Journal of Applied Geophysics 146 (2017) 103–119

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

Improved characterisation and modelling of measurement errors in electrical resistivity tomography (ERT) surveys

Chak-Hau Michael Tso [a,*], Oliver Kuras [b], Paul B. Wilkinson [b], Sebastian Uhlemann [b], Jonathan E. Chambers [b], Philip I. Meldrum [b], James Graham [c], Emma F. Sherlock [d], Andrew Binley [a]

[a] Lancaster Environment Centre, Lancaster University, Lancaster, UK
[b] British Geological Survey, Keyworth, UK
[c] National Nuclear Laboratory, Seascale, UK
[d] Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 15 May 2017
Accepted 8 September 2017
Available online 14 September 2017

Keywords:
ERT
Resistivity
Measurement errors
Uncertainty
Linear mixed effects
Inversion

ABSTRACT

Measurement errors can play a pivotal role in geophysical inversion. Most inverse models require users to prescribe or assume a statistical model of data errors before inversion. Wrongly prescribed errors can lead to over- or under-fitting of data; however, the derivation of models of data errors is often neglected. With the heightening interest in uncertainty estimation within hydrogeophysics, better characterisation and treatment of measurement errors is needed to provide improved image appraisal. Here we focus on the role of measurement errors in electrical resistivity tomography (ERT). We have analysed two time-lapse ERT datasets: one contains 96 sets of direct and reciprocal data collected from a surface ERT line within a 24 h timeframe; the other is a two-year-long cross-borehole survey at a UK nuclear site with 246 sets of over 50,000 measurements. Our study includes the characterisation of the spatial and temporal behaviour of measurement errors using autocorrelation and correlation coefficient analysis. We find that, in addition to well-known proportionality effects, ERT measurements can also be sensitive to the combination of electrodes used, i.e. errors may not be uncorrelated as often assumed. Based on these findings, we develop a new error model that allows grouping based on electrode number in addition to fitting a linear model to transfer resistance. The new model explains the observed measurement errors better and shows superior inversion results and uncertainty estimates in synthetic examples. It is robust, because it groups errors together based on the electrodes used to make the measurements. The new model can be readily applied to the diagonal data weighting matrix widely used in common inversion methods, as well as to the data covariance matrix in a Bayesian inversion framework. We demonstrate its application using extensive ERT monitoring datasets from the two aforementioned sites.

1. Introduction

Measurement errors are an integral part of scientific observations. Properly describing such errors is essential to harness the information about the observed behaviour contained in the measurements. Measurement errors may be random or systematic. In commonly used geophysical inverse methods, measurement errors are assumed to be uncorrelated and random. In the context of an inversion, the total data error is given by the square root of the sum of squares of measurement errors and modelling errors. Sources of modelling errors include inaccuracy of the forward model (e.g. due to discretisation of a numerical model) and appropriateness of a forward model (e.g. representing a 3D problem using a 2D model). Modelling errors are relatively well understood because they can be studied by comparing forward modelling data of a homogeneous domain with analytical solutions (see Binley, 2015). We, therefore, focus here on measurement errors, in particular within the context of electrical resistivity tomography (ERT).

1.1. The role of ERT measurement errors

Measurement error estimates play a critical role in ERT inversion (see more in Section 2.3). They affect the amount of damping imposed on the data and also the point at which convergence is attained. Both of the above are achieved by weighting data in the objective function, and thus, measurement error estimates control whether there will be over-fitting or under-fitting of data during inversion. This concept can be illustrated by comparing various inverted images. Fig. 1 shows the results of inverting synthetic ERT experiments corrupted by 5% Gaussian noise. In the synthetic domain, a resistive target is placed between $x = 15$ m and $x = 20$ m and the topsoil is relatively conductive (Fig. 1a). Inverting the data with 10% assumed Gaussian noise leads to under-fitting and a very smooth resultant image (Fig. 1b), while assuming 2% noise leads to over-fitting and a number of artefacts (Fig. 1d). This

* Corresponding author.
E-mail address: m.tso@lancaster.ac.uk (C.-H.M. Tso).

https://doi.org/10.1016/j.jappgeo.2017.09.009
0926-9851/

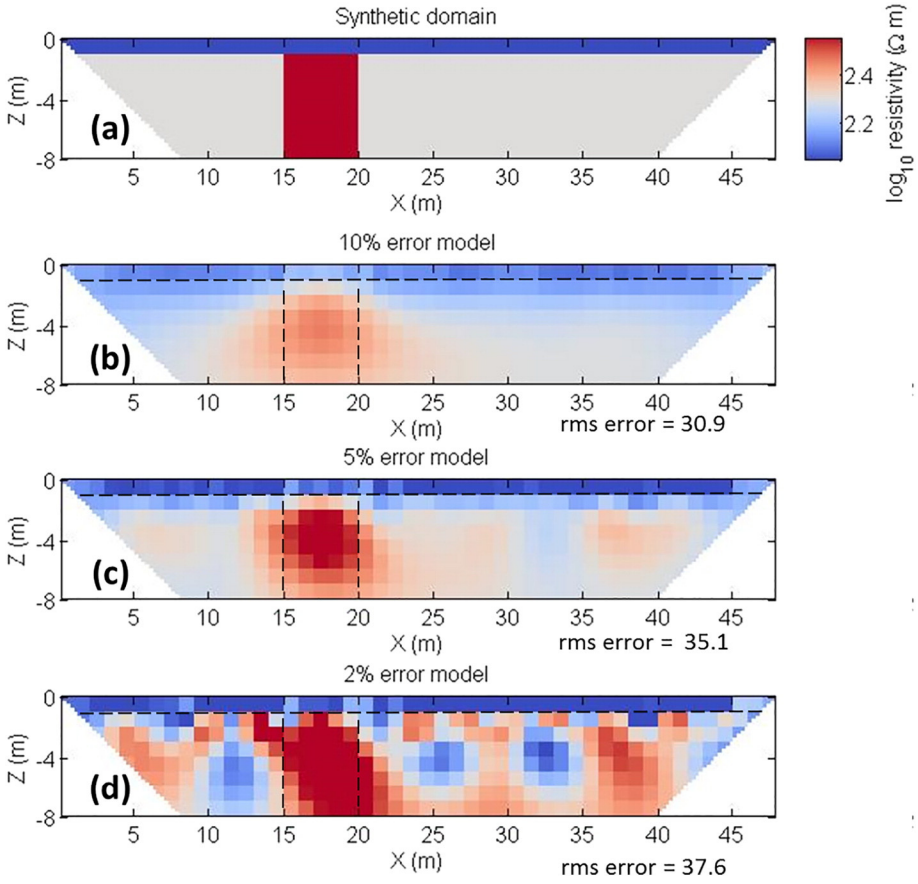

Fig. 1. Synthetic problem for demonstration: (a) synthetic domain with a more conductive layer near the surface and a resistive area between $x = 15$ m and $x = 20$ m. The synthetic data from running a forward model in (a) is perturbed with 5% Gaussian noise and then inverted by assuming (a) 10% linear error model (b) 5% linear error model (c) 2% linear error model. Note that rms error is defined as $\sqrt{\sum_{i=1}^{n}(obs-sim)^2/n}$, where *obs* and *sim* are vectors of observed/true and simulated transferred resistances of length *n* respectively. Note that the convergence target for all the inversions is a chi-squared statistic of 1.

simple example shows that inversion results can be sensitive to the assumed measurement error levels. Failure to prescribe them adequately can significantly change the resultant image.

Attempts have been made to account for data errors in a more sophisticated manner. Robust inversion (Kemna, 2000; Morelli and LaBrecque, 1996) adjusts error weights when there are apparent outliers. It is important, however, to notice that the outliers are linked to a specific error weight derived a priori by the error model—they may not be outliers anymore if a different error model is used. Similarly, in Bayesian inversion (e.g. Irving and Singha, 2010; Ramirez et al., 2005), one needs to prescribe the estimated data uncertainty in the likelihood function. While different inversion strategies handle measurement errors differently, a robust and accurate prescription of measurement errors is essential to obtain reliable and realistic inversion results.

The impact of measurement errors is not limited to inversion—it is a natural extension of stochastic inversion where posterior models are estimates of uncertainty, whereas deterministic inversion results (or an ensemble of them) can further be used to estimate uncertainty via Monte Carlo approaches. Therefore, uncertainties in measurements would propagate to uncertainties of model estimates. Similarly, if the inversion results are used to detect or monitor subsurface processes, or to infer hydrological properties, their associated errors can be traced back to measurement errors. It is apparent that measurement errors propagate through the various stages of a hydrogeophysics study workflow. With the heightening interest in uncertainty estimation within hydrogeophysics (Binley et al., 2015; Huisman et al., 2010; Linde et al., 2015; Rubin and Hubbard, 2005; Vereecken et al., 2006), better characterisation and treatment of measurement errors is necessary to provide better image appraisal.

1.2. Measurement errors in ERT: a review

The handling of measurement errors in ERT surveys, despite its importance as outlined above, is variable. The simplest (but not necessarily the most reliable) method of assessing a measurement error in an ERT measurement is through the use of stacking, i.e. the repeated measurement of transfer resistance through a number of cycles of current injection. Such stacking assessment offers valuable in-field data quality appraisal but, as shown later, may be of limited value in quantifying a data weight for ERT inversion. Alternatively, repeatability errors can be obtained by multiple, separate measurements of transfer resistance over time. Usually this involves a repetition of the entire ERT measurement sequence sometime after the first attempt. Reciprocity checks are another method of measurement error assessment. Reciprocity is the general physical principle where the switching of source/sink and observation locations would yield the same response (Parasnis, 1988). It is, for example, utilised in groundwater hydrology (Barker, 1991; Bruggeman, 1972; Delay et al., 2011; Falade, 1981). Reciprocity checks for ERT are conducted by swapping the current and potential electrodes. Reciprocity breaks down when the ground response is non-linear (i.e. non-ohmic for ERT) or time-dependent (i.e. something changes between forward and reciprocal measurements).

As LaBrecque et al. (1996) point out, both repeated and reciprocal measurements are measures of precision not accuracy. Sources of systematic error are not accounted for explicitly in measurements of precision – some procedures may miss them entirely while others lump them as random errors. Reciprocal errors treat the swapping of electrodes as a way to account for some systematic errors while repeatability errors do not consider them at all. Therefore, reciprocal errors may be more useful to

eliminate bias caused by using a particular pair of electrodes as transmitter and another as receiver. The most commonly used errors in ERT, however, are stacking errors and they are misreferred as repeatability errors (Day-Lewis et al., 2008). Modern ERT instruments are equipped with stacking capability and they automatically return stacking errors. In other words, stacking errors can be obtained without re-running the measurement procedure, which is very attractive in time-sensitive or time-consuming surveys.

We surveyed a number of published ERT studies and report their description of error analysis in Table 1. From Table 1 we see that reciprocity is a commonly used measure, while a small fraction of field and experimental studies do not report their treatment of errors at all. Studies often attribute their exclusion of reciprocal errors to logistical constraints and argue that reporting stacking errors is sufficient. After errors are obtained, an error model (usually a linear relationship linking error to transfer resistance) is established (Binley et al., 1995). Once obtained, such a relationship may be used to predict the errors of individual measurements and thus contribute to the data weight in the inverse modelling. Some authors, however, assign observed errors directly in the inverse modelling, although this is potentially flawed unless statistical robustness of the quantified error is established (recognising that for most surveys errors are only computed from two observations). This practice also makes it impossible to identify "disinformative data" (Beven and Westerberg, 2011). From the reported error models, it is observed that error levels are generally higher for cross-borehole surveys, largely due to more challenging electrode contact conditions (compared to most surface ERT array surveys). Prior to fitting the error model and carrying out inversion, measurements with high errors are often eliminated; sometimes more than 20% of the collected data are removed. For time-lapse studies, it is quite common that the entire time series of an individual resistance measurement is removed if any part of the time series is deemed to be an outlier. For recent work on time-lapse cross-borehole ERT, see Schmidt-Hattenberger et al. (2016) and Yang et al. (2014).

Error models are generally a function of average measured transfer resistance (i.e. the error in a transfer resistance increases with the magnitude of transfer resistance) because of the well-known proportionality effects (Aster et al., 2005) in DC resistivity measurement errors (Binley, 2015; Binley et al., 1995). In studies where errors are accounted for, there is generally a preference to use model-predicted errors rather than individually observed errors since error assessment based on two observations is potentially unreliable. Some studies mitigate this potential issue by binning (or grouping) data with similar transfer resistance together before fitting an error model (Koestel et al., 2008; Robinson et al., 2015; Wehrer and Slater, 2014). This practice should give more robust error estimates, although the error model may vary with the number of bins used.

To better characterise measurement errors, more understanding of the factors that contribute to them is needed. Current practice leaves many of the assumptions in ERT measurement errors modelling unchallenged. For example, do measurement errors show temporal or spatial correlations? Can we improve from using linear measurement error models? Are stacking errors and reciprocal errors comparable indicators of measurement errors? ERT surveys typically use each of the electrodes for multiple measurements. Ramirez et al. (2005) notes that this may increase the probability that measurement errors are correlated, however, there has been no published work addressing this issue.

