Mathematisches Forschungsinstitut Oberwolfach

Report No. 47/2016

DOI: 10.4171/OWR/2016/47

Mathematical and Algorithmic Aspects of Data Assimilation in the Geosciences

Organised by
Andreas Griewank, Berlin
Sebastian Reich, Potsdam
Ian Roulstone, Guildford
Andrew Stuart, Coventry

2 October – 8 October 2016

Abstract. The field of "Data Assimilation " has been driven by applications from the geosciences where complex mathematical models are interfaced with observational data in order to improve model forecasts. Mathematically, data assimilation is closely related to filtering and smoothing on the one hand and inverse problems and statistical inference on the other. Key challenges of data assimilation arise from the high-dimensionality of the underlying models, combined with systematic spatio-temporal model errors, pure model uncertainty quantification and relatively sparse observation networks. Advances in the field of data assimilation will require combination of a broad range of mathematical techniques from differential equations, statistics, machine learning, probability, scientific computing and mathematical modeling, together with insights from practitioners in the field. The workshop brought together a collection of scientists representing this broad spectrum of research strands.

Mathematics Subject Classification (2010): 37-XX, 60-XX.

Introduction by the Organisers

The workshop *Mathematical and Algorithmic Aspects of Data Assimilation in the Geosciences*, organized by Andreas Griewank (Berlin), Sebastian Reich (Potsdam), Ian Roulstone (Surrey), and Andrew Stuart (Warwick) was held 2 October – 8 October 2016. The meeting was attended by nearly 50 participants representing a broad range of mathematical subject areas as well as applications of data assimilation in the geosciences.

A total of 29 talks were presented during the workshop. The talks were selected such as to cover novel mathematical developments on, e.g., particle filters (Dan Crisan, Peter Jan van Leeuwen, Nikolas Kantas), hybrid filter algorithms (Hans-Rudolf Künsch, Roland Potthast), the analysis of sequential filter algorithms in high-dimensions/small sample sizes (Chris Snyder, Daniel Sanz Alonso, Matthias Morzfeld), theoretical and practical aspects of the ensemble Kalman filter (Lars Nerger, Tijana Janjic, Xin Tong, Marc Bocquet, Claudia Schillings) parameter estimation and model comparison (Dean Oliver, Alberto Carrassi, Manfred Opper) multi-level Monte Carlo filtering (Kody Law, Colin Cotter), statistical inference (Youssef Marzouk, Illia Horrenko, John Harlim), variational techniques (Olivier Talagrand, Jochen Bröcker, Henry Abarbanel, Manfred Opper, Nancy Nichols), multi-scale dynamics (Rupert Klein), and Lagrangian data assimilation (John MacLean, Amit Apte) on the one hand and practical advances and challenges arising from the geosciences (Chris Snyder, Roland Potthast, Olivier Talagrand, Nancy Nichols, Lars Nerger, Serge Gratton, Peter Jan van Leeuwen, Hans-Rudolf Künsch) on the other. A poster session was held on Tuesday evening which gave the attending PhD students and postdocs the opportunity to present and discuss their work. The first prize for the best poster went to Yvonne Ruckstuhl.

Throughout the workshop a number of spontaneous discussion groups arose triggered by the many different facets of data assimilation presented during the talks. The following discussion groups in the central lecture hall of the MFO on Monday and Wednesday evening, respectively, shall be mentioned in particular: (i) typical sets and their relevance to particle filters in high dimensions (inspired by the talk by Peter Jan van Leeuwen) and (ii) Localization, adaptivity and cross-validation (inspired by the talk of Hans Rudolf Künsch). These discussions reflect the actuality and broad scientific appeal of data assimilation.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, "US Junior Oberwolfach Fellows".

Mathematical and Algorithmic Aspects of Data Assimilation in the Geosciences

Table of Contents

Abstracts

Particle filters in high dimensions

Chris Snyder

Simple examples that are independent and identically distributed in each degree of freedom illustrate basic difficulties in particle filters for high-dimensional systems. These examples, together with asymptotic results valid when the variance of the unnormalized weights is large and the log likelihood function has an approximately Gaussian distribution, demonstrate the need for sample sizes that grow exponentially with that variance (which is proportional to the system dimension in the simple examples). An approximated, spectral model of turbulence, following Lorenz (1969), shows that the large-variance asymptotics are relevant as the number of observations increases, except when the turbulent flow has a kinetic energy spectrum with a k^{-3} power law.

Adapting the Ensemble Kalman Particle Filter to Large Scale Data Assimilation

Hans Rudolf Künsch
(joint work with Sylvain Robert)

The Ensemble Kalman filter (EnKF) is used in operational data assimilation schemes due to its robustness in high dimensions. It is however valid only for Gaussian background distributions. Particle filters (PF) can deal with non-Gaussian background distributions, but become degenerate in high dimensions. The Ensemble Kalman Particle Filter (EnKPF) of Frei and Künsch ([1]) provides a continuous interpolation between these two methods through a parameter $\gamma \in [0, 1]$ and thus has the potential to combine their strengths. In this talk we discuss the modifications necessary to apply the EnKPF to a near operational regional weather prediction system over Europe. This involves finding a transform version of the filter in ensemble space, reducing discontinuities that stem from a localisation of the filter and choosing the parameter γ . We also show preliminary results from a numerical experiment with cycled assimilation over several days in a period of intense convective activity.

References

[1] M. Frei, H. R. Künsch, *Bridging the ensemble Kalman and particle filters* Biometrika **100** (2013), 781–800.

On Ensemble and Particle Filters for Large-Scale Data Assimilation and Inverse Problems

Roland Potthast

In almost all operational centres for numerical weather prediction around the world ensemble data assimilation techniques are of rapidly growing importance. Ensemble techniques allow to describe and forecast uncertainty of the analysis, but they also improve the assimilation result itself, by allowing estimates of the covariance or, more general, the prior and posterior probability distribution of atmospheric states.

In our talk, we will first give a survey about recent activities of the German Meteorological Service DWD, who is using an Ensemble Kalman Filter both for its new global ICON model as well as for the convective scale high-resolution model COSMO-DE. To be more precise, for the global model a hybrid variational ensemble Kalman filter EnVAR has been developed. We survey the setup of its Ensemble Kalman Filter component, which is based on the LETKF of Hunt, with a range of further features such as relaxation to prior perturbations or random perturbations. The system is run operational since January 2016 and shows scores comparable to state-of-the-art 4D-VAR systems as run by many international centers today. Ensemble data assimilation also provides initial states for ensemble prediction (EPS). We describe ICON EPS and demostrate the high quality of the system.

In the second part of the talk, we present recent work on the further development of the ensemble data assimilation towards a particle filter for large-scale atmospheric systems, which keeps the advantages of the LETKF, but overcomes some of its limitations. We describe a Localized Markov Chain Particle Filter (LMCPF), present its mathematical foundation. A localized particle filter has been implemented for the global ICON model of DWD. We show results of a case study of one week global assimilation for a hybrid particle filter variational assimilation in a quasi-operational setup, showing the huge potential of the method which already in its first simple implementation can achieve better or comparable scores to the operational EnVAR system.

The seamless multilevel ensemble transform filter

Colin Cotter
(joint work with Alistair Gregory, Sebastian Reich)

We report on recent work developing a multilevel ensemble transform filter. This work combines two ideas together, the multilevel Monte Carlo method [1] which has now been applied in many different application areas, and the idea of using optimal transportation to transform between ensembles, which is the foundation of the ensemble transform particle filter (ETPF) described in [2], [3]. The basic idea behind the multilevel Monte Carlo method is to reduce the cost of a Monte

Carlo estimator based on some underlying discretisation method by blending together results from a hierarchy of discretisation resolutions in time and/or space. This is done by using a telescoping sum formula, replacing the high resolution Monte Carlo estimate by a low resolution estimate, combined with a series of difference estimators that estimate the difference in the expected value between two different levels of the discretisation hierarchy. If the number of samples at each level is chosen correctly, there is the possibility to reduce the cost of estimating a statistic with mean-square-error bound of ϵ^2 from $\mathcal{O}(\epsilon^{-3})$ for regular Monte Carlo down to $\mathcal{O}(\epsilon^{-2})$. As described in [1], this reduction is possible, provided that the variance of the difference estimators decays sufficiently quickly with resolution. This requirement depends on $\beta > \gamma$, where the variance of the difference estimators scales with model resolution as $\mathcal{O}(h^{\beta})$ and the cost of the model scales as $\mathcal{O}(h^{-\gamma})$, where h is the model resolution parameter (e.g. the timestep or spatial cell size). This provides a good framework for proving optimal multilevel Monte Carlo algorithms in different application areas. Since ensemble data assimilation algorithms are a form of Monte Carlo method, it is natural to try to build multilevel ensemble data assimilation algorithms, and some other groups have also been working on this [4], [5]. For particle filters, the idea is simple: maintain a large ensemble at the coarsest resolution, together with correlated pairs of ensembles at two consecutive resolutions that form the difference estimators. The correlation is maintained by using the same noise realisations (appropriately coarse-grained) for each pair of samples. The telescoping sum estimator is then used to calculate sample statistics from the filter. The main challenge is finding a way to avoid losing correlation after the filter resampling stage. This leads to the development of coupled resamplers. Since the ETPF already makes use of linear programming to apply the Bayesian ensemble transform to in- corporate new observed data, we thought it would be interesting to experiment with using linear programming to maximise correlation between the pairs of ensembles in the difference estimators. In the algorithm described in [6], we used the ETPF procedure to independently update the two ensembles, and then solved an assignment problem to repair the ensemble members such that the covariance is maximised. For a one dimensional state space, this assignment problem can be solved by a sorting algorithm in $\mathcal{O}(N)$ time, where N is the ensemble size. For dimensions greater than one, we used a direct linear programming algorithm that solves the assignment problem in $\mathcal{O}(N^3 \log N)$ time. In numerical experiments for one dimensional processes we found that $\beta = 2$ scaling could be achieved, leading to optimal scaling of the whole multilevel Monte Carlo algorithm. When considering higher dimensional problems such as stochastic versions of Lorenz '63 or Lorenz '96, we obtained $\beta = 1$ which is still possible to obtain optimal scaling for sufficiently efficient algorithms. Optimal scaling of the whole algorithm is not possible due to the $\mathcal{O}(N^3 \log N)$ scaling of the assignment calculation, but we expect that this part of the computation will be dominated by the forward model cost in e.g. large dimensional geophysical simulations. More recently, we have found an improvement to the algorithm, that we call the "seamless" multilevel ETPF. The problem with the algorithm

described above is that independent resampling of coarse and fine ensembles in a difference estimator decorrelates the two ensembles, and then we have to solve an assignment problem using a direct solver. It would be much better to try to maintain correlation between the coarse and fine ensembles throughout the Bayesian transform step. The idea of the seamless version is to reverse the steps described above. After reweighting the ensemble members in both coarse and fine ensembles in each difference estimator, we compute the optimal coupling between the coarse and fine ensembles, and produce a new coarse ensemble using fine ensemble members (possibly coarse-grained if there is a spatial discretisation) but with different weights. Now we have pairs of ensembles members with the same state vector, but different weights. We then resample the fine ensemble and the new coarse ensemble using the ETPF procedure. In this version of the algorithm, we always solve optimal transportation problems rather than assignment problems (the latter of which seeks integer solutions rather than real valued solutions). This allows for more accurate approximation of the transformed distributions, as well as the possibility of using iterative solvers for the linear programming problems. In numerical experiments, we have obtained $\beta = 2$ scaling for higher dimensional problems, provided that the coarsest ensemble is large enough. Although we do not have rigorous analysis to confirm it, we believe that this is because the seamless version of the transform gives a more accurate estimate of the posterior distribution than the old version of the algorithm. We are currently investigating the application of the seamless algorithm, in combination with ensemble inflation and localisation techniques, to discretisations of nonlinear PDEs such as the quasigeostrophic equations.

References

[1] M. B. Giles, *Multilevel Monte Carlo Path Simulation*, Oper. Res. **56** (3) (2008), 607–617.

[2] S. Reich, *A nonparametric ensemble transform method for bayesian inference*, SIAM Journal on Scientific Computing **35** 4 (2013), A2013–A2024.

