A Multi-Output Gaussian Model Applied to Uncertainty Quantification of Seismic Problems Using Computational Simulation

Erb F. Lins
Fernando A. Rochinha
Alvaro L. G. A. Coutinho
Josias J. Silva

High Performance Computing Center
COPPE/Federal University of Rio de Janeiro
Rio de Janeiro, Brazil

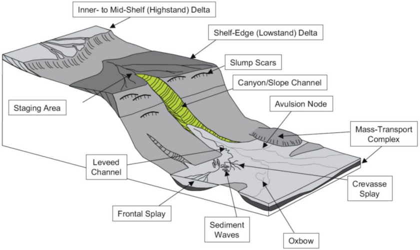
COPPE UFRJ

Outline

- Background and Motivation: High Performance Computing in the realm of Engineering and Applied Sciences
- Scientific Workflows and Uncertainty Quantification: Proof of Concept
- Computational Simulations: Surrogates
- Final Remarks and Future Steps

Motivation: Oil and Gas (and many other) applications: simulation of complex (multiscale - multiphysics) flows

Fluid-Object Interaction - Structural Hydrodynamics



Turbidity Currents - Sediment Deposition

Background

- Large Scale Algorithms + Petascale Computing push the envelope of Computational Science and Engineering
- Confidence (reliability) of simulations predictions make CS&E an effective tool
- Uncertainty Quantification + Validation: decision making
- Chain of codes involving high performance computation and a huge amount of data. Need of a efficient control strategy and tools for the analysis of output like provenance catalogue and queries within heterogeneous data

Our context: Oil and Gas (and many other) applications: simulation of complex (multiscale - multiphysics) flows

- A large amount of Brazilian oil reservoirs (indeed worldwide) were formed by the action of Turbidity Currents;
- Understanding reservoir geological formations may help decision making on well placement;
- Most of the studies in this area are still based on experiments or nature observation. Computer simulations might be transformed in an effective tool (at least simulations can help geologists to deeper analyze theirs theories);
- Highly coupled and non-linear problem: incompressible flow, polydisperse mixture, interaction of sand deposition and bottom morphology;
- Room for improvements in turbulence models and uncertainty quantification (UQ). Now, Seismic will play a role.
- BIG DATA.

Strategy

We are putting together three pieces:

- High Performance CFD code based on Large Eddy Simulation approach: Residual Based Variational Multiscale Method to model Particle Laden Flows.

  G. Guerra et al., Numerical simulation of particle-laden flows by the residual based variational multiscale method. International Journal of Numerical Methods in Fluid Mechanics. DOI: 10.1002/fld.3820

- Uncertainty Propagation: Stochastic Collocation (low stochastic input dimension) and Gaussian modeling – the need for computational surrogates

  J. Rohmer and E. Foerster, Global Sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta-modeling, Computers and Geosciences 37 (2011) 917–927

- Scientific Workflows Managing UQ. Guerra et al., Uncertainty Quantification in Computational Predictive Models for Fuid Dynamics Using a workflow Management Engine. International Journal for Uncertainty Quantification, v. 2, p. 53-71, 2012.

Seismic: Simulation and Uncertainties

Uncertainties are ubiquitous and pervasive.
Computational modeling (+ UQ) can help?

- ... Experimental design is becoming a widely accepted technique to handle geological and production uncertainty in E&P projects ...
  - Uncertainty Quantification to Evaluate the Value of Information in a Deepwater Reservoir. Portella et al. SPE 79707. Texas, 2003
- ... three major sources of uncertainty in the static model ...
  - Stochastic-Aided Design and Bayesian Updating: New Tools to Use Expert Knowledge in Quantitative Models That Incorporate Uncertainty. Alain Galli et al. SPE 90414. Texas, 2004
- ... The modelling performed ... greatly helped us to understand why the images were ...
  - Impact of modelling shallow channels on 3D prestack depth migration, Elgin-Franklin fields. J. Arnaud et al. EXTRACT. First Break, 28 (4), 2010
- Seismic 3D amplitude variation with offset (AVO) data from the Alvheim field in the North Sea are inverted into lithology/fluid classes, elastic properties, and porosity ... The inversion is phrased in a Bayesian setting.
  - Hierarchical Bayesian lithology/fluid prediction: A North Sea case study, Rimstad, Avseth, Omre, GEOPHYSICS, 77(2) 2012)
- 2013 SEG Summer Research Workshop Seismic Uncertainties and Their Impact

Seismic: Simulation and Uncertainties. Our view of UQ

Stochastic framework - parameters or (and) physical quantities are modeled as random variables (fields).

