A Grid-Based Facility for Large-Scale Cross-Correlation of Continuous Seismic Data

John Townend[1], Yannik Behr[2], Kevin Buckley[2], Martha Savage[1], John Hine[2]
[1]School of Geography, Environment, and Earth Sciences, Victoria University of Wellington, Wellington, New Zealand,
john.townend@vuw.ac.nz
[2]School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

INTRODUCTION

Just as radiographers use X-rays and ultrasound to image the internal structure of a human body, so seismologists use seismic waves generated by earthquakes or artificial sources to study the earth's interior. More than 95% of the time, however, seismometers designed to record earthquakes are actually recording continuous, low-pitched noise—the incoherent background hum of the earth. Much of this noise is produced by ocean waves hitting the coastline, and New Zealand's geographic isolation exposes it to a particularly energetic ocean. Recent studies reveal that this noise is not entirely random. By comparing long records of seismic noise recorded at two different locations, a small amount of coherent seismic energy propagating directly between them can be detected. This energy propagates as a seismic wave at a speed governed by the physical properties of the rocks it passes through. By measuring this speed, geophysicists can map Earth's deep structure in much the same way as ultrasound is used to look inside human bodies (Figure 1).

This work described here addressed two complementary goals. The first was to develop a computational workflow — a documented sequence of analytical steps — allowing automated or interactive analysis of continuous raw seismic data using grid-computing resources. The second goal was to develop an interface to the computational workflow that facilitates its use in an efficient and effective manner by researchers in the broader geophysical community.

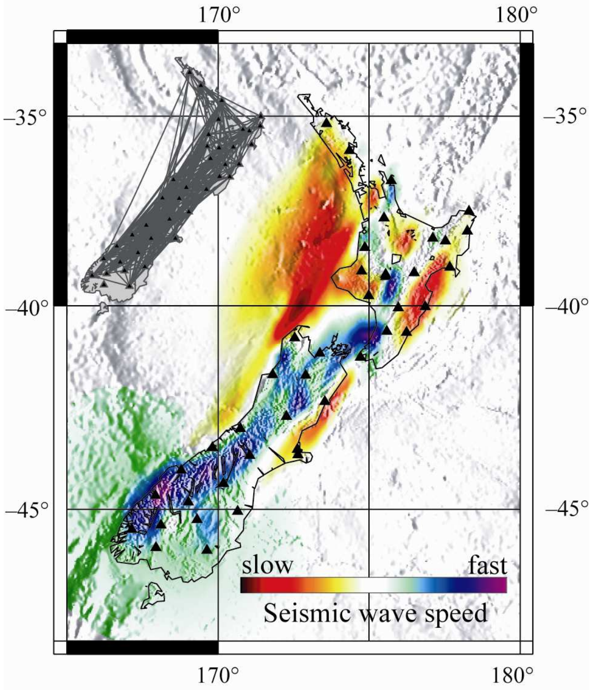

Figure 1: Map of seismic wave speeds at ~15 km depths obtained using one year of continuous seismic noise recorded by the GeoNet network (triangles; http://www.geonet.org.nz/). Red colours indicate slow speeds (e.g. Taranaki and Hikurangi basins west and east of the North Island), and blue colours indicate high speeds. The inset map shows the propagation paths analysed [1].

AMBIENT SEISMIC NOISE TOMOGRAPHY

Modern seismological networks, such as New Zealand's GeoNet system (Figure 1), are designed to record seismic waves generated by earthquakes, but more than 95% of the "signal" recorded by such networks is ostensibly incoherent noise generated by the interaction of ocean waves with the seabed and coastline. It has been shown recently that cross-correlating long noise records from pairs of seismometers reveals a coherent signal corresponding to the propagation of a seismic wave from one instrument to the other [2-5]. Over distances greater than ~10 km, this signal corresponds to a wave propagating just below the earth's surface [6] at speeds determined by the earth's elastic properties. Lower-pitched energy samples greater depths: measurements at periods of 5–25 s provide images of geological structure to depths of

~30 km [4, 7]. By analysing cross-correlation measurements at different frequencies, we can therefore determine the seismic velocity structure of the earth's crust at various depths [1].

