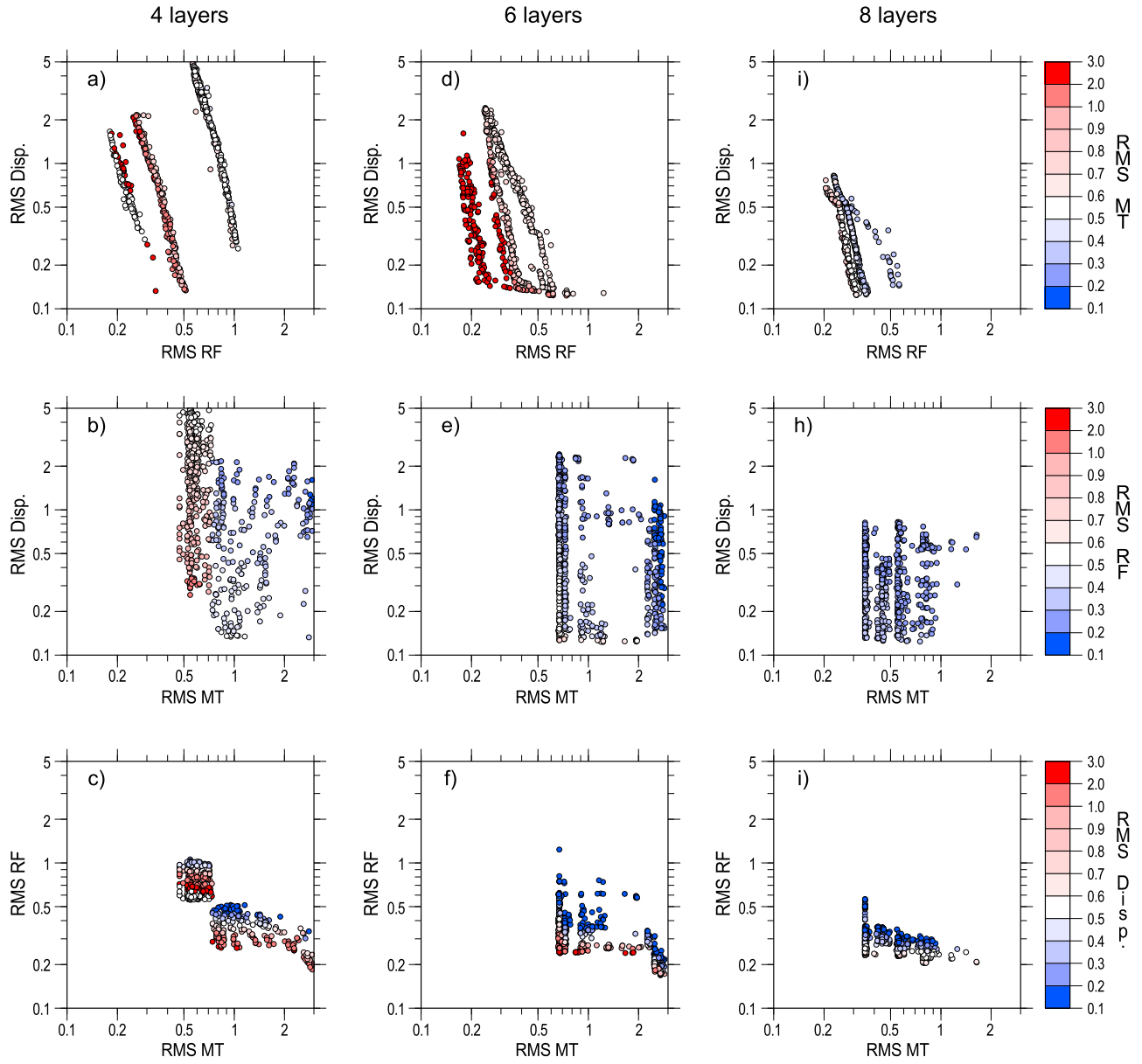

Figure 6. Trade-off plots for incompatible synthetic test problem when running the inversion with (a–c) four layers, (d–f) six layers, and (g–i) eight layers. Within each plot we show two objective function values against each other and color each point by the third objective function value.

it is difficult to infer this reliably and we will quantify this impression below when we apply our selection scheme.

[43] The trade-off plot between receiver function and MT data shows a pronounced difference between the compatible (Figure 4c) and incompatible cases (Figure 4f). In the incompatible case there is a significant gap in the distribution of models in the lower left corner, revealing that there are no models that acceptably fit both data sets simultaneously. In addition, the models with MT misfit below 2 and a RF misfit that approaches the optimum misfit have a high surface wave dispersion misfit. From this plot it appears to be impossible to find a model that fits all three data sets simultaneously.

[44] Although these plots give a good first visual impression of the compatibility, it is important to quantify this, particularly when the Pareto front contains a significant number of models. We apply the scheme we used to identify the optimum model in the compatible case to the incompatible case as well. If we select the models with an optimum receiver function and surface wave misfit, we only find models with a MT misfit >2.5, significantly higher than the best fitting MT models. If conversely we examine the best fitting MT models, both receiver function misfit and surface wave dispersion misfit are significantly higher than the value determined from the optimum trade-off of the seismic data sets. This indicates that the data sets are incompatible.

[45] Next we examine the effect of different numbers of layers within the inversion. In the previous experiment we used the same number of layers to generate the input data

and within the inversion. In practice it is difficult to estimate the optimal number of layers so we need to find out when the coupling between the data sets becomes too weak. We use the incompatible data set for this purpose as it also demonstrates how the term incompatible depends on the number of layers.

[46] Figure 6 shows the trade-off plot for inversion of the incompatible data set with 4 layers, 6 layers and 8 layers. Generally the minimum misfit decreases with increasing number of layers, particularly for receiver functions and MT data. Also the shape of the trade-off curve changes depending on the number of layers. This can be most prominently observed for the trade-off between surface wave dispersion and receiver functions and MT and receiver functions, respectively. For the comparison between the compatible and the incompatible cases we saw that the trade-off between MT and receiver functions allows to distinguish these two cases visually. Compared to the inversion with 4 layers, the inversion with 6 layers finds models that minimize both MT and receiver function data, the region toward the lower left corner of trade-off plot (Figure 6f) that was empty when inverting with 4 layers, now contains some models. However, the surface wave dispersion misfit of these models is much higher than indicated from the trade-off between surface wave dispersion and receiver functions. There is another cluster of models however with acceptable misfit for both seismic data sets, but MT misfit >2. The difference between these models is the location of the crustal conductor that in the inversion results always coincides with the crustal velocity change. In the models used to generate the input data these two structures were at different depths and therefore we identify the incompatibility as we did in the 4 layer case. This is somewhat unexpected as 6 layers are enough to describe the incompatible seismic and electric models with a single model. The inversion, however, uses these additional layers to introduce a gradational velocity change below the Moho that does not exist in the original model. This is a result of adding noise to the receiver function data. The added noise decreases the correlation of the Moho conversion and therefore suggests a smaller impedance contrast across the Moho. In order to reconcile the smaller contrast with the surface wave dispersion data, the inversion has to introduce a sequence of small velocity changes that have little impact on the receiver functions, but raise the velocity in the mantle to a level that matches the dispersion observations.

[47] When we use 8 layers in the inversion we obtain a similar trade-off plot as for the compatible case. Now we have enough degrees of freedom to model the different structures in the electric and seismic models as well as the gradational layers introduced by the noise. Consequently the model contains layer interfaces to accommodate both the seismic and electric data. We obtain a good approximation of the crustal structure and the electric LAB, however the model does not reproduce the different position of the seismic LAB, but places the low-velocity zone with the electric conductor. As it is not a prominent feature in the receiver functions and the surface wave dispersion data does not have sufficient resolution to locate the position of the low-velocity zone, this explains both data sets well within the noise level.

[48] These synthetic tests demonstrate that under realistic conditions we can identify whether the data sets can be described by a joint model. One of the important factors for this success is a sufficient coupling between the models through a small number of layers. As our synthetic tests demonstrate the exact number of layers is not critical, but if we use a large number of layers the data sets effectively decouple and we will be able to accommodate any kind of structure. It it therefore important for the inversion of real data to at least approximately identify the smallest number of layers with which we can model each data set.

5. Application to Data From Northern Canada and South Africa

[49] We apply our approach to data from two Archean cratonic regions, the Slave Craton and the Kaapvaal Craton. Both areas contain diamond-bearing Kimberlites and consequently have been the targets of various geophysical experiments [e.g., *Jones et al.*, 2003, 2009]. Furthermore the similarities and differences between these two cratonic platforms have been a long-standing question [*Jones et al.*, 2003]. A detailed comparison of the structure of the two cratons is beyond the scope of this paper; we will focus on the analysis of one site from each region. While this does not permit a detailed interpretation it provides a first hint at possible similarities.

