Advances in Ensemble Kalman Inversion

Marco Iglesias[a]
[a] University of Nottingham

The Ensemble Kalman filter (EnKF) developed by Evensen and co-workers in the 1990s has had enormous impact on the geosciences and various engineering disciplines. EnKF has been historically used for data assimilation problems, where the objective is to infer the state of a partially observed dynamic system from observational data. More recently, algorithmic ideas from EnKF have been adapted to solve PDE-constrained inverse problems (i.e. to infer inputs from output of a PDE model). In this talk I will provide an algorithmic motivation for Ensemble Kalman Inversion (EKI) from both the Bayesian and the classical (deterministic) perspective. I will further discuss recent advances in EKI which are relevant to the stability, accuracy and applicability of the algorithm to numerous problems. Numerical experiments that show the efficacy (but also possible limitations) of EKI will be provided in the context of electrical resistivity tomography, non-destructive testing of composite materials and thermal characterisation of building walls.

Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection

M. Tso[a], T.C. Johnson[b], X. Song[b], X. Chen[b], O. Kuras[c], A. Binley[a]

[a]Lancaster University
[b]Pacific Northwest National Laboratory
[c]British Geological Survey

Time-lapse electrical resistivity tomography (ERT) measurements provide indirect observations ofhydrological processes in the Earth's shallow subsurface at high spatial and temporal resolution. ERT has been used for decades to detect leaks and monitor the evolution of associated contaminant plumes. However, this has rarely been applied to hazardous environmental sites in a strategic fashion. The ERT data are also rarely tied to groundwater models. Furthermore, an assessment of uncertainty in such applications has thus far been neglected, despite the clear need to provide site managers with appropriate quantitative information for decision making purposes. There is a need to establish a framework that allows leak detection with uncertainty assessment from geophysical observations. Ideally such a framework should allow the incorporation of additional data sources in order to reduce uncertainty in predictions.

To tackle these issues, we propose an ensemble-based data assimilation framework that evaluates proposed hydrological models (i.e. different hydrogeological units, different leak locations and loads) against observed time-lapse ERT measurements. Each proposed hydrological model is run through the parallel coupled hydrogeophysical code PFLOTRAN-E4D [Johnson et al., 2017] to obtain simulated ERT measurements. The ensemble of model proposals is then updated using the ES-MDA method [Emerick and Reynolds, 2013] iteratively. In this contribution, we demonstrate the proposed framework on synthetic data and field data collected from an ERT trial simulating a leak at the Sellafield nuclear facility in the UK [Kuras et al., 2016]. Our results show that it allows joint identification of contaminant source location, initial release time, and solute loading from the cross-borehole time-lapse ERT data, alongside with an assessment of uncertainties in these estimates. We demonstrate a reduction in site-wide uncertainty by comparing the prior and posterior plume mass discharges at a selected plane. This framework is particularly attractive to sites that have previously undergone extensive geological investigation (e.g. nuclear sites). It allows monitoring of unplanned releases of solutes because it effectively combines prior geological knowledge and ERT monitoring data to give improved estimates of leak parameters and their uncertainty bounds.

Particle Filtering for Inverse Problems

S. Dubinkina[a], S. Ruchi[a], M. Iglesias[b]

[a]CWI
[b]University of Nottingham

Particle filtering has been considered impractical for high-dimensional problems for a long time. Recent developments have shown, however, that the curse of dimensionality can be avoided. Localized ensemble transform particle filter [Reich, S. and Cotter, C., Probabilistic forecasting and Bayesian data assimilation, Cambridge University Press, 2015] is an example of such a development. The distinguished feature of particle filtering to represent any probability density function is, however, destroyed by localization. In my talk I will introduce an improved ensemble transform particle filter that does not use any localization and show that it provides excellent results in parameter estimation of high-dimensional problems of subsurface reservoir simulations, which are comparable to MCMC but are significantly less computationally demanding.

Efficient estimation of hydraulic conductivity fields with non-redundant information from head data: A comparison between linearization and sequential sampling methods

B. Carrera[a], C.M. Mok[b], I. Papaiannou[a]

[a]Technical University of Munich
[b]GSI Environmental

The hydraulic conductivity (K), a major parameter governing groundwater flow, varies significantly at all spatial scales. Since the spatial distribution of K cannot be measured directly in the field, inverse modeling based on hydraulic heads monitored during aquifer tests is often used to estimate the spatial K distribution and its uncertainty. The Successive Linear Estimator method (SLE, Yeh et al., An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. Water Resources Research, 1996) has been demonstrated to be a practical inverse technique for estimating the K distribution through employing a random field modeling. It is based on linearizing the relationship between each head observation and each random variable representing the K random field. Due to the linearization approximation, the accuracy in the estimated K field is higher when the K variance is smaller. In addition, the accuracy increases and the corresponding uncertainty decreases as more head data is incorporated. However, large amount of head data, such as the time-variable pressure transducer data automatically logged during aquifer tests, cannot be fully processed by SLE without sacrificing efficiency. As a result, in most cases, only a subset of head data is used, which might lead to loss of important measurement information. In this paper, an alternative approach is proposed to incorporate non-redundant information from a large amount of data, which is typically correlated. In the inversion process, the model predictive covariance matrix associated with all head measurements is formed and the Eigenmodes are obtained through spectral decomposition. Only a subset of the eigenmodes with the largest eigenvalues that adequately represents the model predictive variance are used to perform inversion. In addition, the Sequential Monte Carlo method (SMC, Koutsourelakis, A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters, Journal of Computational Physics, 2009) was adopted for inverse modeling. SMC is a robust sampling method without the need to approximate the groundwater problem by linearization. The performance of both SLE and SMC are compared in a two-dimensional numerical experiment and it is demonstrated that both approaches lead to consistent results in the low variance case.

