SciDAC 2009 IOP Publishing
Journal of Physics: Conference Series **180** (2009) 012065 doi:10.1088/1742-6596/180/1/012065

Application of the SALSSA framework to the validation of smoothed particle hydrodynamics simulations of low Reynolds number flows

Karen Schuchardt[1], Jared Chase[1], Jeff Daily[1], Todd Elsethagen[1], Bruce Palmer[1] and Tim Scheibe[1]

[1]Pacific Northwest National Laboratory, PO Box 999 Richland WA 99352

E-mail: karen.schuchardt@pnl.gov

Abstract. The Support Architecture for Large-Scale Subsurface Analysis (SALSSA) provides an extensible framework, sophisticated graphical user interface (GUI), and underlying data management system that simplifies the process of running subsurface models, tracking provenance information, and analyzing the model results. The SALSSA software framework is currently being applied to validating the Smoothed Particle Hydrodynamics (SPH) model. SPH is a three-dimensional model of flow and transport in porous media at the pore scale. Because fluid flow in porous media at velocities common in natural porous media occur at low Reynolds numbers, it is important to verify that the SPH model is producing accurate flow solutions in this regime. Validating SPH requires performing a series of simulations and comparing these simulation flow solutions to analytical results or numerical results using other methods. This validation study has been greatly aided by the application of the SALSSA framework, which provides capabilities to setup, execute, analyze, and administer these SPH simulations.

1. Introduction

Many subsurface flow and transport problems of importance today involve coupled non-linear flow, transport, and reaction in media exhibiting complex heterogeneity. In particular, problems involving biological mediation of reactions and/or mineral precipitation/dissolution fall into this class of problems. We are developing a hybrid multi-scale modeling framework that integrates models with diverse representations of physics, chemistry, and biology across different scales (sub-pore, pore and continuum). The modeling framework takes advantage of advanced computational technologies including parallel code components such as parallel solvers and gridding. Components of the modeling framework are being constructed using the Common Component Architecture (CCA) toolkit [1]. As part of that effort, we have developed a three-dimensional model of flow and transport in porous media at the pore scale — that is, where the geometry of solid grains and pore spaces is explicitly represented. The numerical method used to solve the flow and transport equations is the mesh-free particle method called Smoothed Particle Hydrodynamics (SPH)[2][3]. The code can accommodate large three-dimensional (3D) simulations with arbitrarily complex pore geometry. However, the complexity of these systems presents challenges for code validation and verification. Therefore, a series of simpler 3D simulations, designed for verification of the SPH solution of the Navier-Stokes equations at low Reynolds numbers typical of porous media flow is currently under way. The *Support Architecture for Large-scale Subsurface Simulation and Analysis* (SALSSA) software is the second component of this effort, addressing the macro modeling processes such as running simulations and analyzing results. SALSSA integrates workflow, data management, and visualization in support of model development and validation activities. In this paper, we describe the application of the SALSSA software to the SPH validation study and present the features and capabilities of SALSSA in this context.

2. Validation of low Reynolds number flows

Fluid flow in porous media at velocities common in natural porous media (e.g., groundwater flow) occur at low Reynolds numbers and therefore it is important to verify that the SPH model is producing accurate flow solutions in this regime. The SPH model contains parameters that do not appear in the corresponding incompressible Navier-Stokes equations and it is important to identify value ranges for these parameters that lead to good solutions for the flow profiles in topologically complex media. These parameters include a speed of sound c, which manifests itself by introducing a finite compressibility into the flow, and a weighting function diameter h that is roughly analogous to the grid spacing in grid-based numerical methods. Ideally, solutions should be independent of these parameters in some limit.

