Mathematisches Forschungsinstitut Oberwolfach

Report No. 58/2012

DOI: 10.4171/OWR/2012/58

Mathematical and Algorithmic Aspects of Atmosphere-Ocean Data Assimilation

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

2 December – 8 December 2012

Abstract. The nomenclature "data assimilation" arises from applications in 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 quantifications 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, probability, scientific computing and mathematical modelling, 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 (2000): 65C05, 62M20, 93E11, 62F15, 86A22, 49N45.

Introduction by the Organisers

The workshop *Mathematical and Algorithmic Aspects of Atmosphere-Ocean Data Assimilation*, organised by Andreas Griewank (Berlin), Sebastian Reich (Potsdam), Ian Roulstone (Surrey), and Andrew Stuart (Warwick) was held 2 December – 8 December 2012. The meeting was attended by over 45 participants representing a broad range of mathematical subject areas as well as practical aspects of atmosphere-ocean data assimilation.

A total of 23 talks were presented during the workshop. The talks were selected both to cover novel mathematical developments and to point towards practical advances and challenges in atmosphere-ocean data assimilation. Talks relating to mathematical developments include those on, e.g., Lagrangian data assimilation (Chris Jones, Amin Apte), particle filters (Chris Snyder, Dan Crisan, Wilhelm Stannat), ensemble Kalman filters (Georg Gottwald, Lars Nerger, Tijana Janjic-Pfander, Roland Potthast), statistical inference (Youssef Marzouk, Andreas Hense, Illia Horenko), variational techniques (Eldad Haber, Philippe Toint, Jim Purser, Arnd Rösch, Michael Hinze), model/representativity errors (Nancy Nichols, Alberto Carrassi) and mathematical fluid dynamics (Edriss Titi). Talks relating to practical advances and challenges in atmosphere-ocean data assimilation include those by Chris Jones, Roland Potthast, Andreas Hense, Chris Snyder, Hendrik Elbern, Georg Craig, Alberto Carassi, Patrick Heimbach). A poster session was held on Tuesday evening which gave the attending PhD students and postdocs the opportunity to present and discuss their work.

Throughout the workshop a large 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 shall be mentioned in particular: (i) filter stability (inspired by the talks by Wilhelm Stannat and Roland Potthast), (ii) Bayesian inference and optimal transportation (inspired by the talk by Youseff Marzouk), and (iii) high-dimensional particle filters (inspired by the talks by Chris Snyder and Dan Crisan). These examples also reflect the actuality and broad scientific appeal of data assimilation.

Workshop: Mathematical and Algorithmic Aspects of Atmosphere-Ocean Data Assimilation

Table of Contents

Abstracts

Challenges of Lagrangian data assimilation

Christopher K.R.T. Jones

Lagrangian data come from observations of the ocean using instruments that, at least in part, move with the ocean flow. These instruments have two great advantages over other ocean measuring instruments: they open up the sub-surface to observations, which is otherwise opaque to satellite based measurements, and they can increase data coverage by virtue of their ability to move. The data have, however, not been systematically nor properly assimilated into ocean models. Part of the reason comes from the technical issues associated with these non-standard measurements. We have addressed these challenges with a scheme known as Lagrangian Data Assimilation (LaDA), see [1, 2].

Much recent progress has been made in understanding LaDA more deeply, making it effective and using it for observational design that optimizes the information content of the Lagrangian data obtained. In particular, recent work has focused on the vertical propagation of information, see [3], integrating and optimizing vehicle control with the assimilation of data obtained by that vehicle, see [4],.

The scheme can be challenged if the observational frequency is not high enough. The result is that the nonlinear character of the Lagrangian dynamics can cause filter divergence. In [5], it is shown that a particle filter can effectively correct this, but it is also pointed out that this does not rescue LaDA from the challenges of dimensionality and these remain to be resolved completely.

References

[1] K.Ide, L.Kuznetsov, and C.Jones, *A method for assimilation of Lagrangian data*, Monthly Weather Review **131** (2003), 2247–2260

[2] H. Salman, L.Kuznetsov, K.Ide, and C.Jones, *A method for assimilating Lagrangian data into a shallow-water equation ocean model* Monthly Weather Review **134** (2006), 1081–1101

[3] N.Santitissekadorn, E.Spiller, and C.Jones, *Cross-layer observability using Lagrangian data assimilation*, submitted

[4] D.McDougall, and C.Jones, *Decreasing flow uncertainty in Lagrangian data assimilation through drifter control*, in preparation

[5] A.Apte, and C.Jones, *The impact of nonlinearity on Lagrangian data assimiltion*, submitted

Performance bounds for particle filters using the optimal proposal

Chris Snyder

Particle filters may be cast in the form of sequential importance sampling and thus allow the choice of a proposal density. This choice is known to be crucial to the performance of the algorithm. In particular, the proposal density influences a phenomenon known as "degeneracy," in which the maximum of the particle weights approaches 1. Degeneracy can occur spuriously when N_e, the number of

particles, is too small and in that case the quality of estimates obtained from the particle filter will be poor.

Often, the proposal is taken to be $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, which we will term the *standard* proposal, where $\mathbf{x}_k$ is the state vector at time t_k. For that choice, Bengtsson et al. (2008) and Snyder et al. (2008) have given asymptotic arguments that quantify, under mild assumptions, when degeneracy occurs.

To summarize those asymptotic results, let

$$V_k^i = -\log(w_k^i/w_{k-1}^i),$$

where w_k^i is the weight of the ith particle $\mathbf{x}_k^i$ at time t_k. For the standard proposal, $w_k^i/w_{k-1}^i = p(\mathbf{y}_k|\mathbf{x}_k^i)$, where $\mathbf{y}_k$ are noisy observations related to $\mathbf{x}_k$. Bengtsson et al. (2008) and Snyder et al. (2008) show that, when the components of $\mathbf{x}_k$ are sufficiently independent (under the proposal distribution) and the dimension N_x of $\mathbf{x}_k$ is large, then V is approximately Gaussian when considered as a function of the random variable $\mathbf{x}_k$ with $\mathbf{y}_k$ known. Since the maximum weight corresponds to the minimum of V_k^i over i, known properties of the sample minimum from a Gaussian and for the tails of the Gaussian density and cumulative distribution function can then be used to show that the maximum weight, $w^{(N_e)}$, behaves according to

$$E(1/w^{(N_e)}) \sim 1 + \frac{\sqrt{2\log N_e}}{\tau}, \tag{1}$$

for N_e, τ large and $\sqrt{\log N_e}/\tau \gg 1$, where $\tau^2 = \mathrm{var}(V_k^i)$. This means that the maximum weight approaches 1 (and degeneracy occurs) when τ^2 is large, unless N_e increases as $\exp(\tau^2/2)$. As discussed in Snyder et al. (2008), this puts considerable restrictions on the problems that can be tackled using the particle filter with the standard proposal.

As already noted, the performance of a particle filter can depend strongly on the choice of proposal. A preferred choice is the "optimal" proposal, $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$, which differs from the standard proposal through the additional conditioning on the most recent observations $\mathbf{y}_k$; not only the weights but also the new particles at time t_k depend on $\mathbf{y}_k$. For the optimal proposal,

$$V_k^i = -\log(w_k^i/w_{k-1}^i) = -\log(p(\mathbf{y}_k|\mathbf{x}_{k-1})). \tag{2}$$

This proposal is optimal in the sense that it minimizes the variance of w_k^i over possible draws for $\mathbf{x}_k^i$ from the proposal. In fact, the variance is zero, since w_k^i depends on $\mathbf{x}_{k-1}^i$ but not on $\mathbf{x}_k^i$.

Use of the optimal proposal does not prohibit degeneracy, however. This can be seen in the special case of a linear, Gaussian system, where the asymptotic result (1) can be easily extended to the optimal proposal. We write the system as

$$\mathbf{x}_k = \mathbf{M}\mathbf{x}_{k-1} + \eta_k, \quad \mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \epsilon_k, \tag{3}$$

with $\eta_k \sim N(0, \mathbf{Q})$ and $\epsilon_k \sim N(0, \mathbf{R})$. Equation (3) implies that

$$\mathbf{y}_k = \mathbf{H}\mathbf{M}\mathbf{x}_{k-1} + \mathbf{H}\eta_k + \epsilon_k = \tilde{\mathbf{H}}\mathbf{x}_{k-1} + \tilde{\varepsilon}_k,$$

where $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{M}$ and $\tilde{\varepsilon}_k = \mathbf{H}\eta_k + \epsilon_k$. Because η_k and ϵ_k are Gaussian, $\tilde{\varepsilon}_k$ is also Gaussian: $\tilde{\varepsilon}_k \sim N(0, \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R})$. This means that $p(\mathbf{y}_k|\mathbf{x}_{k-1})$ has the same

form (as a function of $\mathbf{y}_k$ and $\mathbf{x}_{k-1}$) as $p(\mathbf{y}_k|\mathbf{x}_k)$ but with $\mathbf{H}$ replaced by $\tilde{\mathbf{H}}$ and $\mathbf{R}$ replaced by $\mathbf{HQH}^T + \mathbf{R}$.

Thus, the results of section 5 of Snyder et al. (2008) may also be applied to the optimal proposal and $p(\mathbf{y}_k|\mathbf{x}_{k-1})$ but again with $\mathbf{H}$ replaced $\tilde{\mathbf{H}}$ and $\mathbf{R}$ replaced by $\mathbf{HQH}^T + \mathbf{R}$. This yields

$$\tau^2 = \mathrm{var}(V_k^i) = \sum_{j=1}^{N_y} \lambda_j^2 \left(\lambda_j^2/2 + \tilde{y}_{k,j}^2\right), \tag{4}$$

where $\tilde{y}_{k,j}$ is the jth component of $\tilde{\mathbf{y}}_k = (\mathbf{HQH}^T + \mathbf{R})^{-1/2}\mathbf{y}_k$ and λ_j^2 are eigenvalues of

$$A = \mathrm{cov}\left((\mathbf{HQH}^{\mathrm{T}} + \mathbf{R})^{-1/2}\mathbf{HM}\mathbf{x}_{\mathrm{k-1}}\right).$$

In order for the asymptotic relation (1) between the maximum weight, τ and N_e to hold, V_k^i must be approximately Gaussian, which in turn requires that τ and N_e are large and that for each j

$$\frac{\lambda_j^2}{\sum_{i=1}^{N_y} \lambda_i^2} \ll 1.$$

The last condition ensures that no single direction dominates the variability of $\mathbf{y}_k|\mathbf{x}_{k-1}$.

A simple example provides further insight and helps quantify the difference in the performance of the two proposals. Consider the system

$$\mathbf{x}_k = a\mathbf{x}_{k-1} + \eta_{k-1}, \quad \mathbf{y}_k = \mathbf{x}_k + \epsilon_k, \tag{5}$$

where $\mathbf{x}_{k-1} \sim N(0, \mathbf{I})$, $\eta_{k-1} \sim N(0, q^2\mathbf{I})$ and $\epsilon_k \sim N(0, \mathbf{I})$. Equation (4) can now be applied, either following Snyder et al. (2008) for the standard proposal or the manipulations following (4) for the optimal proposal. [Covariances for $\mathbf{y}_k|\mathbf{x}_k$, $\mathbf{y}_k|\mathbf{x}_{k-1}$ and $\mathbf{x}_{k-1}$ are necessary and can be calculated directly from (5), along with $E(\tilde{y}_{k,j}^2)$.] This gives

$$\tau^2 = \mathrm{var}(V) = \begin{cases} N_y(a^2+q^2)\left(\frac{3}{2}a^2 + \frac{3}{2}q^2 + 1\right), & \text{standard proposal} \\ N_y a^2\left(\frac{3}{2}a^2 + q^2 + 1\right)/(q^2+1)^2, & \text{optimal proposal.} \end{cases}$$

The benefits of using the optimal proposal are clear in this example. First, τ^2 is always greater for the standard proposal than for the optimal proposal. The two proposals give the same τ^2 only in the limit that system dynamics has no noise, $q^2 = 0$. As q increases (or a decreases, also increasing the relative importance of the system noise), the difference in τ^2 between the two proposals increases. For $a = q = 1/2$, which makes the prior variance of $\mathbf{x}_k$ equal to the observation-error variance, τ^2 from the standard proposal is 5 times that from the optimal proposal. Since the sample size N_e must grow as $\exp(\tau^2/2)$, reducing τ^2 by a factor of 5 can mean a reduction of N_e by orders of magnitude.