Uncertainty quantification and characterization of an Italian sedimentary aquifer

E. Bianchi Janetti[a], L. Guadagnini[a], M. Riva[a], A. Guadagnini[a]

[a]Politecnico di Milano

We perform a global sensitivity analysis to assess the impact of the uncertainty associated with (a) the spatial distribution of hydraulic parameters and (b) the conceptual model adopted to describe the system on the characterization of a regional-scale aquifer. The study aquifer lies in Northern Italy and covers a planar extent of approximately 785 km2. We reconstruct the aquifer architecture and parameterize the associated conductivity field following diverse conceptual schemes: (1) Composite Medium model that considers the system composed by distinct lithological units, and (2) Overlapping Continuum approach for which the system consists of a collection of media coexisting in space. Groundwater flow is simulated with the numerical code MODFLOW-2005 for each of the adopted conceptual models. We then quantify the relative contribution of the selected uncertain parameters to the total variability of the piezometric level recorded in a set of 39 monitoring wells by relying on (1) the Morris screening method, (2) the variance-based Sobol indices and (3) the recently proposed AMA metrics. Sobol and AMA indices are derived numerically for the investigated settings through the use of a model-order reduction technique based on the polynomial chaos expansion approach. We find that the three sensitivity measures display consistent results for all the analyzed models. Depending on the conceptual model which has been used to characterize the lithological reconstruction, diverse parameters mainly affect the model outcomes. The proposed methodology can also provide a valuable tool in the context of inverse modeling of the groundwater flow system and parameter estimation.

Uncertainty Quantification of geochemical and mechanical compaction in layered sedimentary basins

I. Colombo[a], F. Nobile[b], G. Porta[c], A. Scotti[c], L. Tamellini[d]

[a]Geolog Srl
[b] Ecole Polytechnique Fédérale de Lausanne
[c] Politecnico di Milano
[d] Consiglio Nazionale delle Ricerche – Istituto di Matematica Applicata e Tecnologie Informatiche

In this work we propose an Uncertainty Quantification methodology for the evolution of sedimentary basins undergoing mechanical and geochemical compaction processes, which we model as a coupled, time-dependent, non-linear, monodimensional (depth-only) system of PDEs with uncertain parameters.
Specifically, we consider multi-layered basins, in which each layer is characterized by a different material. The multi-layered structure gives rise to discontinuities in the dependence of the state variables on the uncertain parameters. Because of these discontinuities, an appropriate treatment is needed for surrogate modeling techniques such as sparse grids to be effective.
To this end, we propose a two-steps methodology which relies on a change of coordinate system to align the discontinuities of the target function within the random parameter space. Once this alignement has been computed, a standard sparse grid approximation of the state variables can be performed. The effectiveness of this procedure is due to the fact that the physical locations of the interfaces among layers feature a smooth dependence on the random parameters and are therefore amenable to sparse grid polynomial approximations.
We showcase the capabilities of our numerical methodologies through some synthetic test cases.

Upscaling particle transport in porous media: PDF and stochastic methods

M. Icardi[a]

[a]University of Nottingham

In this talk, we deal with the upscaling from the pore- to the Darcy-scale of several transport mechanisms, relevant for many environmental and industrial applications. These include the well-known problem of mechanical dispersion, its non-Fickian extension and reaction/mass transfer mechanisms such as deposition, collisions leading to breakage and aggregation, and heterogeneous reactions.

We will show how to frame these problems using Probability Density Function approaches, with respect to ensemble and spatial averages. Dispersion is mainly affected by the velocity heterogeneities and we will therefore study how to characterise particle velocity PDF and its relation with pore-size distributions and flow velocity PDFs.

Reactions and surface processes are instead driven by what can be defined as random collision events within the averaging volume/ensemble, and therefore be represented from a probabilistic perspective.

Random and periodic pore-scale three-dimensional geometries are used to obtain macroscopic effective properties.

Moment Equations for Tracer Solute Transport in Composite Media with uncertain dispersivities

A. Dell'Oca[a], A. Guadagnini[a], M. Riva[a]

[a]Politecnico di Milano

Characterizing dissolved chemical migration in porous media through the Advection Dispersion Equation requires the knowledge of the fluid velocity field and of dispersivity values associated with diverse geomaterials which can make up the internal architecture of the system. Several studies have focused on the assessment of the impact on solute concentration dynamics of an incomplete knowledge of the fluid velocity field, the latter being typically due to uncertainty of hydraulic properties of the hosting media (e.g., permeability). Limited attention has been devoted to analyze the way uncertainty about spatial distribution of dispersivity values can propagate to uncertainty of solute concentration fields. Here, we address this issue by focusing on a simple one-dimensional domain filled with two distinct porous media and subject to a pulse injection of a tracer. We derive and solve numerically the equation governing the expected value and associated variance of solute concentration by considering uncertain dispersivity values and conceptualizing the domain as a random composite medium (where the location of the interface between the two materials can be uncertain). The solutions of such moment equations compare well against corresponding moments evaluated through a numerical Monte Carlo analysis. Our results suggest that in the investigated set-up (i) solute concentration variance exhibits a three peaks behavior, even in the presence of conditioning on a given location of the interface between the two materials and (iii) the actual sequence of the materials traveled by the solute impacts spatial distributions of expected value and variance of concentrations.

