Geophysical Research Abstracts
Vol. 21, EGU2019-12612, 2019
EGU General Assembly 2019

Another look at the treatment of data uncertainty in the presence of outliers

Frederik Tilmann (1,2)

(1) Deutsches GeoForschungsZentrum, GFZ Potsdam, Seismology, Potsdam, Germany (tilmann@gfz-potsdam.de), (2) Institut für Geologische Wissenschaften, Freie Universität Berlin, Berlin, Germany

The data uncertainty in many geophysical inference problems is poorly quantified. In probabilistic Bayesian inversions, the data uncertainty is a crucial parameter for quantifying the nonlinear uncertainties and correlations of the model parameters. Therefore, it is common practice to allow the data uncertainty itself to be a parameter to be determined. In a Markov chain Monte Carlo approach, some uncertainty parameter is varied probabilistically in the same way as the physical model parameters through subsequent steps. Although in principle any arbitrary uncertainty distribution can be assumed, in the vast majority of published studies Gaussian distributions are assumed for the data error, those standard deviation is the unknown parameter to be estimated.However, in this special case, a simple analytical integration is sufficient to marginalise out this uncertainty parameter, reducing the complexity of the model space without compromising the accuracy of the posterior model probability distribution. This approach works equally well if there are different data classes with unknown errors, without having to introduce a priori weighting between them.

On the other hand, it is well known that the distribution of geophysical measurement errors, although superficially similar to a Gaussian distribution, contain far more frequent samples along the tail of the distribution, commonly described as outliers. In linearised inversions these are often removed in subsequent iterations based on some threshold criterion, but in Markov chain Monte Carlo inversions, this is not possible and introduces ad-hoc parameters, whose effect on the model uncertainty is not easily quantified. The presentation will introduce a straightforward approach to deal with outliers in a Bayesian context that is flexible in that no rigid assignments are made, and the effect of arbitrary parameters is (largely) avoided. The techniques will be illustrated with synthetic and observational case studies of two-dimensional tomographic problems, but should be applicable to all geophysical inference problems. By focussing on the likelihood minimisation, it is possible to devise a weighting scheme for iterative cost-function gradient-based solutions of nonlinear geophysical inference problems.