[3] S. Reich, C. Cotter, *Probabilistic forecasting and Bayesian data assimilation*, Cambridge University Press, (2015).

[4] A. Jasra, K. Kamatani, K. Law, Y. Zhou, *Multilevel particle filter*, arXiv:1510.04977, (2015).

[5] H. Hoel, K. Law, R. Tempone, *Multilevel ensemble kalman filtering*, SIAM Journal on Numerical Analysis **54** 3 (2016), 1813–1839.

[6] A. Gregory, C. Cotter, S. Reich, *Multilevel ensemble transform particle filtering*, SIAM Journal on Scientific Computing **38** 3 (2016), A1317–A1338.

Multilevel Monte Carlo for inference

Kody Law

For half a century computational scientists have been numerically simulating complex systems. Uncertainty is recently becoming a requisite consideration in complex applications which have been classically treated deterministically. This has led to an increasing interest in recent years in uncertainty quantification (UQ). Another recent trend is the explosion of available data. Bayesian inference provides

a principled and well-defined approach to the integration of data into an a priori known distribution. The posterior distribution, however, is known only point-wise (possibly with an intractable likelihood) and up to a normalizing constant. Monte Carlo methods have been designed to sample such distributions, such as Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC) samplers. Recently, the multilevel Monte Carlo (MLMC) framework has been extended to some of these cases, so that approximation error can be optimally balanced with statistical sampling error, and ultimately the Bayesian inverse problem can be solved for the same asymptotic cost as solving the deterministic forward problem. This talk will concern the recent development of multilevel SMC (MLSMC) samplers [1] and the resulting estimators for standard quantities of interest as well as normalizing constants [2]. The methods have been applied successfully to nonlocal equations [3], which are used to model anomalous diffusion and fractures in materials. MLMC data assimilation methods have also been developed, which combine dynamical systems with data in an online fashion. Examples are ML particle filters [4] and ensemble Kalman filters [5].

References

[1] A. Beskos, A. Jasra, K. J. H. Law, R. Tempone, Y. Zhou, *Multilevel sequential Monte Carlo samplers.*, Stochastic Processes and Applications (2016).
[2] P. Del Moral, A. Jasra, K. J. H. Law, Y. Zhou, *Multilevel sequential Monte Carlo samplers for normalizing constants.*, arXiv preprint arXiv:1603.01136 (2016).
[3] A. Jasra, K. J. H. Law, Y. Zhou, *Forward and inverse uncertainty quantification using multilevel Monte Carlo algorithms for an elliptic nonlocal equation.*, arXiv preprint arXiv:1603.06381 (2016).
[4] A. Jasra, K. Kamatani, K. J. H. Law, Y. Zhou, *Multilevel particle filter.*, arXiv preprint arXiv:1510.04977 (2015).
[5] H. Hoel, K. J. H. Law, R. Tempone, *Multilevel ensemble Kalman filtering.*, SIAM J. Numer. Anal. **54** (2016), 1813–1839.

What the collapse of the ensemble Kalman filter tells us about particle filters

Matthias Morzfeld
(joint work with Daniel Hodyss, Chris Snyder)

The ensemble Kalman filter (EnKF) is a reliable data assimilation tool for high-dimensional meteorological problems. On the other hand, the EnKF can be interpreted as a particle filter, and particle filters collapse in high-dimensional problems. We explain that these seemingly contradictory state- ments offer insights about how particle filters function in certain high-dimensional problems, and in particular support recent efforts in meteorology to "localize" particle filters, i.e., to restrict the influence of an observation to its neighborhood.

Particle Filters in high dimensions: equal weights on the typical set

Peter Jan van Leeuwen
(joint work with Javier Amezcua, Mengbin Zhu, Mel Ades)

Introduction

Bayesian inference in high-dimensional systems is a major challenge both for practitioners and mathematicians. Intuition obtained from low-dimensional systems is not necessarily useful, and obtaining guidance from numerical experimentation is hard because of computational expense.

We are interested in the Bayesian inference problem for a dynamical system with state evolution equation $x^k = f(x^{k-1}) + \eta^k$ in which the state $x^k \in \Re^n$, $f(..)$ is the nonlinear deterministic model, and $\eta^k \in \Re^n$ is the model noise, and the index k denotes discreet time t_k. Bayesian inference encodes prior knowledge about this system, represented by the state vector in a prior probability density function (pdf) denoted $p(x^k)$. When new information about the system becomes available as new observations of the system $y^k \in \Re^m$, this prior pdf is transformed into the so-called posterior pdf $p(x^k|y^k)$ via Bayes Theorem, given by:

$$p(x^k|y^y) = \frac{p(y|x^y)}{p(y^k)} p(x^k) \tag{1}$$

Note that we denote each pdf by its argument for simplicity of notation. The observations are related to the true state of the system via the so-called observation equation $y = H(x_{true}) + \epsilon$ in which the observation operator H maps an element of the state space to an element of the observation space and the observation error is represented by $\epsilon \in \Re^m$.

It is easy to show that the Bayesian inference problem for e.g. numerical weather forecasting, in which the dimension is $n = 10^9$ at present is to big to be stored in even the biggest super computer. Assuming 10 frequency bins per dimension one would need to store of the order of 10^{10^9} real numbers, a total number much larger than estimates of the number of atoms in the whole universe ($O(10^{80})$). Hence we will always have to make approximations.

Particle filters are a Monte-Carlo approximation to the Bayesian inference problem in which the prior is represented by a set of delta functions centred around the particles x_i^{k-p}, $p > 0$, $i \in \{1, \cdots, N\}$. The problem to be solved can be formulated as follows. Given these particles x_i^{k-p}, find the best representation of $p(x^k|y^k)$. 'Best' is defined by the user, for instance the most accurate first moments of this pdf. While common wisdom is that particle filters fail in high dimensional systems due to the curse of dimensionality a more general class of particle filters has been explored recently that does seem to allow for particle filters that beat the curse of dimensionality (see e.g. [3], [4]). We will discuss one of these particle filters in the following.

The typical set

When sampling from a pdf (probability density function) in a high dimensional

space it is very unlikely to sample close to the mode. To build intuition on why this is the case consider the following example.

Suppose we want to draw samples from a n-dimensional Gaussian $N(0, I)$, in which I is the $n \times n$ dimensional identity matrix. We can simply generate samples from that density by sampling the coordinates of the sample vector as $\xi^{(j)} \sim N(0, 1)$. The square of the length of this vector is a χ^2 variable with mean n and variance $2n$. For N_x large this square is Gaussian distributed with the same mean and variance. This means that more than 99% of the probability mass of the pdf of the square of the length of a random vector lies in the interval $\left[n - 3\sqrt{2n}, n + 3\sqrt{2n}\right]$, which is far away from the mean and mode of the pdf of the random vectors when n is large. Interestingly enough the distance of each element of that random vector to the mean and mode will be small, of order 1, but adding the squares of all these small number will end up as a large number.

This means that the probability mass is not at the mode/mean, but some distance away from that, and random samples from a pdf will end up in the areas of high probability mass. These areas of high probability mass form the so-called Typical Set: for $\epsilon > 0$ and any n, we define the *typical set* A with respect to probability density $p(x)$ as follows:

$$A_\epsilon^{(n)} = \left\{ (x) \in S^n : \left| -\frac{1}{n} \log p(x) - H(X) \right| \leq \epsilon \right\} \tag{2}$$

and it can be shows that for each $\epsilon > 0$ and $\delta > 0$ there exist an n such that

$$Prob\left(X \in A_\epsilon^{(n)} \right) > 1 - \delta \tag{3}$$

The importance for us is that for our particles to represent the posterior pdf well they have to be on the typical set, because that is where the probability mass is. Advanced particle filters like the Implicit Particle Filter ([1]) do bring particles close to the typical set. This has not been proven in general, but it is easy to prove that that filter puts particles on the typical set for a Gaussian posterior pdf for systems of arbitrary dimension, and numerical experimentation shows that the method has good first-order moments in low-dimensional systems ([2]).

The problem with the IPF is that the relative weights of the particles vary widely in high-dimensional systems. such that typically one particle has a weight very close to one, while all other particles have weights close to zero. This is a so-called degenerate configuration because when estimating moments of the pdf, like mean and covariance, the weighted particles need to be used, leading to very poor estimates of these moments. So an extra step is needed to ensure that the particles have (near) equal weights.

Hamiltonian Monte Carlo

The idea is to start from particles near the typical set and use a Markov-Chain Monte-Carlo method to bring them on the set and at the same time ensure they obtain equal weights. Hamiltonian Monte Carlo (HMC) can be used in such a method, with the advantage that large moves in state space can be made while that move is still accepted in the Metropolis-Hastings ratio in the HMC methodology.

Furthermore, it can be shown that HMC moves will keep the particles on the typical set. Because HMC is computationally expensive for each step we propose to perform just one step, as follows:

$$x_i^n = x_i^* - \tau^2 P^{1/2} \frac{dE}{dx} + \alpha_i \tau P^{1/2} v_i \tag{4}$$

in which x_i^* is a particle generated with the Implicit Particle Filter, $E(x) \propto -\log p(x|y)$, τ is the size of the HMC step, and the v_i are the -artificial- velocity variables introduced in the HMC algorithm.

Of special note is the scalar variable α_i, which is used to ensure equal weight for all particles in the following way. First we use the Implicit Particle Filter to find x_i^* for each particle i. Then we calculate $dE(x)/dx$ and sample $v_i \sim N(0, I)$, with identity matrix $I \in \Re^{n \times n}$. This leaves us with $x_i^n(\alpha_i)$. The resulting expression for x_i^n is then put into the expression for the weight of particle i, which is given by:

$$w_i(\alpha_i) \propto \frac{p(y|x_i^n)p(x_i^n|x_i^*)p(x_i^*|x_i^{n-1})}{q(x_i^n|x_{1:N}^*)q(x_i^*|x_i^{n-1}, y^n)} \tag{5}$$

We now set a target weight w_{target}, set $w_i(\alpha_i) = w_{target}$ for each particle and solve for scalar α_i for each particle i. In this way we ensure that the particles are on or close to the typical set while having equal weight.

Challenges

Although the above might be a step in the right direction, huge challenges remain. First and foremost, it is vital that ideas like this get a proper mathematical foundation. Work is ongoing, but no results have been obtained yet. Furthermore, issues arise because we don't have an exact expression for $p(x|y)$ in particle filtering, so $E(x)$ needed in the HMC step is only known approximately via the particles. And then there are several practical issues, for instance related to the fact that one has to prescribe the statistics of the errors in the model equations, which is extremely hard to obtain. It will be clear that this is not the last word on this matter, but hopefully it is a start.

References

[1] A. J. Chorin, X. Tu, *Implicit sampling for particle filters.*, PNAS **106** (2009), 17249–17254

[2] M. Morzfeld, A. J. Chorin, *Implicit Particle Filtering for models with partial noise, and an application to geomagnetic data assimilation.*, Nonlinear process. Geophys.**19** (2012), 365–382.

[3] P. J. van Leeuwen, Y. Cheng, S. Reich, *Nonlinear data assimilation.*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials **2** (2015), Springer.

[4] M. Zhu, P. J. van Leeuwen, J. Amezcua, *Implicit equal-weights particle filter.*, Q. J. R. Meteor. Soc. (2016).

New perspectives in importance sampling

Daniel Sanz Alonso

Importance sampling is a building block of many algorithms in computational statistics, perhaps most notably particle filters. It is the importance sampling step that often limits the accuracy of these algorithms. In this talk I will introduce a new way of understanding importance sampling based on information theory. I will argue that the fundamental problem facing algorithms based on importance sampling can be understood in terms of the distance between certain measures. The results give new understanding on the potential use of importance sampling and particle filters in high (possibly infinite) dimensional spaces.

Towards a stable Particle Filter in High Dimensions (Take Two)

Dan Crisan

The purpose of this talk is twofold: First I will attempt to provide a common platform of communication between two sister areas of research: data assimilation and stochastic filtering, particularly its numerical component. Second, I will describe the major difficulties encountered when solving the filtering problem in high dimensions and will try to dispell some of the misconceptions that are still prevalent in the data assimilation literature.

