THE GNEM R&E PARAMETRIC GRID SOFTWARE SUITE: TOOLS FOR DATA CREATION, ACCESS, MANAGEMENT, VIEWING, AND EXPORT

James R. Hipp, Randall W. Simons, Lee Jensen, Lucinda A. Lindsey, and Mathew N. Chown

Sandia National Laboratories

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. DE-AC04-94AL85000

ABSTRACT

One of most important types of data in the National Nuclear Security Administration (NNSA) Ground-Based Nuclear Explosion Monitoring Research and Engineering (GNEM R&E) Knowledge Base (KB) is Parametric Grid (PG) data. PG data can be used to improve signal detection, signal association, and event discrimination, but so far its greatest use has been for improving event location by providing ground-truth-based corrections to travel-time base models. In this presentation we will discuss the Parametric Grid Software Suite (PGSS), which is the complete suite of GNEM R&E software developed by NNSA to create, access, manage, and view PG data. The PGSS consists of 4 main components, which are described below.

The primary PG population tool is the Knowledge Base Calibration Integration Tool (KBCIT). KBCIT is an interactive computer application whose goal is to produce interpolated calibration-based information that can be used to improve monitoring performance by improving precision of model predictions and by providing proper characterizations of uncertainty. It is used to analyze raw data and produce kriged corrections surfaces that can be included in the Knowledge Base. KBCIT not only produces the surfaces but also records all steps in the analysis for later review and possible revision.

The Parametric Grid Library (PGL) provides the interface to access the data and models stored in a PGL file database. The PGL represents the core software library used by all GNEM R&E tools that read or write PGL data (e.g. KBCIT). The PGL also provides a common access interface [the Common Grid Interface (CGI)] for many event monitoring pipeline codes. The library provides data representations and software models to support accurate and efficient seismic phase association and event location. The PGL currently provides direct support for travel-time, slowness, and azimuth data, as well as for generalized geophysical model data. PGL stores all of its data and model representations as Serialized Binary Stream Objects (SBSOs) that can be conveniently accessed using a set of database key objects. The key objects contain the SBSOs' descriptive meta-data information.

The Data Management Tool (DMT) provides access to PG data for purposes of managing the organization of the generated PGL file database, or for perusing the data for visualization and informational purposes. It is written as a GUI that can access the key file associated with any PGL file database and display it in an easily interpreted visual format. The format allows a user to view the entire object reference and dependency hierarchy, to remove objects from a database, copy objects from one database to another, or create new minimal PGL file databases that use a subset of one or more existing PGL file databases. The DMT also provides a mechanism to textually or visually display the content of any SBSO in a PGL file database. It will provide a means to locate and remove redundancy and to evaluate the overall storage efficiency of each type of supported SBSO.

The Viewing and Extraction Tool (VEXTool) provides a viewing and export capability for geophysical model data stored in the PGL database. VEXTool has four main displays providing different views of the geophysical data: the layer display, the volume display, the cross-section display, and the borehole display. Data from any of the displays can be exported for use by other analysis tools.

OBJECTIVE

Monitoring sensor data for possible nuclear explosions involves four fundamental tasks: signal detection, signal association, event location, and event identification. For all but the first of these, the task is accomplished by minimizing the misfit between observations and theoretical predictions, which in turn are dependent on the model of the Earth used. Thus, improving the quality of the Earth models used for monitoring is of critical importance and this is why it is a primary focus of the GNEM R&E efforts.

Having better Earth models is not enough, however; if the models are to be maintained and used for operational monitoring, we need a cohesive set of software tools to leverage the improved models' benefits. Population tools are needed to make it easy to process the data into the proper format, and to update the models when new data becomes available. The Earth models must be designed and defined into an integrated and extensible software system. An interface is needed around the models to serve predictions to various applications. An efficient, extensible storage format is needed to archive the models on disk. Finally, some sort of data management and viewing utility is needed to easily allow users to browse through the models to better understand them, and to make simple edits (deletions, insertions) as needed.

Researchers at Sandia Labs have been working to meet these needs by developing an integrated suite of tools collectively expressed as the Parametric Grid Software Suite.

RESEARCH ACCOMPLISHED

The PGSS consists of the primary data population tool, KBCIT, the low level storage, representation, and model library, PGL, the database management tool, DMT, and the viewing and export tool, VEXTool. We describe each below.

Knowledge Base Calibration Integration Tool (KBCIT)

The Knowledge Base Calibration Integration Tool (KBCIT) is an interactive computer application whose goal is to produce interpolated calibration based information that can be used to improve monitoring performance by improving precision of model predictions and by providing proper characterizations of uncertainty. It is used to analyze raw data and produce kriged correction surfaces, which can be included in the Knowledge Base. KBCIT not only produces the surfaces, but also records all steps in the analysis for later review and possible revision.

