Probabilistic source inversion of static displacement and seismic waveform data using neural networks

Paul Käufl, Andrew Valentine & Jeannot Trampert
Dep. of Earth Sciences, Utrecht University, The Netherlands
Contact: p.j.kaufl@uu.nl

Universiteit Utrecht

NWO
Netherlands Organisation for Scientific Research

Introduction

We use artificial neural networks to find posterior distributions on point source parameters from seismic waveform data in a Bayesian framework.

As a first step we extract information from static displacement measurements, as provided by GPS instruments or INSAR images. We aim to investigate if static displacements, measured by GPS instruments, can contribute information to better constrain early source estimates.

Our method might be applicable in an earthquake early warning (EEW) context, where it could provide source parameter estimates together with realistic uncertainty bounds rapidly. Assuming that a trained network is available an inversion takes only a few ms on a desktop computer.

Methodology

The solution to the general inverse problem can be written as a posterior probability distribution (Tarantola, 2005)

$$\sigma_M(\mathbf{m}|\mathbf{d}_{obs}) = k\rho(\mathbf{m})L(\mathbf{m}), \qquad (1)$$

with the prior distribution $\rho_M(\mathbf{m})$ and likelihood function $L(\mathbf{m})$ and $\mathbf{m} \in M$ denoting a member of the model space. We directly model marginal distributions

$$P(m_i) = \int \sigma_M(\mathbf{m}) \prod_{m_k \neq m_i} dm_k \qquad (2)$$

using Mixture Density Networks (MDNs).

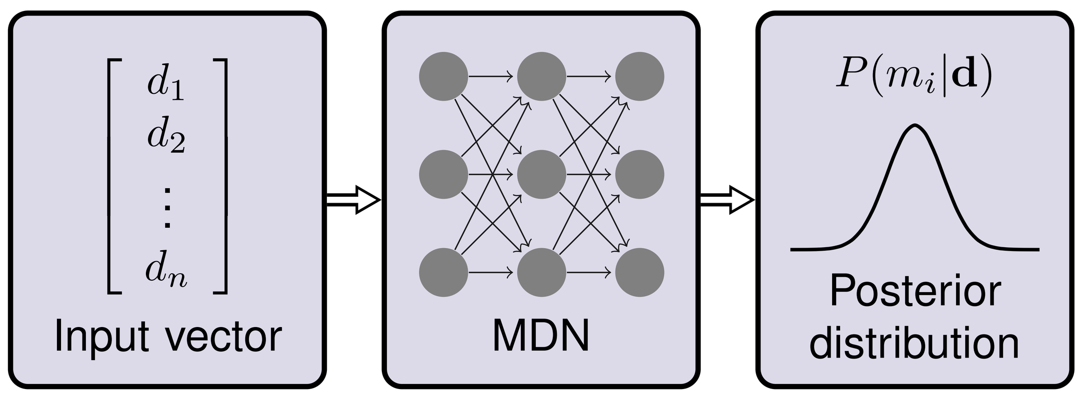

Figure 1: Mixture density networks approximate the inverse mapping and yield posterior distributions.

A mixture density network approximates the inverse mapping from data space D into model space M using an adaptive, weighted sum of non-linear basis functions. The network weights are optimized during the training stage, which requires a large number of example input-output pairs. See also the info-box *Mixture Density Networks*.

Data vectors $\mathbf{d}$ are formed by concatenating the x- and y- components of the static displacement at different stations. We add white Gaussian noise with a standard deviation of $\sigma = 1$ mm to simulate the presence of ambient noise, measurement- and modeling-errors.

Results, Uncertainties and Trade-offs

We trained networks on the 6 moment tensor components, centroid location, strike, dip, rake and scalar moment M_0. The performance of the trained networks is assessed by presenting an independent set of test examples to the network. Figure 3 shows test set predictions for a virtual network of 80 stations (black triangles in Fig. 2).

