Smart Geotechnics for Smart Societies – Zhussupbekov, Sarsembayeva & Kaliakin (Eds)

Developing geodatabases in data-scarce regions: A case study of the Kathmandu Valley, Nepal

P.J. Vardanega, C.E.L. Gilder*, R. De Risi, F. De Luca & M.J. Werner
University of Bristol, Bristol, UK

R.M. Pokhrel*
Earth Investigation and Solution Nepal Pvt Ltd, Kathmandu, Nepal
**formerly, University of Bristol, Bristol, UK*

ABSTRACT: The Seismic Safety and Resilience of Schools (SAFER) project aimed to improve seismic hazard assessments in the Kathmandu Valley, Nepal. There remains scarce geotechnical data in this region. Therefore, the SAFER project team collected and sourced new geotechnical and geophysical data to allow for the development of improved maps showing the distribution of shear wave velocity measurements in the study region. This paper summarises geotechnical and geophysical investigations conducted in the valley and outlines the building of SAFER/GEO-591, which contains data from different sources such as previously drilled boreholes in the Kathmandu Valley. A methodology is proposed for geotechnical engineers and engineering geologists wishing to develop geodatabases in data-scarce regions.

1 INTRODUCTION

1.1 *Background*

In 2015, the M_w 7.8 Gorkha earthquake struck the Himalayan region affecting many communities, including the highly populated Kathmandu Valley in Nepal (Goda *et al.* 2015). Significant ground motion and many instances of damage were observed (e.g. Ohsumi *et al.* 2016). There was considerable investment in new seismic engineering research efforts from various agencies and organisations post-earthquake. This paper reports on some key outputs from the Engineering and Physical Sciences Research Council (EPSRC) funded project 'Seismic Safety and Resilience of Schools in Nepal' (EP/P028926/1), hereafter referred to as the 'SAFER project' (https://www.safernepal.net/). This paper primarily focuses on outputs related to geotechnical earthquake engineering from the 'New generation seismic hazard for Nepal' work package, hereafter referred to as 'WP1'.

1.2 *Study aims*

The SAFER project was comprised of five work packages (https://www.safernepal.net/) and ran from 2017 to 2021. A major focus of WP1 was on the compilation and curation of new geo-data sets for use in the development of new tools, such as new and updated maps for seismic hazard assessments in the Kathmandu Valley. This paper has two objectives: (i) it summarises the key deliverables from WP1 elaborating on how the outputs have contributed to improving seismic hazard assessments; and (ii) it offers a general methodology for geodatabase building for geotechnical engineers and engineering geologists working in data-scarce regions.

DOI: 10.1201/9781003299127-182

2 PROJECT OUTPUTS

This section focuses on the first objective of the paper. Sections 2.1 to 2.3 summarise the main deliverables from WP1 across three broad themes: (i) New geotechnical field testing in the Kathmandu Valley; (ii) Development of an open-source geodatabase for Kathmandu Valley soils and (iii) Updated maps for hazard assessments for the Kathmandu Valley.

2.1 *New geotechnical field testing in the Kathmandu Valley*

The field investigation included two new boreholes (progressed with rotary open hole drilling and sample recovery using the Standard Penetration Test (SPT) split spoon sampler), referred to as BH-1 and BH-2 (Gilder *et al.* 2019a). BH-1 was located at Dillibazar on the Padmakanya school premises at a ground elevation of 1309m. BH-2 was located on a site in Bijeshowari at a ground elevation of 1286m (see Gilder *et al.* 2019a for details on the testing). Seismic downhole testing was used to determine shear wave velocities at BH-1 (Gilder *et al.* 2019a), and Horizontal to Vertical Spectral Ratio (HVSR) testing was used for the same purpose at BH-2 (Pokhrel *et al.* 2019).

On a separate trip, Cone Penetration Testing (CPT) (to the authors' knowledge the first undertaken in the area) was carried out on seven sites to obtain measurements of shear wave velocity (Vs) using the seismic cone and also acquire continuous ground profile records with depth for the soils in the study region (see Gilder *et al.* 2021a and Gilder 2022 for further details on the CPT work). Figure 1 shows the locations of the new CPT readings (along with the location of the SPT readings from the SAFER/GEO-591 database) in the Kathmandu Valley.