KBCIT allows the user to define the boundaries of regions, where the data in each region can be treated differently, including using different base models, trend removals, outlier limits and variograms in each region. (See Figure 1.) KBCIT allows the user to set parameters for blending functions to apply across region boundaries to avoid discontinuities. A declustering operation is available to reduce data set size by removing redundancy, and stabilize kriging by averaging values close to one another.

KBCIT uses the Parametric Grid Library, described below, to implement both stationary and non-stationary Bayesian kriging for interpolation of empirical data with uncertainties. (See Figure 2.) The results of kriging can be cross-validated by comparing empirical values with interpolated predictions based on a kriged surface formed using all but the values to be compared.

To provide a faster alternative to interpolating a kriged surface, the user can specify a tessellation on which KBCIT will pre-compute kriged values. Any combination of kriged and tessellated surfaces can be summed or differenced and viewed with an interactive 3D viewer. (See Figure 3.) Kriged and tessellated surfaces can be saved in a Parametric Grid database for use by other programs such as those doing event location.

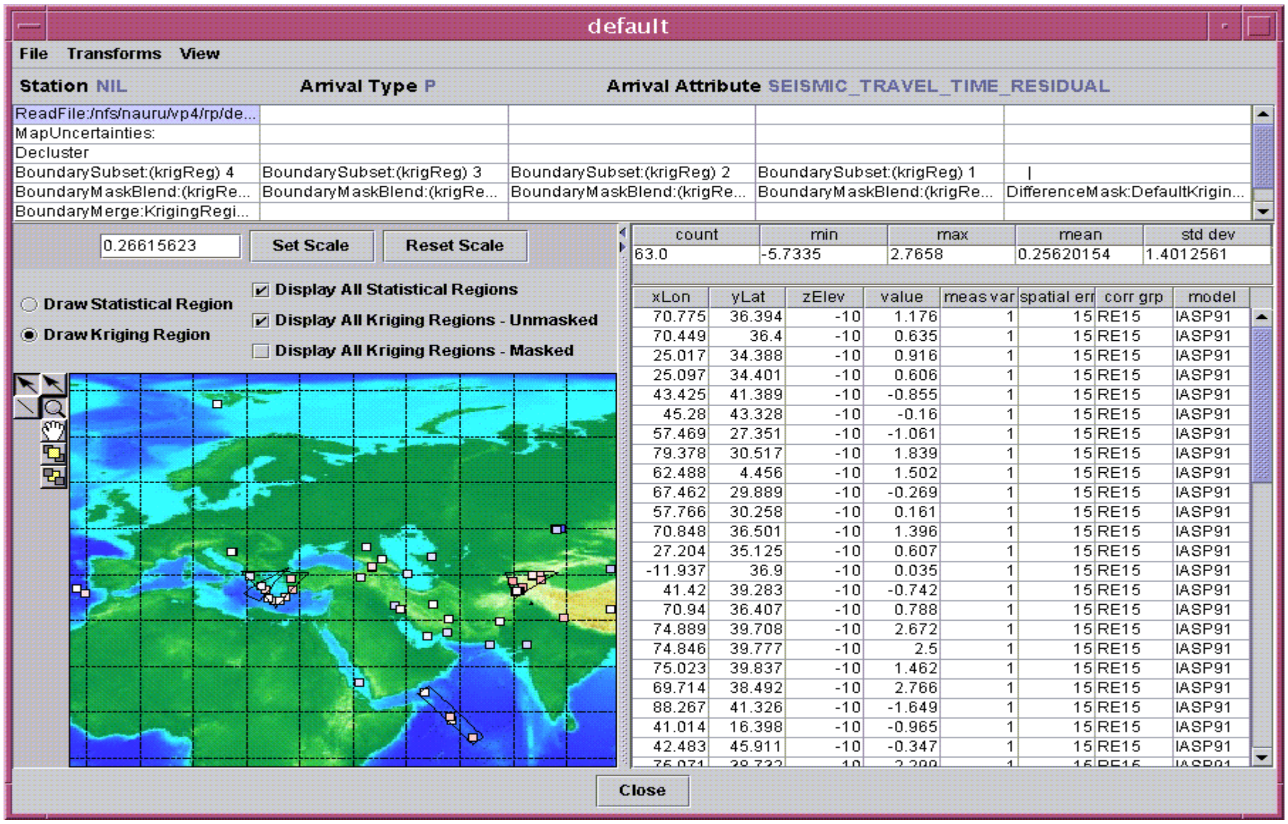

Figure 1. Kriging Set Window

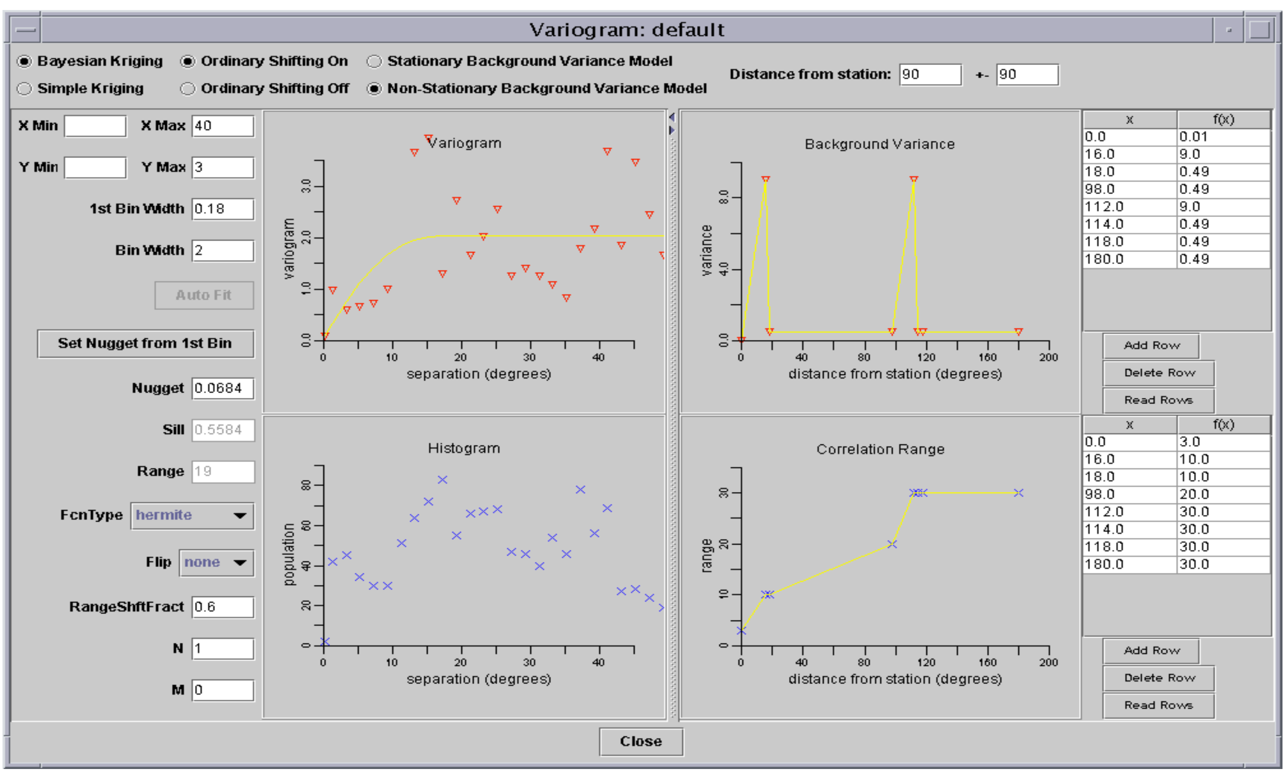

Figure 2. Variogram Window

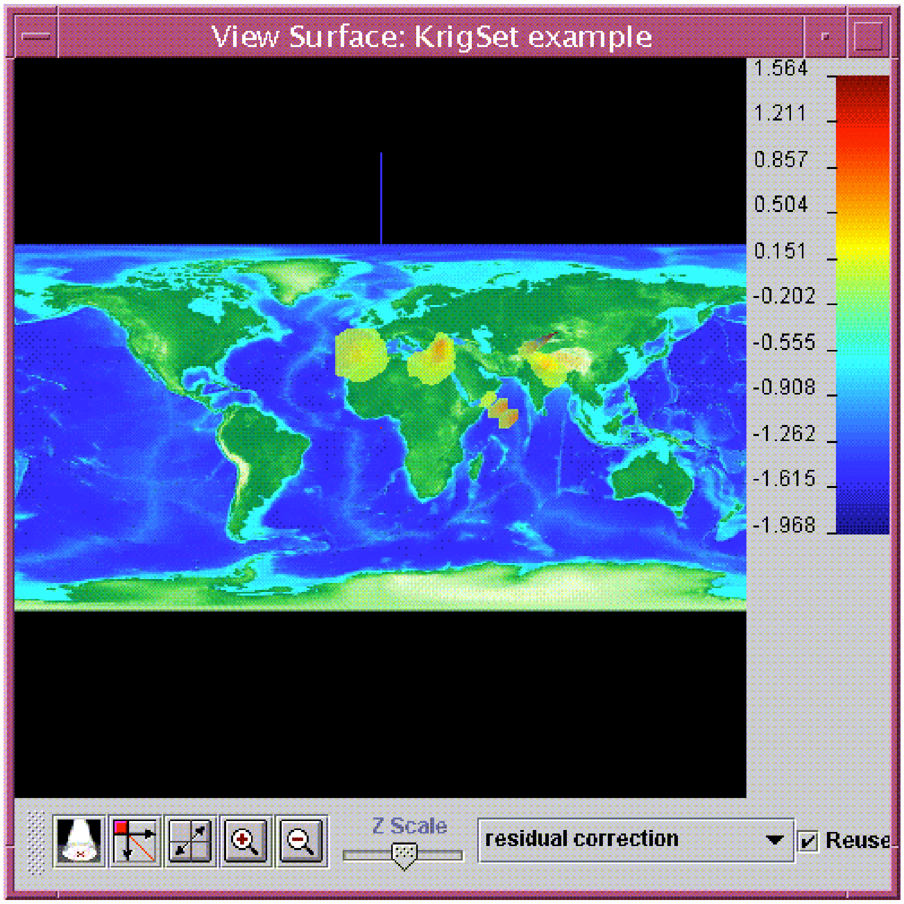

Figure 3. View Surface Window

Parametric Grid Library (PGL)

The Parametric Grid Library (PGL) provides the internal representation and client interfaces to access the data and models utilized by all GNEM R&E tools. This includes tools that provide a means of populating the PGL (e.g. KBCIT described above) and those that require access to its data or models (e.g. U.S. National Data Center (USNDC) codes). The library can be divided into several distinct functional groups that provide all necessary library capability. These groups include:

1) A physical models group that supports all basic and complex seismic modeling, Bayesian kriging for providing enhanced path corrections to the seismic models, and multi-region polygonal representations to provide a means for smoothly transitioning from discrete models defined in different tectonic regions;
2) A geometry group that provides the interface and representation functionality to support a variety of 1, 2, and 3 dimensional representation schemes. These include parametric and tessellated curve, surface, and volume representations, and a 2D point location facility;
3) A tessellation group that provides a means of generating and subsequently representing tessellated data for purposes of data interpolation and spatial searching;
4) A database group that defines the means to store and access arbitrary object data in a rapid manner;
5) An interfaces group that provides a Run-Time Type Identification (RTTI) facility for creating, managing, and destroying all database object types, and various client Application Programming Interfaces (APIs) for interfacing directly with the PGL functionality; and finally
6) A general utility group for providing other types of PGL requirements such as geometric vector support, memory management, and core level base class definitions.

Figure 4, below, depicts the structure of the various PGL groups outlined above.

The supported physical models provided by the PGL include travel-time, slowness, and azimuth. Each model consists of one or more discrete seismic regions. The specific characteristics of each of the regions are collected into a single object called a Reference object. Each Reference object is comprised of a base model definition and an optional 3D regional model. The 3D regional model can be incorporated as a residual model that increments, or improves, the values in the base model object, or as a standalone base model. In the case of travel-time, the Reference object can also contain an ellipticity correction model.

The collection of regional Reference objects is associated within a neighborhood of adjacent polygons. Each polygon defines the boundaries of a unique seismic region. Each edge of each polygon in the neighborhood can be modeled as a transition zone with an inner and outer boundary. The transition zone definition provides a means of smoothly varying both statistical and physical properties from one seismic region to the next. In addition to the one or more Reference objects, which are associated in a polygon neighborhood, the standard physical model will also contain a Station object that defines the station location for which the model was created.

Finally, a model can contain a path correction surface, which improves the accuracy and/or error definition of the physical properties for which the model was constructed. Although the surface is general in structure, it is usually created as a statistical surface representation using the PGL Bayesian kriging model. The Bayesian kriging model (Schultz, et. al., 1998) provides a means of utilizing correlation characteristics of known seismic ground-truth events to interpolate and improve the model values. This method produces the correction while minimizing the variance of the predicted error and provides an estimate of the error associated with the interpolated (or kriged) value. The job of defining the correlation characteristics and related kriging parameters that are used by the Bayesian kriging technique is the primary function of the KBCIT population tool described in the previous section.