1.3. Recent work on ERT measurement errors

Attempts have been made to handle potential systematic effects of measurement errors. Zhou and Dahlin (2003) studied the effect of spacing errors for 8 types of common 2D resistivity arrays. They confirm the common observation that ERT error outliers are often correlated with high contact resistances for some of the electrodes used in a measurement. Wilkinson et al. (2008) developed an approach to filter out configurations that are highly sensitive to geometric error in crosshole ERT surveys. Similarly, Wilkinson et al. (2016, 2010a) developed techniques to recover movements of permanently installed electrodes so that active landslides can be monitored using time-lapse ERT data only. As the popularity of time-lapse surveys increases, specific methods to handle and characterize measurement errors in large time-lapse datasets emerge. Deceuster et al. (2013) developed a method to automate the identification of changes in electrode contact during time-lapse ERT experiments. More recently, Mitchell and Oldenburg (2016) developed a 4-step data quality control methodology for very large ERT datasets.

Meanwhile, Kim et al. (2016) proposed a new measurement protocol in which self-potential (SP) data are obtained immediately prior to measuring DC. It involves swapping the polarities of the two current electrodes in each measurement to obtain a positive and a negative potential (i.e. thus a forward and backward resistance) for each measurement. This protocol claims to account for SP effects in DC measurements and eliminate distortions in the DC resistivity potential field caused by all unknown mechanisms including ambient noise.

1.4. Outline of this work

This paper addresses a number of practical issues related to the treatment of measurement errors in ERT inversion. We compare stacking, repeatability, and reciprocal errors in their utility to describe errors in measurements. We also study whether measurement errors are correlated in time and/or in space. We then hypothesize that measurement errors in ERT are not only linearly dependent on transfer resistances, but that the electrodes used in taking each measurement can be used as a grouping variable to improve error characterisation. We show that using the new error model leads to better inversion results and uncertainty estimates through synthetic and field experiments data. We first describe the datasets and analysis methods in Sections 2.1 and 2.2. Then we describe the ERT inversion and uncertainty estimation methods used in Sections 2.3 and 2.4. Section 3 reports results for the error analysis. We introduce a new error model based on linear mixed effects models and grouping of electrodes in Section 4, and Section 5 shows results of inversion and uncertainty quantification. We then discuss the implications of the results in Section 6, and provide conclusions and recommendations in Section 7.

2. Approach

With recent advances made in the development of automated ERT systems, ERT experiments can be conducted remotely, allowing the collection of a large volume of background ERT measurements for quality control purposes. These rich datasets can be exploited to investigate the behaviour of measurement errors through statistical analysis. They provide opportunities to explore errors in ERT datasets, including the assessment of temporal and spatial correlation of errors. We scrutinize two large field datasets through statistical analysis of different types of measurement errors.

First, we examine the probability density functions for each error type, namely stacking, reciprocal and repeatability errors. This allows us to understand the mean and variance of their distributions. Next, we use autocorrelation and correlation coefficient analysis to study the sequential and spatial correlation of errors between measurements. Insights about the potential correlation in measurement errors can help in developing improved error models. We study the validity of repeatability errors by computing the autocorrelation and correlation coefficient of the departure from the mean of repeated measurements instead of using the repeatability errors directly. If repeatability errors are purely random, using any subset of the set of repeats for each given measurement should give the same errors and thus the departure from the mean should exhibit little correlation. We compare inversion of ERT data using different error types and models on identical datasets

Table 1
Table of error models reported in the literature.
Note: outliers are measurements that are not used for inversion. If the error type is reciprocal, reciprocal pairs are counted as one measurement.

Source	Error model	Error type	Survey type	Outliers/measurements	Outlier rule	Other description of errors
(Cassiani et al., 2006)	Linear model	Reciprocal	12 surface lines, tracer	244 of 744 pairs (long) 44 of 744 (high-resolution)	reciprocal error > 5%	
(Chambers et al., 2014)	Not reported	Reciprocal	21 and 15 parallel survey lines	5% of 20,563 0.5% of 23,164	$\|e\| > 0.05\|R\|$	
(Crook et al., 2010)	Not reported	Reciprocal	A line and 2 boreholes 3 surface lines	~25% of 6022 ~10% of 750	$\|e\| > 0.05\|R\|$ $\|e\| > 0.04\|R\|$	
(Flores-Orozco et al., 2011)	Linear model	Reciprocal	2 surface lines	Not reported	Not reported	See also (Flores-Orozco et al., 2012)
(French et al., 2002)	Direct use	Reciprocal	2 boreholes, snowmelt, tracer	? of 1172 pairs	$\|e\| > 10\%\|R\|$ are removed	
(Gélis et al., 2010)	Not reported (probably direct use)	Repeatability and Reciprocal	3 2.5-km long profiles, tunnel	1777 pairs retained	Stacked $\|e\| > 5\%$	Reciprocity of Wenner array was always better than 10%. Reciprocity in dipole-dipole arrays was not reported.
(Haarder et al., 2015)	Direct use	Reciprocal	5 boreholes, tracer	Not reported	Not reported	
(Hayley et al., 2009)	Not reported (probably direct use)	Repeatability	15 surface lines		Keep only measurements repeatable better than 1%	"It was deemed infeasible to include a reciprocal survey because of the added time and cost."
(Hermans et al., 2012)	$\|e\| = 0.01 + 0.025\|R\|$	Reciprocal (200/823)	Surface (W—S)	0/823	N/A	Mean ($\|e_{repeat}\|$) < 0.1%
(Henderson et al., 2010)	Direct use: If $\|e\| > 10\%\|R\|$, $\|e\| = 200\|e\|$, else $\|e\| = 2.5\|e\|$,	Reciprocal	Marine surface	None	None	Stacked data error during acquisition did not exist 2%
(Johnson et al., 2012)	$\|e\| = 0.001 + 0.05\|R\|$	Reciprocal	4 parallel surface lines next to river	715 of 8660 (first 20 d) 5021 of 8660 (all)	(i) Injected current <0.01 A (ii) Having reciprocal error > 5% At any timeframe during monitoring	
(Johnson et al., 2015)	$\|e\| = 0.01 + 0.05\|R\|$	Look up error model values from the literature	11 surface lines	200 (~0.5%) of 40,454 nonreciprocal measurements	(i) Apparent resistivities >100 Ωm (ii) Injected current <0.01 A If any of the above is found a measurement in the first five surveys, it is removed in the entire dataset	
(Kuras et al., 2016)	Not reported	Reciprocal and stacked	TL 3D, 4 boreholes	0.79% of ~13,400 each day 9.2% of ~53,100	$\|e\| > 0.01\|R\|$ (log10 scale)	
(Koestel et al., 2008)	$\|e\| = 0.026 + 0.003\|R\|$	Reciprocal, binned	TL 3D soil column	?/23,130	Obvious	
(Lesparre et al., 2016)	Direct use	Reciprocal	Surface and borehole, tunnel	8022 or 90% retained	(i) $\|e\| > 0.05\|R\|$	
(Meyerhoff et al., 2014)	Not reported	Repeatability	2 surface lines	Not reported	(ii) Stacked $\|e\| > 0.1\%$ (iii) Active time-constrained inversion: "The first-order operator applied to the sequence of snapshots filters out noisy data that are not correlated in time and is flexible enough to allow relatively abrupt changes to occur on the sequence of inverted tomograms if supported by the data"	
(Nimmer et al., 2007)	Individual difference	Reciprocal	Surface 2D*field	735/2940 557/2830	$\|e\| > 0.05\|R\|$	
(Perri et al., 2012)	Linear	Reciprocal	Surface and borehole tracer test	?/2262 (surface) ?/5179 (borehole)	Not reported	
(Ramirez et al., 1996)	Direct use	Reciprocal	2D*field 3D*field	Not reported	Not reported	Inversions are condcuted with both direct and reciprocal measurements are conducted. It was found that change due to using the either of the two is much smaller than changes due to the leak.

(Revil et al., 2010)	Direct use	Stacked	9 surface lines, volcano	?/9525	Not reported	Each of the resistivity measurements represents the mean of 8 to 16 distinct measurements stacked together with the same set of electrodes.
(Robert et al., 2011)	N(6.3e−4–1.37e−3, 3.2e−4–1.1e−3) $\|e\| = a + b\|R\|$ a = 0–0.0026 b = 0.0031–0.016	Reciprocal binned	6 surface surveys	?/490–760 per site	$\|e_{repeat}\| > 0.01$ Outside $\pm 2\sigma$	"…ERT reciprocal errors are approximately 20 times larger than repetition errors"
(Robinson et al., 2015)	Binned log-linear $\|e\| = 0.131\|R\|^{0.48}$	Reciprocal	8 boreholes, fracture	153 of 4810	$\|e\| > 5\%\ \|R\|$ are removed	
(Singha and Gorelick, 2006)	$\|e\| = 0.01 + 0.05\|R\|$	Reciprocal	TL 3D field	20–50/3150 per day	$\|e\| > 0.05\|R\|$	
(Slater et al., 2000)	$\|e\| = 0.1 + 0.01\|R\|$	Reciprocal	2D* tank, tracer	?/660	$\|e\| > 0.1\|R\|$	
(Slater et al., 2010)	Not reported	Apparent outliers	1-Line continuous waterborne survey	~3.2% out of ~65,000	(i) Very small resistivity close to 0 Ωm (ii) Negative apparent chargeability (iii) Further processing	
(Slater and Binley, 2003)	Not reported	Reciprocal	Surface, IP 2D*field, IP	~25% of 772 0/770*6	$\|e\| > 0.08\|R\|$ $\|e\| > 0.03\|R\|$	
(Slater and Binley, 2006)	N(5e−4, 8e−3) N(8e−4, 6e−3)	Reciprocal	2D* field, IP	1521/3744	$\|e\| > 0.05\|R\|$ & $e_{numerical} > 2\%$	
(Uhlemann et al., 2015)	N/A	Reciprocal	TL surface (// ×5)	8–9% of 4285	$\|e\| > 0.05\|R\|$	
(Uhlemann et al., 2016)	$\|e\| = a + b\|R\|$ a = 1e−5–5e−4 b = 0.001–0.01 a > 0.001 Ω(system malfunction)	Reciprocal Binned or Individual	TL surface (×2)	<5%–10% of (1528 + 516) for $\|e\| > 0.05\|R\|$	$\|e\| > 0.10\|R\|$ (Replaced by values derived from their respective time-series, using an inverse distance weighting interpolation)	
(Wagner et al., 2013)	$\|e\| = 0.01 + 0.039\|R\|$	CV from stacked V	TL 3D field	~4% of 258,947	CV > 0.05 $\|R\| > 1700\ \Omega$	
(Wallin et al., 2013)	$\|e\| = 0.001 + 0.05\|R\|$	Reciprocal	3 surface lines	851 out of 12,359 pairs	(i) injected current <0.01A (ii) having reciprocal error > 5% at any timeframe during monitoring	
(Wehrer and Slater, 2014)	Binned quadratic-log10 $\|e\| = -8e-6\|R\|^2 + 0.062\|R\| + 0.0182$	Reciprocal	36 electrodes on each of the four vertical sides of the lysimeter core	? of 4200 pairs	Not reported	
(Williams et al., 2016)	Direct Use	Stacking	Surface	Not reported	Not reported	Up to three stacks were collectedduring surveys, with a mean stacking error of 0.2%
(Wilkinson et al., 2010a)	Not reported	Reciprocal	4 surface lines	516 pairs	Not reported	Over 90% of data has reciprocal errors <1%
(Wilkinson et al., 2010b)	Direct use	Reciprocal	14 boreholes	Not reported	Not reported	98.7% of the data had errors of <0.3% and maximum error recorded was only 2.7%
(Wilkinson et al., 2016)	Direct use	Reciprocal	5 surface lines	12% of 2580 pairs	(i) Having negative apparent resistivity (ii) Having reciprocal error > 5% (iii) Having contact resistance >2000 Ω (iv) Having positive/negative pulse amplitude ratios <0.75 or >1.33 (a measure of wave-form symmetry)	
(Yeh et al., 2006)	N/A	Compare with homogeneous resistivity field (i.e. only diagnose for possible systematic errors regarding the estimated mean resistivity field)	34 boreholes, leachate injection	? of 2700 for each timeframe	Measurements outside the theorectical voltage variance are removed	N/A

to illustrate how they manifest in inversion results. Lastly, we obtain uncertainty estimates of inversion results using a Monte Carlo simulation procedure. This allows us to visualize how measurement errors propagate into uncertainty in model estimates (in this study we assume there are no other error sources).

2.1. Dataset description

A synthetic dataset, along with two field datasets collected by the British Geological Survey (BGS) are used for this work.

2.1.1. Synthetic dataset

A synthetic dataset was created for use as an illustrative example using the synthetic domain and array of Fig. 1. The array consists of 25 2-m spaced surface electrodes. As seen in Fig. 1(a), the resistivity structure of the domain consists of a 1 m thick, 100 Ωm top layer. Beneath it is a 200 Ωm formation, in which a 500 Ωm unit protrudes vertically.

We created a forward model of dipole-dipole transfer resistances on the synthetic domain using R2 (http://www.es.lancs.ac.uk/people/amb/Freeware/R2/R2.htm) to obtain measurement error-free data. Two sets of synthetic data are generated by adding noise to these data: one with 2% Gaussian noise everywhere, and the other with 10% Gaussian noise on measurements involving three of the electrodes on the left (x = 6 m, 14 m, 22 m) and 2% noise everywhere else. The second noisy dataset was created to simulate the effect of a non-uniform error model that may be typical of surveys in areas with variable electrode contact or quality.