Richards' equation as an initial value problem: a Bayesian point of view

M. Berardi[a], E. Barca[a], F. Di Fonzo[b], M. Vurro[a]

[a]Consiglio Nazionale delle Ricerche – Istituto di Ricerca sulle Acque
[b]Code Architects Automation

The 1-dimensional Richards' equation for modeling water infiltration into unsaturated zone is generally handled by assigning suitable boundary conditions.
Because of the difficulty in measuring the state variables at the low boundary of the domain, here Richards' equation is solved as an initial value problem with respect to space: from a numerical point of view, this feature is faced by the transversal method of lines (TMoL), where, just pressure and flux are assigned as initial conditions at the top of the domain. In the present work, these initial values are perturbed according to a time-dependent family of distributions and the corresponding numerical solution is computed by TMoL . Afterwards, the resulting states at the bottom are assumed to be the classical boundary conditions for Richards' equation, and its solution is now computed by a standard method of lines (MoL).
An accurate insight is devoted to the comparison between these two different approaches; the dependence of bottom boundary condition (in the classical Richards framework) from the flux initial condition (in the TMoL context) is studied from a from Bayesian point of view.

Adaptive error models for efficient uncertainty quantification with application to subsurface fluid flow

C. Fox[a], T. Cui[a]

[a]University of Otago

Sample-based Bayesian inference provides a route to uncertainty quantification in the geosciences, though is very computationally demanding in the naive form that requires simulating an accurate computer model at each iteration. We present a new approach that adaptively builds a stochastic model for the error induced by a reduced model. This enables sampling from the correct target distribution at reduced computational cost, while avoiding appreciable loss of statistical efficiency. We build on recent simplified conditions for adaptive Markov chain Monte Carlo algorithms to give practical approximation schemes and algorithms with guaranteed convergence. We demonstrate the efficacy of our new approach by calibrating a large-scale numerical model of a real geothermal reservoir, showing good computational and statistical efficiencies for measured data sets.

Probabilistic assessment of environmental impacts of hydraulic fracturing scenarios

A. Guadagnini[a], A. Dell'Oca[b] M. Riva[a]

[a]Politecnico di Milano, University of Arizona
[b]Politecnico di Milano,

We illustrate some of the main results stemming from the application of a probabilistic approach for the quantification of the environmental risk associated with hydraulic fracturing operations. Our analyses are grounded on a set of focused scenarios which are developed within the context of the EU H2020 project FracRisk. The latter is keyed to (*a*) the quantification of the environmental impact of hydraulic fracturing technologies and (*b*) the development of decision support tools for the quantification of environmental risk. A significant segment of the study is devoted to the probabilistic modeling of risk scenarios during and following fracturing operations. The main aspects characterizing these scenarios include the pattern of the subsurface circulation of the fracturing fluid, the migration of methane, as well as the geomechanics of the development of fractures.

Here, we consider two well defined scenarios: (*a*) scenario S2 targets the possibility that the fracturing fluid escapes from the fractured schistose formation to migrate towards shallow aquifer bodies; (*b*) scenario S4 analyzes the migration path connecting the source of the fracturing fluid and a potential receptor through discontinuities such as, for example, a fault zone. Due to the computational burden associated with the numerical solution of the complete set of equations governing the dynamics of the scenarios considered, we resort to formulating the system behavior through a low complexity (or surrogate) model. The latter is then used to support global sensitivity analysis aiming at identifying the relative importance of (typically uncertain) model parameters on target model outputs which can be used to characterize the environmental signature of the fracturing operations. In addition to the classical Sobol' indices, we consider here the global sensitivity metrics suggested by Dell'Oca et al. (2017) to quantify the impact of model parameter uncertainties on statistical moments driving key features of the structure of the probability density function (*pdf*) of model outputs.

For the purpose of our showcase analysis, a generalized Polynomial Chaos Expansion (gPCE) is selected as a surrogate model, other choices being compatible with the approach. The quantification of the global sensitivity indices is performed in a Monte Carlo framework, upon relying on the surrogate model results.

The two focused scenarios are analyzed separately and jointly, to enable quantifying the environmental risk associated with the joint occurrence of various sources of uncertainty, i.e., the source of the contaminant (S2 scenario) and its migration (S4 scenario) to a overlying aquifers.

Our results allow identifying the relative importance of the uncertainty of the various model parameters/inputs on the first four (statistical) moments (i.e., mean, variance, skewness and kurtosis) of the *pdf* of of the environmental response of interest. It is suggested that the fracturing fluid flow rate imposed at the well, which is a controllable operating parameter, is key in determining the amount of dissolved methane potentially released by the shale formation under the conditions characterizing the focused scenarios analyzed.

Dell'Oca, A., Riva, M., & Guadagnini, A. (2017). Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrology and Earth System Sciences, 21*(12), 6219. https://doi.org/10.5194/hess-21-6219-2017

Thursday, September 6th, Morning Session

Certified dimension reduction of the input parameter space of vector-valued functions

O. Zahm[a], P. Constantine[b], C. Prieur[c], Y. Marzouk[d]

[a]INRIA
[b]University of Colorado Boulder
[c]Université Grenoble Alpes
[d]Massachussets Institute of Technology

Approximation of multivariate functions is a difficult task when the number of input parameters is large. Identifying the directions where the function does not significantly vary is a key preprocessing step to reduce the complexity of the approximation algorithms.
In this talk, we propose a methodology for dimension reduction which consists in minimizing an upper bound of the approximation error obtained using Poincaré-type inequalities. This approach is fundamentally gradient-based, and generalizes the so-called active subspace method for vector-valued functions, e.g. functions with multiple scalar-valued outputs or functions taking values in function spaces.
We also compare the proposed gradient-based approach with the popular and widely used truncated Karhunen-Loève decomposition (KL). We show that, from a theoretical perspective, the truncated KL can be interpreted as a method which minimizes a looser upper bound of the error compared to the one we derived. Also, numerical comparisons show that better dimension reduction can be obtained provided gradients of the function are available.

