RMS – Rather Meaningless Simplification?

Grayver, A.V., Tietze, K and Ritter, O.
Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences
{agrayver | ktietze | oritter}@gfz-potsdam.de

Abstract

For a long time the root mean square (RMS) error has been used in the EM community:

- to characterize data fit for a particular model;
- as a criterion to compare several models obtained from inversion.

The RMS error appears to be a natural choice since we usually tackle inverse problems in a least-squares sense. Over the years, RMS became a customary criterion and gained ultimate significance. However, on the hunt for low RMS values, one often needs to introduce subjectivity by arbitrarily adjusting error floors or masking "bad" data without referring to the assumptions behind RMS. In this contribution, we revisit basic assumptions behind RMS, demonstrate its deficiency and propose alternative ways, which may provide more insight into our data and allow a more comprehensive assessment of the quality of the modelling result/resistivity model.

Definition of RMS

Let ϕ_d denote the data misfit calculated as

$$\phi_d = \sum_i^N r_i^2, \tag{1}$$

where normalized residuals *r* are:

$$r_i = \frac{d_i^{pred} - d_i^{obs}}{e_i}; \quad i = 1 \dots N. \tag{2}$$

Vectors $\boldsymbol{d}^{pred}, \boldsymbol{d}^{obs} \in \mathbb{R}^N$ represent predicted and observed data, respectively, and $\boldsymbol{e}$ is the vector of the estimated data errors (e.g., standard deviations). Assuming that the residuals in equation (2) are zero-mean normal independent random variables, ϕ_d is an asymptotically χ2-distributed random variable with expected value $E[\phi_d]$ = N. This immediately brings us to the well-known expression

$$\text{RMS} = \sqrt{\frac{\phi_d}{N}}. \tag{3}$$

If the majority of the predicted data are within the data error bounds, RMS ≈ 1.

Deficiency of RMS

The RMS error is based on several assumptions which can be violated in real-world problems: we assume that (i) our error estimations in (2) are correct and (ii) the residuals are zero-mean Gaussian independent random variables.

Ideally, one should be able to obtain good estimates for the standard deviations from data processing. In practice, however, standard deviations obtained from data processing have other shortcomings. For instance, a large number of observations are usually available for

high frequency data, but very few for low frequency measurements. Thus, data variances usually increase with period length. Data points can be strongly biased but with very small uncertainty. Normally noise affects all input- and output channels thereby violating underlying assumptions for our data processing. More uniform data quality can be achieved with artificial or arbitrary error floors, which are often used in real data inversion. Often, the choice of error floors is driven by the desire to achieve an RMS of ~ 1 for the recovered model.

The second point, namely an assumed Gaussian distribution of the residuals, is trickier. If the residuals originate mainly from noise in observed data, this noise has to be normally distributed with zero mean in order to make the RMS measure optimal. Obviously, in reality residuals do not contain idealized cultural noise and measurement error. In addition the modelling results contain numerical errors e.g. related to a particular discretization and modelling.

Furthermore, being non-robust RMS is sensitive to outliers. Few outliers may result in a violation of the normal distribution assumption.

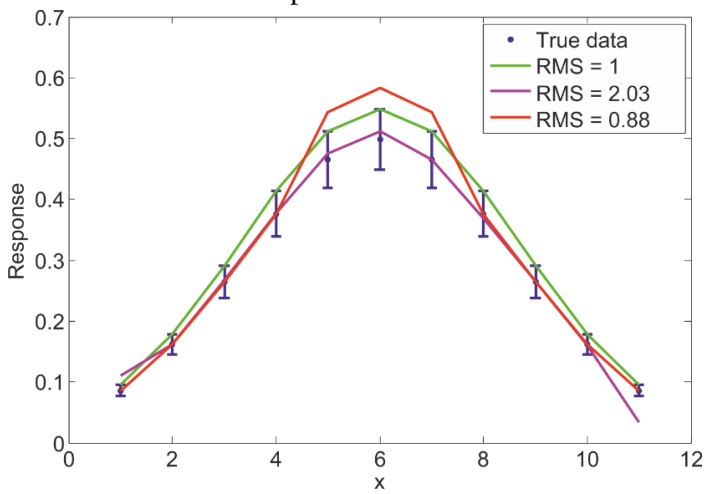

Figure 1: A number of artificial cases to demonstrate that RMS values can be misleading. Blue dots are the true response of a model with error bars corresponding to 10% of the original data amplitudes. The green line is a "true" response systematically shifted to fit the upper limit of the error bars, resulting in an RMS of one. The magenta line shows the ``true'' data perturbed with 1% normal noise and two outliers added at the ends of the curve, which results in an RMS of 2.03. The red line shows ``true'' data with three data points outside the error bars with a total RMS of 0.88.

