IMPROVING MAGNETOTELLURIC IMPEDANCE TENSOR ESTIMATES
BY SELF-ORGANIZING MAPS

R. Carbonari¹, R. Di Maio¹, E. Piegari¹, L. D’Auria², A. Esposito³, Z. Petrillo³

¹ Dipartimento di Scienze della Terra, dell’Ambiente e delle Risorse, Università di Napoli Federico II, Naples, Italy
² Instituto Volcanológico de Canarias, San Cristobál de la Laguna, Spain
³ Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli Osservatorio Vesuviano, Naples, Italy

Introduction. In the last decades, the magnetotelluric (MT) method has been proved to be a useful geophysical tool in different contexts, from geothermal reservoir characterization to crustal structures studies (Simpson and Bahr, 2005). Nevertheless, the MT method is very sensitive to the presence of noise. Indeed, as the method is based on the measurement of both the electric and magnetic components of the natural electromagnetic (EM) field, it fails when these components are affected by noises of different nature. In particular, in industrialized and urbanized context, the MT time series may be strongly affected by man-made noise and, as a consequence, the impedance tensor estimates, given by the ratio between the electric and magnetic components of the MT field, could be unreliable. To improve the reliability of these estimates, most of the proposed approaches rely on the robust evaluation of the impedance tensor (e.g., Sutarno and Vozoff, 1989; Egbert, 1997) as well as on the use of remote reference MT stations (Gamble et al., 1979) or on the combination of both approaches (Jones et al., 1989; Egbert, 1997). However, these methods are not always effective. The robust methods fail when most of the data are affected by noise, giving as result a biased impedance tensor, while the remote reference approach is ineffective when the noise is correlated between reference and local MT station (Ritter et al., 1998). In recent years, alternative procedures have been proposed to obtain reliable estimates of the MT impedance tensor. In particular, Weckmann et al. (2005) proposed a pre-selection scheme of the Fourier transform coefficients followed by a robust processing, while Escalas et al. (2013) and Carbonari et al. (2017) proposed a polarization analysis of the MT signals in the wavelet transform domain. The use of the wavelet transform is motivated by the resolution that it provides in both time and frequency domain, thus allowing to deal with transient components of the MT signal, as the man-made noise usually occurs.

In the present work, a different approach based on the use of Discrete Wavelet Transform (DWT) and Self-Organizing Map (SOM) neural network analysis (D’Auria et al., 2015) is proposed for improving the magnetotelluric impedance tensor estimates. The approach has
been tested by changing type, level and window length of the noise affecting MT time series. Furthermore, in order to identify the most reliable apparent resistivity and phase values among the different impedance tensors clusters provided by the SOM analysis for each analyzed period, a selection criterion is provided and tested on synthetic and field MT data.

**SOM clustering technique.** The self-organizing map (SOM) is a neural network widely used for data exploratory analysis (Kohonen, 1998). A SOM consists of a set of nodes (neurons) distributed on a low-dimensional grid, where each node has an associated D-dimensional weight (or prototype) vector, with D the input vectors dimension. A neighborhood relationship links adjacent nodes characterizing the SOM map structure. The latter identifies the local lattice structure and the global map shape. The weight vectors are initialized with random values before the SOM processing that consists of an iterative procedure: at each step, an input vector is compared with all the weight vectors of the nodes by calculating the Euclidean distance between them. The node whose weight vector is closer to the input vector is called Best Matching Unit (BMU) and its weight vector, as well as those of its neighboring nodes, are adjusted to move towards the input vector by using a neighborhood function. The latter is a decreasing function of the distance between BMU and nth node of the grid. This procedure is repeated for all the input vectors to provide a two-dimensional map with new weights for each node. The final result of the SOM analysis is a map in which each node contains a certain number of input vectors with similar features. It is worth noting that, even if SOM clustering has been widely used in the last two decades for different geophysical applications, from meteorology to seismology (e.g., Esposito et al., 2008; Liu and Weisberg, 2011), its application to MT data is an almost unexplored research field, except for the first attempts proposed by D’Auria et al. (2015) and Carbonari et al. (2017).

**Application to MT data.** SOM clustering of MT data is performed by using as input vectors the normalized impedance tensors obtained after a DWT decomposition of the MT signal. In particular, the latter is firstly decomposed through a Discrete Wavelet procedure (D’Auria et al., 2015; Carbonari et al., 2017), then, for each wavelet scale, the impedance tensor is estimated.