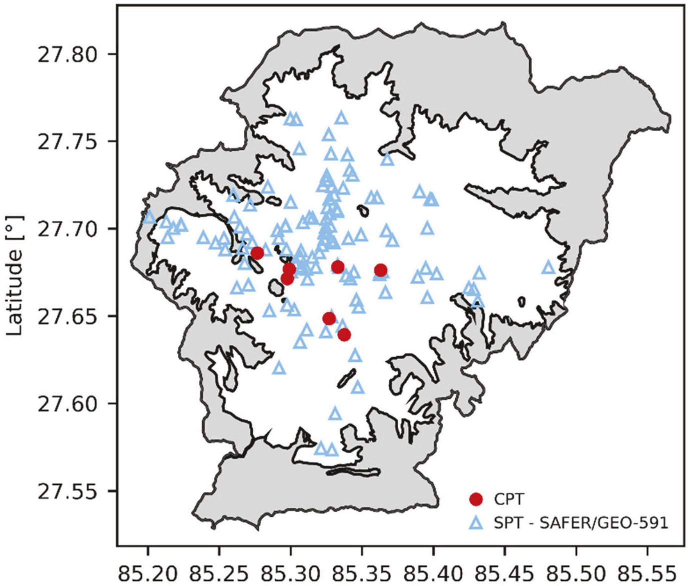

Figure 1. Location of the CPT sites outlined in Gilder *et al.* (2021a) compared with SPT locations from SAFER/GEO-591 (data from Gilder *et al.* 2019b) (figure adapted from Gilder *et al.* 2021a under the terms of the CC BY 4.0 license).

2.2 *Development of an open-source geodatabase for Kathmandu Valley soils*

The construction and data files of the SAFER/GEO-591 database is described in detail in Gilder *et al.* (2020) and the accompanying data record (Gilder *et al.* 2019b). SAFER/GEO-591 is available open-access from the University of Bristol Research Data Repository. While the database contains almost 600 borehole records collected by the project team from the literature, published reports and industry files, an extensive 'data-cleaning' process was necessary in order for the data to be sensibly put into the database structure (Gilder *et al.* 2021b gives an example of this process). As explained in Section 3, adopting a standard format for a geodatabase is essential for effective data transfer but requires considerable time investment and prior geotechnical knowledge. Figure 2 shows an example borehole log from SAFER/GEO-591 (Gilder *et al.* 2019b, 2020). The log shown in Figure 2 is for a site in the Gokarna formation.