5.1. Slave Craton

[50] We use data from site EKTN of the POLARIS network [*Eaton et al.*, 2004] recorded between 2002 and 2005 to calculate P receiver functions. We use receiver functions from 27 events that show a clear initial correlation peak, Moho conversions and multiples and stack them in 4 bins with a width of 0.01 s/km each between 0.04 s/km and 0.07 s/km. The parameters for the events used to calculate the receiver functions are listed in Table A1. The MT data were recorded at a nearby station that was part of the Lithoprobe SNORCLE project [*Jones et al.*, 2003] and are the same data we inverted previously [*Moorkamp et al.*, 2007]. We add fundamental mode Rayleigh wave phase velocity data derived by *Chen et al.* [2007] directly in the joint inversion.

[51] As before we invert for layer thickness t, logarithmic layer resistivity $\log\rho$ and S wave velocity V_S. We calculate V_P with a fixed factor of $\sqrt{3}$ and density ρ using the empirical relationship of *Berteussen* [1977]. We run the genetic algorithm with a population size of 1000 for 200 iterations to ensure sufficient convergence to the true Pareto front [*Corne and Knowles*, 2007]. Comparison of results from different inversion runs confirms that the results are stable with these parameters. Apart from the additional data, another major difference to our previous setup is that we do not prescribe any crustal velocity, but determine the seismic velocity of all layers within the inversion.

[52] Figure 7 shows the trade-off plots for the three data sets produced by a single run of the inversion algorithm for inversions with 7 layers, 9 layers and 11 layers. The trade-off between fitting the receiver function and the dispersion data shows a more pronounced L shape than we observed in the synthetic experiments and shows similar behavior regardless of the number of layers used in the inversion. The optimum trade-off point in the Pareto front yields an RMS of 0.3 for the surface wave dispersion data and 2 for the

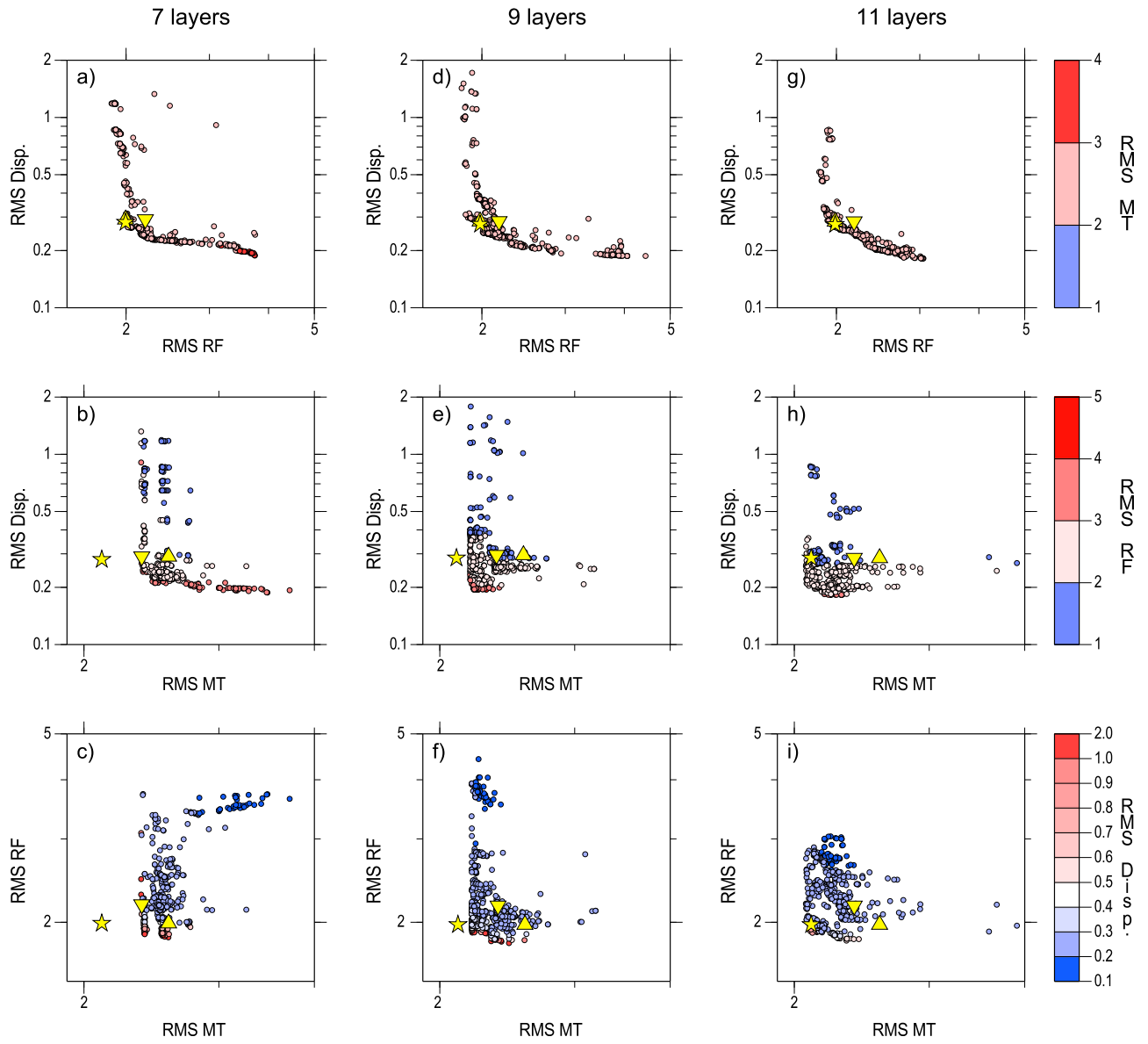

Figure 7. Trade-off plots for the data from site EKTN. Analogous to Figure 4, in each plot we show two objective function values against each other and color the points by the third objective function value. Three selected models are marked by yellow symbols; their meaning is discussed in section 5.1.

receiver function data and indicates that our noise estimate of 2% for the surface data was too pessimistic and 2% for the receiver function data was too optimistic. As we reach this optimum trade-off point for the seismic data sets for all three inversion, this seems to indicate that 7 layers are enough to model both seismic data sets adequately.

[53] However, the trade-off between the two seismic data sets and the MT data shows systematic variation with the number of layers used in the inversion. As expected the minimum RMS for the MT decreases with increasing number of layers. More importantly, the difference between the best fitting MT model and the model that corresponds to the optimal seismic model also decreases with increasing number of layers. As the synthetic tests demonstrate, increasing the number of layers increases the ability of the inversion to accommodate seismic and electric structure through decoupling.

[54] For the inversion with 7 layers we can either find a model with optimal misfit for the seismic data sets, but with an MT misfit significantly higher than the best models (marked by a triangle in Figure 7), or with optimal MT and surface wave dispersion misfit, but significantly higher receiver function misfit (marked by an inverted triangle in Figure 7). This is also true for the inversion with 9 layers and only when we use 11 layers do the differences between the MT misfit of the optimal seismic model and the optimal MT model become insignificant. We mark the optimum model with 11 layers with a star in Figure 7. We can see how the best model changes with the number of layers and which structures are constrained by which data by com-

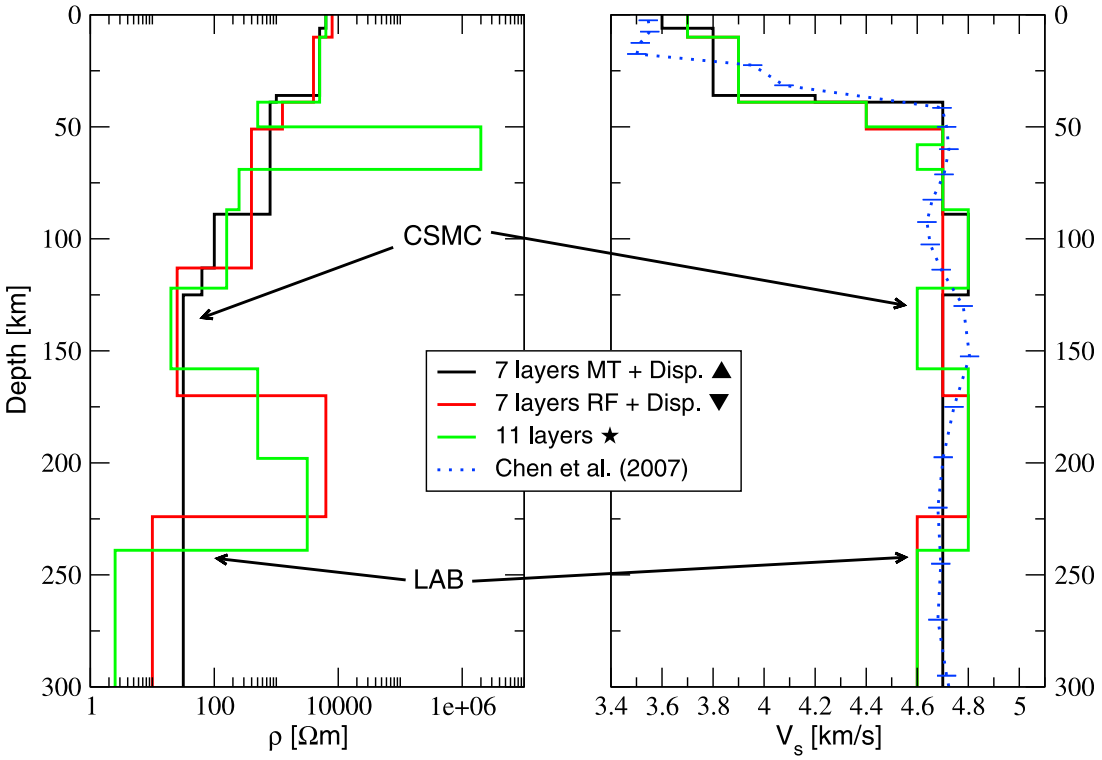

Figure 8. The electric and seismic parameters for the three models for site EKTN marked in Figure 7 plus the seismic surface wave model by *Chen et al.* [2007].

paring the two models for the 7 layer inversion with the best model for the 11 layer inversion (Figure 8).

