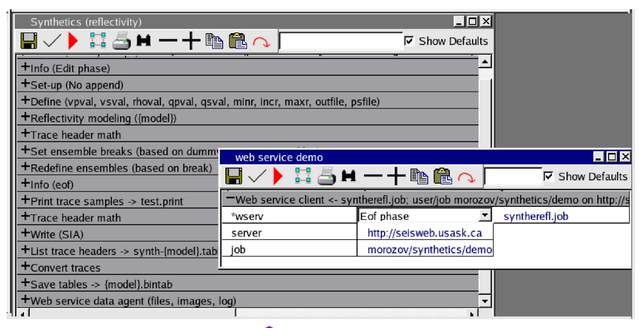
7. Processing web service

A web service model for data manipulation, analysis, and modeling was implemented, based on a generalized data processing system called SIA. The service is not limited to any specific data type or operation and allows the user to combine ~190 tools of the existing package, with new codes easily includable. Although initially started as a replacement to a commercial reflection seismic processing system, the system has now grown into a generalized code design framework and has been applied in several areas of geophysics, such as reflection/refraction and earthquake seismology, potential fields, and visualization. The system is already fully developed and includes a Graphical User Interface, parallel processing capabilities, on-line documentation, on-line software distribution service and automatic code updates.

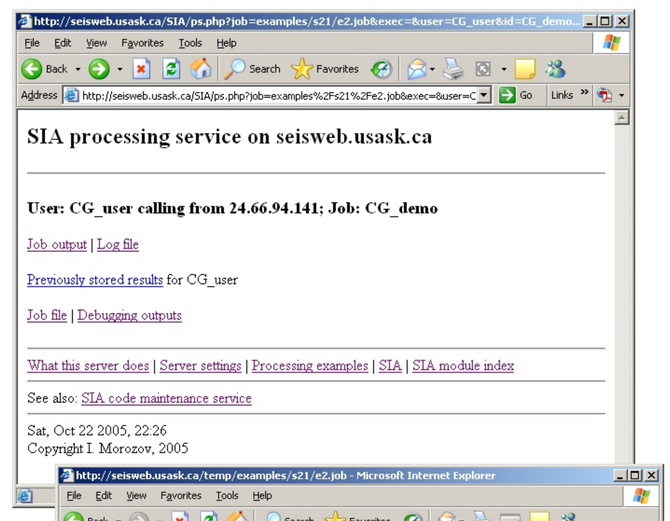
The processing service (http://seisweb.usask.ca/SIA/ps.php) allows remote execution of complex processing flows completely designed and controlled by remote clients who are presented with mirror images of the server processing environment. A characteristic feature of the model is its decentralization, with all the servers and clients running identical copies of SIA software. Clients are also able to post their processing flows at the web server, thereby building a knowledge base of processing expertise shared by the community. Flows in this knowledge base are currently represented by a hierarchy of automatically-generated interactive web forms. These flows can be accessed and the resulting data retrieved by either using a web browser or through API calls from within the clients' applications. Server administrator is thus relieved of the need for development of any content-specific data access mechanisms or codes.

Currently, the processing service is utilized to maintain a library of processing examples (below) including a number of useful web tools (such as UTM coordinate transformations, calculation of travel times of seismic waves in a global Earth model, and generation of GMT color palettes). Examples of important potential applications of this web service model include b intelligent data queries, processing, and modeling of global seismological data.



A fragment of client's SIA GUI window showing (a) client and (b) server processing flows performing a web service. When flow (a) is executed, the "web service client" tool sends flow (b) to the remote server. In this example, the server process performs 1D waveform modeling using the *reflectivity* method (Fuchs and Müller, 1971). Note the complex modeling sequence, with symbolic substitutions, several data transformations, output, and plotting performed. The results (model, output seismic data files, images, and processing log) are posted on the returned web page by the "web service data agent" tool, marked with an arrow in flow (b). Note that if flow b) requires no special system environment (computing power, databases), the client is also able to execute it on its local network by pressing the "Play" button (circled).

Typical format of service results (single-point UTM coordinate conversion example). Top window: HTML document returned by the server. It provides links to the job script, output (index), and log files, user's directory, as well as links to service info and system help. Middle: current job script. Note that the last tool (table) prints the results directly into the output file, `index.html`, represented by symbol '`@`'. Bottom: job output file `index.html`) containing the output string. Other tools, such as `post`, `print`, and `table`, could make additional printouts and add links to the output files.