Multilevel Monte Carlo Acceleration of Seismic Wave Propagation under Uncertainty

M. Ballesio[a], J. Beck[a], A. Pandey[a], L. Parisi[a], E. von Schwerin[a], R. Tempone[a]

[a]King Abdullah University of Science and Technology

We consider forward seismic wave propagation in an inhomogeneous linear viscoelastic media with random wave speeds and densities, subject to deterministic boundary and initial conditions. We study this forward problem as a first step towards the treatment of inverse problems. There the goal is to determine, for example, earthquake source locations from seismograms recorded in a small number of seismic sensors at the Earth's surface. Existing results on earthquake source inversion for a given event show a large variability, which indicates that the inherent uncertainty of the Earth parameters should be taken into account. Here this uncertainty is modeled through random parameters.
We propose multilevel Monte Carlo simulations for computing statistics of quantities of interest which are motivated by the choice of loss function for the corresponding inverse problem, presenting a case study based on experimental seismic data from a passive experiment in Tanzania. This work provides a benchmark for the implementation of multilevel algorithms to accelerate seismic inversion addressing earthquake source estimation as well as inferring Earth structure.

Variational inference and Geodesic Monte Carlo for Polynomial Chaos adaptation

P. Tsifilis[a], R. Ghanem[b]

[a]Ecole Polytechnique Fédérale de Lausanne
[b]University of Southern California

A novel approach is presented for constructing Polynomial Chaos representations of scalar quantities of interest (QoI), that extends previously developed methods for adaptation in Homogeneous Chaos spaces. In this work we develop a Bayesian formulation of the problem, that characterizes the posterior distributions of the series coefficients and the adaptation rotation matrix acting on the Gaussian input variables. The adaptation matrix is thus construed as a new parameter of the map from input to QoI, estimated through Bayesian inference. For the computation of the coefficients' posterior distribution we use a Variational Inference (VI) approach that approximates the posterior with a member of the same exponential family as the prior, such that it minimizes a Kullback-Leibler criterion. On the other hand, the posterior distribution of the rotation matrix is explored by employing a Geodesic Monte Carlo sampling approach, consisting of a variation of the Hamiltonian Monte Carlo algorithm for embedded manifolds, in our case, the Stiefel manifold of orthogonal matrices. The performance of our method is demonstrated on the problem of multiphase flow in heterogeneous porous media.

Efficient reliability analysis for fuzzy cross-correlated random field material parameters of an hydro-mechanical coupled system

A. Schmidt[a], L. Nguyen-Tuan[a], C. Könde[a], T. Lahmer[a]

[a]Bauhaus-Universität Weimar

Realistic analyses require appropriate models of uncertain input data. Polymorphic uncertainty approaches combine basic uncertainty models, like a fuzzy random variable describing a material parameter.
Heterogeneous materials can be described by random fields (RF), which dependent on an auto-correlation function. Material interdependencies can be taken into account using cross-correlated RFs. For geo-science materials, these RF parameters (auto- and cross-correlations) are only vaguely known. Therefore, in the presented paper, they are described by convex fuzzy sets. An appropriate reliability analysis scheme is introduced comprising two main blocks: A surrogate model, approximating the failure probability of the system and an α-level optimization performed on the surrogate model.
A hydro-mechanical coupled system of a masonry gravity dam cross-section, implemented as a 2D plain strain finite element model, serves an application example. The resulting fuzzy sets obtained using the surrogate model are found to be in good agreement with re-calculated failure probabilities estimated with the finite element model itself. The computational costs of the reliability analysis are reduced by a factor of several tens by using the surrogate model. The consideration of material cross-correlation, even as a wide fuzzy description, implies significant differences in the resulting failure probabilities and should not be neglected in a reliability analysis of such multi-field systems. For large finite element models using a polymorphic uncertain material description, a possible domain decomposition approach is introduced.

Uncertainty Quantification for low-frequency Maxwell equations with applications in the controlled-source electromagnetic method

D. Kamilis[a], N. Polydorides[a]

[a] The University of Edinburgh

We discuss a forward and inverse UQ problem for low-frequency Maxwell equations, with uncertainty emanating from the conductivity material parameter. The application of UQ for such models is important in the controlled-source electromagnetic, geophysical exploration method. Emphasis is placed on the computationally efficient statistical characterization of quantities of interest using sparse quadrature and model reduction methods. We present our analytical and numerical results and discuss remaining challenges.

Isogeometric Hierarchical Model Reduction for Bayesian Inverse Problems in CFD

Y.A. Brandes[a], S. Perotto[a], A. Veneziani[b]

[a]Politecnico di Milano
[b]Emory University

In the field of hemodynamics, numerical models have evolved to account for the demands in speed and accuracy of modern diagnostic medicine. Methods have incorporated different reduction techniques to perform, with the same level of precision, the computation of the solution to partial differential equations in the constraints of time and computation power available in most diagnostics centers.

In this context, we studied Hierarchical Model Reduction (HiMod) techniques combined with Isogeometric Analysis (HigaMOD). HigaMod is a reduction procedure to downscale models when the phenomenon at hand exhibits a preferential direction of flow, e.g., when modelling the blood flow in arteries. The method showed a significant improvement in reducing the computational demand associated with multi-query simulations, such as the ones involved in the solution of inverse problems.

Our specific field of interest comprises problems related to the estimation of patient specific data, possibly with a significant lack of information. In more detail, we study the effectiveness of Kalman Filter-based techniques, powered by HigaMOD, to estimate patient-specific parameters in a number of different scenarios, such as lack of knowledge on the boundary conditions, source term and initial condition. In particular, we deal with the nonlinear estimation problem in the framework of steady and unsteady convection-diffusion-reaction problems and the Stokes equations. The goal of this investigation is to analyze the trade-off between accuracy and efficiency when HigaMOD reduction replaces the full model in the main algorithms used in Bayesian inference.