Data assimilation is the process by which observations are incorporated into a computer model of a real system. Applications of data assimilation arise in many fields of geosciences, perhaps most importantly in weather forecasting and hydrology. Data assimilation proceeds by analysis cycles. In each analysis cycle, observations of the current (and possibly past) state of a system are combined with the results from a numerical weather prediction model (the forecast) to produce an analysis, which is considered as the best estimate of the current state of the system. This is called the analysis step. Essentially, the analysis step tries to balance the uncertainty in the data and in the forecast. The model is then advanced in time and its result becomes the forecast in the next analysis cycle.

The stochastic filtering problem involves a pair of processes $(Z, Y) = \{(Z_t, Y_t), t \geq 0\}$ where

- Z is the signal process or the "hidden component"
- Y is the observation process - "the data" and $Y_t = f(Z, \text{“}noise\text{”})$.

and aims to find the conditional distribution of the *signal* Z_t given $\mathcal{Y}_t = \sigma(Y_s, s \in [0, t])$, i.e.,

$$\pi_t(\varphi) = \mathbb{E}[\varphi(Z_t)|\mathcal{Y}_t], \qquad t \geq 0.$$

Both data assimilation (DA) and stochastic filtering are dealing with the same problem of merging models with partial observations. DA focuses on algorithms for large scale problems and large data sets. Stochastic filtering has a wider scope incorporating both theoretical and numerical components. Both fields are moving towards each other.

Ensemble-based algorithms in both areas fit into the following framework:

particle approximations	Gaussian approximations
$\pi_t^M \overset{\text{mutation}}{\underset{\text{model}}{\longrightarrow}} \bar{\pi}_{t+\delta}^M \overset{\text{selection}}{\underset{\{Y_s\}_{s\in[t,t+\delta]}}{\longrightarrow}} \pi_{t+\delta}^M$	$\pi_t^M \overset{\text{forecast}}{\underset{\text{model}}{\longrightarrow}} \bar{\pi}_{t+\delta}^M \overset{\text{assimilation}}{\underset{\{Y_s\}_{s\in[t,t+\delta]}}{\longrightarrow}} \pi_{t+\delta}^M$

The approximations *appear to be* different:

particle approximations	Gaussian approximations
$(\underbrace{a_j(t)}_{\text{weight}}, \underbrace{v_j^1(t),\ldots,v_j^d(t)}_{\text{position}})_{j=1}^M$	$(\underbrace{a_j(t)}_{\text{weight}}, \underbrace{v_j^1(t),\ldots,v_j^d(t)}_{\text{mean}}, \underbrace{\omega_j^{11}(t),\ldots,\omega_j^{dd}(t)}_{\text{covariance matrix}})_{j=1}^M$
$\pi_t \rightsquigarrow \pi_t^M = \sum_{j=1}^M a_j(t)\,\delta_{v_j(t)}$	$\pi_t \rightsquigarrow \pi_t^M = \sum_{j=1}^M a_j(t)\,N(v_j(t),\omega_j(t))$

but the stored information can modelled by, say M stochastic processes

$$\{p_i(t), t>0\} \quad i=1,...,n, \quad p_i(t)\in\mathbb{R}^N.$$

We can think of the processes p_i as the trajectories of M (generalized) particles/ensemble members. Typically $N>d$, where d is the dimension of the state space. The approximation of the conditional distribution of the signal will then take the form

$$\pi_t^M = \Lambda_t^M(p_i(t), t>0 \quad i=1,...,n).$$

I will talk about numerical approximations of the filtering problem in high dimensions, that is, when the hidden state lies in R^d with d large (typical set-up in data assimilation). For low dimensional problems, one of the most popular numerical procedures for consistent inference is the class of approximations termed particle filters or sequential Monte Carlo methods. However, in high dimensions, **standard particle filters (e.g. the bootstrap particle filter)** can have a cost that is exponential in d for the algorithm to be stable in an appropriate sense. I will present a new particle filter, called the *space-time particle filter*, designed for a specific family of state-space models in discrete time. This new class of particle filters provide consistent Monte Carlo estimates for any fixed d , as do standard particle filters. The space time particle filter is expected to scale much better with d than the standard filter. The numerical results suggest that it is indeed possible to tackle some high dimensional filtering problems using the space-time particle filter that standard particle filters cannot handle. The talk is based on the papers [1, 2, 3, 4].

References

[1] A. Beskos, D. Crisan, A. Jasra, *On the stability of sequential Monte Carlo methods in high dimensions*, Ann. Appl. Probab. **24** (2014), no. 4, 1396–1445.

[2] A. Beskos, D. Crisan, A. Jasra, N. Whiteley, *Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions*, Adv. in Appl. Probab. **46** (2014), no. 1, 279–306.

[3] N. Kantas, A. Beskos, A. Jasra, *Sequential Monte Carlo methods for high-dimensional inverse problems: a case study for the Navier-Stokes equations*, SIAM/ASA J. Uncertain. Quantif. **2** (2014), no. 1, 464–489.
[4] A. Beskos, D. Crisan, A. Jasra, K. Kamatani, Y. Zhou, *A Stable Particle Filter in High-Dimensions*, (2016) submitted.

Iterative Ensemble Smoother for Parameter Estimation

Dean Oliver

The key features of inverse problems applied to subsurface flow models are that the number of parameters are large ($\sim 10^5 - 10^7$) and that the relationships between observations and model variables are highly nonlinear and nonlocal. The cost of evaluating the likelihood is relatively high (15 min to 1 day).

I present a method for generating sample realizations that are concentrated in regions of high probability density for the posterior distribution. The samples are generated by minimizing a cost function with perturbed observations and perturbed model parameters. The distribution of the samples can be computed from the Jacobian of the transformation. This distribution is used to evaluate the Metropolis-Hastings test to generate samples from the actual posterior, but they could have been used for importance sampling. The weights are shown to be nearly uniform and the method was shown to scale well even when the prior and the posterior are highly dissimilar.

The method is illustrated with two multimodal toy problems. In actual applications to geoscience problems, an iterative ensemble smoother is used to minimize the cost function. It has not been feasible to compute the weighting of samples in those cases.

Performance of the EnKFs with small ensemble size

Xin Tong

The successful EnKF prediction skill with an ensemble size K much smaller than the dimension d has remained an intriguing mystery. The practitioners often attribute this success to an low effective dimension p, of which the formal definition has never been given. The first part of our framework proposes a natural definition for the effective dimension, using the covariance spectrum of an associated Kalman filter. The second component employs the Mahalanobis norm to quantify the EnKF performance, which is intrinsically dissipative for Kalman type of filter updates. This dissipative mechanism is stable enough to wither the noisy perturbation from model or small sampling error. The low effective dimension plays a vital role here, since when $K > Cp$ for a constant C, the sample forecast covariance matrix can concentrate around its expected value, using a new random matrix theory result. Practical covariance inflation and spectral projection are employed to our EnKF. The fact that these augmentations are necessary for our proof, indicates the theoretical significance of these augmentations, while their practical significance has already been observed and well documented.

On the convergence of (ensemble) Kalman filters and smoothers onto the unstable subspace

Marc Bocquet
(joint work with Alberto Carrassi, Karthik S. Gurumoorthy, Amit Apte, Colin Grudzien, Christopher K. R. T. Jones)

The characteristics of the model dynamics are critical in the performance of (ensemble) Kalman filters. In particular, as emphasized in the seminal work of Anna Trevisan and co-authors, the error covariance matrix is asymptotically supported by the unstable and neutral subspace only, i.e. it is spanned by the backward Lyapunov vectors with non-negative exponents. This behavior is at the heart of algorithms known as Assimilation in the Unstable Subspace, although its formal proof was still missing.

This convergence property, its analytic proof, meaning and implications for the design of efficient reduced-order data assimilation algorithms are the topics of this talk. The structure of the talk is as follows.

Firstly, we provide the analytic proof of the convergence on the unstable and neutral subspace in the linear dynamics and linear observation operator case, along with rigorous results giving the rate of such convergence. The derivation is based on an expression that relates explicitly the covariance matrix at an arbitrary time with the initial error covariance. Numerical results are also shown to illustrate and support the mathematical claims.

Secondly, we discuss how this neat picture is modified when the dynamics become nonlinear and chaotic and when it is not possible to derive analytic formulas. An ensemble Kalman filter (EnKF) is used in this case. We also explain why, in the perfect model setting, the iterative ensemble Kalman smoother (IEnKS), as an efficient filtering and smoothing technique, has an error covariance matrix whose projection is more focused on the unstable-neutral subspace than that of the EnKF.

Conservation laws and the ensemble Kalman filter

Tijana Janjic
(joint work with Yuefei Zeng, Yvonne Ruckstuhl)

Numerical discretization schemes have a long history of incorporating the most important conservation properties of the continuous system in order to improve the prediction of the nonlinear flow. The question arises whether data assimilation algorithms should follow a similar approach and whether the inclusion of conservation laws when computing the initial condition would improve the prediction of the nonlinear flow. To address these issues, we explore for which problems the conservation laws could be beneficial in data assimilation algorithms and in which cases conservation properties are well recovered when using local ensemble Kalman filter (EnKF). We show the behavior of the localized EnKF with respect to preservation of positivity, conservation of mass, energy and enstrophy in toy models that conserve these properties. These toy models are motivated with two

applications of data assimilation, on convective scale for radar data assimilation and on global scale for conventional data assimilation. Both applications show benefit of including some of the conservation laws as constraints. Past studies have shown that EnKF algorithm can be applied to the convective scales since it is capable of handling complex and highly nonlinear processes through use of time evolving error covariances. However, some challenges for the convective scale applications still remain. These include a need to estimate fields that are nonnegative (such as rain, graupel, snow) and to use data sets such as radar reflectivity that have the same property. What underlines these examples are errors that are non-Gaussian in nature causing a problem with the EnKF that uses Gaussian error assumptions to produce the estimates from the previous forecast and the incoming data. Since the proper estimates of hydrometeors are crucial for prediction on convective scales, the question arises whether the EnKF method can be modified to improve these estimates. In this talk, we first illustrated the problem using the non-hydrostatic convection permitting COSMO model, and the localized EnKF (LETKF) [1] as implemented in KENDA (Km-scale Ensemble Data Assimilation) system of German Weather Service [2]. We showed that the analyses of water species would become negative in some grid points of the COSMO model when radar data are assimilated. These values are set to zero after the LETKF analysis step, in order not to give the numerical model unphysical values. The tests done within this setup show that such a procedure introduces a bias in the analysis ensemble with respect to the truth, that increases in time due to the cycled data assimilation. In order to preserve physical properties in the analysis as well as to deal with the non-Gaussianity in an EnKF framework, Janjic et al. 2014 [3] proposed the use of physically based constraints in the analysis step to constrain the solution and therefore change the analysis error statistics. This approach led to the Quadratic Programming Ensemble filter algorithm (QPEns) based on the EnKF and the quadratic programming. In [3] it was shown on a very simple example that for state estimation, the inclusion of the constrained estimation can improve the ensemble Kalman filter results in case of strongly non-Gaussian error distributions. Importantly, only methods that preserved positivity and mass together produced accurate analysis. The QPEns algorithm was further tested on the modified shallow water model [4], which was designed to mimic the most important characteristics of convective motion. Perfect model experiments were performed with observations that are taken every 5 minutes at the grid points where there is rain for all three variables of the modified shallow water model (wind, height of the fluid and rain). It was shown that the mass conservation- and positivity-constrained rain significantly suppresses noise seen in localized EnKF results. This is highly desirable in order to avoid spurious storms from appearing in the forecast starting from this initial condition [5]. In addition, the root mean square error (RMSE) is reduced for all fields and total mass of the rain is correctly simulated. In the second application, using perfect model experiments with mass, total energy and momentum conserving 2D shallow water model that also conserves enstrophy for non-divergent flow, we mimicked with the simple model data

assimilation on global scale over the domain corresponding to Northern Hemisphere with the horizontal resolution of 50 km. The data assimilation used the LETKF with varying localization radius, thinning interval, observed variable and inflation [6]. During assimilation, the total mass remained consistent with that of the nature run and the total energy of the analysis mean converged towards the nature run value. However, enstrophy, divergence, as well as energy spectra were strongly affected by localization radius, thinning interval, and inflation and depended on the variable, which was observed. In this idealized setup, we tested the effects on prediction depending on the type of errors in the initial condition. By measuring nonlinear energy cascade through a scalar, domain averaged, noise term, we show that the accumulated noise during assimilation and the analysis RMSE are good indicators of quality of the prediction. In both applications, the EnKF would violate some of the conservation properties. In convective scale applications, these include preservation of positivity, and, due to setting of the negative values to zero, conservation of mass. In global application, the EnKF was able to recover the correct energy and mass, but it was not able to recover the correct enstrophy and divergence. To this end, imposing the constraints on enstrophy and energy will be tested in this setup in the future.

