the expected earthquake to the accelerogram of the ground motion reaching the location where the structure is aimed at. Schematically, such mapping would enable the establishment of the relation described in Figure 1.

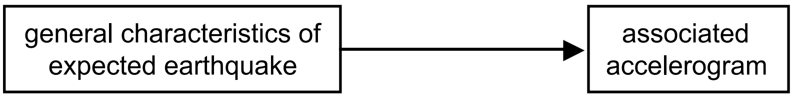

Figure 1 - High level strategy for the undertaken approach

For engineering purposes, the most relevant data describing expected earthquakes can be given by a few parameters such as the magnitude and location. A corresponding accelerogram measured at a given distance of the epicentre is generally described in a more extensive manner by a diagram as the one of Figure 2.

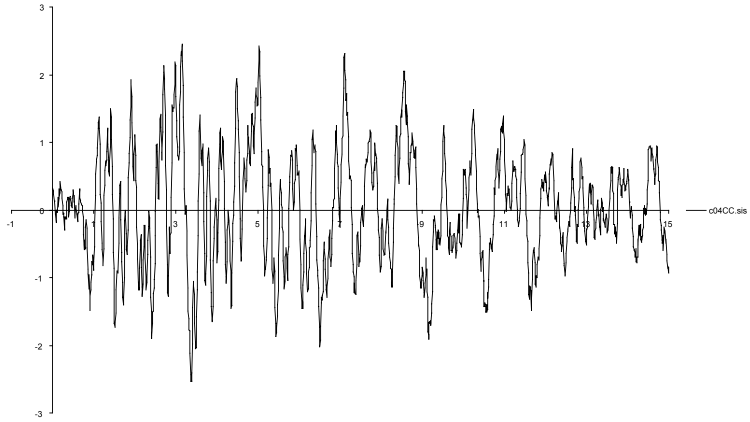

Figure 2 - Typical accelerogram

MAPPING EARTHQUAKES INTO ACCELEROGRAMS

In the absence of models capable of describing such mapping, it is possible to attempt inferring it by means of existing data about past occurrences. The most obvious possibilities would be to consider 1) statistical approaches or 2) the use of machine learning methodologies. However, statistical approaches would require assumptions about the functional form of such mapping, for which there exists no sound knowledge.

A promising alternative way for such endeavour would be the use of a supervised training technique that would be based in descriptions of existing data about past earthquakes and their respective accelerograms. Such was the adopted approach, by endorsing the use of artificial neural networks.

In order to proceed with the training of networks eventually capable of providing the required mapping between expected earthquakes and the corresponding accelerograms, it would be necessary to identify the attributes that could characterise appropriately both an earthquake and an accelerogram. Both were required to be described synthetically, in the sense that the aim was to feed them as input or output nodes of the networks to be trained. Moreover, it is well known that earthquakes can be generically characterised by a few parameters, such as, for example, *magnitude*, *duration*, *distance*, etc. As for accelerograms, they are typically described by very extensive means, taking, generally, the form of a very high number of *time-acceleration* pairs, rendering this type of description inappropriate for representation in a neural network input layer.

Consequently, it is necessary to process existing accelerograms in order to synthesise some of their main characteristics. This may be done resorting to Fourier analysis, as discussed ahead.

It is, thus, possible to establish a scenario in which the problem at hand can be represented as schematically depicted by Figure 3.

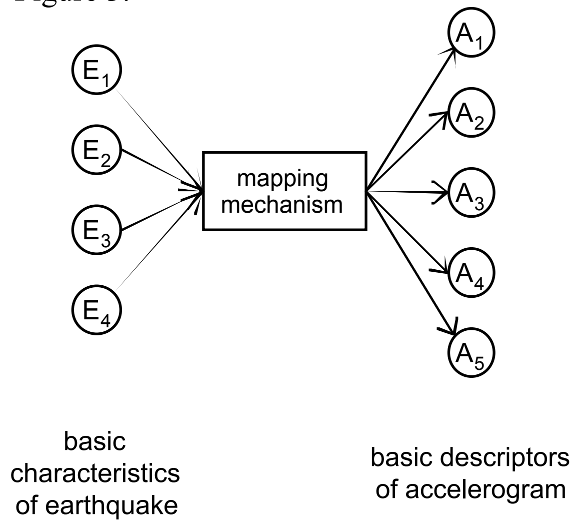

Figure 3 - Conceptual mechanism for mapping between characteristics of earthquakes and associated descriptors