2.1.2. Boxford dataset

The first field dataset is from the Boxford Water Meadows Site of Special Scientific Interest in Berkshire, United Kingdom (Chambers et al., 2014; Musgrave and Binley, 2011; Uhlemann et al., 2016). The collection of the data was automated using the BGS's PRIME system. The ERT array is next to the Northern Array used in Uhlemann et al. (2016), having 32 electrodes spaced at 0.6 m intervals. A dipole-dipole type measurement configuration was chosen with dipole lengths (a) of 0.6 m to 2.4 m, and dipole separation multipliers (n) of 1 to 8. The measurement sequence includes 516 pairs of reciprocal measurements. Less than 15 min was needed to complete the measurement sequence and each of the measurements is obtained by stacking multiple readings from the same cycle of current injection to improve signal-to-noise ratio. The measurement sequence was repeated 96 times within a 24 hour period starting at 5:43 a.m. on 19th November 2015, yielding 96 independent repeats of full reciprocal data. Each of the repeats has 516 measurements (or pairs of reciprocals). During the 24-hour period, the air temperature in the area varied between 7 and 10 °C and there was no recorded precipitation.

2.1.3. Sellafield dataset

A full-scale 3-D time lapse cross-borehole ERT trial to monitor simulated subsurface leakage was undertaken at a UK nuclear licensed site in Sellafield, Cumbria, United Kingdom (Kuras et al., 2016). The data collection setup includes four 40 m deep boreholes and 160 electrodes. The data collection cycle of each ERT frame is less than a day, and each day's data includes 51,302 dipole-dipole measurements, including 12,481 pairs of reciprocals. The monitoring spanned a 2-year period with 246 days of data collection during that time. The first nine months of monitoring includes three stages of conductive leak simulant injection, while the remainder was designed for long-term background monitoring. The collection of data was automated using BGS's ALERT system. In order to be consistent with the autocorrelation analysis of the Boxford dataset, we divide the data into two subsets of 96 days (one encompasses all three injection periods while the other is during long-term background monitoring) for autocorrelation analysis.

2.2. Analysis methods

2.2.1. Definition of measurement error types

Stacking errors are given by the averaging of 'stacks' obtained by the ERT data collection equipment. Usually they can be output alongside the measured transferred resistance from the data collection console.

For reciprocal errors, if R_f is the forward (normal) transfer resistance for a particular quadrupole and R_r is the reciprocal of that measurement where its current and potential dipoles are swapped with the forward measurement, then the mean absolute transfer resistance ($|R|$) and absolute errors ($|e|$) are simply:

$$|\mathrm{R}| = \frac{|| R_f | + | R_r ||}{2} \text{ and } |\mathrm{e}| = \frac{|| R_f | - | R_r ||}{2} \tag{1}$$

As a proxy for repeatability errors, the departure from the mean of the j-th repeated reading for measurement number i ($d_{i,j}$) is given by:

$$e_{i,j} = d_{i,j} - \overline{d_i} \tag{2}$$

where $\overline{d_i}$ is the mean value for the i-th measurement.

2.2.2. Statistical analysis of measurement errors

The probability density function of an error type for a dataset is obtained by fitting a Gaussian distribution to the population of errors. Autocorrelation is defined as the correlation among a sequence of values at a given lag L:

$$\mathrm{autocorr}(L) = \frac{E\left[(X_t - \overline{X})(X_{t+L} - \overline{X})\right]}{var(X_t)} = \frac{\sum_{i=1}^{q-L}(X_t - \overline{X})(X_{t+L} - \overline{X})}{var(X_t)} \tag{3}$$

where $E[]$ is the expected value, X_t is a time-series, X_{t+L} is a time-series shifted by lag L, and $\overline{X}$ and $var(X_t)$ are the mean and variance of the time series respectively. q is the number of repeats for a measurement.

Correlation analysis can be used to study the potential correlation between measurement errors. The correlation coefficient, r, for the correlation between arbitrary variables x and y is defined by the products of standard scores (also known as z-scores or standardized variables) as follows:

$$r = r_{xy} = \frac{1}{q-1}\sum_{i=1}^{q}\left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right) \tag{4}$$

For the purposes of our analysis of measurement error correlations, x and y are series of two measurements that we consider and q is the number of repeats. $\overline{x}$ and $\overline{y}$ are the means of x and y respectively, while s_x and s_y are the standard deviations.

2.3. Inversion methods

To obtain 2D tomograms from electrical measurements from the synthetic study and Boxford site, we use the finite-element based, Occam-type, two-dimensional electrical resistivity inversion program R2 (www.es.lancs.ac.uk/people/amb/Freeware/R2/R2.htm). The three-dimensional inversion (Sellafield dataset) was performed by using the commercial code Res3DInvx64 (Loke and Barker, 1996). The inverse problem is posed as a minimization problem, where the objective function is defined as

$$\Phi = \Phi_d + \Phi_m = (\mathbf{d} - \mathbf{F}(\mathbf{m}))^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}(\mathbf{d} - \mathbf{F}(\mathbf{m})) + \alpha\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}} \tag{5}$$

where $\mathbf{d}$ are the data vector (e.g. measured apparent resistivities), $\mathbf{F}(\mathbf{m})$ is the set of simulated data using the forward model and estimated parameters $\mathbf{m}$. $\mathbf{W}_{\mathrm{d}}$ is a data weight matrix, which, if we consider the uncorrelated measurement error case and ignore forward model errors, is a diagonal matrix with entries equal to the reciprocal of the standard

deviation of each measurement. Forward modelling errors are also added to the diagonal of $\mathbf{W}_d$. Usually a forward model is run for the computational grid using a known homogeneous domain. Any discrepancy between the computed and known apparent resistivity values (i.e. data errors) will be added to the reciprocal of $\mathbf{W}_d$ by means of square root of sum of squares. In this study, we assume measurements errors are the only source of data errors while other sources, such as forward modelling errors and field procedural errors, are negligible. To regularize the minimization problem, a model penalty term $\Phi_m = \alpha \mathbf{W}_m^T \mathbf{W}_m$ is added to impose the spatial connectedness of the parameter cell values. α is a scalar that controls the emphasis of smoothing.

We can state a desired level of data misfit as $\Phi_d = N$, where N is the number of measurements (Binley, 2015). In an Occam's solution, we seek to achieve this desired data misfit subject to the largest possible value of α. The process is achieved by utilizing the Gauss–Newton approach, which results in the iterative solution of

$$\begin{aligned}\left(\mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J} + \alpha\mathbf{W}_m^T\mathbf{W}_m\right)\Delta\mathbf{m} &= \mathbf{J}^T\mathbf{W}_d^T(\mathbf{d}-\mathbf{F}(\mathbf{m})) - \alpha\mathbf{W}_m^T\mathbf{W}_m \\ \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + \Delta\mathbf{m}\end{aligned} \tag{6}$$

where $\boldsymbol{J}$ is the Jacobian (or sensitivity) matrix, given by $J_{i,j} = \partial d_i / \partial m_j$; $\boldsymbol{m}_k$ is the parameter set at iteration k; and $\Delta\mathbf{m}$ is the parameter update at iteration k. For the DC resistivity case, the inverse problem is typically parameterized using log-transformed resistivities.

The resolution matrix for the inversion is given by:

$$\left(\mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J} + \alpha\mathbf{W}_m^T\mathbf{W}_m\right)\Delta\mathbf{m} = \mathbf{J}^T\mathbf{W}_d^T(\mathbf{d}-\mathbf{F}(\mathbf{m})) - \alpha\mathbf{W}_m^T\mathbf{W}_m \tag{7}$$

2.4. Error propagation and uncertainty quantification methods

We follow the Monte Carlo uncertainty propagation procedure of Aster et al. (2005) outlined below. The goal is to simulate a collection of noisy data vectors and then examine the statistics of the corresponding models. The advantage of this method is that it can readily applied to field data where no repeats are available. The procedure is achieved by the following steps:

1 Propagate the inverse solution $\overline{\mathbf{m}}$ into an assumed noise-free baseline $j \times 1$ data vector d (where j is the size of number of measurements) using the forward model $\mathbf{G}$:

$$\mathbf{G}\overline{\mathbf{m}} = \mathbf{d} \tag{8}$$

2 Generate q realizations ($i = 1, \ldots, q$) of noisy data about $\overline{\mathbf{m}}$ using the error model

$$\mathbf{d}_i = \mathbf{d}_b + \boldsymbol{\varepsilon} .* \mathbf{Z} \tag{9}$$

where $\boldsymbol{\varepsilon}$ is the $j \times 1$ vector of error levels predicted by the error model and $\mathbf{Z}$ is the standard normal distribution variable and .* is element-wise multiplication.

3 Invert the q realizations ($i = 1, \ldots, q$) of noisy data using the inverse model

$$\mathbf{G}\boldsymbol{m}_i = \mathbf{d}_b + \boldsymbol{\varepsilon}_i \tag{10}$$

4 Let $\mathbf{A}$ be a $q \times m$ matrix where the i-th row contains the departure of the i-th model from the baseline inverse solution $\overline{\mathbf{m}}$

$$\mathbf{A}_i = \mathbf{m}_i^T - \overline{\mathbf{m}}^T \tag{11}$$

5 An empirical estimate of the model covariance matrix is given by

$$\mathrm{cov}(\overline{\mathbf{m}}) = \frac{\mathbf{A}^T\mathbf{A}}{q} \tag{12}$$

6 95% confidence interval about $\overline{\mathbf{m}}$ is given by

$$\overline{\mathbf{m}} \pm 1.96 \cdot \mathrm{diag}(\mathrm{cov}(\overline{\mathbf{m}}))^{1/2} \tag{13}$$

7 Similarly, the coefficient of variation of the estimate is given by

$$\mathrm{diag}(\mathrm{cov}(\overline{\mathbf{m}}))^{1/2} ./ \overline{\mathbf{m}}^T \tag{14}$$

where ./ is element-wise division.

3. Analysis of errors in field datasets

In this section, we report results from the statistical analysis of different types of errors with the methods outlined in Sections 2.1 and 2.2. Probability density functions (PDFs) show the ranges of these errors, while autocorrelation and correlation coefficient analysis reveals the potential autocorrelation of errors for successive repeated measurements and correlation of errors between pairs of measurements, respectively.

3.1. Probability density function of reciprocal and repeatability errors

Before detailed statistical analysis of measurement errors is performed, we first examine the probability density function of errors obtained from the Boxford dataset. Since the measurements are repeated 96 times, we can define repeatability errors based on averaging different numbers of repeats. Fig. 2 shows the repeatability errors based on measurements obtained with a 30 min, 1 hour, 2 hour, and 24 hour window. They correspond to averaging 2, 4, 8, and 96 repeats. The mean of the PDF increases with greater time windows while the variance first decreases, then increases for the 24 hour repeatability error. When large windows of averaging are used, changes in the subsurface condition such as diurnal changes in temperature can be mistaken as errors. This is supported by the observed increase in the mean. For the 24 hour sampled PDF, the lower tail overlaps that of the 1 hour and 2 hour PDFs while having a much greater spread. Clearly some measurements do not vary much during the 24 hour monitoring period while others do: measurements sensitive to the shallower subsurface will be more susceptible to external influences (e.g. temperature, evaporation, etc.).

Fig. 2 also shows the PDF for stacking errors for each of the measurements as well as the reciprocal errors from individual datasets. The reciprocal errors PDF essentially overlay that of the 30 minute repeatability errors. Their similarities may be explained by the fact that both of them are obtained from averaging pairs of measurements. It is noteworthy, however, that both the mean and variance of the PDF for reciprocal errors (which is collected in a 15 minute timeframe) is slightly higher—which is opposite to our general observation that repeatability errors increases with the size of the averaging window. Reciprocal errors may be sensitive to other error contributions not registered by repeatability errors, or the process of taking a reciprocal measurement introduces an additional source of error.

The stacking errors PDF overlays the low-end of the PDFs of repeatability errors while having a very small variance. In other words, stacking errors do not register any of the high-error measurements that appear in the true assessment of repeatability or in reciprocal errors. For instance, the PDF shows that almost none of the stacking errors are higher than $10^{-4}\ \Omega$, which covers a majority of the area under the other PDFs. This shows that stacking errors are potentially an inadequate measure for describing the true quality of ERT measurements.

The second portion of Fig. 2 shows the PDF of stacking, reciprocal, and 2-week (which correspond to six frames) repeatability errors for the Sellafield dataset. In general, the ranges of magnitude of the errors are greater due to ground conditions and contact resistances. Similar to the Boxford results, we find that the stacking errors are an order of

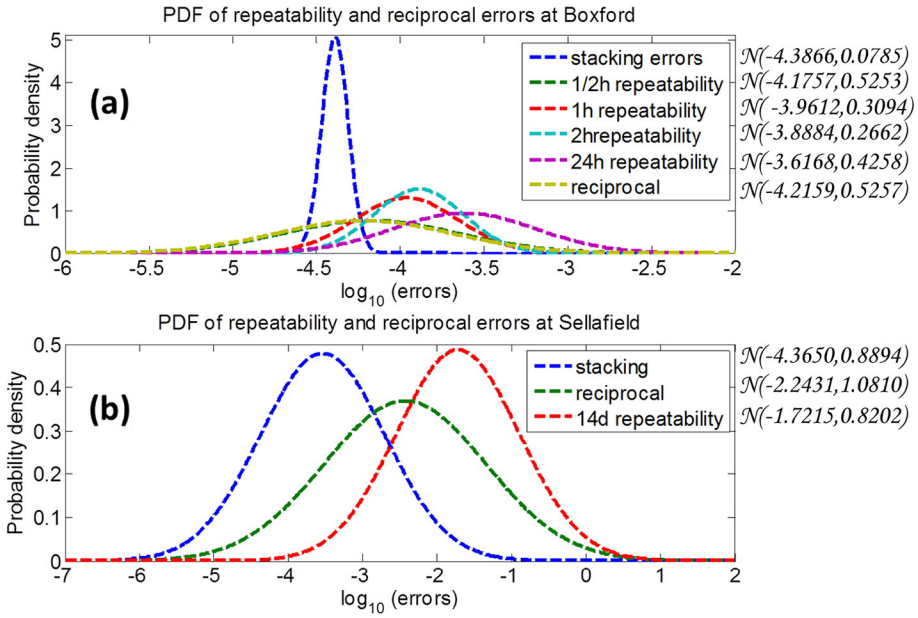

Fig. 2. (a) Comparison of stacking errors, repeatability errors, and reciprocal errors for the Boxford dataset by plotting probability density functions. The PDFs of reciprocal errors and repeatability errors are comparable to each other. The stacking errors PDF, however, show very low mean and low variance. Using stacking errors for measurement errors characterisation may lead to significant underestimation of uncertainty and over-fitting of data. (b) Comparison of stacking errors, repeatability errors, and reciprocal errors for the Sellafield dataset. The PDFs for Sellafield show greater variances than those for Boxford. Since a two-week repeatability cycle is used, the repeatability errors are much greater than reciprocal errors. In general, the stacking errors are more than an order-of-magnitude smaller than the reciprocal errors, indicating there may be significant underestimation of errors if they are used as error weights. The mean and standard deviation of each fitted normal distribution is shown next to the legend.

magnitude smaller than reciprocal errors. Since a larger time window (i.e. days) is used to obtain the repeatability errors, they are significantly greater than the reciprocal errors.

