**Nonlinear Processes
in Geophysics**

Denoising gravity and geomagnetic signals from Etna volcano (Italy) using multivariate methods

C. Del Negro[1], **F. Greco**[1], **R. Napoli**[1], and **G. Nunnari**[2]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Italy
[2]Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, Università di Catania, Italy

Abstract. Multivariate methods were applied to denoise the gravity and geomagnetic signals continuously recorded by the permanent monitoring networks on the Etna volcano. Gravity and geomagnetic signals observed in volcanic areas are severely influenced by meteorological variables (i.e. pressure, temperature and humidity), whose disturbances can make the detection of volcanic source effects more difficult. For volcano monitoring it is necessary, therefore, to reduce the effects of these perturbations. To date filtering noise is a very complex problem since the spectrum of each noise component has wide intervals of superposition and, some times, traditional filtering techniques provide unsatisfactory results. We propose the application of two different approaches, the adaptive neuro-fuzzy inference system (ANFIS) and the Independent Component Analysis (ICA) to remove noise effects from gravity and geomagnetic time series. Results suggest a good efficiency of the two proposed approaches since they are capable of finding and effectively representing the underlying factors or sources, and allow local features of the signal to be detected.

1 Introduction

Over the last decades, new modern techniques of volcano monitoring have been implemented on Mt. Etna in order to improve the knowledge of eruptive processes. In particular, monitoring involves gravity and magnetic techniques that have provided essential information on the eruption mechanism including magma storage and transport within the volcano edifice (Carbone et al., 2003; Del Negro et al., 2004). However, gravity and geomagnetic time series are severely

influenced by meteorological variables (i.e. pressure, temperature and humidity), whose disturbances can not only make the detection of volcanic source effects more difficult but also may lead to misinterpret data. For all practical purposes, volcano monitoring is concerned with detection of gravity and geomagnetic anomalies attributable to the dynamics of a volcano and removal of variations with no geophysical significance.

Temporal gravity changes in volcanic areas are related to sub-surface mass-redistributions and/or surface elevation changes in response to magmatic activity, and their amplitude, wavelength and duration depend on several parameters such as the size, depth and evolution rate of the sources. The expected gravity changes due to volcanic sources range in amplitude between 10 and 1000 μGal ($1\,\mu$Gal$=10^{-8}\,$ms^{-2}) with a spectrum varying from 1–10 s to more than 1 yr. To isolate gravity residuals, due to sub-surface mass-redistribution it is necessary to remove the effects of non volcanic sources (i.e. luni-solar gravitational effects, atmospheric contribution, instrumental drift, ground tilt etc.). Unfortunately, especially when used in the adverse environmental conditions often encountered at active volcanic areas, the behavior of spring gravity meters (the most commonly utilized instruments for microgravity studies) have proven to be severely influenced by meteorological variables (i.e. pressure, temperature and humidity; Andò and Carbone, 2001, 2004, 2006; Carbone et al., 2003; El Wahabi et al., 1997; Warburton and Goodking, 1977). In particular, El Wahabi et al. (1997) showed that, over a yearly period, temperature changes can produce up to 1 mGal (1 mGal$=10^{-5}\,$ms^{-2}) instrumental effect. An admittance up to 0.2 mGal/°C, over changes in periods longer than 1 month, has been evidenced in Carbone et al. (2003). It should also be noted that the temperature effects are evident for longer periods than 1 month (Carbone et al., 2003). The correction formulas are instrument-specific and often

Correspondence to: F. Greco
(greco@ct.ingv.it)

frequency-dependent. Obviously, frequency-domain filters cannot be applied to remove the effects of these perturbations since the spectrum of each noise component has wide intervals of superposition. A wavelet based approach to filter noise components from gravity signals was proposed in Panepinto et al. (2006). Andò and Carbone (2001, 2004, 2006) investigated the possibilities of a Neuro-Fuzzy algorithm as a tool to reduce the effect of meteorological variables from the continuous gravity signal.

Geomagnetic changes attributable to the dynamics of a volcano are usually very small, within $1\sim10$ nT, while changes up to a few hundreds nanoteslas are caused by natural geomagnetic fluctuations of external origin (ionospheric and magnetospheric currents, and secular variations) whose spatial distribution is, generally, considered uniform because of the great distance of their sources. The classical differential technique, based on simultaneous simple differences among the magnetic field amplitudes recorded at several points on a volcano, is the most frequently used and reliable method to remove them. Unfortunately, this simple technique does not allow properly reducing the geomagnetic signal to the level of a few nanoteslas, which is the apparent upper limit of detectability of magnetic anomalies associated with volcanic activity (Davis et al., 1981). Up to now filtering geomagnetic noise is a very complex problem that involves the development of different algorithms to reduce transient fields, which could be of the same order as the volcanomagnetic signal to be detected (Currenti et al., 2004). If very rapid changes are indeed characteristic of volcanomagnetic events (Sasai, 1990), then filtering techniques for removing short-period geomagnetic noise may be very helpful in increasing the detectability of volcano-related magnetic field changes. Methods of predictive filtering (Davis et al., 1981) and adaptive type approach (Currenti et al., 2004) have been suggested to remove changes in the difference fields due to contrasting responses at magnetometer sites. However, even if the effects of external and transitory fields are properly eliminated, both periodic and non-periodic geomagnetic changes have clearly been observed in the magnetic reduced signals. Since fluctuations are present even when no volcanic activity is apparently affecting the signal (e.g. Johnston, 1989; Zlotnicki et al., 2000; Del Negro et al., 2004) it is evident that external sources are responsible for these variations. Hence, it is necessary to correct this component in order to identify significant geophysical signals. Recent and more accurate studies claim that annual periodic variations in the geomagnetic total intensity could be caused by seasonal changes in the heterogeneous magnetization of near-surface rocks due to a diffusion of atmospheric temperature changes into the ground. The intensity of the local magnetic anomaly depends on the heterogeneity of the near-surface rocks and their temperature dependence. Using the method proposed by Utada et al. (2000), the features of annual variations can be quantitatively estimated by a simple one-coefficient filter. However, this simple linear filtering is not able to remove a residual annual component (Del Negro and Currenti, 2003) which is probably due to a non-linear effect of the temperature.

Notwithstanding the problem of filtering gravity and geomagnetic time series recorded in volcanic areas has been addressed by different authors, the lack of standard procedures justify the effort presented in this paper, devoted to describing the application of two different techniques for denoising gravity and geomagnetic data. The first one is a nonlinear autoregressive model based on the application of an Adaptive Neuro-Fuzzy Inference System (ANFIS), whose intrinsic learning features seem to be particularly suitable for such a task. The second one is a method obtained by combining wavelet transform and Independent Component Analysis (ICA), which is able to separate multiple data series into independent data series. Both techniques were applied to data recorded by the gravity and magnetic monitoring networks of Etna volcano (Italy). A comparison between ANFIS and ICA techniques is also reported together with a consideration on their usefulness.

The objective of the work was not the observation of anomalies related to an eruptive event but rather the study of methods for noise reduction in gravity and geomagnetic field measurements in volcanic area. It is important to note that, in this paper, the term noise is used to indicate only the components due to non volcanic sources (i.e. meteorological effects).

2 Mathematical background

To represent the relation between a geophysical variable $y(t)$ (e.g. gravity field and geomagnetic field) and a set of related variables $u_1(t), u_2(t), \ldots u_p(t)$, which represent candidate sources of noise, we can consider both non-linear such as autoregressive models with exogenous inputs (NARX), and linear approaches, such as the Independent Component Analysis (ICA) technique. NARX models are recurrent dynamic structures, with feedback connections, which can be represented as:

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u_1(t), u_1(t-1), \ldots,$$
$$u_1(t-n_1), \ldots u_p(t), u_p(t-1), \ldots u_p(t-n_p)) \quad (1)$$

where t represents the discrete time variable, $y(t)$ is the output model, $u_1, \ldots u_n$ are the model input variables, and f is an unknown non linear function, which can be approximated by using several methods such as the neuro-fuzzy approach considered in this work.

