Microtremor Measurements in the City of Palermo, Italy: Analysis of the Correlation between Local Geology and Damage

by Fabrizio Cara, Giovanna Cultrera, Riccardo Mario Azzara, Valerio De Rubeis, Giuseppe Di Giulio, Maria Stella Giammarinaro, Patrizia Tosi, Paola Vallone, and Antonio Rovelli

Abstract This study presents the results of 90 seismic ambient noise measurements in Palermo, the main city of Sicily (Italy). The dataset has been processed using the horizontal-to-vertical spectral ratio (HVNSR) technique and interpreted in terms of local geology, which is characterized by the presence of alluvial sediments of two riverbeds masked by urbanization since the seventeenth century. HVNSRs show significant variations in the study area: when the transition stiff to soft is crossed, a typical spectral peak appears in the HVNSRs, mostly in the frequency band 1–2 Hz, and exceeding a factor of 3 in amplitude. Using available information on subsurface geological structure, we compute theoretical 1D and 2D transfer functions. The resonance frequencies of soft soils obtained by HVNSR are well reproduced by the fundamental frequencies from numerical modeling.

The distribution of frequency peaks of HVNSR and their amplitudes are also compared with the local damage caused by historical earthquakes. Previous studies demonstrated that damage variations in Palermo were controlled more by near-surface geology than building vulnerability. A uniform vulnerability is an ideal condition to test statistical methods and their capability in seeking correlation between HVNSR and potential damage due to local geological conditions. We apply two well-established multivariate statistical methodologies (factor analysis and canonical correlation) to the HVNSR dataset and macroseismic data (damage grades of the European macroseismic scale). Through these analyses we quantify the significance of the correlation between the HVNSR peak in the low-medium frequency range (0.5–3 Hz) and the occurrence of the highest damage grades. This approach allows us (1) to estimate the threshold value in the resulting linear combination of the HVNSR amplitudes, which separates zones of light damage from zones of significant damage, and therefore (2) to improve the spatial definition of potentially high hazard zones through a denser grid of microtremor measurements.

Introduction

The damage pattern in urban areas during an earthquake depends on the characteristics of the event and on the interaction between site response and vulnerability of the exposed structures. For seismic hazard assessment, the site effect is typically represented by the resonance frequency and the associated amplification of ground motion. Among several methods (e.g., site-to-reference spectral ratios, horizontal-to-vertical spectral ratios [HVNSRs], array-data analysis), many authors use ambient noise vibrations (microtremors, hereinafter) to retrieve these parameters. In particular, following Nakamura (1989), the HVNSR of microtremors has been widely used to determine the resonance frequency of a site and has been applied successfully in urban environments (Duval et al., 2001; Lebrun et al., 2001; Panou et al., 2005). There are many reasons why use of this technique spread very rapidly: (1) it can be applied in regions with low seismicity rates, (2) a short time of acquisition is required in the field, and (3) low computational time and minimal computing resources are needed to analyze data. The seismological community is still debating about several aspects of the method, both practical and theoretical (Fäh et al., 2001; Malischewky and Scherbaum, 2004; Bonnefoy-Claudet, Cornou, et al., 2006). Open questions are: (1) What does the noise wave field consist of? (2) Is the method really effective in every geological condition? (3) Are there standards for operating in the field and for computing HVNSR?

Indeed, successful and unsuccessful experiences have alternated (e.g., Seo, 1998; Bard, 1999; Mucciarelli and Ga...
Microtremor Measurements in the City of Palermo, Italy

New advances and overall guidelines have been given by the project SESAME (site effects assessment using ambient excitations) funded by the European Commission (SESAME Project, 2001). A significant indication that came out from the project is that the HVNSR method is successful when sites have simple geological structures with strong impedance contrast; in such cases the resonance frequency obtained with microtremors tends to be coincident with that obtained using weak motion recordings although the level of amplification is often underestimated (WP04 of SESAME obtained using weak motion recordings). In such cases the resonance frequency obtained with microtremors tends to be coincident with that obtained using weak motion recordings although the level of amplification is often underestimated (WP04 of SESAME project, deliverable D16.04, 2004; see also Satoh et al., 2001). It was also underlined that the absence of resonance frequencies in HVNSRs does not imply itself the absence of site effects during an earthquake. For this reason, the HVNSR approach should not be used stand alone for seismic microzonation purposes.

Part of the SESAME project was dedicated to investigating correlations between the HVNSR results and the observed damage distribution during earthquakes. Some studies (e.g., Lachet et al., 1996; Teves-Costa et al., 2001) find that the HVNSR results are correlated with the spatial distribution of damage when damage variation is controlled by near-surface geology. In this article we present results of an experiment performed in the city of Palermo, southern Italy (see the inlet in Fig. 1) within the framework of the project. The urban area of Palermo is characterized by high spatial variability of geology, partially masked by urban growth. HVNSRs from 90 noise measurements are first compared to the known geology. Second, standard multivariate statistical analyses (Davis, 2002) are used to correlate the HVNSR results with the damage pattern in historical earthquakes under the assumption that there was limited variation in the vulnerability of buildings beforehand. The availability of geological and damage data allows us to test how the HVNSR technique can be used to get information on site response and statistically quantify its correlation with damage.

Geological Setting and Felt Seismicity

The city of Palermo (Fig. 1) was founded on a tectonic depression mostly filled by deposits that are Pleistocene in age. The bedrock of the area is composed of clay and sandstone levels of Oligocene–Miocene age (Numidian Flysch) with local quartzarenite layers. It is unconformably overlain by Plio–Pleistocene sediments (silty-clayey sands and calcarenites). These latter are locally overlain by Holocene sediments that consist of both sea (silts, sands, gravels, and calcirudites) and land deposits (sandy silts with peat layers and organic materials, thin sands and clayey silts).

The substratum is incised by the bed of four rivers flowing from southwest to northeast. Two of them, Papireto and Kemonia, were buried and partially filled during the seventeenth century. They are documented by historical maps and joined in downtown Palermo, flowing into the old harbour (Fig. 1). The high level of urbanization now masks their presence. The width of the infill is typically about 150 m, and the thickness of soft deposits increases towards the confluence of the two rivers, from 10 to more than 20 m for Papireto, and from 15 to more than 30 m for Kemonia. A third stream, namely Passo di Rigano, in the northern part of Palermo, is also buried nowadays. The south–southeast side of Palermo, adjacent to the historical center, has been influenced by the presence of the Oreto river (Fig. 1), which had its path rectified last century. Because of the proximity to its outlet to the sea, the river had a meandering flow, and the position of a twelfth century (Norman) bridge indicates the different river flow until the twentieth century (Fig. 1). Alluvial sediments 30–40-m thick are found in this area. Moreover, it is important to mention the presence of remarkable thicknesses of anthropic fillings in large sectors of the city, due to ruins from wars and building debris, and due to human activities such as quarries.

