**Engineering Strong-Motion database: a gateway to access European strong motion data**

L. Luzi(1), R. Puglia(2), E. Russo(3), M. D’Amico(4), G. Lanzano(5), F. Pacor(6), C. Felicetta(7), ORFEUS-WG5(8)

(1) Senior Researcher, Istituto Nazionale di Geofisica e Vulcanologia, lucia.luzi@ingv.it

(2) Technologist, Istituto Nazionale di Geofisica e Vulcanologia, rodolfo.puglia@ingv.it

(3)  Researcher, Istituto Nazionale di Geofisica e Vulcanologia, emiliano.russo@ingv.it

(4) Researcher, Istituto Nazionale di Geofisica e Vulcanologia, maria.damico@ingv.it

(5) Researcher, Istituto Nazionale di Geofisica e Vulcanologia, giovanni.lanzano@ingv.it

(6) Senior Researcher, Istituto Nazionale di Geofisica e Vulcanologia, francesca.pacor@ingv.it

(7) Researcher, Istituto Nazionale di Geofisica e Vulcanologia, chiara.felicetta@ingv.it

(8) Working group 5, ORFEUS (Observatories and Research Facilities for European Seismology)

# *Abstract*

The repeated attempts of building unified European strong-motion databases are motivated by the increasing demand of strong motion data, that are one of the primary sources of information used by engineering seismologists and earthquake engineers to predict ground shaking.

We describe the Engineering strong-motion database (esm.mi.ingv.it), implemented during the NERA project (www.nera-eu.org, 2010-2014), within the 7th Framework Program of the European Commission, that has recently become an infrastructure of EPOS (European Plate Observing System, www.epos-eu.org).

This infrastructure represents a bridge between present and past; on one side it exploits the most advanced technical solutions adopted by network operators for data distribution, and, on the other, preserves important strong-motion data mainly recorded by analogue instruments before 2000. In this way the database can be continuously updated with real-time data, but also with offline data, as soon as they become available. Moreover, it has been specifically designed to provide end-users (engineers and scientists alike) quality-checked, uniformly processed data and tools for data selection.

ESM is composed by a centralized database, with a core formed by about 80 tables, with a web interface and several tools in order to: i) search data from online archives and catalogues, ii) populate the database, iii) perform quality control and iv) process accelerometric data online. The underlying IT framework is tailored on the peculiarity of strong-motion data, that need the support of a variety of metadata on seismic events and recording stations to be fully exploited for engineering applications (e. g. derivation of ground motion prediction equations, selection of accelerograms compatible with code spectra, derivation of seismic codes).

*Keywords: strong-motion; acceleration; database; strong-motion data processing.*

# 1. Introduction

# Strong-motion data are one of the primary sources of information used by engineering seismologists and earthquake engineers to predict ground shaking and perform structural seismic analysis. Since 1998, when the first European Strong-motion Database, supported by the European Commission (5th Framework Programme), was published, several attempts of building unified engineering strong-motion databases in Europe have been carried out. Unfortunately all of them are the results of single projects and the data collection stopped after the end of resources. Differently from the previous attempts, that were essentially project based, the approach followed to build the Engineering Strong-Motion database (ESM, http://esm.mi.ingv.it) consisted in the networking of European strong-motion data operators with the aim of creating a long-term infrastructure. The resulting database is specifically tailored to serve engineers and scientists alike in the assessment of seismic hazard, as it provides end-users only quality-checked, uniformly processed strong-motion data from events in the magnitude range 4.0 - 7.8 (where the minimum magnitude is selected according to the lower threshold generally employed for engineering applications and hazard assessment). If magnitudes larger than 4 are considered, ESM encompasses all the previous European databases and is continuously updated as it is closely linked to the European Integrated Data Archive, EIDA (http:// http://www.orfeus-eu.org/eida), a key infrastructure aiming at archiving digital waveforms in Europe.

# This paper describes the data collection strategy adopted for ESM, the content of the database and the way of accessing and disseminating data.