In neuro-fuzzy systems, neural networks are used to tune the membership functions of the fuzzy system and to automatically extract fuzzy rules from numerical data. The internal structure of a neuro-fuzzy network is illustrated in Fig. 1. The nodes of the first layer represent the crisp inputs. The activation functions of the second layer nodes are Gaussian and act as membership functions. Each neuron of the third
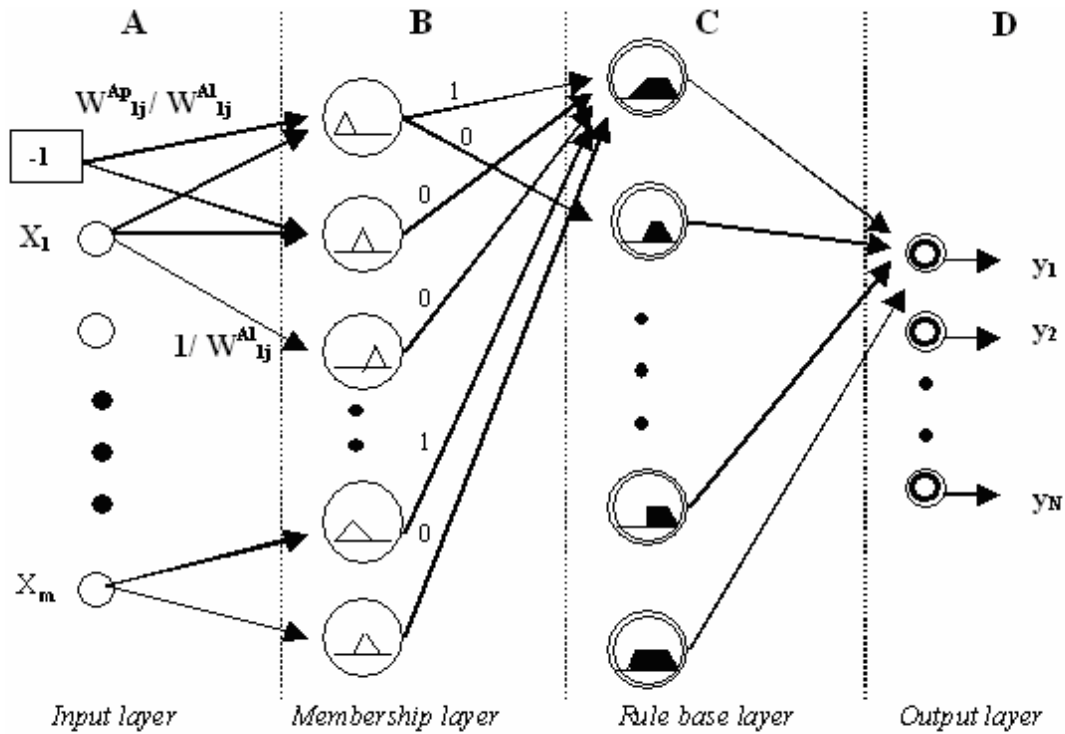
Fig. 1. Neuro-fuzzy network architecture.

layer acts as a rule node so that this layer provides the fuzzy rule base. The output of this layer determines the activation level at the output memberships. As ordinary neural nets, the neuro-fuzzy one learns on a training data set, tuning membership functions and rules, by means of a back-propagation algorithm.

When x_i is the i-th node in layer A, O_j^L is the j-th output of generic layer L and W_{ij}^L is the weight of the link between j-th neuron at layer L+1 and i-th neuron at layer L, each layer output can be described as follows:

Layer B: $O_j^B = \left(1 + \exp(-\dfrac{x_i - w_{ij}^{Ap}}{W_{ij}^{Al}})\right)^{-1}$

Layer C: $O_j^C = \min_i (w_{ij}^B \cdot O_j^B)$

Layer D: $O_j^D = \dfrac{\sum\limits_i w_{ij}^C \cdot O_i^C}{\sum\limits_i O_i^C}$

The second approach taken into account in this study, to express the relation between different variables, considers the static linear model:

$$y(t) = a_1 u_1(t) + a_2 u_2(t) \ldots + a_p u_p(t) \qquad (2)$$

In the model (2), the observed variable is considered to be a linear mixtures of some latent variables, assumed to be nongaussian and mutually independent. The set of unknown coefficients $a_1, \ldots a_p$, can be obtained by using various kinds of methods. By arranging the observations $y_i = a_{i,1} \cdot u_1 + \cdots a_{i,n} \cdot u_n$ of the random variable y into a vector Y it is possible to write

$$Y = \mathbf{A} \cdot u \qquad (3)$$

where $\mathbf{A}$ is the so-called mixing matrix. Thus, given the observation Y the problem is to estimate both the mixing matrix $\mathbf{A}$ and the observed sources u. This is done by adaptively calculating a cost function which either maximizes the nongaussianity or minimized the mutual information.

The original sources u can be recovered by multiplying the observed signals Y with the inverse of the mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, also known as the un-mixing matrix. Of course, in general the matrix A in not square thus the inverse matrix must be interpreted in the sense of the generalized-inverse (Ben-Israel and Greville, 2003).

One of the problems of the ICA approach is that it is not possible to identify the original scaling of the sources (Hyvärinen and Oja, 2000). The reason is that, both u and $\mathbf{A}$ in expression (3) are unknown. Thus, any scalar multiplayer in one of the sources u_i could always be cancelled by dividing the corresponding column a_i of the $\mathbf{A}$ matrix by the same scalar. However, in our applications, the energy of the unknown component, expect the volcanic source component, is measurable. For instance, one typical problem is to remove the meteorological variables (i.e. temperature, pressure, etc.)
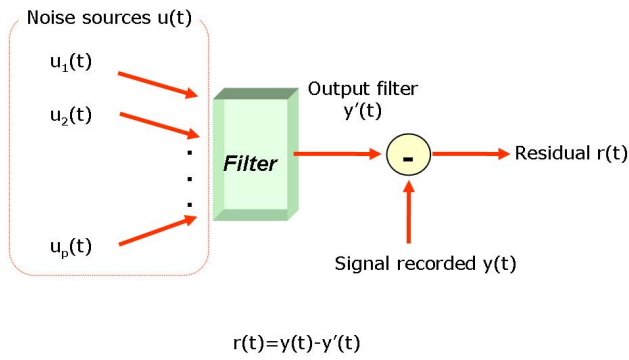
Fig. 2. Filtering strategy used to remove the effects of meteorological variables from gravity and geomagnetic time series considered.

Fig. 3. Schematic map showing the locations of the continuous gravity and magnetic stations operating on Mt. Etna.

effects. Thus, we can compute appropriate scale factor by measuring the oscillation of each component in the observed time period and comparing this oscillations with the corresponding one provide by the ICA approach.

3 Computation of the residuals

Regardless of the adopted representation for the relation among geophysical variables $y(t)$ and a set of candidate noise sources $u_1(t), \ldots u_p(t)$, the denoising scheme (Fig. 2) adopted in this work is based on the idea of existence of a time interval in which the signal $y(t)$ is not affected by a volcanic source. In this hypothesis, a residual signal $r(t)$ can be computed as:

$$r(t) = y(t) - y'(t) \tag{4}$$

where $y'(t)$ is the estimated value of $y(t)$. This residual in absence of effects due to the volcanic sources, that might involve mass redistribution and/or variations of the local geomagnetic field, will be limited in amplitude to typical ranges depending on the considered signals (i.e. ranging between a few μGal for gravity and a few nT for geomagnetic field). Knowledge of the magnitude of the residuals in "quiet periods" allows to recognizing and isolating the anomalies due to volcanic sources.

Let us represent the component x_i of the observed random vector $x=[x_1, x_2, \cdots x_m]^T$ are generated as a sum of the independent components u_k, $k=1, \cdots n$. The data is represented by the random vector $\boldsymbol{x}=[x_1, x_2, \cdots x_m]^T$ and the components as the random vector $\boldsymbol{u}=[u_1, u_2, \cdots u_n]^T$. The task is to transform the observed data x, using a linear static transformation W as $u=W \cdot x$ into maximally independent components u measured by some function of independence. This is done by adaptively calculating the w vectors and setting up a cost function which either maximizes the nongaussianity of the calculated $s_k=W \cdot x$ or minimizes the mutual information.