A high density stratigraphic–geotechnical database has been collected for the area of Palermo (Giammarinaro et al., 2001; Contino et al., 2006). This database contains data of 3000 boreholes with maximum depth up to 180 m. It was the result of a process that required three main steps: (1) identification and classification of the different lithotypes, (2) georeferencing the available boreholes and storing all the information in a geographic information system (CITYGIS) designed to assess natural hazards in urban areas through specific capabilities and queries, and (3) final data consistency control, exploiting the advantages of the GIS technology.

This GIS allows to reconstruct the geological structure beneath Palermo with high spatial resolution, even though the city is locally characterized by a quickly variable local geology. The presence of several zones interested by thick poorly consolidated sediments overlying stiffer materials causes significant amplification of ground motion during earthquakes as documented by Guidoboni et al. (2003) and Di Giulio et al. (2005). The boundaries of the amplification prone zones have been well constrained in downtown Palermo where we have a high density of stratigraphic data, more than 700 boreholes in a area of 2.4 km² (Fig. 1). In these zones, we distinguish three different geological conditions depending on whether the boreholes find (Al) alluvial deposits, (Sd) recent sea deposits, and (As) alluvial deposits overlain by recent sea deposits (Fig. 1).

Although the detailed information on near-surface geology, in situ measurements of shear-wave velocities ($V_s$) are limited in number for each soil type. Other estimates came from laboratory tests, mostly edometric compression (Giammarinaro et al., 2003). The consistency between measured $V_s$ and indirect estimates through laboratory tests is sometimes not strict. Di Giulio et al. (2005) inferred a range of $V_s$ and $Q$ values for the soils through a comparison between experimental transfer functions from small earthquake recordings and 1D simulations.

The city of Palermo is exposed to moderate seismicity. The Catalog of Strong Italian Earthquakes (Boschi et al., 2000) lists 14 events in the last three centuries that caused
Mercalli–Cancani–Sieberg (MCS) intensity $I \geq V$. Documented damage information in Palermo is available from 1726, when the city was struck by a strong earthquake (1 September 1726, $I_{\text{max}} = \text{VIII–IX MCS}$). At that time, the urban imprint of Palermo was the same as delineated under the Spanish domination of the city, from the early seventeenth century up to the first decade of eighteenth century. The important role of Palermo as capital of the vice reign of Sicily led the Spanish to reinforce the ancient walls, build up new streets and squares, and design a new morphology of the area crossed by the Papiro and Kemonia streams. The urban plan was drastically simplified, with the division in four main areas called mandamenti. There the plan of the city was almost the same as we can recognize nowadays. From numerous buildings still standing from that period, we know that the Spanish built one- to two-storey masonry edifices.

Guidoboni et al. (2003) examined the 1726 earthquake and two other earthquakes (5 March 1823, $I_{\text{max}} = \text{VIII MCS}$; 15 January 1940, $I_{\text{max}} = \text{VII MCS}$) that hit the city, causing serious damage. Their analysis of spatial damage variations and near-surface geology showed that the maximum level of damage was well correlated with the presence of the two bur-

![Figure 1](image-url)
ied riverbeds. While for the 1726 earthquake the vulnerability was uniform, for the other two earthquakes the influence of building vulnerability on observed damage is not so well defined. After the 1726 event, many buildings were poorly repaired and inadequately maintained up until the earthquake in 1823. Moreover, numerous buildings were elevated in subsequent stages up to three to four storeys. The 1940 event occurred in war time, and the macroseismic survey was carried out after the end of the war. Therefore, the seismic damage was disguised by the damage caused by the January–July 1943 bombardments.

On 6 September 2002, an $M_w$ 5.9 earthquake (40 km offshore) caused damage of European macroseismic scale (EMS-98, see Grünthal, 1998) intensity V–VI in downtown Palermo and in its southern sector, the Oreto River valley (Azzaro et al., 2004). Apart from some churches and historical buildings in the city center, few structures were seriously damaged. Greater damage occurred in the Oreto River valley, and even there it was sporadic (Giammarinaro et al., 2005). Aftershocks of the 2002 $M_w$ 5.9 earthquake were recorded by eight temporary stations deployed in the urban area on different geological formations (Calderoni et al., 2005; Di Giulio et al., 2005). The locations of some of these stations are shown in Figure 1.

**Microtremor Survey**

Noise measurements were performed at 90 sites in the urban area of Palermo (Cultrera et al., 2004) during five days of May 2003. We used six seismological stations equipped with Lennartz Marslite digital recorders and Lennartz Le3D-5sec triaxial velocity sensors. To ensure reliable experimental conditions, we followed the guidelines proposed by Koller et al. (2004) in the framework of SESAME (WP02 of SESAME Project, deliverables D01.02, 2002; D08.02, 2003; and D23.12, 2004). The minimum recording time was 30 min. A longer time window was used whenever clear nonstationary sources of noise (cars, pedestrians, markets) occurred. We tried to avoid roads with heavy traffic and underground structures. However, it is important to stress the difficulty in measuring microtremor under controlled experimental conditions in urban environments.

Figure 1 shows the area where microtremor measurements were performed. In downtown Palermo, we recorded along four north-northwest–south-southeast profiles (DD', AA', BB', and EE') crossing the Kemonia and Papireto riverbeds, with spatial sampling from 25–100 m between measurement points. The higher sampling there of noise measurements was chosen to increase the resolution across the sharp geological transitions at the edges of the riverbeds. Measurements were performed in a grid configuration in the confluence area, expected to show the largest spatial geology variations. In addition, twelve microtremor measurements were made along a profile crossing the Oreto River valley (CC'). The CITY-GIS provided by Giammarinaro et al. (2001) allowed us to quickly locate the measurement points taking into account buildings, streets, and borehole information.

Microtremor data were analyzed using software implemented within the SESAME consortium (WP03 of SESAME Project, deliverable D09.03, 2003). We selected the stationary portion of the whole signal using at least 10 windows of 25–60-sec length. Each window of signal was detrended and 5% cosine tapered, fast Fourier transformed (FFT), and then smoothed using the Konno and Omachi (1998) algorithm with a $b$-value of 40. We tried several procedures of smoothing, without significant differences in the results. No filtering before FFT was applied. A check was made to guarantee the absence of anthropic disturbances, that is, monochromatic sources (see Fig. 2). The two horizontal components spectra were combined into root mean square values before dividing by the vertical component spectrum. Finally, the geometric means and standard deviations of the HVNSRs were calculated for frequencies greater than 0.4 Hz, which was chosen to avoid the eigenfrequency of the sensor (0.2 Hz).