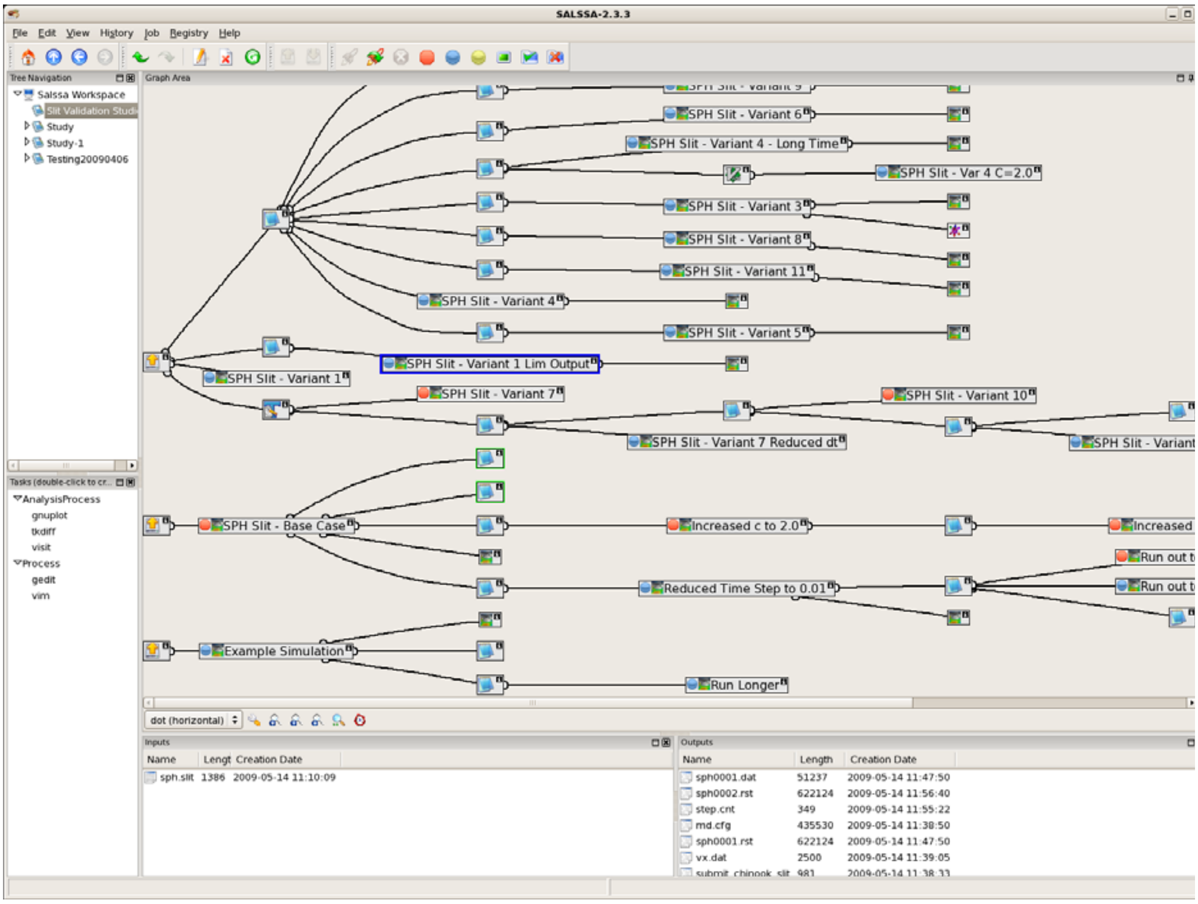
The approach taken in this study is to perform a series of simulations using relatively simple geometries where the flow solutions can be compared against analytical results or numerical results obtained using other methods, such as finite elements. The two flow geometries are slit flow between two parallel walls and flow in a cubic array of periodic spheres. The slit flow problem leads to a parabolic flow profile for which there is an analytic solution and the periodic sphere problem can be modeled using standard grid-based numerical methods. Both of the configurations will be simulated using SPH over a range of parameters and compared to validated solutions obtained by alternative means. The SPH solutions will be required to accurately reproduce the flow field and to introduce minimal density fluctuations. The second requirement is imposed to confirm that solutions correspond to relatively incompressible flow.

3. SALSSA workflow environment

For the SPH validation study considered here, users begin with an initial input file, launch the SPH code on a remote compute machine, and examine the output. Based on their analysis of the output, the input file of their previous run is edited, the model rerun, and the new outputs are analyzed. This process repeats until a conclusion is derived. The SALSSA workflow environment is composed of four primary components that support and facilitate this process: an Organizer tool component; a simulation execution component; a data management component; and an analysis and visualization component. Each of these components is described in more detail in the following sections, including how they specifically contribute to the SPH validation study.

3.1. Organizer

The main user interaction tool, the “Organizer”, provides access to the capabilities needed to conduct the validation study. This includes invocation of tools and applications used to manage and analyze simulation files, access to a workflow engine for simulation execution, provenance graph as well as spreadsheet views of tasks as they are conducted, and tabular views of simulation input and output data. It is structured around the concept of a “study” where a single study is a collection of setup, execution and analysis tasks. Figure 1 shows the Organizer tool with a provenance-graph representation of an SPH slit flow validation study. Each icon in the graph represents a task performed by the user. All information regarding these tasks, including inputs and outputs, time stamps, job information for simulation tasks, and an extensible set of additional metadata, is automatically recorded by the SALSSA data management component and presented in the Organizer tool. A text file-based registry system allows for the recognition and integration of additional desktop or remote applications. The Organizer emphasizes the use of existing software such as text editors, spreadsheets, and visualization tools though we provide a Python based programming model to add tools with custom interfaces and have developed a few including a grid visualizer and parameter study tools.