In this figure, E_1 represents the *magnitude*, E_2 the *epicentral distance*, E_3 the *focal depth* and E_4 the *type of soil*. On the other handA1 denotes the *peak acceleration*, A_2 the *effective length* (duration), A_3 the *predominant frequency*, A_4 the *radius of gyration of the power spectral density function* and A_5 the *frequency value for which the power spectral density function reaches its maximum*

FROM EXPECTED EARTHQUAKES TO EXPECTED ACCELEROGRAMS

The process briefly described by Figure 3, when successful, provides a means for computing a number of basic descriptors of accelerograms (A_i) as a function of the basic characteristics of the expected earthquake (E_i). Since a given A_i set is not in itself the expected accelerogram, it is, then, necessary to convert such set of parameters into an artificially generated earthquake.

The whole process may be schematically depicted as in Figure 4.

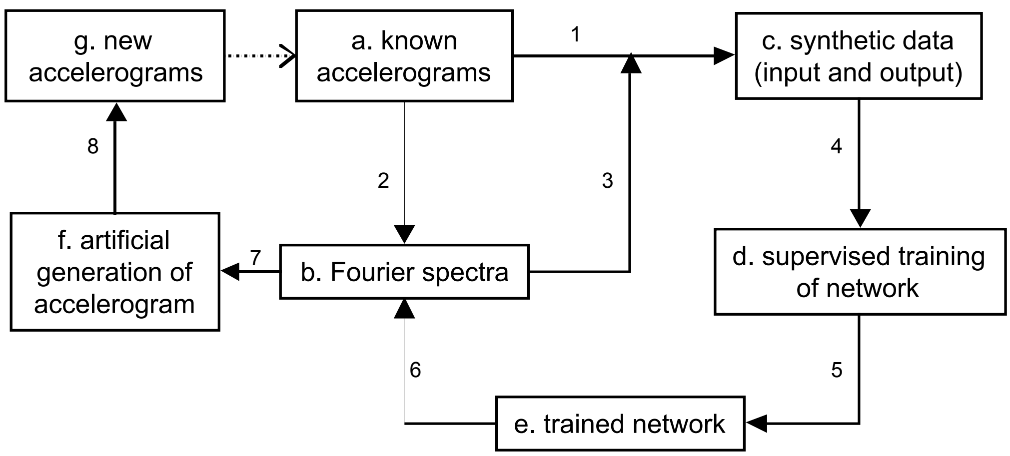

Figure 4 - Overall strategy

DATA GATHERED FOR TRAINING

Given the experimental nature of the present venture, it was decided to resort to a relatively small number of records showing a reasonable level of uniformity. For the present exercise to get a broader scope of application it will b necessary to carefully select a large and significant number of strong motion records.

Given the availability in terms of quantity and of quality of strong motion records, Southern and Central California were chosen as the target area.

Because no single source of data was found to process all the relevant data, both the *Earthquake Strong Motion* (NGDC, 1997) and the *Seismicity Catalog* CD-ROMs (Whiteside et al., 1996) had to be used.

Some of the data collected from the latter source referred to the date (year, month, day, hour, minutes and seconds), the coordinates (latitude and longitude), the focal depth and magnitude. *Table 1* illustrates the result of a query to the source, containing the information described, for a given geographical polygon previously defined.

Table 1 - Result of query to the Seismicity Catalog

```
                                …
1970  1  1 5154179 37N1459121W4311    272  13   8
1970  1  113 11705 37N1449121W4304    235   5   3
1970  1  121583853 36N4623121W2282    952   9  10
1970  1  122201370 36N4681121W2286    817  14  22
1970  1  2 0 24186 36N4621121W2272    948  11   9
1970  1  2 1403440 36N4560121W2404    599  10   8
1970  1  2 3204244 36N4598121W2228    922   9   7
1970  1  2 3503078 36N4637121W2326    778  12   6
1970  1  2 8465093 37N2045121W4293    371  10   3
1970  1  211553609 36N4625121W2239   1203  10   5
1970  1  219234890 37N1937122W 577      0  16  16
1970  1  3 0 51002 36N5661121W4055    868  13   5
1970  1  3 3213048 37N1920122W 476    484  13   7
1970  1  3 3313252 37N1891122W 422    489  14   9
1970  1  3 4542752 37N1863122W 358    476  12   7
1970  1  618151511 37N1776121W5072      0   7  10
1970  1  619263362 37N1947122W 613      0  14  18
1970  1 1114495033 36N 513120W5969    961  18  22
1970  1 1219292369 37N1959122W 642      0  14  22
1970  1 1221204555 37N5188122W  77   1171  13  13
1970  1 13 3284434 35N4835121W2042    448  24  24
1970  1 14 9175312 37N3270121W5172    424  12   5
1970  1 1419271305 37N1934122W 590      0  16  18
1970  1 16 630 824 37N3285121W5182    602  15   3
1970  1 1811305467 37N2249121W4401    663  17   9
1970  1 2219243646 37N1934122W 624      0  11  19
1970  1 251917 466 37N 883121W5619   1439  12  1
1970  1 31 5392348 36N3252121W 451   1160  15  14
1970  1 31 8553024 36N5212121W3039     99   9   7
                                …
```