Following the SESAME recommendations (WP02 of SESAME project, deliverable D23.12, 2004), we visually selected frequency peaks with amplitude greater than 2. This threshold value was adequate in Palermo, because we found that flat HVNSRs had no clear frequency peaks with amplitude below 2. Further details of the criteria for the selection of the peaks are discussed by Cultrera et al. (2004).

**Spectral Analysis: Correlation between HVNSRs and Geology**

Results show large variability in the shape of HVNSRs. Based on the frequency of the HVNSR peak and its amplitude, the measurement sites were classified into two classes of frequencies (1.4–1.9 Hz and 2.0–2.8 Hz) and three classes of amplitudes (2–3, 3–4.5, and 4.5–6). Figures 3 and 4 show the comparison between HVNSR results and geological setting in downtown Palermo. Geological cross sections in Figure 4 were reconstructed using at least one well log every 90 m, not farther than 10 m from the profile for the Papireto River, and 30–40 m for the Kemonia River. Note that these distances are to be compared with the order 100–150-m widths of the river valleys.

The spatial distribution of the classes involves (1) occurrence of frequency peaks in the areas of the Kemonia and Papireto rivers, and (2) absence of clear peaks, or presence of peaks with HVNSR amplitude below 2 at sites without soft soils at the surface. Peak frequencies within the Kemonia and Papireto riverbeds (profiles AA' and DD') are in class 2 (see Fig. 4, where details of each profile are displayed). From profile DD' in the west to profile EE' in the east, the sediments deepen and the frequency peak values decrease to class 1: we found the lowest values of frequency in the confluence zone (profiles BB' and EE'). A possible explanation is the major thickness of soft soils due to sea and alluvial deposits, but lower wave velocity of the soils may contribute as well.
Moving from the middle to the edges of the riverbeds, we did not find any increase in the resonance frequency. Even where we increased the number of sites to improve the resolution in these zones, the spatial sampling we used was not adequate to map the transition from frequency class 1 to frequency class 2.

The distribution of amplitudes of the HVNSR peaks confirms the relationship with the presence of soft sediments: in this case, the threshold of 2 appears to be appropriate to distinguish stiff and soft sediments. Highest values (amplitude class 3) are limited to few small areas of the confluence zone and close to the harbour (profiles BB and EE, respectively). Overall, the distribution of HVNSR amplitudes within the three classes is very irregular with amplitudes in different classes at closely spaced sites. Beyond the fixed threshold, amplitude of HVNSR peaks is an indicator of the occurrence of a resonance frequency peak but at the moment no theoretical basis exists to attribute a physical meaning to variations of amplitude (see previous studies: Lachet et al., 1996; Satoh et al., 2001). It is worth noting that Di Giulio et al. (2005) found lower amplifications from HVNSR, at the resonance frequencies than from analysis of earthquakes. For example, station PAL5, along the profile BB' (see Fig. 1), has a resonance frequency from HVNSR that, consistently with close measurements, is in frequency class 1. In the earthquake analysis using classical site-to-reference spectral ratios, the amplification of ground motion at PAL5 reaches values up to 10 in the same frequency class.

According to the criterion of Bard and Bouchon (1985), the $V_s$ values proposed by Di Giulio et al. (2005) and the widths of the two river valleys indicate that 2D resonance is possible in areas where the sediment thickness is larger than 20 m. To investigate whether the HVNSRs are representative of the S-wave earthquake amplification, we performed 2D simulations along the AA' profile, and we compared the results with those of the HVNSR analysis (Fig. 5). We used a finite-difference code (Caserta et al., 2002), which computes displacement time series at the free surface using a vertically incident deltalike (Gabor) function as $SH$-plane wave input in the frequency band of 0–7 Hz. To obtain the 2D theoretical transfer functions, we smoothed the amplitude Fourier spectrum of each synthetic seismogram with a 0.5-Hz window and divided it by the corresponding reference spectrum for a half-space model. The geological setting and geometry of profile AA' are shown in Figure 5a. In order to fit the resonance frequencies of HVNSR in the middle of the valleys,
we had to increase the $V_s$ of alluvial sediments used by Di Giulio et al. (2005) by 25% (from 200 to 250 m/sec; see Table 1). Note that this increment in $V_s$ is compatible with the uncertainty in the resonance frequency depicted through HVNSR (Bonnefoy-Claudet, Cotton, et al., 2006). Using this velocity model, the comparison of 2D transfer functions with HVNSR contour maps is shown in Figure 5b,c as a function of frequency and position along the profile. The fundamental resonance frequencies of the 2D simulation reproduce fairly well the HVNSR peaks of the sites in the middle of the riverbeds. Note the higher spectral complexity in the 2D simulation compared to the HVNSR results. Figure 5d shows the comparison of the amplification curves (2D and HVNSR) at sites A6, A3, and A16. For station A6, in the middle of the Papiro river valley, the fundamental frequency from HVNSR agrees with that estimated using 2D modeling. Stations A3 and A16 are very close to seismic stations PAL2 and PAL6, respectively, for which horizontal-to-vertical spectral ratios from earthquakes are available (see Fig. 1). At the edges of the riverbeds (sites A3 and A16), the resonance frequencies from the 2D simulation differ from those obtained by HVNSR. Earthquake transfer function for PAL6 is more similar to HVNSR than the 2D result. In contrast, station PAL2 shows earthquake amplification at high frequencies whereas the HVNSR is almost flat (see Fig. 5d). This result suggests that amplification of earthquake ground shaking is possible also at sites with flat HVNSR (Koller et al., 2004, WP02 of SESAME project, deliverable D23.12, 2004).

Taking into account all these limitations, the uncertainties of profile reconstruction in such a complicated geological framework could lead to bias in the modeling results, such as the shift in the zone of maximum amplitude observed along profile AA’ (Fig. 5b,c). 1D transfer functions (not shown in the figure) indicate slight differences from the 2D simulation. However, our results suggest that (1) within the riverbeds, the HVNSR technique is able to detect the fundamental resonance frequencies predicted by 1D and 2D modeling, whereas it does not show the higher modes of the theoretical transfer functions (see Bard, 1999), and (2) at sites close to the edges of the riverbeds, the differences between HVNSR spectral ratios in earthquakes and numerical transfer functions indicate that the complexity in site response due to the complicated geology cannot be fully represented by 1D or 2D numerical simulations.