Physics - based models phrased as stochastic partial differential equations (SPDEs).

Bayesian techniques emerging as leading tools for analysis

Decision (risk analysis) making (Bayesian) workflow (suggested in Model-Based Uncertainty Quantification and Seismic Information Value. R. Gibson et al., SEG San Antonio Annual Meeting, 2007)

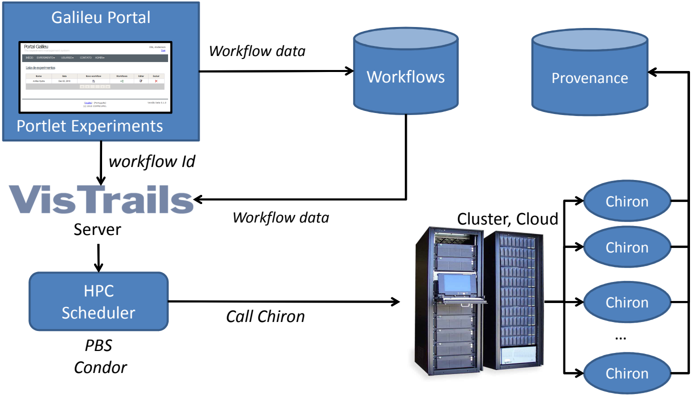
$$\pi_{post} := \pi(\mathbf{m}|\mathbf{d}) \propto L(\mathbf{d}|\mathbf{m})\ \pi_{prior}(\mathbf{m})$$

$\mathbf{d}$	seismic data
$\mathbf{m}$	petrophysical parameters (eg. porosity, thickness, water saturation)
L (likelihood)	composition of rock-physics (uncertain) modeling and a forward (uncertain) wave propagation solver

Scientific Workflows supporting High Performance Computing

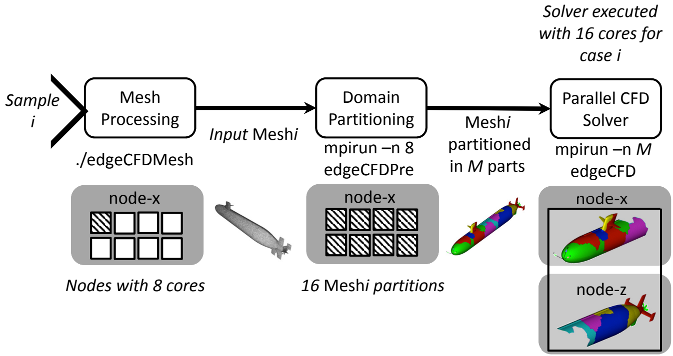
- Scientific/Engineering Computational Experiments Modeled as Scientific Workflows
- Simulations generate a lot of data: understanding how to manage and query simulation data in runtime
- Track who performed the computational experiment and who is responsible for its results Provenance data is automatically registered by SWfMS

Workflow (Sequential) Execution (previous version)



Chiron is a data-centric scientific workflow engine to support the execution of scientific workflows using an algebraic approach. Chiron allows for the parallel execution of workflows in clusters and it supports runtime provenance queries and workflow optimization.
Available at: http://sourceforge.net/projects/chironengine/