# 2. ESM data collection strategy

The state of the art of strong-motion data collection in Europe indicates that too many transitions were made after the initial database attempt in 1998. Therefore, before undertaking the construction of the ESM database, a step back was made to retrieve the unprocessed data, from the original databases. To this aim, three main regional databases were identified:

− Unified HEllenic Accelerogram Database (HEAD), released in 2004 and containing Greek waveforms and metadata from 1973 to 1999 [1];

− ITalian ACcelerometric Archive (ITACA), the database of Italian strong-motion data from 1972 to 2015 [2], [3];

− Strong-motion database of Turkey (TR-NSMN), containing the Turkish data set from 1976 to 2007 [4].

Finally, pan-European data from 1972 to 2008, not included in regional databases, are extracted from the European Strong-motion Database (http://www.isesd.hi.is/).

Raw waveforms extracted from these databases form the initial core of the ESM data collection, and represent the following percentages of the current dataset: HEAD ~2%, ITACA ~20%, TR-NSMN ~7% and ESD ~1%.

A common characteristic of “historical” strong-motion data is the lack of standardized metadata and formats for their full integration with actual seismological standards. Therefore, a major effort has been spent towards the homogenization and standardization of metadata. Standardized network codes, assigned by the international Federation of Digital Seismograph Networks, are assigned to all providers, including the ones no longer operating (e.g. networks belonging to the former Yugoslavia), whereas waveforms are named following the SEED convention.

Nevertheless, the innovation of ESM is not only the preservation of the existing patrimony of “historical” strong-motion data and their harmonization with actual seismological standards. A major innovation is the full exploitation of the potential of the new generation data centers.

Currently, new generation instruments can record weak-to-strong-motions and transmit waveforms in real-time to data centers, soon after the occurrence of a seismic event. A recent initiative within Observatories and Research Facilities for European Seismology (ORFEUS, http://www.orfeus-eu.org) is the European Integrated Data Archive, EIDA, structured as a distributed data center, to securely archive seismic waveforms gathered by European research infrastructures, and providing transparent access to the archives. Several European data centers act as EIDA nodes, collecting and archiving data from seismic networks. The availability of continuous data streams allows to access seismic signals in quasi-real time and to progressively populate the ESM database through semi-automatic procedures.

After the occurrence of any event with magnitude larger than or equal to 4, reported by the European-Mediterranean Seismological Centre (EMSC), an automatic procedure for signal windowing is applied to the continuous streams available through EIDA and waveforms are uploaded into ESM together with their metadata. At this stage, only preliminary location parameters are assigned to the earthquake. Offline waveforms, e.g. data not available through web-services, are then uploaded after the network operators make them available.

Before the manual revision, waveforms are automatically processed using the procedure proposed for the Italian strong-motion database ITACA [5]. Automatically processed waveforms are stored in the database, although not accessible to users, and only the peak ground acceleration (PGA) is published. Subsequently, waveforms are manually revised and processed acceleration, velocity and displacement time series are released together with acceleration and displacement response spectra at 5% damping. Bad quality records (e.g. noisy records, or waveforms containing spikes) are made available to users only in the unprocessed version. The release time of processed waveforms ranges from few hours (relevant events) to few days.

Event and station metadata contained in the database are periodically revised. The event information collected from earthquake-specific literature studies are always ranked as the primary reference for large seismic events. For moderate to small events the sources of information are regional catalogues (e.g. the INGV Bulletin) or the Bulletin of the International Seismological Centre, ISC, in case regional catalogues are unavailable. The ISC bulletin relies on the contributions from worldwide seismological agencies and is typically 24 months behind real-time. Different magnitudes (e.g. Mw, Ml, Mb, Ms) are reported in the database as well as moment tensor solutions from different agencies. Information on the geometries of the seismic sources comes from regional or international catalogues or from specific source-model studies.

Station metadata are periodically updated, after specific studies are published in the literature or after the results of national and international projects. The station information actually contained in ESM is obtained from regional databases (ITACA, TR-NSMN, HEAD) or specific literature studies [6], [7], [8].

# 3. Description of data contained in ESM

The ESM database contains 23,000 three-component waveforms (March 2016), of which about 13.000 are manually processed. Manually processed waveforms derive from 1,929 seismic events (M≥4.0) and have been recorded by 1,901 sites operated by 38 networks. The following statistics is relative to this subset of waveforms.