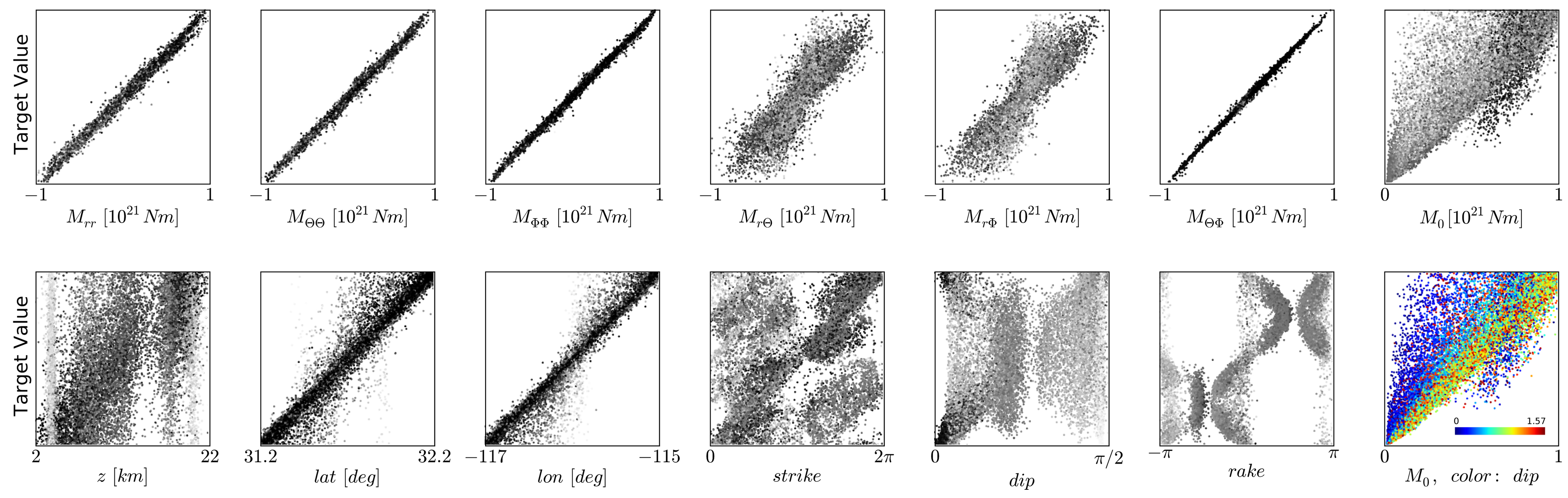

Figure 3: Every dot corresponds to an example of the test set. Horizontal axis: position of the maximum posterior Gaussian kernel, vertical axis: Desired target value. color axis: width of the maximum posterior kernel. Ideally the plots would show a straight diagonal. The plot on the bottom right indicates that M_0 for extreme dip angles is systematically underestimated.