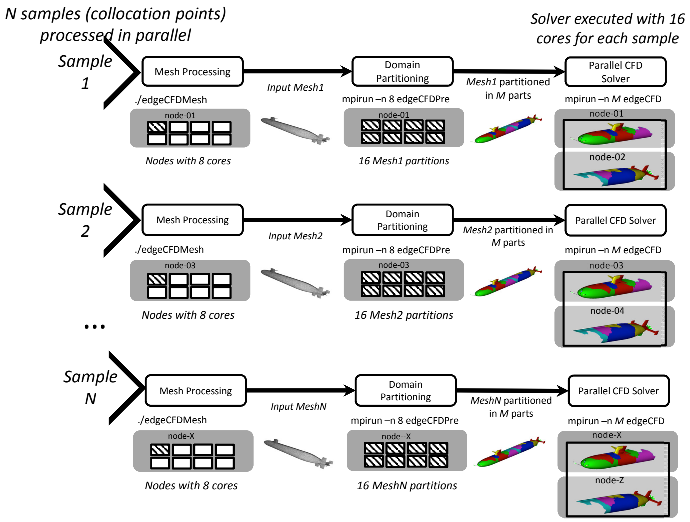
Provenance



Chiron is running in each core of each node:managing scheduling, fault-tolerance, provenance data gathering

Typical queries: check for convergence of the deterministic solver ; computation on the fly of high order statistics (two point correlation represents important QoI)for checking convergence regarding stochastic components

EdgeCFD + CHIRON: Two level parallelism



Proof of Concept Prototype (ongoing research and implementation)

- Non-intrusive UQ strategies: allow the use of complex already implemented codes or comercial codes
- Stochastic Collocation: low stochastic dimension. Multi-output Gaussian process (bayesian framework) already implemented but not opitmized
- Double level parallelization: exploring the stochastic space ; exploring built-in parallel Edge-CFD features
- Still more: space-time-stochastic adaptivity (provenance data and online queries); computing solution statistics (post-processing)
- Uncertainty on the initial conditions (initial scenario of the currents – Lesshaff et al., Towards inverse modeling of turbidity currents: The inverse lock-exchange problem. Computer & Geosciences, 2011) and on settling velocity

Proof of Concept: Turbidity Currents Uncertainty Propagation

UQ Techniques

Forward Solvers:

- Monte Carlo
- Non-intrusive Stochastic Galerkin
- Adaptive Sparse Grid Collocation
- Bayesian Techniques (using Multi-output Gauss Processes)

Inverse Analysis:

- Markov Chain Monte Carlo (MCMC)
- Langevin Methods
- Stochastic Newton
- Particle Filters and Sequential Monte Carlo

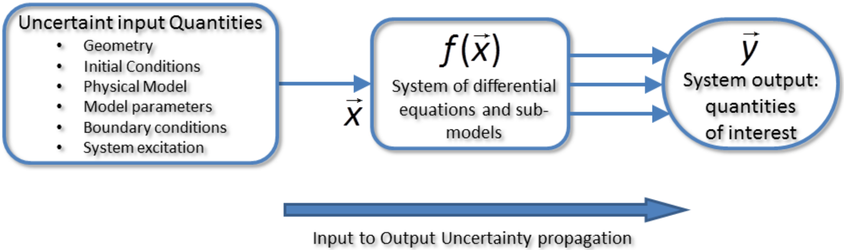
All quite computationally intensive. Specially in high-dimensions ("Curse of Dimensionality").

$$\pi_{post} := \pi(\mathbf{m}|\mathbf{d}) \propto \underbrace{L(\mathbf{d}|\mathbf{m})}_{need\ \ for\ \ surrogates} \pi_{prior}(\mathbf{m})$$

Computational Surrogates

Objective

Compute output statistics using as few as possible deterministic solver calls.



Surrogates: a Bayesian Approach

- Construct a probability measure over the possible surrogate for the computational code using a finite number of "observations" (code outputs). Gaussian processes (local or global)
- Sample the surrogate and estimate error bars
- Compute statistics of the outputs (MCMC sampling method

J. Oakley, A. O'Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika 89 (4) (2002) 769-784

I.Bilionis, N. Zabaras. Multi-output local Gaussian process regression: applications to uncertainty quantification. Journal of Computational Physics 231 (17) (2012) 5718 - 5746.