Fig. 1 shows the magnitude-distance sampling of the records, whereas in Fig. 2 the distributions of event depths and style of faulting are shown in terms of number of records. The ESM dataset is well sampled in the magnitude range 4.0-6.0 and in the epicentral distance range 10 - 200 km (Fig. 1), with a significant number of waveforms related to strong events (6.0<Mw<7.5) recorded at epicentral distances larger than 10 km. The strongest events have been recorded in Italy (1980 Irpinia, Mw = 6.9), Turkey (e.g. 1999 Izmit, Mw = 7.4; 1999 Duzce, Mw = 7.1) and in Iran (1990 Western Iran, MW = 7.4). As shown in Fig. 2, most of the events in the database are shallow crustal earthquakes with hypocentral depth lower than 40 km. The majority of records is related to normal (~30%) and reverse (~25%) faulting, although a considerable amount of records is still not associated to a style of faulting (~30%).

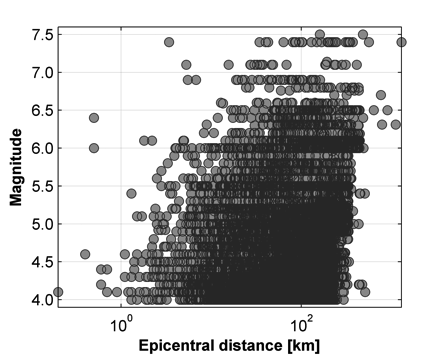


Fig. 1 - Magnitude versus distance distribution of the manually processed waveforms in the ESM database.

Several recording stations are characterized by geotechnical and geophysical measurements (available for several Albanian, French, Iranian, Italian, Swiss and Turkish stations). Geophysical measurements, in particular, consist of about 340 1D velocity profiles estimated through different geophysical prospection methods. The average shear wave velocity of the uppermost 30 meters (VS,30), a fundamental information for engineering applications, has been calculated from the available velocity profile or, in a few cases, is taken from literature. Fig. 3 (left panel) shows that the majority of stations have VS,30 in the range 200-600 m/s, whereas very few stations are characterized by VS,30 > 800 m/s (rock conditions). As the percentage of sites associated to a VS,30 value is quite low (18%), in order to characterize a larger number of sites, we make use of the support of surface geology. Fig. 3 (right panel) shows the distribution of the EC8 subsoil categories, evaluated from the measured VS,30 or inferred from surface geology. Despite the support of surface geology, information for a remarkable number of sites (32%) is still missing.

|  |  |
| --- | --- |
| Depth.png | SoF.png |

Fig. 2 - Distribution of number of waveforms in function of: event depth (top left); style of faulting (top right; NF: normal faults; SS: strike slip faults; TF: reverse and thrust faults; UN: undefined fault mechanism).

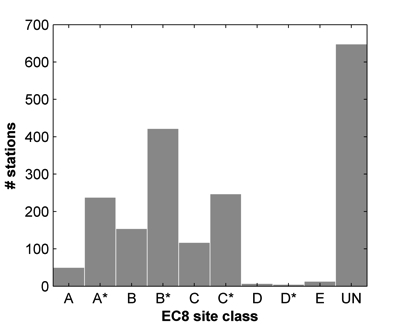
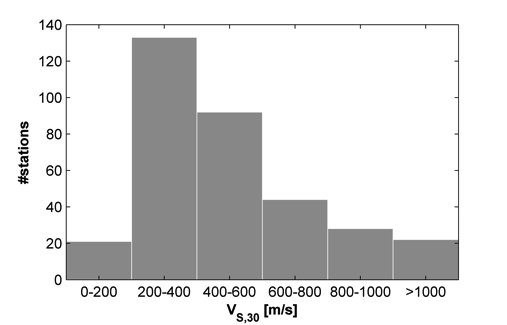


Fig. 3 - Left: Distribution of the VS,30 values for the ESM stations; right: distribution of the EC8 soil categories (classes denoted by an asterisk are inferred from surface geology, e.g. 1:100.000 scale maps).