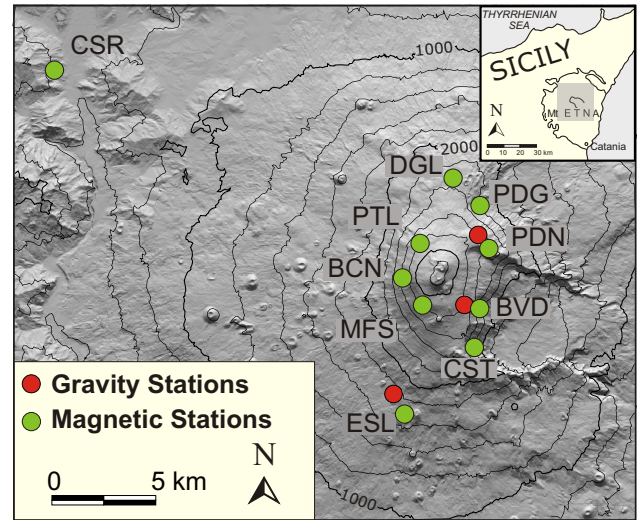## 4 Data set and preliminary analysis

The long and high-quality gravity and geomagnetic sequences recorded at Etna volcano during the last two decades represent an essential starting point to develop and validate analysis techniques to remove effects caused by meteorological variables. The continuously monitoring systems running at Etna were set up in 1998 (Del Negro et al., 2002; Carbone et al., 2003) and improved during recent years. At present, they consist of 3 gravity remote stations, a network of 6 scalar magnetometers and 3 magnetic gradiometers. Stations are located at elevations ranging between 1700 and 3000 m a.s.l. along a North-South profile crossing the summit craters. The magnetic reference station (CSR) is installed further west (about 50 km) on the Nebrodi Mountains (Fig. 3). The continuous recording stations were devised using innovative technologies which guarantee uninterrupted working under harsh environmental conditions.

The gravity stations are equipped with LaCoste and Romberg (L&R) spring gravimeters, featuring analog feedback systems, which are installed in partially buried concrete cases at ESL and BVD, while at PDN the gravimeter is located inside the observatory building. Data are recorded at 1 datum/min sampling rate through a CR10X Campbell Scientific datalogger. All magnetic stations are equipped with a GSM-90 Overhauser effect magnetometer (0.01 nT sensitivity). Each station synchronously samples the Earth's magnetic field every 5 s. A Global Positioning System (GPS) receiver controls the synchronization of readings. Simultaneously with gravity and magnetic signals, atmospheric and ground temperature, pressure and humidity are acquired at each station.
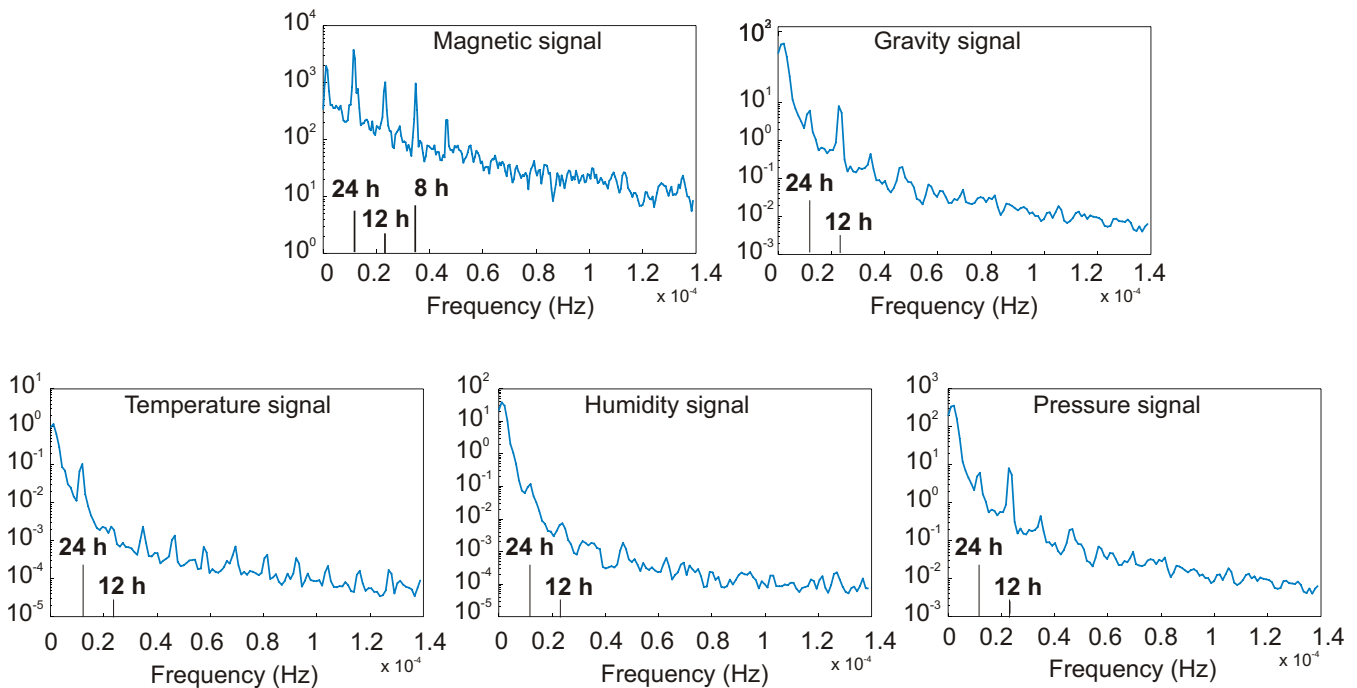
Fig. 4. Power spectral densities of analyzed signals: **(a)** geomagnetic; **(b)** gravity; **(c)** temperature; **(d)** humidity; **(e)** pressure.

The definition of a background level for each signal requires a long and continuous data series acquired in unperturbed (i.e. quiescent period) conditions, therefore we used data gathered on Etna when no significant volcanic activity was observed. As a first step, we reduced gravity data for the effect of Earth Tide and instrumental drift (Torge, 1989). The effect of Earth Tides (amplitude up to 250 μGal peak-to-peak depending on latitude, elevation and stage in the tidal cycle) is modeled through the Eterna 3.30 data processing package (Wenzel, 1996). The accuracies of the prediction model is within ±1%, implying tidal residuals affecting the gravity signal up to 1–2 μGal peak-to-peak over the most relevant tidal families (diurnal and semidiurnal). To correct the data for the main effect of instrumental drift a best linear fit was removed from the sequences. For geomagnetic signals we have used both the raw signals and the difference of the geomagnetic fields measured by magnetic stations located in the volcanic area with respect to reference station (CSR).

Considering that effect of meteorological variables mainly affects the long period components (Carbone et al., 2003; Del Negro et al., 2004), we computed hourly averages and used time series generally of about 6 months or more. In order to define the correlation between gravity, geomagnetic data, and meteorological variables (temperature, pressure and humidity), we performed the analysis in the frequency domain for each time series investigated. The power spectra analysis for each signal, obviously, reveals the presence of harmonics with their fundamental oscillations (see peaks in Fig. 4). In particular, besides seasonal components, the dominant periodic components of raw gravity data are centred around 12 and 24 h. The same semidiurnal and diurnal components were found in the temperature, pressure and humidity signals. In the same way, power spectra of geomagnetic data show prominent peaks at the period of 8, 12 and 24 h.

Cross-correlation analysis in the time domain between geophysical signals and meteorological variables was also performed. Although the results of this analysis are strongly dependent on the period chosen and on the length of the considered window time, they provide useful information on the general relationship between different signals. Both gravity and geomagnetic signals show a strong correlation with temperature. In particular, the correlation coefficient for different gravity and geomagnetic time series is up to −0.64 and 0.68, respectively. Moreover, gravity data are anti-correlated and show a considerable time lag up of to 900 h. As regards geomagnetic data, it is worth noting that the significant correlations found between differences of geomagnetic signals and temperature, over a time lag of about 10 h, are not so marked when we considered the raw signals.