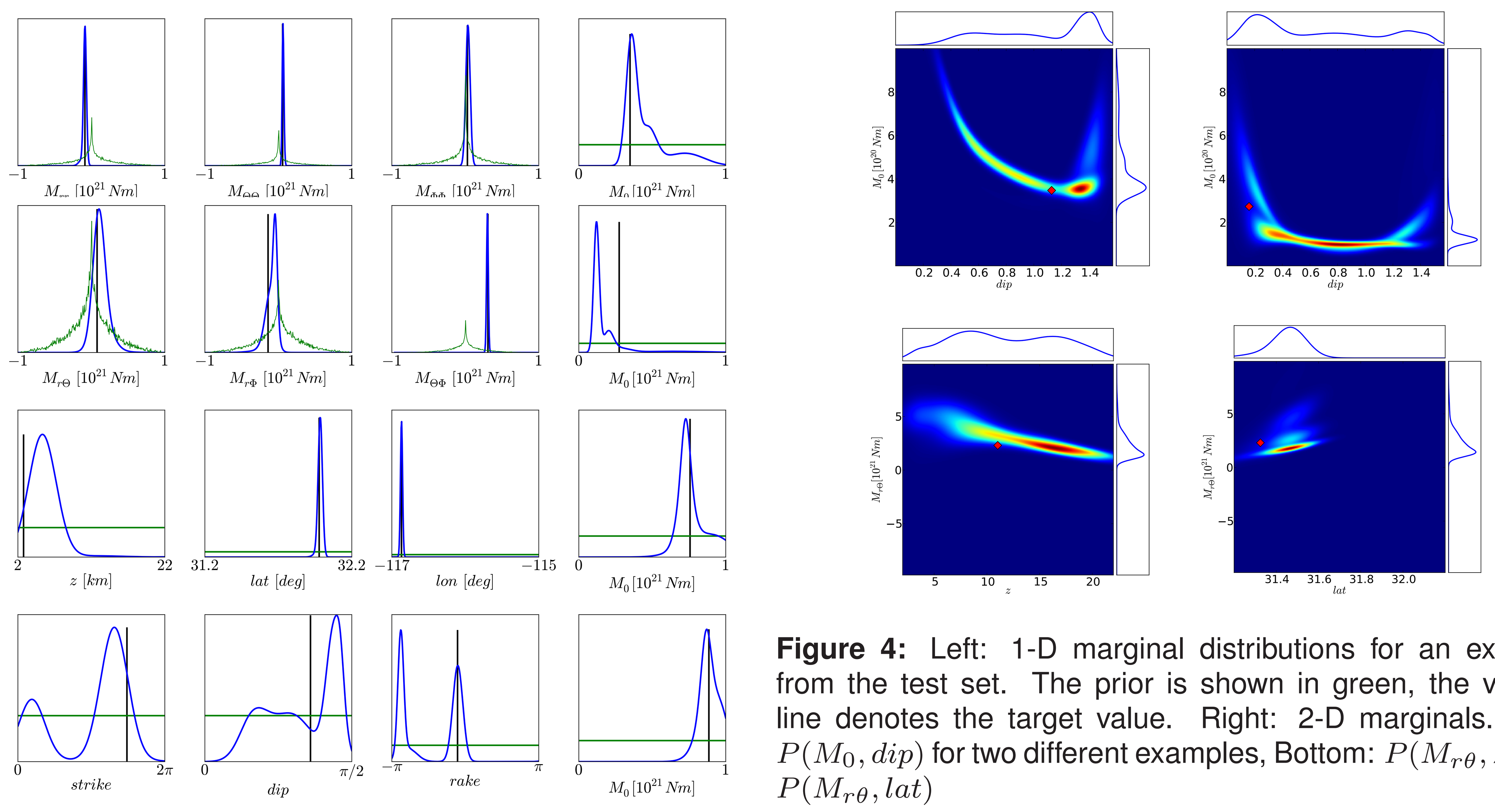

Figure 4: Left: 1-D marginal distributions for an example from the test set. The prior is shown in green, the vertical line denotes the target value. Right: 2-D marginals. Top: $P(M_0, dip)$ for two different examples, Bottom: $P(M_{r\theta}, z)$ and $P(M_{r\theta}, lat)$

Static Displacement Measurements

Static displacements are the zero-frequency-limit solution to the seismic wave equation, and as such only detectable by long-period displacement instruments such as GPS stations. The use of GPS data for moment-tensor point source inversions is subject of ongoing research (e.g. O'Toole et al., 2012a; Melgar et al., 2012) and may have potential for earthquake early warning applications.

We calculate coseismic displacements in a layered, isotropic, elastic half-space using a propagator matrix method developed by O'Toole and Woodhouse (2011).

Prior Source Distribution

We draw 50.000 double-couple sources of random magnitude and orientation from the prior distribution (Figure 2).

