

TC Corner

TC304 (in collaboration with TC309) – Publically-available databases “304dB”

The technical committee on Engineering Practice of Risk Assessment & Management (TC304, risk) has recently compiled publically-available databases (nickname: 304dB), as the first step of the collaboration between TC304 and newly established technical committee on Machine Learning and Big Data in Geotechnics (TC309, machine learning). The main goal of this collaboration is to explore the use of advanced statistical methods and machine learning techniques for determining relationships within the data, and computing parameters for analytical models that apply those relationships to the use case at hand.

At this time, the following databases are available:

1. CPT databases
2. Multivariate soil/rock property databases
3. Geospatial databases

The databases can be freely downloaded at the following link:

http://140.112.12.21/issmge/Database_2010.htm

CPT databases

Four sets of vertical CPT clusters, i.e., multiple vertical CPTs are conducted in a local site, and one set of horizontal CPT are available for download. Some clusters have closely spacing vertical CPTs with small horizontal separation distances. These datasets are ideal for the study of spatial variability and underground stratification. One site (Baytown) is also with performance measurements (footing settlements & bearing capacities). This dataset is ideal for the study of performance-based and reliability-based design. Figure 1 shows an example of data available for the Hollywood, SC test site.

Multivariate soil/rock property databases

Eight multivariate datasets are available. These databases are for soil/rock samples with simultaneously measured properties (e.g., two clay samples at the same depth in the same site are tested, one to obtain Atterberg's limits and the other to obtain undrained shear strength). They may be used for estimating soil/rock design parameters and for the study of correlations among different soil/rock properties. Figure 2 shows the correlation plots produced by the data in a multivariate clay database.

Geospatial databases

Differently from the other two databases, the information essentially consists of a list of links to datasets available elsewhere. They can be either resources freely accessible on-line (e.g. Italian National Geoportal, UK National Geoscience Data Centre) or geospatial datasets compiled and managed for specific reasons (e.g. datasets used for landslide risk management in Hong Kong). In the latter case, the accessibility option can either be a public web page or the email of the person responsible for the data, who preliminary agreed to share them for academic studies upon request. Figure 3 shows an Italian seismic hazard.

TC Corner (Con't)

TC304 (in collaboration with TC309) – Publically-available databases
 “304dB”

