Environmental Modelling & Software 55 (2014) 176–189

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A high-performance workflow system for subsurface simulation

Vicky L. Freedman [a,*], Xingyuan Chen [a], Stefan Finsterle [b], Mark D. Freshley [a], Ian Gorton [d], Luke J. Gosink [a], Elizabeth H. Keating [c], Carina S. Lansing [a], William A.M. Moeglein [a], Christopher J. Murray [a], George S.H. Pau [b], Ellen Porter [a], Sumit Purohit [a], Mark Rockhold [a], Karen L. Schuchardt [a], Chandrika Sivaramakrishnan [a], Velimir V. Vessilinov [c], Scott R. Waichler [a]

[a] *Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99352, USA*
[b] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA*
[c] *Los Alamos National Laboratory, 30 Bikini Atoll Rd, Los Alamos, NM 87545 94720, USA*
[d] *Carnegie Mellon Software Engineering Institute, 4500 Fifth Avenue, Pittsburgh, PA 15213, USA*

A R T I C L E   I N F O

Article history:
Received 2 October 2013
Received in revised form
15 January 2014
Accepted 25 January 2014
Available online 14 February 2014

Keywords:
Workflows
ASCEM
Akuna
Amanzi
Model calibration
Uncertainty analysis
Contaminant transport
Vadose zone

A B S T R A C T

The U.S. Department of Energy (DOE) recently invested in developing a numerical modeling toolset called ASCEM (Advanced Simulation Capability for Environmental Management) to support modeling analyses at legacy waste sites. This investment includes the development of an open-source user environment called Akuna that manages subsurface simulation workflows. Core toolsets accessible through the Akuna user interface include model setup, grid generation, sensitivity analysis, model calibration, and uncertainty quantification. Additional toolsets are used to manage simulation data and visualize results. This new workflow technology is demonstrated by streamlining model setup, calibration, and uncertainty analysis using high performance computation for the BC Cribs Site, a legacy waste area at the Hanford Site in Washington State. For technetium-99 transport, the uncertainty assessment for potential remedial actions (e.g., surface infiltration covers) demonstrates that using multiple realizations of the geologic conceptual model results in greater variation in concentration predictions than when a single model is used.

Software availability

Name of software: Akuna
Developers: Consortium of National Laboratories (PNNL, LBNL, LANL)
Contact: tim.scheibe@pnnl.gov, vicky.freedman@pnnl.gov
Hardware requirements: Desktop computer with at least 4 GB memory
Software requirements: 32-bit Windows XP/Windows 7, 64-bit Windows 7 and 64-bit Mac OS
Programming language: 64-bit Java 6
Availability: Web links for downloading executable and tutorial files at http://akuna.labworks.org/download.html
Cost: Free

1. Introduction

Significant complexity is involved in computational simulation, including preparing data for input, executing multiple simulations, visualizing results and tracking the data that evolve from multiple analyses. The overall process is not readily amenable to automation, since each step usually requires that the modeler examine the results before proceeding to the next step in the analysis. Moreover, the process of data preparation, execution, analysis and decision-making is often followed by even more data preparation, execution, analysis and decision-making as the investigation proceeds. This process can occur over long time periods, and can involve significant user interaction. For example, an environmental computational analysis can require a repetitive cycle of moving data to a supercomputer or workstation for analysis and simulation, launching the simulations, and managing the storage of the output results. To step through this workflow, modelers typically make extensive use of batch files, shell scripts and scripting-language

* Corresponding author.
E-mail address: vicky.freedman@pnnl.gov (V.L. Freedman).

1364-8152/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2014.01.030

programs to link the sequence of applications needed to complete the analysis. For large data sets, data reduction techniques and parallel visualization may be needed to analyze results generated from the simulations.

Numerical models are frequently used to assess future risks, support remediation and monitoring program decisions, and assist in design of specific remedial actions for complex systems. These decisions are often made with incomplete information, and the impacts of knowledge gaps need to be quantified. Subsurface science is not the only environmental discipline that faces the challenge of making management decisions in the presence of significant uncertainty. Modeling is used in policy and decision-making for other disciplines, such as climate change (e.g., Li et al., 2014; Stainforth et al., 2006), sustainable development (e.g., Mortberg et al., 2013; De Lara and Marinet, 2009) and future energy supplies (e.g., Arnette, 2013; Jebaraj and Iniyan, 2006). Given the importance of identifying uncertainty, several freely available software packages, such as PEST (Doherty, 2010a, 2010b) and UCODE (Poeter et al., 2005) have emerged for uncertainty quantification. Web-based distributed modeling architectures (Bastin et al., 2013) have also emerged to assist modelers in quantifying uncertainty. However, uncertainty quantification can be extremely computationally intensive, requiring many model runs for their implementation, thus making their deployment difficult without high performance computing (i.e., supercomputers).