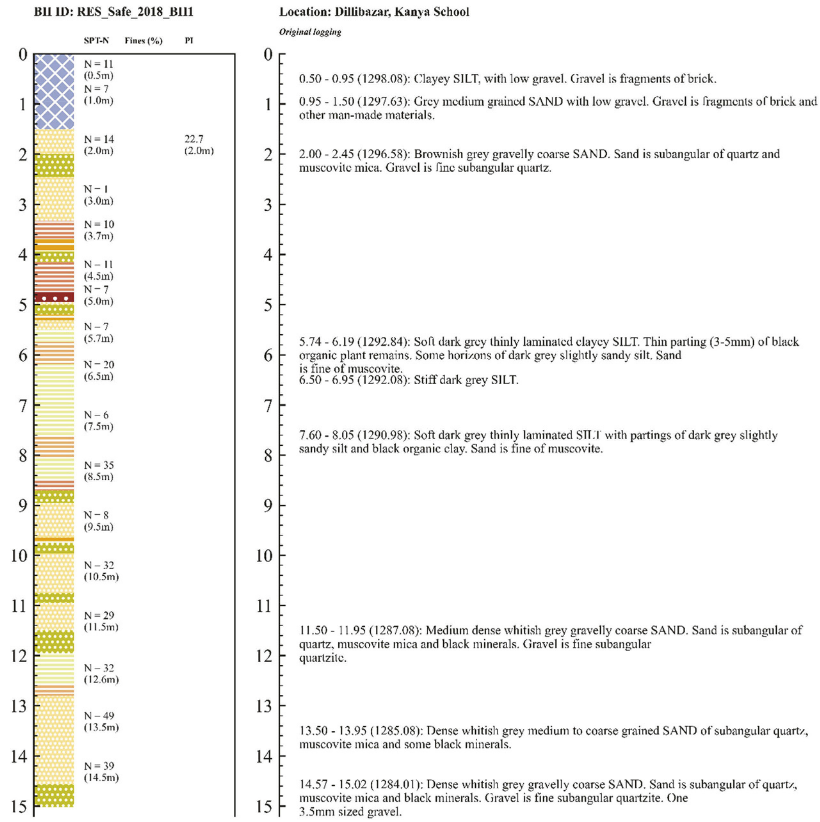

Figure 2. Example SAFER/GEO-591 borehole log from the Gokarna formation: upper 15m of the log shown (data from Gilder *et al.* 2019b).

2.3 *Updated maps for hazard assessments for the Kathmandu Valley*

Shrestha *et al.* (1998) produced a key engineering and environmental geological map for the Kathmandu Valley. During the SAFER project, data digitised from this map was often used as a base for the presentation of new analysis results and as a source of comparison for further findings. An important proxy for site amplification studies is the soil Vs averaged across the top 30m depth of ground ($V_{S,30}$), (e.g. Stewart *et al.* 2003). Kriging analysis is often used to estimate $V_{S,30}$ across geographical regions. De Risi *et al.* (2021) presented the results of a newly developed Bayesian-Kriging study making use of the V_{s30} measurements available (not including the new CPT data from Gilder *et al.* 2021a). Then the new V_{s30} measurements from the CPT work as well as those from the original database, were combined and analysed using the further enhanced multi-Gaussian Bayesian updating framework, showing how different kinds of V_{s30} proxies can be used to supplement local data and capture key geological features of the basin (see Gilder 2022 and Gilder *et al.* 2022).

3 DATA COLLECTION & CURATION IN DATA-SCARCE REGIONS

Geodatabases are vital for geotechnical engineers to benchmark results and make a-priori predictions of key design parameters. Thompson (2016) presented a case for public databases and described examples of such databases from various jurisdictions around the world. Open geodatabases allow the technical community to add new data to existing datasets and avoid duplicating digitisation and curation work. Phoon (2019, p.187) explained the 'MUSIC' concept i.e. that geotechnical data is often 'Multivariate, Uncertain and Unique, Sparse, Incomplete, and potentially Corrupted (MUSIC)'.

While the case for open geodatabases can be easily made in the authors' view, there are barriers to their creation and curation. Gilder *et al.* (2021c) presented a detailed review of the current barriers to geotechnical and geological data-sharing, with time and resources often cited as key constraints (as well as incentive) (see also the discussion in Thompson 2016). Based on the learnings from the SAFER project, a general methodology is proposed for those professionals starting work in data-scarce regions (see also Shepheard *et al.* (2019), where similar efforts have been made to develop new databases using a similar methodology, as well as efforts to build a pile load test database for the UK (Vardanega *et al.* 2021) and Thompson (2016) who also articulated many pertinent issues with developing public geodatabases). Phoon (2018) outlines some key challenges and important aspects of 'probablisitic site characterization' an activity for which geodatabases are useful (cf. Phoon 2018).

3.1 *Assess the geo-data landscape*

As with desk studies during site investigation projects, historical soil maps, hazard maps, and other geo-resources should be sourced by the project team. These resources (if available) can serve as useful benchmarks against future analyses (see Section 3.5). During this phase, the literature should be consulted to see if other databases have been compiled for the region of study.

3.2 *Local knowledge*

Liaise with local practitioners to determine who the main 'data owners' are. If they easily identifiable, approach these individuals/entities to determine if contributions to the new database are possible. Discuss at the outset how data records may need to be anonymised, e.g. exact locations removed, project names redacted etc. During this phase, agree the license and terms of use for the compiled database and ensure that those donating data aware of these conditions related to future data release.

3.3 *Local practice*

Unless the project team is familiar with site investigation practices in the area and, if budget permits, arrange to visit local projects or laboratories to determine which testing standards are used and which testing equipment is commonly utilised for laboratory and field studies conducted in the study region.