Uncertainty quantification analysis in discrete fracture network flow simulations

S. Berrone[a], C. Canuto[a], S. Pieraccini[a], S. Scialò[a], F. Vicini[a]

[a]Politecnico di Torino

In the framework of underground flow simulations in fractured media, fractures may act as preferential pathways, and may have a strong impact on the flow, in particular on its directionality. The discrete fracture network (DFN) model allows for an explicit representation of the interconnected fractures. Flux on each fracture is assumed to be driven by Darcy Law, and suitable matching conditions are imposed along fracture intersections, ensuring flux balance and head continuity.
Networks for simulations are typically generated sampling geometrical and hydro-geological properties from probabilistic distributions; this stochastic generation is likely to generate geometrical configurations very challenging for the meshing process, which is a major issue in the framework of DFN simulations. Furthermore, stochasticity in geometrical parameters my largely impact on connectivity properties of the network, thus resulting in a possible non-smooth behavior of the quantity of interest with respect to the stochastic parameters.
In order to quantify the influence of these stochastic parameters on the output of DFN models, we propose an approach based on the geometric Multi Level Monte Carlo method, applied in conjunction with a well assessed underlying solver for performing DFN flow simulations. Key properties of the solver are its capability of circumventing the need of conforming meshes, and its consequent extreme robustness with respect to geometrical complexities in the network. These features make the solver quite suitable to be used in conjunction with MLMC for the effective application of UQ strategies, as they allow to tackle complex geometrical configurations even with very coarse meshes.

Multifidelity methods for estimating the influence of overpressure on observed induced seismicity sequences

A. Quaglino[a], D. Karvounis[b], C. Tomasi[a], M. Favino[a], S. Wiemer[b], R. Krause[a]

[a]Università della Svizzera Italiana
[b]ETH Zurich

Enhanced Geothermal Systems (EGS) are a deep geothermal technology expected to assist countries both in decentralizing energy and in meeting climate goals. Crucial step for achieving commercially competitive power generation from an EGS power plant is the stimulation of its reservoir before operating it, which increases the productivity of its wells. Nevertheless, the risk associated with induced seismicity increases with the duration of the stimulation. The operators have to restrict to procedures of low hazard, but at the cost of decreasing the productivity of the wells. Future systems are expected to move past the passive monitoring of induced seismicity and towards probabilistic assessments both of seismicity (PSA) and of reservoir's performance (PRPA). To achieve this, a 3D discrete fracture hybrid model for HFR-Sim, the in-house EGS simulator (Karvounis and Jenny 2016) was developed. HFR-Sim was designed for modeling flow and heat transport inside dynamically-changing fracture networks (Karvounis and Wiemer 2017) Geo-mechanical phenomena are modeled using a stochastic mechanical model, where the probability of an induced seismic event increases with the overpressure due to the injection Karvounis and Wiemer. The model receives as input the stimulation strategy that needs to be assessed and it samples a probable fracture network based on the observed fracture density in the host rock. It simulates the evolution of pressure, locates the fractures for which their failure criterion is satisfied, and models the resulting flow. Eventually, the model computes not only the subset of the fractures that are expected to hydro-shear due to the stimulation strategy, but also the sequence of these events. Discrete fracture based hybrid models are computationally very expensive. In order to accelerate the computations, HFR-Sim employs a discrete representation of the subset of fractures whose surface size is larger than a threshold value. The rest of the domain is represented by a continuum with a low effective permeability wherever seismicity occurs and higher otherwise. Moreover the following numerical methods have been employed: (i) an adaptive step-size time integration, (ii) an efficient linear solver, (iii) an adaptive semi-implicit method for the continuity equation, where only a subset of the total degrees of freedoms is treated implicitly. Despite such improvements, performing a full Monte Carlo study still requires a large computational effort, preventing the assessment of a strategy within a desirable time frame. In order to overcome this limitation, we propose to take advantage of the availability of numerous models of varying complexity that perform the PSA (Kiraly-Proag et al. 2016) and model the induced seismicity in an EGS. We will show how such a model fusion program can be achieved within the framework of multifidelity Monte Carlo methods (Peherstorfer et al. 2016), where the correlation between models is exploited in order to accelerate the computation of the relevant statistics of the quantities-of-interest. In particular, we will remark how, even when very inaccurate models are employed, it is the correlation that allows to retain the accuracy of the full model.

Peherstorfer, B., Willcox, K. and Gunzburger, M. Optimal model management for multifidelity Monte Carlo estimation. SIAM Journal on Scientific Computing 38.5 (2016): A3163-A3194.

Kiraly-Proag, E., Zechar, J. D., Gischig, V., Wiemer, S., Karvounis, D., and Doetsch, J. Validating induced seismicity forecast models-Induced Seismicity Test Bench. Journal of Geophysical Research: Solid Earth, 121(8), (2016)
Karvounis, D. C. and Jenny, P. Adaptive Hierarchical Fracture Model for Enhanced Geothermal Systems. Multiscale Modeling and Simulation, 14(1), 207-231. (2016)
Karvounis, D. and Wiemer, S. Probabilistically Assessing the Efficacy of Stimulation Strategies. In 42nd Workshop on Geothermal Reservoir Engineering, Stanford University, (2017).

Thursday, September 6th, Afternoon Session

Uncertainty Quantification of fracture characterization in Naturally Fractured Reservoir

E.R. Russo[a], I. Nicolotti[a], I. Colombo[a]

[a]Geolog Srl

Scope of this work is to show an application of an uncertainty quantification procedure applied to a reservoir scale problem to identify and characterize formation fractures occurring during drilling operation.
The characterization of formation fractures, in terms of fracture width estimation, is performed by means of analytical and numerical models which take as input a well-defined set of parameters as drilling fluid rheological properties and flow rates, pore and dynamic drilling fluid pressure, wellbore geometry and drilling parameters. Mentioned parameters are generally affected by different sources of uncertainty associated with the measurement instruments accuracy, the high variability of investigated phenomena, the difficulties to assess the sample representability and to follow a standard operational procedure at rig site. In such a context, adopting a methodology to quantify the uncertainty in model responses, i.e. fracture width, becomes fundamental to get reliable high-quality results.
In this work, both analytical and numerical models, suitable for estimating the fracture width, are considered and numerical coefficients appearing therein are taken as uncertain parameters. The uncertainty propagation from input parameters to numerical model outputs is investigated and quantified through a straightforward procedure implemented into a Monte Carlo framework. This procedure is applied to few real cases and leads to probability density function of model output.
Results emphasize the relevance of employed methodology in tackling uncertainty in formation fractures characterization improving Fractured Reservoir Characterization and, at the same time, allowing to point out the wide range of applicability of uncertainty quantification techniques.