3.2. Autocorrelation analysis

Autocorrelation analysis is used to investigate whether there is "memory" (i.e. correlation in time) in ERT measurement errors. We compare autocorrelation plots between the (i) departure from the mean and (ii) reciprocal errors of individual measurements for the Boxford dataset in Fig. 3. Each grey translucent line represents the autocorrelation function of a measurement, while the red line is the mean averaged across all measurements. The red hashed regions highlights the area with a autocorrelation value below the critical Pearson's correlation coefficient (Pearson and Hartley, 1970), which is around 0.2 for 96 timesteps. For the departure from the mean, the autocorrelation drops to 0.5 on average at lag 1 and then decreases steadily. This is likely to be due to the presence of diurnal temperature effects within the 24 hour data collection cycle. Individual reciprocal errors, show negligible autocorrelation for all lag numbers (i.e. within the hashed region). Thus, we can conclude the individual reciprocal errors between any two repeated measurements are independent from one another for this survey. From the above, we see that the assumption of uncorrelated errors is appropriate for reciprocal errors but not so much for long-term repeatability errors.

Fig. 4 shows the autocorrelation of (a) departure from the mean and (b) reciprocal errors for the 96 datasets collected continuously at the Sellafield site encompassing the three injection periods (22th Jan

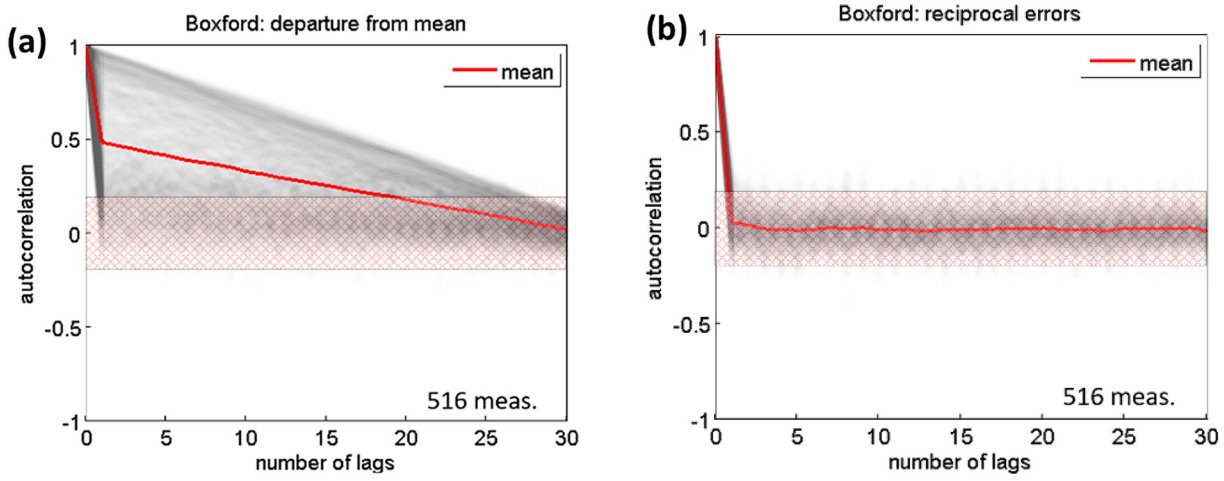

Fig. 3. Autocorrelation of (a) departure from the mean (as a measure of repeatability errors) and (b) reciprocal errors for the 96 datasets collected continuously within 24 h at the Boxford site. The number of lags is on the horizontal axis (here 1 lag = 15 min). Each grey translucent line plots the autocorrelation of one of the 516 ERT measurements as a function of lag. The red line denotes the mean autocorrelation. For each autocorrelation plot, 96 datasets are considered. The hashed region has insignificant correlation according to the critical Pearson's test (around ±0.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2013–3rd Nov 2013) and those for another 96 datasets during the long-term background monitoring period (i.e. no injection, 5th Nov 2013–31st Mar 2014). We can see much greater autocorrelation of errors at Sellafield than at Boxford. Like in the Boxford dataset, the departure from the mean shows greater autocorrelation than individual reciprocal errors, both for injection and long-term background monitoring. In general, however, the departure from the mean and reciprocal errors during background monitoring reach insignificant autocorrelation sooner than during injections. While the 96 datasets at Boxford were collected in less than 24 h, the two groups of 96 datasets examined above were collected over a period of months. It is certain that the subsurface condition had changed during the monitoring period due to injection, dilution and dispersal of tracer, as well as regional groundwater and vadose zone changes (see Kuras et al., 2016 for details).

3.3. Correlation coefficient analysis

Although measurement errors are commonly assumed to be uncorrelated in ERT, previous studies have highlighted the potential of correlation in measurement errors because ERT surveys typically use the same electrodes for multiple measurements (Ramirez et al., 2005). We have computed the correlation coefficient matrix for departure from the mean and reciprocal errors for the Boxford dataset. We subdivide all the correlation coefficients into two groups: one group consists of pairs of measurements that share one or more electrodes and the other consists of all measurement pairs. Next, we grouped departure from the mean or reciprocal errors as a function of dipole-dipole separation multiplier n and plot the mean of each group. We show in Fig. 5 that for all n used for the Boxford dataset, the mean correlation coefficients for measurement pairs that share one or more electrodes are always higher than the means for all pairs. The mean correlation coefficients for reciprocal errors are orders of magnitude smaller than those of departure from the mean. The effect of electrode sharing is also pronounced for reciprocal errors—the mean correlations of all reciprocal errors pairs are negligible while those for pairs that share one or more electrodes are consistently higher. Note that electrode sharing only occurs in ~10% of all pairs. Fig. 5 shows that by taking into account the correlation of the electrodes used to make multiple measurements, ERT measurement errors may be better modelled. With the autocorrelation results, we also show that the departure from the mean exhibits more spatial and temporal correlation than the reciprocal errors.

4. A new error model

4.1. Model definition and implementation

Our error analysis reported in Section 3 revealed that the combination of electrodes used appears to influence ERT measurement errors.

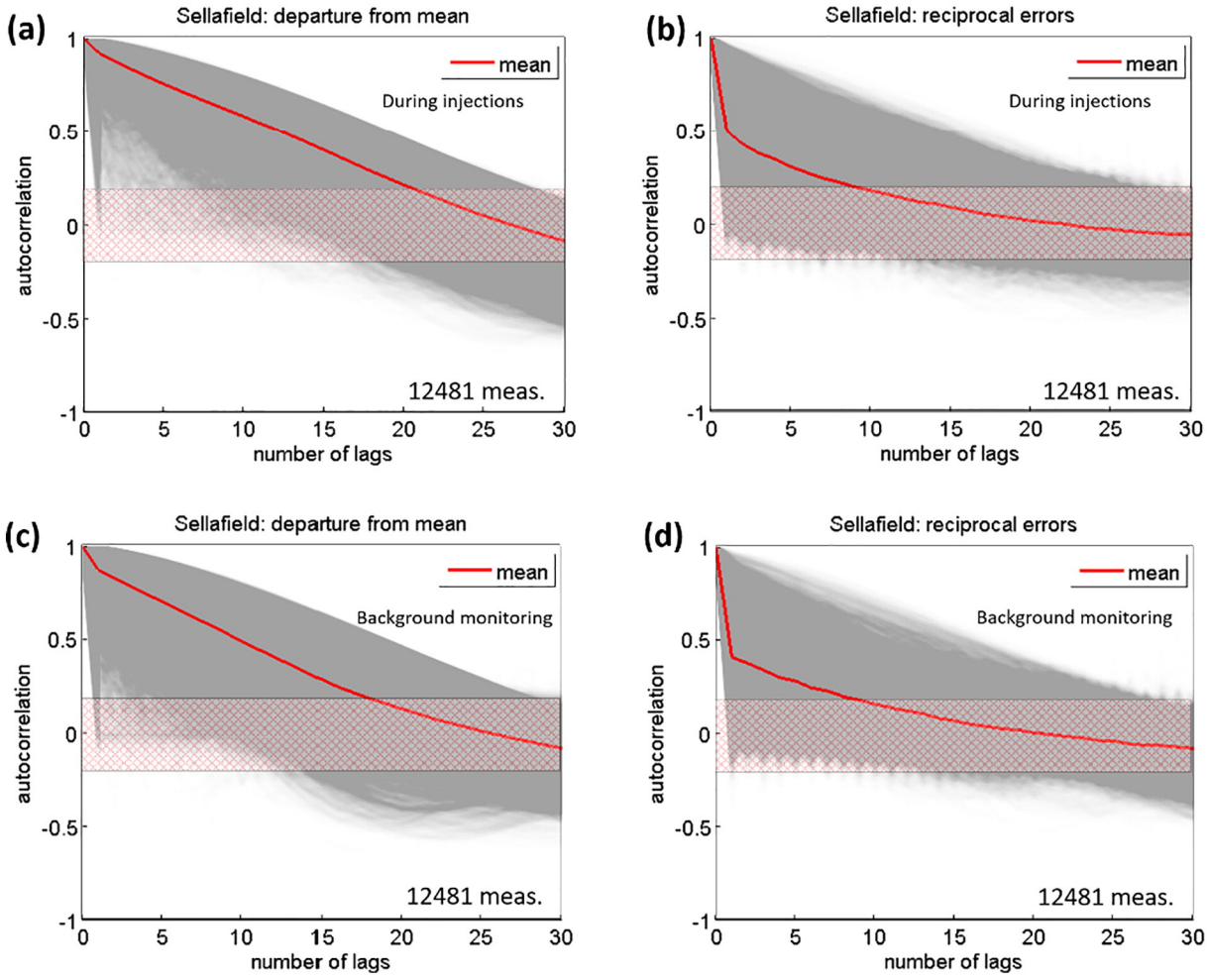

Fig. 4. Autocorrelation of (a) departure from the mean (as a measure of repeatability errors) and (b) reciprocal errors for the 96 datasets collected continuously at the Sellafield site encompassing the three injection periods (22/1/2013–3/11/2013). The number of lags is on the horizontal axis (here 1 lag = ~2 to 3 days). Each grey translucent line plots the autocorrelation of one of the 12,481 ERT measurements as a function of lag. The red line denotes the mean autocorrelation. For each autocorrelation plot, 96 datasets are considered. The hashed region has insignificant correlation according to the critical Pearson's test (around ±0.2). Similarly, (c) and (d) show the same for the long-term background monitoring period (i.e. no injection, 5/11/2013–31/3/2014).

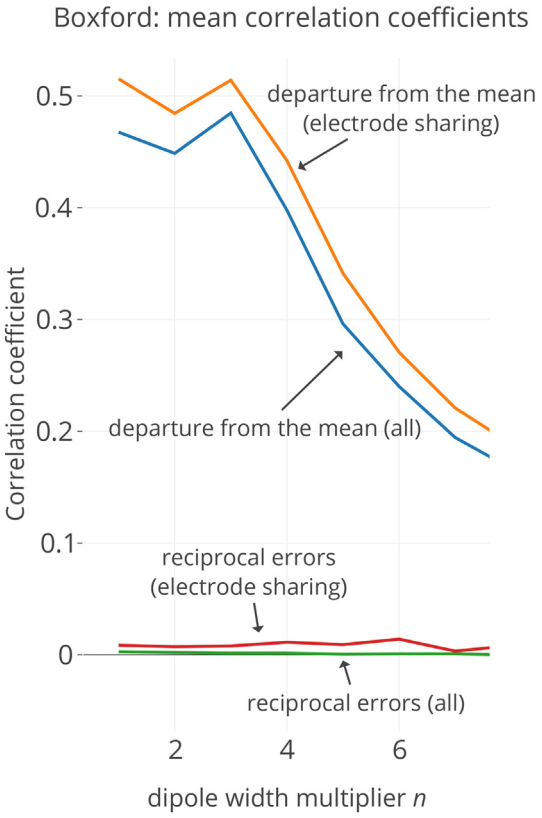

Fig. 5. Mean correlation coefficient of departure from the mean (as a measure of measurement errors) and reciprocal errors for measurement pairs from the Boxford dataset as a function of dipole separation multiplier *n*. For both departure from the mean and reciprocal errors, mean correlation coefficients are distinctively higher for measurements that share electrode(s) in their quadrupoles than the mean correlation coefficients for all measurements, indicating by considering the effect of using each electrode to make multiple measurements may improve error models. Also, note that the reciprocal errors have strikingly lower correlation coefficients than the departure from the mean. Electrode sharing occurs in ~10% of all pairs.