[55] All three models have a number of common features. We can identify a clear seismic Moho at a depth of 39 km in the models that fit the receiver function data well and at 36 km for the 7 layer model that fits the MT data better. In our previous study [*Moorkamp et al.*, 2007] we located the Moho at a depth of 35 km. Without surface wave data we had to prescribe a crustal velocity and we chose 3.4 km/s based on results by *Perry et al.* [2002]. Our new results require a significantly larger average crustal velocity of 3.8 km/s. Attempts to invert the data with a fixed crustal velocity of 3.4 km/s as before resulted in much larger misfits for at least one of the seismic data sets. With the updated crustal velocity the depth of the Moho shifts to 39 km in accordance with previous studies [*Bank et al.*, 2000].

[56] Two low-velocity and high-conductivity zones between 120 km and 160 km and below 220 km, respectively, appear in various expressions in all three models. These are the Central Slave Mantle Conductor (CSMC) and the lithosphere-asthenosphere boundary (LAB) that we also observed in our previous study. The 7 layer models show either the top of the CSMC or the bottom and only one shows the LAB in the seismic velocities. The 11 layer model includes all these features. There are some variations in the interface positions for the CSMC and LAB between the 7 layers models and the 11 layer models, but these are within the limits of resolution we can expect for the inversion. Compared to our previous results the negative velocity anomaly of the CSMC is less pronounced. The change in velocity is due to the influence of the dispersion wave data that regularizes the inversion and does not permit very low velocities, but also to the additional receiver function data we are using here. Figure 9 shows the data we use in the inversion. We can see that the receiver function stacks with a ray parameter of 0.04 s/km and 0.07 s/km show only a very weak negative conversion associated with the CSMC, while the other two stacks show a much more pronounced conversion that we do not explain with our inversion result. Consequently our inversion result is a compromise between the different aspects of the seismic data.

[57] Another difference between the models is the location of the lowermost low-velocity-low-resistivity zone that we previously interpreted as the LAB. Given the generally higher seismic velocities compared to our previous study it is not surprising that the seismic interface is located deeper in both models at 225 to 235 km. For the MT and the dispersion data this depth range corresponds to data at the longest periods available and the impact on the predicted data is low. We therefore can only infer that the LAB has to be located between 220 and 235 km as we did previously. Despite some differences between the different inversion results in this study and the results of our previous study, the general features, Moho, CSMC and LAB remain the same. We prefer the 11 layer model, as it explains all data sets simultaneously, but a comparison of the two 7 layer models that we presented leads to the same conclusions, demonstrating the robustness of our approach and the usefulness of a general indicator of compatibility of the inversion results.

[58] The main source of incompatibility between the seismic and the MT data in the 7 layer case is the CSMC that cannot be described adequately by such a model. However, there is also a difference in the location of the first conductive layer that appears to be associated with the Moho in the

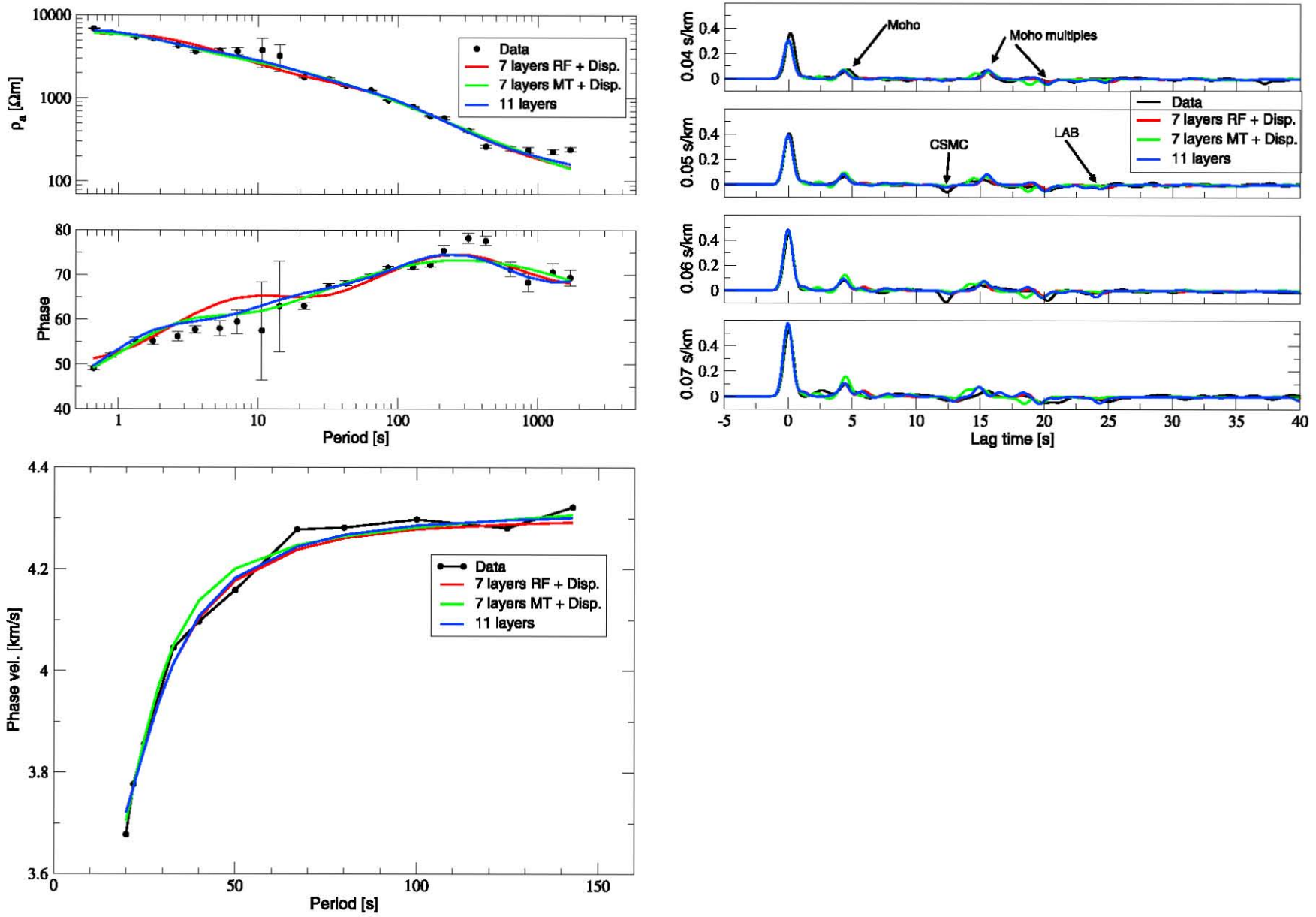

Figure 9. Comparison between measured data (black) and predicted data for the models shown in Figure 8.