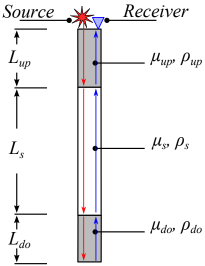
I.Bilionis, N. Zabaras, B. A. Konomi, G. Lin. Multi-output separable Gaussian process: Towards an efficient fully Bayesian paradigm for uncertainty quantification. Journal of Computational Physics 241 (2013) 212–239.

Application in Seismic Reflection

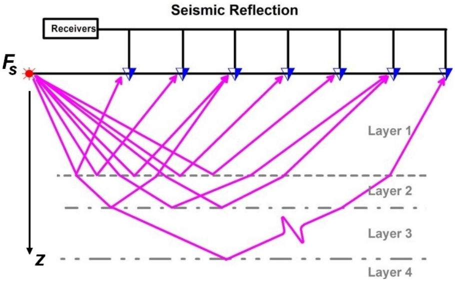
Applications

Given some uncertainty in a multilayered domain, how much uncertainty is expected in measurements?

We treat the following problems in uncertainty framework:



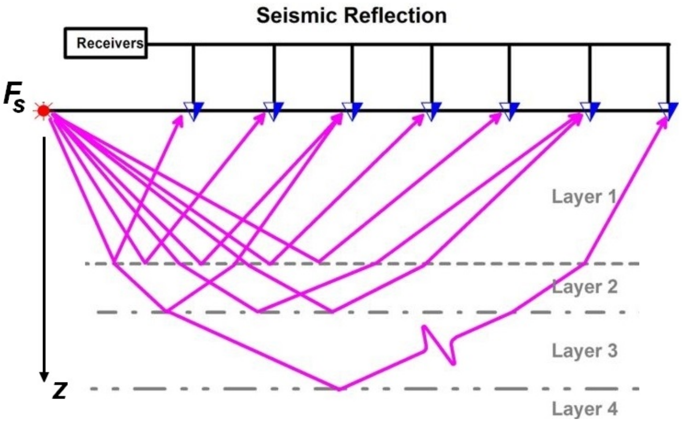
1D Problem



2D Problem

Seismic Forward Problem: the General Setting

Consider the seismic problem of determining the travel time of a seismic wave in a multilayered elastic medium.



Problem Definition

This can be modeled using the wave equation

$$\rho(x)\frac{\partial^2 \mathbf{u}}{\partial t^2} - \frac{\partial}{\partial \mathbf{x}}\left(\mu(\mathbf{x})\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\right) = F_s(\mathbf{x}, t) \tag{1}$$

where $\mathbf{u} \equiv \mathbf{u}(\mathbf{x}, t)$, set over time $t = [0, \infty[$, and spatial independent variable $\mathbf{x} \in \mathbb{R}^3$.
The propagation media is described by material constants $\rho(\mathbf{x})$ and $\mu(\mathbf{x})$, the materials density and stiffness. The source is defined by $F_s(x_s, t)$.

Stochastic Variables

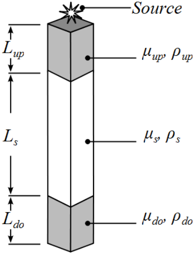
One representation for this problem stochastic variables could be set as

$$\chi = \bar{\chi} \times (1 + \sigma^2 \xi)$$

where $\bar{\chi}$ is the mean value of χ, σ^2 is the variance and ξ is any random variable with zero mean. Here, we assumed $\xi \in [-1, 1]$ is a random uniform variable.

Random Variables Model

Then we could have $\mu_{up}(\xi_1)$, $\rho_{up}(\xi_2)$,$L_{up}(\xi_3)$, for the first layer, and similarly to second and third layer. There is 8 i.i.d. random vars.