During the 6 September 2002 earthquake, buildings in the Oreto River valley (see Fig. 1) suffered scattered instances of damage (Giammarinaro et al., 2005). Significant amplifications in this area (see Fig. 1, station PAL9) were found by Di Giulio et al. (2005). We performed noise measurements along a profile crossing the river (CC’ in Fig. 1). The top panel of Figure 6 shows the distribution of boreholes around the measurement profile. These were too sparse for reconstruction of the geology to adequately represent the complexity of this region (Fig. 1 and Fig. 6, bottom). Indeed, the number of logs is high only in restricted areas close to sites 42, 45, 46, 47, and 48. A further issue is the rectification of the Oreto river during the twentieth century. Topographic variations and the presence of ruins of a Norman bridge (twelfth century) reveal that the original flow of the river was meandering and indicate that its previous path was southward of the present one.

The results of the HVNSR analysis for profile CC’ are shown in Figure 6. Sites from 39 to 44 show similar broad peaks around 1–1.2 Hz with low amplitude values of 2–3, even though they have different geological conditions: the first three sites (39–41) are in an area with 5–8-m-thick layers of sea deposits overlaying calcarenites and silty sands, whereas the others (42–44) are in the alluvial deposits of
Figure 4. Comparison between the HVNSR results and the geological sections aa', bb', dd', ee'. For each noise measurement profile (AA', BB', DD', EE'), representative HVNSR curves, the distribution of the resonance frequency (horizontal dash lines indicate the separation between classes) and the geological cross section are shown. Black vertical lines in the geological cross sections indicate the boreholes.
Figure 5. (a) 2D model of profile AA’. Triangles are measurement sites; crosses and circles indicate the depths from boreholes of the man-made infill and alluvial deposits, respectively. Refer to Table 1 for the seismic parameters of soil 1 (Flysh), 2 (calcarenites), 3 (alluvial deposits), and 4 (fills). (b) Contour plots of HVNSRs along the profile AA’. (c) Response for deltalike SH waves from the 2D model of profile AA’. (d) Comparison of 2D (black line) theoretical transfer functions with HVNSRs (average ±1 standard deviation, gray lines) for sites 3, 6, and 16. For sites 3 and 16, the empirical transfer functions obtained from earthquakes by Di Giulio et al. (2005) for stations PAL2 and PAL6, respectively, are overlain (black dashed line).
the river. Narrow frequency peaks around 1.2–1.8 Hz with large amplitude values (up to 6) occur on the right-hand side of the current river, from sites 45 to 48 (see HVNSR of site 45 in Fig. 6). These peaks are related to low S-wave velocity associated either with alluvium infilling of the river valley or the presence of a quarry filled in historical times with soft materials.

In view of the lack of detailed geological information along the profile, we computed only the 1D transfer functions of vertically incident SH waves (Thomson, 1950; Haskell, 1953) at specific sites using nearby borehole data. The correlation between HVNSR peaks and thickness and S-wave velocity of the underlying materials (Table 1) was investigated in particular at site 45, which is representative also of 46 and 47. Results there show that HVNSR peaks are in good agreement with the fundamental frequency of the 1D transfer function (Fig. 6).

Finally, the last two sites (49 and 50) have broad HVNSR peaks around 1–2 Hz with amplitudes of 3 but with very large standard deviations (see HVNSR of site 49 in Fig. 6).

The joint analysis of frequency peaks and amplification of HVNSRs with subsurface geology outlines the capability of the HVNSR method to reveal the presence of soft sediments in Palermo, especially when there is a strong impedance contrast at the underlying bedrock. This conclusion is particularly important because it is very difficult to achieve the same results in urban areas using other low-cost geophysical techniques. The geological configuration of the Kemonia, Papireto, and Oreto valleys permits the mapping of fundamental frequencies of sites using the HVNSR method and from that the estimation of sediments thickness. Less reliable results are found in zones where the spatial variability of sediments is high, such as at the edges of riverbeds and the transition zones between fluvial and sea deposits. Moreover, the comparison of HVNSR from microtremors with SH-wave simulation of 1D or 2D models shows that HVNSR fails in reproducing the higher modes of the simulated transfer functions. This is commonly observed at soft sites (e.g., see Cara et al., 2003). The amplitude of HVNSR peaks in Palermo underestimates the S-wave amplification both from synthetics and earthquake data, as also observed by Di Giulio et al. (2005). Therefore, the amplitude information can improve knowledge of local geological conditions but cannot be used as a direct estimate of site amplification.

Table 1
Velocity Model Used for the Simulations

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Vs (m/sec)</th>
<th>Density (g/cm³)</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Flysch (bedrock)</td>
<td>1200</td>
<td>2.2</td>
<td>1000</td>
</tr>
<tr>
<td>2 Calcarenites</td>
<td>400</td>
<td>2.1</td>
<td>33</td>
</tr>
<tr>
<td>3 Alluvial deposits</td>
<td>200–250</td>
<td>1.9</td>
<td>10–20</td>
</tr>
<tr>
<td>4 Fills</td>
<td>100</td>
<td>1.8</td>
<td>5–10</td>
</tr>
<tr>
<td>5 Quarry fills</td>
<td>200</td>
<td>1.9</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 6. Results of the HVNSR analysis in the Oreto River valley. Top panel: black squares represent HVNSR without frequency peaks or with amplitude values less than 2; white squares represent HVNSR with significant frequency peaks (class 1 means 1.2–1.8 Hz). Size of squares is proportional to the value of HVNSR peak. Colored dots are the available boreholes in the area, indicating presence (red) or absence (blue) of alluvial deposits. Note the position of the Norman bridge with respect to the current position of the Oreto River. Dashed red line is the trace of the reconstructed geological cross section CC’ (in the bottom panel). Bottom panel: the three HVNSR curves (mean and standard deviations for sites 41, 45, and 49) are representative of the three spectral shapes found west, inside, and east of the Oreto River valley. For site 45, the 1D theoretical transfer function is also shown (dashed line). The geological model used to perform the 1D simulation consists of 7 m of fill and 42 m of alluvial deposits overlaying the bedrock. Refer to Table 1 for shear-wave velocities of the soils.
Statistical Analyses: Correlation between HVNSRs and Damage

A further aim of this study is to verify whether the damage pattern in strong earthquakes can be associated with the HVNSR variations. Because of the scattered nature of the damage in the Oreto River valley during the 2002 earthquake (Giammarinaro et al., 2005) and the absence of information there for past earthquakes, we restrict this analysis to the downtown Palermo area.