3.4 *Data harmonization and cleaning*

If data from multiple laboratories or site investigation companies will comprise the database, ensure that a 'data harmonisation' and 'cleaning' process is carried out by an experienced project team member. During this time, determine the data format to be used (see the discussions in Thompson (2016) and Gilder *et al.* (2021c) regarding file formats and workflow) and prepare a database manual with the data sources and notes from the cleaning and harmonisation process.

3.5 *Benchmarking analysis*

The new database should be checked against existing resources (if available), e.g. soil maps, local 'transformation models' (cf. Phoon & Kulhawy 1999a, b) or previously published parameter distributions, to determine if the new database is sufficiently representative of the study area. Also ascertain if the database contains sufficient data for meaningful statistical analyses. Some engineering judgement will be needed to determine if this is the case and during this step, potential outliers may be identified and noted in the database manual. It may be that the database will need further expansion as new data becomes available after the initial data release. This stage is also where major gaps in the database can be identified (cf. Phoon 2019: MUSIC) and these should be noted in the database manual.

3.6 *Data release*

While there are barriers to this stage (e.g. commercial interests), if possible, the database should be released open-access with appropriate disclaimers. Later corrections and clarifications may be needed as practitioners and researchers start using the database, which can be incorporated in future versions. 'Living databases' may, become a useful technical resource for the local and international geotechnical communities, allowing for improved design and model calibration efforts.

4 SUMMARY AND CONCLUSIONS

This paper has presented some of the key outputs from WP1 from the SAFER project. The outputs discussed covered three broad themes: (i) New geotechnical field testing in the Kathmandu Valley; (ii) Development of an open-source geodatabase for Kathmandu Valley soils and (iii) Updated maps for hazard assessments for the Kathmandu Valley. The focus of the SAFER project was on improving the safety and resilience of school buildings in Nepal; however, many of the project findings should be useful for other infrastructure classes. The paper proposed a 'step-by-step' methodology for those developing open-access databases in data-scarce regions.

In conclusion, the lessons from WP1 of the SAFER project have revealed many challenges in building and releasing open-access geodatabases, however such resources are essential for regional statistical analyses (e.g. using Kriging-based methodologies), to produce new maps and information for improved hazard assessments.

ACKNOWLEDGEMENTS

The second author acknowledges the support of EPSRC (EP/R51245X/1). The authors acknowledge the support of: the EPSRC project SAFER (EP/P028926/1), the UoB Faculty of Engineering; IN SITU Site Investigation and Professor Anastasios Sextos for leading the SAFER project.

DATA AVAILABILITY STATEMENT

This study has not generated new experimental data.

REFERENCES

De Risi, R., De Luca, F., Gilder, C.E.L., Pokhrel, R.M. & Vardanega, P.J. 2021. The SAFER geodatabase for the Kathmandu valley: Bayesian kriging for data-scarce regions. *Earthquake Spectra* 37(2): 1108–1126.

Gilder, C.E.L. 2022. Geotechnical data curation and a geostatistical multivariate framework for V_S prediction in data scarce contexts. *Ph.D. thesis*, Bristol, UK: University of Bristol.

Gilder, C.E.L., Pokhrel, R.M. & Vardanega, P.J. 2019a. A ground investigation to inform earthquake hazard assessment in the Kathmandu Valley, Nepal. In H. Sigursteinsson *et al.* (eds.), *Proc. XVII*

