Geophysical Research Abstracts
Vol. 13, EGU2011-11777, 2011
EGU General Assembly 2011

A novel automated approach to ambient noise data processing using the ADMIRE framework

Luca Trani (1), Alessandro Spinuso (1), Michelle Galea (1), and Ian Main (2)

(1) University of Edinburgh, School of Informatics, United Kingdom , (2) University of Edinburgh, School of GeoSciences, United Kingdom

Recording seismological data volumes have dramatically increased in the last decade. For example 500 modern standard broadband seismometers aimed at detecting natural earth vibrations including earthquakes yield ≈17 gigabytes per day and 6.2 terabytes per year.

These large, continuously-recorded datasets (in the form of time series at discrete spatial points) can be exploited to address several fundamental questions related to diverse phenomena: the Earth's seismic velocity structure and its time-dependence from seismic interferometry, its free oscillations and their relation to the Earth's deep internal structure and composition, its 'hum' (low frequency constant background noise) and its origin in storms in the ocean, the nature of recently discovered 'slow' seismic events and their relation to the nature of plate coupling at subduction zones, the origin of non-volcanic tremors, and the episodic nature of glacier movement.

Processing primary seismic data for these applications increasingly requires the use of high-performance technologies, that are not always available either at the data-centre's facilities or within individual research groups. Thus there is a need for an accessible facility of "software as infrastructure" capable of "data-rich" and "cpu-rich" tasks. Such an infrastructure should provide, in a flexible and maintainable way, tools and capabilities for data and model sharing, information and knowledge extraction, data mining, processing and visualization. In this context is desirable to have a general framework which has been professionally engineered to help accomplish most of the common tasks needed to achieve the data-intensive research objectives.

The ADMIRE project (Advanced Data Mining and Integration Research for Europe) aims to deliver the first such framework in Europe, providing a consistent and easy-to-use technology for extracting information and knowledge from data coming from multiple heterogeneous and distributed resources.

This work shows how the ADMIRE framework can significantly help and support scientists to achieve their goals and to improve their research quality. In particular it focuses on a specific use case: cross-correlation of ambient seismic noise recorded by seismographs installed all over the globe and stored within a network of distributed data-centres.

Although the format of the raw data is usually represented in a well established community standard (MSEED), each data centre can have a different data access and management system. Thus to be as flexible as possible a framework has to deal, for example, with legacy distributed databases and file system archives, heterogeneity in data and meta-data storage. In this respect ADMIRE proved itself to be a very smooth and easily configurable solution, capable of dealing with large data volumes.

The ADMIRE architecture separates the concerns of domain and data-intensive experts, so that researcher can focus purely on their scientific tasks. Individual analysis units or Processing Elements (PEs) are available to users via a registry with the correspondent semantic description. By combining these PEs domain experts formalise their knowledge into workflows that are independently and transparently enacted. In this particular study functionalities for raw data gathering, pre-processing (filtering, normalization, instrument response removal, etc), data aggregation, cross-correlation and stacking have been implemented.

An important effort has been dedicated to integrating already available domain specific API (ObsPy) as the basis of the analysis code embedded into PEs.
The framework ensures also a high grade of quality measurement through stability, robustness and repeatability of

the processing procedures. Workflows can be independently (re-)submitted with different input parameters giving the possibility to compare intermediate results or to look at time-dependent processes.
In summary the novelty of our approach is in the flexibility given to users to reach their scientific goals, one immediate application for example could be, with a few adjustments of the workflow, real time seismic interferometry in volcanic areas.