Correlations were also observed between gravity signal and pressure (factor is up to −0.367) with a time delay of about 200 h, and between geomagnetic data and humidity (index is up to −0.54). Conversely low correlation coefficients were obtained between gravity sequences and humidity (especially at PDN station), and between geomagnetic data and pressure for all values of the time delay. As an example of the correlations identified in different periods at different stations, we report two representative series in Fig. 5.
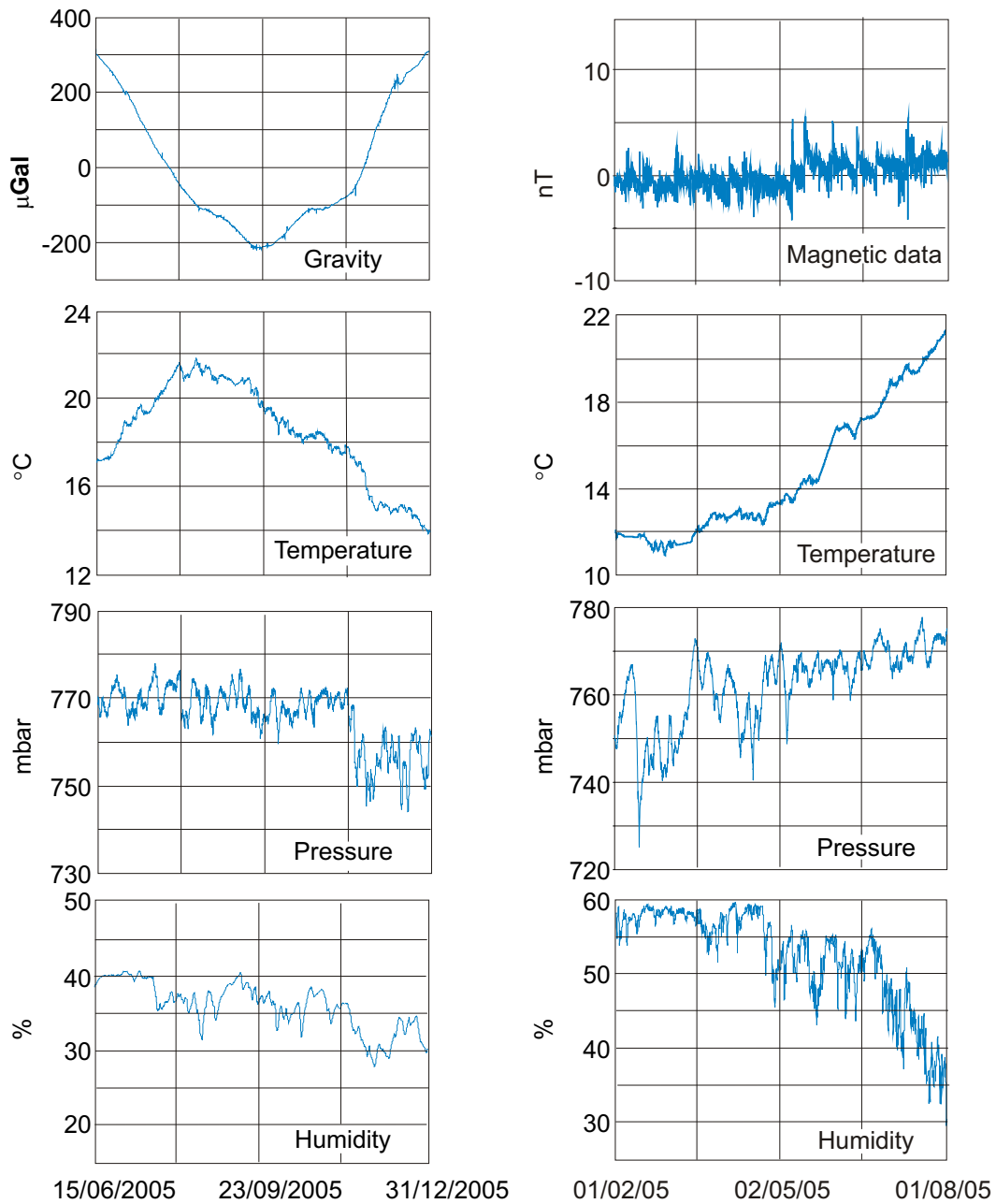
Fig. 5. Gravity, geomagnetic and meteorological variables (temperature, humidity and pressure) recorded on Mt. Etna from February to December 2005. The correlation coefficients between gravity and geomagnetic signals and meteorological variables are reported at the bottom.

Table 1. Three different structures to model the effects of temperature and pressure in the gravity signals.

Model 1	$y(t) = f(T(t), P(t))$
Model 2	$y(t) = f(T(t), T(t-m), P(t), P(t-n))$
Model 3	$y(t) = f(T(t), T(t-m), T_{\text{mean}}(t-m), P(t), P(t-n), P_{\text{mean}}(t-n))$

These correlations were taken into consideration during the implementation of the model for reducing geophysical signals for the effect of meteorological variables.

5 The ANFIS non linear approach

The adaptive network based on fuzzy inference system (AN-FIS) is a specific approach in neuro-fuzzy development. Neuro Fuzzy systems make use of neural networks (ANNs) in order to determine the parameters of a fuzzy rule base from fuzzy sets and fuzzy rules by processing data examples (patterns). These systems are able to capture the benefits of fuzzy logic and ANNs in a single framework. ANFIS is based on a fuzzy Sugeno's model which has shown significant capability in modeling nonlinear systems. It can simulate and analyze the mapping relation between the input and output data through a learning procedure to implement a set of fuzzy rules in "if-then" form to determine the optimal distribution of the membership functions. In ANFIS, the membership function parameters are extracted from a dataset that describes the system behaviour and successively optimized according to a given error criterion (Jang, 1993; Ubeyli and Guler, 2006) during the learning process. The optimization is accomplished using a hybrid algorithm combining the least squares method and the gradient descent method. The training process aims to minimize the training error between the real target and the ANFIS output. This allows ANFIS to learn features from observed data, and represents the final model in the form of linguistic rules.

We implemented ANFIS autoregressive non linear models to denoise gravity and geomagnetic signals from effects of temperature, pressure and humidity. The development environment adopted was the Matlab® Neuro-Fuzzy tool. The model uses three bell shaped membership functions and a Sugeno model structure of the rule base. The available datasets were divided into training and testing subsets in order to ensure the validation of the model on fresh data, i.e. data not considered during the training phase. This avoids the well known problem of over-fitting. Based on results of correlation analysis the most promising candidates as input variables were the temperature (T) and pressure (P) for the gravity signal and the temperature and humidity (H) for geomagnetic signals. Once the input variables have been chosen, it is necessary to define the structure of the autoregressive model, i.e. the values of n_y, $n_{1...}n_p$ which appear in expression (1). We tested three different configurations for model-

Table 2. Standard deviation of the residuals obtained by three different models described in Table 1.

	Model 1	Model 2	Model 3
Standard deviation	70.07 μGal	13.35 μGal	4.86 μGal

ing the effects of temperature and pressure in the gravity signals, as indicated in Table 1, where f is the non-linear model estimated by neuro-fuzzy algorithm; t is the present time index; T and P represent temperature and pressure, respectively; m and n represent appropriate time delays obtained by the cross-correlation analysis; $T_{\text{mean}}(t-m)$ and $P_{\text{mean}}(t-n)$ are the mean values of temperature and pressure within the intervals $[t-m, t]$ and $[t-n, t]$.

To estimate the goodness of the models presented in Table 1, we calculated the standard deviation of the residuals as performance index and results are shown in Table 2. It is evident that the estimation capabilities of the first two models are worse than model 3, this is, probably due to the lack of information on the average behaviour of temperature $T_{\text{mean}}(t-m)$ and pressure $P_{\text{mean}}(t-n)$.

Model 3 was practically applied to a real case study consisting of the gravity sequences recorded from January to December 2005 at BVD station and from June to December 2005 at PDN station (Fig. 6). These gravity stations, the only working during the 2005 on Etna, are equipped with LaCoste and Romberg D-185 and PET 1081 gravimeters, respectively. After removing the theoretical Earth Tide and the instrumental drift, as described above, large components with amplitude of about 600 (BVD) and about 500 μGal peak-to-peak (PDN) are strongly dominant in both gravity sequences (Fig. 6). The amplitude of the residual signals, calculated as the difference between the instrumental effect due to atmospheric temperature and pressure, estimated by model 3 and gravimeters output, is very low compared to the original signals. These results highlight that the instrumental effects of atmospheric temperature and pressure are the most significant components of all the original signals, and are strongly confirmed also through the low correlation coefficients between residual gravity sequences and temperature and pressure signals (see Table 3).