![Fig. 1 - Apparent resistivities and phases retrieved from the SOM clustering of the impedance tensors respectively in absence (a) and presence (b) of noise. Green and orange circles indicate apparent resistivity and phase retrieved from, respectively, the $\text{xy}$ and $\text{yx}$ components of the impedance tensor, while the continuous blue line shows the noise-free synthetic MT curve.](image)
for different subsets of the DWT coefficients. The choice of the coefficient subsets is made by the Monte Carlo technique, which operates by selecting, in each wavelet scale \( s \), \( N \) random subsets of \( k \) coefficients from the whole set consisting of \( 2^s \) DWT coefficients. For each set of coefficients, the impedance tensor is then estimated through a least square procedure. Finally, the obtained impedance vectors are normalized using a logistic transformation that scales all possible values between 0 and 1, and then, each normalized impedance tensor is transformed into a vector of eight elements used as input vector for the SOM clustering. Once the clustering procedure is over, for each cluster, the apparent resistivity and phase are evaluated in each wavelet scale by using all the wavelets coefficients that have generated the impedance tensor estimates falling in the cluster. As an example, Fig. 1 shows the results for the apparent resistivity and phase clusters by using clean data (Fig. 1a) and noisy data (Fig. 1b) obtained after the addition of a Gaussian noise to one-fifth of the entire MT time series of the electric component. As it can be seen, in both images, most of the resistivities and phases lie on the synthetic noise-free curve.

Fig. 2 - Apparent resistivity and phase curves for the \( xy \) (a) and \( yx \) (b) components of the impedance tensor. These curves have been obtained by applying the SOM filter procedure on a MT signal affected by a Gaussian white random noise, with a SNR of 0.83 (83%), that has been added to one-fifth of the entire MT time series of the electric component. As it can be seen, in both images, most of the resistivities and phases lie on the synthetic noise-free curve.

**MT data denoising through a selection criterion.** The scattering of the apparent resistivity and phase values in presence of noise (see Fig. 1b) raises the problem of identifying the most reliable apparent resistivity and phase curve, especially when there is no additional information about it, which is the most common case in MT prospecting. This means that a selection criterion able to recognize, for each frequency, the most reliable resistivity and phase values is needed. The MT impedances, or equivalently the apparent resistivity and phase curves, exhibit smooth trends according to the diffusive nature of the MT field (Weidelt, 1972). Thus, basing on the idea to maximize the smoothness of apparent resistivity and phase curves, the proposed selection criterion chooses among all the resistivities (or phases) provided by the SOM clustering for each frequency, the one that minimizes the difference with the apparent
resistivity (or phase) value estimated for the previous period. In other words, for each frequency 
k, the “best” resistivity (or phase) value, \( k \), is the one that satisfies the following relation:

\[
\min | \log(\rho_{i,k-1}) - \log(\rho_{i,k}) | ,
\]

where \( i \) is the index of the \( i \)th resistivity obtained for the specific frequency \( k \).

As an example, Fig. 2 shows the results of the application of the selection criterion to synthetic MT data affected by Gaussian white random noise (SNR=0.83) in a window of length equals to one-fifth of the entire time series length. As it can be seen, most of the resistivities \( \rho_{xy} \) and \( \rho_{yx} \) and corresponding phases, retrieved after the application of the filtering procedure (diamond symbols), fit well with the synthetic noise-free curves (blue straight lines). Similar results have been found by applying noise with a chi-square distribution as well as by varying the noise window length affecting the original synthetic dataset.

Finally, the proposed filtering procedure has been applied to real MT data acquired in the Yellowstone caldera (Yellowstone National Park, Wyoming, United States). In Fig. 3, a comparison between apparent resistivities and phases retrieved with and without the application of the SOM filtering procedure is shown. As it can be seen in Fig. 3a, the curves estimated after the filtering have a generally smooth shape. In particular, the filter clearly improves the quality of the apparent resistivity estimates in the range of periods between 3 s and 20 s. Indeed, in this range, the apparent resistivities obtained without the clustering filter are very scattered and definitely unreliable, because, as mentioned above, the apparent resistivity variations generally show a smooth pattern. The filter application improves also the estimates of the \( yx \) component (Fig. 3b), particularly in the range of periods between 0.05 s and 0.3 s, where the resistivity values obtained without the SOM filter tend to decrease faster than those obtained after filtering. This behaviour is likely due to the presence of noise, as it is suggested by the occurrence of the two large clusters below the red crosses, which indicate the \( \rho_{yx} \) values obtained without the SOM filter. Indeed, these clusters refer to portions of signal that generate apparent resistivities lower than those obtained in the remaining part of the signal.

Fig. 3 - Comparison between the apparent resistivity and phases estimates retrieved with (circles) and without (cross) the application of the clustering procedure for both \( xy \) (a) and \( yx \) (a) components.
References