References

[1] B. R. Hunt, E. J. Kostelich, I. Szunyogh, *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter*, Physica D, **230** (2007), 112–126.

[2] C. Schraff, H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periáñez, R. Potthast, *Kilometre-scale ensemble data assimilation for the COSMO model (KENDA)*, Q. J. R. Meteor. Soc., **142** (2016), 1453–1472.

[3] T. Janjic, D. McLaughlin, S. E. Cohn, M. Verlaan, *Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms*, Monthly Weather Review, **142** 2 (2014), 755–773.

[4] M. Würsch, G. C. Craig, *A simple dynamical model of cumulus convection for data assimilation research*, Meteorol. Z., **23** (2014), 483–490.

[5] H. Lange, G. C. Craig, *The Impact of Data Assimilation Length Scales on Analysis and Prediction of Convective Storms*, Monthly Weather Review, **142** (2014), 3781–3808.

[6] Y. Zeng, T. Janjić, *Study of conservation laws with the Local Ensemble Transform Kalman Filter*, Q. J. R. Meteor. Soc., **142** (2016), 2359–2372.

Balanced data assimilation via blended models from an asymptotic hierarchy

Rupert Klein

The state of the atmosphere is nearly balanced in that it is free of acoustic oscillations of sizeable amplitude and in that it is essentially free of very long-wave internal gravity waves. Numerical weather forecasting schemes are designed to maintain the associated acoustic, hydrostatic, and geostrophic near-balances very well, even if they approximate the full compressible flow dynamics of the atmosphere.

The assimilation of observational data into numerical weather forecasting codes implies adjustments of the model state that are generally not compatible with the model dynamics. In particular, they will perturb the state substantially away from the abovementioned intricate balances. The development of assimilation schemes that avoid such imbalances is an active field of research.

In this presentation I have described a strategy for balanced data assimilation which, instead of focusing on the core data assimilation algorithms, considers the problem from the side of the dynamical model [1]. The acoustic, hydrostatic, and geostrophic balances emerge analytically from asymptotic limits for low Mach, Froude, and Rossby numbers. Reduced dynamical equations for the respective limit situations maintain these balances exactly. The idea promoted in [1] is to utilize numerical discretizations of the full compressible dynamics that robustly default to solvers of the limit equations when the respective dimensionless parameters become small or are set to zero. Depending on the scientific community, such schemes are called "well-balanced", "asymptotic preserving" or"asymptotically adaptive".

The strategy for balanced data assimilation proposed in [1] and presented here consists of first performing the adjustment of the model state according to some chosen data assimilation scheme. The subsequent temporal evolution of the state is then, however, not done right away with the full compressible model. Rather, for a few time steps after assimilation the forecast model is run in an appropriate asymptotic limit mode. During these time steps, the pressure field and flow divergence get adjusted automatically to balanced distributions. Subsequently the constraint is relaxed continuously over a few more time steps and from then on the scheme is run again as a full compressible flow solver.

Various tests including a double pendulum, the slow-fast Lorentz '96 model, and an abruptly started buoyancy-driven flow in a vertical 2-dimensional slice of the atmosphere yield favorable results (see also the poster-contribution by M. Reinhardt, G. Hastermann).

The main advantage of the approach is that the task of balanced data assimilation can be taken care of without any computational overhead, provided the dynamical forecast model's discretization belongs to the class of asymptotically adaptive schemes.

References

[1] T. Benacchio, W. O'Neill, R. Klein, *A blended soundproof-to-compressible numerical model for atmospheric dynamics*, Mon. Wea. Rev., **142** (2014), 4416–4438

Ensemble variational assimilation and Bayesian Estimation

Olivier Talagrand
(joint work with Mohamed Jardak)

Ensemble Variational Assimilation (EnsVAR) consists in perturbing the data to be assimilated according to their error probability distribution, and then perform variational assimilation on the perturbed data. EnsVAR achieves exact Bayesian estimation in the linear and Gaussian case.
EnsVAR is implemented on the Lorenz'96 small-dimension chaotic nonlinear system. It numerically achieves as high a degree of bayesianity (as measured by statistical reliability) as in a purely linear and Gaussian situation. For assimilation windows longer than the predictability time defined by the dominant Lyapunov exponent (10 days or more), high reliability requires the use of Quasi Static Variational Assimilation (QSVA), in which the length of the assimilation window is gradually increased. The performance of EnsVAR compares favourably with that of other ensemble assimilation algorithms, such as Ensemble Kalman Filter and Particle Filter.
Quantitavely similar results are obtained with the Kuramoto-Sivashinsky equation. Similar results are also obtained in the case of a noisy model, assimilation being implemented in weak-constraint form.

Issues in making the weakly-constrained 4DVar formulation computationally efficient

Serge Gratton
(joint work with Selime Gürol, Ehouarn Simon, Philippe Toint)

The main challenges in applying weakly constrained 4D-Var to large scale data assimilation problems are considered. In a first part, three formulations are reviewed. The first two, called state and forcing formulations, take the form of an unconstrained optimization problem and differ in the parametrization of the degrees of freedom. The third instead reformulates the problem as a saddle point calculation. All three formulations lead to a sequence of linear systems which must be solved iteratively, and the crucial ingredient of computational efficiency depends on the time-paralellizable matrix-vector products both in the problem formulation and also in the design of an efficient preconditioner. In our approach, the preconditioner relies on the solution of a very large scale lower triangular system with block bi-diagonal structure. The use of parallel computing for solving this system is therefore a prerequisite for efficient application of the weakly constrained 4D-Var approach. While methods for solving this kind of systems in parallel have long been known for tridiagonal systems, new challenges arise in the structured bi-diagonal case where the entries are possibly very large matrices. A complexity analysis is reported providing a bound on the length of the critical path in a generalization of the block-cyclic reduction algorithm which exploits the special structure of the involved system. It is shown that this method can only be

efficient in parallel if the product of two large matrices representing the model can be computed at a cost bounded by a moderate multiple of that of a matrix-vector product.

Data assimilation has long been an integral and important part of weather forecasting, as new (and often incomplete) meteorological observations are integrated in the ongoing process of predicting the weather for the next few days [1]. The question we consider here is that of using the data to determine a "best" current state of the weather system from which elaborate models may then be evolved in time, providing the desired predictions. Among the possible techniques for this task, variational methods have been applied extensively, typically weighting the use of a priori knowledge (often materialized by the specification of a background state x_0) with the quality of the fit to the data. This is the case, in particular, for the well-known 4D-Var formulation [4, 2]. In recent years, it has also become necessary to take possible model errors into account, thus weighting a priori knowledge, data fitting and model error reduction, an approach which leads to the "weakly-constrained 4D-Var" formulation of the relevant data assimilation problem [14, 15]. In this formulation, the total time horizon (assimilation window) considered is split into a number (N_{sw}) of time sub-windows, and the problem can be be written as

$$\min_{x\in\Re^\ell} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{j=0}^{N_{sw}} \left\|\mathcal{H}_j(x_j) - y_j\right\|^2_{R_j^{-1}} + \frac{1}{2}\sum_{j=1}^{N_{sw}} \|x_j - \mathcal{M}_j(x_{j-1})\|^2_{Q_j^{-1}} \tag{1}$$

where

- $x = (x_0, x_1, \ldots, x_{N_{sw}})^T \in \Re^\ell$ is the control variable (with $x_j = x(t_j)$),
- $x_b \in \Re^n$ is the background given at the initial time (t_0).
- $y_j \in \Re^{m_j}$ is the observation vector over a given time interval
- $\mathcal{H}_j$ maps the state vector x_j from model space to observation space
- $\mathcal{M}_j$ represents an integration of the numerical model from time t_{j-1} to t_j
- B, R_j and Q_j are the covariance matrices for background, observation and model error, respectively.

The incorporation of possible model errors is achieved by the presence of the third term in the objective function.

As it is the case for the standard 4D-Var (consisting of the first two terms in 1), the general unconstrained nonlinear least-squares problem is solved by applying the Gauss-Newton algorithm [3, 7], which iteratively proceeds by linearizing $\mathcal{H}$ and $\mathcal{M}$ at the current iterate and then, often very approximately, minimizing the resulting quadratic function. If the operators M_j are the linearized $\mathcal{M}_j$ and H_j are the linearized $\mathcal{H}_j$, then the problem can be expressed in terms of $\delta x_j = (x - x_0)_j$ as

$$\min_{\delta x\in\Re^\ell} \frac{1}{2}\|L\,\delta x - b\|^2_{D^{-1}} + \frac{1}{2}\|H\,\delta x - d\|^2_{R^{-1}} \tag{2}$$

where the matrices are defined as:

$$L = \begin{pmatrix} I & & & & \\ -M_1 & I & & & \\ & -M_2 & I & & \\ & & \ddots & \ddots & \\ & & & -M_{N_{sw}} & I \end{pmatrix}, \tag{3}$$

$$H = \mathrm{diag}(H_0, H_1, \ldots, H_{N_{sw}}), \quad D = \mathrm{diag}(B, Q_1, \ldots, Q_{N_{sw}}),$$

and

$$R = \mathrm{diag}(R_0, R_1, \ldots, R_{N_{sw}}),$$

and where the vectors are defined as:

$$d = (d_0, d_1, \ldots, d_{N_{sw}})^T \quad \text{and} \quad b = (b_0, c_1, \ldots, c_{N_{sw}})^T,$$

with $b_0 = x_b - x_0$, $\forall j \in [1, \ldots, N_{sw}]$ $c_j = \mathcal{M}_j(x_{j-1}) - x_j$ and $\forall j \in [0, \ldots, N_{sw}]$ $d_j = y_j - mathcal H_j(x_j)$.

(Note the incorporation of the background covariance matrix B in D). The approximate minimization of the quadratic subproblem is itself carried out using a Krylov method (often conjugate gradients [10] or a specialized version of it such as RPCG [9] or RSFOM [8], see also [6]), which typically requires preconditioning for achieving reasonable computational efficiency.

Three variants of the above problem can then be defined. In the form presented above, the formulation is called the "state formulation" and requires solving the linear system of the form

$$(L^T D^{-1} L + H^T R^{-1} H)\, \delta x = L^T D^{-1} b + H^T R^{-1} d.$$

Another version (called the "forcing formulation") may be obtained by making the change of variables $\delta p = L\delta x$, then requiring the solution of the linear system

$$(D^{-1} + L^{-T} H^T R^{-1} L^{-1} H)\, \delta p = D^{-1} b + L^{-T} H^T R^{-1} d.$$

A third version (the "saddle" formulation) is obtained by transforming the terms in 2 in a set of equality constraints and writing the Karush-Kuhn-Tucker conditions for the resulting constrained problem, leading to the large "saddle" linear system

$$\begin{pmatrix} D & 0 & L \\ 0 & R & H \\ L^T & H^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \\ \delta x \end{pmatrix} = \begin{pmatrix} b \\ d \\ 0 \end{pmatrix}$$

where the control vector $[\lambda, \mu, \delta x]^T$ is a $(2\ell + m)$-dimensional vector. For the sake of brevity, we do not cover the details of this latter derivation here, see [13, 12].

