XIX International Conference on Water Resources
CMWR 2012
University of Illinois at Urbana-Champaign
June 17-22, 2012

AGNI: COUPLING MODEL ANALYSIS TOOLS AND HIGH-PERFORMANCE SUBSURFACE FLOW AND TRANSPORT SIMULATORS FOR RISK AND PERFORMANCE ASSESSMENTS

Velimir V. Vesselinov[*], George Pau[†], and Stefan Finsterle[†]

[*] Los Alamos National Laboratory (LANL)
Computational Earth Science Group, Earth and Environmental Sciences Division
EES-16, MS T003, Los Alamos, NM 87545, USA
e-mail: vvv@lanl.gov web page: http://ees.lanl.gov/staff/monty, http://ascemdoe.org

[†] Lawrence Berkeley National Laboratory
Earth Sciences Division
1 Cyclotron Rd., MS 90-1116, Berkeley, CA 94720, USA
e-mail: gpau@lbl.gov, SAFinsterle@lbl.gov

Key words: high-performance computing, parallelization, model coupling, sensitivity analysis, uncertainty quantification, parameter estimation, risk assessment, decision support

Summary. Advanced Simulation Capability for Environmental Management (ASCEM) is an emerging state-of-the-art scientific methodological and computational framework for analyzing and predicting subsurface flow and contaminant transport in natural and engineered systems. ASCEM is supported by the U.S. Department of Energy Office of Environmental Management (DOE-EM). The ASCEM framework is designed to provide a modular, object-oriented, and open-source high-performance computational environment that will facilitate assessments of risk and performance for DOE-EM contaminated site cleanup as well as remediation and closure decisions. The ASCEM framework includes three major codes: *Akuna*, *Agni* and *Amanzi*. *Amanzi* is a parallel high-performance computing subsurface flow and contaminant transport simulator. *Akuna* is an advanced Graphic User Interface (GUI) that incorporates various data management and model development tools. *Agni* provides the coupling of *Akuna* and *Amanzi* by managing model input and output interactions between the user interface and the simulator. *Agni* drives *Amanzi* simulations based on a series of model analysis tools including parameter estimation, geostatistical simulation, uncertainty quantification, risk assessment and decision support. *Agni* offers efficient parallelization (using system calls, threading or MPI) of the model analysis and simulation tasks in multiprocessor heterogeneous clusters by balancing the utilization of computational resources based on runtime performance. *Agni* allows coupling not only with *Amanzi* but also with any other existing flow and transport simulator. *Agni* is an object-oriented code written in C++ that allows for incorporation of additional model analyses tools.

1 INTRODUCTION

Restoration of U.S. Department of Energy (DOE) legacy waste sites for protection of important environmental assets remains an essential and substantial national effort. To date, the U.S. Department of Energy Office of Environmental Management (DOE-EM) has closed more than 80 sites, but still faces a daunting challenge at its most difficult sites. There is a general agreement that transformational changes are needed to develop tools and approaches to effectively guide and assess long-term cleanup priorities.

Recent advances in field technologies and scientific research have yielded new approaches to characterize subsurface systems, generated large volumes of scientific data, and driven significant advancements in fundamental understanding of subsurface multi-phase fluid flow and contaminant transport. According to the January 2009 Report to Congress on the Status of Environmental Management Initiatives to Accelerate the Reduction of Environmental Risks and Challenges Posed by the Legacy of the Cold War[1], DOE-EM is "implementing a planning approach that is national in scale and uses risk reduction as a major prioritization factor." This report also states that "the biggest challenges EM faces are those that have few precedents and fewer off-the-shelf technologies and processes to address them". There is a challenge to maximize the reduction of environment, safety, and health risks across the complex in a safe, secure, compliant, scientifically-defensible, and cost-effective manner.