Hydrogeological model selection with complex spatial priors

C. Brunetti[a], M. Bianchi[b], M. Pirot[a], N. Linde[a]

[a]University of Lausanne
[b]British Geological Survey

Hydrogeological studies often rely on a single conceptual representation of the subsurface. However, the impact of a poorly chosen conceptual model might be significantly larger than the impact of parameter uncertainty within a given conceptual model. Also, it is well recognized that conceptual models need to include geologically realistic features in order to provide meaningful uncertainty quantification. Therefore, when several conceptual models are defined, it is useful to know how to rank, select, or falsify some of them. Here, we propose a methodology for Bayesian model selection among conceptual models built from complex spatial priors using training images and concepts from multiple-point statistics (MPS). Our aim is to rank conceptual models describing aquifer heterogeneity and, as a field case, we consider a small-scale tracer test at the heterogeneous Macrodispersion Experiment (MADE) site in Mississippi, USA. We compare conceptual models built from five different training images: (1) multi-Gaussian, (2) multi-Gaussian with connected channels, (3) outcrop-based, (4) analog-based from the Herten site in Germany and (5) lithofacies-based. The evidence and subsequent Bayes factors used for model selection are derived using the thermodynamic integration method with an underlying Markov chain Monte Carlo (MCMC) inversion based on a sequential Gibbs algorithm. Compared with most MCMC-based methods for Bayesian model selection, thermodynamic integration has the advantage to be compatible with MPS-based inversion. The flow and transport simulations are presently 2-D, but future work will account for modeling errors caused by the 2-D approximation.

A Case Study of Seismic Wave Propagation with Random Parameters

A. Pandey[a], M. Ballesio [a], J. Beck [a], L. Parisi [a], E. von Schwerin [a], R. Tempone [a]

[a]King Abdullah University of Science and Technology

We will present results from a case study based on an earthquake with seismograms recorded on a small dense seismic network in the Ngorongoro Conservation Area in Tanzania. We consider forward seismic wave propagation in an inhomogeneous linear viscoelastic media with random wave speeds and densities, subject to deterministic boundary and initial conditions. The random parameters model the inherent uncertainty of the Earth parameters. We use multilevel Monte Carlo simulations for computing statistics of quantities of interest chosen to formulate a suitable loss function for the corresponding source inversion problem. We use recorded seismograms to study a noise model for use in Bayesian inverse problems. This work provides a benchmark for the implementation of Multilevel algorithms to accelerate Seismic Inversion addressing earthquake source estimation as well as inferring Earth structure.

Thursday, September 6th, Poster Session

A Multi-Level Markov Chain Monte Carlo Sampler With Applications to a Seismic Source Inversion Problem

J.P. Madrigal Cianci[a], F. Nobile[a], S. Krumscheid[a], R. Tempone[b]

[a]Ecole Polytechnique Fédérale de Lausanne
[b]King Abdullah University of Science and Technology

We focus on implementing a sampling strategy using Bayesian inversion, in particular Markov Chain Monte Carlo (MCMC) methods, to recover the probability distribution of the spatial location of the epicenter of an earthquake in a bounded domain. To this end, we first postulate a computational elasticity model (usually arising from the discretization of a partial differential equation) that describes the earthquake dynamics. In particular, this so called forward problem is described by the elastodynamic wave equation. We can then use this forward model together with MCMC methods in order to obtain samples from the posterior probability distribution of the epicenter of the earthquake. This represents a challenge from a computational perspective, as it implies that the computational model for the elastodynamic wave equation needs to be run a large number of times. Based on the ideas presented in [1,2] we present a Multi-Level Markov Chain Monte Carlo (MLMCMC) algorithm to alleviate the computational burden associated with the problem at hand. A brief description of the algorithm is as follows. we introduce a hierarchy of levels $\ell = 0,\ldots,L$ for which there corresponds a discretization parameter $\{h_\ell\}_{\ell=0}^{L}$ of the underlying PDE, with an associated posterior distribution $\{\pi_\ell\}_{\ell=0}^{L}$. At each level we use MCMC to simultaneously generate two chains $\chi_{\ell,\ell}$ and $\chi_{\ell,\ell-1}$, of length N_ℓ, targeting π_ℓ and $\pi_{\ell-1}$ in such a way that we can compute multi-level estimators as the ones used in [2]. The crux of the algorithm lies on the construction of the proposal distribution that is being used to generate the chains $\chi_{\ell,\ell}$ and $\chi_{\ell,\ell-1}$. In particular, we propose to use an independent sampler strategy where the candidate for $\chi_{\ell,\ell}$ and $\chi_{\ell,\ell-1}$ is drawn from an estimated density $\tilde{\pi}_{\ell-1}$, e.g, by kernel density estimation, built from the collected samples $\chi_{\ell,\ell-1}$ at the previous level. Other estimation procedures will also be discussed. Experimental results suggest that the benefits provided by this algorithm are two-fold. In particular, the majority of the computations are performed at a coarse discretization level, where each run of the solver is much cheaper to compute. Additionally, experimental results have also shown that the samples generated at higher levels are (almost) independent and identically distributed.