Since the *Earthquake Strong Motion* source does not provide a query mechanism, the task of matching between records of the first source and the equivalent ones in this one was very complex. Furthermore, most of the matching records were not complete, reason why most of them had to be abandoned. It was eventually possible to gather a set of 71 complete records, which was considered sufficient for the exercise at hand.

As a matter of illustration, Table 2. describes a typical description found in this source.

Table 2 - Sample of strong motion record source data

```
CORRECTED ACCELEROGRAM  IIS257   71.177.    COMP N82W   FILE  8 DERIVED FROM
FILE   8  OF UNCORRECTED ACCELEROGRAM DATA OF VOL. I-S, EERL 73-24
SAN FERNANDO EARTHQUAKE
FEBRUARY 9, 1971 - 0600 PST
IS 257 - 71.177.0 EN                                                    20
STATION NO. 445    34 03 47N, 118 21 43W                                40
6200 WILSHIRE BOULEVARD, 17TH FLOOR, LOS ANGELES, CAL.                  54
COMP N82W                                                               9
SAN FERNANDO EARTHQUAKE   FEB 9, 1971 - 0600 PST                        48
EPICENTER  34 24 00N, 118 23 42W                                        32
INSTR PERIOD % 0.030 SEC   DAMPING % 0.592    SENSITIVITY % 1.57 CM/G   69
NO. OF POINTS % 1922  DURATION % 40.942 SEC                             43
UNITS ARE SEC AND G/10                                                  22
RMS ACCLN. OF COMPLETE RECORD %  0.6300 G/10                            42
ACCELEROGRAM IS BAND-PASS FILTERED BETWEEN   0.125  AND  25.000  CYC/SEC
  2048  INSTRUMENT AND BASELINE CORRECTED DATA
AT EQUALLY-SPACED INTERVALS OF   0.02   SEC.
PEAK ACCELERATION %  -248.94943  CMS/SEC/SEC AT   7.0600 SEC
PEAK VELOCITY     %    39.89056  CMS/SEC     AT   6.8600 SEC
PEAK DISPLACEMENT %    16.44511  CMS         AT   7.1000 SEC
INITIAL VELOCITY  %     2.21538  CMS/SEC     INITIAL DISP. %   -0.47811  CMS
SAN FERNANDO EARTHQUAKE   FEB 9, 1971 - 0600 PST

IIS257 71.177. 6200 WILSHIRE BOULEVARD,17TH FLOOR,LOS ANGELES,CAL.  COMP N82W
    8        1       19     257       71   177       0       6     445      34
    3       47      118      21       43    34      24       0     118      23
   42        2        9    1971      600     0     278    1922      24      54
    0        0        0       0        0     0       0       0       0       0
    0        0        0       0        0     0       0       0       0       0
 1922     1922     2048       2       10    10       1       0      48      85
   10       10        2    1024        5   410       0       0       0       0
    0        0        0       0        0     0       0       0       0       0
    0        0        0       0        0     0       0       0       0       0
    0        0        0       0        0     0       0       0       0       0
0.03000 0.59200 40.94199 0.63000  0.10000  0.0     0.0      0.0     0.0     0.0
0.0      0.0     0.0      0.0      0.0     0.0     0.0      0.0     0.0     0.0
0.0      0.0     0.0      0.0      0.0     0.0     0.0      0.0     0.0     0.0
0.0      0.0     0.0      0.0      0.0     0.0     0.0      0.0     0.0     0.0
0.0      0.0     0.0      0.0      0.0     0.0     0.0      0.0     0.0     0.0
1.000  98.0665  0.0100  40.94199 209.43948 1.00000 1.0000 27.00000  2.0000 40.93999
0.02000 0.12500 0.02000   0.0  7.06000 -248.9494  6.8600  39.89056 7.10000 16.44511
2.21538 0.1250 25.0000 0.20000 0.20000  -0.47811  6.84000 39.78506 7.10000 16.44511
0.0      0.0     0.0      0.0      0.0     0.0      0.0     0.0      0.0     0.0
0.0      0.0     0.0      0.0      0.0     0.0      0.0     0.0      0.0     0.0

 (2048 aceleration points (cm/seg2) - fixed time interval of 0.02 s)
 240     181    119     19      -86     -149     -187   -176     174      -12
-176    -156   -102     -4       16       29       24     -2     -55      -98
-133    -186   -259   -334     -409      -18      -74   -263    -150       -7
-35      -13    -42    -64       -4       24      115    188     221        2
278      276    258    230       23      189      194     37     303       32
                                    …
(1024 velocity points (cm/seg) - fixed time interval of 0.04 s)
 2.215   2.936   3.006   2.433    1.720    1.008    0.419   0.246   0.343   0.308
-0.074  -0.838  -2.173  -3.793   -4.844   -5.173   -5.278  -5.497  -5.382  -4.670
-3.636  -2.549  -1.627  -0.851    0.120    1.476    2.904   4.453   6.020   7.451
 8.379   8.464   8.076   8.588    9.897   11.389   12.298  10.641   6.945   3.942
 2.772   1.979   1.460   1.441    0.628    -2.12   -4.923  -6.188  -5.203  -1.793
                                    …
(410 displacement points (cm) - fixed time interval of 0.10s)
 4.513   4.192   3.838    .297   2.865   2.498   1.810   0.736   -0.333   -1.014
-1.341  -1.453  -1.353  -1.261  -1.431  -2.023  -3.085  -4.280   -5.348   -6.176
-6.794  -7.075  -6.997  -6.700  -6.546   6.676  -6.839  -6.855   -6.827   -6.650
-5.998  -4.951  -3.725  -2.443  -0.999   0.574   2.115   3.756    5.707    7.762
                                    …
```