Advanced Simulation Capability for Environmental Management (ASCEM) is a project supported by DOE-EM for development of a state-of-the-art and emerging scientific methodological and computational framework for analyzing and predicting subsurface multi-phase fluid flow and contaminant transport in natural and engineered systems. The ASCEM framework is developed through a collaborative effort across Department of Energy National Laboratories; the Los Alamos, Lawrence Berkeley and Pacific Northwest National Laboratories are the key participants. The computational framework is designed to be a modular, object-oriented, and open-source high-performance computational environment that will facilitate standardized assessments of risk and performance for DOE-EM contaminated site cleanup and closure decisions. The ASCEM computational environment includes three major codes: *Akuna*, *Agni* and *Amanzi*. *Amanzi* is a parallel high-performance computing (HPC) subsurface multi-phase fluid flow and contaminant transport simulator. *Akuna* is an advanced graphic user interface that incorporates various data management and model development tools based on Eclipse[2,3]. The architecture, design and development of *Agni* is described below. Currently, the ASCEM computational framework is still under development; however, some of the developed capabilities and tools have been successfully tested and applied for real world applications[3].

2 AGNI OVERVIEW

Agni is developed to provide the coupling of *Akuna* and *Amanzi* by managing model input and output interactions between the user interface and the simulator (Figure 1). The code is written in C++. Similar to other ASCEM modules, the code incorporates a series of Third-Party Libraries (TPLs) to simplify and generalize the code development effort. *Agni* is an object-oriented code based on a series general classes and libraries that allows for future incorporation of additional

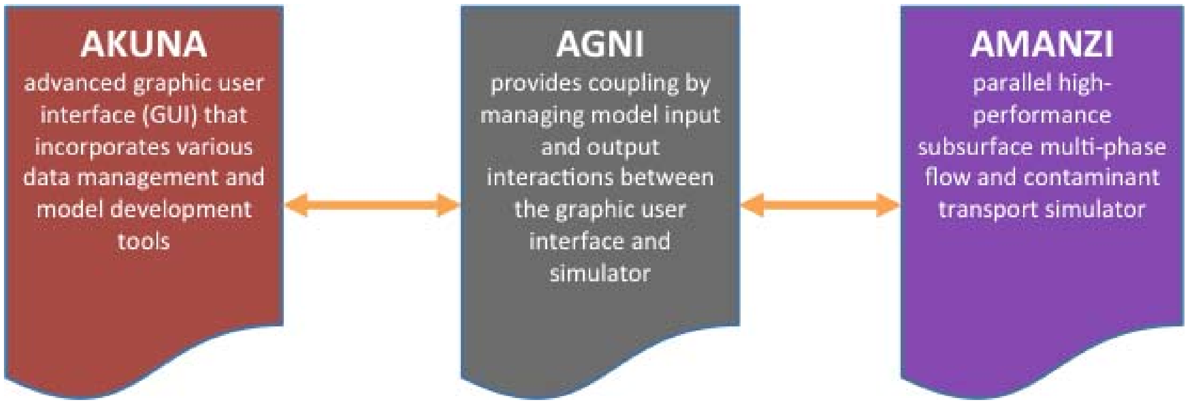

Figure 1: ASCEM computational framework includes Akuna, Agni and Amanzi.

model analyses tools.

Currently, the coupling provided by *Agni* is accomplished through a series of input and output files. All these files are prepared automatically by the ASCEM codes using Extensible Markup Language (XML). The *Agni* input files are generated by the graphic user interface (*Akuna*). These files provide information about (1) the selected numerical model, (2) various model parameters, (3) the model-based analyses that will be performed. The analyses utilize a series of tools integrated in *Agni* which include sensitivity analysis, parameter estimation, uncertainty quantification, risk assessment and decision support. All these analyses require the performance of a sequence of model simulations. In these simulations, model parameters are varied to estimate their impact on the model simulations. *Agni* is designed to propagate a single or a series of model parameter sets **p** to the simulator and collect the respective information from the obtained simulation results in the form of model outputs and corresponding observations **z(p)**.