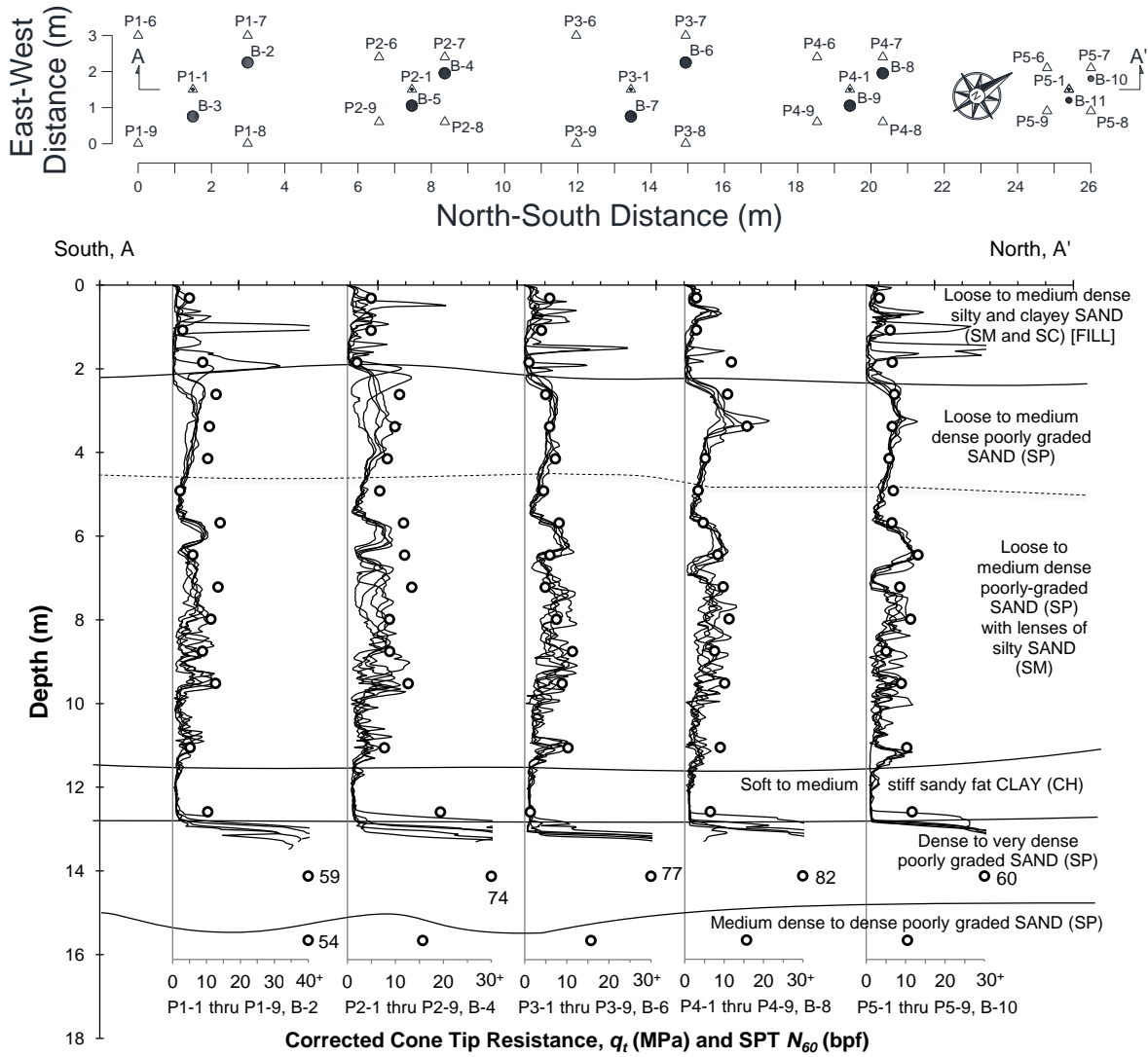


Figure 1. Example of data available for the Hollywood, SC test site: site and exploration plan (top), and subsurface cross-section indicating spatial variability of soil (bottom).

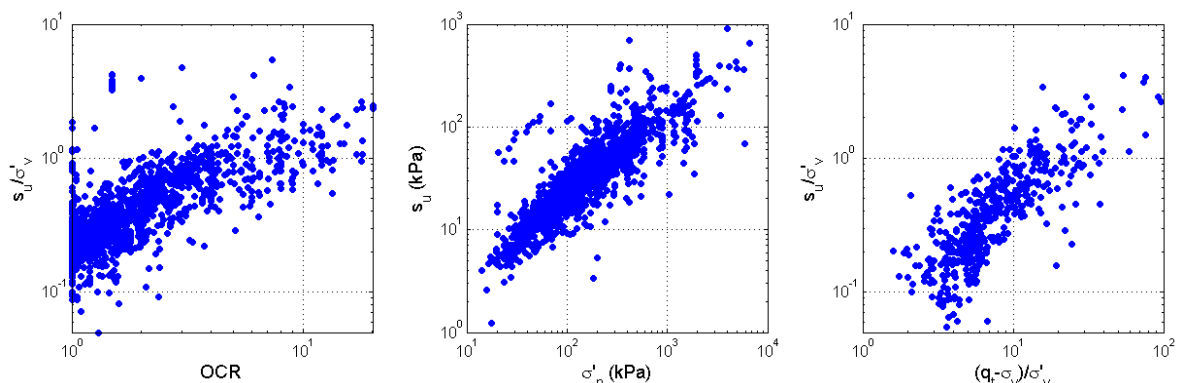


Figure 2. Correlation plots produced by the data in a multivariate clay database in 304dB.

TC Corner

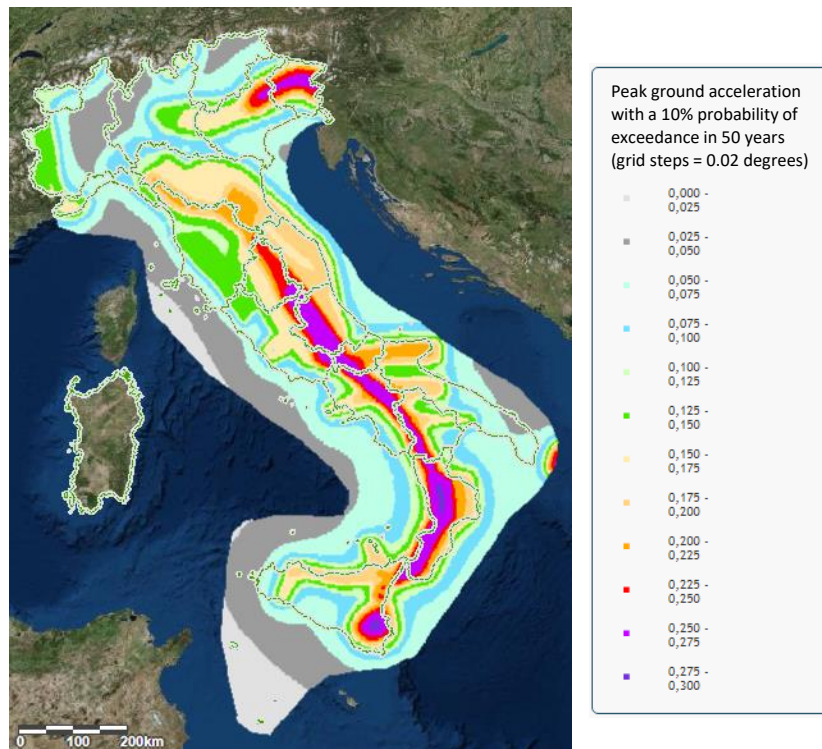
TC304 (in collaboration with TC309) – Publically-available databases
“304dB”

Figure 3. Italian seismic hazard: peak ground acceleration with a 10% probability of exceedance in 50 years on a grid spaced at 0.02° (Source: Italian National Geoportal, <http://www.pcn.minambiente.it/mattm/en/>)

The complete list for the database leaders & current contributors (contributors in alphabetic order):

CPT databases (leader: Armin Stuedlein):

Arjan Grashuis, Mark Jaksa, Armin Stuedlein, Zhongqiang Liu (NGI)

Multivariate soil/rock property databases (leader: Yu Wang):

Guojun Cai, Jieru Chen, Jianye Ching, Marco D'Ignazio & Tim Länsivaara

Geospatial databases (leader: Michele Calvello):

Michele Calvello, Wing Sun (Hong Kong GEO)