With the above information at hand, it was necessary an intense pre-processing activity in order to establish the remaining required parameters: effective strength, type of soil, predominant frequency and radius of gyration of the power spectra, as well as the maximum spectral acceleration.

PRE-PROCESSING OF DATA

A first concern relating to pre-processing of data was the conversion of the existing coordinates to longitudinal and transversal components. The vertical component was abandoned and, at a later stage of training with this data, it became clear that the only relevant component was, for this purpose, the longitudinal one.

A brief description of the parameters extracted or computed for training of the network follows.

Epicentral distance and focal depth

The region of the globe where the earthquake occurs receives the name of *focus* or *hypocentre*. Its projection to the surface of the earth is called *epicentre*. The *epicentral distance* is, therefore, the distance between the epicentre and the station where the motion was recorded, and constitutes one of the attributes that was used to characterise each earthquake. The value of this attribute is usually given in *km* and may be calculated as indicated bellow.

$$\text{epic.dist.} = \sqrt{\left[\left(\text{Lat}_{\text{epi}} - \text{Lat}_{\text{st}}\right) \times 111\right]^2 + \left\{\cos\left[\left(\text{Lat}_{\text{epi}} + \text{Lat}_{\text{st}}\right)/2\right] \times \left(\text{Long}_{\text{epi}} - \text{Long}_{\text{st}}\right) \times 111\right\}^2}$$

The *focal depth* is no more than the actual depth of the hypocentre, and is usually given in *km*. This parameter was directly extracted from the data source.

Magnitude

The *magnitude* is a normalised measure of the amplitude of the earthquake waves measured at a given distance of the epicentre. There are several scales of magnitude, but the Richter scale is the most commonly used and provides a 1 to 10 scale. For the present study the values of the magnitude were directly extracted form the data source.

Type of soil

The description of the soil in terms of a class or type is a common procedure in earthquake engineering. This parameter is of utmost importance, because the type of soil where a structure is located is extremely relevant for the definition of the ground motion felt at its base. For the consideration of this parameter it was, thus, necessary to have a classification of the soil where each base station was located. Since the available data sources were not only incomplete, regarding this parameter, but were also using different classification scales (e.g. *soil/rock* or *rock/stiff/soft*) for different cases, it was necessary to complete and harmonise the available classifications to a simplified one considering only two soil classes – *soil* or *rock.*

Peak acceleration

The *peak ground accelerations* may be directly obtained from the record of the ground motion, after conversion to longitudinal and transversal components. As it will be discussed ahead, after experimentation it was decided to use only the longitudinal component of the acceleration, reason why the peak acceleration will refer to the longitudinal component only.