This brief report has outlined how asymptotic arguments of Bengtsson et al. (2008) and Snyder et al. (2008) can be extended to the optimal proposal. This demonstrates that the optimal proposal does not avoid the exponential increase

of the required ensemble size as the system dimension grows. The asymptotic result (1) also serves as a quantitative measure of how much the optimal proposal improves over the standard proposal. In essence, use of the optimal proposal reduces the factor in the exponent in the relation between the ensemble size and the state dimension and so can dramatically reduce the required ensemble size. [Note that (1) does not actually involve the state dimension N_x directly, but rather τ^2. The variance τ^2 is proportional to the state dimension in the example given here but may have a more complicated relation to N_x in the general case when there is dependency among the components of $\mathbf{x}_k$.] The optimal proposal should thus permit the use of particle filters for systems of moderate effective dimension (a few tens or hundreds), even though it does not immediately provide a path to a truly high-dimensional particle filter.

References

[1] T. Bengtsson, P. Bickel and B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, IMS Collections, **2** (2008), 316–334. doi: 10.1214/193940307000000518.

[2] C. Snyder, T. Bengtsson, P. Bickel and J. Anderson, *Obstacles to high-dimensional particle filtering*, Mon. Wea. Rev., **136** (2008), 4629–4640. doi:10.1175/2008MWR2529.1

Particle versus Gaussian Approximations: What is the difference?

Dan Crisan

(joint work with Alex Beskos, Ajay Jasra, Nick Kantas, Kai Li, Salvador Ortiz-Latorre and Joaquin Miguez)

The filtering problem involves the estimation of the current state of an evolving dynamical system based on partial observation. The evolution of the dynamical system is modelled by a stochastic process $X = \{X_t,\ t \geq 0\}$ called the *signal*. The signal process X can not be measured directly. However, a partial measurement of the signal can be obtained. This measurement is modelled by another continuous time process $Y = \{Y_t,\ t \geq 0\}$ which is called the *observation* process. The observation process is a function of X and a measurement noise. The measurement noise is modelled by a stochastic process $W = \{W_t,\ t \geq 0\}$. Hence,

$$Y_t = f_t(X_t, W_t) \qquad t \in [0, \infty).$$

Let $\mathcal{Y} = \{\mathcal{Y}_t,\ t \geq 0\}$ be the filtration generated by the observation process Y; namely, $\mathcal{Y}_t = \sigma\left(Y_s,\ s \in [0,t]\right)$, for $t \geq 0$. Then the filtering problem consists in computing π_t, the conditional distribution of X_t given $\mathcal{Y}_t$. The process $\pi = \{\pi_t,\ t \geq 0\}$ is a $\mathcal{Y}_t$-adapted probability measure valued process, so that

$$\mathbb{E}\left[\varphi(X_t) \mid \mathcal{Y}_t\right] = \int \varphi(x)\pi_t(dx),$$

for all statistics φ for which both terms of the above identity make sense.

Generally speaking the filtering problem can not be solved analytically: An explicit formula cannot be obtained for the conditional distribution π_t. Only in

specific cases such as the Kalman-Bucy filter and the Benes filter (see, eg Chapter 6 in [1]) is this not true. Numerical methods are thus employed to obtain approximations to the solution of the filtering problem. The description of a numerical approximation for the solution of the filtering problem should contain three parts:

1. The method of recording the approximation. The following table contains a succinct description of such a method for particle and respectively, Gaussian approximations:

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

2. The law of evolution of the approximation:

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

3. The measure of the approximating error:

$$\sup_{\{\varphi\in C_b,\ \ \|\varphi\|\le 1\}} \mathbb{E}\left[|\pi_t^n(\varphi) - \pi_t(\varphi)|\right], \quad \hat{\pi}_t - \hat{\pi}_t^n, \quad \|\pi_t^n - \pi_t\|_{TV}.$$

In effect, an approximation can be seen as a way to *quantize* the information available at the current time. The quantized information is modelled by n stochastic processes

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

Typically $N > d$, where d is the dimension of the state space. We think of the processes p_i as the trajectories of n (generalized) particles. Then the approximation will be a function of $\{p_i(t), t>0\}$ $i=1,...,n$, $p_i(t)\in\mathbb{R}^N$. That is

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

Examples of generalized particle filters include classical particle filters, Gaussian approximations, wavelets, grid based methods. In this talk I will discuss a number of results pertaining to the first two categories of generalized particle filters.

In the first part of the talk, I will introduce a class of Gaussian approximations (GAs) for which rigorous convergence rates are deduced. Numerical experiments confirm that for small number of generalized particles, GAs outperform particle filters. I also discuss a class of particle filters based on cubature methods[1], see [4] for details.

In the second part of the talk, I discuss a number of results where particle filters can be adapted to produce an approximation of the posterior density and functions of the posterior density, including the posterior entropy. The adaptation involves a (possibly Gaussian) smoothing procedure, for details see [5]

[1]These methods where originally developed to replace Gaussian measure with discrete measure by matching their moments, see, e.g. [6] and the references therein.

Finally, I look at the filtering problem for high dimensional systems and present a stability result as the dimension of the underlying state space tends to infinity for a class of particle approximations with *fixed* number of particles, for details see [2] and [3]. I conclude with an application of these results to the observed Navier-Stokes equation.

References

[1] A. Bain, D. Crisan, *Fundamentals of stochastic filtering*, Stochastic Modelling and Applied Probability, 60. Springer, New York, 2009.

[2] A. Beskos, D. Crisan, A. Jasra, *On the Stability of Sequential Monte Carlo Methods in High Dimensions.* http://arxiv.org/abs/1103.3965.

[3] A. Beskos, D. Crisan, A. Jasra, N. Whiteley, *Error Bounds and Normalizing Constants for Sequential Monte Carlo in High Dimensions,* http://arxiv.org/abs/1112.1544

[4] D. Crisan, S. Ortiz-Latorre, *A KLV particle filter*, http://www2.imperial.ac.uk/ dcrisan/ Crisan-Ortiz-Latorre-TBBACubFiltering.pdf

[5] D. Crisan, J. Miguez *Particle Approximation of the Filtering Density for State-Space Markov Models in Discrete Time*, http://arxiv.org/abs/1111.5866.

[6] N. Victoir, *Asymmetric cubature formulae with few points in high dimension for symmetric measures* SIAM J. Numer. Anal. 42, no. 1, 209–227, 2004.

Geophysical imaging and reservoir history matching

Eldad Haber

Managing hydrocarbons reservoirs is an important field of reservoir engineering as it aids in more efficient recovery of oil and gas. In principle, the dynamics of the reservoir is governed by the flow in porous media equations. Given the reservoir parameters and the initial conditions, these equations can predict the time evolution of the reservoir and therefore aid in its management.

The challenge in this field stems from the large number of unknown parameter functions in the flow equations. Quantities such as porosity, hydraulic conductivity, mobility functions and capillary pressure functions are, in many cases, specially dependent and unknown. There parameter functions are typically sampled sparsely in space and have a multi-scale structure. To alleviate this problem, historic flow data can be used in order to estimate the unknown parameters. Such a process is often refers to as history matching.

The basic idea of history matching is to intelligently modify the physical properties of the reservoir such that simulated flow data fits the measured one. Technically, the process is rather challenging as it involves with the solution of the forward problem (reservoir simulation), the computation of the gradients of the simulator with respect to the parameters and solving a regularized optimization problem for the coefficients.

However, even though it is possible to estimate some physical properties that fit the flow data these physical properties can be a highly inaccurate representation of the reservoir. The reason being the large null space associated with the inverse problem that is associated with a highly sparse sampling of the reservoir in space. To overcome this problem subspace techniques have been used. In these techniques

one restricts the physical properties to "live" in a small subspace spanned by a (relatively) small number of vectors. This approach decreases the variance in the recovery by increasing the bias of the estimated physical properties.

A different approach to reduce the variance of the recovered earth models is to simply add data and further sample the reservoir. While this is clearly one of the better ways to reduce the uncertainty in the recovery, it is not a practical one as it is unlikely that more wells are drilled just to improve the simulation capability. However, while direct measurements are difficult and expensive to obtain, indirect measurements of reservoir properties are cheap and relatively easy to get. Such measurements include time laps seismic and electromagnetic imaging data. Our interest is give to electromagnetic data as it is sensitive to fluids in the reservoir and has a better likelihood to estimate the flow.

The field of reservoir monitoring using electromagnetic (EM) methods is rather new and it presents a few technological challenges. The main question is how should we fuse the EM monitoring data with standard history matching data? In this work we suggest a new framework to achieve such a goal. We show that the problem of recovery of reservoir parameters can be treated as a joint inversion problem that can be solved by making "reasonable" assumptions on the electromagnetic and reservoir parameters. This in turns, lead to a recovery algorithm where the EM and reservoir parameters are jointly estimated and substantially better recovery is obtained.

Krylov methods in the observation space for data assimilation

Philippe Toint

(joint work with Serge Gratton, Selime Gürol, Jean Tshimanga, Anthony Weaver)

The numerical solution of data assimilation problems in oceanography or weather prediction is often obtained by solving, for a given time widow, the well-known 4DVAR problem [5], which leads to minimizing the nonlinear least-squares function given by

$$\min_{x_0} \|x_0 - x_b\|^2_{B^{-1}} + \sum_{i=1}^{p} \|H(\mathcal{M}(x_0, t_i)) - d_i\|^2_{R^{-1}}, \tag{1}$$

where the state x of the dynamical system under study is described at time t by a dynamical model $\mathcal{M}(x_0, t)$ whose initial condition is given by x_0, where H is the observation operator, x_b is the so-called "background state", d_i is the observation at time t_i, and the matrices B and R are suitable covariance matrices. This minimization attempts, by a suitable choice of initial condition, to balance fitting the evolution of the dynamical system to observed data and the distance of the associated inital condition x_0 from the background state. In the case of interest in this paper, we make the assumption that p, the dimension of the observation space, is much smaller (by orders of magnitude) than that of the space containing the state x.

The nonlinear least-squares problem 1 is usually algorithmically tackled by using a truncated version of the standard Gauss-Newton method, which itself proceeds by solving a sequence of successive linearized subproblems. After suitable transformations, the subproblem solution boils down to computing an approximate solution to a set of equations of the form

$$(I + KL^T)x = b, \tag{2}$$

where b is a suitable right-hans side,often but necessarily belonging to the range space of K, where I is the identity matrix and where the matrices K and L have a number of columns given by the observation space dimension (the case where $L = K$ is of particular interest). This suggests that the term KL^T may be interpreted as a "low rank" modification of the identity (the notion of "low rank" is of course relative, since this rank can be as high as 10^6). Unfortunately, a simple solution consisting in using the Sherman-Morrison-Woodbury formula on the linear system and then applying an Krylov-space iterative method on the resulting $p \times p$ linear system (or variants thereof such as the PSAS method [1, 3] or the equivalent "representer" technique [2]) may often be problematic (see [4, 7]) in the sense that they does not preserve the monotonicity of the local quadratic approximation to 1, a crucial property if one wishes to truncate the solution of the linear system while, at the same time, maintaining global convergence of the overall nonlinear Gauss-Newton process. As it turns out, the structure of 2 may be exploited by a new "observation-space based" family of iterative solvers exploiting Krylov spaces structure (see [7, 6], whose numerical recurrences also involves short (dimension p) vectors. The corresponding variants of conjugate-gradients (CG), full orthogonalization method (FOM), MINRES or GMRES may in particular be derived and may be very elegantly interpreted in terms of a non-standard inner product on the observation space. One crucial property is that the iterates produced by these new variants are mathematically equivalent to those generated by the corresponding standard full state-space Krylov method. As a consequence the new observation-space CG algorithm (RPCG) does maintain the desired monotonicity property of the full space original, even when truncated. The lower iteration costs of the new methods makes them attractive in the context of data assimilation applications.