4. Data dissemination

The ESM web site is organized in three main blocks, relevant to waveforms, recording stations and seismic events. Seismic events can be retrieved entering the page “Events” of the portal. User can select 13 parameters including date and time of the event, magnitude range, hypocentral coordinates or style of faulting. The query returns a list of earthquakes that can be individually accessed.

Fig. 4 shows the example of one of the largest events occurred in Europe on 1999-08-17 at 00:01:38 UTC (Mw = 7.4) at Izmit, Turkey. The event has been recorded by 23 stations belonging to the AFAD network (code TK) and is associated to the ID’s of major international catalogues in order to provide complete information to the user. In this example the preferred location is by the International Seismological Centre, while several magnitude determinations are attributed from international agencies. Focal mechanism solutions and style of faulting are also provided. In case of important events, like this one, the geometry of the seismic source is also reported. The list of stations that have recorded the event is reported at the bottom of the page and includes metadata such as EC8 site class, epicentral distance and the maximum peak ground motion (e.g. PGA, PGV, PGD) of the three components.

Station information can be accessed entering the page “Stations”, where recording sites can be retrieved according to 14 parameters including location, network and station code, and parameters related to the site characterization, such as the average velocity in the uppermost 30 m (VS,30). Fig. 5 shows the example for the station CLF, belonging to the network IT (Italian accelerometric network, operated by the Italian Department of Civil Protection). The station is displayed on a topographic map and information related to location, housing and site class are provided. Metadata related to each station are also available in the form of a report (“Monography” button on the station page) containing detailed information, such as stratigraphic and geophysical logs or the horizontal to vertical spectral ratio obtained from noise measurements.

Waveforms information can be accessed entering the page “Waveforms”, where 35 parameters can be specified related to stations, events, or waveform metadata. Waveforms can be explored with the aid of a visualization tool, that allows zooming and exporting the time-series as images.

Upon user registration, time-series can be downloaded in ASCII format, as unprocessed acceleration time series or processed acceleration, velocity and displacement time series, or acceleration, pseudo-velocity and displacement response spectra (5% damping) calculated at 105 periods (0.01 - 10s). A client, written in Python language, can be downloaded from the ESM home page in order to convert ASCII files in standard seismological formats (e.g. sac or mseed).

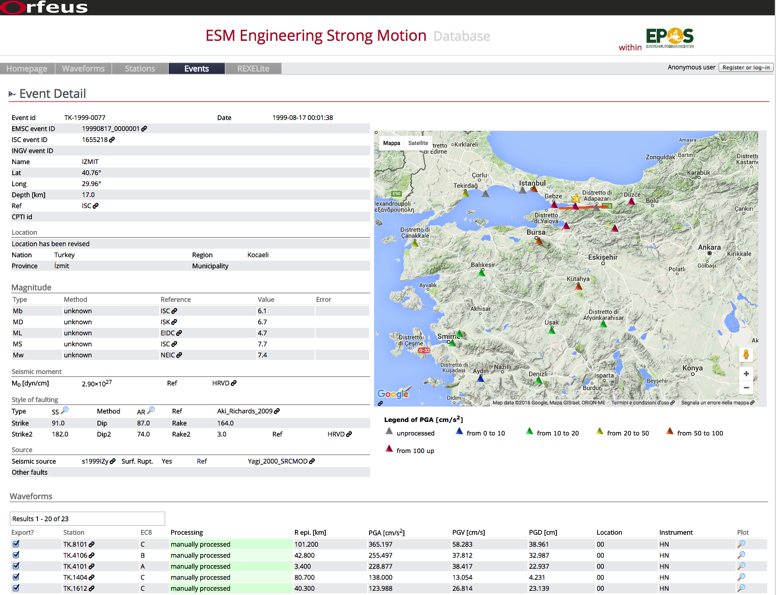


Fig. 4 - Detail of the Mw = 7.4 event occurred on 1999-08-17 at 00:01:38 UTC (http://goo.gl/73Qvrj).