We can see that the ANFIS method removes satisfactorily the long period components of the gravity sequences,
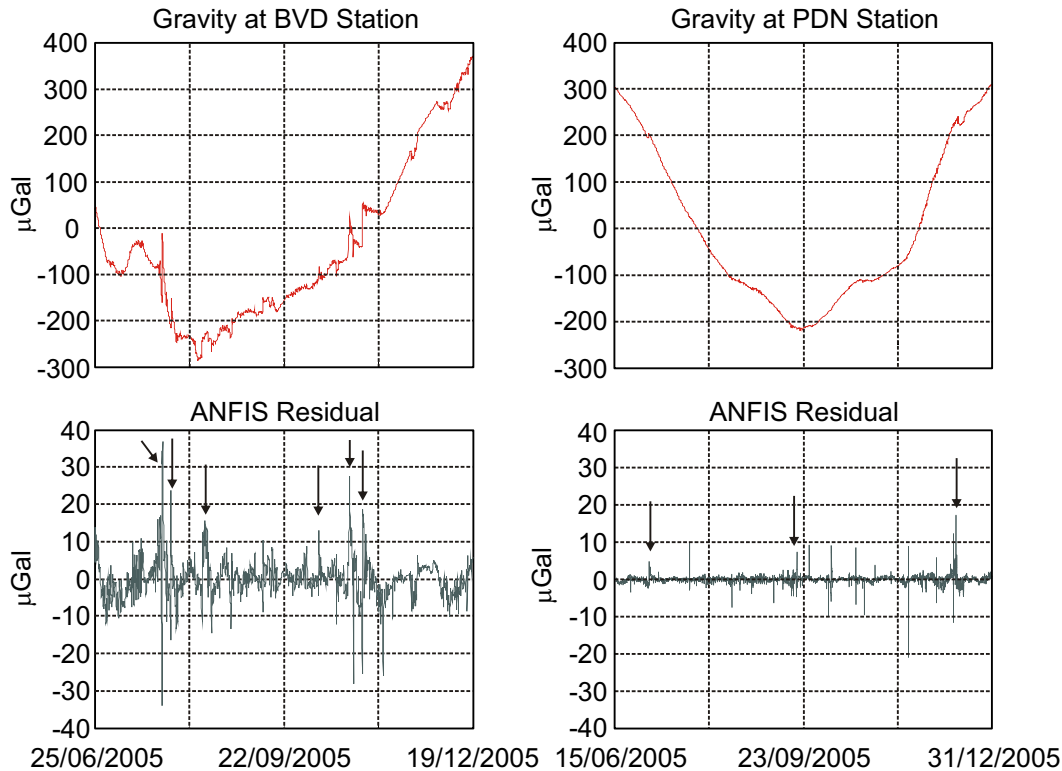
Fig. 6. Left panel: at the top, gravity signal after removal of the best linear fit and the theoretical Earth Tide observed at BVD station from June to December 2005. At the bottom, gravity residual after removing the meteorological effects estimated by ANFIS model. Right panel: at the top, gravity signal after removing the best linear fit and the theoretical Earth Tide observed at PDN station from June to December 2005. At the bottom, gravity residual after removing the meteorological effects estimated by ANFIS model. Black arrows indicate anomalies which are not related to the meteorological variables.

Table 3. Correlation coefficients between raw gravity data, residual gravity signals and temperature and pressure after removing the corresponding temperature and pressure effects estimated by AN-FIS model from each gravity sequence. It is important to note the large difference of the relative values of the correlation coefficients before and after the filtering approach were performed.

	Temperature	Pressure
Raw gravity data at BVD	−0.570	−0.350
Raw gravity data at PDN	−0.643	−0.367
Gravity residual at BVD	−0.055	−0.043
Gravity residual at PDN	−0.029	−0.035

considered harmonic of the annual oscillation due to the influence of the seasonal atmospheric variables changes. The magnitude of the residuals (BVD and PDN) as well as diurnal and semidiurnal components are highly comparable each other and are in the order of $2 \div 3\,\mu$Gal (well-matched also with the uncertain of the Earth Tide model used). Instead, a significant component with a maximal amplitude range of

about $\pm 10\,\mu$Gal peak-to-peak and a period of about 20 days strongly emerges in the residual of BVD (Fig. 6). This discrepancy is probably due to the position of the BVD station, located very close to the SE Crater (about 700 m). The residual may reflect changes in the local gravity field due to the "normal" activity of the Crater.

For geomagnetic data, since the results of the simulations obtained by using the model 2 are better than the results gained by applying the model 3, we used the structure of the model 2, but pressure was substituted by humidity (H) on the grounds of the cross correlation analysis; consequently the model is defined as follows:

$$y(t) = f(T(t), T(t-m), H(t), H(t-n)) \qquad (5)$$

This model was applied to the hourly averages of total intensity variations from February to August 2005 observed at PDN and DGL stations, relative to the reference station (CSR). It is worth stressing that, though external magnetic fields were previously removed by differential technique, and at that time no significant volcanic activity occurred, a clear trend is still evident especially at DGL station (Fig. 7a). In order to evaluate the capability of the filtering process, we compared the estimated residuals with the differences of total
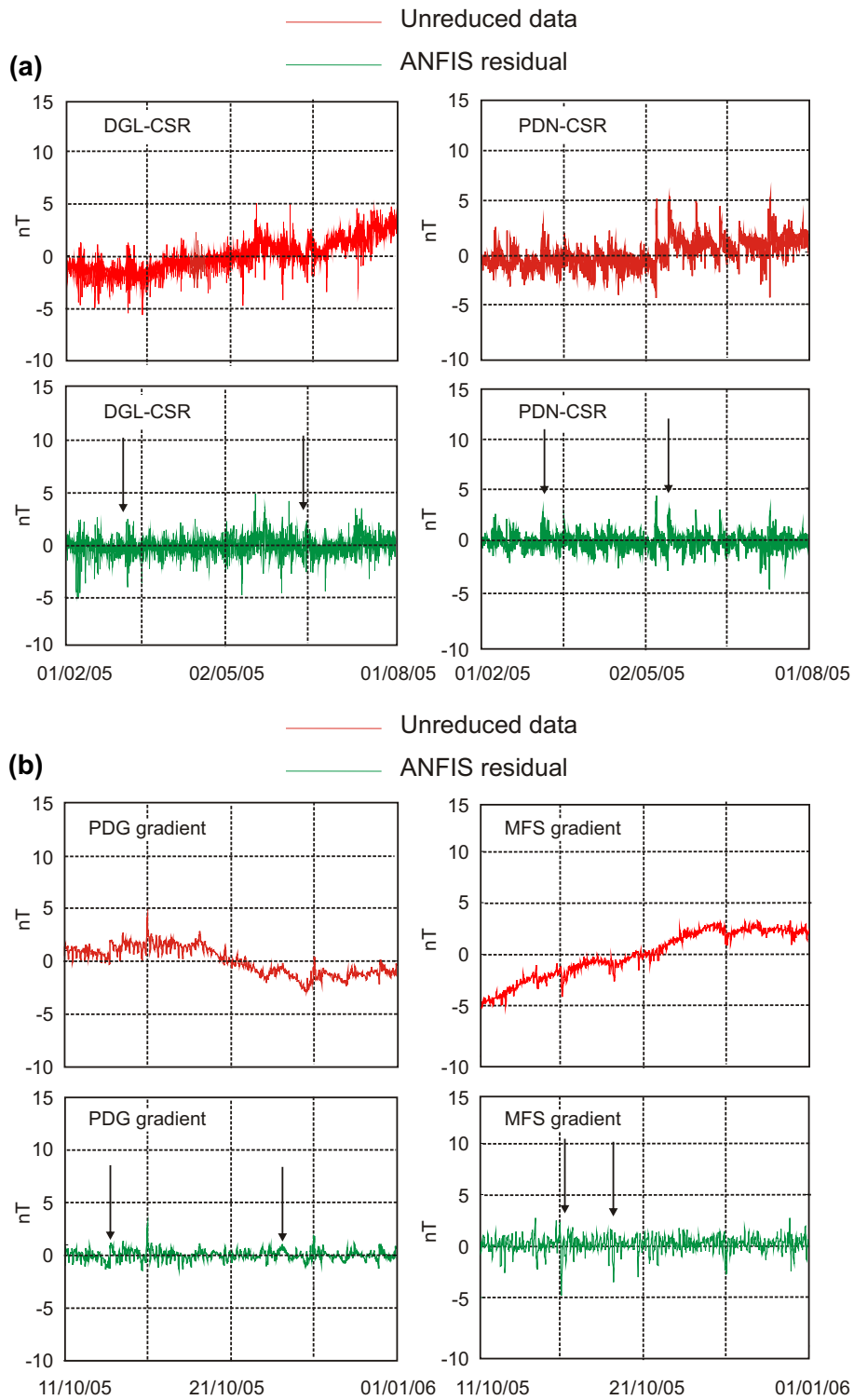
Fig. 7. Comparison between the unreduced geomagnetic data (in red) and the residuals (in green) estimated by ANFIS model considering **(a)** the differences of the geomagnetic signals (DGL-CSR and PDN-CSR) and **(b)** raw signals recorded at the gradiometric stations of PDG and MFS. Black arrows indicate anomalies which are not related to the meteorological variables.