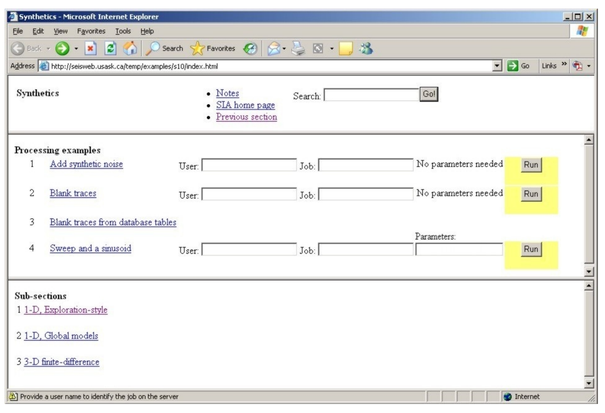


8. Sharing community processing expertise

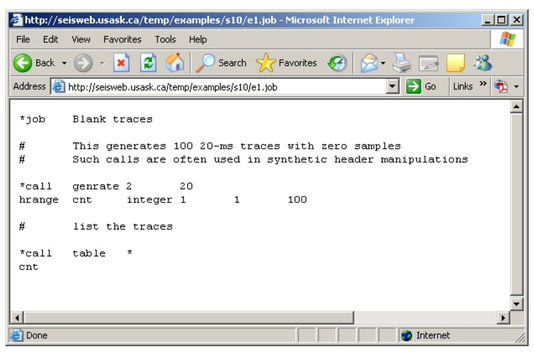
Using the web processing capability above, we created a prototype for "community" shared data processing. At present, this model is used for organizing and sharing processing expertise by the students of our Seismology group.

By submitting jobs containing tool "expert" to the web processing server, any user can contribute to the library of processing examples (http://seisweb.usask.ca/temp/examples). This library includes executable jobs as well as fragments illustrating the use of specific tools and their combinations. Posted job scripts can be browsed and pasted into the SIA flow editors or executed on the server, with the results provided in the form of printouts, images, or data files.

Ongoing development of this functionality includes the functionality to allow the clients to post web forms and complete web pages offering to the community a complex custom functionality, such as pre-processing or modeling performed at the data center facility. With adoption of such a model, the "data center" staff would become relieved of many responsibilities for development of user interfaces.



Sample page (section "Synthetics") in the current library of processing examples (http://seisweb.usask.ca/temp/examples).
In the upper frame, comments to the section (link "Notes"), a link to system on-line documentation, and a link to the parent section are provided. The search tool is implemented by executing a predefined processing flow `search.job` on the server.

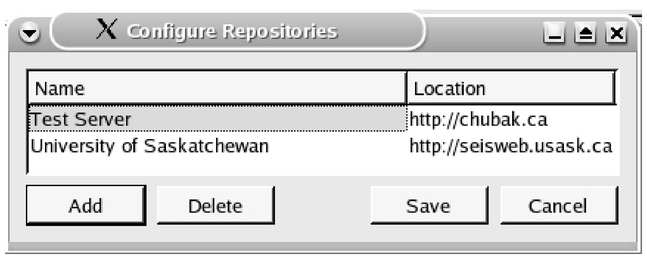


Browser window invoked by following the "Blank traces" link above. This script can be dragged and dropped into the flow editor or executed on the server by pressing the "Run" button.

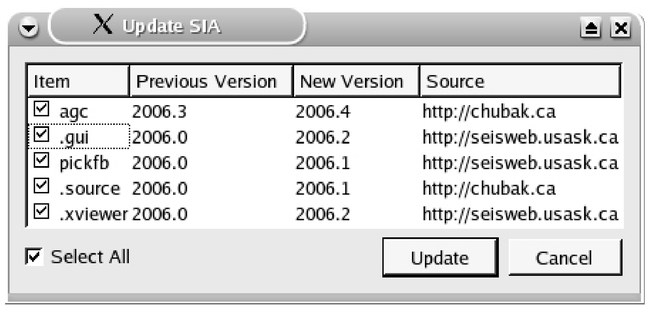
9. Automated software distribution and updates

In addition to the processing service, a companion code distribution and automatic update service (http://seisweb.usask.ca/SIA/cs.php) was also designed. The service utilizes the SIA system maintenance utilities to synchronize its source code with multiple (user-selectable) software repositories. The program performs automatic checking for the available updates, downloads them from the web, and installs them from source code. Notably, any SIA distribution (used for data processing or code development) on a system providing Apache web services would automatically become a code distribution server.

A characteristic feature of our model is that the SIA code is configured identically on all the code repository servers and processing clients. Code servers are therefore also capable of performing full data processing, including remote processing as a web service (see above). Conversely, if a standard Apache web server is available on a system used, for example, for specialized data processing and development of the corresponding tools, it can automatically share these tools with others. As a code is updated or added, all clients which connect to this repository will immediately (as soon as the version number is advanced) have access to it. Such symmetrical design makes installation and maintenance of multiple copies of the package easy and reliable.



Configuration of SIA software repositories. Note that only the root entry points are shown as the web addresses. For example, the actual code server for the selected line is http://chubak.ca/SIA/cs.php. The buttons below allow the user to edit the list.



Choosing software components to update. Note that the components whose names begin with a period are system libraries or configuration directories, and the rest are plug-in processing tools. For each component, its current and updated version numbers, and the source of the update are displayed. The user can select some or all components which will be downloaded, compiled, and installed.

10. Conclusion

The ongoing development of the SIA code framework shows that the entire scope of critical issues facing geophysical data management and processing can be solved in a consistent manner. The codes are highly integrated, streamlined for data- or computationally-intensive seismic and non-seismic processing and modeling, make broad provisions for parallelization and remote (web service) operation, and incorporate some of the key community software. With the new web distribution service, the codes can also be developed by multiple authors and seamlessly and automatically maintained up to date.

UNIVERSITY OF SASKATCHEWAN