As already mentioned, the highest levels of earthquake damage in downtown Palermo are due to the presence of sea and alluvial sediments, and these have been identified using HVNSRs in the present study. In this section we use statistical methods to assess whether and to what extent the occurrence of damage in the city of Palermo is related to the frequency distribution of HVNSR peaks and/or their amplitudes.

In Figure 7, we overlay the results of the HVNSR analysis and the damage distribution of the strongest event analyzed by Guidoboni et al. (2003), that is, the 1 September 1726 earthquake ($I_{\text{max}} = \text{VIII–IX MCS}$). We have translated the damage description of these authors to damage grades of the EMS according to the Grünthal (1998) prescriptions. This operation was possible because the type of construction was quite homogenous at the time of the earthquake, with the vulnerability of the buildings being moderate to high but without strong spatial variations (Table 2).

The presence of a resonance peak in the HVNSRs coincides well with the occurrence of damage and the largest amplitudes of HVNSR peaks are observed close to the zones where the highest grade of damage occurred (Fig. 7). However, the change of the frequency class does not always result in a change of the damage class. This is probably due to the range of the two frequency classes (1.4 to 2.8 Hz), which includes the building resonances of the types of structure in downtown Palermo (three to five storeys).

To quantify the extent of the correlation, for each class of damage in the 1726 earthquake, we calculated the histograms showing the number of sites where the different frequency peaks and amplitudes are found (Fig. 8). For comparison we also show corresponding results for the cumulative effect of the 1726, 1823, and 1940 earthquakes studied by Guidoboni et al. (2003). Some features emerge from this analysis:

1. Most of the HVNSR measurements show insignificant amplification ($<2$, also denoted as frequency $<1$) where damage grades 0 to 2 are observed; this is more evident when considering the cumulative damage of the three earthquakes studied by Guidoboni et al. (2003) (Fig. 8b). However, a frequency peak with amplitude over 6 and a few up to 5 appears for the 1726 earthquake (Fig. 8a) in the same range of damage grades.

2. For the damage class 3, several frequency peaks in the range 1–2 Hz and even more in 2–3 Hz have amplitude...
of 2–4. Of the 28 HVNSR measurements in this class 5 do not show a peaked frequency (amplitude <2, frequency <1).

3. All sites in damage class 4 show a peaked frequency. Most of the observed HVNSR peaks are in the 1–2 Hz class, with amplification of 3–5.

This kind of semiquantitative analysis does not provide a quantitative relation between HVNSR and damage. To give a more detailed statistical interpretation, we applied two multivariate statistical analyses (factor and canonical correlation analyses).

We considered EMS damage grades and HVNSR of the 78 sites in downtown Palermo as elements of a multivariate dataset. For damage we used the values of the 1726 and 1823 earthquakes and the cumulative effect quantified by Guidoboni et al. (2003) as the prevalent maximum level of damage (above 50%) of the three earthquakes in 1726, 1823, and 1940, occurring in buildings within each 100 by 100-m square cell. In our analysis we did not use the damage in 1940 event as a stand-alone variable because of the uncertainties in the macroseismic survey. Similarly, we have not considered damage in the 2002 event because of the generally low level of damage in the downtown area, where macroseismic information is also poor (Giammarinaro et al., 2005). Furthermore, it is important to note that both the damage pattern in the 1823 event and the cumulative effect are significantly driven by the effects of the 1726 earthquake, which was the strongest of the events and before which the buildings had comparatively uniform vulnerability to damage.

For the statistical analyses that follow, the continuous spectra of HVNSRs from microtremor data have been segmented into nine 0.5-Hz-wide frequency intervals, from 0.5 to 5.0 Hz. In each interval, the logarithm of the mean amplitude has been computed. We have checked that these mean-amplitude values for each site have in general a log-normal distribution, which is a requirement for the statistical analyses.

Therefore, each site has 12 variables: nine of them represent the logarithm of the HVNSR mean amplitude in the 0.5-Hz frequency intervals; the other three variables are the damage degree of the two earthquakes and the cumulative damage value of the three events. The whole dataset has been organized in a \((78 \times 12)\) matrix, \(\mathbf{X}_{\text{raw}}\).

The first multivariate approach we used is factor analysis (Davis, 2002). This method is suitable for finding interrelationships among variables. The result of the analysis is a representation of the original dataset as linear combination of uncorrelated new variables, called factors. The elements of the factors \(a_{ij}\) are referred to as factor loadings and define a \((12 \times 12)\) matrix \(\mathbf{A}^{\text{F}}\) in which each column \(j\) represents a factor and each row refers to an original variable \(i\). A description of the factor analysis and the definitions of the associated statistical quantities are given in Appendix A.

Because of the way the factors are calculated, each factor loading is weighted in proportion to the square root of the variance contributed by that variable to the factor. In Figure 9a, the proportion of the total variance accounted for by each factor is shown. Note that the first two factors each accounts for more than 30% of the total variance, and almost 70% when summed together (Fig. 9b). This result strongly indicates that our data reflect the behavior of only two independent factors (factor 1 and factor 2). Therefore, we can reduce the effective size of \(\mathbf{A}^{\text{F}}\) to \(12 \times 2\).

Figure 10 shows the factor loadings of the original variables for both factors. Factor 1 represents very well a direct correlation of HVNSR log amplitudes within the 0.5–3.0-Hz frequency range with all the three seismic damage variables. Factor 2 represents the higher frequency intervals (3.0–5.0 Hz) with no correlation with damage. Note that the slightly higher correlation given by the cumulative damage in factor 1 may be due to an effective reduction in noise.