Preconditioning RPCG, the observation-space version of CG, is also possible while remaining in this space. In particular, limited-memory quasi-Newton preconditioners are shown to have low-dimensional equivalents which satisfy variational properties similar to those of their full-space counterparts. Moreover, their analysis in the observation space provides sharper bounds than those obtained in the full state space. These preconditioners extract information from one iteration of the Gauss-Newton process for use in the next one, which may sometimes be problematic when strong nonlinearity is present. Indeed, the change in the underlying Jacobian matrices over a single iteration may cause loss of symmetry with respect to the non-standard inner product, and strategies must (and can) then be designed to overcome this difficulty while maintaining good performance.

While these new Krylov methods are definitely of interest, their existence suggests the question of whether their dominant cost, the necessary matrix-vector

products typically involving model integration (as is the case for all Krylov based techniques), can be made less expensive, possibly by giving up somme accuracy in the computation of this very product. Such an idea was already explored for general Krylov methods by Simoncini and Szyld [8] and van den Eschof and Sleijpen [9], but it is refined in that formal bounds on residual error evolution are presented, which allow distinguishing between forward error and backward error models for the product. In the first model, the error on the product is interpreted as an error of the result of this calculation, while it is interpreted as a perturbation on the data (the matrix/operator) in the second. It is shown that constant nonzero accuracy threshold can be accommodated in this context, and that the forward error model allows a looser accuracy criterion to be used. It is also argued that Krylov methods with long recurrences (FOM, GMRES) might be preferable to methods using short recurrences, such as CG and MINRES. Finally, the formal bounds on residual convergence for inexact product also ensure numerical stability of the new methods in the presence of round-off errors and resulting errors in the definition of the observation space itself.

References

[1] L. Amodei, *Solution approchée pour un problème d'assimilation de donnéees météorologiques avec prise en compte de l'erreur modèle*, C. R. Acad. Sci., Paris, **321** (1995), 1087–1094.

[2] A. Bennett and M. Thornburn, *The generalized inverse of a non-linear quasi-geostrophic ocean circulation model*, J. Phys. Oceanogr., **3** (1992), 213–230.

[3] S. Cohn, A. Da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, *Assessing the effects of data selection with the DAO physical-space statistical analysis system*, Mon. Weather. Rev., **126** (1998), 2913–2926.

[4] A. El Akkroui, P. Gauthier, S. Pellerin, and S. Buis, *Intercomparison of the primal and dual formulations of variational data assimilation*, Q. J. R. Meteorol. Soc., **134** (2008), 1015–1025.

[5] Ph. Courtier, J.-N. Thépaut, and A. Hollingsworth, *A strategy for operational implementation of 4D-Var using an incremental approach*, Q. J. R. Meteorol. Soc., **120** (1994), 1367–1388.

[6] S. Gratton, Ph. L. Toint, and J. Tshimanga, *Inexact range-space Krylov solvers for linear systems arising from inverse problems*, SIAM Journal on Matrix Analysis and Applications (to appear) (2012).

[7] S. Gratton and J. Tshimanga, *An observation-space formulation of variational assimilation using a restricted preconditioned conjugate-gradient algorithm*, Q. J. R. Meteorol. Soc., **135** (2009),1573–1585.

[8] V. Simoncini and D. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM Journal on Scientific Computing, **25** (2003), 454–477.

[9] J. van den Eshof and G. L. G. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM Journal on Matrix Analysis and Applications, **26** (2004), 125–153.

Constraining overestimation of error covariances in ensemble Kalman filters

Georg A. Gottwald
(joint work with Andy Majda, Lewis Mitchell, Sebastian Reich)

We consider the problem of an ensemble Kalman filter when only partial observations are available. For small ensemble sizes this may lead to an overestimation of the error covariances. We show that by incorporating climatic information of the unobserved variables the variance can be controlled and superior analysis skill is obtained. We then apply this Variance Controlling Kalman Filter to

- sparse observational networks
- balance
- model error
- controlling catastrophic filter divergence

We assume that we have access to proper (noisy) observations are available for some variables (*observables*) but not for other unresolved variables, for which only their statistical climatic behaviour such as their variance and their mean is available (*pseudo-observables*). Observations $\mathbf{x}_{\rm obs}(t_i) = \mathbf{H}\mathbf{z}(t_i) + \mathbf{r}_{\rm obs}(t_i)$ are taken at equidistant observation times $t_n = n\Delta t_{\rm obs}$. Here $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$ is the observation operator, and the the observational noise is assumed to be Gaussian with error covariance matrix $\mathbf{R}_{\rm obs}$. For the pseudo-observables we assume climatic knowledge about the mean $\mathbf{a}_{\rm target}$ and variance $\mathbf{A}_{\rm target}$. We introduce the pseudo-observation operator $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^m$ and the error covariance matrix $\mathbf{R}_{\rm w}$ associated with the pseudo-observables. By requiring that the projected analysis error covariance assumes climatology, we can determine $\mathbf{R}_{\rm w}$.

The algorithm for the filter is as follows:

Step 1: Forecast step

$$\mathbf{Z}_f = F(\mathbf{Z}_b)$$

$$\mathbf{P}_f = \frac{1}{k-1}\mathbf{Z}'_f(t)[\mathbf{Z}'_f(t)]^T$$

Step 2: Analysis step

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f - \mathbf{K}_{\rm obs}(\mathbf{H}\bar{\mathbf{z}}_f - \mathbf{x}_{\rm obs}) - \mathbf{K}_{\rm w}(\mathbf{h}\bar{\mathbf{z}}_f - \mathbf{a}_{\rm target})$$

$$\mathbf{K}_{\rm obs} = \mathbf{P}_f\mathbf{H}^T(\mathbf{H}\mathbf{P}_f\mathbf{H} + \mathbf{R}_{\rm obs})^{-1}\,, \quad \mathbf{K}_{\rm w} = \mathbf{P}_f\mathbf{h}^T(\mathbf{h}\mathbf{P}_f\mathbf{h} + \mathbf{R}_{\rm w})^{-1}$$

$$\mathbf{R}_{\rm w}^{-1} = \mathbf{A}_{\rm target}^{-1} - (\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T)^{-1} \qquad .$$

To assure that $\mathbf{R}_{\rm w}$ is positive definite, we diagonalize and project onto the overestimating subspace.

Step 3: Update of the ensemble
The ensemble needs to be consistent with

$$\mathbf{P}_a = [\mathbf{I} - \mathbf{K}_{\rm obs}\mathbf{H} - \mathbf{K}_{\rm w}\mathbf{h}]\ \mathbf{P}_f \ = \ \frac{1}{k-1}\mathbf{Z}'_a\,[\mathbf{Z}'_a]^T$$

using ensemble square root filters.

Step 4: Update of the forecast
Set $\mathbf{Z}_b = \mathbf{Z}_a$ to propagate the ensemble forward again with the full dynamics to the next observation time.

We have shown in simulations of the Lorenz-96 model that using this filter better skill in the data assimilation procedure is obtained compared to the classical ETKF in sparse observational networks for small observation intervals $\Delta t_{\rm ons} \leq 6$hrs, and for sufficiently large observational noise. For large observational intervals the ensemble of ETKF will have acquired climatological covariance and our constraint is not needed.

Furthermore we have presented results on a genesis for catastrophic filter divergence in the situation of sparse observations with small observational noise. Machine-infinity blow-up of the forecast model was explained by the finite-size sampling effect of large cross-covariances pushing the analysis off the attractor with subsequent rapid attraction back to the attractor. The associated stiffness causes the numerical integrator to blow-up.

References

[1] G. A. Gottwald, L. Mitchell and S. Reich, *Controlling overestimation of error covariance in ensemble Kalman filters with sparse observations: A variance limiting Kalman filter*, Monthly Weather Review **139** (2011), 2650–2667.

[2] G. A. Gottwald and L. Mitchell *Controlling model error of underdamped forecast models in sparse observational networks using a variance limiting Kalman filter*, QJRMS (in press).

Practical aspects of ensemble-based Kalman filters

Lars Nerger

In real applications, ensemble-based Kalman filters are applied to large-scale numerical models. Due to their high computing cost, one has to operate with very small ensembles of model states. In the application, there is some freedom in the design of the different parts of ensemble Kalman filters. There are different techniques for the initial ensemble sampling, like second-order exact sampling [11] or breeding [12]. In the forecast phase, several possibilities to simulate model error can be included. For the analysis step in which the state estimate is corrected, there are different formulations acting either in the observation space, like the original ensemble Kalman filter (EnKF) [3] or more efficiently in the space spanned by the ensemble, like the ensemble transform Kalman filter (ETKF) [2],

or the singular evolutive interpolated Kalman filter (SEIK) [10]. Finally, there are different possibilities to resample the ensemble to represent the analysis state and its corresponding error covariance matrix estimates. Here, either mean-preserving random transformations or deterministic transformation are possible. The ETKF usually applies a minimum transformation, in which the transformation matrix is closest to the identity matrix in the Frobenius norm. As pure filter formulations fail for large-scale systems, e.g. due to the computational cost that limits the usable ensemble size is limited to $\mathcal{O}(100)$, 'fixes' like inflation [1] or the forgetting factor [10] as well as localisation methods (see [6]) are commonly applied. All these methods need tuning by running several numerical experiments for optimal performance.

To this end, the aim is to find the filter formulation that needs the least computational cost and tuning, while delivering the most accurate state and corresponding error estimates. My talk focuses on the analysis step. Here, the introduction of ensemble square-root Kalman filters (like the ETKF) already led to a significant reduction of the computational complexity. This also holds for the SEIK filter, which computes the state correction in the error subspace spanned by the state ensemble (see [7]) and only recently has been classified as an ensemble square-root Kalman filter [8]. The SEIK filter has a slightly lower computational complexity than the ETKF, which is today one of the most widely used algorithms. However, when a deterministic ensemble transformation is applied, the SEIK filter does not yield the minimum transformation, which is desirable in order to minimise possible violations of physical balances in the analysis model state. In particular, we show that the transformation used with the SEIK filter depends on the order in which the ensemble members are stored in the ensemble matrix. While the statistics (mean and covariance matrix) are correct, the ensemble transformations show small deviations when the ensemble order is changed. This issue is related to inconsistent projections between the state space and the error subspace (see [7]). A new projection is derived, which avoids the dependence on the ensemble ordering. The new algorithm, termed Error Subspace Transform Kalman Filter (ESTKF, [8]) provides the unique minimum transformation, while exhibiting a slightly lower computational cost than the ETKF.

Common choices for the localisation of the analysis step are the covariance localisation [4], which modifies the forecast state error covariance matrix, and observation localisation [5], which modifies the observation error covariance matrix. The localisation effect of both methods is different when the same localisation function is used. Observation localisation results in a wider effective localisation length. The widening is a function of the ratio of the estimated errors int he state and the observations. A new localisation function is derived that results in the same effective localisation length scale for both covariance and observation localisation and improves the performance of the filter with observation localisation. The new regulated localisation function [9] counters the spurious widening of the localisation that would occur in situations, when the ensemble spread estimates a large error in the state estimate, e.g. due to nonlinear dynamics when

some ensemble members develop small-scale features, like ocean eddies at distinct locations.

References

[1] J.L. Anderson, S.L. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Mon. Wea. Rev. **127** (1999) 2741–2758

[2] 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.

[3] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. **99 (C5)** (1994) 10 143–10 162.

[4] P. L. Houtekamer and H. L. Mitchell, *Data assimilation using an Ensemble Kalman Filter technique*, Mon. Wea. Rev. **126** (1998) 796–811.