[1] TJ Dodwell, C Ketelsen, R Scheichl, and AL Teckentrup. A Hierarchical Multilevel markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. arXiv preprint arXiv:1303.7343, 2013.
[2] Michael B Giles. Multilevel Monte Carlo Path Simulation. Operations Research, 56(3):607–617, 2008.

Accounting for model error using a local basis approach

C. Koepke[a], J. Irving[a], A. El-Sheikh[b]

[a]Univeristy of Lausanne
[b]Heriot-Watt University

Inverse problems in the fields of geophysics and hydrology typically involve high-dimensional spatial distributions of subsurface model parameters and computationally expensive forward solvers. As a result, uncertainty quantification is extremely difficult, and often relies upon fast forward approximations based on simplified physics, coarsened discretizations, or response-surface proxies in order to make the stochastic inverse solution, obtained for example via Bayesian-MCMC methods, computationally tractable. In most cases, however, the errors resulting from these computationally efficient forward approximations are not taken into account into the inversion procedure. This can lead to strongly biased posterior parameter estimates and erroneous predictions.
Here, we present ongoing work on a methodology for dealing with model errors in Bayesian solutions to hydrogeophysical inverse problems that is geared towards the common case where these errors cannot be effectively characterized either (i) globally through some parametric statistical distribution, or (ii) locally based on interpolation between a small number of computed realizations. Rather than focusing on the construction of a global or local error model, we instead work towards identification of the model-error component of the residual using a learning-based approach. Pairs of approximate and detailed model runs are stored in a dictionary that grows at a specified rate during the inversion procedure. At each iteration, a local model-error basis is constructed for the current test set of model parameters using the K-nearest neighbour entries in the dictionary, which is then used to separate the model error from the other error sources before computing the likelihood of the proposed set of model parameters. We show the application of the developed technique to spatially distributed inverse problems in geophysics of varying degrees of complexity. In each case, accounting for the model-error enables us to reduce posterior parameter bias and obtain a more realistic characterization of uncertainty.

Assessment of parametric uncertainty and calibration of basin-scale depositional models

S. Patani[a], G. Porta[a], E. Bianchi Janetti[a], F. H. Maina[a], A. Guadagnini[b]

[a]Politecnico di Milano
[b]Politecnico di Milano, University of Arizona

The study is geared towards the development of a strategy to manage uncertainty arising from the variety of parameters embedded in three-dimensional multi-lithology stratigraphic models which allow modelling the development and formation of large scale stratigraphic sequences. These models are associated with a high computational cost due to (a) the richness and complexity of processes included and (b) the large number of input parameters which are typically affected by uncertainty.

Here, we design and implement a procedure which enables us to perform (a) a preliminary screening leading to the delineation of the most influential model parameters, (b) a global sensitivity analysis yielding the relative contribution of uncertain model parameters to key (statistical) moments of model output and (c) model calibration.

Quantities of interest forming the output of a stratigraphic modeling application include thickness of depositional sequences and spatial distributions of sediments, the latter being categorized in terms of a set of lithological fractions (e.g. siliclastic and carbonate sediments). Our approach starts by focusing on the selection of the most influential parameters, identified within the space of the model parameters. We do so by relying on the joint use of Morris indices and Principal Component Analysis (PCA) and by considering model parameters as independent and identically distributed (iid) random variables, uniformly distributed across a given support, whose width is assessed on the basis of literature information and modelers' experience.

The resulting reduced parameter set is then used for the construction of a surrogate (or reduced complexity) model mimicking basin scale deposition and grounded on the Polynomial Chaos Expansion (PCE) technique. This step enables us to (a) evaluating model output quantities of interest at an affordable computational cost and (b) obtain directly the Sobol Indices, which provide a quantification of the relative contribution of each model parameter to the variance of target outputs.

The surrogate model is the basis for model calibration, which is conducted by way of the Particle Swarm Optimization (PSO) technique. This yields the full (posterior) probability density function of the selected model parameters, thus allowing a complete characterization of the way uncertainty is propagated from unknown model parameters to target outputs through conditioning on available observations.

The procedure outlined above is applied to a synthetic test bed, considering input parameter effects on variables of the system such as local sediment compositions (in terms of fractions of sand, shale, and carbonates) and thickness of the basin. Our study allows assessing the potential of this procedure to (1) identify important model parameters, (2) analyze spatially distributed sensitivities to a variety of uncertain parameters in a complex sedimentological environment and (3) identify parameter combinations compatible with a given set of observations typically available in basin-scale exploration campaigns (i.e. vertical distributions of volumetric fractions and sediment thickness in deep exploration boreholes). Current developments include application of the procedure in a full field scenario.

Uncertainty quantification of fault stability: a simple Monte-Carlo model

D. Rossi[a], L. Valdacca[a]

[a]Politecnico di Milano

The need to assess the stability of preexisting faults and their response to changes in the stresses and other physical parameters is of paramount importance for public security. Therefore, there is a strong need for mathematical tools, such as direct numerical simulations and data analysis, to study this problem. Since subsurface data are usually affected by uncertainty this problem is a natural subject for the uncertainty quantification analysis. Fault stability depends on the ratio between tangential and normal stress, which must be smaller than the friction coefficient to avoid fault slip. The aim of this work is to study the dependence of the tangential and normal components of the stress tensor on uncertainties in the reference angle (depending on fault strike and dip) and pore pressures for some idealized cases. We consider different levels of anisotropy in the original stress field, using a Monte-Carlo method. We show the results for a homogenous case and then progress to include effects of uncertainty in the fault position as a perturbation of a predefined fault geometry.