The crucial observation is that matrix-vector products with L^{-1} appear sequential at first sight, which is a serious drawback in the context of modern computer architectures for high-performance computing. In this respect, using the forcing formulation is problematic and, even if the state and saddle point formulations allow performing matrix-vector products with L in parallel, their suitable preconditioners involve the operator $\tilde{L}^{-1}$, where $\tilde{L}$ is a bi-diagonal formulation-dependent approximation of L.

The use of this operator therefore centrally depends on the efficient (and possibly approximate) solution of a linear system

$$\tilde{L}u = v, \tag{4}$$

involving the bi-diagonal matrix $\tilde{L}$ whose form is structurally the same as that defined for L in 3 and suitable vectors u and v. While a simple recurrence of the form

$$u_0 = v_0, \quad u_i = v_i + M_i u_{i-1} \quad (i = 1, \ldots, N_{sw}) \tag{5}$$

is obviously possible, this technique is inherently sequential, which is a computational drawback as mentioned before. One is thus naturally lead to considering parallel algorithms for the solution of tridiagonal systems (of which our system is a specially structured instance) and, in particular, the "cyclic reduction" method [5, pp. 177-190]. Broadly speaking and when applied to 4 with $N_{sw} = 5$, this method transforms the system

$$\begin{pmatrix} I & & & & & \\ -M_1 & I & & & & \\ & -M_2 & I & & & \\ & & -M_3 & I & & \\ & & & -M_4 & I & \\ & & & & -M_5 & I \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} \Longrightarrow \begin{array}{l} x_0 = b_0 \\ x_1 = b_1 + M_1 x_0 \\ x_2 = b_2 + M_2 x_1 \\ x_3 = b_3 + M_3 x_2 \\ x_4 = b_4 + M_4 x_3 \\ x_5 = b_5 + M_5 x_4 \end{array}$$

in two independent sub-systems

$$\begin{pmatrix} I & & \\ -M_{21} & I & \\ & -M_{43} & I \end{pmatrix} \begin{pmatrix} x_0 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_2 + M_2 b_1 \\ b_4 + M_4 b_3 \end{pmatrix} \tag{6}$$

and

$$\begin{pmatrix} I & & \\ -M_{32} & I & \\ & -M_{54} & I \end{pmatrix} \begin{pmatrix} x_1 \\ x_3 \\ x_5 \end{pmatrix} = \begin{pmatrix} b_1 + M_1 x_0 \\ b_3 + M_3 b_2 \\ b_5 + M_5 b_4 \end{pmatrix} \tag{7}$$

where $M_{i,j} \stackrel{\text{def}}{=} M_i M_{i-1} \ldots M_j$. The systems 6 and 7 can then be solved in parallel (recursively using cyclic reduction).

While the parallel performance of cyclic reduction has already been analyzed for tridiagonal systems (see [11], for instance), its performance for structured block bi-diagonal systems appears to be unstudied so far. As it is clear from the description above, it hinges on making the product $M_{i+1}M_i$ efficient. Focussing on the simple case (corresponding to one-dimensional models) where the matrix M_i is banded with semi-bandwidth p and where the computation $M_{i+1}M_i$ makes use of this banded structure, it turns out that the relative parallel complexity of solving 4 exactly using N_{sw} processors is given by

$$\frac{\text{wall-clocktime for cyclic red.}}{\text{wall-clocktime for sequential}} \leq \left(\frac{2pN_{sw}}{3} + \frac{\log_2 N_{sw}}{\max[p,1]N_{sw}} + 3 \right)$$

where the "sequential" mode refers to the substitution 5. The wall-clocktime is considered in an idealized perfect scenario: infinite number of processors, infinite memory, and neither I/O time nor communication delay is taken into account. As a consequence, the use of block cyclic reduction cannot be as efficient in parallel as that of the sequential substitution, unless the products $M_{i+1}M_i$ are approximated in a manner that makes the cost of such a matrix-matrix product comparable to that of a matrix-vector product. This may for instance be achieved if each of the products $M_{i+1}M_i$ is itself approximated by a matrix of semi-bandwidth p, in which case the relative parallel complexities become

$$\frac{\text{wall-clocktime for cyclic red.}}{\text{wall-clocktime for sequential}} \leq \frac{(2p+2)\log_2 N_{sw}}{N_{sw}}.$$

This then becomes advantageous when N_{sw} grows. There are several ways to obtain a limited bandwidth approximation of $M_{i+1}M_i$ at various levels of the cyclic reduction recursion. The simplest is just to truncate the product by ignoring entries outside the band, but one can also think of integrating the models with a coarser time step (stability allowing) or using simplified physics models.

All the above assumes that valuable preconditioning information is contained in the model matrices M_i. An alternative and somewhat radical point of view is to ignore this information, in which case approximations such as

$$\tilde{L} = I, \quad \text{or} \quad \tilde{M}_i = \alpha_i I \quad (i = 1, \ldots, N_{sw})$$

for well-chosen scalar α_i may be considered, leading to very cheap and fully-parallel but less informative preconditioners [13]. In particular the use of secant information to update the scalars α_i is an attractive option.

Limited numerical experiments show that the choice of a best strategy among all these options is far from straightforward, and significantly depends on the particular formulation used, the conditioning of the various operator involved as well as on the number of time sub-windows and available processors. Further experiments are therefore necessary to reach stable conclusions on the feasibility of using efficient parallel computations in the context of weakly-constrained 4D-Var data assimilation problem.

References

[1] F. Bouttier and P. Courtier. Data assimilation concepts and methods. Technical report, ECMWF, Reading, England, (1999). ECMWF Meteorological Training Course Lecture Series.

[2] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society* **123** (1997), 2449–2461.

[3] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* **120** (1994), 1367–1388.

[4] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A** (1986), 97–110.

[5] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edition, (1996).

[6] S. Gratton, S. Gürol, Ph. L. Toint, J. Tshimanga, and A. Weaver. Krylov methods in the observation space for data assimilation. Oberwolfach Reports, (2012).
[7] S. Gratton, A. Lawless, and N. K. Nichols. Approximate Gauss-Newton methods for nonlinear least-squares problems. *SIAM Journal on Optimization* **18** (2007), 106–132.
[8] S. Gratton, Ph. L. Toint, and J. Tshimanga. Range-space variants and inexact matrix-vector products in Krylov solvers for linear systems arising from inverse problems. *SIAM Journal on Matrix Analysis* **32(3)** (2011), 969–986.
[9] S. Gratton and J. Tshimanga. An observation-space formulation of variational assimilation using a restricted preconditioned conjugate-gradient algorithm. *Quarterly Journal of the Royal Meteorological Society* **135** (2009), 1573–1585.
[10] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of the National Bureau of Standards* **49** (1952), 409–436.
[11] S.P. Hirshman, K.S. Perumalla, V.E. Lynch, and R. Sanchez. BCYCLIC: A parallel block tridiagonal matrix cyclic solver. *Journal of Computational Physics* **229(18)** (2010). 6392–6404.
[12] M. Fisher, S. Gratton, S. Gürol, X. Vasseur, and Y. Trémolet. Low rank updates in preconditioning the saddle point systems arising from data assimilation problems. *Optimization Methods and Software*, (2016).
[13] M. Fisher and S. Gürol. Parallelisation in the time dimension of four-dimensional variational data assimilation, (2016). submitted to Quarterly Journal of the Royal Meteorological Society.
[14] Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society* **132(621)** (2006) 2483–2504.
[15] Y. Trémolet. Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society* **133(626)** (2007), 1267–1280.

On existence and uniqueness of solutions for variational data assimilation in discrete time

Jochen Bröcker

Data assimilation is a term from the geosciences and refers to methods for estimating orbits of dynamical models from observations. Variational techniques for data assimilation estimate these orbits by minimising an appropriate cost functional which takes the error with respect to the observations but also deviations of the orbits from the model equations into account. Such techniques are very important in practice. In this contribution, the problem of existence and uniqueness of solutions to variational data assimilation is investigated. Under mild hypotheses a solution to this problem exists. The problem of uniqueness is investigated as well, and several results (which all have analogues in optimal control) are established in the present context. The value function is introduced as the cost of an optimal trajectory starting from a given initial condition. The necessary conditions in combination with an envelope theorem can be used to demonstrate that the solution is unique if and only if the value function is differentiable at the given initial condition. This occurs for all initial conditions except maybe on a set of Lebesgue measure zero. Several examples are studied which demonstrate that non–uniqueness of solutions cannot be ruled out altogether though, which has important consequences in practice.

References

[1] J. Bröcker, "Existence and Uniqueness For Four Dimensional Variational Data Assimilation in Discrete Time.", SIAM Journal of Applied Dynamical Systems (accepted).

Annealing method to find the smallest minimum action (cost function) in variational methods

Henry Abarbanel

Formulating the statistical data assimilation problem as a path integral [1] provides a powerful framework for estimating moments of the conditional distribution, states and parameters of a model conditioned on observations. The path integral also provides a structure within which one can systematically evaluate the corrections to ideas about estimating the integral and a guide to using the information in the measurements more efficiently. This talk focuses on using this framework to identify the global maximum of the conditional probability distribution in the space of paths of the model system through time.

References

[1] H. Abarbanel, *Predicting the Future.*, Springer, (2013).

Variational approximation to inference for stochastic differential equations

Manfred Opper

In this talk I will discuss an approximation method for inference of multivariate stochastic differential equations of the form $dX_t = f_\theta(X_t)dt + D^{1/2}(X_t)dW_t$, where θ is a set of parameters. We assume that we have discrete time, noisy observations $\mathcal{Y} \doteq (y_1, \ldots, y_n)$ and that these observations are described by a likelihood $p(\mathcal{Y}|X_{0:T}) = \prod_{k=1}^{N} p(y_k|X_{t_k})$ where $X_{0:T}$ denotes the path of the process over a time window T. The goal will be to estimate the parameters θ by a maximum likelihood or Bayesian approach, and, given the parameters, to predict the path $X_{0:T}$, e.g. by the computation of the conditional expectation $E[X_t|\mathcal{Y}]$. It is possible to solve this problem using Markov chain Monte Carlo methods by treating the unobserved process as a latent random variable. However, this approach could be time consuming for higher dimensional problems, because one has to draw samples of the entire path conditioned on the observations. Here I will discuss a different approximate approach which has its origin in the field of statistical physics and which has been applied in recent years to inference in probabilistic models in the area of machine learning. In this method, one tries to approximate the (analytically or computationally intractable) conditional measure over paths $P(X_{0:T}|\mathcal{Y})$ by an approximation $Q(X_{0:T})$ which is a member of a tractable family of measures, e.g. Gaussian ones. Q is chosen by minimising the Kullback–Leibler divergence (relative entropy) $D(Q\|P) = \int dQ \ln(\frac{dQ}{dP})$ over the tractable family of measures.

One can show that the optimal Gaussian measure corresponds to an approximate linear stochastic differential equation $dX_t = (A_t X_t + b_t)dt + D^{1/2}(X_t)dW_t$ where the explicit time dependency of the variational parameter functions A_t and b_t take the fact into account that the conditional process (through the observations) is no longer homogeneous. The variational problem can be solved using forward–backward iterations for small dimensionalities. The approach also yields an approximation for the evidence (likelihood) $P(X_{0:T}|\theta)$ which can be used for parameter estimation. Higher dimensional problems can be treated efficiently using further approximation, e.g. a factorising one, where the individual components of the process X_t are treated as independent.