Effective duration

The effective duration is the period corresponding to occurrence of the higher components of the motion; it provides a better description of the duration of the ground motion, for it removes the initial and final parts of the accelerogram, during which the values of the acceleration show lesser significance. For the effective duration to be computed, it is necessary to define an accumulated function of the acceleration that is then used to establish the beginning and the end of the *effective* period – usually set to the first and the final 5%, as described by *Figure 5*:

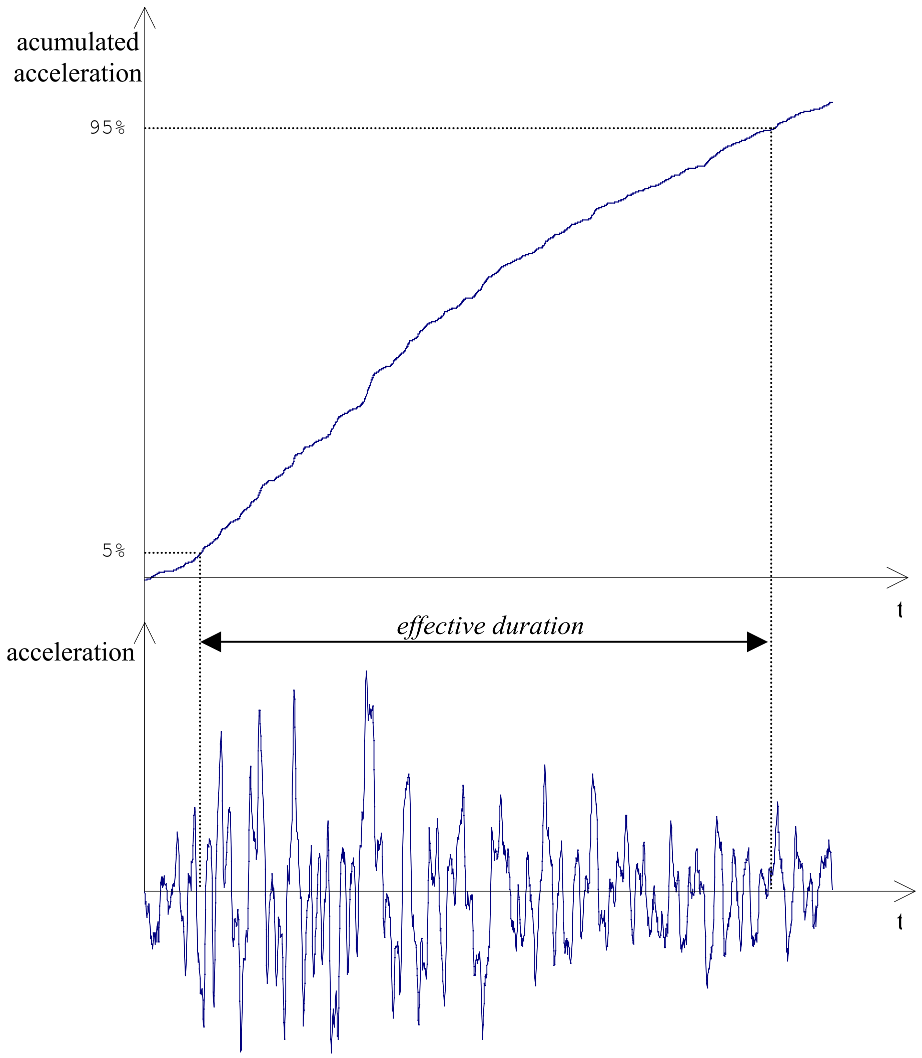

Figure 5 - Effective duration of an earthquake

Predominant frequency

The description of an earthquake in the time domain may be adequately given by an accelerogram, as already seen and depicted. However, in order to describe an earthquake in the frequency domain, i.e., in order to describe the contents of an earthquake in each of its constituent frequencies, it is necessary to resort to a function of spectral density usually named *power spectrum*, like the one of *Figure 6*. Therefore, the attribute used for training called *predominant frequency* is obtained from the spectral density function that had to be computed for each of the records used for training. It represents the central value of that function.