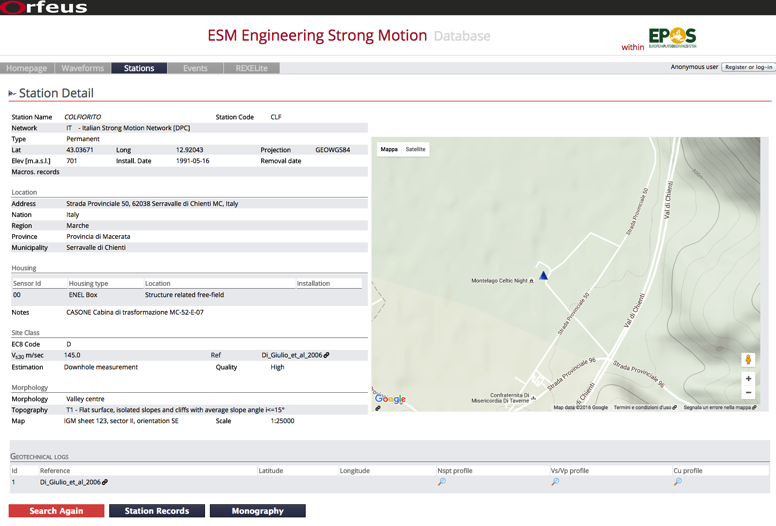


Fig. 5 - Details of the station CLF, belonging to the network IT, operated by the Italian Civil Protection (http://goo.gl/KPMGd0).

# 5. Available tools

*5.1 Waveform processing*

A waveform processing web front-end is available at http://esm.mi.ingv.it/processing, providing access to all waveforms included in the ESM database. Individual acceleration time-series are processed manually following the general procedure, described in [5], that consists of:

− linear detrend of the uncorrected acceleration signal (subtraction of a first order polynomial);

− application of a cosine taper, at begin and end of the signal, with percentage fixed to 5% of the signal length, with the possibility of being modified by the user;

− visual inspection of the Fourier spectrum to select the band-pass frequency range (band-pass frequency may be different for the three components);

− application of a 2nd order acausal time-domain Butterworth filter to the acceleration time-series; zero-pads are added at the beginning and end of the signal before the acausal filter is applied [9];

− removal of zero pads from the acceleration trace;

− (begin / end) taper of the acceleration signal, with percentage fixed to 5%;

− computation of the velocity signal and linear detrend;

− (begin / end) taper of the velocity signal, with percentage fixed to 5%;

− computation of displacement signal and linear detrend;

− (begin / end) taper of the displacement signal, with percentage fixed to 5%;

− recursive differentiation to obtain velocity and acceleration time-series, respectively.

Acausal filters are preferred to causal, as elastic and inelastic response spectra are sensitive to the corner periods used in causal filtering at periods much shorter than the corner periods [10].

Fig. 7 shows two screenshots representing the processing web front-end of the three components of the ground motion recorded at station AVJ (Iran) on 2002-06-22 at 02:58:21 GMT (Mw = 6.5). Fig. 7a displays the three components of the unprocessed and processed acceleration time-series, whereas Fig. 7b displays the corresponding Fourier spectra, useful to check, in the frequency domain, the result of the filtering.

Registered users can select waveforms, perform customized processing and save the results.

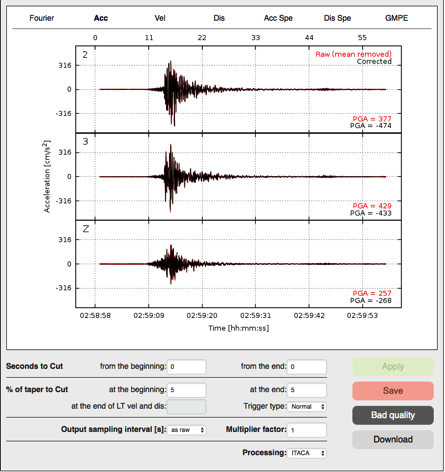
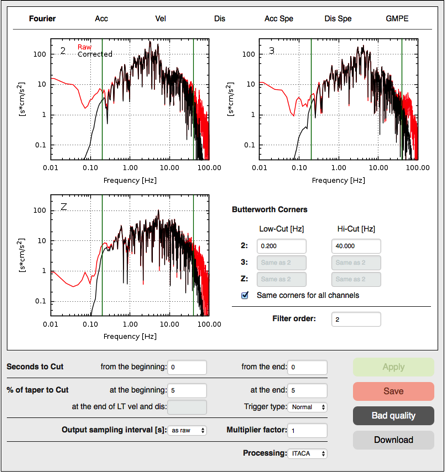
 

Fig. 7 - a) Three components of the ground motion recorded at the station AVJ (Iranian Strong Motion Network, I1) on 2002-06-22 at 02:58:21 GMT (Mw = 6.5); b) Fourier spectra of the three components.