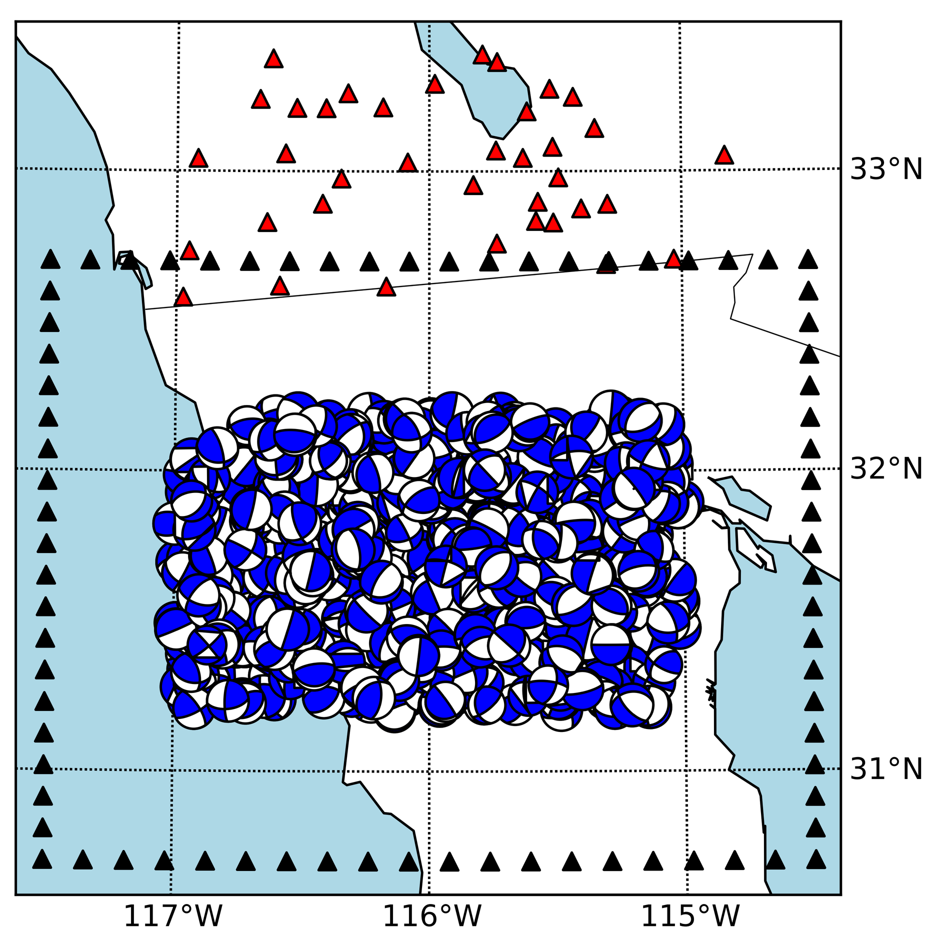

Figure 2: A set of random double-couple sources drawn from the prior. Triangles indicate positions of virtual (black) and California Real Time Network (red) receivers.

Outlook & Conclusion

A probabilistic neural network inversion can act as a powerful tool to analyze complex non-linear inverse mappings, including realistic uncertainty estimates and trade-offs. Computational demands are low, once a trained network is available, which makes the method suitable for EEW purposes.

Our preliminary results suggest, that static displacement data alone does not constrain source depth and magnitude better than seismic data. There might be potential, however, in joint inversions of seismic and co-seismic near-field data, which we can readily incorporate into a neural network inversion due to the flexible treatment of input data.

A More Realistic Example - The 2010 El Mayor-Cucapah Event

Networks are trained using 37 stations of the CRTN GPS network (red triangles in Figure 2). Figure 11 shows the test set performance,

Subsequently we presented real observations for the 2010, M_w 7.2 El Mayor-Cucapah event to the trained network.

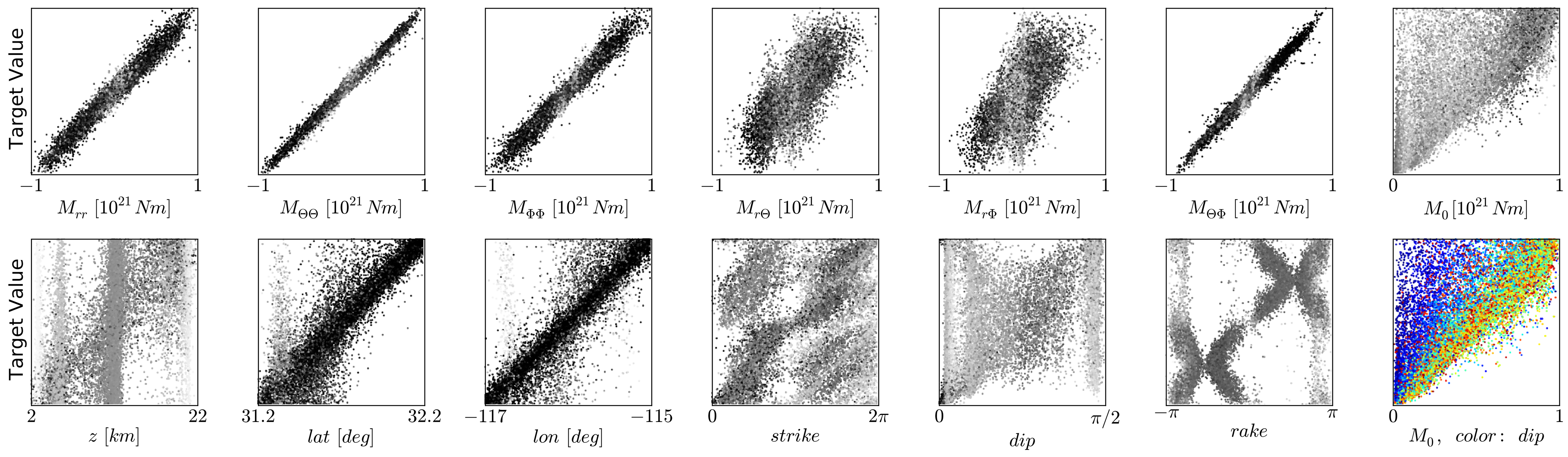

Figure 5: As opposed to Figure 3 the scatter is slightly broader, due to the uneven station distribution. Also note that sources that are further away from the GPS network (smaller latitude) are predicted less well.

