

## **Dataset and Query Characteristics in Integration of Seismic and Simulation Data in Subsurface Modeling**

by **Tahsin Kurc** (Center for Comprehensive Informatics, Emory University)

Simulation based studies of subsurface fields (such as oil reservoirs) involve integration of data obtained from numerical models of the field with data collected from sensors. In the case of reservoir studies, data may come from seismic surveys of the subsurface field, especially if said field is located at the bottom of the sea.

### Datasets from Simulation of Numerical Models:

The subsurface field is represented by a 3-dimensional mesh. The mesh may be regular (i.e., mesh points are spaced equally from each other in the space) or irregular (i.e., mesh points may be unevenly and irregularly distributed in the space with dense population of mesh points in certain regions of the space and fewer mesh points in other regions).

A number of variables are computed at each mesh point. These variables represent the physical entities and phenomena simulated by the numerical model. For example, a mesh point may hold values for oil saturation, pressure, water density, fluid velocity in x, y, and z dimensions. Simulations are carried out over many time steps. Thus a simulation dataset consists of multiple 3D meshes, each mesh representing the simulation output at a given time step.

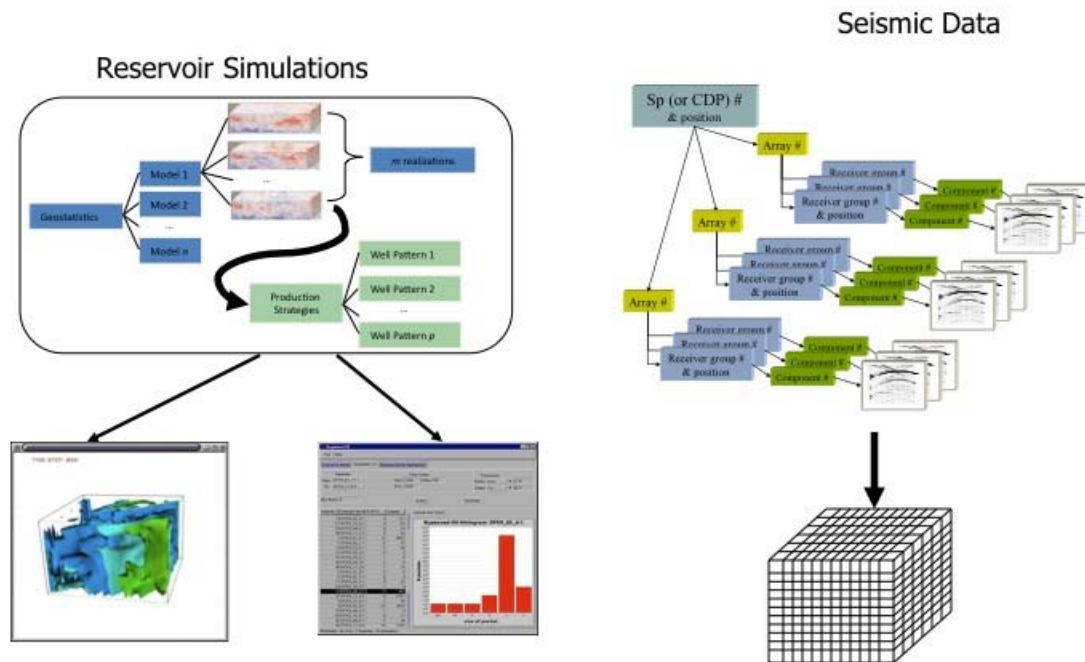
The types of queries into these datasets involve selection of one or more attributes from a subregion in the mesh. The queries may also select subregions over a specific range of time steps or over a specific set of time steps. Queries may also involve aggregation operations on data, e.g., sum or average of oil pressure at a set of mesh points over N time steps, or maximum value of oil pressure at a set of mesh points over N time steps.

In some studies, multiple simulations of the numerical models may be executed using different numerical model parameter values or different input parameters (e.g., locations and number of oil extraction wells in a reservoir simulation study). Thus, there is metadata associated with datasets in a database of this type of simulation data. Queries may involve this metadata as well as spatio-temporal aspects of the datasets. For example, a query may ask for simulations with a particular simulation parameter value and extract data from subregions and/or subsets of times steps from these simulations.

### Datasets from Sensor Measurements:

Sensor measurements collect data at multiple locations in the region under study. For example, seismic surveys of a reservoir field in the sea may gather data using an array of sound sources and receivers placed at the bottom of the sea. The receivers will collect reflections from subsurface rocks, etc., of sounds generated from sound sources over a period of time. The sensors are organized into lines and arrays, forming in a sense a 2-dimensional array. Each sensor reading is associated with attributes such as survey no (there may be multiple surveys of the field), line number, array number, receiver group no and position (in the array or in x and y dimensions), source number and position.

A query may request sound traces from a particular group of receivers (identified based on their receiver number or position in the array or their x-y position) and received from a set of sound producers. The query may also specify from which survey(s) the requested data should be extracted.



**Left:** Query of reservoir simulations. Multiple simulations using different numerical model parameters or input parameters (well patterns). Data is selected from a subset of simulations over a subset of time steps and subregion. The query results may be further analyzed and filtered using application specific filtering operations to computer such values as average oil pressure, regions of by-passed oil, etc. **Right:** Query of seismic data. Data subsets are selected based on source id and position, array number, receiver group and position (not shown in the figure is the query may also specify which survey(s) to choose data from). The results may be processed to form a three dimensional representation of the reservoir being studied.

### References:

1. M. Parashar, V. Matossian, H. Klie, S. G. Thomas, M. F. Wheeler, T. Kurc, J. Saltz, and R. Versteeg. "Towards Dynamic Data-Driven Management of the Ruby Gulch Waste Repository.", Computational Science – ICCS 2006: 6th International Conference, Reading, UK, May 2006. Proceedings, Part III. Lecture Notes in Computer Science, Volume 3993, 2006.
2. X. Zhang, B. Rutt, U. Catalyurek, T. Kurc, P. Stoffa, M.Sen, and J. Saltz, "Supporting Scalable and Distributed Data Subsetting and Aggregation in Large-Scale Seismic Data Analysis", The International Journal of High Performance Computing Applications, Vol. 20, No 3, pp. 423-438, 2006.

3. T. Kurc, U. Catalyurek, X. Zhang, J. Saltz, R. Martino, M. Wheeler, M. Peszyńska, A. Sussman, C. Hansen, M. Sen, R. Seifoullaev, P. Stoffa, C. Torres-Verdin, M. Parashar. "A Simulation and Data Analysis System for Large Scale, Data-Driven Oil Reservoir Simulation Studies", *Concurrency and Computation: Practice and Experience*, Vol. 17, No. 11, pp. 1441-1467, 2005

4. M. Parashar, H. Klie, U. Catalyurek, T. Kurc, W. Bangerth, V. Matossian, J. Saltz, M. Wheeler. "Application of Grid-enabled technologies for solving optimization problems in data-driven reservoir studies", *Future Generation Computer Systems*, Vol. 21, No. 1, pp. 19–26, 2005

5. S. Narayanan, T. Kurc, U. Catalyurek, J. Saltz. "Database Support for Data-driven Scientific Applications in the Grid", *Parallel Processing Letters*, Vol. 13, No. 2, pp. 245–273, June 2003.