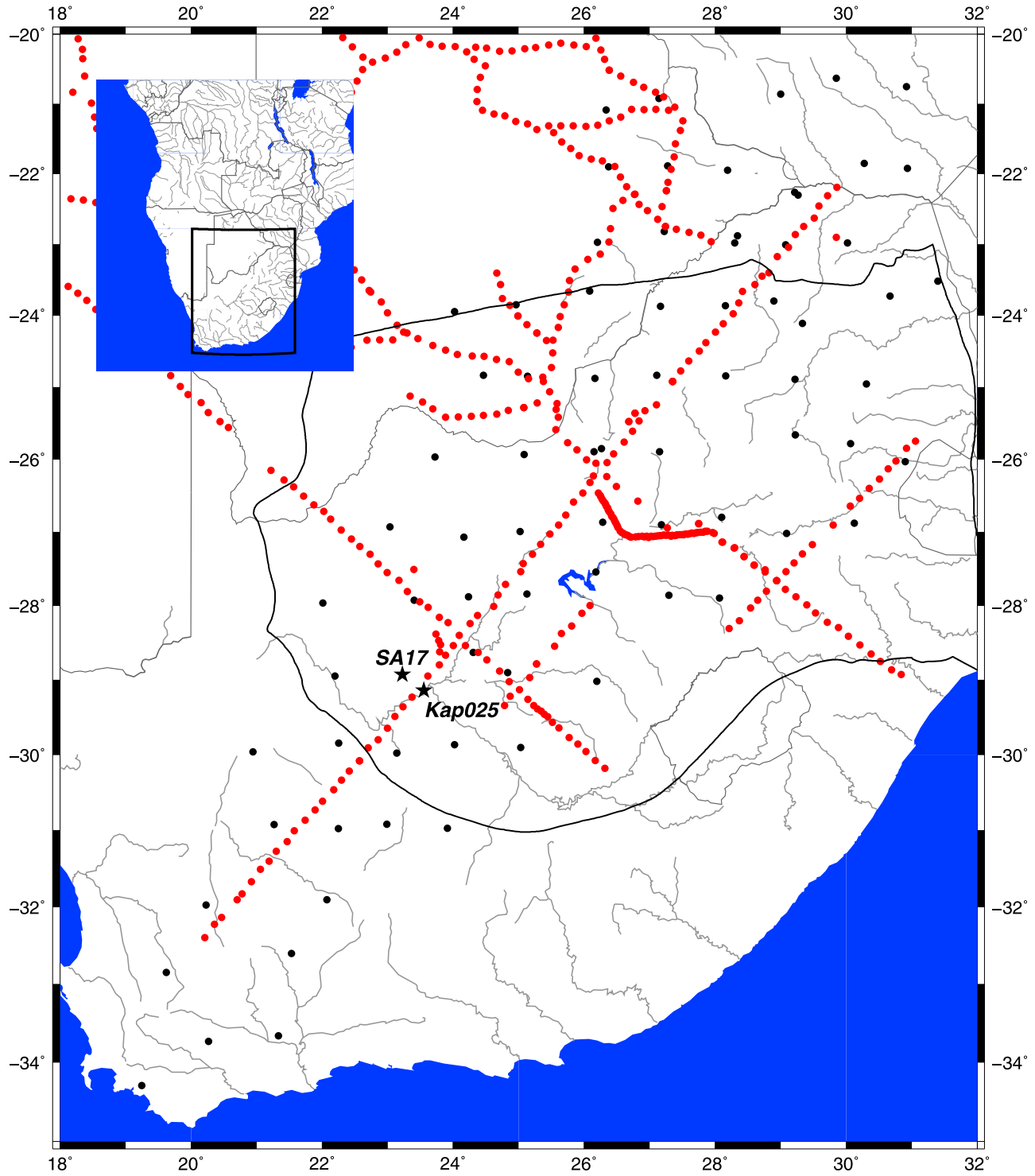

Figure 10. Locations of the sites for the SAMTEX and SASE experiments. The red dots mark the locations of SAMTEX magnetotelluric measurements, while the black dots show the locations of SASE seismic stations. We selected the MT site KAP25 and the seismic site SA17 for our analysis.

7 layer model that fits the RF data better and in the 11 layer model, but appears to start just above the Moho in the other 7 layer model. In our previous study we speculated whether this layer is associated with the Moho that we then located at 35 km. We also note that there is a consistent discrepancy between observed and predicted MT phases between 1 and 10 s (Figure 9).

[59] Genetic algorithms are very good at finding stable overall solutions but can fail to reproduce secondary features that are located at extreme ends of the trade-off curve

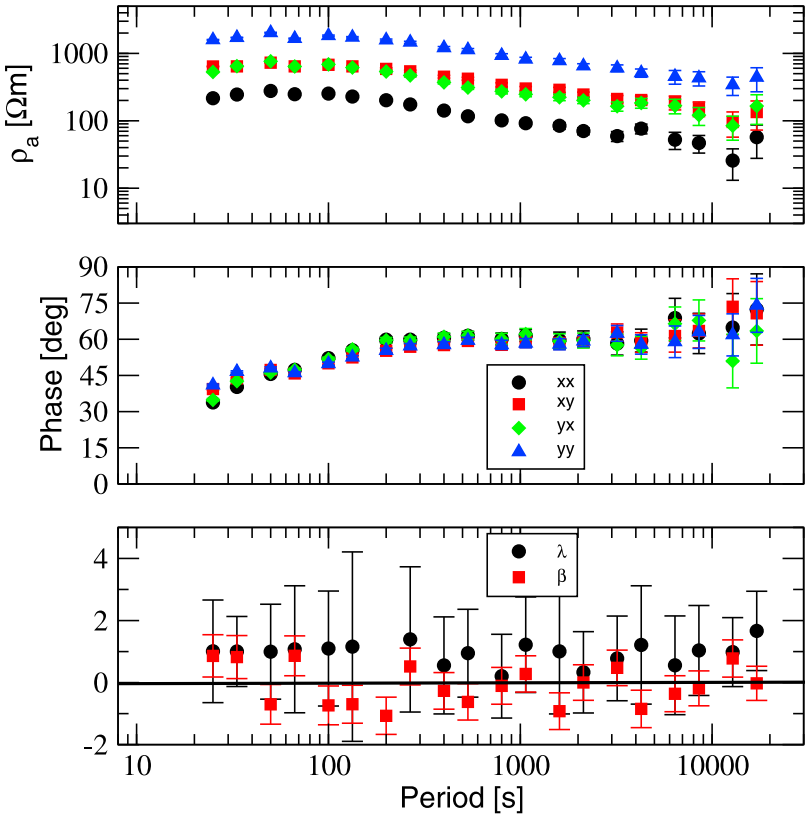

Figure 11. The magnetotelluric data for site KAP25. We show (top) apparent resistivity and (middle) phase for all four tensor components. (bottom) The phase tensor ellipticity λ and the phase tensor skew β with error bars calculated with a bootstrap approach.

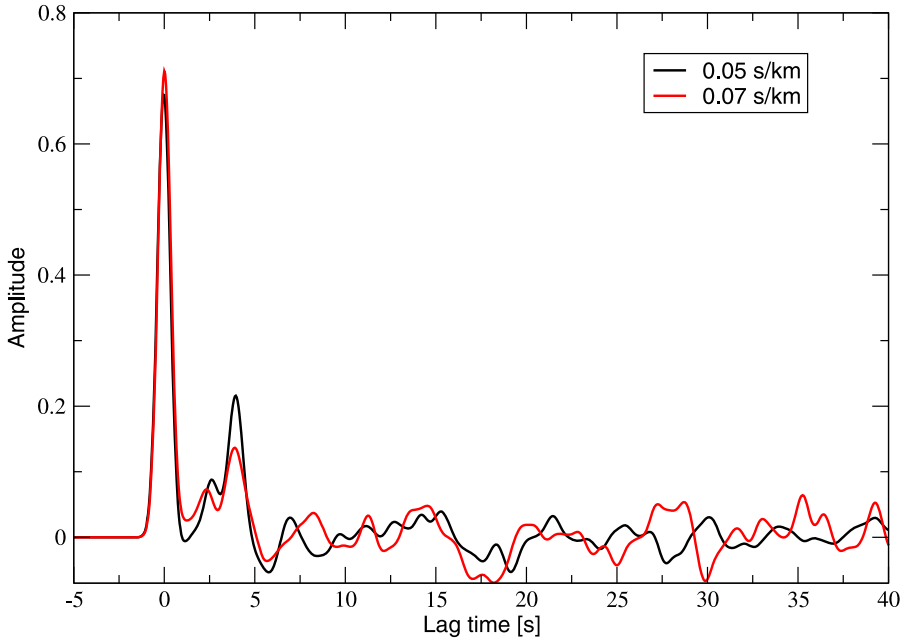

Figure 12. The stacked receiver functions for site SA17 for two different ray parameters. The parameters for the events in these stacks are listed in Table A2.