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

[6] T. Janjić, L. Nerger, A. Albertella, J. Schröter, and S. Skachko, *On domain localization in ensemble based Kalman filter algorithms*, Mon. Wea. Rev. **139** (2011) 2046–2060.

[7] L. Nerger, W. Hiller, and J. Schröter, *A comparison of error subspace Kalman filters*, Tellus **57A** (2005) 715–735.

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

[9] L. Nerger, T. Janjić, J. Schröter, and W. Hiller, *A regulated localization scheme for ensemble-based kalman filters*, Q. J. Roy. Meteor. Soc. **138** (2012) 802–812.

[10] D. T. Pham, J. Verron, and L. Gourdeau, *Singular evolutive Kalman filters for data assimilation in oceanography*, C. R. Acad. Sci., Ser. II **326 (4)** (1998) 255–260.

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

[12] X. Wang and C. H. Bishop, *A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes*, J. Atm. Sci. **60** (2003) 1140–1158

Preservation of physical properties with ensemble based Kalman filter algorithms

Tijana Janjic

(joint work with Dennis B. McLaughlin, Stephen E. Cohn)

One of the principles used in developing discretization schemes for geophysical fluid dynamics is to maintain conservation properties that characterize the flow. The most basic of these integral constraints is the conservation of total mass. In this work we focus on ensemble-based sequential data assimilation algorithms and show that even for linear dynamics preservation of the physical properties can not be guaranteed. For example, total mass is changed for positive scalar quantity if an estimate (analysis) is produced that has negative values. The new approach is developed that ensures conservation of two basic properties during data assimilation, the positivity and total mass, without resorting to a posteriori adjustments. We examine the extent to which imposing mass conservation requirements in data assimilation algorithms changes the assimilation results and demonstrate the impact of imposing mass conservation on prediction in simple experiments.

1. Introduction

In this talk we will show that if the physical conditions, that characterize the underlying geophysical fluid dynamics, are included as additional constraints within ensemble based Kalman filter algorithm, we will be a) able to preserve some basic physical properties of the estimate, a prerequisite for a valid model initial condition and b) improve predictions in case of perfect and non perfect model experiments. Constraints that we will be imposing on the estimate are both equality and inequality constraints and present the important source of information about the system. Two properties that we chose in this paper to preserve during the data assimilation are positivity and the total mass. Both of these properties have been important for design of numerical weather prediction schemes [Arakawa(1972), Arakawa and Lamb(1977), Lin and Rood(1996)] and have not been discussed thus far as such for analysis updates.

In order to preserve both mass and positivity we introduce a new ensemble based Kalman filter algorithm. The new algorithm imposes strictly equality and inequality constraints for each ensemble member. In order to show possible benefits of the new algorithm on prediction we consider linear dynamics and the impact of imposing conservation constraints on prediction within solid body rotation setup, an often used example for testing numerical schemes.

2. Problem Formulation

The dynamics that we consider will not have any source and sink terms, as well as no fluxes through the boundaries. Total mass will not be considered a stochastic quantity. The change in mass (if it happens) will happen in analysis or resampling steps of ensemble based Kalman filter algorithm. The ensemble transform Kalman filter (ETKF) algorithm will be considered as a prototype of square root ensemble Kalman filters and will be used for comparison. In the rest of what follows, we will assume that our numerical model, not necessarily perfect, will have a numerical scheme that conserves global integral of quantity to be estimated $\mathbf{w}$.

We require that the error covariance function $P^a(x_1, x_2, t)$ of continuous version, $w^a(x,t) =< w|\mathbf{w}^o >$ of an estimate $\mathbf{w}_k^a$

$$P^a(x_1, x_2, t) \equiv \langle (w(x_1,t) - w^a(x_1,t))(w(x_2,t) - w^a(x_2,t)) \rangle \tag{1}$$

satisfy

$$\int_D P^a(x_1, x_2, t)\, dx_1 = 0. \tag{2}$$

Such an error covariance will be called **mass conserving**. In discrete case, the analysis error covariance matrix $\mathbf{P}_k^a$ will be **mass conserving** if

$$\mathbf{P}_k^a \mathbf{e} = 0 \tag{3}$$

and $\mathbf{e} = \mathbf{e}_{n\times 1} = [1\,1\,\ldots 1]^T$. The form of $\mathbf{e}$ is chosen for simplicity. The exact form of the definition (3) will depend on the grid of our numerical model and quadrature chosen in (2).

We show that ETKF algorithm conserves mass as defined above if no localization is applied. In case localization is applied on the ensemble derived error covariance, by Schur multiplication of ensemble derived covariance and pre chosen covariance matrix, mass will not be conserved. In this case, following simple transformation $\mathbf{I} - \mathbf{e}\mathbf{e}^T/n$ can be applied to localized covariance in order to make it mass conserving.

3. Problem Solution

The ETKF equations can be derived by minimization of the cost function that takes into the account inverse of ensemble derived forecast error covariance and the observation error. Additional added constraints to the cost function will ensure analysis that satisfies physical conditions needed. To ensure that each ensemble member has desired properties same algorithm will be used for every ensemble member. Therefore each member of the analysis ensemble is constructed from the corresponding forecast ensemble member, perturbed observations and the analysis error covariance. The minimization with constraints can be done using the quadratic programming algorithm for both equality and inequality constraints as strong constraints. This method will be called QPEns.

In order to estimate the effect through time and on the prediction, we use two dimensional solid body rotation experiment. The initial ensemble was specified in the form of a cone with the basis with the radius of 100 km, and the height of 100 arbitrary units. Eight member ensemble is generated around the solution by uniformly perturbing the central location of the cone. The initial forecast is then computed as average over ensemble members, producing the field with maximum value of 42.7 and minimum value of zero. We note that although every ensemble member is the perfect cone with the desired properties, due to the averaging over the ensemble with uncertainty in the location of the disturbance, the structure of the cone is not seen in the initial conditions. Seven analysis were performed with observations that are true solution perturbed by normally distributed noise.

The computational dispersion even for numerically conservative schemes will produce field that does not perfectly maintain the initial condition. We will use the second order accurate scheme for the advection of the cone [Janjić et al.(2011)]. The RMS error after second full revolution is 4.1, minimum and maximum value are -14 and 81 without any data assimilation. The ETKF algorithm would slow down the drift of the numerical model, but would not be able to correct the negative values, leading to the solution with the minimum -11, maximum 78 and RMS value of 2.54 after second rotation. Even in the case of imperfect model, the QPEns algorithm produces positive and mass conservative analysis and in this way significantly improves the forecast result. The experimental results show that even in the case of non perfect model experiments incorporating the constraints on positivity and mass significantly reduces the error of the analysis. Therefore, if appropriate data assimilation scheme is chosen, the drift of the numerical model due to numerical errors will be reduced, indicating the importance of the constraints that we are imposing for prediction.

References

[Arakawa(1972)] Arakawa, A., 1972: Design of the UCLA general circulation model. Technical report, Department of Meteorology, University of California, Los Angeles, tech. Report No. 7.

[Arakawa and Lamb(1977)] Arakawa, A. and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods of Computational Physics*, **17**, 173–265.

[Janjić et al.(2011)] Janjić, Z., T. Janjić, and R. Vasić, 2011: A class of conservative fourth-order advection schemes and impact of enhanced formal accuracy on extended-range forecasts. *Mon. Wea. Rev.*, **139**, 1556–1568, doi:10.1175/2010MWR3448.1.

[Lin and Rood(1996)] Lin, S.-J. and R. B. Rood, 1996: Multidimensional flux-form semi-lagrangian transport schemes. *Mon. Wea. Rev*, **124**, 2046–2070.

Finite Number of Determining Parameters for the Navier-Stokes Equations with Applications in Data Assimilation and Feedback Control

Edriss S. Titi

(joint work with Abderrahim Azouani and Eric Oslon)

Dissipative evolutionary equations, such as the Navier-Stokes equations and certain reaction-diffusion systems, are known to posses finite number of determining parameters and finite-dimensional global attractors (cf. [3, 4, 6, 7, 10, 11, 12, 15] and references therein). In [9, 13, 14] (see also [8]) we take advantage of this finite-dimensional long-term dynamics and use the notion of finite number of determining modes to design an algorithm to recover the unknown infinitely many high Fourier modes (fine scales) of the exact solution of the Navier-Stokes equations from the given finitely many low Fourier modes (coarse grid measurements). This algorithm fits within the scope and strategy of data assimilation. Recently, we have extended this result and designed a new algorithm (cf. [2]) for recovering the exact solutions of the Navier-Stokes equations from their nodal values on finite, but fine enough, spatial grid. Our algorithm is general enough to cover the most general approximate interpolants, in the spirit of finitely many determining parameters/prjections as it is reported in [3, 7, 10, 11, 12] and references therein. Our algorithm also provides a finite-dimensional feedback control strategy to stabilize solutions of the Navier-Stokes equations and reaction-diffusion systems (cf. [1]).

On the more theoretical dynamical systems aspect of dissipative evolution equations, the determining modes for the two-dimensional incompressible Navier-Stokes equations can also be shown (cf. [5]) to satisfy an ordinary differential equation of the form $dv/dt = F(v)$, in the Banach space, X, of all bounded continuous functions of the variable $s \in \mathbb{R}$ with values in certain finite-dimensional linear space. This new evolution ODE, named *determining form*, induces an infinite-dimensional dynamical system in the space X which is noteworthy for two reasons. One is that F is globally Lipschitz from X into itself. The other is that the long-term dynamics of the determining form contains that of the Navier-Stokes equations; the traveling wave solutions of the determining form, i.e., those of the form $v(t, s) = v_0(t + s)$,

correspond exactly to initial data v_0 that are projections of solutions of the global attractor of the Navier-Stokes equations onto the determining modes. The determining form is also shown (cf. [5]) to be dissipative; an estimate for the radius of an absorbing ball is derived in terms of the number of determining modes and the Grashof number (a dimensionless physical parameter which behaves like the square of Reynolds number). Finally, a unified approach inspired by the algorithm given in [2, 1] is outlined for an ODE satisfied by a variety of other determining parameters such as nodal values, finite volumes, and finite elements.

References

[1] A. Azouani and E.S. Titi, *Feedback control of nonlinear dissipative systems by finite determining parameters - a reaction-diffusion paradigm*, (Preprint).

[2] A. Azouani, E. Olson and E.S. Titi, *Continuous data assimilation using general interpolant observables*, (Preprint).

[3] B. Cockburn, D.A., Jones, E.S., Titi, *Estimating the number of asymptotic degrees of freedom for nonlinear dissipative systems*, Math. Comput. **66** (1997), 1073–1087.

[4] P. Constantin and C. Foias, *Navier-Stokes Equations*, Univ. of Chicago Press (1988).

[5] C. Foias, M. Jolly, R. Kravchenko and E.S. Titi, *A determining form for the 2D Navier-Stokes equations - the Fourier modes case*, J. Math. Phys., **53** (2012), 115623.

[6] C. Foias, O. Manley, R. Rosa, R. Temam, *Navier–Stokes Equations and Turbulence*, Encyclopedia of Mathematics and Its Applications 83, Cambridge University Press, 2001.

[7] C. Foias and E.S. Titi, *Determining nodes, finite difference schemes and inertial manifolds*, Nonlinearity **4** (1991), 135–153.

[8] W.D. Henshaw, H.O. Kreiss, J. Yström, *Numerical experiments on the interaction between the large and small-scale motions of the Navier–Stokes equations*, Multiscale Model. Sim. **1:1** (2003), 119–149.

[9] K. Hayden, E. Olson and E.S. Titi, *Discrete data assimilation in the Lorenz and* 2*D Navier-Stokes equations*, Physica D **240** (2011), 1416–1425.

[10] D. Jones and E.S. Titi, *On the number of determining nodes for the 2-D Navier-Stokes equations*, J. Math. Anal. Appl. **168:1** (1992), 72–88.