Therefore, we developed a new error model based on linear mixed effects (LME) models to group measurement errors by the electrodes used to obtain them, which allows us to incorporate the effects of electrode combinations.

The linear mixed effect model is a powerful statistical tool in settings where repeated measurements are made on the same statistical units (longitudinal study), or where measurements are made on clusters of related statistical units (Bates et al., 2015; Diggle et al., 2015; Pinheiro and Bates, 1988; West et al., 2007). It is especially useful to group qualitative variables that influence the data. In general, a mixed effect model is given by

$$\mathbf{y} = \underbrace{\boldsymbol{X\beta}}_{fixed} + \underbrace{\boldsymbol{Zb}}_{random} + \underbrace{\boldsymbol{\varepsilon}}_{error} \tag{15}$$

where $\mathbf{y}$ is the n-by-1 response vector, and n is the number of observations; $\boldsymbol{X}$ is an n-by-p fixed-effects design matrix, and p is the number of fixed effect variables; $\boldsymbol{\beta}$ is a p-by-1 fixed-effects vector and q is the number of random effect variables; $\mathbf{Z}$ is an n-by-q random-effects design matrix; and b is a q-by-1 random-effects vector. $\boldsymbol{\varepsilon}$ is an n-by-1 unknown vector of random, independent and identically distributed errors.

Linear mixed effect models can now be readily implemented using the MATLAB® statistics and machine learning toolbox and the lme4 package for R (Bates et al., 2015). In this paper, we model measurement errors in ERT by treating transfer resistances as fixed effects and each of the electrodes used (A, B, M, N) as grouping variables. The above model was implemented in MATLAB® (see Supplementary information for more details).

The linear mixed effect model essentially establishes a hierarchy or grouping when fitting the measurement errors. Fitting is achieved by both optimizing fit within each cluster, while the covariate vectors link the fixed and random effects between clusters. The clustering introduces additional degrees of freedom that allow a better fit of measurement errors than commonly used linear models. An illustrative example of the LME grouping formulation can be found in the Supplementary information, along with details for fitting the LME error model to the Boxford and Sellafield field datasets. The evolution of the error model coefficients with time is also described.

4.2. LME error model behaviour for time-lapse ERT measurements

A longitudinal survey is a correlational research study that involves repeated observations of the same variables over long periods of time. One of the original uses of LME models is to handle longitudinal data in tracking studies to eliminate potential bias of using the same samples. For example, in a drug study the health indicators of the same group of patients are sampled multiple times during a long period. The times at which they are sampled can be used as an additional grouping variable in the LME model. With the increased popularity of long-term monitoring using ERT and other geophysical methods, it may be beneficial to treat measurement errors as longitudinal data too. In Fig. 6, we compare fitting observed measurement errors in the 96 repeat datasets from Boxford individually (i.e. obtaining 96 LME equations) and as longitudinal data (i.e. obtaining one LME equation, with the repeat number as an additional grouping variable). The scatter plots show that a much better fit is obtained by fitting each of the 96 datasets individually. In other words, treating ERT measurement errors as longitudinal data does not better characterise them. Measurement errors for time-lapse datasets should instead be characterised on a frame-by-frame basis.

5. Comparison of error models using image appraisal

Improvements in the measurement error model are only useful if they can lead to better inversion results. We applied the new error model to the synthetic data and field data from the Boxford and Sellafield sites. Also, we will consider the resolution matrix and model variance from Monte Carlo simulations to see whether using the new error model can give additional insight to data and reduce uncertainty.

5.1. Synthetic data

Using the synthetic domain introduced in Section 2.1.1, we compared the inversion results and the corresponding resolution matrices and uncertainty estimates using different error models. Note that since Fig. 1 and Fig. 7 use the same domain and have the same resistivity structure, Fig. 1(c) can be seen as a benchmark case where the data is inverted with perfect knowledge of measurement errors.

Fig. 7 (a–c) shows inversion results for synthetic data where measurements involving three “bad electrodes” are corrupted by 10% noise and others by 2% noise. We first compare the inversion with two linear error model—one assumes there are no bad electrodes (i.e. the 2% error model), while the other is obtained by fitting the corrupted data with the Koestel et al. (2008) model (i.e. the 4.52% linear model). We see that the resultant resistivity model from assuming the 2% linear error model is very noisy while that from assuming 4.52% linear error model is smoother. With the LME error model, however, the inversion result is the most similar to that of the benchmark case (Fig. 1c) (see also rms errors printed on plots). The effect of better characterisation of measurement errors by the LME model is manifested in the inversion results.

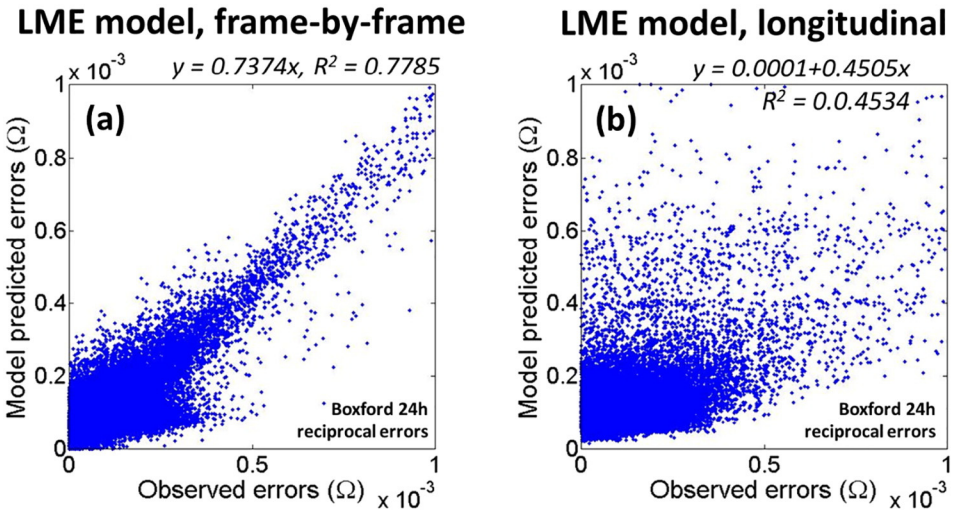

Fig. 6. Comparison of fitting reciprocal errors of time-lapse data as (a) individual datasets (fitting each dataset individually with a different LME model) and (b) longitudinal data (fitting all data with one LME model). The above shows that it is much better not to treat errors as longitudinal data.

Fig. 8 (a–c) shows the diagonal terms of the resolution matrices for the inversion using (a) 2% linear, (b) 4.52% linear, and (c) LME error models. In general, the resolution patterns are uniform laterally yet decreases with depth. For the 2% linear error model, we see that some of the artefacts from the inversion results is also shown on the resolution pattern. For the LME error model, the resolution on the right is somewhat higher than on the left for the top layer, where the bad electrodes are located. The resolution values are between that of (a) and (b) in most of the cells, although some of the cells near the surface show very high resolution. The above shows that while the resolution from the linear error model is purely a function of distance away from sources and sensors and therefore cannot distinguish quality between measurements, the LME error model allows the inversion to resolve areas unaffected by the bad electrodes better.

Subsequently, we ran Monte Carlo experiments using the procedure in Section 2.4 to understand how uncertainty in measurement errors propagates to affect uncertainty in the parameter estimates. The Monte Carlo experiment results can be used to form empirical model covariance matrices. This matrix shows how information is shared between parameters (i.e. model estimate of different elements). In the ideal, noise-free and well-defined case, the model covariance matrix should be a zero matrix, meaning the parameter is deterministically known and the parameters are not correlated with one another. Fig. 9 (a–b) show that assuming a 2% linear error model yields lower model covariances than the 4.52% model, which is expected because lower percentage error implies less sharing of data. Also, the band of high covariances is also narrower. With the LME model (Fig. 9c), however, we notice that the model covariances values are lower than those of the

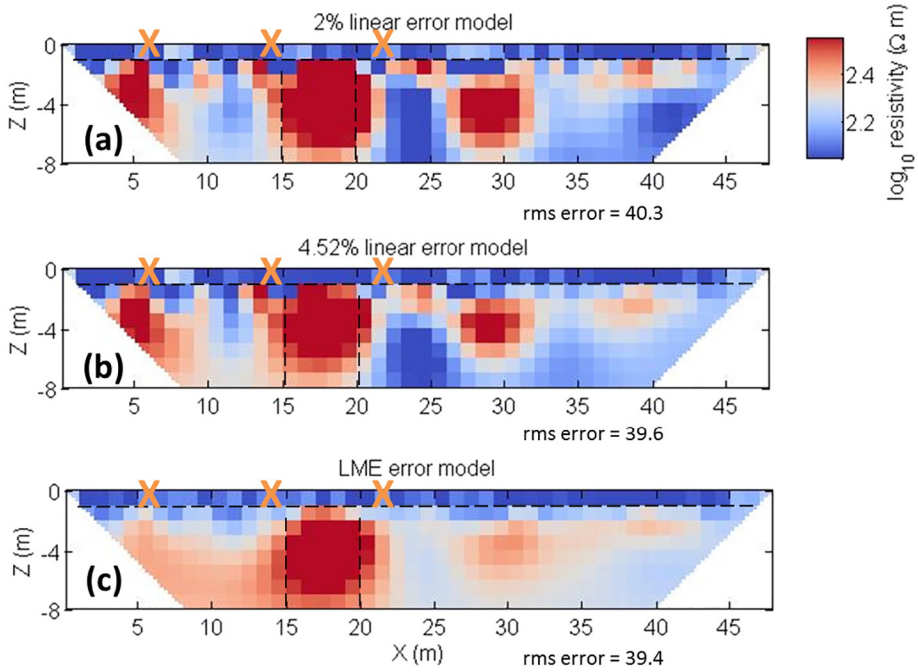

Fig. 7. Synthetic surface ERT experiments to demonstrate the performance of the error models. For data involving 3 bad electrodes(marked by "X"), data is corrupted by 10% white noise while for the rest of the data 2% white noise is added. (a) Inverted resistivity distribution using the 2% linear error model (a) Inverted resistivity distribution using a 4.52% (obtained from the Koestel et al., 2008 method) linear error model (c) Inverted resistivity distribution using the LME error model. Note that the convergence target for all the inversions is a chi-squared statistic of 1.

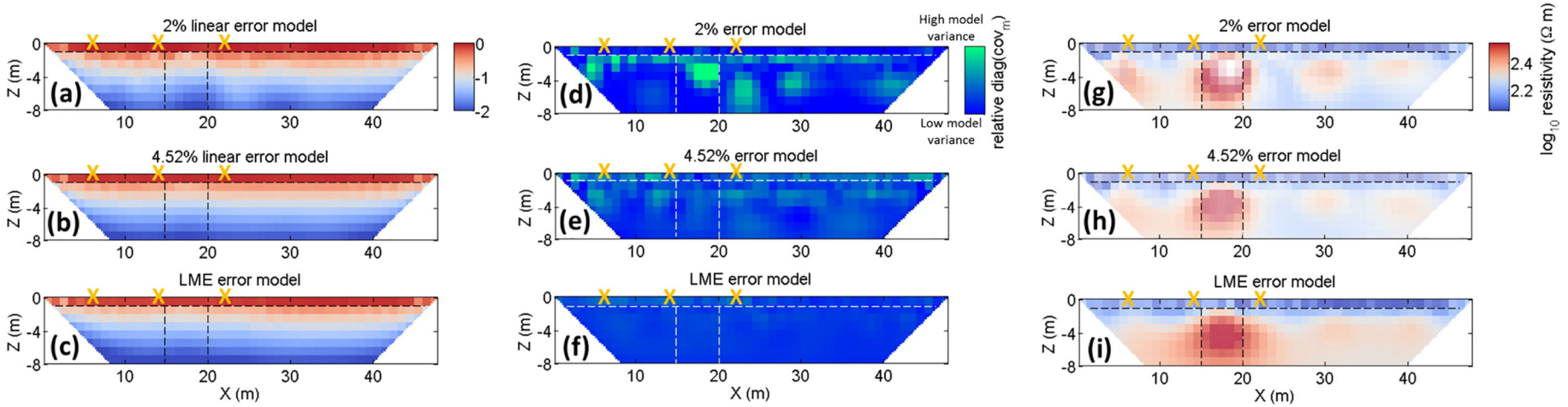

Fig. 8. (a–c) Diagonal of resolution matrix for inversion using the following error models for inverting the synthetic data corrupted by "bad electrodes": (a) 2% linear model (b) 4.52% linear model (c) LME model. (d–f) variance of element-wise log-resistivity estimates using each of the error models obtained from Monte Carlo experiments. The colour scale is the same for all three error models. Darker cells indicate more similar model estimates among Monte Carlo estimates. (g–i) mean model estimates from Monte Carlo experiments. The transparency is controlled linearly the variance shown in (d–f). With model averaging, the mean estimates of the three error models agree. It is noted, however, the deterministic results from the LME model agrees the best with its model-averaged results.

2% and 4.52% models. More importantly, the spread of the high covariance region is less uniform than the linear models, meaning that only measurements affected by the bad electrodes share information heavily with others. The above agrees with the comparison of resolution matrices—the new error model is able to exploit information in noisy data without increasing the overall noise level.