COMPUTATIONAL WORKFLOW MODELLING

As in many other data-intensive fields of science, geophysical research often involves a large number of incremental processing steps, during each of which decisions must be made regarding the particular choice of parameters or even algorithms to use. We have been working on methods of combining the increasingly routine geophysical task of cross-correlating large data sets to image the earth, with modern e-research approaches to systematising and documenting the research process itself.

The idea of encapsulating within a particular research output all of the processing parameters used in obtaining that output (a figure, table, or parameter value) from the original input (raw data) is not new. Examples within the geophysical realm include the Complete PostScript System [8], an archival and exchange format that incorporates details of the processing parameters and algorithms used to generate it within an output image; the SINEX format [Solution-INdependent Exchange Format; 9] used to exchange geodetic locations and the modelling parameters they depend on; and the Stanford Exploration Project reproducible electronic document protocol [10], which consists of rules used to reproduce entire books from the original source code and data. What each of these examples — which are often referred to as "reproducible research" — provides, to varying degrees, is the potential for researchers to ask questions of their own or others' research such as:

- Could we plot that data a different way to see things more clearly?
- What effect on the results would changing that parameter have?
- Would she have got the same result if she'd used his algorithm at that step instead of mine?

At its simplest, the implementation of a workflow is a computer program. However, when there are large amounts of data and possibly different computational resources involved the problem of tracking and documenting input parameters and processing requirements becomes more complex.

CROSS-CORRELATION OF CONTINUOUS SEISMIC DATA IN A GRID-COMPUTING ENVIRONMENT

We have developed methods of extracting continuous seismic waveform data from the GeoNet archives via the Kiwi Advanced Research and Education Network (KAREN) using web services, and of distributing the data pre-processing and cross-correlation tasks amongst processors operating within a Sun Grid Engine environment.

To date, our research has focussed on tailoring geophysical demands to the available computational resources and protocols. The collaborative nature of this project has highlighted how important close interaction between end-users (geophysicists) and computing specialists is in ensuring that complex problems are described and represented in the most computationally efficient way possible. We have successfully demonstrated the suitability of a two-stage processing sequence for cross-correlation jobs involving data sets of a range of sizes. Future work will address questions such as how to most efficiently allocate tasks between the processors available at any one time, and how to monitor resources and job progress.

REFERENCES

1. Lin, F.C., et al., *Ambient noise Rayleigh wave tomography of New Zealand.* Geophysical Journal International, 2007. **170**(2): p. 649-666.
2. Shapiro, N.M. and M. Campillo, *Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise.* Geophysical Research Letters, 2004. **31**(7): p. L07614 1–4.
3. Sabra, K.G., et al., *Extracting time-domain Green's function estimates from ambient seismic noise.* Geophysical Research Letters, 2005. **32**(3): p. 1–5.
4. Shapiro, N.M., et al., *High-resolution surface-wave tomography from ambient seismic noise.* Science, 2005. **307**(5715): p. 1615–1618.
5. Weaver, R.L., *Information from seismic noise.* Science, 2005. **307**(5715): p. 1568–1569.
6. Roux, P., et al., *P-waves from cross-correlation of seismic noise.* Geophysical Research Letters, 2005. **32**(19): p. 1–4.
7. Sabra, K.G., et al., *Surface wave tomography from microseisms in Southern California.* Geophysical Research Letters, 2005. **32**: p. 1–4.
8. Wessel, P., *Complete PostScript: an archival and exchange format for the sciences?* EOS, 2003. **84**: p. 351.
9. Blewitt, G., Y. Bock, and J. Kouba, *Constraining the IGS polyhedron by distributed processing*, in *IGS Analysis Workshop Proceedings: Densification of ITRF through regional GPS networks*, I.C. Bureau, Editor. 1994, Jet Propulsion Laboratory: Pasadena. p. 21-27.
10. Schwab, M., M. Karrenbach, and J. Claerbout, *Making Scientific Computations Reproducible*, in *Computing in Science and Engineering*. 1997. p. 61-67.