[11] D. Jones and E.S. Titi, *Determining finite volume elements for the 2-D Navier–Stokes equations*, Physica D **60** (1992), 165–174.

[12] D. Jones and E.S. Titi, *Upper bounds on the number of determining modes, nodes, and volume elements for the Navier-Stokes equations*, Indiana U. Math. J., **42** (1993), 875–887.

[13] E. Olson and E.S. Titi, *Determining modes for continuous data assimilation in* $2-D$ *turbulence*, J. Stat. Phys. **113** (2003), 799–840.

[14] E. Olson and E.S. Titi, *Determining modes and Grashof number in 2D turbulence: A numerical case study*, Theor. Comp. Fluid Dyn. **22(5)** (2008), 327–339.

[15] R. Temam, *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, New York: Springer (2nd augmented edition, 1997).

Accounting for model error in data assimilation

Alberto Carrassi

(joint work with Stephane Vannitsem)

The prediction problem in geophysical fluid dynamics typically relies on two complementary elements: the model and the data. The mathematical model, and its discretized version, embodies our knowledge about the laws governing the system

evolution, while the data are samples of the system's state. They give complementary information on the same object. The sequence of operations that merges model and data to obtain a possibly improved estimate of the flow's state is usually known as data assimilation. The physical and dynamical complexity of geophysical systems makes the data assimilation problem particularly involved.

The different information entering the data assimilation procedure, usually the model, the data and a background field representing the state estimate prior to the assimilation of new observations, are weighted according to their respective accuracy. Data assimilation in geophysics, particularly in numerical weather prediction, has experienced a long and fruitful stream of research in the last decades which has led to a number of advanced methods able to take full advantages of the increasing amount of available observations and to efficiently track and reduce the dynamical instabilities. As a result the overall accuracy of the Earth's system estimate and prediction, particularly the atmosphere, has improved dramatically.

Despite this trend of improvement, the treatment of model error in data assimilation procedures is still, in most instances, done following simple assumptions such as the absence of time correlation [7]. The lack of attention on model error is in part justified by the fact that on the time scale of numerical weather prediction, where most of the geophysical data assimilation advancements have occurred, its influence is reasonably considered small as compared to the initial condition error that grows in view of the chaotic nature of the dynamics. Nevertheless, the improvement in data assimilation techniques and observational networks on the one hand, and the recent grow of interest in seasonal-to-decadal prediction on the other, has placed model error, and its treatment in data assimilation, as a main concern and a key priority. A number of studies have appeared, in the context of sequential and variational schemes (see e.g. [6, 9, 8]).

Two main obstacles are the huge size of the geophysical models and the wide range of possible model error sources. The former problem implies the need to estimate large error covariance matrices on the basis of a limited amount of available observations. Then, given that modelling errors can arise from a number of causes including incorrect parametrisation, numerical discretization, the lack of description of some relevant scale of motion, it is difficult to outline a general model error dynamical framework.

In the present contribution we describe a new approach, referred to as deterministic model error treatment, in which the deterministic evolution of the model error is described based on a short-time approximation suitable for realistic applications and used to estimate the model error contribution in the state estimate, *i.e.* based on covariances and correlations. We have distinguished two situations: first assuming that model error originates only from a misspecification of the parameters, and second from the presence of unresolved scales. This deterministic formulation has been applied in the context of sequential and variational approach, for state and parameter estimation. The duration of the short-time approximation depends on the type and size of the model error. In the case of uncertain parameters, it scales with the parametric error but also with the functional dependence of the

model dynamics on the uncertain parameters. When the error comes from the presence of unresolved scales, the short-time model error evolution in the resolved scales depends on the difference between the truth (unknown) velocity field and the modeled one: a solution to estimate this term in practice has been proposed [3].

It is the difference between the truth and modelled advection fields that regulates its evolution when model error comes from the presence of unresolved scales.

The deterministic model error treatment has been proven competitive in a number of different applications with prototypical chaotic dynamics, in the framework of sequential [1, 3] and variational schemes [2] as well as for parameter estimation [4, 5]. Research is currently undergone for the application of this approach in soil data assimilation where the soil temperature and moisture content is estimated on the basis of atmospheric observations close to the ground.

References

[1] Carrassi, A., S. Vannitsem, and C. Nicolis, *Model error and sequential data assimilation: A deterministic formulation*, Quart. J. Roy. Meteor. Soc., **134** (2008), 1297–1313.

[2] Carrassi A. and S. Vannitsem, *Accounting for Model Error in Variational Data Assimilation: A Deterministic Formulation*, Mon. Wea. Rev., **138** (2010), 3369–3386.

[3] Carrassi, A. and S. Vannitsem, *Treatment of the error due to the unresolved scales in sequential data assimilation*, International Journal of Bifurcation and Chaos, **21** (2011a), 3619–3626.

[4] Carrassi, A. and S. Vannitsem, *State and parameter estimation with the extended Kalman filter: An alternative formulation of the model error dynamics*, Q. J. Roy. Meteorol. Soc., **137** (2011b), 435–451.

[5] Carrassi, A., R. Hamdi, P. Termonia and S. Vannitsem, *Short time augmented extended Kalman filter for soil analysis: a feasibility study*, Atmos. Sci. Let. (2012), DOI: 10.1002/asl.394

[6] Dee, D. and A. da Silva, *Data assimilation in the presence of forecast bias*, Q. J. R. Meteorol. Soc., **117** (1998), 269–295.

[7] Jazwinski, A.H., *Stochastic Processes and Filtering Theory*, (Academic Press) (1970), New York.

[8] Kondrashov D, Sun C, Ghil M., *Data assimilation for a coupled ocean-atmosphere model. Part II: Parameter estimation*, Mon. Weather Rev.,**136** (2008), 5062–5076.

[9] Trémolet, Y., *Model-error estimation in 4D-Var*, Q. J. Roy. Meteorol. Soc., **133** (2007), 1267–128

Bayesian inference and data assimilation with optimal maps

Youssef M. Marzouk

(joint work with Tarek A. El Moselhy)

Predictive simulation of complex physical systems increasingly rests on the interplay of experimental observations with computational models. Indeed, the assimilation of observational data has become an essential task in fields ranging from weather prediction to subsurface modeling. In this context, Bayesian statistics provides a natural framework for quantifying uncertainty in parameter estimates

and model predictions, for fusing heterogeneous sources of information, and for optimally selecting experiments or observations. Yet the computational expense and convergence issues associated with rigorous Bayesian methods present significant bottlenecks in large-scale problems.

We present a new approach to Bayesian inference that entirely avoids Markov chain simulation or sequential importance resampling, by constructing a map that pushes forward the prior measure to the posterior measure [1]. Consider a *prior* probability measure μ on $\mathbb{R}^n$ and a *posterior* probability measure ν on $\mathbb{R}^n$, related according to a Bayesian inference problem where the likelihood function is proportional to $d\nu/d\mu$. Then we seek a deterministic coupling or *map* $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $\nu = f_\sharp \mu$. For simplicity, suppose that both the prior and posterior have densities with respect to Lebesgue measure, denoted by p and π^d/β, respectively. In most inference problems of interest, the normalizing constant β is not known. We define $T(f) \equiv \log\left(\pi^d \circ f\right) + \log|\det Df| - \log p$ and show that satisfying the measure transformation condition is equivalent to finding a monotone f such that T is a constant. It is straightforward to show that the value of the constant thus attained is in fact $\log\beta$.

Now consider the problem

$$\min_f \mathbb{E}\,\|X - f(X)\|^2 \text{ s.t. } \nu = f_\sharp\mu \tag{1}$$

with $X \sim \mu$. The seminal work of Brenier and McCann (see [2]) shows that, under relatively weak conditions on μ and ν, this problem has a solution f that is uniquely determined μ-a.e and is the gradient of a convex scalar function. We relax this problem somewhat and instead solve

$$\min_f \left\{\mathbf{V}T(f;X) + \lambda\mathbb{E}\,\|X - f(X)\|^2\right\}, \tag{2}$$

with appropriate control on λ. Alternatively, we can look for a Knothe-Rosenblatt re-arrangement by restricting the ith component of the map to depend only on the first i inputs, $z_i = f_i(x_1, \ldots, x_i)$, $i = 1 \ldots n$; we say this map has "triangular" structure. Then the optimization problem becomes:

$$\min_f \mathbf{V}T(f;X) \text{ s.t. } f \text{ has triangular structure.} \tag{3}$$

Additional ways of formulating the optimization problem are given in [1].

These optimization problems are the point of departure for numerical schemes. Note that $\mathbf{V}[T]$ involves an expectation with respect to the prior measure, which we can approximate using a fixed set of samples, thus addressing the stochastic optimization problem via a *sample average approximation* (SAA) approach. Sample average approximation allows one to apply the full machinery of deterministic optimization; in particular, we employ gradient information from the forward model in a quasi-Newton method.

Another numerical issue centers on how to represent the map f. There is a great deal of latitude in the choice of representation, but in our current work we choose to parameterize the map using polynomials orthogonal with respect to the reference (prior) measure, so that posterior moments may be computed analytically from

the polynomial coefficients. A polynomial representation also allows for a natural enrichment strategy, where first we solve the minimization problem using a linear map, then enrich as needed to polynomials of degree three, and so on—monitoring the distance of $\mathbf{V}[T]$ from zero to assess convergence.

The resulting inference scheme overcomes many of the computational bottlenecks associated with Markov chain Monte Carlo (MCMC). With a map in hand, one can generate arbitrary numbers of independent and uniformly weighted posterior samples without additional posterior or model evaluations. The approach also provides clear convergence criteria for posterior approximation, and facilitates model selection through automatic evaluation of the marginal likelihood. Comparisons on elliptic inverse problems and inference in spatial statistical models show meaningful gains in performance and accuracy over current state-of-the-art MCMC methods.

We also discuss extensions of the map to problems of sequential data assimilation, i.e., filtering and smoothing. First, we demonstrate how a sequence of low-order maps may be *composed* to capture the transition from prior to posterior; this construction allows a complex change of measure to be captured more economically than with a single map. Then we discuss filtering with nontrivial transition dynamics, e.g., some nonlinear dynamical model (potentially stochastic) giving rise to a state transition distribution $\mu_{t+1|t}(A) = \mathbb{P}\left(X_{t+1} \in A | X_t\right)$. For the latter problems, one scheme involves pushing forward a fixed reference measure to each filtered state distribution, while an alternative scheme computes maps that push forward the filtering distribution from one stage to the other. The "reference measure" approach is generally superior, though the essential challenge now involves approximating the prior distribution at each assimilation step. To this end, we propose a scheme that involves a nonlinear mapping from the forecast distribution (e.g., with density $p(x_{t+1}|y_{1:t})$) to a reference Gaussian, then performing the Bayesian update on this Gaussian space.

References

[1] T. Moselhy, Y. Marzouk, *Bayesian inference with optimal maps*, Journal of Computational Physics **231** (2012), 7815–7850.

[2] C. Villani, *Optimal Transport: Old and New*, Springer (2009).

Statistical modelling in meteorological DA beyond the normal, some ideas

Andreas Hense
(joint work with Jessica Keune, Andreas Röpnack, Daniel Simonis, Christoph Gebhardt, Tanja Zerenner)

1. Introduction

The aim of the presentation is to present three topic to the meteorological data assimilation community which are studied in other areas of meteorological and climatological research. These points could be of potential interest in data assimilation applications. The following three topics will be covered

(1) Verification, ensemble postprocessing, proper scores and optimisation
(2) Paleo climate reconstruction from binary information of presence/absence of vegetation
(3) Sparse (inverse, non-singular) covariance matrices

We will shortly describe these topics and summarise the presentation with suggestions how the results could be applied in data assimilation (DA).