Inputs

Name	Lengt	Creation Date
sph.slit	1386	2009-05-14 11:10:09

Outputs

Name	Length	Creation Date
sph0001.dat	51237	2009-05-14 11:47:50
sph0002.rst	622124	2009-05-14 11:56:40
step.cnt	349	2009-05-14 11:55:22
md.cfg	435530	2009-05-14 11:38:50
sph0001.rst	622124	2009-05-14 11:47:50
vx.dat	2500	2009-05-14 11:39:05
submit_chinook_slit	981	2009-05-14 11:38:33

Figure 1. Screen snapshot of the Organizer and its use in the validation study. The main window contains an interactive graph that shows the provenance and relationships between individual tasks. The primary tasks used in this study were editors to adjust inputs, SPH simulations, and analysis sessions. Metadata, automatically extracted from data files, is available for each task along with a listing of all input and output files. The Organizer also supports spreadsheet style views of simulations and graph decluttering methods.

3.2. Simulation execution

The Simulation execution component includes a job Launcher tool, a customizable machine registration and job script generation mechanism, job execution workflows, job monitoring, and the ability to terminate individual jobs. The Launcher tool, invoked on simulations in the Organizer, collects user job submission inputs such as number of nodes and processors, and time, memory, and disk space limits (depending on the requirements of the target machine). This information, along with references to job input files is sent to the a workflow engine that stages files, submits the job requests, monitors the status of the jobs, and manages the disposition of outputs as directed by the user. Changes to the job states and data files are automatically sent to the Organizer. The Organizer also provides capabilities to monitor a running job through viewing the output file while it is being generated and terminate the job if necessary.

The simulation execution engine was implemented using the Kepler [4] open-source workflow system. Kepler has been applied to several scientific domains, and significant effort has gone into building components tailored for scientific workflow. By developing a custom Kepler workflow and augmenting it with a machine registry, we are able to submit and monitor arbitrary simulations run on a variety of machines ranging from Linux workstations, clusters, and supercomputers. For this study, simulations were run both at the Environmental and Molecular Sciences Laboratory (EMSL) and the National Energy Research Scientific Computing Center (NERSC). Our workflow is capable of launching multiple jobs to multiple machines while performing basic load balancing. This enables the automatic execution of a series of jobs such as parameter sensitivity studies.

SPH simulations may generate many very large files, which are typically left in place for analysis and often eventually moved to an archive. The user can specify the disposition of output files from within the Launcher tool. Later, these files can be staged to/from archives automatically as needed or under explicit user control. This capability is implemented as an additional workflow. The workflow system is also used to perform remote analysis tasks on data that resides on compute servers.

3.3. Data management

The data management component automatically tracks provenance (relationships between tasks) and inputs and outputs of each task, performs metadata extraction on data files to support search, browse, and graphical display, and performs MIME type detection. An extended version of the Open Provenance Model (OPM) [5] is used to represent provenance information. As a user executes tasks in the Organizer, relationships between data and processes are recorded using the OPM model, providing a complete record of a user’s activities. The provenance information is used by the Organizer to create the interactive graph of tasks, as shown in figure 1. The provenance is stored in an open-source Resource Description Framework (RDF) store called Sesame [6]. RDF is a directed, labeled graph data format, which provides a flexible platform to describe complex relationships. The provenance graph and metadata is queried using the RDF query language, SPARQL [7]. Data artifacts associated with each task are stored in a content management system based on Alfresco [8]. The data model used to track inputs and outputs also supports references to files in the form of a URI. These are used for large simulation outputs that may be archived or left on the compute server.

In addition to the provenance, additional metadata is extracted from input and output files as they’re introduced into the SALSSA environment. This capability relies on a registry system that provides rules for first establishing the MIME type of a file based on its name or, if needed, by examining the contents of the file. Once the type is known, the registry provides rules, in the form of regular expressions, to capture contextual metadata from the files. Both of these capabilities are readily extensible, by end users, through small Python scripts.