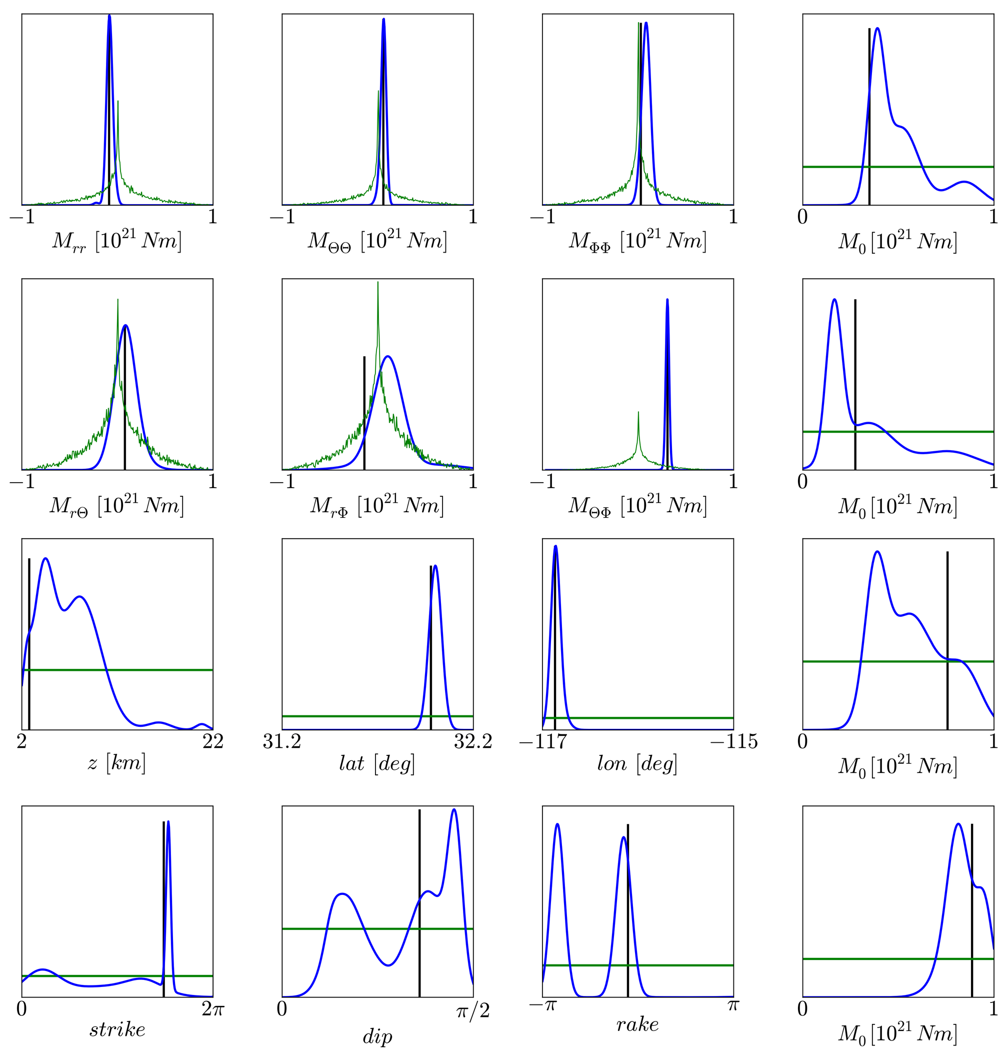

Figure 6: 1-D marginal distributions for the same test set example as in Figure 4. Distributions are in general broader and less pronounced due to the uneven station distribution.