Estimating model evidence using data assimilation

Alberto Carrassi
(joint work with Marc Bocquet, Alexis Hannart, Michael Ghil)

This study focuses on the problem of quantifying the resulting performance of a state inference by estimating the so-called marginal likelihood – also sometimes referred to as model evidence – which quantifies the "goodness-of-fit" between the data and the chosen state-space model. Model evidence can be used as a general metric for model selection and comparison, relevant to many different purposes frequently faced by both scientists and practitioners: e.g. calibrating the model parameters, comparing the skill of several candidate models (or model settings, or boundary conditions) in representing the observed signal, or even evidencing the existence (or non-existence) of a causal relationship between an external forcing and an observed response. We review the field of data assimilation (DA) from a Bayesian perspective and show that, in addition to its by now common application to state estimation, DA may be used for model selection. An important special case of the latter is the discrimination between a factual model – which corresponds, to the best of the modeler's knowledge, to the situation in the actual world in which a sequence of events has occurred – and a counterfactual model, in which a particular forcing or process might be absent or just quantitatively different from the actual world. Three different Gaussian ensemble-DA methods are reviewed for this purpose: the ensemble Kalman filter (EnKF), the ensemble four-dimensional variational smoother (En-4D-Var), and the iterative ensemble Kalman smoother (IEnKS). An original contextual formulation of model evidence (CME) is introduced. It is shown how to apply these three methods to compute CME, using the approximated time-dependent probability distribution functions (pdfs) each of them provide in the process of state estimation. The theoretical formulae so derived are applied to two simplified nonlinear and chaotic models: (i) the Lorenz three- variable convection (L63) model, and (ii) the Lorenz 40-variable mid-latitude atmospheric dynamics model (L95). The numerical results of these three DA-based methods and those of an integration based on importance sampling are compared. It is found that better CME estimates are obtained by using DA, and the IEnKS method appears to be best among the DA methods.

Differences among the performance of the three DA-based methods are discussed as a function of model properties. Finally, the methodology is implemented for parameter estimation and for event attribution. The next step in applying the present DA-based approach to more realistic models and observational scenarios is to consider climate models of intermediate complexity. This application-oriented research activity has to be supported and accompanied by two theoretical lines of investigations, namely the extension of the present results (i) in the presence of model error, and (ii) in conjunction with spatial localization techniques. Preliminary results in the context of an intermediate complex primitive equation model of the atmosphere are presented in which the model evidence formulae are re-derived to account for the spatial localization.

Nonparametric probabilistic modeling and data assimilation

John Harlim

I will discuss a nonparametric modeling approach for forecasting stochastic dynamical systems on smooth manifolds embedded in Euclidean space. In the limit of large data, this approach converges to a Galerkin projection of the semigroup solution of the backward Kolmogorov equation of the underlying dynamics on a basis adapted to the invariant measure. This approach, which we called the "diffusion forecast", allows one to evolve the probability distribution of non-trivial dynamical systems with an equation-free modeling. I will also discuss a data-driven nonparametric method to estimate likelihood functions which can be used to estimate observation model error distribution.

Low-dimensional transports for Bayesian filtering and smoothing

Youssef Marzouk
(joint work with Alessio Spantini, Daniele Bigoni)

In this talk, we present a variational approach to characterizing complex (e.g., high-dimensional, non-Gaussian) probability distributions, and discuss its particular application to problems of sequential Bayesian inference—filtering, smoothing, and joint state/parameter inference. The central idea is to approximate the target distribution using *transportation of measures* [8].

Let $\boldsymbol{Z}$ be a random variable on $\mathbb{R}^n$ endowed with an intractable continuous density π that we wish to simulate. We assume that π is available only up to a normalizing constant, as is typical in Bayesian inference. One possible approach to the problem of sampling is to seek a deterministic transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ that couples $\boldsymbol{Z}$ with a tractable reference random variable $\boldsymbol{X}$ (e.g., a standard normal) [4]. If we let η denote the density of $\boldsymbol{X}$, then we can say that T pushes forward η to π, i.e., $T_\sharp \eta = \pi$, where, for any invertible map T we have $T_\sharp \eta(\boldsymbol{z}) = \eta(T^{-1}(\boldsymbol{z})) \cdot |\det \nabla T^{-1}(\boldsymbol{z})|$. (Here we abuse notation slightly, by writing the notion of a pushforward measure in terms of densities with respect to the Lebesgue measure on $\mathbb{R}^n$.) A map satisfying this coupling condition enables

the generation of cheap, independent, and unweighted samples from π; one simply has to evaluate T on a collection of iid reference samples.

A transport map between distributions on $\mathbb{R}^n$ exists under very weak conditions. For instance, in the example above it suffices that the law of $\boldsymbol{X}$ vanish on subsets of (Hausdorff) dimension $n-1$. As shown in [5], the transport map can be computed via deterministic optimization by minimizing the Kullback–Leibler divergence $\mathcal{D}_{\mathrm{KL}}(\, T_\sharp \eta \,||\, \pi\,)$ over a suitable function space for the map, i.e., for $T \in \mathcal{T}$. The approach adopted in [5] seeks a parametric transport map within a finite dimensional approximation space, $\mathcal{T}_h \subset \mathcal{T}$. The resulting variational problem reads as $\min_{T\in\mathcal{T}_h} \mathcal{D}_{\mathrm{KL}}(\, T_\sharp \eta \,||\, \pi\,)$. The particular approximation spaces we consider are subsets of the cone $\mathcal{T}_\triangle$ of "triangular" maps that are monotone increasing with respect to the lexicographic order on $\mathbb{R}^n$ [4]. $\mathcal{T}_\triangle$ includes the Knothe-Rosenblatt (KR) rearrangement [6] between η and π. In fact, the Rosenblatt map can be characterized as the unique minimizer of $\mathcal{D}_{\mathrm{KL}}(\, T_\sharp \eta \,||\, \pi\,)$ over $\mathcal{T}_\triangle$ for absolutely continuous η and π.

The key observation of this work is that a transport map is not just *any* multivariate function on $\mathbb{R}^n$. There exist transports that inherit low-dimensional parameterizations from the Markov structure [3] of the underlying target density. By considering recursive graph decompositions of a (non-complete) Markov network for π, it is possible to prove the existence of transports T that factorize *exactly* as the composition of k low-dimensional maps, $T = T_1 \circ \cdots \circ T_k$, for some finite k, where each map T_j differs from the identity function only along few components and is triangular up to a permutation of the input and output space. We call such transports *decomposable*. Clearly, a decomposable transport is easier to parameterize than a regular one. Moreover, the decomposition $T = T_1 \circ \cdots \circ T_k$ suggests that the computation of T may be broken into multiple simpler steps, each associated with the computation of a low-dimensional map T_j that accounts only for *local* features of π.

We instantiate this observation in the context of Bayesian filtering and smoothing [7]. Consider the problem of sequential Bayesian inference for a discrete-time, continuous-state, nonlinear, and non-Gaussian state-space model, in a general formulation that also includes hyperparameters (i.e., static parameters) to be inferred. Let $(\boldsymbol{Z}_k)_{k\geq 0}$ denote the unobserved latent Markov process (each $\boldsymbol{Z}_k$ is a random variable on $\mathbb{R}^n$), $(\boldsymbol{Y}_k)_{k\geq 0}$ denote the observed process, and $\boldsymbol{\Theta}$ represent the hyperparameters of the model, which are treated as an $\mathbb{R}^p$-valued random variable. The state-space model is then fully specified in terms of the conditional densities $\pi_{\boldsymbol{Y}_k|\boldsymbol{Z}_k,\boldsymbol{\Theta}}$, $\pi_{\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k,\boldsymbol{\Theta}}$, $\pi_{\boldsymbol{Z}_0|\boldsymbol{\Theta}}$, and the marginal density $\pi_{\boldsymbol{\Theta}}$, together with the observed data $(\boldsymbol{y}_k)_{k\geq 0}$. We assume these are given. Our goal is then to characterize, sequentially, the posterior distribution $\pi_{\boldsymbol{\Theta},\boldsymbol{Z}_{0:k}|\boldsymbol{Y}_{0:k}}$ for all $k \geq 0$, from which any filtering distributions $\pi_{\boldsymbol{Z}_k|\boldsymbol{Y}_{0:k}}$ or smoothing distributions $\pi_{\boldsymbol{Z}_j|\boldsymbol{Y}_{0:k}}$ with $j < k$, along with the parameter marginals $\pi_{\boldsymbol{\Theta}|\boldsymbol{Y}_{0:k}}$, are readily available.

Based on the recursive decomposition of the Markov network associated with the posterior distribution above, we propose a new *deterministic* and recursive algorithm for online inference with transport maps. In a *single forward pass*,

the algorithm computes a sequence of triangular maps $(\mathfrak{M}_j)_{j\geq 0}$, each of fixed dimension $(2n+p)$, that, properly composed, are capable of sampling $\pi_{\boldsymbol{\Theta},\boldsymbol{Z}_{0:k}|\boldsymbol{Y}_{0:k}}$ for all $k \geq 0$. Unlike most smoothing algorithms, the present algorithm does not resort to any *backward pass* that touches the state-space model. The composition step involves embedding the submaps $(\mathfrak{M}_j)$ into higher-dimensional identity maps to form the sequence $(T_k)_{k\geq 0}$. We can then evaluate $\mathfrak{T}_k := T_0 \circ \cdots \circ T_k$ to sample directly from $\pi_{\boldsymbol{\Theta},\boldsymbol{Z}_{0:k+1}|\boldsymbol{Y}_{0:k+1}}$ and obtain information about any smoothing or filtering distribution of interest. Successive transports in the sequence $(\mathfrak{T}_k)_{k\geq 0}$ are *nested* and thus ideal for online inference. The variational yet online character of the algorithm seems to distinguish it from existing approaches to nonlinear and non-Gaussian smoothing and joint parameter inference [1, 2].

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. "Particle Markov chain Monte Carlo methods." *J. R. Stat. Soc. B* **72** (2010), 269–342.

[2] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin. "On particle methods for parameter estimation in state-space models." *Statist. Sci.* **30** (2015), 328–351.

[3] S. L. Lauritzen. *Graphical Models*. Oxford University Press, (1996).

[4] Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, "Sampling via measure transport: an introduction," In *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, H. Owhadi, eds., Springer (2016). Also arXiv:1602.05023.

[5] T. Moselhy, Y. Marzouk, "Bayesian inference with optimal maps," *J. Comp. Phys.* **231** (2012), 7815–7850.

[6] M. Rosenblatt. "Remarks on a multivariate transformation." *Ann. Math. Statist.* **23** (1952), 470–472.

[7] A. Spantini, D. Bigoni, Y. Marzouk. "Variational inference via decomposable transports: algorithms for Bayesian filtering and smoothing." In Neural Information Processing Systems (NIPS) Workshop on Advances in Approximate Inference, (2016).

[8] C. Villani. *Optimal Transport: Old and New*. Springer, (2008).

A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation

Amit Apte

(joint work with Laura Slivinski, Elaine Spiller, Bjorn Sandsted)

Lagrangian data assimilation (LaDA) refers to the use of observations provided by (pseudo-)Lagrangian instruments such as drifters, floats, and gliders, which are important sources of surface and subsurface data for the oceans. After giving a brief introduction to augmented state approach to LaDA and issues specific to this problem, I will describe a recent proposal for a hybrid particle-Kalman filter method for LaDA. The main motivation for the hybrid filter is as follows: particle filters work well for nonlinear systems but the computational effort required grows exponentially with dimensions of the system, whereas Kalman filters are successful in dealing with large dimensional problems which are close to being linear. Thus neither of them work well for high dimensional, highly nonlinear systems. On the other hand, in LaDA, the dynamics of the Lagrangian drifters is typically

low dimensional but highly nonlinear, whereas the flow in which they move is high dimensional but less nonlinear. The hybrid filter attempts to combine the strengths of both these filters and the specific structure of the Lagrangian dynamics, by using an ensemble Kalman filter for the velocity flow while at the same time using a particle filter for the Lagrangian drifters. I present some promising results about the efficacy of this proposed method and discuss its shortcomings.

Bridging sequences for Sequential Monte Carlo

Nikolas Kantas
(joint work with Alex Beskos, Ajay Jasra, Alex Thierry, Dan Crisan, Francesc Pons-Llopis)

Traditional particle filtering or Sequential Monte Carlo (SMC) methodology has been extremely successful in low dimensional non-linear non-Gaussian applications (e.g [1]), but their application in high dimensional settings has been very challenging mainly due to the difficulty to perform importance sampling efficiently in high dimensions ([2], [3], [4]). Despite this challenge a few successful high dimensional particle filtering implementations have appeared recently for data assimilation applications when the hidden signal obeys discrete time dynamics ([5], [6], [7], [8], [9], [10]). In this talk we present past and ongoing work for addressing problems related to dimensionality or more generally the mismatch between successive posterior distributions in the SMC target sequence. We will focus on the use of tempering and MCMC steps and also outline how these methods can be tuned adaptively. In addition we present recent theoretical results on the method and relate them with the issue of dimensionality. Finally we present numerics related to high dimensional inverse problems or filtering problems when the signal of interest obeys the deterministic or stochastic Navier Stokes equation that is observed at discrete times with noise.