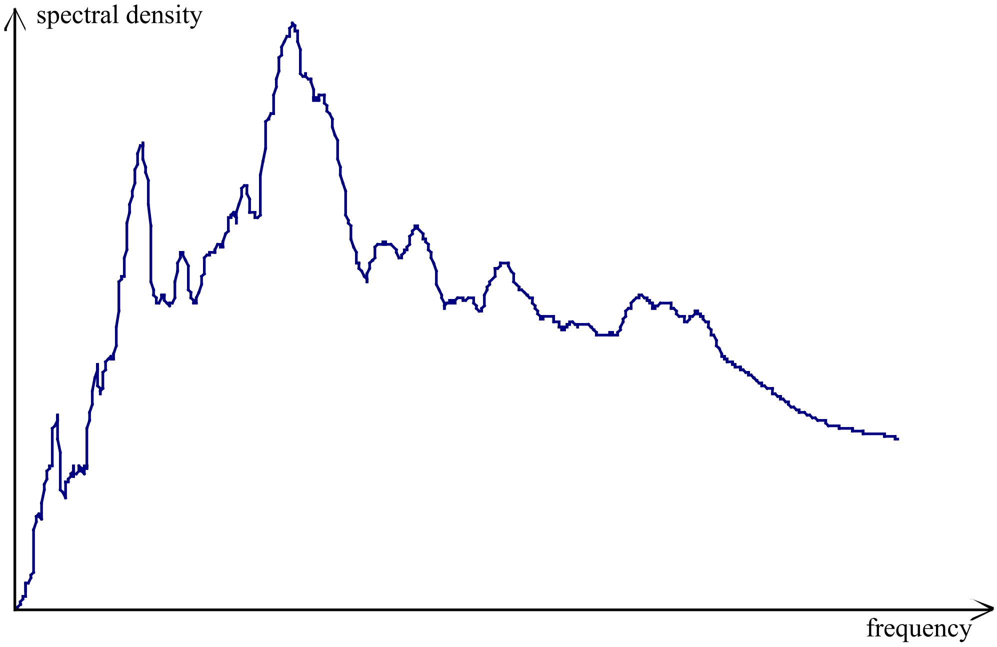

Figure 6 - Sample spectral density function

Radius of gyration of the power spectrum

Because the shape of the power spectral density function is a relevant factor for the generation of artificial accelerograms, it was necessary to provide a measure of the concentration or dispersion of the frequency contents around the central value, i.e. around the main frequency.

Maximum spectral frequency

Because sometimes the central frequency of a strong motion does not coincide with the central value, the *maximum spectral frequency*, which obviously represents the maximum of the spectral density function, was also computed and taken into account for the training of the networks. This parameter allows for taking into account some eventual skewing of the spectra around the central value.

TESTED ARCHITECTURES

With the various attributes characterising the earthquakes adequately acquired or computed, it was possible to test a number of different architectures of artificial neural networks for training. It was possible to vary the number of input and output attributes, both for reasons related with the belief that a given attribute could improve the quality of the network, or simply on the grounds of numerical convenience (i.e., having more or less nodes at a given value).

On the other hand, it was possible to experiment with one or more hidden layers and it was considered relevant to try various numbers of nodes in each of such layers, as depicted by *Figure 7*.

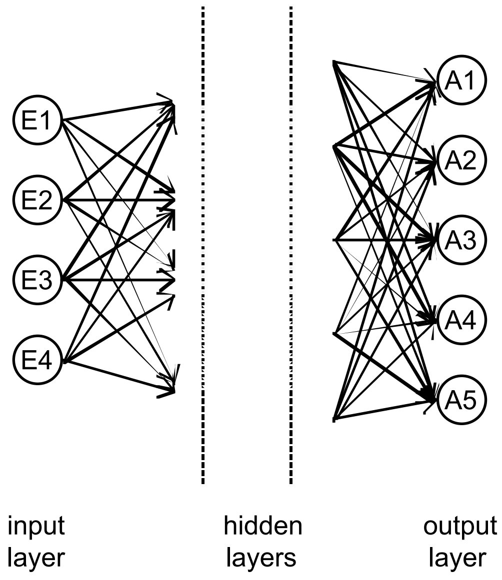

Figure 7 - General architecture used for training (input and output nodes fixed)

LEARNING RESULTS

The described feedforward network was trained using a standard backpropagation algorithm for a set of 64 training vectors. 7 vectors were set aside for testing through the learning phase.

For the training set, after convergence of the network, the results obtained may be considered particularly good for the peak acceleration, effective duration and maximum spectral density, with average errors of 5%, 7% and 5%, respectively. A higher dispersion was found for the predominant frequency and the radius of gyration of the power spectra that have shown average errors of 10.9% and 10.1%, respectively.