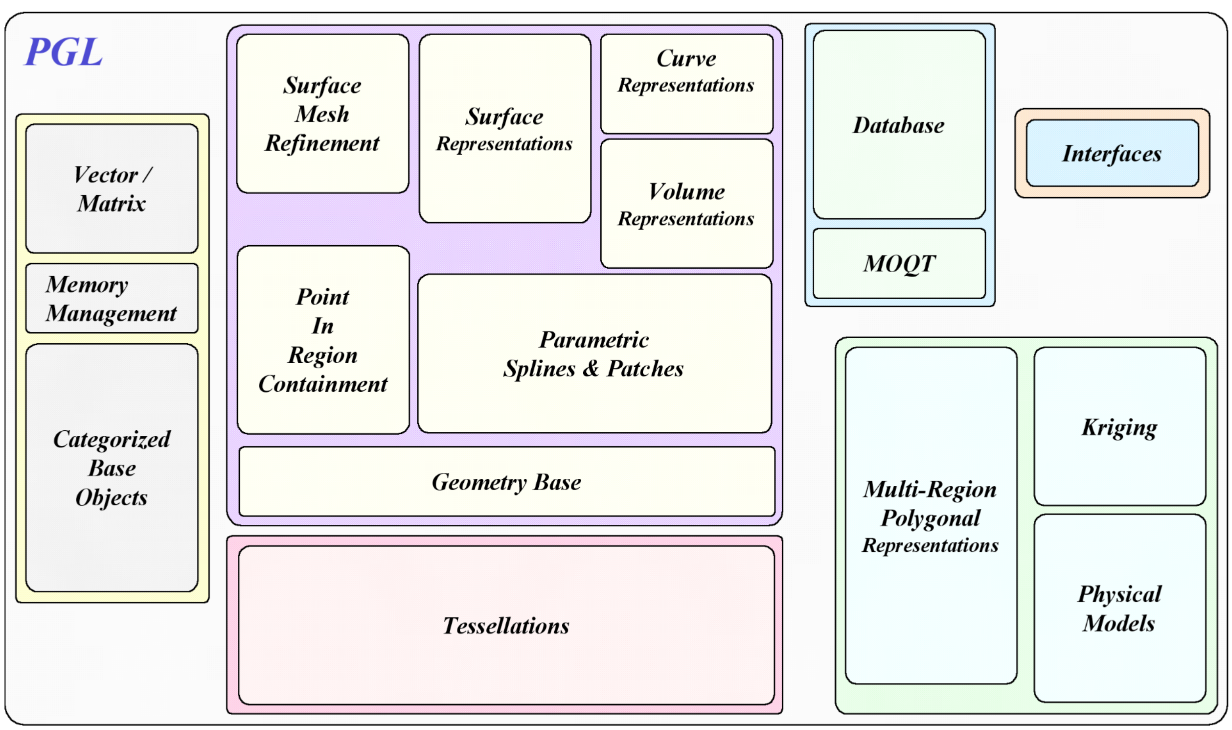

Figure 4 This Figure illustrates the six major components of the PG software library and various subcomponents of each group.

The PG Geometry group provides a means of representing data-centric curves, surfaces, and volumes. The primary means for representing geometric entities include parametrics and tessellations, and less often, a layered volume. Parametric representations can be defined when gridded data are available. A parametric representation can be formulated into linear, cubic, or quintic order polynomials in 1, 2, or 3 dimensions. A single parametric object can support multiple value definitions on the same grid. This provides an efficient means for storing multiple curves, surfaces, and volumes on a single unique grid. Full 1^{st} and 2^{nd} order derivative support is provided for all polynomial orders and spatial dimensions described above.

Unlike the parametric representation, the tessellated surface is used in cases where only unstructured 2 dimensional data are provided. The tessellated surface can also be used when high resolution is required in limited areas throughout a broader domain. Under these circumstances a tessellated surface can coarsen the unstructured mesh spacing in regions where low resolution is sufficient; and refine, or densify, the mesh spacing in regions where higher resolution is required. All PGL tessellations are Delaunay triangular (maximum angle minimized) tessellations. They can support constrained boundaries, which are guaranteed boundary representations, while maintaining the Delaunay criteria. This is important when exact representations of arbitrary boundaries (e.g. tectonic province boundaries) are required. Parametric representations are only capable of approximating unstructured boundaries.

Another major geometric object is the layered volume. This object utilizes a series of surfaces (parametric or tessellated) to represent a set of arbitrary layers. A typical use for a layered volume is to represent different stratified layers within the Earth. Layered volume objects are convenient when an independent parameters data cannot be easily or accurately represented in a rectilinear fashion (e.g. depth for the example above).