Thursday, September 6th, Poster Session

The Diffusion Equation with Random Diffusion Coefficient given by (Transformed) Lévy Fields

M. Reese[a], H. Gottschalk[a], O. Ernst[b], T. Kowalewitz[b]

[a]University of Wuppertal
[b]Chemnitz University of Technology

We consider diffusion equations which model several problems e.g. the groundwater flow in heterogeneous media. As only partial information on the conductivity is available, such equations are often modeled with a random diffusion coefficient. This random diffusion coefficient is mostly considered to be a log-Gaussian field, i.e. the exponential transformation of a Gaussian field.
This poster presents diffusion equations with transformed smoothed Lévy noise fields as random diffusion coefficients as the natural generalization of Gaussian fields. We investigate the unique weak solution to this elliptic PDE wrt its integrability and approximability by estimating the extreme values of these fields. With this approach, we can show that solutions possess moments of all order of the Sobolev norm.
We approximate random diffusion coefficient by applying Mercer's theoremon the smoothing kernel of the smoothed Lévy noise field. This approximation provides us the convergence of the appropriate weak solution in the n-th mean to the real weak solution.

Nonlinear parametrization of geological models using generative deep-learning techniques

A. H. El-Sheikh[a], S. Chan[a]
[a]Heriot-Watt University

Generative models have the ability to reconstruct high-dimensional data using a low-dimensional latent space. The compressed latent space is not directly observed but is rather inferred from the data after defining the mathematical structure of the generative model. In this talk, I will we will focus on generative models formulated as a deep neural network. The two most common neural network based generative models are Variational Auto-Encoders (VAE) and Generative Adversarial Networks (GAN). I will review the advantages and disadvantages of both techniques and then present some recent results of using GANs as a general sample-based parameterization tool for stochastic inputs to partial differential equations with an emphasis on geological modeling. I will present results for unconditional and conditional realizations of binary channelized geological models, and perform experiments in uncertainty quantification and parameter estimation for subsurface flow models. Our results show that GAN based parametrization is very effective in preserving the visual realism as well as high order statistics of the flow responses [1]. Additionally, I will introduce a novel post-hoc conditioning technique of pre-trained generative models to produce conditional geological realizations [2].

References:
[1] Shing Chan, Ahmed H. Elsheikh, Parametrization and Generation of Geological Models with Generative Adversarial Networks https://arxiv.org/abs/1708.01810
[2] Shing Chan, Ahmed H. Elsheikh, Parametric generation of conditional geological realizations using generative neural networks, https://arxiv.org/abs/1807.05207

The Generalized Sub-Gaussian Model: Theory and Applications

M. Riva[a]
[a]Politecnico di Milano, University of Arizona

Geostatistical analysis has been introduced over half a century ago to allow quantifying seemingly random spatial variations in earth quantities such as rock mineral content or permeability. The traditional approach has been to view such quantities, $Y(\mathbf{x})$, as multivariate Gaussian random functions characterized by one or a few well-defined spatial correlation scales. There is, however, mounting evidence that many spatially varying quantities exhibit non-Gaussian behavior over a multiplicity of scales. A statistical model that captures these disparate, scale-dependent distributions of Y and their (spatial or temporal) increments, ΔY, in a unified and consistent manner is the generalized sub-Gaussian (GSG)" model. GSG has the form $Y(\mathbf{x}) = U(\mathbf{x}) G(\mathbf{x})$ where $G(\mathbf{x})$ is (generally, but not necessarily) a multi-scale Gaussian random field and $U(\mathbf{x})$ is a non-negative subordinator independent of G. In our model the peak and tails of the ΔY pdf scale with lag in line with observed behavior. The model allows one to estimate, accurately and efficiently, all relevant parameters by analyzing jointly sample moments of Y and ΔY.

Robust information divergences for model-form uncertainty arising from sparse data in random PDE

E. Hall[a], M. Katsoulakis[a]
[a]University of Massachusetts Amherst

Random partial differential equations of elliptic type arise in the modeling of steady-state subsurface flow in groundwater hydrology, carbon sequestration, and petroleum engineering. One goal is to predict quantities of interest, that depend on model outputs, to inform decision support tasks such as regulatory assessment and risk management. A key challenge in this context is the analysis of model-form or epistemic uncertainty arising from sparse data. In this talk we present a novel application of a hybrid information divergence, based on the Donsker-Varadhan variational formula, to make robust data-informed predictions of quantities of interest. In particular, we highlight two uncertainty quantification tasks: sensitivity analysis and estimating misspecification errors due to using computationally advantageous surrogates in place of fully resolved models.

Model tests for flow predictions in ungauged basins

C. Prieto[a], N. Le Vine[b], D. Kavetsky[c], E. Garcia[d], C. Alvarez[d], R. Medina[d]

[a]University of Cantabria, Imperial College London, Bristol University
[b]Imperial College London, Swiss Re
[c]The University of Adelaide
[d]University of Cantabria

A catchment is a complex system where a multitude of interrelated energy, water and vegetation processes occur at different temporal and spatial scales. A rainfall-runoff model is a simplified representation of the system, and serves as a hypothesis about catchment inner workings. When observations of flow data are not available, a regionalization model is used to extrapolate knowledge from data-rich regions to ungauged regions. A common practice when making predictions in ungauged basins is to use a pre-selected hydrological model structure and regionalization procedure, usually with little justification of their adequacy (due to the lack of observed flows). This work proposes two statistical metrics for testing hydrological and regionalization model applications in ungauged basins: a) DistanceTest assesses how well do given hydrological model and regionalized model reproduce the available information; and b) InfoTest quantifies the information gain offered by the hydrological and regionalization models over what is already known a priori. The proposed techniques are applied to 92 basins in northern Spain. The results show that the proposed adequacy tests allow to identify the main source of predictive streamflow uncertainty, and suggest avenues for reducing it. Specifically, results suggest that the contribution of the hydrological model – including its structure and inputs – to streamflow predictive uncertainty exceeds the contribution of the regionalization model. Based on these findings, priority should be given to the improvement of the hydrological model, followed by improvement of the regionalization model, not the other way around.