Emerging Applications of Data Assimilation in the Geosciences

13-17 March 2017 @ Oort

Emerging Applications of Data Assimilation in the Geosciences Goal

The goal of this workshop was to bring together experts on data assimilation, Bayesian statistics, seismology, and (paleo-)climatology to exchange knowledge, suggest new data-assimilation methods suitable for applications in each field, and to stimulate future collaborations. Data assimilation is a well-established methodology in meteorology to estimate the complex behaviour of a dynamical system using available data. Seismology and climatology share the same goal as meteorology though the mathematical methods differ depending on each field due to limited data, models, and computing power. With improvement of models and wider availability of data the traditional distinction between these fields is, however, blurring, and considerable convergence can be observed. Moreover, the current increase of computing power allows to quantify the uncertainty of an estimate using Bayesian statistics, which serves as a unifying framework for data assimilation in all fields.

Focus Groups

The presentations and discussions during the week where organized around four overarching themes: *Formulating the Prior, Model-Data Mismatch, Uncertainty Quantification, and Joint Parameter and State Estimation* . Below are brief summaries of the outcomes of each focus group.

Formulating the Prior

The prior probability density function encodes the prior assumptions on the quantity we are estimating. It plays different roles in (paleo-)climatology and seismology. In seismology, due to the iterative nature of the data assimilation algorithms, the prior refers to the initial prior information about the parameters. In climatology, data assimilation algorithms are sequential and the prior mainly refers to the updated posterior. Moreover, due to the ill-posed nature of the inverse problem in seismology, so-called regularization terms are added to force a unique solution. The choice of regularization is often dictated by the availability of efficient data assimilation algorithms rather than realistic assumptions on the parameters. In climatology, the chaotic nature of the system results in a rather quick memory loss of the initial prior and thus perturbation of the updated posterior using for example simple inflation is more essential, which could be replaced by a recent employment of *non-informativ* e priors in data assimilation. For some applications, such as paleoclimate, the Lyapunov time could be rather long and thus sophisticated perturbations, such as breeding, are required. Another avenue of investigation that came up is the use of techniques from machine-learning to distill priors in data-rich problems or based on more sophisticated (fine resolution) models.

Model-Data Mismatch

Quantifying the mismatch between the predictions and observations is relevant in both (paleo-)climatology and seismology. In seismology, the primary goal is to identify features that lead to a simple relation between the parameters of interest and some observations. An example is the travel-time of a certain wave from their source to the location where observations are obtained. In (paleo-)climatology, the first concern is to incorporate any biases introduced while acquiring or processing the observations. During the discussion some common problems were identified. How to identify features in the observations that are optimally sensitive to parameters of interest is an open question. Transparency and reproducibility when processing the data was also identified as an important issue.

Uncertainty Quantification

Quantifying the uncertainty (UQ) in the (best) estimate is of crucial importance when interpreting the results. The Bayesian point-of-view provides a natural framework to think about uncertainty. We can identify several sources of uncertainty, such as measurement errors, modelling errors and the inherent non-uniqueness of the inverse problem. The relative importance of these various contributing factors differs per application, leading to distinct approaches for UQ in each field. In (paleo-)climatology, measurement and modelling errors are most prominent. Using a purely Bayesian framework, estimates of the uncertainty follow naturally. In seismology, the inherent non-uniqueness is the biggest problem and uncertainty is usually formulated in terms of the resolution of the final image. Porting the Bayesian interpretation of uncertainty to seismology is challenging because of the bias introduced by the problem formulation (prior and likelihood). Quantifying the influence of the choice of prior/likelihood is identified as a highly relevant open question.

Joint Parameter and State Estimation

Estimation of the parameters and states jointly can be formulated in a common framework. Starting from this generic formulation, various existing approaches can be derived and compared. In both seismology and (paleo-)climatology there is interested in extended current methodologies in this direction. Several participants have expressed interest in drafting a review paper on this subject.

Outcomes

- Planned review paper on joint parameter and state estimation.
- Several investigative collaborations.
- Emerging interest in seismology of sharing and combining fine regional models in order to produce a 'Collaborative Seismic Earth Model' could profit from the climatological community where such collaborations are well established.
- Multi-model approach (similar to an Ensemble Kalman Filter) used in climatology could be applied in seismology as well in order to obtain successive model updates, including uncertainties.
- Improvement in the comparison of model estimations and observations in paleoclimatology is inspired by the sophisticated methods applied in seismology to isolate the most useful signal.
- Automatic quality control of the input data in paleoclimatology could be based on the expertise developed in other fields.

Svetlana Dubinkina Amsterdam, The Netherlands
Andreas Fichtner Zürich, Switzerland
Hugues Goosse Leuven, Belgium
Peter Jan van Leeuwen Reading, UK
Tristan van Leeuwen Utrecht, The Netherlands