The PG tessellation group provides the functionality to construct, interpolate, and search tessellation objects. The interpolation aspects of tessellations are used to support geometric tessellated surfaces described in a previous paragraph. All tessellation interpolations utilize the gradient-based Natural Neighbor Interpolation (NNI) technique (Watson, 1992). Gradient-based NNI provides a continuous and smooth interpolation over any arbitrary unstructured Delaunay mesh.

The searching characteristics of the tessellation are utilized throughout the PGL hierarchy. Searching is used by the multi-region polygon neighborhood, described above, to locate all polygon transition boundaries that influence an arbitrary point location. The influencing point transitions are used to calculate a blending weight for each of the properties that are being interpolated in a physical model. Searching is used by the NNI interpolation technique to locate all triangles whose circum-circles contain the interpolation location. The surrounding set of triangle nodes are used to calculate NNI weights, which multiply the node values, whose summation produce the desired interpolant. Tessellated surface mesh refinement uses searching to locate new node insertion locations. The new nodes are added during mesh densification and mesh refinement operations. The Bayesian kriging algorithm uses tessellation searching to locate nearest neighbor ground-truth observations relative to a kriging interpolation location. The nearest neighbor observations are used in a "shifting" algorithm to transition between simple and neighborhood based kriging results depending on the proximity and spatial distributions of the observations about the interpolation location.

The PG database group provides the functionality to store and retrieve PGL objects from a flat-file or database in a generalized manner. Each object stored in the database is decomposed into a Meta-Data Key (MDK) and a Serialized Binary Stream (SBS). The MDK is used to contain summary information about the object including it's object type (class name), date of creation, location in the database of the associated SBS, pertinent attributes (station, phase, base model, frequency, etc.), a description string, and a unique 16 byte (128 bit) hash string, called the Key String Identifier (KSI). The KSI uniquely identifies an object's MDK for which it was created. Additionally, the MDK contains all dependency KSIs that are required to build the object associated with the MDK. All MDK objects are loaded from the database immediately after connecting to a database platform.

The SBS contains the actual concrete data necessary to construct the object. This data is stored separately in the database as a single binary blob. It is not accessed until the object is requested for creation from a client. The SBS is provided to the client (or the PG CGI object described below) to perform the instantiation.

The PGL interfaces group supports all current interfaces to PGL. These include the Common Grid Interface (CGI), a U.S. National Data Center (USNDC) code interface into the libloc library, an interface for the KBCIT population tool, and an interface for the Sandia seismic location code, LOCOO. Interfaces for the Data Management Tool (DMT) and the Viewing and Extraction Tool (VEXTool) are defined using the Java Native Interface (JNI) that connects directly into the PGL library. DMT and VEXTool are described below in the next two sections.

The CGI is used by all interfaces to create, evaluate, interpolate, and destroy all PGL maintained objects. Object instantiation requests are processed through CGI by providing CGI with the object's KSI. CGI will attempt to create the object associated with the KSI by first creating any of its dependent objects and ultimately by instantiating the requested object. The object can be evaluated or interpolated directly through the CGI generic interface, supported for all geometric objects, or by accessing the actual object using the Run-Time Type Identification (RTTI) facility provided by the CGI.

The last primary PG functionality group provides common utilitarian services to the all PG resources. This functionality includes a set of base classes used to divide the functionality into the groups defined here. The services also include a memory management object that allows pool allocation and destruction for small objects. This service is useful for efficiency reasons when many (near millions) small objects are repeatedly allocated and destroyed. Finally, the utility group provides a set of geometric vector classes for performing vector algebra in planar, spherical, and ellipsoidal surface geometries. These have proven to be useful and efficient constructs and are used throughout the entire PGL.

Data Management Tool (DMT)

The Data Management Tool (DMT) provides access to the PG data for purposes of managing the organization of the generated PG database, and for perusing the data for visualization and information purposes. The tool is written in Java providing a GUI front end and a JNI interface into PGL. The visualization component is primarily handled by the VEXTool application, which is the topic of the next section.

The database management facilities of the DMT include a means for inspecting the entire object reference and dependency hierarchy for any object in multiple databases simultaneously. These facilities provide mechanisms to copy or move objects and their dependents from one database to another or to simply remove them from their containing database entirely. These services allow a user to construct new databases from existing databases that use a subset of the objects contained in any of the existing databases. This provides a means for users to build databases that are tailored for spatially interesting regions or that contain objects that are specific for certain types of problems (e.g. seismic location). Figure 5, below illustrates a typical hierarchical dependency view of a database including object MDK listings and their dependents.