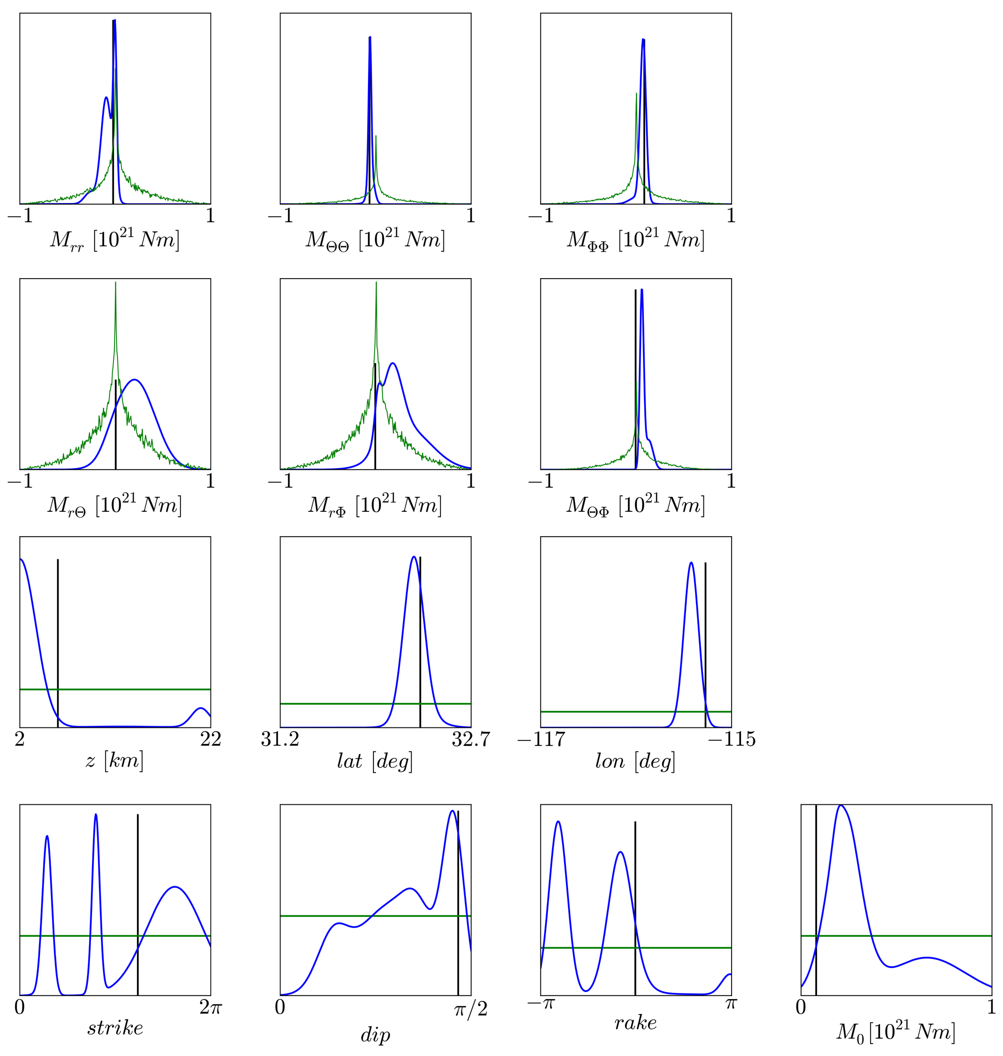

Figure 7: Preliminary results for the 2010 M_w 7.2 El Mayor-Cucapah event. Black vertical lines denote the position of the CMT catalogue solution.

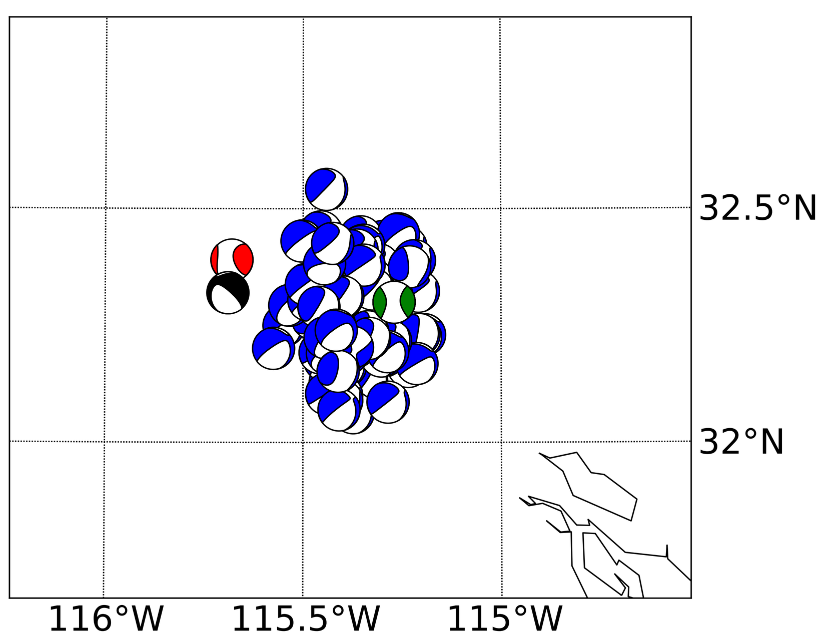

Figure 8: 100 Samples taken from the posterior distribution under the assumption of independent model parameters. Other solutions plotted are gCMT (green), a solution by Melgar et al. (2012) (red) and by O'Toole et al. (2012b) (black).