Since most of these model runs driven by the model-based analyses are independent, it is computationally efficient to perform these runs in parallel. *Agni* offers a series of parallelization techniques based on external system calls, internal threading and MPI calls. The parallelization methods can be selected by the user depending on the computational environment. In addition, *Agni* will include methods that will provide automated load balancing based on runtime performance of individual jobs to improve the utilization of computational resources (currently under development).

Agni allows coupling not only with *Amanzi* but also with other existing flow and transport simulators. The coupling in this case is performed through a series of template and instruction files. The template files provide the current values of the model parameters to the external simulator. The template files are created by modification of the input files required by the simulator; the locations in the file where model parameters **p** that will be provided by *Agni* are replaced with special fields that will be filled with numerical data. Similarly, the instruction files are applied to read from the output files the model observations (predictions) **z(p)** obtained from the external simulator for the given set of model parameters **p**. The instruction file contains instructions that direct *Agni* to identify and extract from the simulator's output files the observable values **z(p)**. The coupling approach is similar to the one provided in PEST[4]. The external coupling subroutines are based on functionalities developed for MADS[5].

External Third-Party Libraries (TPLs) currently included in *Agni* are:

- MPI[6] – parallel execution;
- Teuchos[7] (part of Trilinos[8]) – importing and exporting Extensible Markup Language (XML) files;
- Blas[9], LAPack[10] and GSL[11] – linear-algebra and mathematical subroutines;
- GSLib[12] – geostatistical simulations;
- LevMar[13] – Levenberg Marquardt simulations;
- Evolving Objects (EO)[14] – an Evolutionary Computation Framework.

The building of the code uses CMake, a cross-platform, open-source build system that simplifies the build process and incorporation of TPLs. The code has been tested on various Unix platforms (Linux, Mac OS X, Cygwin MS Windows).

3 AGNI ARCHETECTURE AND INTERNAL DESIGN

The *Agni* architecture can be subdivided into three tightly interconnected components: **Toolset Driver**, **Model-Analysis Toolsets**, and **Simulation Controller** (Figure 2). The **Toolset Driver (TD)** provides communication between *Akuna* and the user-selected **Model-Analysis Toolsets (MAT)**. The **MAT** performs the requested analyses related to model parameters and predictions by interfacing with the simulator through the **Simulation Controller (SC)**. The **SC** performs the model simulations demanded by **MAT** and provides back to **MAT** the obtained simulation results. During the model executions, **SC** also provides information to **MAT** and *Akuna* about the current status of the model simulations (intermediate results, error messages, execution failures). In addition to controlling the communication between the MAT and SC, TD provides a suite of libraries that provide flexible data structures for describing the parameters and observations. They are designed to fulfill the various needs of MAT.

The **Sensitivity Analysis (SA) Toolset** provides techniques to estimate the sensitivity of model predictions to conceptual model elements and model parameters. The SA techniques include local and global methods.

The **Parameter Estimation (PE) Toolset** provides techniques to estimate conceptual model elements and model parameters based on site observation data. It determines model parameters that produce simulated outputs that best match the observation data. The PE techniques include local and global methods. Most of the PE methods (especially the global ones) are computationally intensive and require large numbers of forward model simulations.

The **Uncertainty Quantification (UQ) Toolset** provides techniques to estimate the uncertainty of model predictions caused by uncertainties in conceptual model elements and model parameters. Most of the UQ methods are computationally intensive and require large