*5.2 Spectrum-compatible data selection*

The ESM database is coupled with the REXELite application, which is the online version of the computer program REXEL [11], [12], for the selection of ground motion suites for code-based seismic structural analyses. REXELite allows searching for combinations of seven 1- or 2-components strong-motion records, compatible, in average, with a specified code spectrum. More specifically, REXELite: i) automatically builds code spectra for any limit state according to Eurocode 8 (EC8, ENV 1998) and ii) finds the set of seven records having the most similar spectral shape with respect to that of the code, and whose average also matches the target spectrum, in a user-specified period range and with a desired tolerance. The records are pre-selected by the user according to specific features, such as magnitude and source-to-site distance ranges, style of faulting and soil conditions, codified as EC8 site classes. The resulting set of accelerograms may include unscaled (original) or amplitude-scaled records and may be used for code-compliant non-linear time history analyses of structures.

# 6. Conclusions

We described the Engineering Strong-Motion database (ESM), tailored to enable users to fully exploit pan-European strong-motion data recorded since 1969, relative to events with magnitude larger than or equal to 4. The database has been designed for a large variety of stakeholders (expert seismologists, earthquake engineers, but also students and professional) and for this reason the web interface is friendly and straightforward. In addition, expert users may benefit from specific tools for data processing and data selection.

The core of ESM has been built from existing regional databases (~30% of the actual waveforms) and it is constantly growing thanks to the continuous supply of waveforms gathered from the European Integrated Data Archive (EIDA) or offline archives by several European providers. The rate of growth of the database is about 3,000 waveforms per year, if we exclude seismic sequences that could double the estimated rate.

ESM has a long term-vision for the distribution of strong-motion data in Europe. It is distributed within ORFEUS and it relies on a large network of strong-motion data operators in Europe. They are directly involved in the decisional process (e.g. setting rules for data dissemination) and are continuously updated on the technological progress and on the state of the art of techniques for metadata compilation and data processing.

Moreover, the Engineering Strong-Motion database has been selected as one of the infrastructures of EPOS, the European Plate Observing System (www.epos-eu.org), a long-term plan for the integration of national and transnational Research Infrastructures for solid Earth science in Europe, to provide seamless access to data, services and facilities.

# 7. Acknowledgements

The ESM interface was designed and implemented by INGV using funding from EC-funded FP7 project NERA (Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation, http://www.nera-eu.org). We are thankful to the NERA-NA3 working group (coordinated by prof. Sinan Akkar) for supporting the development of this work.

We thank the EIDA and ORFEUS Working Group 5 (Acceleration and strong-motion data) communities for providing open data to the community and for the fruitful discussions on data policy.

We acknowledge different institutions and single researchers for providing local databases, original waveforms or metadata or being part of the ORFEUS-WG5:

* Laboratory of Institute of Engineering Seismology and Earthquake Engineering, Thessaloniki (ITSAK), Greece, that provided the HEAD data (1973-2000);
* National Observatory of Athens (NOA), Greece: HEAD, the Unified HEllenic Accelerogram Database;
* Italian Department of Civil Protection (DPC) and Istituto Nazionale di Geofisica e Vulcanologia (INGV): ITACA, the ITalian ACcelerometric Archive;
* Disaster and Emergency Management Presidency of Turkey (AFAD): strong-motion database of Turkey (TR-NSMN) from 1976 to 2007;
* Earthquake Engineering Research Centre of the University of Iceland (EERC): Internet Site for European Strong-Motion data (ESD);
* Building and Housing Research Center, Iran (BHRC);
* Institute of GeoSciences, Energy, Water and Environment (IGEWE) Tirana, Albania;
* Helmholtz Centre Potsdam (GFZ) German Research Centre for Geosciences;
* French seismologic and geodetic network (RESIF);
* Bogazici University Kandilli Observatory And Earthquake Research Institute (KOERI), Turkey;
* National Institute for Earth Physics (NIEP), Bucharest, Romania;
* Swiss Seismological Service (SED);
* Friuli accelerometric network (RAF), Trieste, Italy;
* National Institute of Oceanography and Geophysics (OGS) Trieste, Italy