Mean Values to random parameters

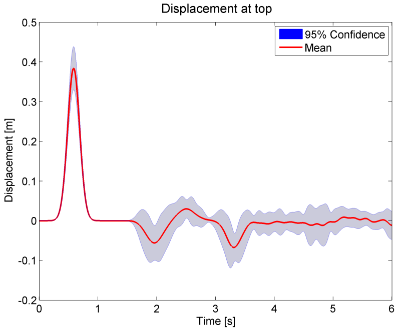
Stiffness $[\times 10^9 kg/(ms^2)]$	Density $[kg/(m^3)]$	Length $[km]$
$\bar{\mu}_{up} = 3.63$	$\bar{\rho}_{up} = 2440$	$\bar{L}_{up} = 1.4$
$\bar{\mu}_s = 21$	$\bar{\rho}_s = 2300$	$\bar{L}_s = 1.3$
$\bar{\mu}_{do} = 10.3$	$\bar{\rho}_{do} = 2650$	$\bar{L}_{do} = 3.7 - \bar{L}_{up} - \bar{L}_s$

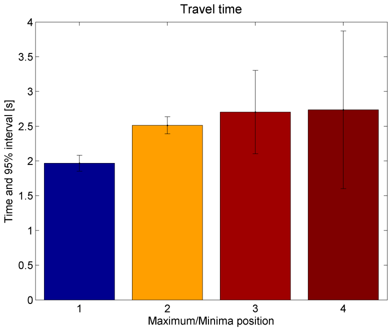
We set $\sigma^2 = 0.1$ to all vars.

What uncertainty could we expect to pressure sensors and to travel time?

Results

Displacement at top sensor. The chart at right shows the time position of 4 max/min values after direct wave arrival.

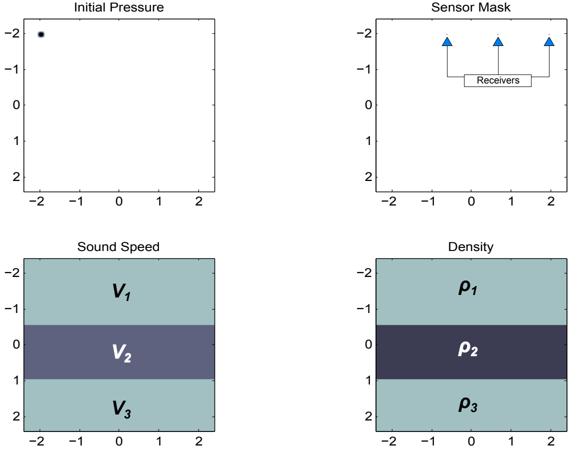




This information is useful to estimate the position of layers interface. Given this data, the uncertainty in first interface position is $\sim 230m$ and in second interface $\sim 580m$.

Problem Definition - 2D

In this work we consider a two dimensional model with source and receiver at the same vertical position, 3 sensors and three layer with different material and sizes.

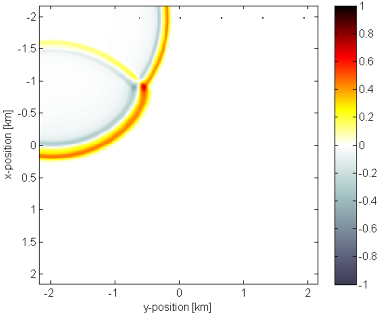


Deterministic solution

- A FD scheme in space and time can be used to solve this problem. With this we will have the pressure at each point x, y over time t.

- We solve the following PDE at each timestep (and set PML boundary conditions at the domain boundaries.)

$$
\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} &= \frac{1}{\rho_0} \nabla p \\
\frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \cdot \mathbf{u} \\
p &= c_0^2 \rho
\end{aligned}
$$



Deterministic solver

- A deterministic solver is necessary to compute the pressure at sensors and this data will be used to create the *meta model*.
- We used a free available toolbox for Matlab to compute the pressure distribution (k-Wave toolbox).
- The approach used here is totally non intrusive: there is no need to change your current solver.
- In this work, we used a 256×256 grid with a 10 points band to PML.