When considering the test set, most of the attributes were satisfactorily predicted for the whole set, as described by *Table 3* and *Figure 8*, where all values are non-dimensional and have been scaled to the interval [0,1].

Table 3 - Test values and average errors for test set

Peak acceleration		**Effective duration**		**Predominant frequency**		**Radius of gyration**		**Max. spectral density**	
Real	Predicted	Real	Predicted	Real	Predicted	Real	Predicted	Real	Predicted
0.350	0.484	0.174	0.165	0.028	0.205	0.372	0.430	0.141	0.238
0.365	0.487	0.153	0.272	0.149	0.068	0.359	0.288	0.060	0.093
0.458	0.464	0.459	0.491	0.121	0.080	0.164	0.086	0.000	0.027
0.469	0.431	0.271	0.208	0.167	0.194	0.392	0.315	0.000	0.009
0.504	0.501	0.736	0.269	0.037	0.122	0.053	0.298	0.045	0.082
0.397	0.433	0.109	0.079	0.111	0.461	0.603	0.629	0.036	0.125
0.359	0.469	0.086	0.166	0.067	0.221	0.230	0.445	0.130	0.086
average error	**6%**		**11%**		**13%**		**10%**		**4%**

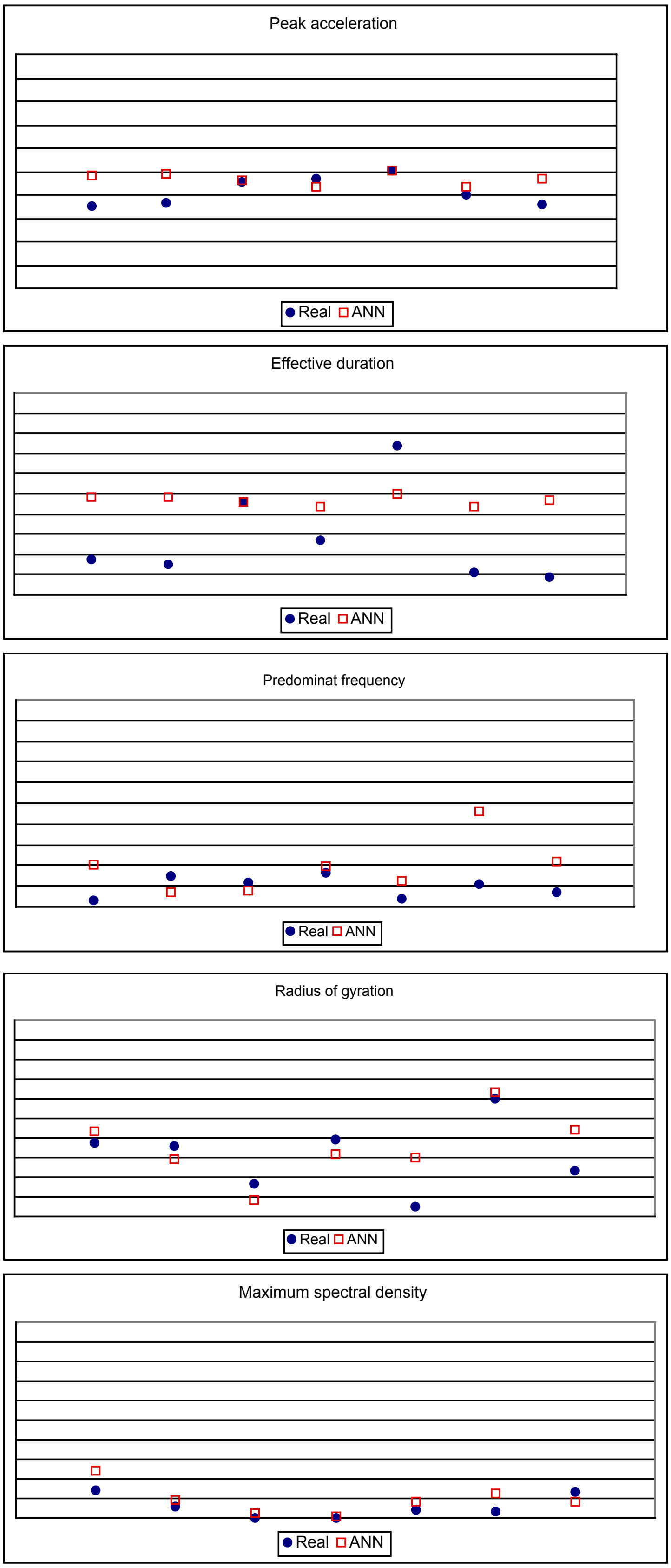

Figure 8 - Real vs. predicted results for test set

EVALUATION OF RESULTS FOR STRUCTURAL ENGINEERING PURPOSES

Artificial generation of accelerograms

Once the basic descriptors (A_1 to A_5) used to artificially generate an accelerogram corresponding to a given earthquake have been obtained, the generation procedure is as follows:

First, given descriptors A_3 to A_5, a power spectral density function is defined. A possible way is to use a combination of Gaussian functions whose mean value corresponds to descriptor A_3, whose standard deviation is related to descriptor A_4 and whose maximum value corresponds to descriptor A_5.

Based on discrete values ($PSD(w_k)$) of this power spectral density function, stationary time series ($X_{st}(t)$) can be generated using the following expression

$$X_{st}(t) = \sum_{k=1,n} \sqrt{2 \;\; \Delta f \;\; PSD(w_k)} \;\; \sin(2\pi w_k t + \phi_k)$$

where

Δf is the discretization interval in terms of frequencies

ω_κ are the frequency values for which PSD was discretized and

ϕ_κ are phases associated to each harmonic (usually assumed as random).

To obtain a synthetic accelerogram ($X(t)$) the obtained stationary accelerogram needs to be modulated using the information given by descriptors A_1 and A_2.

For that purpose, the stationary signal is multiplied by a modulating function $F_M(t)$ such that

$$X(t) = X_{st}(t) \;\; F_M(t)$$

where $F_M(t)$ can be assumed to be a Beta function of the type

$$F_M(t) = \left(\frac{t}{t_0}\right) \exp\left[m\left(\frac{t_0 - t}{t_0}\right)\right]$$

being m a parameter that can be linked to the duration of the earthquake and t_0 the time at which the function reaches its maximum value.

Overall assessment of agreement

The overall agreement of the generation process can be assessed at four different levels, comparing, for a selected sub-set of the used earthquakes, data gathered by means of the training procedure and the corresponding original one.

The first level assessment is by means of comparing the obtained descriptors with similar values gathered from the used accelerograms. This first level assessment has already been successfully achieved.
The second level assessment consists on making a similar comparison in terms of the power spectral density functions, assuring that there is a good description of the frequency content of the motion.
The third level assessment would be a similar comparison in terms of the accelerograms themselves, allowing assuring also a good description of the time domain evolution of the ground motion, that is especially important in case the generated accelerograms are to be used for the purpose of non-linear structural analysis.
Finally, a fourth level assessment can be done by means of elastic response spectra. These spectra, which are currently used for structural design purposes, are measures of the maximum response of a generic single degree of freedom system to the ground motion described by means of an accelerogram. This assessment does hence allow a direct comparison in terms of their effects on structures, between the real and the trained systems.
Studies at the three higher levels for the assessment of the overall agreement are currently under way.

CONCLUSIONS

The problem of predicting the description of the effects of an expected earthquake at a given site, through the associated strong ground motion record has been tackled from a rather unconventional way, by virtue of an artificial neural network approach. It was demonstrated that the proposed strategy may lead to interesting results, for it was possible to show that, even for a small number of ground motion records, it was possible to train a network to be able to predict the parameters required for the generation of artificial accelerograms, equivalent to those that have been actually recorded.

ACKNOWLEDGEMENTS

The paper has benefited from the valuable contributions of Profs. Luís Guerreiro and Rita Bento, of ICIST, through a number of interesting discussions, suggestions and materials. The authors are also indebted to Dr. Julian Bommer of the Department of Civil Engineering of Imperial College, London, for providing the soil classification of the whole records.

The work was developed in the framework of the *NeuroSis* research project funded by the Portuguese Foundation for Science and Technology (FCT) – PBICT/P/CEG/2380/95.Data analysis and processing was partly developed by Andreia Frade and Catarina Graça in the framework of their Final Year project in Applied Mathematics.

The authors have also benefited from regular and programmatic funding to ICIST by the Portuguese Foundation for Science and Technology (FCT).

REFERENCES

Gere, J; Shah, H., 1984, *Terra Non Firma*, The Portable Stanford.

NGDC, 1997, *The Earthquake Strong Motion CD-ROM Collection*, NOAA National Data Centre.

WHITESIDE, L; DATER, David and others, 1996, *Earthquake Sesmicity Catalog - Volumes 1 & 2*, National Geophisical Data Center, Boulder, CO