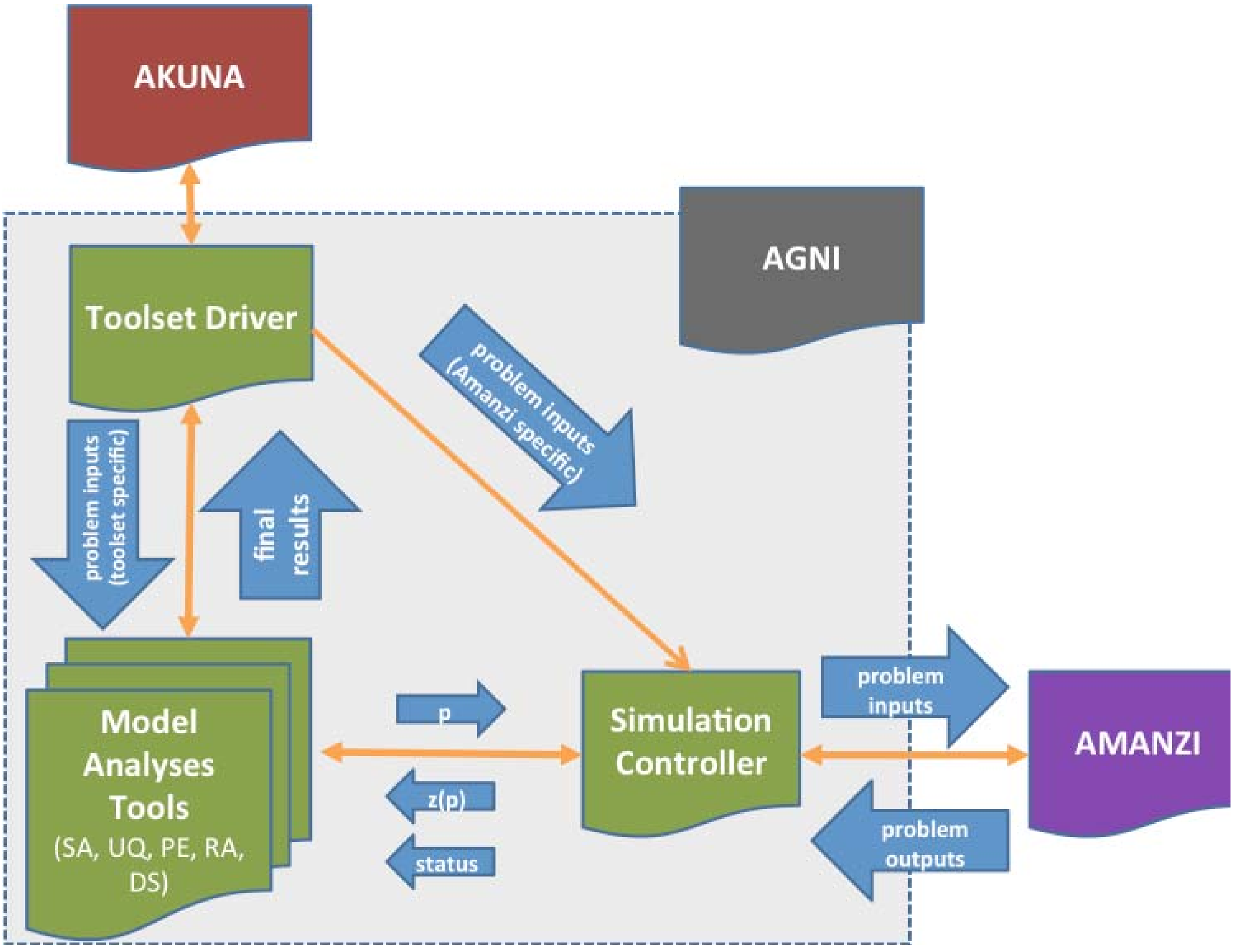

Figure 2: Internal Agni framework showing interconnection between Toolset Driver, Model-Analysis Toolsets and Simulation Controller. Model-Analysis Toolsets include Sensitivity Analysis (SA), Uncertainty Quantification (UQ), Parameter Estimation (PE), Risk Assessment (RA) and Decision Support (DS).

numbers of forward model simulations.