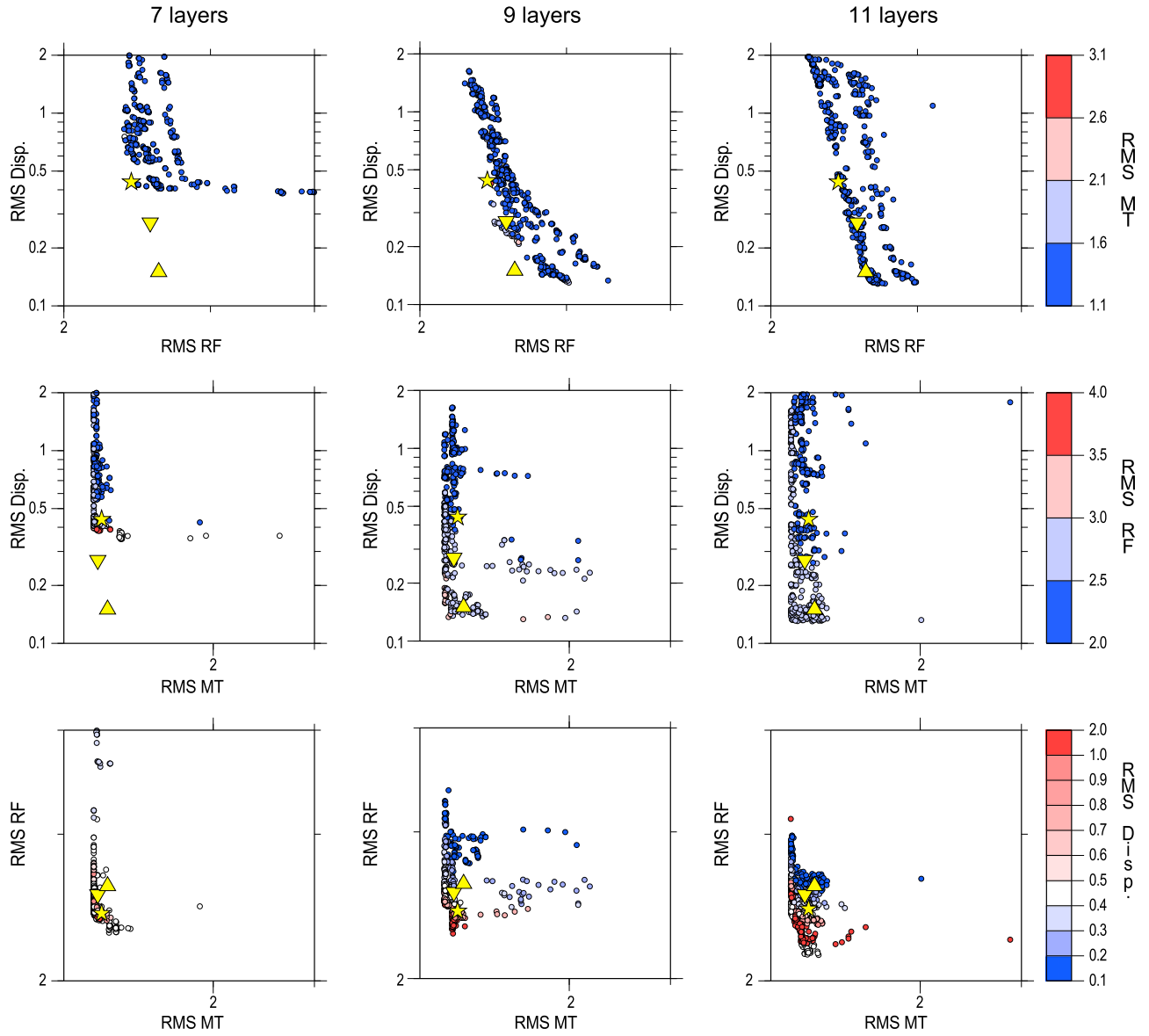

Figure 13. Trade-off plots for site KAP25 for 7 layers, 9 layers, and 11 layers. As for the Slave Craton we mark three representative models that we discuss in section 5.2.

as mentioned above. We therefore run a test inversion in which we fix all structure below 50 km depth but introduce an additional crustal layer and search with smaller discretization steps of 0.5 km for layer thickness, 0.05 for logarithmic resistivity and 0.01 km/s for seismic velocities. The resulting models (not shown) have a significantly lower MT misfit, and even slightly lower misfit for the seismic data sets. The decrease in misfit for the MT is due to a better agreement in the critical period range for the phases and these models show that the top of the conductor should be at a depth between 25 km and 35 km. This conductive layer is also associated with an increase in seismic velocity. Although the influence on the predicted seismic data is barely visible, the reduced misfit at least supports the possibility that the Moho is associated with a broader transition in velocities [*Hale and Thompson*, 1982; *Owens and Zandt*, 1985] and not necessarily a sharp transition.

5.2. Kaapvaal Craton

[60] For the Kaapvaal Craton we analyze seismic data from the SASE experiment [*Silver et al.*, 2001] for the receiver functions and magnetotelluric data from the recent SAMTEX experiment [*Hamilton et al.*, 2006; *Jones et al.*, 2009]. The locations of the measurement sites are shown in Figure 10. We use the Rayleigh wave dispersion data by *Li and Burke* [2006]. They use a two-plane wave inversion technique [*Forsyth and Li*, 2005] to estimate phase velocities that are the input to our inversion.

[61] One problem when inverting various data sets together is that the requirements for each data set and the necessity to have closely located measurements strongly limit the number of candidate sites. Although the SAMTEX experiment covers the whole area of the previous SASE experiment, it was not designed for a joint inversion approach and only few MT sites are located closer than 40 km to a seismic site. At this distance

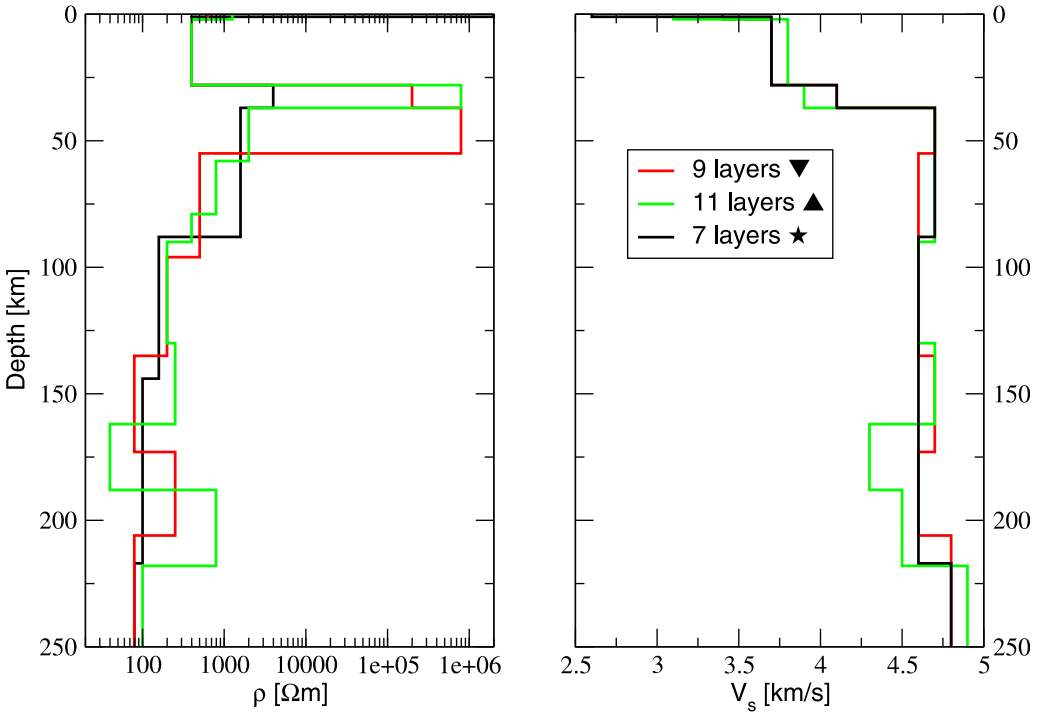

Figure 14. The electric and seismic parameters for the three models for site KAP25 marked in Figure 13.

features in the upper mantle should be sensed similarly at both sites. Considering the data quality and the requirements on electromagnetic dimension of the more than 550 SAMTEX sites and 80 SASE sites only about 10 remain that can be considered suitable. We present here the results from site KAP25 that is located on the Kaapvaal Craton near Kimberley. For this site the distance between the seismic and magnetotelluric measurements is 35 km. We show the apparent resistivities and phases for all impedance elements and the phase tensor dimensionality indicators in Figure 11.

[62] From Figure 11 it is clear that the MT data can be described by a regional 1-D model with local 3-D distortion [*Larsen*, 1975]. Within error the phases of all four impedance elements are equal while the apparent resistivities have the same shape, but are offset against each other. For most frequencies the phase tensor dimensionality indicators are not significantly different from zero although the errors are surprisingly large. The reason for this is that the phase tensor elements are poorly defined as det(Re**Z**) is small at most frequencies. The equal apparent resistivity values for off-diagonal elements suggests that the data are not affected by static shift.

[63] For our receiver function analysis we calculate receiver functions for 58 events with $M_w > 6$ recorded between 1997 and 1999. Of those events 13 show a clear initial correlation peak, Moho conversion and multiples (see Table A2). We stack these receiver functions in two bins, one with ray parameters between 0.048 s/km and 0.052 s/km and one with ray parameters between 0.074 s/km and 0.08 s/km. We plot the stacked receiver functions in Figure 12. In general the data quality is lower than for the Slave Craton receiver functions. We can identify a coherent conversion at 4 s lag time corresponding to the Moho, and there are a number of negative amplitude conversions that appear on both stacked receiver functions. Still there is also a large portion of incoherent amplitude variations that are likely due to noise.

[64] We invert the three data sets with the same settings as for site EKTN. Separate inversion of the MT data suggests that a four-layer model can capture the main features of the electrical structure. However we cannot achieve a satisfactory fit for the seismic data even with 5 layers. We therefore vary the number of layers for the joint inversion between 7 and 11 again. Figure 13 shows the resulting trade-offs generated by our joint inversion code for 7, 9 and 11 layers. The minimum misfit for the MT is approximately 1.2 and essentially the same for all number of layers. Also there is a large number of models with the same MT misfit in each inversion. This is not surprising considering that we over-parameterize the problem in terms of the electrical structure without applying additional regularization. However, we found that the regularization of the electrical parameters interacts with the surface wave data and the models that fit all data sets were essentially unregularized. We therefore only show the results of the unregularized inversions for which the trade-off plots are also easier to analyze.