k-Wave
A MATLAB toolbox for the time-domain simulation of acoustic wave fields

k-Wave is an open source acoustics toolbox for MATLAB developed by Bradley Treeby (Australian National University) and Ben Cox (University College London). The software is designed for time domain acoustic and ultrasound simulations. The simulation functions are based on the k-space pseudospectral method.

www.k-wave.org

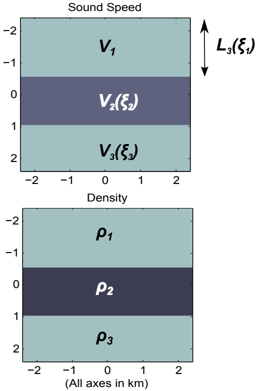
Deterministic solution

Pressure field over time

Random Variables Model

- The parameters regarding this problem are most adequately represented in a uncertainty framework.
- Consider velocities and position of interface in this problem as random vars. The values to material are set as

$$
\begin{aligned}
L_1(\xi_1) &= 2 \times (1 + \sigma^2 \xi_1) km \\
V_1 &= 4000 m/s \\
V_2(\xi_2) &= 6 \times (1 + \sigma^2 \xi_2) km/s \\
V_3(\xi_3) &= 4 \times (1 + \sigma^2 \xi_3) km/s \\
\rho_1 &= 2000 kg/m^3 \\
\rho_2 &= 2500 kg/m^3 \\
\rho_3 &= 2000 kg/m^3 \\
\sigma^2 &= 0.1
\end{aligned}
$$



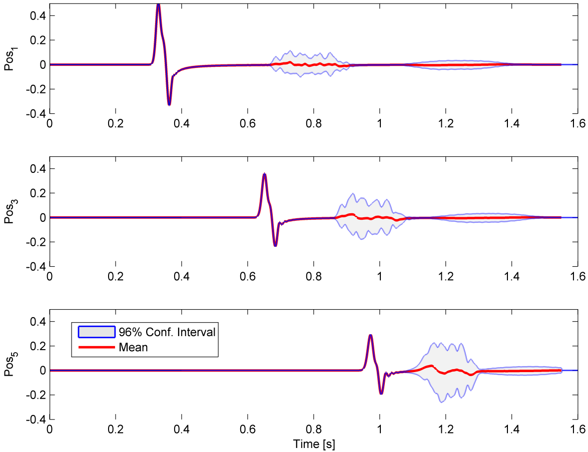
- where ξ_i's are uniform random variables in range $[-1, 1]$.

Surrogate setup

- The design set (surrogate training) are 729 points taken from a uniform cartesian grid in random space $d_i \in [-1, 1]^3$.
  - Most of the time spent in simulation is used to obtain the training data.
- Each input point corresponds to the one tuple of $\{\xi_1, \xi_2, \xi_3\}$.
- The output are the pressure at 3 sensors along 1811 time steps.
- We then sample the hyperparameters 1000 times. The Markov Chain is properly mixed after less than one hundred steps.
- With surrogate trained, we can take statistics using Monte Carlo. In this work, 10000 samples from initial space are used.
- The time required to sample the surrogate are 1 to 2 orders of magnitude smaller than forward problem!

Results

Synthetic seismogram - Pressure variance and mean values at recorded positions



Final Thoughts

- Multi-output Gaussian + Bayesian results in a promising venue for building efficient surrogates to be employed in Seismic analysis (forward and inverse)
  - The surrogate is much faster than using the forward solver.
  - The approach is totally non intrusive.
- It is just the beginning for us ... what are our next steps?
- Design an workflow for Reverse Time Migration (or even full wave inversion) taking into consideration uncertainties and employing surrogates
- Integrate this workflow in Chiron
- Adapt the Multi-output Gaussian to non stationary situations (time correlations)
- Solve the more challenging problems

Acknowledgments

We acknowledge the fruitful discussions about UQ with Prof. Nicholas Zabaras and Ilias Bilinois (University of Cornell) .

We are thankful for Prof. Marta Mattoso and her crew for continuous support with Chiron Environment.

Computer resources were provided by the High Performance Computer Center, NACAD/UFRJ.