<table>
<thead>
<tr>
<th>Class</th>
<th>Description of Damage to Masonry Buildings</th>
<th>Color Scale</th>
<th>EMS Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>No damage</td>
<td>Negligible to slight damage (no structural damage, slight nonstructural damage). Hairline cracks in very few walls. Fall of small pieces of plaster only. Fall of loose stones from upper parts of buildings in very few cases.</td>
<td>White</td>
<td>—</td>
</tr>
<tr>
<td>Slight damage; cracks to the plastering, collapse of ornamental or protruding elements</td>
<td>Moderate damage (slight structural damage, moderate nonstructural damage). Cracks in many walls. Fall of fairly large pieces of plaster. Partial collapse of chimneys.</td>
<td>Yellow</td>
<td>Grade 1</td>
</tr>
<tr>
<td>Moderate damage; fissures to inside walls, collapse of vaults or ceilings</td>
<td>Substantial to heavy damage (moderate structural damage, heavy nonstructural damage). Large and extensive cracks in most walls. Roof tiles detach. Chimneys fracture at the roof line; failure of individual nonstructural elements (partitions, gable walls).</td>
<td>Orange</td>
<td>Grade 2</td>
</tr>
<tr>
<td>Serious damage; total collapses, deep fissures, disconnections, leaning walls</td>
<td>Very heavy damage (heavy structural damage, very heavy nonstructural damage). Serious failure of walls; partial structural failure of roofs and floors.</td>
<td>Red</td>
<td>Grade 3</td>
</tr>
<tr>
<td>Very serious damage; total collapse of the buildings or most of it</td>
<td>Destruction (very heavy structural damage). Total or near total collapse.</td>
<td>Purple</td>
<td>Grade 4</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2: Translation of the Damage Color Scale of Guidoboni et al. (2003) into the EMS 1998 Scale
Figure 8. Histograms of occurrence of frequency peaks and related amplitudes as a function of EMS damage grade for (a) the 1726 earthquake and (b) the cumulative damage of three earthquakes (Guidoboni et al., 2003), both computed in 100 × 100 m cells.

Figure 9. (a) Proportion of total variance accounted for by each factor, derived from the corresponding eigenvalue; (b) cumulative variance.
generated by merging the damage datasets from the three earthquakes.

These results can be summarized as follows: (1) damage grades are stable, meaning that different events have the same behavior at each site (Fig. 10a), (2) damage grades are directly related to the lower-medium frequency portion (0.5 Hz < f < 3.0 Hz) of log amplitude of HVNSR (Fig. 10a), and (3) high frequencies (3.0 Hz < f < 5.0 Hz) also have a coherent behavior, but there is no relation with damage (Fig. 10b). Finally, the two main factors represent the metavariables that drive our site behavior both for damage and microtremor.

It is also possible to plot the geographic distribution of each factor (factor scores; see Appendix A). In Figure 11, we show the spatial distribution of factor 1: high values of the factor scores represent both high values of damage grades and high values of log amplitude in the 0.5–3.0-Hz frequency band (Fig. 10), and in particular they delineate the zones of fluvial deposits in the west of the study area.

Factor analysis is effective in exhibiting relationships among variables. It does not provide statistical tests of the significance of results, even though an indication is given by the variance accounted for by each factor. Another limitation of factor analysis is that variables are grouped together with no quantification of their relationships.

To further validate results and to fulfill the interpretation gap left by factor analysis, we performed canonical correlation analysis. A description of this analysis and the definitions of the associated statistical quantities are given in Appendix B. The goal is to relate a set of dependent variables (EMS damage grades) to a set of independent variables (HVNSR mean amplitude in 0.5-Hz-wide frequency bands) by finding linear combinations of the variables that give the highest correlations between the two sets.

Let us call [\(\mathbf{Y}\)] the matrix of damage, composed of 78 observations of 3 variables, and [\(\mathbf{X}\)] the matrix of log amplitudes of 0.5-Hz bands, composed of 78 observations of 9 variables. The aim of this analysis is to find two vectors [\(\mathbf{A}\)] and [\(\mathbf{B}\)] of three and nine elements, respectively, in a way that the linear transformations

\[
\mathbf{Y}_{\text{can}} = [\mathbf{Y}] \cdot [\mathbf{A}] \\
\mathbf{X}_{\text{can}} = [\mathbf{X}] \cdot [\mathbf{B}]
\]

will produce the highest possible correlation between them. Moreover, the variances of the two new sets of variables are constrained to be equal to one, so that the maximized covariances become maximized correlations.

With [\(\mathbf{Y}\)] composed of observations of three variables, the analysis allows to transform the two original variable groups ([\(\mathbf{X}\)] and [\(\mathbf{Y}\)]) into three couples of canonical variables ([\(\mathbf{X}_{\text{can}}\)] and [\(\mathbf{Y}_{\text{can}}\)]; they are 78-element vectors), each having a correlation coefficient \(r\). The significance of the three canonical correlations is assessed by means of the Rao’s scoring test (Rao, 1973). Table 3 shows that, among the three possible canonical correlations, only the highest one (equal to 0.717) is statistically significant, with 0.01% probability of being a correlation without meaning. Figure 12a,b shows...
The vectors \([A]\) and \([B]\) for the couple with this best correlation. They represent the weights used to transform the original variables into the best canonical variables. Consistent with the behavior of factor 1 in factor analysis, low-medium frequencies are directly correlated to damage.

The selected couple of canonical variables for each site \(j\) are plotted in Figure 13a. They are expressed through the combined damage \(Y_{\text{can},j} = \sum_{i=1}^{3} Y_{ji}A_i\), where \(A_i\) is the canonical correlation coefficient relative to the original damage variable \(Y_{ji}\) at each site \(j\) (Figure 12a), and the combined HVNSR variables \(X_{\text{can},j} = \sum_{k=1}^{3} X_{jk}B_k\), where \(B_k\) is the canonical correlation coefficient relative to the original variable \(X_{jk}\) also at site \(j\) (Fig. 12b). The distribution of the elements of \([X_{\text{can}}]\) and \([Y_{\text{can}}]\) is fitted best by the regression line

\[
y_{\text{can}} = 0.72x_{\text{can}} + 1.62.
\]

Figure 13b shows that \([Y_{\text{can}}]\) follows a bimodal distribution; this behavior reflects the distribution of the original damage variables (Fig. 8), which are mostly concentrated in two classes: no or light damage and significant damage. This leads us to define a threshold value of \(y_{\text{can}}\) corresponding to the separation of the two modes and equal to 2.7 (Fig. 13a,b).

This makes it possible to assess whether a site within the studied area of Palermo was likely to have suffered a negligible or important level of damage in the earthquakes considered. From HVNSR of microtremors we can retrieve a value of the canonical variable \(x_{\text{can}}\) for the given site. The corresponding canonical variable \(y_{\text{can}}\) is then computed using the regression line stated previously: if the computed \(y_{\text{can}}\) value is below the defined threshold, we can expect negligible damage, whereas beyond it we expect there to have been an important level of damage. Note that by definition, \(y_{\text{can}}\) does not follow the same scale as the original damage variables: we are not able to use it directly to estimate the EMS damage grade but we can estimate relative damage behavior.

Taking into account the threshold of \(y_{\text{can}}\), we can look at the geographical distribution of this canonical variable.