The diagonal term of the empirical model covariance matrix (i.e. variance) shows the variability among parameter estimates from multiple Monte Carlo simulation realizations. Specifically, the higher a diagonal term, the more uncertain is the estimate. We plot their ranges in Fig. 8 (d–f). For all three error models, the variability is always the smallest at depths because deeper regions are less well resolved for surface arrays. As a result, the model estimates at greater depths are closer to the initial guess values and therefore, there is less difference between the realizations of Monte Carlo model estimates. In Fig. 8 (g–i), we plot the model-averaged parameter estimates. The transparency of each element is inversely proportional to its model variance, as shown in Fig. 8 (d–f). In other words, the elements that have more variable or uncertain estimates have greater transparency. The inversion results from assuming a 2% error model are less variable than for the 4.52% model. Model averaging also smooths out the noisy artefacts from deterministic inversion (compare Fig. 8g and Fig. 7a). The LME error model gives the most reliable model estimates among the three error models tested. Also, it is worth noting that the model-averaged parameter estimates are comparable to that obtained from deterministic inversion. This means that with the LME error model, there is no need to run many realizations of the inverse model in order to obtain reliable parameter estimates. Importantly, inversion using the LME error model gives the highest resolution and the least model variance (Fig. 8), and reduces uncertainty in inversion results.

5.2. Boxford field data

In Fig. 10, we compare the inversion results of field data for the Boxford datasets. When using reciprocal data, we only consider one of the 96 available datasets (i.e. the first of the 96 repeats). The resultant image from using linear or LME error models for reciprocal or 24-h repeatability errors (not shown) for the Boxford dataset are effectively identical. When the linear model is applied to the stacking errors, the resultant image becomes quite noisy. Surprisingly, when the LME model is applied to the stacking errors, there is no distinguishable difference between its result and those using reciprocal or 24-h repeatability errors. This shows that although we have shown above and warn against the potential underestimation of measurement errors caused by using the stacking errors, the LME error model is capable of minimizing such effects. We suspect that because of the low mean and low variance of the stacking errors, the linear error model is forced to assign very low errors across the dataset. The LME error model, in contrast, has more degrees of freedom to better fit the observed stacking errors.

This finding has significant implications because all modern ERT equipments output stacking errors and these do not require additional data collection time. For many existing datasets where only stacking errors are available or in applications where the collection of repeats and reciprocal is prohibitive, we recommend using a LME error model instead of a linear model for the stacking errors.

5.3. Sellafield field data

We inverted the Sellafield data collected on 5th February 2013, which was two days before the first tracer injection (Kuras et al., 2016). Of the 51,302 measurements in the sequence, there are 12,412 pairs of valid reciprocal measurements. We fitted them with the LME error model. Note that we have not removed any high-error outliers. Fig. 11 shows the resultant 3-D static inversion image and its associated uncertainty estimates (model standard deviation and model coefficient of variation) derived from Monte Carlo simulations. The resultant model clearly delineates zones of high and low resistivities. In terms of uncertainty, regions next to the borehole and towards the top of the monitoring array have significantly higher model standard deviation. Compared with the absolute images of resistivity reported in Kuras et al. (2016) (note that we use the same mesh and inversion code), Fig. 11a shows similar patterns but with a smaller range and variations in resistivities.

6. Discussion

In the present study, we have used statistical methods to explore ways to improve the current practice of modelling measurement errors in ERT. Among them, we have found that the correlation coefficients of measurement pairs that share some of the electrodes are consistently higher than average. Therefore, we have developped a new error model that consider such effects in ERT surveys by adding electrode-specific fitting terms (i.e. the LME error model).

The proposed error model based on the linear mixed effect (LME) model shows superior performance in terms of characterising errors when compared against a unknown linear error model. The LME model assumes that errors are linearly dependent on transfer resistances and employs the electrodes used to make each measurement as grouping variables. The LME error model can more accurately predict observed measurement errors. However, as we have already argued in Section 4.2, individual errors should not be used directly for inversion because in most practical situations they are only averages between two points. To improve the robustness of the linear error models, errors can be grouped by the magnitude of transfer resistances (Koestel et al., 2008). Such binning, however, is arbitrary and the resultant error

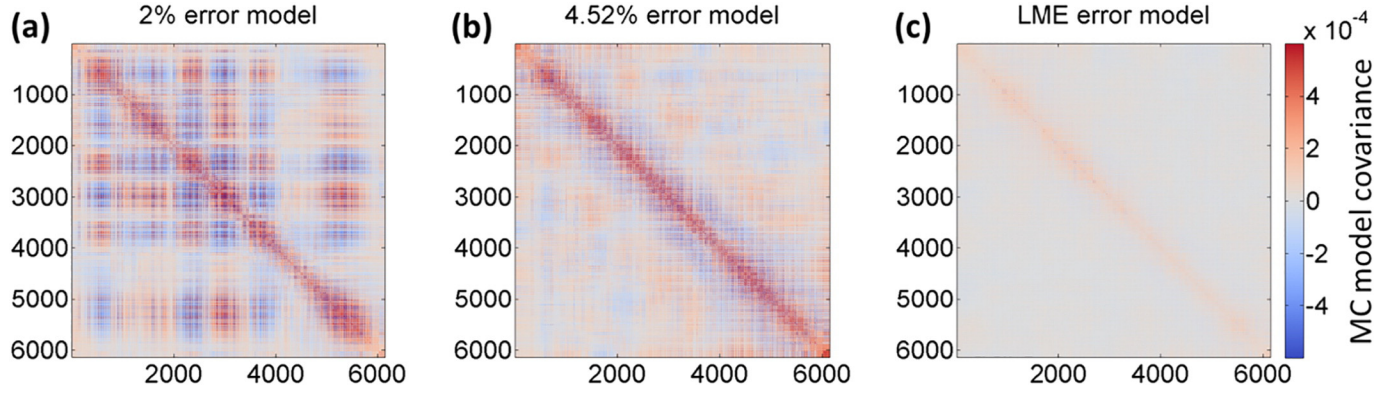

Fig. 9. Empirical model covariance matrix using the Monte Carlo uncertainty propagation procedure and the following error models: (a) 2% linear model (b) 4.52% linear model (c) LME model. The size of the matrix is $m \times m$, where m is the number of model parameters. By comparing (a) and (b), it is shown that assuming higher error levels, there is higher covariance between model parameters. With the LME error model, the model covariance is the lowest. While the spread of high covariance entries are quite even throughout the matrix, we can see that the spread for (c) is quite uneven: generally, elements on the left of the domain have higher spread.

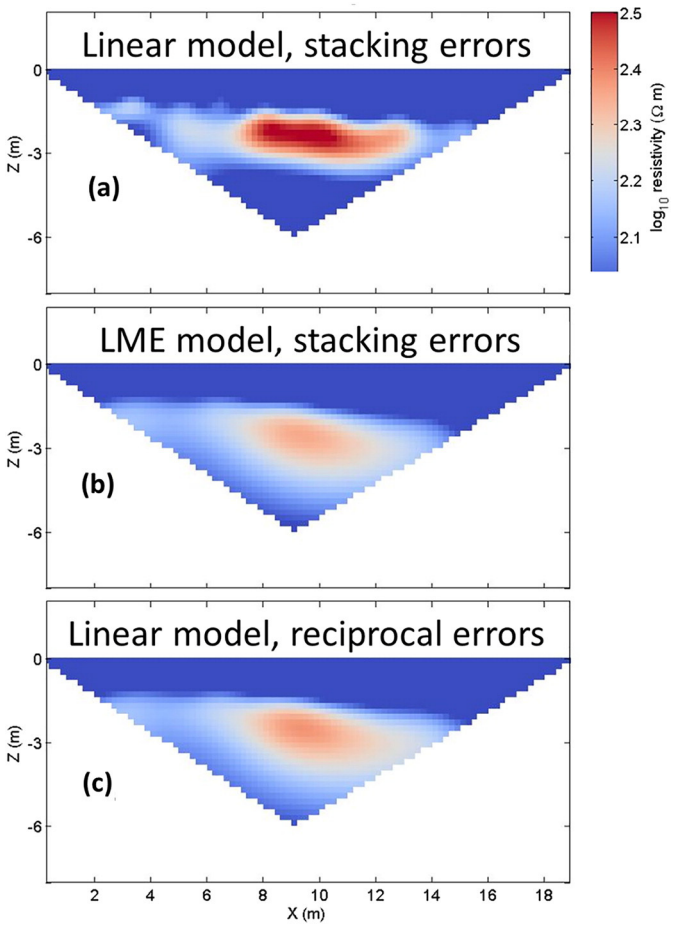

Fig. 10. Inversion results from Boxford using (a) linear error model for stacking errors, (b) LME error model for stacking errors, (c) linear error model for reciprocal errors.

models can be sensitive to the number of bins used. The LME error model is based on the same idea of grouping, yet it considers all of the four electrodes that are used to make each dipole-dipole quadrupole measurement and uses them as the grouping variable. Electrode number is a qualitative variable and it is a reasonable assumption that each electrode has slightly different quality.

The patterns of resolution matrix and model covariance matrix associated with using the LME error model are different from those using the linear model. This has important implications for inversion and uncertainty estimation because it shows that the LME model is capable of detecting poorer measurements and downweighting them in an inversion. Most inversion schemes are capable of weighting data according to their quality. Yet in common ERT practice, either uniform percentage errors (i.e. a linear model) are assumed or the errors are not characterised at all. The LME error model is one of the first statistical tools to characterise the variable quality of ERT measurements (while not using individual errors directly) so that the data weighting schemes in inverse models can be fully utilized.

While fitting a LME model for each set of reciprocal errors gives promising estimates, fitting time series of reciprocal errors with a single LME model and using the sequence of data collection as an additional grouping variable (i.e. as longitudinal data) can yield inferior results. Evaluation of the individual resultant LME error models reveals that, for the dataset considered here, the fixed and random effects coefficients vary over the 24-hour period. Such results challenge our common assumption that electrode quality is extremely stable. The laboratory study by LaBrecque and Daily (2008) on the measurement errors of 15 electrode materials showed many possibilities for electrode quality to evolve during the course of a ERT experiment, some even in the timescale of minutes. Therefore, taking many repeats for measurements probably will not provide better error estimates because electrode quality may evolve during the process. In summary, we recommend the collection of reciprocal measurements at each timeframe and fit a LME model based on the measured transfer resistance and electrodes used to capture the minor variations in electrode quality during ERT experiments.

We have found in the Boxford inversion results that there is no distinguishable difference between using repeatability and reciprocal errors in inversions (figures not shown). From the PDFs, the stacking

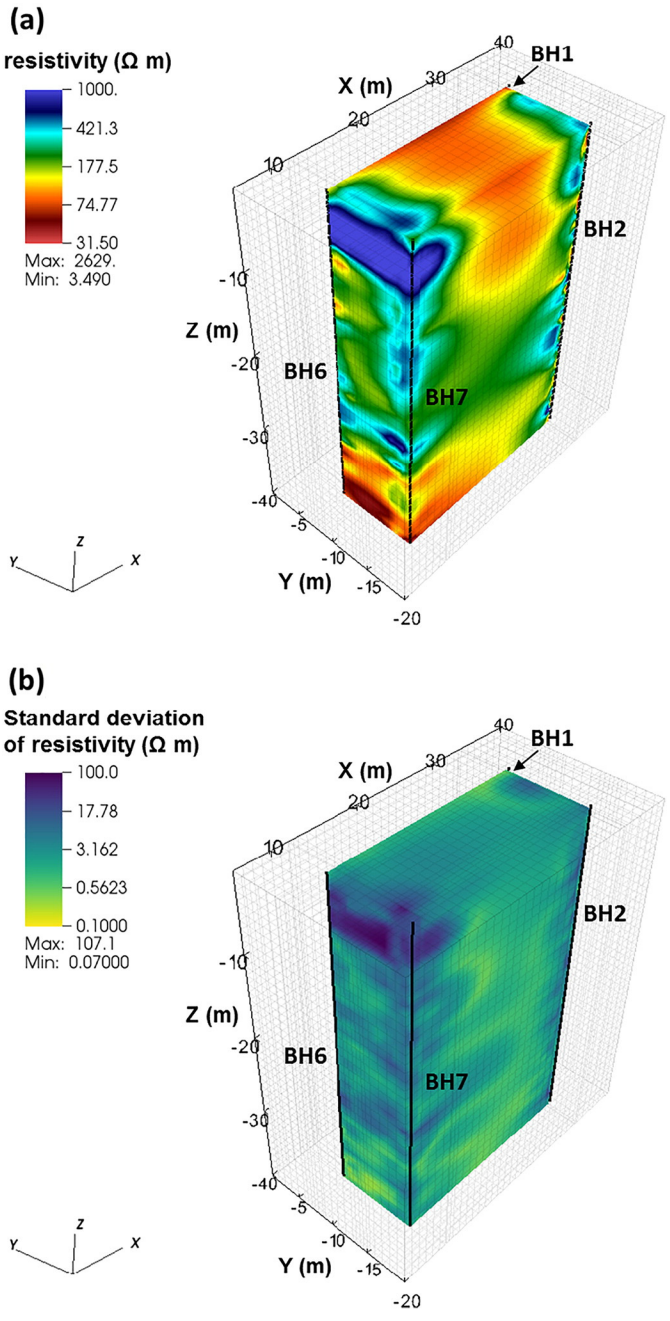

Fig. 11. (a) 3-D static deterministic inversion results from Sellafield on 5th February 2013. Error weights are prescribed by fitting an LME error model. Black lines are boreholes installed with electrodes. (b) The corresponding uncertainty estimates obtained from Monte Carlo simulations, given by model standard deviation from Monte Carlo experiments. (c) The corresponding coefficient of variation of Monte Carlo model estimates.