The **Risk Assessment (RA) Toolset** provides risk assessors, environmental data generation personnel, fate and transport modelers, managers, and stakeholders with an ecological and human health risk interface to the other ASCEM toolsets to support the overarching decision criteria of the environmental problem under consideration. The RA toolset provides a comprehensive set of risk data and tools for analyzing risk from concentration data generated by ASCEM (both measured and modeled) within a flexible framework that allows application to any decision driver, such as regulatory compliance.

The **Decision Support (DS) Toolset** provides techniques to support the iterative characterization and decision-making feedback to the site application users (managers, stakeholders, regulators) based on predefined performance criteria (data quality objective, compliance points, budget, execution timeline, etc.). Typically, prioritization of data collection activities and optimization of the level of complexity used for modeling are primary challenges for the efficient and cost-effective implementation of a science-defensible risk analysis and performance assessment. The DS objective is to identify and address key uncertainties in model predictions and environmental risk and reduce costs as a part of the iterative characterization and decision-making approach.

Currently, most of the toolsets are still under development, with substantial progress achieved in the development of the PE and UQ Toolsets**.** ASCEM HPC capabilities make possible the use of high-fidelity complex physics models to be applied for Sensitivity Analysis (SA), Uncertainty Quantification (UQ), Parameter Estimation (PE), Risk Assessment (RA) and Decision Support (DS) without the need for model oversimplification.

4 CONCLUSIONS

Agni enables the efficient utilization of advance simulation codes such as *Amanzi* in model analysis tools including sensitivity analyses, parameter estimation, uncertainty quantification, risk assessment and decision support. It also offers efficient parallelization (using system calls, threading or MPI) of these tools in multiprocessor heterogeneous clusters by balancing the utilization of computational resources based on runtime performance. For increased flexibility, *Agni* is also able to invoke other existing flow and transport simulators through a protocol that is similar to PEST. *Agni* has been carefully designed such that the toolsets can be easily expanded and new toolset can be added. It thus encourages future users or developers to contribute to the development of this community code.

5 ACKNOWLEDGEMENTS

This work was funded by the DOE EM-32 Office of Technology Innovation and Development.

REFERENCES

[1] DOE-EM, Report to Congress: Status of Environmental Management Initiatives to Accelerate the Reduction of Environmental Risks and Challenges, DOE/EM-0004, January, (2009).

[2] Eclipse Foundation, http://www.eclipse.org/, 2012.

[3] K.L. Schuchardt, D.A. Agarwal, S. Finsterle, C. Gable, I. Gorton, L. Gosink, E. Keating, C. Lansing, J. Meyer, W. Moeglein, G. Pau, E. Porter, S. Purohit, M Rockhold, A. Shoshani, C. Sivaramakrishnan, AKUNA – Integrated toolsets supporting advanced subsurface flow and transport simulations for environmental management, CMWR XIX, (submitted), (2012).

[4] J. Doherty, “Ground Water Model Calibration Using Pilot Points and Regularization”, *Ground Water, 10.1111/j.1745-6584.2003.tb02580.x*, (2003).

[5] V.V. Vesselinov, D. Harp, “MADS: Model analysis and decision support in C”, *http://mads.lanl.gov*, (2012).

[6] Open MPI: Open Source High Performance Computing, http://www.open-mpi.org/, (2012).

[7] Teuchos, http://trilinos.sandia.gov/packages/teuchos/, (2012).

[8] Trilinos, http://trilinos.sandia.gov, (2012).
[9] BLAS (Basic Linear Algebra Subprograms), http://www.netlib.org/blas/, (2012).
[10] LAPACK: Linear Algebra PACKage, http://www.netlib.org/lapack/, (2012).
[11] GSL: GNU Scientific Libraryhttp://www.gnu.org/software/gsl/, (2012).
[12] GSLIB: Geostatistical Software Library, http://www.gslib.com/, (2012).
[13] M.I.A. Lourakis, LevMar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++, http://www.ics.forth.gr/~lourakis/levmar/, (2004).
[14] Evolving Objects (EO): an Evolutionary Computation Framework, http://eodev.sourceforge.net/, (2012).