![Figure 11.](image-url)  
Spatial distribution of factor 1 values. The pattern reflects the position of the two filled riverbeds (compare Fig. 1).

<table>
<thead>
<tr>
<th>Canonical Variable</th>
<th>Canonical Correlation</th>
<th>Rao's F</th>
<th>Numerator Degrees of Freedom</th>
<th>Denominator Degrees of Freedom</th>
<th>Larger F Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7173</td>
<td>2.632</td>
<td>27</td>
<td>193.4</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.3298</td>
<td>0.842</td>
<td>16</td>
<td>134.0</td>
<td>0.6365</td>
</tr>
<tr>
<td>3</td>
<td>0.2713</td>
<td>0.772</td>
<td>7</td>
<td>68.0</td>
<td>0.6130</td>
</tr>
</tbody>
</table>
(Fig. 14). This closely confirms the general result from the factor analysis, of damage and HVNSR being strictly related to the presence of fluvial deposits.

Discussion and Conclusions

In the present study we have investigated the use of HVNSR analysis using microtremors in the city of Palermo to map the fundamental frequency of soils, which depends on the ratio between shear-wave velocity and thickness of the upper soft layers. We have also sought possible statistical correlations between HVNSRs and the damage experienced during moderate earthquakes.

Palermo represents a typical Italian urban context, with very important historical heritage damaged by several earthquakes both in the past and in recent times. Previous investigations, based on a dense set of macroseismic data, indicated that variations of near-surface geology play an im-

Figure 12. Results of the best canonical correlation between damage and HVNSR amplitude variables. (a) Values of canonical correlation vector [A], representing the weights of the damage variables; (b) values of vector [B], representing the weights of the HVNSR variables.

Figure 13. (a) Correlation between the canonical variable [X\text{can}] representing a linear combination of original HVNSR variables and the canonical variable [Y\text{can}] representing a linear combination of the EMS damage grades of the earthquakes studied. Linear fit equation: \(y_{\text{can}} = (0.72 \pm 0.08)x_{\text{can}} + (1.62 \pm 0.14)\); goodness of fit \(R^2 = 0.51\); \(F\) test on regression: \(F = 80.53\) significant at the \(\alpha = 0.01\) level of significance (degrees of freedom 1 and 76). (b) Distribution of values of \([Y_{\text{can}}]\), the bimodal pattern permits the definition of a threshold (vertical straight line) that separates negligible from important damage.
important role in increasing the seismic hazard in the area of downtown Palermo crossed by two buried riverbeds (Guidoboni et al., 2003). This study benefits from abundant data on local geology and damage distribution. A very large number of boreholes and geological surveys were respectively collected and conducted by the University of Palermo (Giammarinaro et al., 2001), and these data have been used in this study to retrieve very local information and reconstruct the geological structure beneath the study area.

Ninety microtremor measurements were performed along several profiles crossing the soft sediment bodies. The results indicate that the ambient noise analysis can be an efficient tool for exploring soil conditions in urban areas under the hypothesis that the velocity contrast between soft sediments and the bedrock is large. In particular, we have assessed the link of HVNSR with soft soil characteristics (thickness and shear-wave velocity), comparing the HVNSR resonance frequencies with the results of 1D and 2D numerical models. Although the amplitude of HVNSR in itself is not a measure of local transfer function amplitude, the pattern of peaked frequencies and amplitudes of HVNSRs are successful in recognizing the zones in Palermo where sediment amplification occurs.

Moreover, we have investigated the correlation between HVNSR and the damage distribution from past earthquakes. This comparison is generally biased by several factors that affect the damage pattern, such as the influence of the earthquake source, the weakness of the existing buildings, and vulnerability changes following a previous earthquake. However, in the case of downtown Palermo, Guidoboni et al. (2003) showed that the damage distribution is mostly controlled by the presence of soft deposits, given that the vulnerability originally was essentially uniform over a large area. We then applied two statistical tools (factor analysis and canonical correlation) to retrieve a quantitative correlation of the resonance frequency and the amplitude of HVNSRs with the record of damage. The factor analysis leads to a simplification of the problem, combining the original variables into only two factors that in our case appear to be significant. One is the logarithm of HVNSR mean amplitude in the low–medium frequency range (0.5–3 Hz) that is directly correlated with the level of damage, the second one is the logarithm of HVNSR mean amplitude in the high frequency range (3–5 Hz) that does not show any correlation with the level of damage. In general, the building resonance frequency ($f_r$) depends on the number of floors. In Palermo, we have mainly three–five-storey buildings and we expect an $f_r$ of about 2–3 Hz. This is confirmed by Di Giulio et al. (2005), who found similar consistent values of $f_r$ (2–3 Hz) by analyzing simultaneous microtremors at the top and at the bottom of some buildings in downtown Palermo. Therefore, the low–medium frequency range includes the resonance frequency of the predominant building type in Palermo.

After the factor analysis, the canonical correlation analysis gave further insights with a statistical test on the significance of the correlation by grouping the variables according to their mutual relationships. For our problem, this approach constructed a pair of related canonical variables which are a linear combination of the HVNSR mean amplitudes and a linear combination of the EMS damage grades, respectively. Their correlation is statistically significant. The distribution of the damage canonical variable follows a bimodal distribution, like that observed for the original damage values from each earthquake, mostly concentrated into two classes (no or light damage and significant damage). The threshold between the two sets of damage values led us to separate the HVNSR canonical variable values into two groups, which correspond to either negligible or important level of damage during the earthquakes considered. In principle, by extending the HVNSR measurements to other sites where we do not have damage information, we can assess the probability that those sites are exposed to significant damage. It has to be underlined that the validity of the proposed statistical methodology is constrained by the damage in the earthquakes used for the analysis and reflects the vulnerability conditions at the times of the earthquakes. Therefore, hazard considerations can be translated into risk only if vulnerability conditions do not change in time. Moreover, the larger the number of the macroseismic datasets, the more reliable the relationships between the canonical variables.

In conclusion, we have established a statistical link of HVNSRs with soft soil characteristics and damage distribution in downtown Palermo. Although there is no theoretical basis to assume that amplitude of HVNSR is related to real amplification, we find a positive correlation between amplitude of the HVNSR peaks and level of damage. A possible

![Figure 14. Spatial distribution of canonical variable $Y_{can}$. White and black dots represent values respectively below and above the damage threshold (2.7). As in Figure 11, the canonical analysis confirms the high correlation among local geology, observed damage and the HVNSR results.](image)
explanation could be that sites characterized by the largest amplification during earthquakes also show the highest amplitudes of HVNSR.