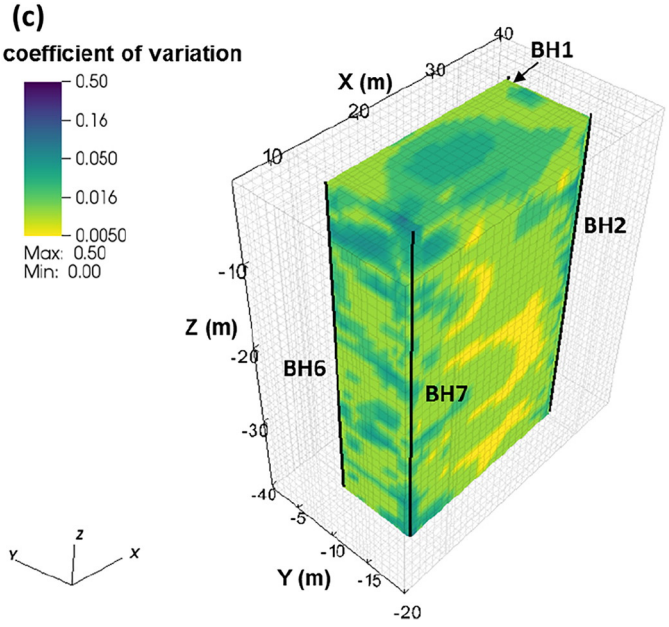

Fig. 11 (*continued*).

errors are much smaller and much less variable than the repeatability or reciprocal errors at Boxford. With the linear error model, the resultant image for using stacking errors is noisy. With the LME error model, however, the inversion image is comparable to that obtained from using repeatability or reciprocal errors. We attributed its better results to the better handling of spurious and overly optimistic estimates of errors by the LME error model.

For the Sellafield dataset, we demonstrated the application of the new LME error model to model reciprocal errors and used its predicted errors for 3-D inversion and uncertainty quantification (i.e. model variance). Such uncertainty estimates are useful as they visualize how uncertainties in measurements propagate to uncertainties in the inverse model estimate.

We have highlighted in the previous section that the new LME error model can be widely applied to essentially any ERT inversion algorithms. It better predicts errors that are used to prescribe weights of the data weighting or covariance matrix. The resultant matrix remains diagonal so that it does not increase computation costs during inversion. Unlike the data quality control strategies recently proposed by Deceuster et al. (2013) and Mitchell and Oldenburg (2016), the new LME error model can be applied to any static and time-lapse ERT problems regardless of their size. Since the model considers the effect of the variable quality of electrodes, it requires minimum culling of data or re-inversion. Alternatively, the new LME error model can be used alongside with other data quality control strategies.

The flexibility of the LME model allows it to be applied to characterisation of errors in other geophysical measurements. For example, geophones used in seismic tomography can be used as grouping variables for their errors. A straightforward next step for future study would be to extend the LME error model to induced polarisation (IP) studies. It has been reported in the literature that IP surveys are even more sensitive to electrode configuration than ERT. Much recent work has been done to improve quality of IP measurements. For example, Dahlin et al. (2013) conducted a duplicate IP survey for a planned tunnel using two types of cable spreads: one with standard multi-core cables and the other with separate cable spreads for transmitting current and measuring potentials. They suggest that the single cable spread is sufficient to give good IP data but suggest the use of separate cable spread for spectral IP inversion and recovery of Cole-Cole parameters. Flores-Orozco et al. (2012) quantified the measurement errors in spectral IP imaging and established a new phase error model. It is an extension of previous models where the discrepancy between normal and reciprocal measurements is analysed (Binley et al., 1995; LaBrecque et al., 1996; Slater and Binley, 2006). They also conducted a bin analysis to ensure the assumption of a normal distribution of errors is valid and showed that, for spectral IP measurements, phase error discrepancies show a consistent behaviour for all frequencies. They proposed an inverse power-law relationship between the error on phase and the corresponding resistance. This brief review highlights the similarities between ERT and IP measurement error models and we believe that the proposed LME model can improve IP measurement errors characterisation, too. Future studies should consider applying the LME error model.

Finally, the proposed LME models can be used readily in Bayesian formulations for ERT inversion. The LME error model can be used to prescribe entries of the data covariance matrix in their likelihood functions, which are usually assumed to be diagonal for computational convenience. Note that the LME method considers errors due to electrodes used as a grouping variable rather than enforcing a correlation function, which would lead to a full data covariance matrix that is computationally difficult to invert. By treating the potential correlation of electrode effects as grouping variables instead, the data weighting or covariance matrix remains diagonal; furthermore, strict and unnecessary assumptions on the correlation between measurements are avoided.

7. Conclusion and recommendations

Our analysis of field datasets shows that short-term repeatability and reciprocal errors are very comparable, while stacking errors are significantly lower. Repeatability errors, however, may increase over time because of subsurface changes between repeats. Repeatability errors also tend to show greater autocorrelation in time for the same measurements, as well as correlation between measurements, than reciprocal errors. Stacking errors are found to have significantly lower magnitude and variability, indicating it may be an overly optimistic measure of measurement error. Correlation coefficients between pairs of measurements that share some of the electrodes used are higher than pairs that use completely different electrodes. This confirms speculations from previous studies that the common use of electrodes may contribute to some correlation in errors (Ramirez et al., 2005).

Based on our error analysis, we confirm the value of collecting reciprocal data in ERT studies, although when making reciprocal measurements, care should be taken to avoid electrode charge-up effects (Dahlin, 2000; Wilkinson et al., 2012). If it is too difficult to set up reciprocal measurements, we recommend running a duplicate survey immediately after the completion of the original survey. Long-term repeatability data does not bring extra benefits for fitting error models because subsurface conditions may change over time. But they may be very useful for long-term quality assurance, for example, detecting instrument drift or abnormal system behaviour. Stacking errors should be avoided when assigning error weights because of their low magnitude and low variability. For modelling the measurement errors, we recommend fitting a linear mixed effect (LME) model over the commonly used linear model. The new LME error model uses both the combination of electrodes used for making ERT measurements and the proportional relationship between errors and transferred resistance in order to better characterize measurement errors. Our synthetic example shows that the LME error model is capable of picking up errors due to the varying quality of electrodes and adjusts resolutions in the inverse model accordingly. This is different from the traditional linear model approach where the resolution everywhere in the entire inverse model domain has to reduce. The new LME model not only improves the inversion results, but also reduces the uncertainty (i.e. variance) in the model estimates. For time-lapse data, we recommend fitting a LME model for each time step because its coefficients change over time and fitting all the data from the different time steps with a single LME model (i.e. as longitudinal data) yields inferior results. We have demonstrated the

applicability of the above-recommended procedure by fitting the LME model to errors observed in two field datasets and inverting the data. This procedure is easy to implement and requires minimal changes to the current practice. Widely implementing this procedure in future geophysical studies can greatly improve their overall reliability—a necessary step towards obtaining more quantitative information from geophysical methods across a range of disciplines and applications.

Acknowledgements

This work is supported by a Nuclear Decommissioning Authority PhD Bursary awarded to the first author and an accompanying PhD studentship provided by Lancaster University. This paper is published with the permission of the Nuclear Decommissioning Authority, Sellafield Ltd., and the Executive Director of the British Geological Survey (NERC). We thank Paul Mclachlan (Lancaster University) for assistance on using R. We are grateful to the two anonymous reviewers for their constructive reviews of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jappgeo.2017.09.009.

References

Aster, R., Borchers, B., Thurber, C.H., 2005. Parameter Estimation and Inverse Problems. Elsevier, Burlington, Massachusetts, USA.
Barker, J.A., 1991. The reciprocity principle and an analytical solution for Darcian flow in a network. Water Resour. Res. 27:743–746. https://doi.org/10.1029/91WR00258.
Bates, D., Maechler, M., Bolker, B.M., Walker, S., 2015. Fitting linear mixed-effects models using {lme4}. J. Stat. Softw. 67:1–48. https://doi.org/10.18637/jss.v067.i01.
Beven, K., Westerberg, I., 2011. On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process. 25:1676–1680. https://doi.org/10.1002/hyp.7963.
Binley, A., 2015. Tools and techniques: electrical methods. Treatise Geophys. https://doi.org/10.1016/B978-0-444-53802-4.00192-5.
Binley, A., Ramirez, A.L., Daily, W., 1995. Regularised image reconstruction of noisy electrical resistance data. Proceedings of the 4th Workshop of the European Concerted Action on Process Tomography, pp. 401–410 (Bergen).
Binley, A.M., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K., Slater, L.D., 2015. The emergence of hydrogeophsyics for improved understanding of subsurface processes over multiple scales. Water Resour. Res. 51:3837–3866. https://doi.org/10.1002/2015WR017016.
Bruggeman, G.A., 1972. The reciprocity principle in flow through heterogeneous porous media. Fundamentals of Transport Phenomena in Porous Media. Elsevier, New York, p. 136.
Cassiani, G., Bruno, V., Villa, A., Fusi, N., Binley, A.M., 2006. A saline tracer test monitored via time-lapse surface electrical resistivity tomography. J. Appl. Geophys. 59: 244–259. https://doi.org/10.1016/j.jappgeo.2005.10.007.
Chambers, J.E., Wilkinson, P.B., Uhlemann, S., Sorensen, J.P.R., Roberts, C., Newell, A.J., Ward, W.O.C., Binley, A., Williams, P.J., Gooddy, D.C., Old, G., Bai, L., 2014. Derivation of lowland riparian wetland deposit architecture using geophysical image analysis and interface detection. Water Resour. Res. 50:5886–5905. https://doi.org/10.1002/2014WR015643.
Crook, N., Binley, A., Knight, R., Robinson, D.A., Zarnetske, J., Haggerty, R., 2010. Electrical resistivity imaging of the architecture of substream sediments. Water Resour. Res. 46: W00D13. https://doi.org/10.1029/2008WR006968.
Dahlin, T., 2000. Short note on electrode charge-up effects in DC resistivity data acquistiion using multi-electrode arrays. Geophys. Prospect. 48:181–187. https://doi.org/10.1046/j.1365-2478.2000.00172.x.
Dahlin, T., Dalsegg, E., Sandstrom, T., 2013. Data quality quantification for time domain IP data acquired along a planned tunnel near Oslo, Norway. Near Surface Geoscience 2013—19th European Meeting of Environmental and Engineering Geophysics. Bochum, Germany.
Day-Lewis, F.D., Johnson, C.D., Singha, K., Lane Jr., J.W., 2008. Best practices in electrical resistivity imaging: Data collection and processing, and application to data from Corinna, Maine. EPA Report, Boston, MA.
Deceuster, J., Kaufmann, O., Camp, M. Van, 2013. Automated identification of changes in electrode contact properties for long-term permanent ERT monitoring experiments. Geophysics 78:E79–E94. https://doi.org/10.1190/GEO2012-0088.1.
Delay, F., Ackerer, P., Guadagnini, A., 2011. Theoretical analysis and field evidence of reciprocity gaps during interference pumping tests. Adv. Water Resour. 34:592–606. https://doi.org/10.1016/j.advwatres.2011.02.006.
Diggle, P.J., Heagerty, P., Liang, K.-Y., Zeger, S., 2015. Analysis of longitudinal data. 2nd ed. Oxford University Press.
Falade, G.K., 1981. Analysis of the reciprocity concept in a porous medium. Water Resour. Res. 17:918–920. https://doi.org/10.1029/WR017i004p00918.
Flores-Orozco, A., Williams, K.H., Long, P.E., Hubbard, S.S., Kemna, A., 2011. Using complex resistivity imaging to infer biogeochemical processes associated with bioremediation of an uranium-contaminated aquifer. J. Geophys. Res. Biogeosciences 116:1–17. https://doi.org/10.1029/2010JG001591.
Flores-Orozco, A., Kemna, A., Zimmermann, E., 2012. Data error quantification in spectral induced polarization imaging. Geophysics 77, E227. https://doi.org/10.1190/geo2010-0194.1.
French, H.K., Hardbattle, C., Binley, A., Winship, P., Jakobsen, L., 2002. Monitoring snowmelt induced unsaturated flow and transport using electrical resistivity tomography. J. Hydrol. 267:273–284. https://doi.org/10.1016/S0022-1694(02)00156-7.
Gélis, C., Revil, A., Cushing, M.E., Jougnot, D., Lemeille, F., Cabrera, J., de Hoyos, A., Rocher, M., 2010. Potential of electrical resistivity tomography to detect fault zones in limestone and argillaceous formations in the experimental platform of Tournemire, France. Pure Appl. Geophys. 167:1405–1418. https://doi.org/10.1007/s00024-010-0097-x.
Haarder, E.B., Jensen, K.H., Binley, A.M., Nielsen, L., Uglebjerg, T.B., Looms, M.C., 2015. Estimation of recharge from long-term monitoring of saline tracer transport using electrical resistivity tomography. Vadose Zo. J. https://doi.org/10.2136/vzj2014.08.0110.
Hayley, K., Bentley, L.R., Gharibi, M., 2009. Time-lapse electrical resistivity monitoring of salt-affected soil and groundwater. Water Resour. Res. 45:1–14. https://doi.org/10.1029/2008WR007616.
Henderson, R.D., Day-Lewis, F.D., Abarca, E., Harvey, C.F., Karam, H.N., Liu, L., Lane, J.W.J., 2010. Marine electrical resistivity imaging of submarine groundwater discharge: sensitivity analysis and application in Waquoit Bay, Massachusetts, USA. Hydrogeol. J. 18: 173–185. https://doi.org/10.1007/s10040-009-0498-z.
Hermans, T., Vandenbohede, A., Lebbe, L., Nguyen, F., 2012. A shallow geothermal experiment in a sandy aquifer monitored using electric resistivity tomography. Geophysics 77:B11. https://doi.org/10.1190/geo2011-0199.1.
Huisman, J.A., Rings, J., Vrugt, J.A., Sorg, J., Vereecken, H., 2010. Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion. J. Hydrol. 380:62–73. https://doi.org/10.1016/j.jhydrol.2009.10.023.
Irving, J., Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resour. Res. 46, W11514. https://doi.org/10.1029/2009WR008340.
Johnson, T.C., Slater, L.D., Ntarlagiannis, D., Day-Lewis, F.D., Elwaseif, M., 2012. Monitoring groundwater-surface water interaction using time-series and time-frequency analysis of transient three-dimensional electrical resistivity changes. Water Resour. Res. 48:1–13. https://doi.org/10.1029/2012WR011893.
Johnson, T.C., Versteeg, R.J., Thomle, J., Hammond, G., Chen, X., Zachara, J., 2015. Four-dimensional electrical conductivitymonitoring of stage-driven river water intrusion: accounting for water table effects using a transient mesh boundary and conditional inversion constraints. Water Resour. Res. 51:6177–6196. https://doi.org/10.1002/2014WR016259.
Kemna, A., 2000. Tomographic Inversion of Complex Resistivity. Ruhr-Universität Bochum.
Kim, J., Supper, R., Ottowitz, D., Jochum, B., Yi, M., 2016. A new measurement protocol of direct current resistivity data. Geophysics 81.
Koestel, J., Kemna, A., Javaux, M., Binley, A., Vereecken, H., 2008. Quantitative imaging of solute transport in an unsaturated and undisturbed soil monolith with 3-D ERT and TDR. Water Resour. Res. 44:1–17. https://doi.org/10.1029/2007WR006755.
Kuras, O., Wilkinson, P.B., Meldrum, P.I., Oxby, L.S., Uhlemann, S., Chambers, J.E., Binley, A., Graham, J., Smith, N.T., Atherton, N., 2016. Geoelectrical monitoring of simulated subsurface leakage to support high-hazard nuclear decommissioning at the Sellafield Site, UK. Sci. Total Environ. 566–567:350–359. https://doi.org/10.1016/j.scitotenv.2016.04.212.
LaBrecque, D., Daily, W., 2008. Assessment of measurement errors for galvanic-resistivity electrodes of different composition. Geophysics 73 (2):F55–F64. https://doi.org/10.1190/1.2823457.
LaBrecque, D.J., Mletto, M., Daily, W., Ramirez, A.L., Owen, E., 1996. The effects of noise on Occam's inversion of resistivity tomography data. Geophysics 61:538. https://doi.org/10.1190/1.1443980.
Lesparre, N., Boyle, A., Grychtol, B., Cabrera, J., Marteau, J., Adler, A., 2016. Electrical resistivity imaging in transmission between surface and underground tunnel for fault characterization. J. Appl. Geophys. https://doi.org/10.1016/j.jappgeo.2016.03.004.
Linde, N., Renard, P., Mukerji, T., Caers, J., 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review. Adv. Water Resour. 86:86–101. https://doi.org/10.1016/j.advwatres.2015.09.019.
Loke, M.H., Barker, R.D., 1996. Practical techniques for 3D resistivity surveys and data inversion. Geophys. Prospect. 44:499–523. https://doi.org/10.1111/j.1365-2478.1996.tb00162.x.
Meyerhoff, S.B., Maxwell, R.M., Revil, A., Martin, J.B., Karaoulis, M., Graham, W.D., 2014. Characterization of groundwater and surface water mixing in a semiconfined karst aquifer using time-lapse electrical resistivity tomography. Water Resour. Res. 50: 2566–2585. https://doi.org/10.1002/2013WR013991.
Mitchell, M.A., Oldenburg, D.W., 2016. Data quality control methodology for large, non-conventional DC resistivity datasets. J. Appl. Geophys. 3:948–953. https://doi.org/10.1016/j.jappgeo.2016.09.018.
Morelli, G., LaBrecque, D.J., 1996. Advances in ERT inverse modeling. Eur. J. Environ. Eng. Geophys. 1, 171–186.
Musgrave, H., Binley, A., 2011. Revealing the temporal dynamics of subsurface temperature in a wetland using time-lapse geophysics. J. Hydrol. 396:258–266. https://doi.org/10.1016/j.jhydrol.2010.11.008.
Nimmer, R.E., Osiensky, J.L., Binley, A.M., Sprenke, K.F., Williams, B.C., 2007. Electrical resistivity imaging of conductive plume dilution in fractured rock. Hydrogeol. J. 15: 877–890. https://doi.org/10.1007/s10040-007-0159-z.