2. Verification of probabilistic forecasts

Running a meteorological or climate ensemble system no matter if this is an ensemble based data assimilation system or a system for free runs unaffected by observations, the outcome has to be interpreted as a statistical sample drawn from the ergodic measure of the underlying model attractor. A necessary condition is that the initial conditions are sufficiently independent. To obtain an estimate of the probability density a postprocessing is inevitable. This even holds when working with the original, raw simulations. A nonparametric order statistics is implied when univariate realisations e.g. of the forecasted temperature at one point in space and time are sorted and quantiles are estimated. But also methods of fitting Gaussian densities of univariate and multivariate structures can be done with the information on the location parameter mean and uncertainty parameter (co-)variance. Note that the presentation of the ensemble mean is not a sufficient information because the mean is not a realisation ("weather"). The fitted density is referred to as the predictive probability density function (pdf).

This is the information which has to be verified by observations.Note that the assimilation step in sequential ensemble data assimilation approach is essentially also such a confrontation of a predictive density produced by the forward model with the newly available observations A review paper by [1] shows that so called proper scores form the basis of such a verification. The definition of a **proper score** is based

- on a **score function** or **scoring rule** which evaluates a specific predictive density or distribution $f_M(m)$ in view of the data b as $s(f_M(m), b)$

- Then taking the expectation with respect to B defines the **score** $S(f_M(m), f_B(b)) = E_B[s(f_M(m), b)]$, which is also called a generalised entropy
- the score and its associate score function is called **proper** if $S(f_M(m), f_B(b)) \leq S(f_B(b), f_B(b))$ meaning that the highest values of the score is obtained if the predicted probability density is equal to the (unknown) probability density of the observations.
- here we have agreed on the convention "the higher the better" which is a reasonable choice if the prediction is coupled to the income of the predictor such that
- a proper score can not be cheated by freak forecasts to obtain for wrong reasons a higher income

This convexity allows to optimize the (estimated) score if parameters of the predictive pdf can be identified such that the relation between the predictive pdf and the observations can be improved.

3. Paleo climate reconstruction

Paleo climate reconstruction of e.g. near surface temperature is often forced to rely on highly non-Gaussian variables like binary data of the absence/presence of climate sensitive plant species. One way of using such a type of data is based on a probabilistic interpretation of the binary data e.g. in terms of conditional probability densities ([2]). This allows the construction of a likelihood function for a second step: the spatial aggregation of data from different paleo proxy sites provided that they are dated to have occurred at approximately the same time in the past. In [3] a simple two dimensional advection diffusion model is used as a weak constraint to regularise the pure vegetation based spatial aggregation. The actual reconstruction is done in the modal space of the singular modes of the simplified model. This defines a spatial filter which makes the spatial reconstruction compatible to the scales which are simulated by global climate models.

4. Precision matrices

Discrete Gaussian Markov random fields can be interpreted as a Gaussian random process on a graph of spatially distributed vertices connected edges or grid lines. E.g. in [4] it is presented that such Gaussian random fields are characterised by their expectation and more important by the invers covariance matrix the precision matrix. This is a direct consequence of the Hammersley - Clifford theorem which establishes a relation between Gaussian Markov random fields and Gibbs measures. Therefore the hotly debated localisation of the covariance matrix should actually be done on the precision matrix. The precision matrix itself can be estimated from the data by various recently developed methods. One of the new methods is the GLASSO estimator. The standard maximum likelihood estimator (MLE) for the precision matrix is extended by an L_1 penalty term which selects the sparsest precision matrix which inverse is best compatible with the standard MLE covariance matrix.

5. Summary

Based on the results related to the above three topics the following points summarise the suggestion to the DA community

- Verification of predictive densities also for the multivariate case is possible using proper scores
- Suggestion to DA: Apply proper scores to ensemble based sequential filters and use the option for optimisation
- Paleo climate research forces the use of highly non-Gaussian variables. Use a probabilistic descriptions
- Suggestion to DA: assimilation of binary data is possible
- The relevant matrix in treating Gaussian random fields is the precision matrix and not the covariance matrix
- Suggestion to DA: do localisation on the precision matrix.

References

[1] Gneiting T., Raftery A. E. *Strictly Proper Scoring Rules, Prediction, and Estimation*, Journal of the American Statistical Association, 102, 359-378 (2007)

[2] Kühl N, Gebhardt C, Litt T., Hense A., *Probability Density Functions as Botanical-Climatological Transfer Functions for Climate Reconstruction*, Quaternary Research 58, 381–392 (2002)

[3] Simonis D., Hense A., Litt T. *Reconstruction of late Glacial and Early Holocene near surface temperature anomalies in Europe and their statistical interpretation* Quaternary International, 274, 233-250 (2012)

[4] Wainwright M.J. and Jordan M.I.; *Graphical models, Exponential Families and Variational inference* Foundations and Trends in Machine Learning,1, 1–2, (2008) 305pp

Nonstationarity and nonhomogeneity in GFD: data analysis, model discrimination beyond the standard probabilistic framework and missing data assimilation

Illia Horenko

Results of mathematical/statistical analysis and assimilation of the weather/climate data are inherently biased by the implicit assumptions about the underlying processes. These assumptions are imposed by the respective analysis methods: e.g., Bayesian learning approaches impose the strong implicit probabilistic assumptions like Gaussianity (for Gaussian Mixture Models) and homogenous Markovianity (for Hidden Markov Models and related approaches). More generally, all of the standard learning approaches involve in some form implicit assumptions about the memory in the data, about the probabilistic model formulation (usually in form of the parametric probability distribution function or the parametric stochastic model), some form of sequential statistical independence (in time and space) and stationarity/homogeneity assumptions (i.e., the assumption that the inferred probabilistic model parameters, e.g. the mean value and the variance, are not changing significantly for the whole available data set).

Generic scenarios where these implicit assumptions of available methods may become violated will be discussed and some alternative approaches to numerical data modeling and model discrimination beyond the standard probabilistic framework will be described. Computational framework for nonstationary identification and assimilation of the joint missing data impact on resolved variables will be given and compared to the popular methods of machine learning on generic data examples. The comparison will be given in terms of: (i) the character of mathematical/statistical assumptions involved, (ii) numerical/computational cost, and (iii) the quality of the obtained online predictions.

References

[1] Ph. Metzner, L. Putzig, I. Horenko *Analysis of persistent non-stationary time series and applications*, Communications in Applied Mathematics and Computational Science **7**(2) (2012), 175–229.

[2] J. de Wiljes, L. Putzig, I. Horenko, *Discrete non-homogenous and non-stationary logistic and Markov regression models for spatio-temporal data with unresolved external influences*, Communications in Applied Mathematics and Computational Science, submitted (2012).

Stability of the optimal filter - a variational approach

Wilhelm Stannat

We consider the problem of estimating a Markovian signal (X_t) on $\mathbb{R}^d$, in discrete or continuous time, observed with independent additive noise. Denote by (Y_t) the observation process, so that

$$\pi_t^{Y_{0:t}}(A) := P\left[X_t \in A \mid Y_{0:t}\right]$$

is the conditional distribution of the state X_t of the signal, also called the posterior distribution in the following, based on the observations $Y_{0:t} = \{Y_s \mid s \in [0,t]\}$ up to time t. We are interested in the stability of $\pi_t^{Y_{0:t}}$ w.r.t. the initial distribution μ_0 of the signal. This problem is of central importance in real applications, since the true state of the unobserved signal, let alone its true distribution, are unknown. Because of its importance, the problem has attracted the attention of many researchers during the last decade and a detailed account on the state of the art can be found in Part III of the monograph [1].

Roughly speaking the stability properties of $\pi_t^{Y_{0:t}}$ depend on the one hand on the mixing properties of the signal (X_t) and on the other hand on the precision of the measurement leading to the observations (Y_t). Note that the two mechanisms have opposite effects on the posterior. Whereas the mixing properties of the signal distribute the mass throughout the whole state space, the measurement has the tendency to concentrate the mass of the posterior distribution in some neighborhood of the observation. It is therefore in particular difficult to quantify the stability of $\pi_t^{Y_{0:t}}$. With our variational approach, outlined below, we can provide a way to measure stability of $\pi_t^{Y_{0:t}}$ w.r.t. the initial condition explicitely in the total variational distance taking into account both ingredients.

Let us first illustrate the approach in discrete time for the Kalman filter of a linear signal and linear observation

$$X_n = BX_{n-1} + CW_{n-1}\,, \qquad Y_n = GX_n + \Gamma e_n$$

with (W_n) (resp. (e_n)) independent $N(0,I)$-distributed on $\mathbb{R}^d$ (resp. $\mathbb{R}^p$) and matrices B, C , G and Γ with appropriate dimensions. The main feature of the optimal filter in this special case is the well-known fact that a Gaussian initial condition $\mu_0 = N(m_0, E_0)$ implies that $\pi_n^{Y_{0:n}} = N\left(m_{n+1}^{Y_{1:n+1}}, E_{n+1,n+1}\right)$, where the mean and the covariance matrix are determined by the well-known recursive Kalman filtering equations. However, if the initial distribution of the true signal, denoted by $\tilde{\mu}_0$ in the following, is not Gaussian, one would like to estimate the error between the optimal filter $\tilde{\pi}_n^{Y_{0:n}}$ of the true signal and its Kalman approximation $\pi_n^{Y_{0:n}} = N\left(m_{n+1}^{Y_{1:n+1}}, E_{n+1,n+1}\right)$.

To answer this question, suppose that the true initial distribution $\tilde{\mu}_0$ is absolutely continuous w.r.t. the Gaussian distribution, $d\tilde{\mu}_0 = h_0\, d\mu_0$. This implies that $d\tilde{\pi}_n^{Y_{0:n}} = h_n\, d\pi_n^{Y_{0:n}}$, where the density h_n is defined recursively as

$$h_n(x) = \int h_{n-1}\left(Q_n^*\left(B^TQ^{-1}x + E_{n_1,n-1}m_{n-1}^{Y_{1:n}}\right) + x'\right)\, N(0,Q_n^*)(dx')$$

with $Q_n^* + \left(E_{n-1,n-1}^{-1} + B^TQ^{-1}B\right)^{-1}$. This implies immediately the following recursive control on the density

$$\|h_n\|_{\mathrm{Lip}} \le \|Q_n^*B^TQ^{-1}\|_{\mathrm{op}}\|h_{n-1}\|_{\mathrm{Lip}}$$

w.r.t. the Lipschitz norm $\|f\|_{\mathrm{Lip}} = \sup_{x_1\neq x_2}\frac{|f(x_1)-f(x_2)|}{\|x_1-x_2\|}$. Here, $\|\cdot\|_{\mathrm{op}}$ denotes the usual operator norm. From this one can deduce the following results on asymptotic stability of the optimal filter in the total variational norm $\|\cdot\|_{\mathrm{tv}}$ (cf. [4]):

Theorem 0.1. *Under the above assumptions:*

$$\|\pi_n^{Y_{0:n}} - \tilde{\pi}_n^{Y_{0:n}}\|_{tv} \le \frac{1+tr(Q_n^*)}{2\int h_n\, d\pi_n^{Y_{1:n}}}\left(\prod_{k=1}^{n}\|Q_n^*B^TQ^{-1}\|_{op}\right)\|h_0\|_{Lip}\,, n = 1,2,\ldots$$

Corollary 0.2. *Assume in addition that (C,B) is controllable and (G,B) observable. Then*

$$\lim_{n\to\infty}\frac{1}{n}\log\|\pi_n^{Y_{0:n}} - \tilde{\pi}_n^{Y_{0:n}}\|_{tv} \le \log\chi_* < 0\,,$$

where $\chi_* = \max\{|\lambda| \mid \lambda \text{ eigenvalue of } B(I-K_\infty G)\}$ and

$$K_\infty = \lim_{n\to\infty}E_{n,n-1}G^T\left(GE_{n,n-1}G^T + \Gamma\Gamma^T\right)^{-1}$$

is the limiting Kalman gain matrix.