Mixture Density Networks

Neural networks are function approximators able to learn by example a general class of mappings to arbitrary accuracy. They consist of multiple layers of connected computational units (neurons), which apply a non-linear activation function $g(\cdot)$ to their input.

Neuron-inputs are weighted sums over the inbound connections, such that the output of a two-layer, **feed-forward network** reads

$$y_k(\mathbf{x}) = \sum_{h=1}^{H} w_{hk} g\left(\sum_{i=1}^{I} w_{ih} x_i\right), \tag{3}$$

where H and I denote the number of hidden and input nodes, respectively, and $\mathbf{w}$ is a vector of network weights.

The connection weights $\mathbf{w}$ of a given network are optimized during the **training stage**. A set of examples $\{\mathbf{x}_n\}$ for which the desired outputs (targets) $\{\mathbf{t}_n\}$ are known is presented and network weights are updated iteratively to minimize an error function. Weight updates are determined using the L-BFGS quasi-Newton method (Liu and Nocedal, 1989), while partial derivatives are efficiently evaluated using error backpropagation.

Mixture-Density-Networks (MDNs) output the parameters of a Gaussian mixture model, thereby directly parametrizing posterior marginal probability distributions. This is achieved by minimizing the following error function based on the maximum likelihood principle

$$E = -\sum_n \ln\left\{\sum_{j=1}^{M} \alpha_j(\mathbf{x}^n)\Phi_j(\mathbf{t}^n|\mathbf{x}^n)\right\}, \tag{4}$$

where α_j are mixing coefficients for the Gaussian kernels

$$\Phi_j(\mathbf{t}^n|\mathbf{x}^n) \propto exp\left\{-\frac{||\mathbf{t}-\mu_j(\mathbf{x})||^2}{2\sigma_j^2}\right\} \tag{5}$$

with means μ_j and variances σ_j^2.

An MDN can thus approximate any arbitrary conditional pdf as a linear combination of Gaussian kernels (Bishop, 1995).

Preview: Inversion of Tele-Seismic Body Waves

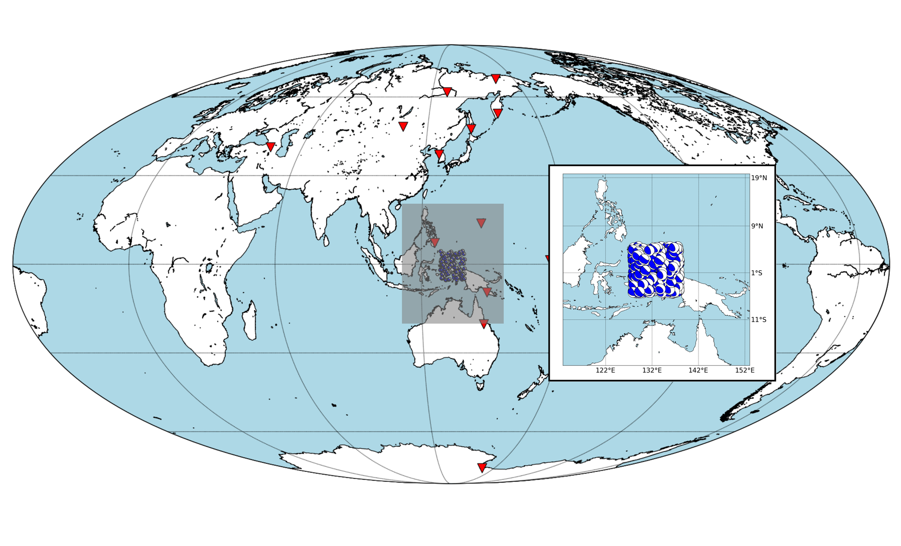

Figure 9: The box shows the prior source distribution, that has been centered around a 2009 Papua New Guinea event from the global CMT catalogue. Synthetic seismograms at the stations denoted by red triangles were generated using normal mode summation in a spherically symmetric Earth model.