References

[1] A. Doucet, J. F. G. de Freitas, N. J. Gordon *Sequential Monte Carlo Methods in Practice.*, Springer, (2001).

[2] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson *Obstacles to high-dimensional particle filtering.*, Monthly Weather Review **136** (2008), 4629–4640.

[3] T. Bengtsson, P. Bickel, B. Li *Curse-ofdimensionality revisited: Collapse of the particle filter in very large scale systems.*, Probability and Statistics: Essays in Honor of David A. Freedman, D. Nolan and T. Speeds, Eds., Institute of Mathematical Statistics **2** (2008), 316–334.

[4] P. B. Quang, C. Musso, F. Le Gland *An insight into the issue of dimensionality in particle filtering.*, 13th Information Fusion (FUSION) conference, (2010).

[5] A. J. Chorin, X. Tu, *Implicit sampling for particle filters.*, Proc. Nat. Acad. Sc. USA **106** (2008), 17249–17254.

[6] N. Papadakis, E. Mmin, A. Cuzol, N. Gengembre *Data assimilation with the weighted ensemble Kalman filter.*, Tellus **62A** (2010), 673–697.

[7] S. Reich *A guided sequential Monte Carlo method for the assimilation of data into stochastic dynamical systems.*, Recent Trends in Dynamical Systems, Springer Proceedings in Mathematics and Statistics **35** (2013), 205–220.
[8] P. J. van Leeuwen *Nonlinear data assimilation in geosciences: An extremely efficient particle filter.*, Q. J. R. Meteor. Soc. **136** (2009), 1991–1999.
[9] M. Ades, P. J. van Leeuwen *An exploration of the equivalent weights particle filter.*, Q. J. R. Meteor. Soc. **139** (2013), 820–840.
[10] J. Weare *Particle filtering with path sampling and an application to a bimodal ocean current model.*, J. Comput. Phys. **228** (2009), 4312–4331.

Localization-induced filter instability and a simple adaptive localization method

Lars Nerger
(joint work with Paul Kirchgessner, Angelika Bunse-Gerstner)

Localization is a standard method for the application of ensemble-based Kalman filters with high-dimensional models, e.g. [1]. While localization is widely used, it theoretical basis is not yet fully understood. The application of localization can lead to surprising results as is demonstrated for the case of ensemble Kalman filters that use serial processing of observations [2, 3]. Based on experiments using the Lorenz-96 model [4] it is shown that the interaction of serial observation processing with localization can result in a filter instability. It is found that if the assimilation has a strong influence, the error of the state estimate can jump during the data assimilation series to a level significantly larger than the error without assimilating any observations. This instability results from an inconsistent update of the state error covariance matrix. The resulting error in the covariance matrix accumulates during the serial assimilation of the scalar observations. While the instability is visible only in filters with serial observation, the inconsistent matrix updates also exist in filters that assimilate all observations at once, e.g. [5, 6, 7]. However, in this case the forecast phase directly following the analysis update hides this inconsistency [8].

Localization methods require usually a costly tuning for optimal filtering performance. The second part of the talk discusses an adaptive localization method, which avoids the need for tuning for the case that dense observations (e.g. observerations of one field at all grid points) are assimilated. Numerical experiments show that a minimal estimation error is obtained when the number of assimilated observations is approximately equal to the ensemble size. In case of localization with weighted observations, this holds for the effective observation number given by the sum of the weights. This finding is used to formulate an adaptive localization method by choosing localization radii such that the effective observation number is equal to the ensemble size for each local analysis domain. This method works successfully, as is demonstrated by an experiment assimilating sea surface height into a global ocean model [9]. Also, the German Weather Service (DWD) has implemented this method in their ensemble assimilation system [10].

References

[1] T. M. Hamill and J. S. Whitaker and C. Snyder, *Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter*, Mon. Wea. Rev. **129** (2001), 2776–1790.

[2] J. L. Anderson *An Ensemble Adjustment Kalman Filter for Data Assimilation*, Mon. Wea. Rev. **129** (2001), 2884–2903.

[3] J. S. Whitaker and T. M. Hamill, *Ensemble Data Assimilation without Perturbed Observations*, Mon. Wea. Rev. **130** (2002), 1913–1927.

[4] E. N. Lorenz, *Predictability - A problem partly solved* in Proceedings Seminar on Predictability, ECMWF, Reading, UK, (1996), pp. 1–18.

[5] C. H. Bishop, B. J. Etherton, and S. J. Majumdar, *Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects*, Mon. Wea. Rev. **129** (2001), 420–436.

[6] D. T. Pham, *Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems*, Mon. Wea. Rev. **129** (2001), 1194–1207.

[7] L. Nerger and T. Janjić and J. Schröter and W. Hiller, *A unification of ensemble square root Kalman filters*, Mon. Wea. Rev. **140** (2012), 2335–2345.

[8] L. Nerger, *On serial observation processing on localized ensemble Kalman filters*, Mon. Wea. Rev. **143** (2015), 1554–1567.

[9] P. Kirchgessner and L. Nerger and A. Bunse-Gerstner, *On the Choice of an Optimal Localization Radius in Ensemble Kalman Filter Methods*, Mon. Wea. Rev. **142** (2014), 2165–2175.

[10] C. Schraff and H. Reich and A. Rhodin and A. Schomburg and K. Stephan and A. Periánez and R. Potthast, *Kilometre-scale ensemble data assimilation for the COSMO model (KENDA)*, Q. J. R. Meteorol. Soc. **142** (2016), 1453–1472.

Correlated Observation Errors in Data Assimilation

Nancy Nichols

(joint work with Joanne A. Waller, Jemima M. Tabeart, Sarah L. Dance, Amos S. Lawless, David Simonin, Susan P. Ballard, Graeme A. Kelly)

With the development of convection-permitting numerical weather prediction, the efficient use of high resolution observations in data assimilation is becoming increasingly important. In current practice, however, the observation errors are assumed to be uncorrelated and about 80% of the available observations are discarded, partly in order to remove observation error correlations from the assimilation. Observation uncertainty arises from a number of sources, including instrument errors, pre-processing errors, observation operator uncertainty and representativity errors (scale mismatch between observations and model data). Calculation of observation error statistics is difficult as they cannot be measured directly. Nevertheless, idealized and operational studies have shown that a better treatment of observation error correlations gives improved forecast skill. Using a diagnostic that makes use of statistical averages of background and analysis innovations [1], we derive observation error statistics for different data types. We develop new theory on the effect of changes in the assumed error statistics used in the assimilation on the estimated observation error covariance matrix [2]. If it is assumed in the assimilation that the correlation matrix is diagonal but the observation errors are

in fact correlated, then it is likely that the diagnostic will underestimate the variance and the correlation length-scales of the observation errors; nevertheless the diagnostic will provide an improved estimate of the observation error covariance matrix. We apply the diagnostic to determine horizontal and along-range observation error covariances for Doppler-radar radial winds [3]. The diagnostic has also been applied to SEVIRI instrument data and to Atmospheric Motion Vectors [4],[5].

In practice the implementation of correlated observation errors in data assimilation schemes is challenging. We investigate the application of correlated observation errors in a 1D-variational system used operationally for satellite retrievals and examine the conditioning of the optimization problem. The condition number measures the sensitivity of the solution to perturbations in the data of the problem and the computational work needed to solve the problem. We present theoretical bounds on the condition number and show experimentally that the bounds are tight and that the condition number increases as: the observations become more accurate; the observation spacing decreases; the prior (background) becomes less accurate; the prior error correlation length scales increase; and the observation error covariance matrix becomes ill-conditioned [6],[7]. In particular we find that the conditioning of the assimilation scheme depends on the minimum eigenvalue of the observation error correlation matrix. To improve the conditioning and reduce the operational costs of the assimilation, we recondition the observation error correlation matrix by altering its eigenstructure. Experiments demonstrate that incorporating the reconditioned observation error correlation matrices in the assimilation has minimal effect on temperature retrievals but does have impact on humidity retrievals. The application to 4D variational systems and techniques for preconditioning the assimilation problem are currently under investigation.

References

[1] G. Desroziers, L. Berre, B. Chapnik, P. Poli, *Diagnosis of observation, background and analysis-error* ***131*** (2005), 3385–3396.

[2] J. A. Waller, S. L. Dance, N. K. Nichols, *Theoretical insight into diagnosing observation error correlations using background and analysis innovation statistics*, Q. J. R. Meteor. Soc. **142 (694)** (2016a), 418–431.

[3] J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, S. P. Ballard, *Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics*, Monthly Weather Review **144** (2016b), 3533–3551.

[4] J. A. Waller, S. P. Ballard, S. L. Dance, G. A. Kelly, N. K. Nichols, D. Simonin, *Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics*, Remote Sensing **8** (2016c), 581 (14pp).

[5] M. Cordoba, S. L. Dance, G. A. Kelly, N. K. Nichols, J. A. Waller, *Diagnosing Atmospheric Motion Vector observation errors in an operational high resolution data assimilation system*, Q. J. R. Meteor. Soc. (2016).

[6] S. A. Haben, *Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation*, PhD thesis, Dept. of Mathematics & Statistics, University of Reading, (2011).

[7] J. M. Tabeart, *On the variational data assimilation problem with non-diagonal observation weighting matrices*, MRes thesis, Dept of Mathematics & Statistics, University of Reading, (2016).

A coherent structure approach to data assimilation

John Maclean

(joint work with Naratip Santitissadeekorn, Chris KRT Jones)

We introduce a Lagrangian data assimilation method to estimate model parameters by directly assimilating Lagrangian Coherent Structures, where they exist. Our approach is different from previous research, where parameters are estimated based on tracer trajectories. Our numerical approach is based on the Approximate Bayesian Computation (ABC) framework and does not require knowledge of the likelihood function of the coherent structure, which is usually unavailable and produces filter degeneracy when the observations are sufficiently precise [1]. We solve the ABC by a Sequential Monte Carlo (SMC) method [2], and use Principle Component Analysis (PCA) to identify the coherent patterns from tracer trajectory data. Our new method shows remarkably improved results compared to the bootstrap particle filter when the number of (passive) tracers is large, the dynamical system is noisy and the observations are highly informative (i.e. there is small variance in observational error statistics).

References

[1] P. Del Moral, L. M. Murray, *Sequential Monte Carlo with highly informative observations*, SIAM/ASA J. on Uncertainty Quant,**3** (2015), 969–997.

[2] P. Del Moral, A. Doucet, and A. Jasra, *An adaptive sequential Monte Carlo method for approximate Bayesian computation*, Statistics and Computing **22** (2012), 1009–1020.

On a direct data-driven reduction of Bayesian models

Ilia Horenko

The applicability of many computational approaches is dwelling on identification of reduced dynamical models defined on a small set of collective variables (colvars). The popular approaches to Bayesian model reduction rely on the knowledge of the full matrix of relations between the systems components. In many application areas these matrices are not directly available and must first be estimated from the data, resulting in the uncertainty of the obtained models and colvars. A simple-to-implement clustering methodology for probability-preserving identification of reduced Bayesian models and colvars directly from the data is presented - not relying on the availability of the full relation matrices at any stage of the resulting algorithm. Methodology is demonstrated on several applications from climate research, computational fluid mechanics and molecular dynamics - covering various scales relevant for the multiscale geophysical problems.

Analysis of the Ensemble Kalman Filter for Inverse Problems

Claudia Schillings
(joint work with Andrew M. Stuart)

The Ensemble Kalman filter (EnKF) has had enormous impact on the applied sciences since its introduction in the 1990s by Evensen and coworkers. The low computational costs, the straightforward implementation and the non-intrusive nature make the EnKF appealing in various areas of application, but, on the downside, the method is underpinned by very limited theoretical understanding. We will present an analysis of the EnKF based on the continuous time scaling limits, which allows to study the properties of the EnKF for fixed ensemble size.