The main idea for a generalization of the above theorem to nonlinear signals is to replace the Gaussian distribution by the family of uniformly strictly log-concave

probability measures $d\mu = e^{-V}\,dx$ and to use the following Brascamp-Lieb inequality

$$\mathrm{Var}_{e^{-V}\,dx}(h) \le \int \langle \nabla h, V_{xx}^{-1}\nabla h\rangle\, e^{-V}\,dx$$

to quantify the stability properties of the posterior distribution.

As a typical example in continuous time consider a signal driven by the following stochastic differential equation

$$\text{(S)} \qquad dX_t = B(X_t)\,dt + C dW_t \qquad \text{on } \mathbb{R}^d$$

where the drift B admits the representation

$$B(x) = B_1 x + Q\nabla \log\varphi(x) \qquad \text{with } Q = CC^T > 0$$

for some matrix B_1 and some smooth polynomially bounded potential $\varphi > 0$. Here (W_t) is a d-dimensional Brownian motion. Suppose that (X_t) is observed with independent additive noise

$$\text{(O)} \qquad dY_t = GX_t\,dt + \Gamma de_t\,, \quad Y_0 = 0 \qquad \text{on } \mathbb{R}^p\,,$$

where (e_t) is a p-dimensional Brownian motion and suppose that $R := \Gamma\Gamma^T > 0$. The Kallianpur-Striebel formula provides the following path-space representation

$$\int F\,d\pi_t^{Y_{0:t}} = \frac{E\left[F(X_t)\exp\left(\int_0^t R^{-1}GX_s dY_s - \frac{1}{2}\int_0^t \langle R^{-1}GX_s, GX_s\rangle\,ds\right)\right]}{E\left[\exp\left(\int_0^t R^{-1}GX_s dY_s - \frac{1}{2}\int_0^t \langle R^{-1}GX_s, GX_s\rangle\,ds\right)\right]}$$

of the conditional expectation $E\left[F(X_t)\mid Y_{0:t}\right]$ and a robust regular conditional distribution $\pi_t^{y_{0:t}}$ of $\pi_t^{Y_{0:t}}$ can be constructed for Hölder-continuous paths $y.$ with Hölder-exponent $\frac{1}{3}$. The stability result can then be formulated as follows (see [4]):

Theorem 0.3. *Assume that*

$$\text{(1)} \qquad V(x) = \langle \nabla\log\varphi(x), B_1 x\rangle + \frac{1}{2}\frac{tr(Q\varphi_{xx}(x))}{\varphi(x)} + \frac{1}{2}\langle R^{-1}Gx, Gx\rangle$$

is uniformly strictly convex with $V_{xx} \ge K$ for some positive definite symmetric matrix K. Let $d\mu_0 = \left(\int \hat m_0\,dx\right)^{-1}\hat m_0\,dx$ with $\varphi^{-1}\hat m$ uniformly log-concave with lower bound $> K_*$, where K_* is the unique positive definite symmetric solution of

$$K_* Q K_* + B_1^T K_* + K_* B_1 - K = 0\,.$$

Suppose that the initial condition $\tilde\mu_0$ of the true signal has a Lipschitz continuous density h_0 w.r.t. μ_0, bounded away from zero by $\delta > 0$. Then

$$\|\pi_t^{y_{0:t}} - \tilde\pi_t^{y_{0:t}}\|_{tv} \le \frac{1 + tr(\sqrt{Q^{-1}}K_*^{-1}\sqrt{Q^{-1}})}{2\delta} e^{-\chi_* t}\|h_0\|_{Lip_{\sqrt{Q^{-1}}}}$$

for any Hölder-continuous path $y.$ with Hölder-exponent $\frac{1}{3}$. Here,

$$\chi_* := \min\left\{\lambda \mid \lambda \textit{ eigenvalue of } \sqrt{Q}K_*\sqrt{Q} + \frac{1}{2}\left(B_1 + B_1^T\right)\right\}$$

The Theorem identifies a criterion for a pathwise asymptotic stability of the optimal filter in terms of convexity properties of the potential V (see (1)). Note that V incorporates both mechanisms, mixing properties of the signal implied by convexity properties of the first two terms $\langle \nabla \log \varphi(x), B_1 x\rangle + \frac{1}{2}\frac{\mathrm{tr}(Q\varphi_{xx}(x))}{\varphi(x)}$, and precision of the measurement, encoded in the third term $\frac{1}{2}\langle R^{-1}Gx, Gx\rangle$.

References

[1] D. Crisan, B. Rozovskii, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, Oxford (2011).

[2] W. Stannat, *Stability of the pathwise filter equation for a time-dependent signal on* $\mathbb{R}^d$, Appl. Math. Optim. **52** (2005), 113–132.

[3] W. Stannat, *Stability of the optimal filter via pointwise gradient estimates*, In: Stochastic Partial Differential Equations and Applications II, Editors: G. da Prato et al., Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, New York (2005), 281–293.

[4] W. Stannat, *Stability of the optimal filter for nonergodic signals - a variational approach*, In: The Oxford Handbook of Nonlinear Filtering, Editors: D. Crisan, B. Rozovskii, Oxford University Press, Oxford (2011), 374–399.

Stability and Error Analysis for Variational and Ensemble Data Assimilation

Roland Potthast

(joint work with África Periáñez, Hendrik Reich, Alexander Moodey, Peter Jan van Leeuwen, Amos Lawless, Boris Marx,)

In our first part of the talk, we investigate the *error dynamics* for cycled data assimilation systems, such that the *inverse problem* of *state estimation* is solved at t_k, $k = 1, 2, 3, \ldots$ with a first guess given by the state propagated via a dynamical system model M_k from time t_{k-1} to time t_k. In particular, for *nonlinear dynamical systems* M_k that are Lipschitz continuous with respect to their initial states, we provide deterministic estimates for the development of the error $\|\mathbf{e}_k\| := \|\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)}\|$ between the estimated state $\mathbf{x}^{(a)}$ and the true state $\mathbf{x}^{(t)}$ over time.

Clearly, observation error of size $\delta > 0$ leads to an estimation error in every assimilation step. These errors can accumulate, if they are not (a) controlled in the reconstruction and (b) damped by the dynamical system M_k under consideration. A data assimilation method is called stable, if the error in the estimate is bounded in time by some constant C.

The key task of this work is to provide estimates for the error $\|\mathbf{e}_k\|$, depending on the size δ of the observation error, the reconstruction operator R_α, the observation operator H and the Lipschitz constants $K^{(1)}$ and $K^{(2)}$ on the lower and higher modes of M_k controlling the damping behaviour of the dynamics. We show that systems can be stabilised by choosing α sufficiently small, but the bound C will then depend on the data error δ in the form $c\|R_\alpha\|\delta$ with some constant c. Since $\|R_\alpha\| \to \infty$ for $\alpha \to 0$, the constant might be large. Numerical examples for

this behaviour in the nonlinear case are provided using simple models in dynamic magnetic tomography and by a (low-dimensional) Lorenz '63 system.

In the second part of the talk we study *ensemble data assimilation systems*. They provide a flexible alternative to large-scale variational systems. Many different versions of ensemble filters have been suggested and tested over the last years.

However, although Ensemble-DA has high potential, its theoretical justification is still under development, both on the *application side* testing the development and operational setup of EnKF systems for different data types as well as the *mathematical analysis* looking into *approximation properties* of ensembles with ensemble-size small compared to the total number of degrees of freedom in the model. It is very common to use *space localization* in order to reduce the effect of spurious long range correlations.

It is the goal of our analysis to understand the basic properties of localization. We derive *deterministic error estimates* for the EnKF and study its dependence on localization. In particular, we investigate the *convergence* of the localized EnKF when the localization radius tends to zero.

Then, we demonstrate the practical meaning of the analysis by derivation of a formula for an *optimal localization radius* depending on the *observation density* and the *observation error*. The validity of the formula is also demonstrated by numerical experiments for some simple assimilation systems.

References

[1] Boris Marx and Roland W.E. Potthast, *On Instabilities for Data Assimilation Algorithms*, EM - International Journal on Geomathematics (2012), DOI: 10.1007/s13137-012-0034-5.

[2] Roland W.E. Potthast, Alexander J.F. Moodey, Amos S. Lawless, Peter Jan van Leeuwen, *On Error Dynamics and Instability in Data Assimilation.* under review.

[3] Alexander J.F. Moodey, Amos S. Lawless, Roland W.E. Potthast, Peter Jan van Leeuwen *Nonlinear error dynamics for cycled data assimilation methods.* Inverse Problems, in Press.

[4] Africa Perianez, Hendrik Reich and Roland Potthast, *Error Analysis and Adaptive Localization for Ensemble Methods in Data Assimilation*, Preprint.

Incorporating Representativity Error in Data Assimilation

Nancy K. Nichols

(joint work with J.A. Waller, S.L. Dance, A.S. Lawless and J.R. Eyre)

Observations used in combination with model predictions for data assimilation can contain information at smaller scales than the model can resolve. Errors of representativity are the result of small scale observational information being incorrectly represented in the model. In previous work, error correlations in satellite observations from various instruments have been identified that may be caused by representativity error [7]. A better understanding of these errors would allow them to be incorporated into the observation error statistics to provide more accurate analyses and enable us to make better use of available observations. The aims

of the research described here are to investigate the structure and properties of representativity errors and to incorporate and assess the impact of these errors in data assimilation.

Initially we have used a method first proposed by Daley [1] and then by Liu and Rabier [4] to diagnose static representativity error covariances. In this method it is assumed that the observations can be written as the mapping of a high resolution state into observation space and that the model state is a truncation of the high resolution state. We have applied this technique to determine the structure of representativity errors in a nonlinear advection-diffusion model with multi-scale chaotic behaviour (the Kuromoto-Sivashinsky equation [6, 3]) and to analyse time independent forward errors for different types of observation. The procedure has also been implemented with temperature and humidity data from the Met Office high resolution (UKV) system, as part of a co-operative PhD project. The data is taken from the system under different atmospheric conditions where small scale features can be identified in the humidity fields.

The results from these experiments have shown that errors of representativity

- are correlated and are state and time dependent;
- are reduced where the resolution of the model is increased or where the observation length scale is increased;
- vary with height throughout the atmosphere;
- are more significant for humidity than temperature.

The magnitude and structure of the covariance is, however, dependent only on the distance between observations and not the number of observations available [5].

From this work we have concluded that although the diagnostics of Liu and Rabier can be used to reveal the structure of forward observation errors, the results are not readily applicable in data assimilation due to the underlying assumptions that the errors are static and are determined through a non-invertible transform and truncation operator.

We are now developing a new method for diagnosing and incorporating time-dependent representativity error in an ensemble data assimilation system. The method is based on the diagnostics of Desroziers et al. [2], together with a symmtrization and localization technique, and has been investigated using the Kuromoto-Sivashinsky multi-scale model. The procedure provides an estimate of the correlated forward error covariance, from which the representativity error covariance can be obtained by removing the uncorrelated instrument error from the result.

Using this approach we have been able to recover the true observation error covariance in cases where the initial error covariance used is incorrect, and also to follow time-varying observation error covariances where the length scale of the true covariance is changing slowly in time. We have also shown that including the estimated representativity error in the assimilation improves the analysis. [8].

References

[1] R. Daley, *Estimating observation error statistics for atmospheric data assimilation*, Ann. Geophysicae 11: (1993) 634–647.

[2] G. Desroziers, L. Berre, B. Chapnik, and P. Poli, *Diagnosis of observation, background and analysis error statistics in observation space.*, Quarterly Journal of the Royal Meteorological Society, 131: (2005) 3385–3396.

[3] Y. Kuramoto, *Diffusion-induced chaos in reaction systems*, Progress of Theoretical Physics, 64: (1978) 346–367.

[4] Z. Q. Liu, F. Rabier, *The interaction between model resolution, observation resolution and observation density in data assimilation: A one dimensional study*, Quarterly Journal of the Royal Meteorological Society 128: (2002) 1367–1386.

[5] J. A. Pocock, S. L. Dance, A. S. Lawless, N. K. Nichols and J. R. Eyre, *Representativity error for temperature and humidity using the Met Office high resolution model*, University of Reading, Department of Mathematics, Preprint Series MPS-2012-18, 2012, (*submitted for publlication*).