Several peculiarities of the case study of Palermo helped in this analysis: (1) the resonance frequency of soils estimated from HVNSR and the damage distribution are well controlled by site amplifications due to the presence of soft sediments; (2) the vulnerability was substantially uniform in the analyzed area when the strongest earthquake occurred; and (3) exhaustive macroseismic datasets were collected for more than one earthquake. The quantitative and robust correlations found in this study are limited to downtown Palermo and cannot be extrapolated to assess the damage distribution in other areas with different site conditions. However, an approach combining damage reports and microtremor measurements seems to be potentially successful, suggesting interesting low-cost perspectives in the practice of territory planning and zoning in urban areas.

Acknowledgments

This work was performed within the framework of the SESAME project, supported by the Environment and Sustainable Development Program of the European Commission Research Directorate General (Contract Number EVG1-CT-2000-00026). We thank the students of Palermo University, who helped us during the experiment. Special thanks are due to Professor John Haines of Cambridge University for reading and checking the text and to the two anonymous reviewers of the manuscript. Their constructive comments considerably improved many parts of this article.

References


The factor analysis is a multivariate approach suitable to find mutual relation among variables. The original dataset is defined as the ensemble of n observations of m variables, arranged in an \( (n \times m) \) matrix \( [X_{\text{raw}}] \). In factor analysis, the common assumption is that all m variables of the original dataset descend from a smaller set of \( p \) independent factors (\( p < m \)).

We followed the principal component approach to factor analysis which consists of the following steps. First, the original raw data \( [X_{\text{raw}}] \) are standardized by subtracting the means and dividing by the standard deviations related to each variable, obtaining the matrix \( [X] \). Second, the standardized variance–covariance (correlation) matrix \( [S^2] \) is calculated; it is symmetric and the diagonal elements are equal to 1, that is standardized data \( [X] \) have zero means and unit standard deviations. Third, eigenvectors and eigenvalues are extracted from \( [S^2] \); the eigenvectors define the matrix \( [U] \), and the square roots of eigenvalues define the diagonal matrix of singular values \( [\Lambda] \); the sum of the eigenvalues represents the total variance (information content) of our data. Fourth, the eigenvectors are converted into factors by using the relation

\[
[A^R] = [U] \cdot [\Lambda],
\]

where \( [A^R] \) is an \( (m \times m) \)-sized matrix, each column representing a factor whose elements are referred to as factor loadings. Because of the way the factors are calculated, each factor loading is weighted proportionally to the square root of the amount of variance contributed by that variable to the factor. Looking at the proportion of variance accounted for by each factor, it is possible to select a smaller number of factors \( (p) \) accounting for most of the total variance in the dataset. The independence of all factors is guaranteed by orthogonal decomposition of the correlation matrix, given by the Eckart–Young theorem, through the extraction of eigenvalues and eigenvectors. Therefore, we can reduce the size of \( [A^R] \) to \( m \times p \).

Factor analysis results in factor scores \( \hat{S}^R \), which represent the values of any single factor for every observation, as follows:

\[
\hat{S}^R = [X] \cdot [A^R] \cdot ([A^R]^T \cdot [A^R])^{-1}.
\]

In this way it is possible to associate with each observation the value of a factor score related to each one of the \( p \) factors. Adding the geographical coordinates, we obtain the distribution over the territory of these factor scores.

Appendix B

Canonical Correlation Analysis (Davis, 2002)

The goal of the canonical correlation analysis is to relate a set of dependent variables \( [Y] \) to a set of independent variables \( [X] \) by finding linear combinations of the variables that give the highest correlations between the two sets. The matrix \( [Y] \) consists of \( n \) observations of \( i \) variables, and \( [X] \) of \( n \) observations of \( j \) variables. The matrix of variance–covariance of the composition of both groups of variables can be expressed as...
The linear transformations of variances and covariances of \( i \) original variable groups into \( i \) couples of canonical variables can be found by maximizing the correlations. If we solve the equation for the correlation between the two canonical variables \( X_{can} \) and \( Y_{can} \), we obtain the vector \( \mathbf{B} \) as

\[
\mathbf{B} = \frac{\mathbf{B}^T \mathbf{S}_{xy}}{\sqrt{\mathbf{B}^T \mathbf{B}}},
\]

where \( \mathbf{B} \) is a \( n \times 1 \) vector of the canonical correlations, \( \mathbf{S}_{xy} \) is the \( n \times n \) matrix containing the covariances between \( X \) and \( Y \), and \( \mathbf{S}_{xx} \) is the \( n \times n \) matrix containing the variances of \( X \) and \( Y \).

Moreover, the variances of the two new sets of variables are constrained to be equal to one, that is

\[
\mathbf{A}^T \mathbf{S}_{xy} \mathbf{A} = 1, \quad \mathbf{B}^T \mathbf{S}_{xx} \mathbf{B} = 1,
\]

where \( \mathbf{B}^T \) is the transpose of vector \( \mathbf{B} \).

In such a way the maximized covariances become maximized correlations. If we solve the equation

\[
|\mathbf{A} - \lambda \mathbf{I}| = 0,
\]

where \( \mathbf{A} = \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{xy}^{-1} \) is \( i \times i \) sized, we find \( i \) eigenvalues \( \lambda_i \) that are numerically equal to the square of the correlation between the two canonical variables \( X_{can} \) and \( Y_{can} \). Substituting them into

\[
(\mathbf{A} - \lambda \mathbf{I}) \mathbf{B}^T = 0,
\]

we obtain the vector \( \mathbf{B} \).

The vector \( \mathbf{A} \) is found as

\[
\mathbf{A} = \mathbf{S}_{xy}^{-1} \cdot \mathbf{B} \sqrt{\lambda}^T.
\]

Substituting the three values of \( \lambda \) into relation (B5) and (B6) and then the resulting \( \mathbf{A} \) and \( \mathbf{B} \) into (B3) we transform the two original variable groups into \( i \) couples of canonical variables \( \{X_{can}\} \) and \( \{Y_{can}\} \), \( n \)-elements vectors having correlation \( r = \sqrt{\lambda} \). Note that we have reduced the size of the problem to the minimum number of variables in the two sets of original variables \( \{X\} \) and \( \{Y\} \), equal to \( i \).

Moreover, we can investigate the significance of the three canonical correlations by means of specific statistical tools. In particular we use Rao’s scoring test (Rao, 1973), which gives the probability to randomly have a score \( F \) (associated to the respective degrees of freedom), that is the probability to have a correlation with no meaning. This test allows to select the best couple of canonical variables whose canonical correlations are statistically significant.