Let $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ denote the forward response operator mapping from the parameter space $\mathcal{X}$ to the observation space $\mathcal{Y} = \mathbb{R}^K$. We assume that the operator $\mathcal{G}$ is continuous and $\mathcal{X}$ is a separable Hilbert space. The inverse problem is then to recover the unknown $u \in \mathcal{X}$ from noisy observation $y \in \mathcal{Y}$ where

$$y = \mathcal{G}(u) + \eta. \tag{1}$$

The observational noise η is assumed to be Gaussian, i.e. $\eta \sim \mathcal{N}(0, \Gamma)$ for a given symmetric, positive definite matrix $\Gamma \in \mathbb{R}^{K\times K}$. We define the least-squares functional by

$$\Phi(u; y) = \frac{1}{2}\|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))\|^2_{\mathcal{Y}}. \tag{2}$$

In the Bayesian framework, we view (u, y) as a jointly varying random variable in $\mathcal{X} \times \mathcal{Y}$ and, under the assumption that $u \sim \mu_0$, the solution to the inverse problem is the $\mathcal{X}-$valued random variable $u|y$ distributed according to measure

$$\mu(du) = \frac{1}{Z}\exp\big(-\Phi(u; y)\big)\mu_0(du), \tag{3}$$

where Z denotes the normalization constant $Z := \int_{\mathcal{X}} \exp\big(-\Phi(u; y)\big)\mu_0(du)$. The EnKF is derived within the Bayesian framework and, through its ensemble properties, is viewed as approximating the posterior distribution on the random variable $u|y$. However, except in the large sample limit for linear Gaussian problems, there is little to substantiate this viewpoint. Here we will choose a different perspective viewing the EnKF as a regularization technique for minimization of the least-squares misfit functional Φ. The basic variant of the EnKF for inverse problems can be summarized as follows: the EnKF maps the ensemble $\{u^j_n\}_{j=1}^J$, where $J \in \mathbb{N}$ is the number of ensemble particles, in the n-th iteration according to

$$u^{(j)}_{n+1} = u^{(j)}_n + C^{up}(u_n)(C^{pp}(u_n) + h^{-1}\Gamma)^{-1}\big(y^{(j)}_{n+1} - \mathcal{G}(u^{(j)}_n)\big), \quad j = 1, \cdots, J, \tag{4}$$

where

$$y^{(j)}_{n+1} = y + \xi^{(j)}_{n+1}$$

and, for $u = \{u^{(j)}\}_{j=1}^J$, we define the operators C^{pp} and C^{up} by

$$C^{pp}(u) = \frac{1}{J}\sum_{j=1}^{J}\big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big) \otimes \big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big), \tag{5}$$

$$C^{up}(u) = \frac{1}{J}\sum_{j=1}^{J}\big(u^{(j)} - \overline{u}\big) \otimes \big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big), \tag{6}$$

$$\overline{u} = \frac{1}{J}\sum_{j=1}^{J} u^{(j)}, \quad \overline{\mathcal{G}} = \frac{1}{J}\sum_{j=1}^{J} \mathcal{G}(u^{(j)}). \tag{7}$$

Interpreting the update step of the EnKF as a time-stepping scheme, i.e. assuming that $u_n \approx u(nh)$, we take the limit $h \to 0$

$$\frac{\mathrm{d}u^{(j)}}{\mathrm{d}t} = C^{up}(u)\Gamma^{-1}(y - \mathcal{G}(u^{(j)})) + C^{up}(u)\Gamma^{-1}\sqrt{\Sigma}\frac{dW^{(j)}}{dt}, \tag{8}$$

where $W^{(j)}$ are independent cylindrical Brownian motions on $\mathcal{X}$.

The study of the linear forward problem gives us useful insights into properties of the method and allows us to establish well-posedness results, quantitative bounds on the ensemble collapse and convergence results for a fixed ensemble size. Furthermore, variants such as variance inflation and localization, together with new ideas borrowing from the use of sequential Monte Carlo (SMC) method for inverse problems can be investigated in this framework.

Numerical results indicate that the conclusions observed for linear problems carry over to nonlinear problems. Theoretical considerations for the nonlinear problem will be subject of future work.

References

[1] C. Schillings, A. M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, arXiv:1602.02020v1 (2016).

Reporter: Maria Reinhardt

Participants

Prof. Dr. Henry Abarbanel
Department of Physics
Scripps Institution of Oceanography
University of California, San Diego
9500 Gilman Drive
La Jolla CA 92093-0374
UNITED STATES

Dr. Walter Acevedo
Institut für Mathematik
Universität Potsdam
Postfach 601553
14415 Potsdam
GERMANY

Prof. Dr. Amit Apte
International Centre for Theoretical Sciences (ICTS)
Survey No. 151, Shivakote
Hesaraghatta Hobli
Bengaluru 560 089
INDIA

Dr. Elisabeth Bauernschubert
German Meteorological Service
Deutscher Wetterdienst
Research and Development Division FE 12
(Data Assimilation)
Frankfurter Straße 135
63067 Offenbach am Main
GERMANY

Prof. Dr. Marc Bocquet
CEREA
Laboratory of Ecole des Ponts
6-8 Avenue Blaise Pascal
77455 Marne-la-Vallée Cedex 2
FRANCE

Dr. Jochen Bröcker
Department of Mathematics
University of Reading
Whiteknights
P.O.Box 220
Reading RG6 6AX
UNITED KINGDOM

Prof. Dr. Angelika Bunse-Gerstner
FB 3 - Mathematik u. Informatik
Universität Bremen
Postfach 330440
28334 Bremen
GERMANY

Dr. Alberto Carrassi
Mohn-Sverdrup Center
NERSC
Thormoehlengsgt. 47
5006 Bergen
NORWAY

Dr. Colin Cotter
Department of Mathematics
Imperial College London
Huxley Building
London SW7 2AZ
UNITED KINGDOM

Prof. Dr. Dan Crisan
Department of Mathematics
Imperial College of Science, Technology and Medicine
180 Queen's Gate, Huxley Bldg.
London SW7 2BZ
UNITED KINGDOM

Angwenyi Nyachae David
Institut für Mathematik
Universität Potsdam
Karl-Liebknecht-Str. 24/25
14476 Potsdam
GERMANY

Dr. Jana de Wiljes
Institut für Mathematik
Universität Potsdam
Karl-Liebknecht-Straße 24-25
14476 Potsdam
GERMANY

Prof. Dr. Jason E. Frank
Mathematisch Instituut
Universiteit Utrecht
P.O. Box 80.010
3508 TA Utrecht
NETHERLANDS

Prof. Dr. Georg A. Gottwald
School of Mathematics and Statistics
The University of Sydney
Sydney NSW 2006
AUSTRALIA

Dr. Serge Gratton
CERFACS
42, Avenue Gustave Coriolis
31057 Toulouse Cedex
FRANCE

Prof. Andreas Griewank
School of Mathematical Sciences and
Information Technology
Yachay Tech
Hacienda San José
Urcuquí, Imbabura
ECUADOR

Prof. Dr. John Harlim
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
UNITED STATES

Gottfried Hastermann
Fachbereich Mathematik & Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Prof. Dr. Illia Horenko
Facoltà di Scienze Informatiche
Università della Svizzera Italiana
Via Giuseppe Buffi 13
6904 Lugano
SWITZERLAND

Dr. Tijana Janjic Pfander
Hans-Ertel-Centre for Weather Research
Institute of Meteorology
Ludwig-Maximilians-Universität
München
Theresienstrasse 37
80333 München
GERMANY

Dr. Nicolas Kantas
Department of Mathematics
Imperial College London
Huxley Building
180 Queen's Gate
London SW7 2AZ
UNITED KINGDOM

Prof. Dr. Rupert Klein
FB Mathematik und Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Dr. Peter Korn
Max Planck Institut für Meteorologie
Universität Hamburg
Bundesstrasse 55
20146 Hamburg
GERMANY

Prof. Dr. Hans Rudolf Künsch
Seminar für Statistik
ETH-Zentrum Zürich
HG G 14.2
Rämistrasse 101
8092 Zürich
SWITZERLAND

Dr. Kody J.H. Law
Oak Ridge National Laboratory
Building 920 A
P. O. Box Y
Oak Ridge, TN 37831
UNITED STATES

Dr. Hernan Leövey
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Dr. John Maclean
Department of Mathematics
University of North Carolina
at Chapel Hill
Phillips Hall
Chapel Hill, NC 27599-3250
UNITED STATES

Prof. Dr. Youssef M. Marzouk
Department of Aeronautics &
Astronautic
Massachusetts Institute of Technology
Room 33-305
77 Massachusetts Avenue
Cambridge MA 02139
UNITED STATES

Prof. Dr. Matthias Morzfeld
Department of Mathematics
University of Arizona
P.O.Box 210089
Tucson AZ 85721-0089
UNITED STATES

Dr. Lars Nerger
Alfred-Wegener-Institut
Helmholtz-Zentrum f. Polar- u.
Meeresforschung
Am Handelshafen 12
27570 Bremerhaven
GERMANY

Prof. Dr. Nancy Nichols
Department of Mathematics
University of Reading
Whiteknights
P.O.Box 220
Reading RG6 6AX
UNITED KINGDOM

Dr. Dean S. Oliver
Uni CIPR
Centre for Integrated Petroleum
Research
Realfagbygget 4010, Allégaten 41
5007 Bergen
NORWAY

Prof. Dr. Manfred Opper
Methoden der Künstlichen Intelligenz
Technische Universität Berlin
Sekr. MA 4-2
Marchstraße 23
10587 Berlin
GERMANY

Prof. Dr. Roland Potthast
Deutscher Wetterdienst
Research and Development
Head Section FE12
Frankfurter Straße 135
63067 Offenbach am Main
GERMANY

Prof. Dr. Sebastian Reich
Institut für Mathematik
Universität Potsdam
Karl-Liebknecht-Straße 24-25
14476 Potsdam
GERMANY

Maria Reinhardt
Institut für Mathematik
Universität Potsdam
Postfach 601553
14415 Potsdam
GERMANY

Dr. Sylvain Robert
Seminar für Statistik
ETH-Zentrum Zürich
HG G 14.2
Rämistrasse 101
8092 Zürich
SWITZERLAND

Prof. Dr. Ian Roulstone
Department of Mathematics
University of Surrey
Guildford GU2 7XH
UNITED KINGDOM

Yvonne Ruckstuhl
Meteorological Institute
Ludwig-Maximilians-Universität
Theresienstrasse 37
80333 München
GERMANY

Dr. Naratip Santitissadeekorn
Department of Mathematics &
Computer Science
University of Surrey
Guildford GU2 7XH
UNITED KINGDOM

Daniel Sanz-Alonso
Department of Applied Mathematics
Brown University
Room 204
170 Hope Street
Providence, RI 02912
UNITED STATES

Prof. Dr. Claudia Schillings
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM

Prof. Dr. Chris Snyder
National Center for Atmospheric
Research
(NCAR)
Foothill Laboratory
1850 Table Mesa Drive
Boulder CO 80305-3000
UNITED STATES

Alessio Spantini
Department of Aeronautics &
Astronautics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307
UNITED STATES

Prof. Dr. Wilhelm Stannat
Fachbereich Mathematik
Technische Universität Berlin
Sekr. MA 7-2
Strasse des 17. Juni 136
10623 Berlin
GERMANY

Dr. Olivier Talagrand
Laboratoire de Meteorologie Dynamique
E N S
24, rue Lhomond
75231 Paris Cedex 05
FRANCE

Prof. Dr. Edriss S. Titi
Department of Computer Science
and Applied Mathematics
The Weizmann Institute of Science
Rehovot 76100
ISRAEL

Prof. Dr. Philippe L. Toint
Department of Mathematics
The University of Namur
FUNDP
61, rue de Bruxelles
5000 Namur
BELGIUM

Prof. Dr. Xin Tong
Department of Mathematics
National University of Singapore
Room 806, Block S17
10 Lower Kent Ridge Road
Singapore 119 076
SINGAPORE

Prof. Dr. Peter Jan van Leeuwen
Department of Meteorology
University of Reading
Earley Gate
PO Box 243
Reading RG6 6BB
UNITED KINGDOM