Table 4. Correlation coefficients between meteorological variables (temperature and humidity), recorded geomagnetic signals and residuals obtained by ANFIS filtering.

Unreduced data	Temperature	Humidity	Residuals	Temperature	Humidity
PDN-CSR	0.68	0.54	PDN-CSR	0.003	0.008
DGL-CSR	0.66	0.20	DGL-CSR	0.00001	0.002
MFS gradient	0.87	0.42	MFS gradient	0.0016	−0.02
PDG gradient	0.84	−0.18	PDG gradient	0.021	−0.003

Table 5. Correlation coefficients between meteorological variables (temperature and pressure), raw gravity signals and their related Independent Components extracted by FastICA method, and the residual gravity signal after filtering with the related Independent Components identified by FastICA method.

Raw gravity	Temperature	Pressure	Residual gravity	Temperature	Pressure
BVD	−0.570	−0.350	BVD	0.052	0.135
IC_{BVD}	−0.572	−0.359			
PDN	−0.643	−0.367	PDN	−0.020	−0.052
IC_{PDN}	−0.602	−0.401			

magnetic intensity with respect to CSR station. The magnitude of the residuals is lower than the original ones and no evident trend appears (Fig. 7a). At the same time, we used hourly averages of total intensity variations recorded at the MFS and PDG gradiometric stations. Each station consists of two sensors, (namely MFSnorth, MFSsouth, PDGnorth and PDGsouth) spaced horizontally by about 50 m, which simultaneously sample the Earth's magnetic field. In this case, we initially applied the model to the raw signals recorded by each sensor and then the residual signals were differentiated to obtain the gradient (MFSnorth-MFSsouth and PDGnorth – PDGsouth). Figure 7b shows the comparison between the hourly averages of the unreduced magnetic gradients recorded from October 2005 to January 2006 at MFS and PDG and the difference of residuals estimated by ANFIS model. It is evident that the long period fluctuations affecting original signals, probably due to the joint effects of temperature and humidity, are successfully removed. To better estimate the validity of the model, as well as for the gravity case, a correlation analysis between each residual and temperature and humidity was calculated. Also in this case, the correlation coefficients significantly decreased, ranging between 0.02 and 0.00001 for temperature and between 0.02 and −0.003 for humidity (Table 4). These results confirm that the applied model removes both the effect of temperature and the humidity from magnetic data.

6　The ICA linear approach

The Independent Component Analysis (Bell and Sejnowski, 1995) is an algorithm which can be used for blind source separation and feature extraction from mixed signals. ICA outputs a set of linearly independent signals, given a set of the original multi-channel input signals. This method can be used directly for feature extraction but requires more than one time series (at least from two separate sensors). To overcome this constraint, we propose a method to generate multiple time series from the single available time series. According to Ming et al. (2005), we use the wavelet transform to pre-process data recorded by single gravity and magnetic sensor and then use the obtained information as input for an apposite ICA tool developed in Matlab® language. We assume that each time series recorded by gravity and magnetic sensors is affected by multiple sources, and we are interested in estimating the mixing ratios of the source signals in the collected data in order to obtain the independent source signals. In particular, we apply ICA to remove from gravity and geomagnetic time series the noise signal (due to meteorological effects) characterized through the frequency analysis.

The FastICA algorithm, based on the approach proposed by Hyvärinen and Oja (1997), seeks to find a set of independent components (IC) by estimating the maximum negentropy (Hyvärinen, 1999). After the pre-processing step, the FastICA analysis starts choosing the indices of the largest and smallest eigenvalues of the covariance matrix of the observed signals to be included in the reduced data. Once the eigenvalues of the covariance matrix of data are computed, we chose the number of Independent Components (IC) related with the ICA algorithm according to the subset of significant eigenvalues. Gravity and geomagnetic residuals are obtained by removing the denormalized independent components, which show a strong correlation with one or more
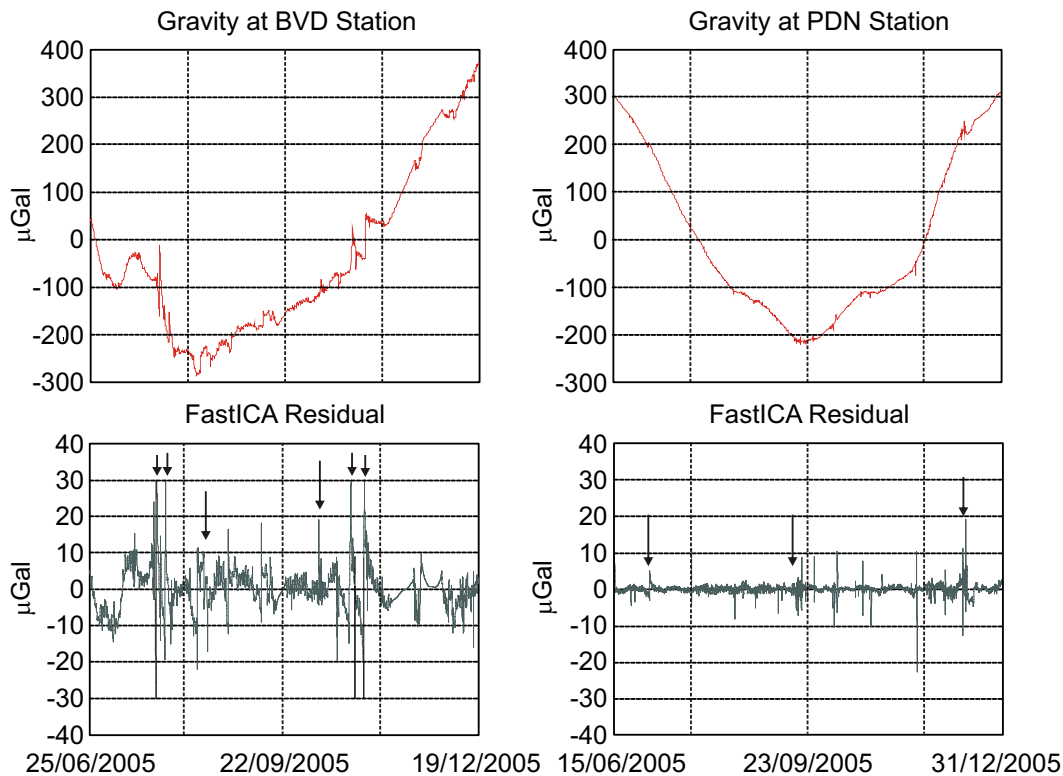
Fig. 8. Left panel: at the top, gravity signal after removal of the best linear fit and the theoretical Earth Tide observed BVD station from January to December 2005. At the bottom, gravity residual after removing the meteorological effects estimated by wavelet/ICA combined method. Right panel: at the top, gravity signal after removing the best linear fit and the theoretical Earth Tide observed at PDN station from June to December 2005. At the bottom, gravity residual after removing the meteorological effects estimated by wavelet/ICA combined method. Black arrows indicate anomalies which are not related to the meteorological variables.

meteorological signals, from the observed gravity and geomagnetic signals. As stated before, the scaling factor for each ICs was computed by comparing the oscillation of each measurable meteorological variables with the magnitude of each ICs component.