3.4. Analysis and visualization

SALSSA enables the analysis and visualization of the SPH validation runs through its registration of standard and custom applications. To date, we have integrated a range of analysis tools including TecPlot, General Mesh Viewer (GMV), Gnuplot, and Microsoft Excel, with current work focused on integrating the VisIt scientific visualization package. Depending on the analysis tool being used, the

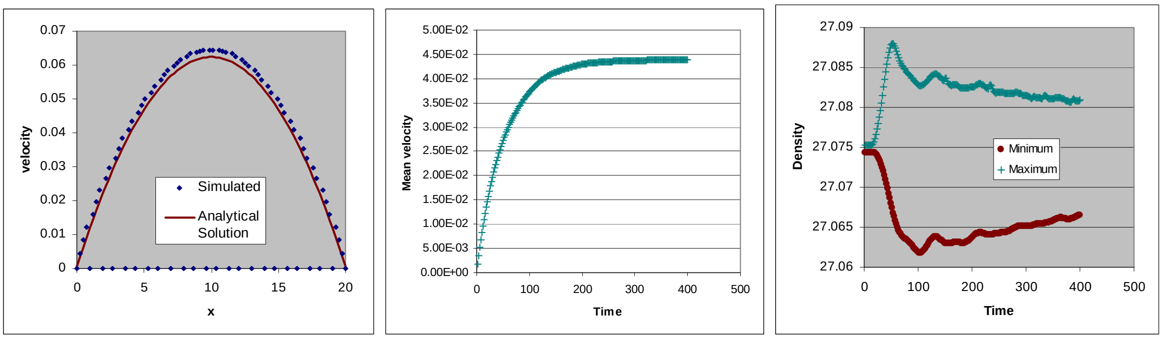


Figure 2. Results of a selected SPH validation simulation run. Left: comparison of a simulated velocity profile along a cross-section through the slit; Center: Time variation of simulated mean velocity as steady-state flow is approached; Right: Density variations within the simulated flow field as a function of time. All variables are presented in non-dimensional units.

SPH output may need to be translated to a different format. These translation scripts are also registered as tools. In this case, a custom Perl script parses the model output file and generates a time step based column-formatted text file. This file can be used within Microsoft Excel or Gnuplot to generate plots or perform further numerical analysis. Figure 2 shows examples of the plots generated from Microsoft Excel. The Gnuplot integration with SALSSA includes a custom interface that supports interactive scripting. Particle-based output from the SPH runs can be viewed directly, such as with GMV or VisIt [9]. Critical due to the size of the data, these analysis tasks can be accomplished from the user's desktop with the data residing where it was originally computed without having to stage it to a local workstation.

4. Summary

Validation of the SPH model is currently under way with tens of parameter configurations already executed and analyzed within the SALSSA workflow environment. SALSSA automates many tasks associated with running the simulation while transparently tracking data location and the provenance associated with tasks as they are executed. It emphasizes the use of familiar desktop tools while introducing new analysis tools specifically designed to handle large-scale simulation data. This case study has identified opportunities for ongoing research including improving the graphical presentation of large and complex workflows, automatic re-execution of a sequence of tasks, and automatic staging of large outputs from archival storage for analysis.

Acknowledgements

This research is supported by the U. S. Department of Energy's Office of Science under the Scientific Discovery through Advanced Computing (SciDAC) program. A portion of this research was performed using the Molecular Science Computing Facility at EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. The work was conducted at the Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RLO1830.

References

[1] Allan B A *et al*. 2006 A component architecture for high-performance scientific computing *The International Journal of High Performance Computing Applications* **20** 163-202
[2] Tartakovsky A M, Meakin P, Scheibe T D and Eichler West R M 2007 Simulations of reactive transport and precipitation with smoothed particle hydrodynamics *Journal of Computational Physics*. **222** 654-72
[3] Tartakovsky A M, Meakin P, Scheibe T D and Wood B D 2007 A smoothed particle hydrodynamics model for reactive transport and mineral precipitation in porous and fractured porous media *Water Resources Research.* **43**
[4] Ludäscher B, Altintas I, Berkley C, Higgins D, Jaeger E, Jones M, Lee MA, Tao J, Zhao Y. Scientific Workflow and the Kepler System. Concurrency and Computation: Practice and Experience. Volume 18, Issue 10, Pages 1039 – 1065.
[5] Moreau L, J Freire, J Myers, J Futrelle, and PR Paulson. 2007. "The Open Provenance Model." Presented by Luc Moreau at Workshop on Principles of Provenance, Edinburgh, Scotland on November 20, 2007
[6] The Sesame open-source RDF Framework. http://www.openrdf.org/
[7] The SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/
[8] The Alfresco Enterprise Content Management System. http://www.alfresco.com/
[9] The VisIt Visualization System. https://wci.llnl.gov/codes/visit/