[65] The two seismic data sets show considerable trade-off, indicating the presence of noise. The minimum misfit for the surface wave data decreases by a factor of two when increasing the number of layers from 7 to 9. When we increase the number of layers further to 11 the misfit for the surface waves improves slightly and the point of optimal trade-off between surface waves and receiver functions moves toward the left of the diagram, indicating an improved receiver function fit. Considering that we only

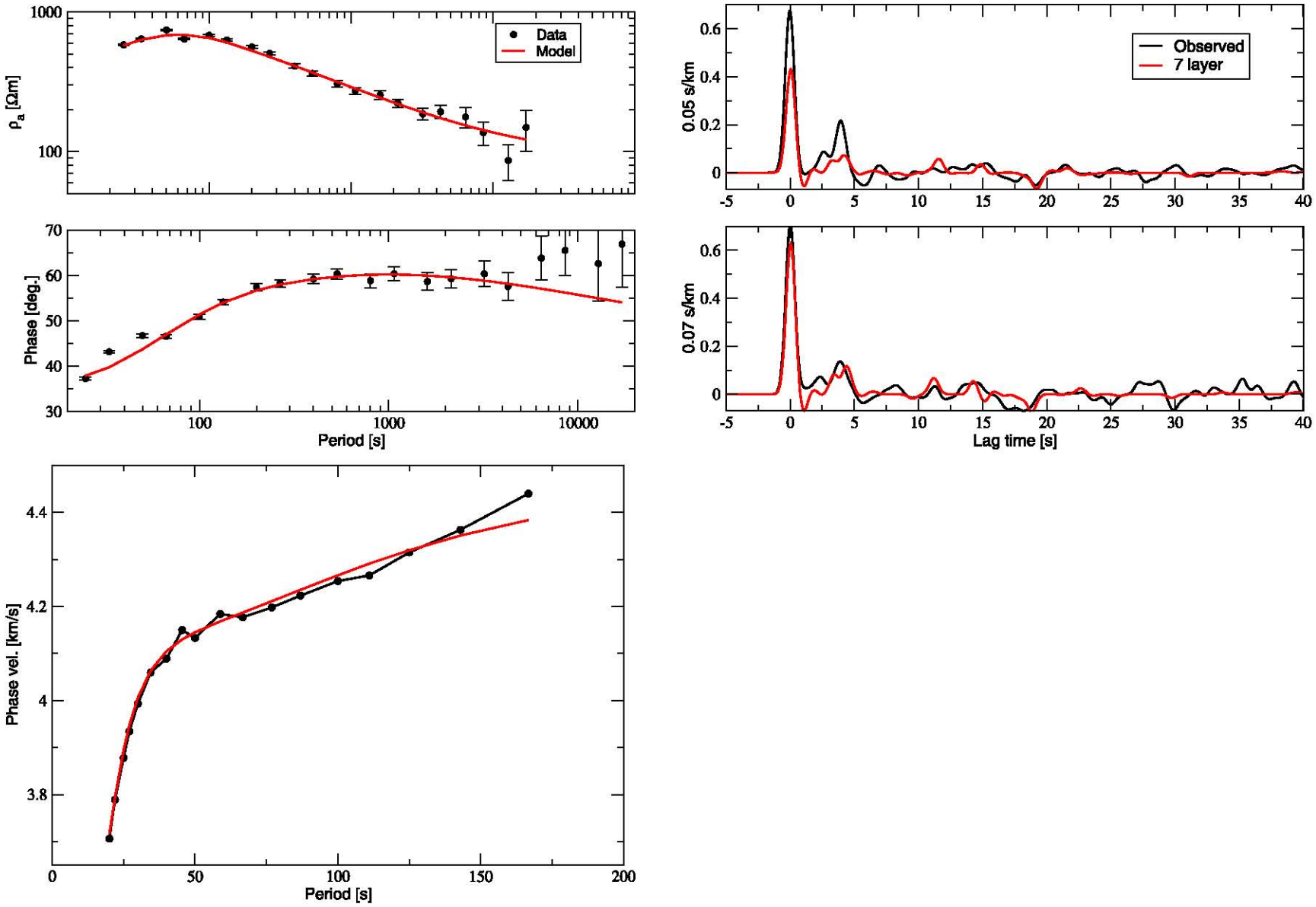

Figure 15. The measured and predicted data for site KAP25.

have 18 phase velocity estimates but 21 degrees of freedom in the 11 layer case, this can be expected.

[66] We mark the models with an optimal trade-off between the seismic data sets with a star for the inversion with 7 layers, with an inverted triangle for 9 layers and a triangle for 11 layers. All three models have approximately the same misfit for the MT data and for each inversion there are some models with a lower MT misfit. Comparison between the marked models and the best models with respect to the MT data shows that in all three cases the improvement in misfit is only marginal and is mainly due to one or two data points. We therefore regard this improvement as not significant and consider the data sets compatible within the resolution.

[67] Figure 14 shows a comparison of the three models marked in the trade-off plots. All three model show the Moho at a depth of 37 km, in agreement with previous studies [*Nguuri et al.*, 2001; *Nair et al.*, 2006; *Yang et al.*, 2008]. Apart from this there is considerable variability between the three models although some common characteristics remain. The resistivity model shows a sharp increase in resistivity that roughly coincides with the seismic Moho. The resistivity gradually decreases from more than 100,000 Ωm to 80 Ωm between 60 km and 100 km depth. In all three cases we observe a relatively conductive zone below 80 km that coincides with a low-velocity zone. With increasing number of layers the models contain additional variations below 120 km. As the misfit for the MT data and for the receiver function data is similar for all three models we regard the 7 layer model as the most conservative and reliable representation. Here we observe a single low-velocity zone between 80 km and 220 km, while the other models suggest that this zone might be split into two parts. Using only surface wave data, a broad region of low shear wave velocities at this depth have been reported [*Li and Burke*, 2006; *Yang et al.*, 2008; *Priestley et al.*, 2008]. Recent studies using receiver functions have also tended to place a discontinuity from faster to slower velocities at around 150 km depth [e.g., *Savage and Silver*, 2008; *Hansen et al.*, 2009; *Vinnik et al.*, 2009], similar to our models constrained by a greater number of layers. Our joint inversion results suggest that this zone coincides with a region of low resistivity. However, in contrast to the Slave Craton, the discontinuities marking any boundaries of the low-velocity zone are not clearly defined due to the higher noise levels. Also the site is located near the boundary of the craton, so it is not clear in how far the results are representative for the area. We therefore refrain from any detailed interpretations of the Kaapvaal results in terms of geological structures or processes.

[68] The problematic nature of the receiver function data in particular can be recognized from the comparison between the predicted data and the observed data in Figure 15. While the MT data and dispersion curves are matched well, we have problems to obtain a satisfactory fit for the receiver functions. While the predicted data matches the initial correlation peak and Moho conversion well for a ray parameter of 0.07 s/km, the predicted amplitude for a ray parameter of 0.05 s/km is too low. Also, there are other features that are matched for one stack but not for the other, for example the second Moho multiple. Still, the low-velocity zone appears in all inversion runs and with different inversion settings, so we regard it as a robust feature of our joint inversion.

6. Conclusions

[69] The results of the synthetic tests and the inversion of the two data sets demonstrate that our joint inversion approach not only finds a suitable model where appropriate, but also gives some clear indication of the compatibility of the data sets and realistic noise levels. We particularly regard the last two points as a clear advantage over linearized approaches. In principle, a similar analysis can be performed by varying the weights of the data sets in a linearized inversion scheme, However, not all points of the Pareto front are accessible for linearized schemes [*Kozlovskaya et al.*, 2007] and the range of weights and their sampling is unclear and different for every problem. Both of these difficulties are solved by our GA-based approach. In addition, the computational cost of multiple linearized inversion runs with varying weights is on the order of the computational cost of our GA based inversion, but requires additional user intervention.

[70] A critical component for any joint inversion algorithm is the parametrization and the interaction of the different data sets within the inversion. We chose the least restricting approach by only requiring coincident interfaces for the electrical and seismic models. As long as we keep the number of inversion layers sufficiently small, we achieve a good coupling of the data sets. The Slave Craton inversion example demonstrates the types of hypothesis we can test with joint inversion. The Central Slave Mantle Conductor can be modeled as a coincident low resistivity, low-velocity zone. This is a robust result from individual and joint inversions and indicates a spatial correlation.

[71] For the Kaapvaal Craton the data quality precludes a detailed interpretation of the models. We can fit both the MT data and the dispersion data well and reproduce some of the main features of the receiver functions with a joint model. The comparison of different models from the Pareto front, the structure of the trade-off and the expression of model features in the synthetic data provide a valuable aid to assess the validity of the joint models and give a qualitative indication of model errors.

[72] The increased information comes at the cost of high computational cost and simultaneous quality requirements on all data sets. While the first point becomes less and less of an issue with increasing computational abilities, the quality requirements limit the number of sites this approach can be applied to. With the tendency toward integrated studies and combined seismic and electromagnetic experiments, more data suitable for joint inversion will become available.

Appendix A: Event Parameters

[73] Here we show the parameters for the events we used in the receiver function analysis. Table A1 shows the event parameters for station EKTN, and Table A2 shows the event parameters for station SA17.

Appendix B: Software

[74] The joint inversion program, associated tools and documentation can be downloaded from http://gplib.sourceforge.