The FastICA algorithm was applied to the same gravity sequences analyzed by the ANFIS non-linear model. The components obtained from the wavelet decomposition of gravity sequences were used as input to FastICA. Only one independent component (IC) was found to be suitable in describing the effects of meteorological variables for each data set considered. The correlation coefficients between gravity signals recorded at PDN and BVD stations and the only one Independent Component extracted from both signals and temperature and pressure are reported in Table 5.

The magnitude of the residual signals, after the independent component was subtracted from the time series recorded at BVD and PDN stations, is less than 98% of the original signals (Fig. 8). The correlation coefficients between gravity residuals and meteorological variables are negligible (see Table 5). This means that the combined method wavelet/FastICA is able to recognize the main components induced by the meteorological variables in the gravity sig-

nals. Analysis performed on residuals obtained through this combined ICA-wavelet method reveals that residuals are similar to those obtained by the non-linear ANFIS approach. Thus both the proposed methods are able to remove meteorological effects from gravity signals.

The ICA approach in combination with wavelet transform was also applied to the same geomagnetic data sets used for validating the ANFIS model previously described. First of all, we performed the wavelet decomposition both of the differences in the geomagnetic signals (PDN-CSR, DGL-CSR) and raw signals gathered at the gradiometric stations of PDG and MFS. Geomagnetic signals were decomposed from scales 1 to 7 and 1 to 9 on the basis of the length of the considered time window. Therefore, ICA was used to process the obtained matrix of wavelet coefficients for detecting single independent sources. The analysis performed on the differences of geomagnetic signals identified 4 Independent Components (IC), while 3 IC were detected for the raw signals of gradiometric stations. These components are associated with the principal eigenvalues of the covariance matrix. It is worth stressing that for all cases only one IC is correlated both with temperature and humidity, while the others are uncorrelated with the meteorological variables considered. Furthermore,
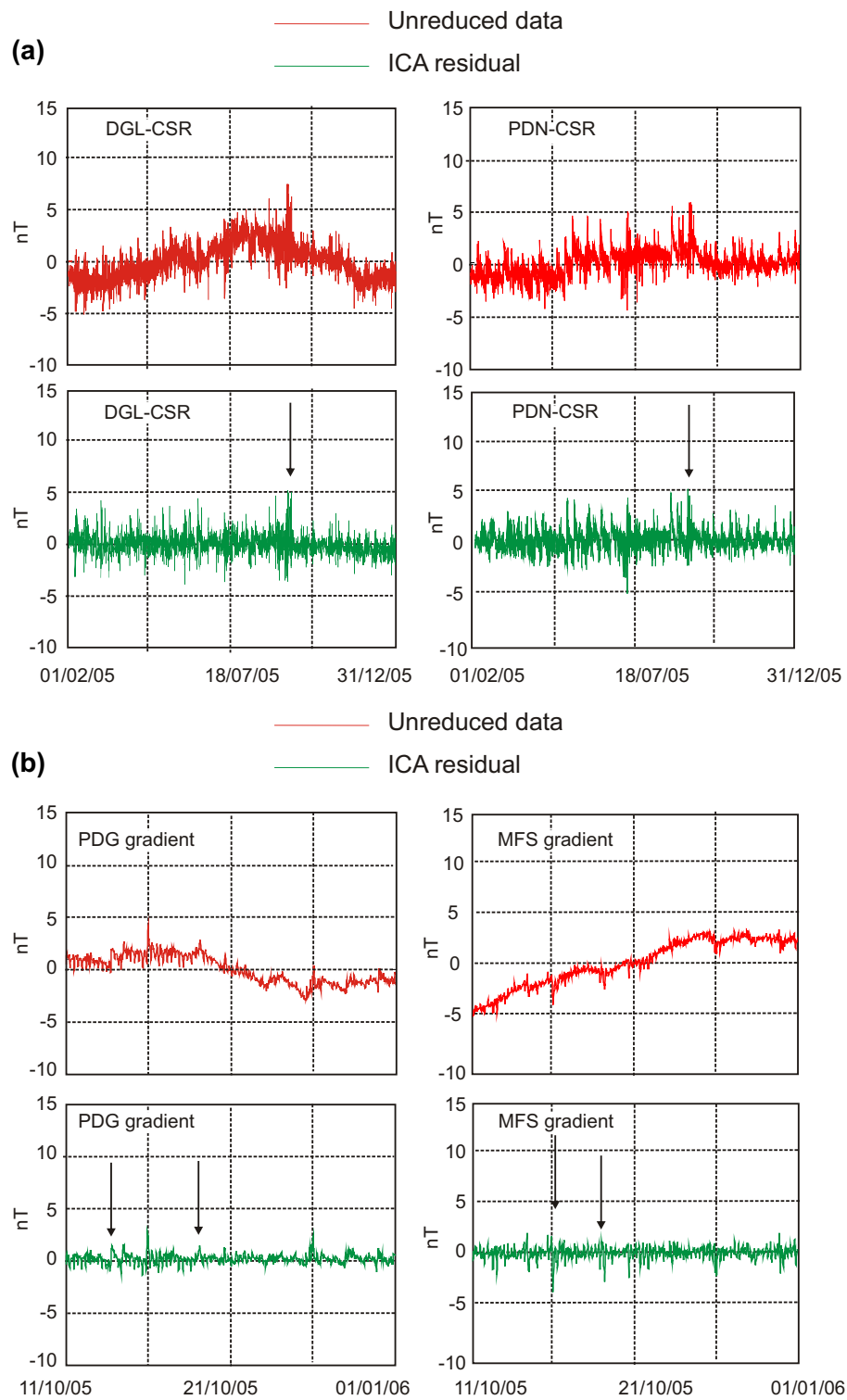
Fig. 9. Comparison between the unreduced geomagnetic data (in red) and the residuals estimated by ICA approach (in green) considering **(a)** the differences of the geomagnetic signals (DGL-CSR and PDN-CSR) and **(b)** raw signals recorded at the gradiometric stations of PDG and MFS. Black arrows indicate anomalies which are not related to the meteorological variables.

Table 6. Correlation coefficients between meteorological variables (temperature and humidity), observed geomagnetic signals, and their related Independent Components extracted by FastICA method.

	Temperature	Humidity		Temperature	Humidity
PDN-CSR	0.68	0.54	DGL-CSR	0.66	0.20
IC1	0.016	−0.0006	IC1	0.002	−0.0003
IC2	0.001	0.013	IC2	0.01	0.018
IC3	−0.73	0.68	IC3	0.003	−0.005
IC4	0.011	−0.035	IC4	0.91	−0.42
MFSn	−0.21	0.08	MFSs	0.016	−0.02
IC1	−0.19	0.07	IC1	0.066	0.06
IC2	−0.12	0.01	IC2	0.013	−0.035
IC3	−0.55	0.19	IC3	0.02	−0.04
PDGn	−0.31	0.08	PDGs	−0.41	0.1
IC1	−0.25	0.07	IC1	0.17	0.06
IC2	0.14	−0.06	IC2	−0.19	0.027
IC3	0.66	−0.24	IC3	−0.83	0.25

the correlation coefficient computed between the meteorological variables and the selected IC is always higher than that obtained between the observed signal and meteorological variables (Table 6).

The extracted IC was then denormalized, as stated before, and removed from the observed signals. The residuals obtained for the differences PDN-CSR and DGL-CSR, and for MFS and PDG gradients were compared with the unreduced signals (Fig. 9). The comparison shows that the long period variations affecting the original signals were efficiently removed. Moreover, the cross-correlation analysis in the time domain between residual signals and meteorological variables provided very low correlation coefficients (Table 7). These results confirm that the method of combining wavelet transform and ICA is a valuable tool for simultaneous separation of ICs affected by hidden meteorological effects in the observed geomagnetic signals.