The REXELite web interface is developed in collaboration with the Laboratories University Network of seismic engineering (ReLUIS, www.reluis.it/); we are grateful to prof. I. Iervolino, G. Baltzopoulos and E. Chioccarelli for adapting REXELite to the ESM database.

# 8. Reference

1. Theodulidis, N, Kalogeras I, Papazachos C, Karastathis V, Margaris BM, Papaioannu C, Skarlatoudis A (2004): HEAD 1.O: A Unified HEllenic Accelerogram Database. *Seismological Research Letters*, **75** (1), 36–45.
2. Luzi, L, Hailemikael S, Bindi D, Pacor, F, Mele F, Sabetta F (2008): ITACA (ITalian ACcelerometric Archive): A Web Portal for the Dissemination of Italian Strong-motion Data. *Seismological Research Letters*, **79** (5), 716–722, Doi: 10.1785/gssrl.79.5.716
3. Pacor F, Paolucci R, Luzi L, Sabetta F, Spinelli A, Gorini A, Nicoletti M, Marcucci S, Filippi L, Dolce M (2011): Overview of the Italian strong-motion database ITACA 1.0. *Bulletin of Earthquake Engineering*, **9** (6), 1723–1739, doi: 10.1007/s10518-011-9327-6
4. Akkar, S, Çağnan Z, Yenier E, Erdoğan Ö, Sandıkkaya MA, Gülkan P (2010): The recently compiled Turkish strong-motion database: preliminary investigation for seismological parameters. *Journal of Seismoogy*, **14** (3), 457–479, doi:10.1007/s10950-009-9176-9.
5. Paolucci R, Pacor F, Puglia R, Ameri G, Cauzzi C, Massa M (2011): Record Processing in ITACA, the New Italian Strong-Motion Database. Earthquake Data in Engineering Seismology - Predictive Models, Data Management and Networks, S. Akkar, P. Gülkan, T. van Eck (Editors), ISBN: 978-94-007-0151-9 (Printed version) 978-94-007-0152-6 (E-book version), Springer, 99-113.
6. Zare M, Bard PY, Ghafory-Ashtianya M (1999): Site characterizations for the Iranian strong-motion network. *Soil Dynamics and Earthquake Engineering*, 18, 101–123.
7. Régnier, J, Laurendeau A, Duval AM, Guéguen P (2010): From heterogeneous set of soil data to VS profile: Application on the French permanent accelerometric network (RAP) sites. *14 EECE,* Ohrid, FYROM.
8. Michel, C, Edwards B, Poggi V, Burjanek J, Roten D, Cauzzi C, Fah D (2014): Assessment of Site Effects in Alpine Regions through Systematic Site Characterization of Seismic Stations. *Bulletin of the Seismological Society of America*, **104** (6), 2809-2826, doi:10.1785/0120140097.
9. Boore, DM (2005): On Pads and Filters: Processing Strong-Motion Data. *Bulletin of the Seismological Society of America*, **95** (2), 745–750.
10. Boore DM, Akkar S (2003): Effect of causal and acausal filters on elastic and inelastic response spectra. *Earthquake Engineering Structural Dynamics*, **32** (11), 1729–1748, doi:10.1002/eqe.299.
11. Iervolino, I, Galasso C, Cosenza E (2009): REXEL: computer aided record selection for code-based seismic structural analysis. *Bulletin of Earthquake Engineering*, **8** (2), 339–362, doi:10.1007/s10518-009-9146-1.
12. Iervolino I, Galasso C, Paolucci R, Pacor F (2011): Engineering ground motion record selection in the ITalian ACcelerometric Archive. *Bulletin of Earthquake Engineering*, **9** (6):1761-1778.