European Conf. on Soil Mechanics and Geotechnical Engineering - Reykjavík, Iceland, 1–6 of September 2019. Reykjavík, Iceland: The Icelandic Geotechnical Society (IGS).
Gilder, C.E.L., Pokhrel, R.M. & Vardanega, P.J. 2019b. The SAFER Borehole Database (SAFER/GEO-591_v1.1). Bristol: University of Bristol. https://doi.org/10.5523/bris.3gjcvx51lnpuv269xsa1yrb0rw
Gilder, C.E.L., Pokhrel, R.M., Vardanega, P.J., De Luca, F., De Risi, R., Werner, M.J., Asimaki, D. Maskey, P.N. & Sextos, A. 2020. The SAFER Geodatabase for the Kathmandu Valley: Geotechnical and geological variability. *Earthquake Spectra* 36(3): 1549–1569.
Gilder, C.E.L., Pokhrel, R.M., De Luca, F., & Vardanega, P.J. 2021a. Insights from CPTu and Seismic Cone Penetration Testing in the Kathmandu Valley, Nepal. *Frontiers in Built Environment* 7: [646009].
Gilder, C.E.L., Vardanega, P.J., Pokhrel, R.M., De Risi, R. & De Luca, F. 2021b. Assessing Transformation Models Using a Geo-Database of Site Investigation Data for the Kathmandu Valley, Nepal. *Lecture Notes in Civil Engineering* 125: 331–338.
Gilder, C.E.L., Geach, M., Vardanega, P.J., Holcombe, E.A. & Nowak, P. 2021c. Capturing the views of geoscientists on data sharing: a focus on the geotechnical community. *Quarterly Journal of Engineering Geology and Hydrogeology* 54 (2):[qjegh2019-138].
Gilder, C.E.L., De Risi, R., De Luca, F., Pokhrel, R.M. & Vardanega, P.J. 2022. Geostatistical framework for estimation of V_{S30} in data-scarce regions. *Bulletin of the Seismological Society of America* 112 (6): 2981–3000.
Goda, K., Kiyota, T., Pokhrel, R.M., Chiaro, G., Katagiri, T., Sharma, K. & Wilkinson, S. 2015. The 2015 Gorkha Nepal earthquake: Insights from earthquake damage survey. *Frontiers in Built Environment* 1: [8].
Ohsumi, T., Mukai, Y., & Fujitani, H. 2016. Investigation of Damage in and Around Kathmandu Valley Related to the 2015 Gorkha, Nepal Earthquake and Beyond. *Geotechnical and Geological Engineering* 34(4): 1223–1245.
Phoon, K-K. 2018. Special Collection Announcement: Probabilisitc Site Characterisation. *ASCE-ASME Journal of Risk and Uncertainty Engineering Systems, Part A: Civil Engineering* 4 (4):[02018002].
Phoon, K-K. 2019. Editorial: Flow and transport in porous media in the face of uncertainty, part I. *Environmental Geotechnics* 6(4): 186–187.
Phoon, K-K. & Kulhawy, F.H. 1999a. Characterization of geotechnical variability. *Canadian Geotechnical Journal* 36(4): 612–624.
Phoon, K-K. & Kulhawy, F.H. 1999b. Evaluation of geotechnical property variability. *Canadian Geotechnical Journal* 36(4): 625–639.
Pokhrel, R.M., Gilder, C.E.L., Vardanega, P.J., De Luca, F., Werner, M.J. & Maskey, P.N. 2019. Estimation of V_{S30} by the HVSR Method at a Site in the Kathmandu Valley, Nepal. In *2nd Int. Conf. on Earthquake Engineering and Post Disaster Reconstruction Planning*, *25-27 April*, *2019*: 52–60.
Shepheard, C.J., Vardanega, P.J., Holcombe, E.A., Hen-Jones, R. & De Luca, F. 2019. Minding the geotechnical data gap: appraisal of the variability of key soil parameters for slope stability modelling in Saint Lucia. *Bulletin of Engineering Geology and the Environment* 78(7): 4851–4864.
Shrestha, O.M., Koirala, A., Karmacharya, S.L., Pradhananga, U.B., Pradhan, R. & Karmacharya, R. 1998. *Engineering and environmental geological map of the Kathmandu Valley, Scale 1: 50 000*. Kathmandu, Nepal: Department of Mines and Geology, Lainchaur.
Stewart, J.P., Liu, A.H. & Choi, Y. 2003. Amplification factors for spectral acceleration in tectonically active regions. *Bulletin of the Seismological Society of America* 93(1): 332–352.
Thompson, T. 2016. A 2016 case for public geotechnical databases. In: *Proc. of the 5th Int. Conf. on Geotechnical and Geophysical Site Characterisation (ISC'5) Gold Coast, Queensland Australia. September 2016*. Available from: < https://www.issmge.org/uploads/publications/25/26/ISC5_154.pdf> (20 January 2023).
Vardanega, P.J., Crispin, J.J., Gilder, C.E.L., Voyagaki, E. & Ntassiou, K. 2021. DINGO: A Pile Load Test Database. In K.G. Higgins ***et al.*** (eds). *Piling 2020: Proceedings of the Piling 2020 Conference*: 229–234. London, UK: ICE Publishing.