7 Discussion and conclusion

With the aim of developing a novel approach to analyze gravity and geomagnetic time series recorded in volcanic areas, we proposed two different methods, namely the ANFIS and the ICA. Results presented throughout this paper show the denoising capability of the two considered approaches for removing noise from both gravity and geomagnetic signals. In particular, the very low correlation coefficients between residuals and the set of explaining variables confirm that both approaches are able to efficiently remove the effects of meteorological variables from considered geophysical data. It is important to note the presence of small anomalies (such as amplitude) in the gravity and magnetic time series in the

Table 7. Correlation coefficients between meteorological variables (temperature and humidity), and residual signals obtained after removal of the related Independent Component identified by FastICA method.

	Temperature	Humidity
PDN-CSR	0.0015	0.005
DGL-CSR	0.00008	0.007
MFS gradient	−0.0062	0.023
PDG gradient	0.013	−0.001

unreduced signals, which are not due to changes in the meteorological variables (Figs. 6 to 9).

The standard deviations of residuals were also assessed and compared to verify the obtained results from ANFIS and ICA approaches (Table 8). It should be noted, that the standard deviation of residuals is much decreased both for gravity and geomagnetic data when compared to that of the unreduced signals. In particular, gravity residuals show standard deviations lower than 98÷99% in comparison with original data, while the standard deviation of geomagnetic residuals decreased by about 40% (for PDN, DGL and MFS stations) and more than 60% for the PDG station.

The ANFIS and ICA techniques remove efficiently noise components showing themselves a valid approach to the general problem of denoising geophysical data. The results are highly promising, and in our view the proposed techniques outperform traditional time series filtering in terms of efficiency. This is an important chance since the gravity and magnetic signals could include volcanic effects with a wide range of evolution rates. Moreover, frequency-domain filters

Table 8. Standard deviations of recorded gravity, geomagnetic signals and residuals obtained by the ANFIS and the ICA approaches.

Gravity data	Unreduced signal	ANFIS residual	ICA residual
PDN	165.36 μGal	1.01 μGal	1.18 μGal
BVD	155.98 μGal	4.86 μGal	6.19 μGal
Geomagnetic data	Unreduced signal	ANFIS residual	ICA residual
PDN-CSR	1.25 nT	0.66 nT	0.72 nT
DGL-CSR	1.56 nT	0.84 nT	0.79 nT
MFS gradient	2.25 nT	1.25 nT	1.22 nT
PDG gradient	1.34 nT	0.47 nT	0.41 nT

cannot be efficiently applied to remove the effect of these perturbations since the spectrum of each component of various origins has wide intervals of superposition. Furthermore, frequency domain filtering does not always work well because: (i) it globally removes frequencies causing a generalized smoothing effect that substantially broadens features of interest; (ii) depending on both cut off frequency and filter order it also could introduce edge effects and distortions of the original signal; (iii) it does not allow to study local features of the signal in the time domain.

Finally, on comparing the standard deviations of ANFIS and ICA residuals, it appears that the efficacy of the two approaches is very similar. Thus, the criteria for choosing one rather than the other should be only based on considerations such as computational speed and degree of difficulty in implementing and applying the proposed filtering scheme. Our research suggests that the ICA approach is more suitable in this respect and is thus recommended to solve the considered filtering problem.

References

Andò, B. and Carbone, D.: A methodology for reducing a continuously recording gravity meter for the effect of meteorological parameters, IEEE Trans. Instrum. Meas., 50(5), 1248–1254, 2001.

Andò, B. and Carbone, D.: A test on a Neuro-Fuzzy algorithm used to reduce continuous gravity records for the effect of meteorological parameters, Phys. Earth Planet. Int., 142, 37–47, 2004.

Andò, B. and Carbone, D.: A new computational approach to reduce the signal from continuously recording gravimeters for the effect of atmospheric temperature, Phys. Earth Planet. Int., 159, 247–256, 2006.

Bell, A. J. and Sejnowski, T. J.: An information-maximization approach to blind separation and blind deconvolution, Neural Comput., 7, 1129–1159, 1995.

Ben-Israel, A. and Greville, T. N. E.: Generalized Inverses: Theory and Applications, (2nd edition), Springer-Verlag, New York, 400 pp., 2003.

Carbone, D., Budetta, G., Greco, F., and Rymer, H.: Combined discrete and continuous gravity observations at Mt. Etna, J. Volcanol. Geotherm. Res., 123, 123–135, 2003.

Currenti, G., Del Negro, C., Fortuna, L., Napoli, R., and Vicari, A.: Non-linear analysis of geomagnetic time series from Etna volcano, Nonlin. Processes Geophys., 11, 119–125, 2004, http://www.nonlin-processes-geophys.net/11/119/2004/.

Davis, P. M., Jackson, D. D., Searls, C. A., and McPhernon, R. L.: Detection of tectonomagnetic events using multichannel predictive filtering, J. Geophys. Res., 86, 1731–1737, 1981.

Del Negro, C., Napoli, R., and Sicali, A.: Automated system for magnetic monitoring of active volcanoes, B. Volcanol., 64, 94–99, 2002.

Del Negro, C. and Currenti, G.: Volcanomagnetic signals associated with the 2001 flank eruption of Mt. Etna (Italy), Geophys. Res. Lett., 30(7), 1357, doi:10.1029/2002GL015481, 2003.

Del Negro, C., Currenti, G., Napoli, R., and Vicari, A.: Volcano-magnetic changes accompanying the onset of the 2002-2003 eruption of Mt. Etna (Italy), Earth Planet. Sci. Lett., 229, 1–14, 2004.

El Wahabi, A., Ducarme, B., Van Ruymbeke, M., d'Oreyè, N., and Somerhausen, A.: Continuous gravity observations at Mount Etna (Sicily) and correlations between temperature and gravimetric records, Cah. Centre Eur. Géodyn. Séismol., 14, 105–119, 1997.

Hyvärinen, A. and Oja, E.: A fast fixed-point algorithm for independent component analysis, Neural Comput., 9(7) 1483–1492, 1997.

Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis, IEEE T. Neural Networ., 10(3), 626–634, 1999.

Hyvärinen, A. and Oja, E.: Independent Component Analysis: Algorithms and Applications, Neural Network, 13(4–5), 411–430, 2000.

Jang, J. S. R.: ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybernet., 23(3), 665–685, 1993.

Johnston, M. J. S.: Review of magnetic and electric field effects near active faults and volcanoes in the U.S.A, Earth Planet. Sci.

Lett., 57, 47–63, 1989.

Ming Zuo, J., Jing, L., and Xianfeng, F.: Feature separation using ICA for a one-dimensional time series and its application in fault detection, J. Sound Vib., 287(2005), 614–624, 2005.

Panepinto, S., Greco, F., Luzio, D., and Ducarme, B.: An overview on wavelet multi-resolution decomposition compared with traditional frequency domain filtering for continuous gravity data denoising, B. Inform. Marèes Terrestres, 141, 11 213–11 224, 2006.

Torge, W.: Gravimetry, Walter de Gruyter, Berlin-New York, 465 pp., 1989.

Ubeyli, E. and Guler, D.: Adaptive neuro-fuzzy inference system to compute quasi-TEM characteristic parameters of microshield lines with practical cavity sidewall profiles, Neurocomputing, 70, 296–304, 2006.

Utada, H., Neki, M., and Kagiyama, T.: A study of annual variations in geomagnetic total intensity with special attention to detecting volcanomagnetic signals, Earth Planets Space, 52, 91–103, 2000.

Warburton, R. J. and Goodkind, J. M.: The influence of barometric-pressure variations on gravity, Geophys. J. R. Astron. Soc., 48, 281–292, 1997.

Wenzel, H. G.: The nanogal software: Earth tide data processing package ETERNA 3.30, B. Inf. Marees Terrestres, 124, 9425–9439, 1996.

Zlotnicki, J., Bof, M., Perdereau, L., Yvetot, P., Tjetjep, W., Sukhyar, R., Purbawinata, M., and Suharno, A.: Magnetic monitoring at Merapi volcano, Indonesia, J. Volcanol. Geotherm. Res., 100, 321–336, 2000.