Comparing models (MT example)

The RMS is often used as a criterion to rank a series of inversion models (when dealing with real data). However, being a single number the RMS cannot reliably describe a complicated model nor does it give any insight into structural differences between models. Figure 2 shows results for three MT 3D inversions runs for the same synthetic data set, but using different starting/prior models. Visual inspection suggests that the first inversion using a 10 Ωm starting model recovers the underlying resistivity structure best (upper panel). If we judge the results only by RMS the model obtained with a starting model of 20 Ωm should be preferred since it results in the lowest RMS; second would be the result using a 50 Ωm half-space. However, in both results the prominent conductive anomaly is missing. This mystery is

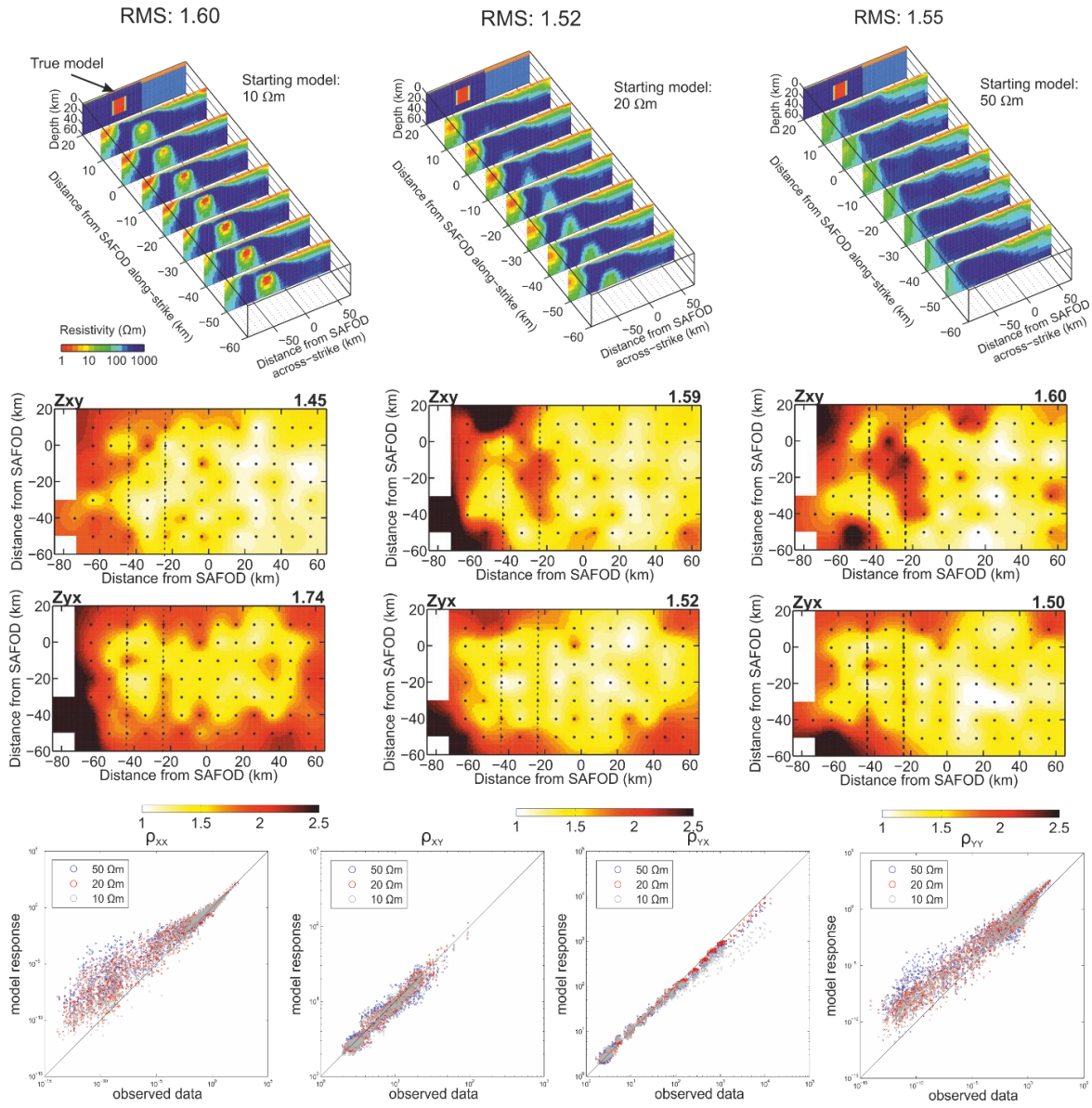

Figure 2: ***Top panel****: Comparison of 3D MT inversion results for three different starting model resistivities (10, 20, 50 Ω*m*); data were generated for a synthetic model shown as slice in the back. Note, the resistivity structure is predominantly 2D with the coarse 3D bathymetry of the Pacific Ocean added; the coastline is located at the left edge of the profiles. The conductive anomaly at depth disappears only for starting resistivities above 10 Ω*m*.*
Middle panel*: Systematic analysis of data misfits in the frequency-space domain of the inversion results reveals large-scale structural deviations of 3D inversion results, which are not reflected in the overall RMS.*
Bottom panel*: Cross-plot of observed and modelled apparent resistivities for all four impedance components. In the ideal case, points would lie on the black line. Scattering around the black line is lowest for model responses obtained with 10 Ω*m* starting model, in particular for the xy-component. The plots also show that the very high apparent resistivities (> 1000 Ω*m*) of the yx-component are difficult to recover for all shown inversions.*

solved when the RMS distribution over the array is assessed in more detail (middle panel of Figgure 2) . For starting models of 20 and 50 Ωm we observe systematically higher RMS values for the Zxy component – the component which is most sensitive to the conductive anomaly – at stations above the conductive anomaly; in return high resistivities of the Zyx component close to the ocean (left edge of the profiles) are better fit. For both components RMS values of the 10 Ωm inversion result show a more uniform distribution of the data misfit. The choice of the model obtained with a starting model of 10 Ωm can also be justified when comparing scatter plots of apparent resistivities for all impedance tensor components. They reveal that the 10 Ωm model (shown in grey) follows the observed data better and has smallest variance.