Table A1. Parameters of the Events Used for Receiver Function Inversion at Site EKTN

Event	Date	Time (UT)	Latitude	Longitude	Depth (km)	M_w
1	12 Oct 2002	2009:11.46	−8.2950	−71.7380	534.30	6.90
2	17 Nov 2002	0453:53.54	47.8240	146.2090	459.10	7.30
3	17 Mar 2003	1636:17.31	51.2720	177.9780	33.00	7.10
4	19 May 2003	1627:10.20	17.5460	−105.4730	10.00	6.10
5	21 May 2003	1844:20.10	36.9640	3.6340	12.00	6.80
6	26 May 2003	0924:33.40	38.8490	141.5680	68.00	7.00
7	23 Jun 2003	1212:34.47	51.4390	176.7830	20.00	6.90
8	27 Jul 2003	0625:31.95	47.1510	139.2480	470.30	6.80
9	25 Sep 2003	1950:06.36	41.8150	143.9100	27.00	8.30
10	17 Nov 2003	0643:06.80	51.1460	178.6500	33.00	7.80
11	5 Dec 2003	2126:09.48	55.5380	165.7800	10.00	6.70
12	14 Apr 2004	2307:39.94	71.0670	−7.7470	12.20	6.00
13	29 May 2004	2056:09.60	34.2510	141.4060	16.00	6.50
14	10 Jun 2004	1519:57.75	55.6820	160.0030	188.60	6.90
15	14 Jun 2004	2254:21.32	16.3370	−97.8450	10.00	5.90
16	6 Sep 2004	2329:35.09	33.2050	137.2270	10.00	6.60
17	9 Oct 2004	2126:53.69	11.4220	−86.6650	35.00	7.00
18	23 Oct 2004	0856:00.86	37.2260	138.7790	16.00	6.60
19	15 Nov 2004	0906:56.56	4.6950	−77.5080	15.00	7.20
20	28 Nov 2004	1832:14.13	43.0060	145.1190	39.00	7.00
21	13 Jun 2005	2244:33.90	−19.9870	−69.1970	115.60	7.80
22	26 Jul 2005	1217:14.27	52.8710	160.1050	27.60	5.80
23	16 Aug 2005	0246:28.40	38.2760	142.0390	36.00	7.20
24	21 Sep 2005	0225:08.11	43.8920	146.1450	103.00	6.10
25	26 Sep 2005	0155:37.67	−5.6780	−76.3980	115.00	7.50
26	8 Oct 2005	0350:40.80	34.5390	73.5880	26.00	7.60
27	12 Dec 2005	2147:46.07	36.3570	71.0930	224.60	6.50

Table A2. Parameters of the Events Used for Receiver Function Inversion at Site SA17

Event	Date	Time (UT)	Latitude	Longitude	Depth (km)	M_w
1	27 Dec 1997	2011:01.34	−55.7830	−4.2180	10.00	6.20
2	3 Jan 1998	0610:08.38	−35.4740	−16.1910	10.00	6.30
3	12 Jan 1998	1014:07.630	−30.985	−71.41	34	6.6
4	30 Jan 1998	1216:08.69	−23.9130	−70.2070	42.00	7.10
5	20 Feb 1998	1218:06.23	36.4790	71.0860	235.60	6.40
6	1 Apr 1998	1756:23.36	−0.5440	99.2610	55.70	7.00
7	1 Apr 1998	2242:56.90	−40.3160	−74.8740	9.00	6.70
8	25 Apr 1998	0607:23.44	−35.2660	−17.3260	10.00	6.30
9	18 Jun 1998	0417:54.98	−11.5720	−13.8940	10.00	6.30
10	29 Jul 1998	0714:24.08	−32.3120	−71.2860	51.10	6.40
11	3 Sep 1998	1737:58.24	−29.4500	−71.7150	27.00	6.60
12	5 Mar 1999	0335:14.71	−34.6730	−69.6000	10.00	6.00
13	28 Mar 1999	1905:11.03	30.5120	79.4030	15.00	6.60

net. We used sac2000 [*Goldstein et al.*, 2003] to process seismograms for receiver function analysis and GMT [*Wessel and Smith*, 1991] for Figures 4, 6, 7, 10, and 13.

[75] **Acknowledgments.** Part of this work was carried out under the HPC-EUROPA project (RII3-CT-2003-506079), with the support of the European Community–Research Infrastructure Action under the FP6 Structuring the European Research Area program. S.F. was supported by the Natural Environment Research Council (NERC grant NE/G000859/1). We would like to thank S. Lebedev for comments on an early version of the manuscript. C. Chen and A. Li kindly made their Rayleigh wave dispersion data available for this study. The MT data for the Slave Craton were part of the Lithoprobe project, and the data for the Kaapvaal Craton were part of the SAMTEX project. All those acknowledged by *Jones et al.* [2003, 2009] are acknowledged here.

References

Ammon, C. J., G. E. Randall, and G. Zandt (1990), On the nonuniqueness of receiver function inversions, *J. Geophys. Res.*, *95*, 15,303–15,318.

Bahr, K. (1988), Interpretation of the magnetotelluric impedance tensor: Regional induction and local telluric distortion, *J. Geophys.*, *62*, 119–127.

Bahr, K., and F. Simpson (2002), Electrical anisotropy below slow- and fast-moving plates: Paleoflow in the upper mantle?, *Science*, *295*, 1270–1272.

Bank, C. G., M. G. Bostock, R. M. Ellis, and J. F. Cassidy (2000), A reconnaissance teleseismic study of the upper mantle and transition zone beneath the archean Slave Craton in NW Canada, *Tectonophysics*, *319*(3), 151–166.

Berdichevskiy, M. N., and V. I. Dmitriev (1976), Basic principles of interpretation of magnetotelluric sounding curves, in *Geoelectric and Geothermal Studies*, edited by A. Adam, pp. 165–221, Akad. Kiado, Budapest.

Berteussen, K. A. (1977), Moho depth determinations based on spectral-ratio analysis of NORSAR long-period P waves, *Phys. Earth Planet. Inter.*, *15*, 13–27.

Bibby, H. M., T. G. Caldwell, and C. Brown (2005), Determinable and non-determinable parameters of galvanic distortion in magnetotellurics, *Geophys. J. Int.*, *163*, 915–930.

Caldwell, T. G., H. M. Bibby, and C. Brown (2004), The magnetotelluric phase tensor, *Geophys. J. Int.*, *158*, 457–469.

Carcione, J. M., B. Ursin, and J. I. Nordskag (2007), Cross-property relations between electrical conductivity and the seismic velocity of rocks, *Geophysics*, *72*(5), E193–E204.

Chen, C.-W., S. Rondenay, D. S. Weeraratne, and D. B. Snyder (2007), New constraints on the upper mantle structure of the Slave craton from Rayleigh wave inversion, *Geophys. Res. Lett.*, *34*, L10301, doi:10.1029/2007GL029535.

Constable, S. C., R. L. Parker, and C. G. Constable (1987), Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*(3), 289–300.

Corne, D., and J. Knowles (2007), Techniques for highly multiobjective optimisation: Some nondominated points are better than others, in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pp. 773–780, Assoc. for Comput. Mach., New York.

Davis, W. J., A. G. Jones, W. Bleeker, and H. Grutter (2003), Lithosphere development in the Slave Craton: A linked crustal and mantle perspective, *Lithos*, *71*(2–4), 575–589.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002), A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, *6*(2), 182–197.

Debayle, E. (1999), *SV*-wave azimuthal anisotropy in the Australian upper mantle: Preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, *137*, 747–754.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference earth model, *Phys. Earth Planet. Inter.*, *25*(4), 297–356.

Eaton, D. W., A. G. Jones, and I. J. Ferguson (2004), Lithospheric anisotropy structure inferred from collocated teleseismic and magnetotelluric observations: Great Slave Lake shear zone, northern Canada, *Geophys. Res. Lett.*, *31*, L19614, doi:10.1029/2004GL020939.

Farquharson, C. G., and D. W. Oldenburg (2004), A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems, *Geophys. J. Int.*, *156*(1), 411–425.

Forsyth, D. W., and A. Li (2005), Array analysis of two-dimensional variations in surface wave phase velocity and azimuthal anisotropy in the presence of multipathing interference, in *Seismic Earth: Array Analysis of Broadband Seismograms*, *Geophys. Monogr. Ser.*, vol. 157, edited by A. Levander and G. Nolet, pp. 81–97, AGU, Washington, D. C.

Gallardo, L. A., and M. A. Meju (2003), Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data, *Geophys. Res. Lett.*, *30*(13), 1658, doi:10.1029/2003GL017370.

Gatzemeier, A., and M. Moorkamp (2005), 3D modelling of electrical anisotropy from electromagnetic array data: Hypothesis testing for different upper mantle conduction mechanisms, *Phys. Earth Planet. Inter.*, *149*, 225–242.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Mass.

Goldstein, P., D. Dodge, M. Firpo, and L. Minner (2003), Sac2000: Signal processing and analysis tools for seismologists and engineers, in *The IASPEI International Handbook of Earthquake and Engineering Seismology*, edited by W. H. K. Lee et al., pp. 1613–1614, Academic, London.

Groom, R. W., and R. C. Bailey (1989), Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion, *J. Geophys. Res.*, *94*(B2), 1913–1925.

Hale, L. D., and G. A. Thompson (1982), The seismic reflection character of the continental Mohorovicic discontinuity, *J. Geophys. Res.*, *87*, 4625–4635.