In addition to database object construction, the DMT also provides services for managing MDK modifications. These services permit a database administrator to correct errors or deficiencies in an MDK object in the database. Obviously, some of the MDK data cannot be changed such as object type, location of the SBS binary in the database, dependent object KSIs, etc. However, attribute and descriptive data can by modified and replaced back into the database without harming the overall database structure. This service is a key feature for performing routine maintenance on existing PG databases.

Another maintenance service involves generic searching facilities. Searching allows a user to discover specific patterns of data within the database. Search requests are formulated by supplying a generic search MDK object with the attributes of interest filled with desired patterns, and unimportant or irrelevant patterns left blank. All matches are returned to the user for inspection.

Finally, the database can be inspected for duplicate binaries. These are objects that have two or more MDKs that all point to SBSs that are duplicates of one-another. This can happened when identical objects are created at different times producing new MDKs with a different KSIs. These duplicates can be discovered and all but one may be removed. The DMT automatically updates the database dependency hierarchy to reflect the changes.

Aside from these features the DMT can also load individual SBSs for examining the contents of the serialized object in a textual format. This data is not available for change, as only a population tool is capable of modifying the data, but it does provide a means for a researcher or analyst to examine the defining data from which the object is composed. Also, as mentioned above, the DMT can request a visual inspection of an SBS by calling the primary VEXTool visualization service. This is described in more detail in the next section.

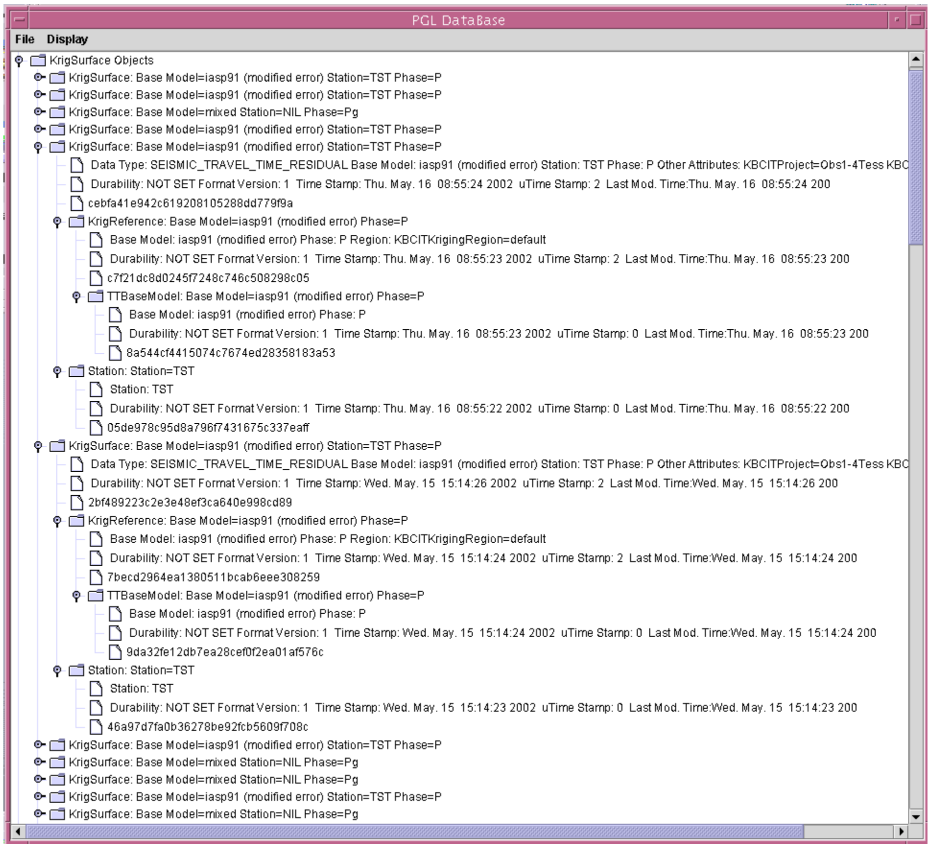

Figure 5. This Figure shows a typical screen shot of the DMT GUI displaying a hierarchical view of the database objects and their dependents.

Viewing And Extraction Tool (VEXTool)

The primary role of the VEXTool is to visualize the properties of any geometric object in a PG database, regardless of the objects type or dimension. It is also used to export interpolated model data to a number of third party tools utilized by GNEM researchers. The VEXTool interface was written in Java and provides a user friendly GUI for manipulating complex objects graphically. The tool utilizes a JNI interface for accessing the PG data and models directly through the PGL CGI object described above.