Estimating data fit (CSEM example)

Usually, we argue that a high RMS value means poor data fit and we infer that the inversion result is not reliable. Given the possibility to choose error floors arbitrarily in (2) one can artificially reduce RMS values by taking/applying larger error floors. Figure 3 shows a section along a profile from 3D inversion models of real CSEM data, and histograms of initial and final normalized residuals for a number of inversion runs with increasing error floors. All inversions were stopped after 30 iterations. All models are very similar. The final RMS shown above each model indicates that increasing the error floor results in lower RMS values. The histograms show that the distribution of residuals is not normal. Increasing error floors shrinks the starting residuals, but has little to do with actual data fit. In addition, red numbers in brackets indicate RMS values calculated with 10% of the most poorly fit data dropped. Without these data values, the RMS values are reduced by more than 40% which indicates the non-robust nature of the root mean square.

Conclusions

The aspects discussed in the previous sections suggest the following conclusions:

- If residuals do not obey a standard normal distribution, RMS is not an optimal measure. It can still be used, but does not need to approach a value of 1 even for geologically justified models.
- The choice of error floors and manual selection of “good” data are subjective and we should try to avoid such assumptions.
- We recommend using more elaborate tools instead of a single RMS number such as investigation of the RMS distribution in the frequency-space domain; analysis of histograms of normalized residuals or scatter plots of modeled vs. observed data.

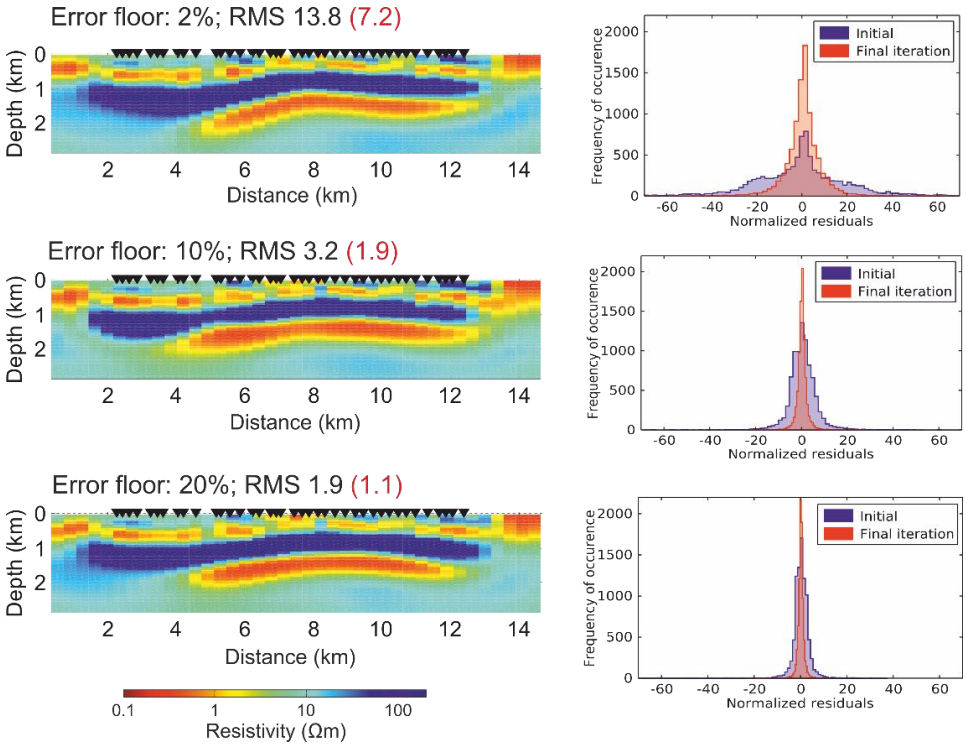

Figure 3. Conductivity sections along the profile line (left) and histograms of initial and final normalized residuals (right) for predefined error floors of (a) 2, (b) 10 and (c) 20\%. The final RMS is given in the title of each plot. Red numbers in brackets indicate RMS values calculated with 10\% of the most poorly fit data values dropped. All three inversions were stopped after 30 iterations.

References

Grayver, A. V., 2013, Three-dimensional controlled-source electromagnetic inversion using modern computational concepts, PhD Thesis, FU Berlin.

Tietze, K., 2012, Investigating the electrical conductivity structure of the San Andreas fault system in the Parkfield-Cholame region, central California, with 3D magnetotelluric inversion, PhD Thesis, FU Berlin.