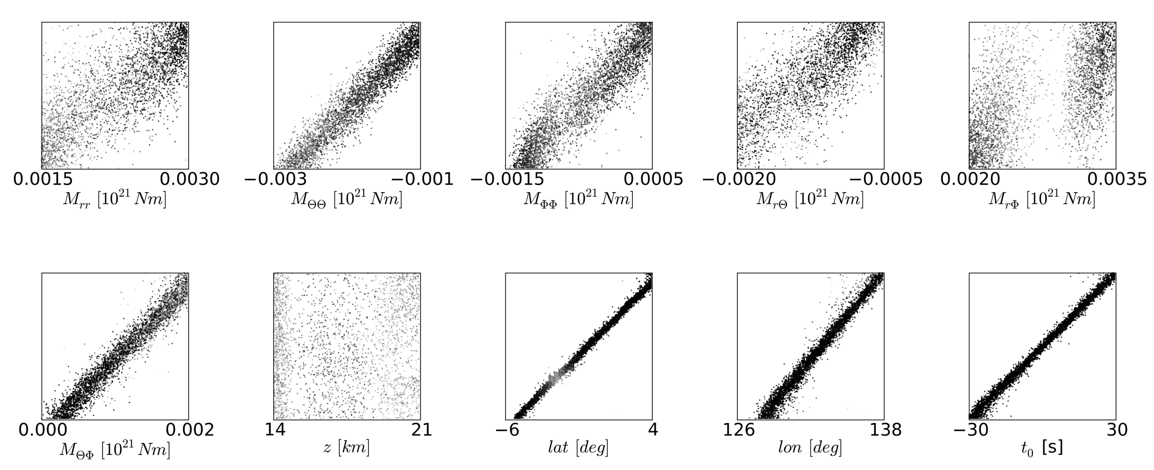

Figure 10: Test set performance for a network trained on tele-seismic body-waves. Network input vectors $\mathbf{d}$ are formed by concatenating long-period vertical component seismograms at several stations.

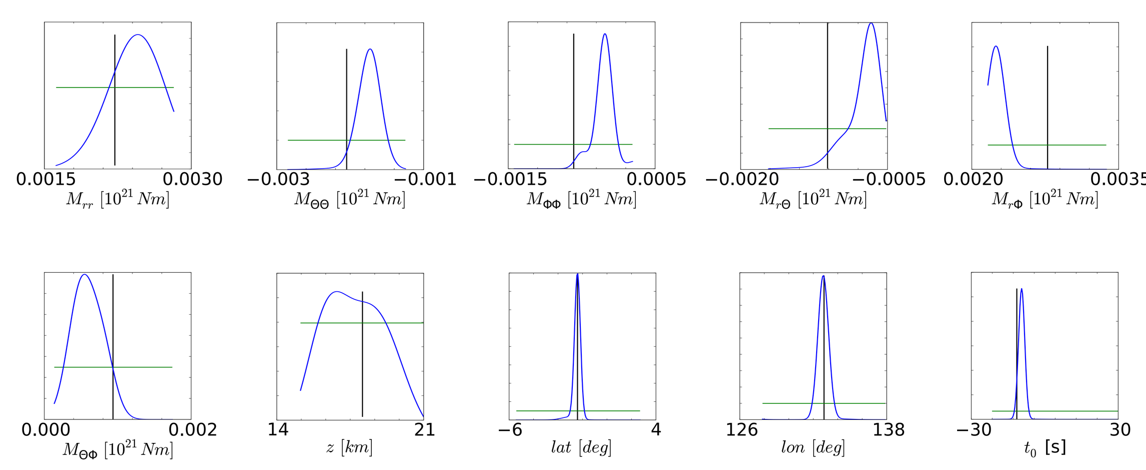

Figure 11: Preliminary results for the 2009 Papua New Guinea event. The black vertical line denotes the catalogue solution, the prior distributions are shown in green.

References

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Volume 92. Oxford University Press.

Liu, D. C. and J. Nocedal (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming 45*(1-3), 503–528.

Melgar, D., Y. Bock, and B. W. Crowell (2012, February). Real-time centroid moment tensor determination for large earthquakes from local and regional displacement records. *Geophysical Journal International 188*(2), 703–718.

O'Toole, T. B., A. P. Valentine, and J. H. Woodhouse (2012a). Centroid-moment tensor inversions using high-rate GPS waveforms. *Geophysical Journal International 191*(1), 257–270.

O'Toole, T. B., A. P. Valentine, and J. H. Woodhouse (2012b). Earthquake source parameters from GPS-measured static displacements with potential for real-time application. *Geophysical Research Letters (in press)*.

O'Toole, T. B. and J. H. Woodhouse (2011). Numerically stable computation of complete synthetic seismograms including the static displacement in plane layered media. *Geophysical Journal International 187*(3), 1516–1536.

Tarantola, A. (2005). *Inverse Problem Theory*, Volume 130. SIAM.