VEXTool can display an individual property of an object in 1, 2, or 3 dimensions or a range of properties as a set of curves or surfaces. Additionally, VEXTool can display multiple views of the same object simultaneously. Figure 6, illustrates a typical VEXTool screen dump where a property of a layered volume object is displayed simultaneously with a two dimensional cross-section and a 1-dimensional positional bore-hole.

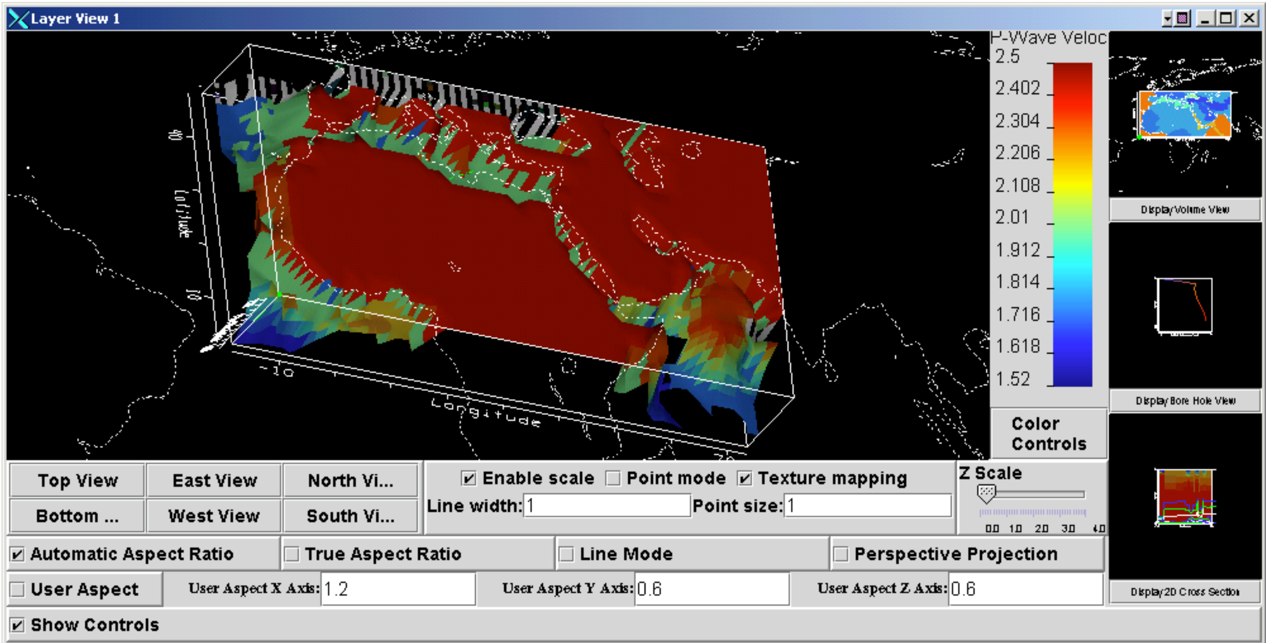

Figure 6. This Figure illustrates a typical VEXTool screen shot where a 3 dimensional layered volume object is shown displaying various geophysical properties in 3D, 2D cross-section, and 1D bore hole depictions.

CONCLUSIONS AND RECOMMENDATIONS

In this paper, researchers at Sandia Labs have defined a set of integrated software tools developed to take advantage of the improvements in Earth models available for seismic event monitoring. This integrated set of software, known as the Parametric Grid Software Suite, consists of data population, storage, modeling, and data management and viewing tools.

The population tool, KBCIT, provides interpolated calibration based information that can be used to improve monitoring performance by improving precision of model predictions and by providing proper characterizations of uncertainty. The underlying software representation of the Earth model is provided by the PGL. The PGL also provides a generalized object database for storing model data and a set of API interfaces utilized by the population, data management, and viewing tools. The PG data management facilities are handled by the DMT. The DMT is primarily responsible for maintaining and validating PG databases and for creating subset databases from existing ones. Finally, the VEXTool provides all model object viewing and data export facilities.

Future versions of the Parametric Grid Software Suite will continue to track with new enhancements in Earth models available for event monitoring. The rapid commitment of these new developments to software will ensure that the Earth model improvements are immediately available for use by GNEM researchers and analysts.

REFERENCES

Schultz, C., S. Myers, J. Hipp, and C. Young (1998), Nonstationary Bayesian Kriging: Application of Spatial Corrections to Improve Seismic Detection, Location and Identification, *Bull. Seism. Soc. Am.*, 1275-1288.

Watson, D. F., (1992), Contouring: A Guide To The Analysis and Display of Spatial Data, *Pergamon*, 321pp.