The U.S. Department of Energy (DOE) has recognized the need for high performance computing and has recently made investments in developing computational tools that can be used to predict the long-term behavior of subsurface contaminant plumes. Remediation of legacy DOE wastes is one of the most complex and technically challenging cleanup efforts in the world, with costs over the next few decades projected to be $265-305 billion (USDOE, 2008). The Advanced Simulation Capability for Environmental Management (ASCEM) program currently underway uses state-of-the-art scientific tools for integrating data, scientific understanding and software. One of the key features of ASCEM is the user environment, Akuna, which is a customized interface for managing subsurface modeling workflows. Akuna provides users with a range of tools to manage environmental and simulator data sets, translate conceptual models to numerical models (including grid generation), execute simulations, and visualize results. Additional toolsets provide users with methods for sensitivity analysis, model calibration and uncertainty quantification.

Several different scientific workflow systems exist [e.g.,Triana (Churches et al., 2006), Pegasus (Deelman et al., 2005), Kepler (Ludascher et al., 2006) and Taverna (Oinn et al., 2004)]. Some of these systems target a particular scientific domain (e.g., Taverna) while others are more generic (e.g., Triana and Kepler). For example, the Kepler system (http://www.kepler-project.org), is used to create, coordinate and execute scientific workflows that can be customized to the user's needs. Typical domain scientists, however, do not have the programming expertise needed to customize Kepler to fit their workflows, and assistance from computer programmers is usually required.

In addition to standalone scientific workflow systems, the Linked Environments for Atmospheric Discovery (LEAD; Plale et al., 2006) project demonstrates how workflows can be used to solve problems specific to Earth system science by integrating different technologies such as web and grid services and workflow systems. Integration of data and model workflows is demonstrated in Turuncoglu et al. (2013), who discuss coupling an Earth System Modeling Framework (ESMF) with the Regional Ocean Modeling System (ROMS) and Weather Research and Forecasting Model (WRF). The focus of their work is on the development of portable and replicable simulation workflows to create self-describing models with common model component interfaces.

User interfaces are an important component of the workflow system. Commercial user interfaces (UIs) (e.g. GMS (2012), Visual MODFLOW (2012), and Groundwater Vistas (2012)) have been developed specifically for groundwater flow and transport using the MODFLOW (Harbaugh, 2005) family of codes, a U.S. Geological Survey simulator that is the de facto standard code for aquifer simulation. Akuna, however, is unique in four major aspects. The first is in its ability to facilitate both serial and high-performance computation (HPC) in a workflow environment already customized for subsurface modeling. Although it is specifically designed to work with the ASCEM simulator, Amanzi, it can be used with other simulators as long as it is set up to read and write that simulator's file formats. Second, unlike many of the UIs for MODFLOW, Akuna provides an interface for variably saturated and multiphase flow simulators, and is not restricted to groundwater only applications. A third distinguishing characteristic is that Akuna is an open-source, platform-independent UI that integrates with other open-source software (e.g., WorldWind (2012), VisIt (2012)) for providing the user with all of the tools needed to perform a complete modeling analysis from model setup, calibration and uncertainty quantification. Finally, Akuna provides a client-server architecture and collaborative user interface, enabling users to perform their modeling analysis cooperatively from disparate locations.

The primary objective of this paper is to demonstrate Akuna capabilities that have been developed to date. This is accomplished by using the BC Cribs Site as an example application for the workflow system. To this end, the large-scale disposal of liquid inorganic waste is simulated for this site, which is located at the Hanford Site in southeastern Washington State. These subsurface discharges were a byproduct of nuclear weapons production during the Cold War. The BC Cribs Site received nearly 140 Ci of technetium-99 (^{99}Tc) in approximately 39 million liters of water (Kincaid et al., 2006). To date, this contamination has migrated to approximately 70 m below ground surface (bgs) into a 107 m thick vadose zone. Remediation of the recalcitrant ^{99}Tc is receiving increased attention in recent years because of its long half-life (2.13×10^5 years), the difficulty posed by its location in the deep vadose zone, and its near-term threat to groundwater.

The Akuna software is used to demonstrate model setup, calibration and uncertainty analysis for the BC Cribs Site, and to develop a model that can be used for evaluating potential remediation alternatives. The impact of accounting for multiple geologic realizations in an analysis of future boundary conditions is evaluated, and could be potentially important for future remedial actions at the site. Throughout the example application, it is demonstrated that use of high-performance computing makes execution of multiple simulations feasible, and the Akuna toolset streamlines the process.

2. Akuna user environment

Akuna is an open-source, platform-independent user environment. It includes features for basic model setup, sensitivity analysis, parameter estimation, uncertainty quantification, launching and monitoring simulations, and visualization of both model setup and simulation results. Features of the model setup tool include visualizing wells and lithologic contacts, generating surfaces or loading surfaces produced by other geologic modeling software (e.g., Petrel (2012), EarthVision (2012)), and specifying material properties, initial and boundary conditions, and model output. Currently, the model setup tool is equipped with a rectilinear grid generator for generating structured grids (orthogonal elements with a uniform