Hamilton, M., A. G. Jones, R. L. Evans, S. Evans, C. J. S. Fourie, X. Garcia, A. Mountford, J. E. Spratt, and the SAMTEX Team (2006), Electrical anisotropy of South African lithosphere compared with seismic anisotropy from shear-wave splitting analysis, *Phys. Earth Planet. Inter.*, *158*, 226–239.

Hansen, P. C. (1992), Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Rev.*, *34*(4), 561–580.

Hansen, S. E., A. A. Nyblade, J. Julià, P. H. G. M. Dirks, and R. J. Durrheim (2009), Upper-mantle low-velocity zone structure beneath the Kaapvaal craton from S-wave receiver functions, *Geophys. J. Int.*, *178*, 1021–1027.

Jones, A. G. (1988), Static shift of magnetotelluric data and its removal in a sedimentary basin environment, *Geophysics*, *53*(7), 967–978.

Jones, A. G. (1992), Electrical properties of the lower continental crust, in *Continental Lower Crust*, edited by D. M. Fountain, pp. 81–131, Elsevier, Amsterdam.

Jones, A. G. (1999), Imaging the continental upper mantle using electromagnetic methods, *Lithos*, *48*(1–4), 57–80.

Jones, A. G., and I. J. Ferguson (1997), The electric Moho, *Nature*, *390*, 331–333.

Jones, A. G., D. Snyder, S. Hanmer, I. Asudeh, D. White, D. Eaton, and G. Clarke (2002), Magnetotelluric and teleseismic study across the Snowbird Tectonic Zone, Canadian Shield: A Neoarchean mantle suture?, *Geophys. Res. Lett.*, *29*(17), 1829, doi:10.1029/2002GL015359.

Jones, A. G., P. Lezaeta, I. J. Ferguson, A. D. Chave, R. L. Evans, X. Garcia, and J. Spratt (2003), The electrical structure of the slave craton, *Lithos*, *71*, 505–527.

Jones, A. G., R. L. Evans, and D. W. Eaton (2009), Velocity-conductivity relationships for mantle mineral assemblages in Archean cratonic lithosphere based on a review of laboratory data and Hashin-Shtrikman extremal bounds, *Lithos*, *109*, 131–143.

Julia, J., C. J. Ammon, R. B. Herrmann, and A. M. Correig (2000), Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.*, *143*(1), 99–112.

Kennett, B. L. N., and K. Yoshizawa (2002), A reappraisal of regional surface wave tomography, *Geophys. J. Int.*, *150*, 37–444.

Kind, R., and L. P. Vinnik (1988), The upper-mantle discontinuities underneath the GRF array from P-to-S converted waves, *J. Geophys.*, *62*, 138–147.

Kozlovskaya, E., L. Vecsey, J. Plomerová, and T. Raita (2007), Joint inversion of multiple data types with the use of multiobjective optimization: problem formulation and application to the seismic anisotropy investigations, *Geophys. J. Int.*, *171*, 761–779.

Lahti, I., T. Korja, P. Kaikkonen, and K. Vaittinen (2005), Decomposition analysis of the BEAR magnetotelluric data: Implications for the upper mantle conductivity in the Fennoscandian Shield, *Geophys. J. Int.*, *163*, 900–914.

Langston, C. A. (1979), Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. Geophys. Res.*, *84*, 4749–4762.

Langston, C. A., and J. K. Hammer (2001), The vertical component P-wave receiver function, *Bull. Seismol. Soc. Am.*, *91*(6), 1805–1819.

Larsen, J. C. (1975), Low frequency (0.1–6.0 cpd) electromagnetic study of deep mantle electrical conductivity beneath the Hawaiian islands, *Geophys. J. R. Astron. Soc.*, *43*, 17–46.

Lebedev, S., G. Nolet, T. Meier, and R. D. van der Hilst (2005), Automated multimode inversion of surface and S waveforms, *Geophys. J. Int.*, *162*, 951–964.

Li, A., and K. Burke (2006), Upper mantle structure of southern Africa from Rayleigh wave tomography, *J. Geophys. Res.*, *111*, B10303, doi:10.1029/2006JB004321.

Ligorria, J. P., and C. J. Ammon (1999), Iterative deconvolution and receiver-function estimation, *Bull. Seismol. Soc. Am.*, *89*(5), 1395–1400.
Linde, N., A. Binley, A. Tryggvason, L. B. Pedersen, and A. Revil (2006), Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data, *Water Resour. Res.*, *42*, W12404, doi:10.1029/2006WR005131.
Marquis, G., A. G. Jones, and R. D. Hyndman (1995), Coincident conductive and reflective lower crust across a thermal boundary in southern British Columbia, Canada, *Geophys. J. Int.*, *120*, 111–131.
Moorkamp, M., A. G. Jones, and D. W. Eaton (2007), Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: Are seismic velocities and electrical conductivities compatible?, *Geophys. Res. Lett.*, *34*, L16311, doi:10.1029/2007GL030519.
Nair, S. K., S. S. Gao, K. H. Liu, and P. G. Silver (2006), Southern African crustal evolution and composition: Constraints from receiver function studies, *J. Geophys. Res.*, *111*, B02304, doi:10.1029/2005JB003802.
Nguuri, T. K., J. Gore, D. E. James, S. J. Webb, C. Wright, T. G. Zengeni, O. Gwavava, and J. A. Snoke (2001), Crustal structure beneath southern Africa and its implications for the formation and evolution of the Kaapvaal and Zimbabwe cratons, *Geophys. Res. Lett.*, *28*, 2501–2504.
Owens, T. J., and G. Zandt (1985), The response of the continental crust-mantle boundary observed on broadband teleseismic receiver functions, *Geophys. Res. Lett.*, *12*, 705–708.
Park, S. K., and D. W. Livelybrooks (1989), Quantitative interpretation of rotationally invariant parameters in magnetotellurics, *Geophysics*, *54*(11), 1483–1490.
Perry, H. K. C., D. W. Eaton, and A. M. Forte (2002), LITH5.0: A revised crustal model for Canada based on Lithoprobe results, *Geophys. J. Int.*, *150*, 285–294.
Priestley, K., D. McKenzie, E. Debayle, and S. Pilidou (2008), The African upper mantle and its relationship to tectonics and surface geology, *Geophys. J. Int.*, *175*, 1108–1126.
Randall, G. E. (1989), Efficient calculation of differential seismograms for lithospheric receiver functions, *Geophys. J. Int.*, *99*, 469–481.
Ritzwoller, M. H., and A. L. Levshin (1998), Eurasian surface wave tomography: Group velocities, *J. Geophys. Res.*, *103*, 4839–4878.
Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, *40*(3), 1009, doi:10.1029/2000RG000089.
Savage, B., and P. G. Silver (2008), Evidence for a compositional boundary within the lithospheric mantle beneath the Kalahari craton from S receiver functions, *Earth Planet. Sci. Lett.*, *272*, 600–609.
Silver, P. G., S. S. Gao, and K. H. Liu (2001), Mantle deformation beneath southern Africa, *Geophys. Res. Lett.*, *28*, 2493–2496.
Siripunvaraporn, W., G. Egbert, Y. Lenbury, and M. Uyeshima (2005), Three-dimensional magnetotelluric inversion: Data-space method, *Phys. Earth Planet. Inter.*, *150*, 3–14.
Snyder, D. B., and M. Bruneton (2007), Seismic anisotropy of the Slave Craton, NW Canada, from joint interpretation of SKS and Raleigh waves, *Geophys. J. Int.*, *169*, 170–188.
Snyder, D. B., S. Rondenay, M. G. Bostock, and G. D. Lockhart (2004), Mapping the mantle lithosphere for diamond potential using teleseismic methods, *Lithos*, *77*, 859–872.
Tournerie, B., and M. Chouteau (2005), Three-dimensional magnetotelluric survey to image structure and stratigraphy of a sedimentary basin in Hungary, *Phys. Earth Planet. Inter.*, *150*, 197–202.
Vinnik, L., S. Oreshin, G. Kosarev, S. Kiselev, and L. Makeyeva (2009), Mantle anomalies beneath southern Africa: Evidence from seismic S and P receiver functions, *Geophys. J. Int.*, *179*, 279–298.
Wait, J. R. (1954), On the relation between telluric currents and the Earth's magnetic field, *Geophysics*, *19*, 281–289.
Wessel, P., and W. H. F. Smith (1991), Free software helps map and display data, *Eos Trans. AGU*, *72*, 441–441.
Yang, Y., A. Li, and M. H. Ritzwoller (2008), Crustal and uppermost mantle structure in southern Africa revealed from ambient noise and teleseismic tomography, *Geophys. J. Int.*, *174*, 235–248.
Yuan, X., R. Kind, X. Li, and R. Wang (2006), The S receiver functions: Synthetics and data example, *Geophys. J. Int.*, *165*(1), 555–564.

S. Fishwick, Department of Geology, University of Leicester, University Road, Leicester, LE1 7RH, UK.
A. G. Jones, School of Cosmic Physics, Dublin Institute for Advanced Studies, 5 Merrion Sq., Dublin, Ireland.
M. Moorkamp, Leibniz-Institut für Meereswissenschaften an der Universität Kiel (IFM-GEOMAR), Wischhofstr. 1-3, D-24149 Kiel, Germany. (mmoorkamp@ifm-geomar.de)