Parasnis, D.S., 1988. Reciprocity theorems in geoelectric and geoelectromagnetic work. Geoexploration 25:177–198. https://doi.org/10.1016/0016-7142(88)90014-2.
Pearson, E.S., Hartley, H.O. (Eds.), 1970. Biometrika Tables for Statisticians. Cambridge University Press.
Perri, M.T., Cassiani, G., Gervasio, I., Deiana, R., Binley, A., 2012. A saline tracer test monitored via both surface and cross-borehole electrical resistivity tomography: comparison of time-lapse results. J. Appl. Geophys. 79:6–16. https://doi.org/10.1016/j.jappgeo.2011.12.011.
Pinheiro, J., Bates, D., 1988. Unconstrained parameterizations for variance-covariance matrices. Stat. Comput. 6, 289–296.
Ramirez, A., Daily, W., Binley, A., Labrecque, D., Roelant, D., 1996. Detection of leaks in underground storage tanks using electrical resistance methods. J. Environ. Eng. Geophys. 1:189–203. https://doi.org/10.4133/JEEG1.3.189.
Ramirez, A.L., Nitao, J.J., Hanley, W.G., Aines, R., Glaser, R.E., Sengupta, S.K., Dyer, K.M., Hickling, T.L., Daily, W.D., 2005. Stochastic inversion of electrical resistivity changes using a Markov Chain Monte Carlo approach. J. Geophys. Res. 110, B02101. https://doi.org/10.1029/2004JB003449.
Revil, A., Johnson, T.C., Finizola, A., 2010. Three-dimensional resistivity tomography of Vulcan's forge, Vulcano Island, southern Italy. Geophys. Res. Lett. 37:1–5. https://doi.org/10.1029/2010GL043983.
Robert, T., Dassargues, A., Brouyère, S., Kaufmann, O., Hallet, V., Nguyen, F., 2011. Assessing the contribution of electrical resistivity tomography (ERT) and self-potential (SP) methods for a water well drilling program in fractured/karstified limestones. J. Appl. Geophys. 75:42–53. https://doi.org/10.1016/j.jappgeo.2011.06.008.
Robinson, J., Johnson, T., Slater, L., 2015. Challenges and opportunities for fractured rock imaging using 3D cross borehole electrical resistivity. Geophysics 80:E49–E61. https://doi.org/10.1190/GEO2014-0138.1.
Rubin, Y., Hubbard, S.S. (Eds.), 2005. Hydrogeophysics. Springer.
Schmidt-Hattenberger, C., Bergmann, P., Labitzke, T., Wagner, F., Rippe, D., 2016. Permanent crosshole electrical resistivity tomography (ERT) as an established method for the long-term CO2 monitoring at the Ketzin pilot site. Int. J. Greenhouse Gas Control 52:432–448. https://doi.org/10.1016/j.ijggc.2016.07.024.
Singha, K., Gorelick, S.M., 2006. Effects of spatially variable resolution on field-scale estimates of tracer concentration from electrical inversions using Archie's law. Geophysics 71:G83. https://doi.org/10.1190/1.2194900.
Slater, L., Binley, A., 2003. Evaluation of permeable reactive barrier (PRB) integrity using electrical imaging methods. Geophysics 68:911–921. https://doi.org/10.1190/1.1581043.
Slater, L., Binley, A., 2006. Synthetic and field-based electrical imaging of a zerovalent iron barrier: Implications for monitoring long-term barrier performance. Geophysics 71:B129–B137. https://doi.org/10.1190/1.2235931.
Slater, L., Binley, A.M., Daily, W., Johnson, R., 2000. Cross-hole electrical imaging of a controlled saline tracer injection. J. Appl. Geophys. 44:85–102. https://doi.org/10.1016/S0926-9851(00)00002-1.
Slater, L.D., Ntarlagiannis, D., Day-Lewis, F.D., Mwakanyamale, K., Versteeg, R.J., Ward, A., Strickland, C., Johnson, C.D., Lane, J.W., 2010. Use of electrical imaging and distributed temperature sensing methods to characterize surface water-groundwater exchange regulating uranium transport at the Hanford 300 Area, Washington. Water Resour. Res. 46:1–13. https://doi.org/10.1029/2010WR009110.
Uhlemann, S., Wilkinson, P., Chambers, J., Maurer, H., Merritt, A., Gunn, D., Meldrum, P., 2015. Interpolation of landslide movements to improve the accuracy of 4D geoelectrical monitoring. J. Appl. Geophys. 121:93–105. https://doi.org/10.1016/j.jappgeo.2015.07.003.
Uhlemann, S.S., Sorensen, J.P.R., House, A.R., Wilkinson, P.B., Roberts, C., Gooddy, D.C., 2016. Integrated time-lapse geoelectrical imaging of wetland hydrological processes. Water Resour. Res. 52:1607–1625. https://doi.org/10.1002/2015WR017932.
Vereecken, H., Binley, A., Cassiani, G., Revil, A., Titov, K. (Eds.), 2006. Applied hydrogeophysics. NATO Science Series. Springer, Netherlands, Dordrecht https://doi.org/10.1007/978-1-4020-4912-5.
Wagner, F.M., Möller, M., Schmidt-Hattenberger, C., Kempka, T., Maurer, H., 2013. Monitoring freshwater salinization in analog transport models by time-lapse electrical resistivity tomography. J. Appl. Geophys. 89:84–95. https://doi.org/10.1016/j.jappgeo.2012.11.013.
Wallin, E.L., Johnson, T.C., Greenwood, W.J., Zachara, J.M., 2013. Imaging high stage river-water intrusion into a contaminated aquifer along a major river corridor using 2-D time-lapse surface electrical resistivity tomography. Water Resour. Res. 49:1693–1708. https://doi.org/10.1002/wrcr.20119.
Wehrer, M., Slater, L.D., 2014. Characterization of water content dynamics and tracer breakthrough by electrical resistivity tomography (ERT) under transient unsaturated conditions. Water Resour. Res. 51. https://doi.org/10.1002/2014WR016131.
West, B., Welch, K.E., Galecki, A.T., 2007. Linear Mixed Models: A Practical Guide Using Statistical Software. 2nd ed. Chapman and Hall/CRC.
Wilkinson, P.B., Chambers, J.E., Lelliott, M., Wealthall, G.P., Ogilvy, R.D., 2008. Extreme sensitivity of crosshole electrical resistivity tomography measurements to geometric errors. Geophys. J. Int. 173:49–62. https://doi.org/10.1111/j.1365-246X.2008.03725.x.
Wilkinson, P.B., Chambers, J.E., Meldrum, P.I., Gunn, D.A., Ogilvy, R.D., Kuras, O., 2010a. Predicting the movements of permanently installed electrodes on an active landslide using time-lapse geoelectrical resistivity data only. Geophys. J. Int. 183:543–556. https://doi.org/10.1111/j.1365-246X.2010.04760.x.
Wilkinson, P.B., Meldrum, P.I., Kuras, O., Chambers, J.E., Holyoake, S.J., Ogilvy, R.D., 2010b. High-resolution electrical resistivity tomography monitoring of a tracer test in a confined aquifer. J. Appl. Geophys. 70:268–276. https://doi.org/10.1016/j.jappgeo.2009.08.001.
Wilkinson, P.B., Loke, M.H., Meldrum, P.I., Chambers, J.E., Kuras, O., Gunn, D.A., Ogilvy, R.D., 2012. Practical aspects of applied optimized survey design for electrical resistivity tomography. Geophys. J. Int. 189:428–440. https://doi.org/10.1111/j.1365-246X.2012.05372.x.
Wilkinson, P., Chambers, J., Uhlemann, S., Meldrum, P., Smith, A., Dixon, N., Loke, M.H., 2016. Reconstruction of landslide movements by inversion of 4-D electrical resistivity tomography monitoring data. Geophys. Res. Lett. 43:1166–1174. https://doi.org/10.1002/2015GL067494.
Williams, M.R., Buda, A.R., Singha, K., Folmar, G.J., Elliott, H.A., Schmidt, J.P., 2016. Imaging hydrological processes in headwater riparian seeps with time-lapse electrical resistivity. Groundwater:1–13 https://doi.org/10.1111/gwat.12461.
Yang, X., Chen, X., Carrigan, C.R., Ramirez, A.L., 2014. Uncertainty quantification of CO2 saturation estimated from electrical resistance tomography data at the Cranfield site. Int. J. Greenhouse Gas Control 27:59–68. https://doi.org/10.1016/j.ijggc.2014.05.006.
Yeh, T.-C.J., Zhu, J., Englert, A., Guzman, A., Flaherty, S., 2006. A successive linear estimator for electrical resistivity tomography. In: Vereecken, H., Binley, A.M., Cassini, G., Revil, A., Titov, K. (Eds.), Applied Hydrogeophysics. Springer, Netherlands:pp. 45–74 https://doi.org/10.1007/978-1-4020-4912-5_3.
Zhou, B., Dahlin, T., 2003. Properties and effects of measurement errors on. Near Surf. Geophys. 105–117.