[6] G. Sivashinsky, *Nonlinear analysis of hydrodynamic instability in laminar flames*, Acta Astronautica, 4: (1977) 1177–1206.

[7] L. M. Stewart, J. Cameron, S. L. Dance, S. English, J. R. Eyre, and N. K. Nichols, *Observation error correlations in IASI radiance data*, University of Reading, Department of Mathematics, Mathematics Report Series 1/2009, 2009.

[8] J.A. Waller, *Using observations at different spatial scales in data assimilation for environmental prediction*, University of Reading, Department of Mathematics, PhD Thesis, (*in preparation).*

Decadal ocean (and ice) state estimation for climate research: What are the needs?

Patrick Heimbach

(joint work with Carl Wunsch)

The advent of the World Ocean Circulation Experiment (WOCE) and satellite altimetry in the early 1990s called for a method for synthesizing the new, globally distributed yet sparse observations of very diverse types into a coherent dynamical framework, one that would enable the calculation of accurate budgets of heat, freshwater, momentum, vorticity, their exchange with the atmosphere, and their evolution through time. A major goal was understanding the mechanistic processes underlying large-scale (multi-)decadal climate variability and change.

In contrast to the practice of data assimilation (DA) in numerical weather prediction (NWP) where the problem is one of optimal forecasting or filtering (extrapolation), we are faced in ocean climate research with a smoothing or state and parameter estimation problem (interpolation). A serious consequence is that techniques used in NWP cannot be readily adopted. However, most so-called ocean "reanalysis" projects are relying on just such filtering methods. The states they produce face the same problems as the atmospheric reanalyses where assimilation increments incur unphysical sources or sinks in the conservation equations. Such artificial sources or sinks of heat or mass at the analysis time severely limit their utility for climate research. Although attention has been called to this issue repeatedly, it is being ignored by a large part of the ocean/atmosphere/climate

research and DA community. Examples are atmosphere-to-ocean global net freshwater flux imbalances of the order of 5 to 10 cm per year, in stark conflict with satellite altimetric observations which place an upper bound on such global mean fluxes of roughly 3 mm per year (the satellite-altimetric estimate of global mean sea level rise since 1992). Similar conflicts exist between net surface heat flux imbalances in atmospheric reanalysis products of the order of 10 W/m^2 compared to roughly 1 (± 1) W/m^2 inferred from independent radiation balance estimates. Such violation of global and local conservation of properties are of little concern to weather prediction, but they are central to the global climate problem [14].

Over the last decade a consortium called "Estimating the Circulation and Climate of the Ocean" (ECCO) has developed a smoother approach for synthesizing much of the available oceanographic (and more recently sea ice) observations into a dynamical framework represented by the Massachusetts Institute of Technology general circulation model (MITgcm) [1]. Estimates have been produced both for the global problem [12, 17] as well as for regional domains, such as an Arctic subdomain coupled ocean-sea ice state [3] and the Southern Ocean [11]. The GCM is fit in a least-squares sense to the observations by means of the Lagrange multiplier or adjoint method [18]. A major algorithmic breakthrough that has made this approach computationally tractable is the rigorous use of automatic differentiation (AD) to generate and exact, efficient, scalable, and up-to-date adjoint model of the nonlinear forward model [7].

The adjoint computes the gradient of the least-squares model-data misfit cost function with respect to a very high-dimensional (10^7 to 10^9) space of independent, uncertain, adjustable control variables, consisting of three-dimensional ocean (and sea ice) initial conditions and two-dimensional time-varying corrections to the atmospheric boundary conditions, which are known to contain significant errors that are spatially and temporally inhomogeneous. Including three-dimensional (time-mean) model parameters to the control space as well, such as vertical diffusivity and eddy-induced mixing parameters, provides a significant step forward toward dealing with internal model (or parameterization) errors in a way that does not introduce artificial source or sink terms [4, 13]. The state estimates produced by ECCO, covering the satellite altimetric record (1992-present) have enabled accurate budget calculations for various applications, such as understanding the causes of the strong regional variations in sea level trends over the last two decades [19], or the spatio-temporal structure of the Global and Atlantic Meridional Overturning Circulation (MOC) [16, 15].

As a by-product of the estimation project, the availability of the adjoint or dual state (i.e., elements of the model's co-tangent space) has enabled detailed sensitivity studies of climate-relevant indices to the time-evolving state and boundary values (examples include Atlantic poleward volume and heat transport [10, 8, 5], sensitivity of Drake Passage volume transport to changes in bottom topography [9], Arctic sea ice export sensitivities to changes in atmospheric state [6], and ocean biological productivity sensitivity to nutrient supply [2]).

Challenges facing the community today are (1) improving the accuracy of these smoother-based state estimates for use in climate research, (2) the provision of formal error estimates (posterior uncertainties) on these state estimates and derived quantities in the presence of a high-dimensional uncertainty space, (3) the need to improve physical consistency through a coupled atmosphere/ocean/ice estimation system which involves highly disparate time scales, (4) the move towards improved horizontal resolution with its own issues (observed versus represented scales, nonlinear regime of the flow field, poor sampling). At least for the climate estimation problem (i.e. for the purpose of understanding the evolution of the climate state over the past few decades), filtering approaches borrowed from DA as practiced in NWP will likely be of limited use for the forseeable future.

References

[1] Adcroft, A., C. Hill, J-M. Campin, J. Marshall and P. Heimbach: *Overview of the formulation and numerics of the MIT GCM*, In: Proceedings of the ECMWF Seminar on "Recent developments in numerical methods for atmospheric and ocean modelling", 6-10 September 2004, Shinfield Park, Reading, UK, pp. 139-150.

[2] Dutkiewicz, S., M.J. Follows, P. Heimbach and J. Marshall: *Controls on ocean productivity and air-sea carbon flux: An adjoint model sensitivity study*, Geophysical Research Letters **33** (2006), L02603, doi:10.1029/2005GL024987.

[3] Fenty, I.G. and P. Heimbach: *Coupled Sea Ice-Ocean State Estimation in the Labrador Sea and Baffin Bay*, Journal of Physical Oceanography (2013), doi:10.1175/JPO-D-12-065.1.

[4] Ferreira, D., J. Marshall and P. Heimbach: *Estimating eddy stresses by fitting dynamics to observations using a residual-mean ocean circulation model and its adjoint*, Journal of Physical Oceanography **35** (2005), 1891–1910, doi:10.1175/JPO2785.1.

[5] Heimbach, P., C. Wunsch, R.M. Ponte, G. Forget, C. Hill, and J. Utke: *Timescales and regions of the sensitivity of Atlantic meridional volume and heat transport magnitudes: toward observing system design*, Deep Sea Research II **58(17-18)** (2011), 1858–1879, doi:10.1016/j.dsr2.2010.10.065.

[6] Heimbach, P., D. Menemenlis, M. Losch, J.M. Campin, and C. Hill: *On the formulation of sea-ice models. Part 2: Lessons from multi-year adjoint sea ice export sensitivities through the Canadian Arctic Archipelago*, Ocean Modelling **33(1-2)** (2010), 145–158, doi:10.1016/j.ocemod.2010.02.002.

[7] Heimbach, P., C. Hill and R. Giering: *An efficient exact adjoint of the parallel MIT general circulation model, generated via automatic differentiation*, Future Generation Computer Systems **21(8)** (2007), 1356–1371, doi:10.1016/j.future.2004.11.010.

[8] Köhl, A.: *Anomalies of Meridional Overturning: Mechanisms in the North Atlantic*, Journal of Physical Oceanography **35** (2005), 1455–1472, doi:10.1175/JPO2767.1.

[9] Losch, M. and P. Heimbach: *Adjoint sensitivity of an ocean general circulation model to bottom topography*, Journal of Physical Oceanography **37(2)** (2007), 377–393, doi:10.1175/JPO3017.1.

[10] Marotzke, J., R. Giering, K.Q. Zhang, D. Stammer, C. Hill, and T. Lee: *Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport variability*, Journal of Geophysical Research **104(C12)** (1999), 29,529–29,547.

[11] Mazloff, M.R., P. Heimbach and C. Wunsch: *An eddy-permitting Southern Ocean State Estimate (SOSE)*, Journal of Physical Oceanography **40(5)** (2010), 880–899, doi:10.1175/2009JPO4236.1.

[12] Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C. Hill, J. and J. Marshall: *The global ocean circulation during 1992-1997, estimated*

from ocean observations and a general circulation model, Journal of Geophysical Research **107(C9)** (2002), 3118, doi:10.1029/2001JC000888.

[13] Stammer, D.: *Adjusting internal model errors through ocean state estimation*, Journal of Physical Oceanography **35** (2005), 1143–1153.

[14] Wunsch, C. and P. Heimbach: *Dynamically and kinematically consistent global ocean circulation and ice state estimates*, In: G. Siedler, J. Church, J. Gould and S. Griffies, eds.: Ocean circulation and climate: observing and modelling the global ocean, 2nd Ed. (2013), in press, Elsevier.

[15] Wunsch, C. and P. Heimbach: *Two decades of the Atlantic Meridional Overturning Circulation: anatomy, variations, extremes, prediction, and overcoming its limitations*, Journal of Climate (2012), in revision.

[16] Wunsch, C. and P. Heimbach: *The globally integrated ocean circulation (MOC), 1992-2006: seasonal and decadal variability*, Journal of Physical Oceanography **39(2)** (2009), 351–368, doi:10.1175/2008JPO4012.1.

[17] Wunsch, C., P. Heimbach, R. Ponte, and I. Fukumori: *The global general circulation of the ocean estimated by the ECCO Consortium*, Oceanography **22(2)** (2009), 88–103.

[18] Wunsch, C. and P. Heimbach: *Practical global ocean state estimation*, Physica D **230(1-2)** (2007), 197–208, doi:10.1016/j.physd.2006.09.040.

[19] Wunsch, C., R. M. Ponte, and P. Heimbach: *Decadal trends in sea level patterns: 1993-2004*, Journal of Climate **20(24)** (2007), 5889–5911, doi:10.1175/2007JCLI1840.1.

Challenges Associated with the Efficient Synthesis and Normalization of Gridded Anisotropic Covariances by Line Filters

R. James Purser

Recursive spatial numerical filters can be used to spread information along grid lines in emulation of a rank-1 Gaussian convolution. The Gaussian shape is significant, being the only shape for which, under sequential applications in the sufficient transversal directions, there is no imprint of the grid's orientation upon the final smooth result (which is then a Gaussian convolution of full rank). An efficient sequence of such line filters is easily made positive and self-adjoint, and therefore serves as a model for a scalar error field's spatial covariance. Realistic covariances of a scalar error field may not have a purely Gaussian profile, but a large repertoire of realistic covariances *can* be built up by linear superpositions of Gaussians in a systematic way, a process that resembles a discrete inverse Laplace transformation.

The characteristics of spatial dispersion for a purely Gaussian filtering sequence can be represented by the *aspect tensor* which, in spatially homogeneous conditions, generalizes to the local diffusivity for a simulated diffusion process of duration $t = 1/2$. The diffusion model for covariances was made explicit by Derber and Rosati [1] and by Weaver and Courtier [6] for ocean data assimilation.

In a smooth gridding of $n \geq 2$ spatial dimensions it can be shown that the rank-1 aspect tensors associated with the set of minimal grid generators has a convex hull in the $C(n+1,2)$-dimensional *aspect space* whose boundary is a polyhedral shell containing all the images of these generators at *its* boundary. Moreover, it is complete in the sense that all proper aspect tensors centrally-project into this shell. Therefore, it is always possible to linearly resolve any desired aspect tensor into rank-1 components associated with a set of only $C(n+1,2)$ generators. In

$n = 2$ and $n = 3$ dimensions this results in a unique *triad* or *hexad* of line filters respectively. In both cases, we avoid numerical conflicts during the execution of the associated triad or hexad of line filters that cross the same physical grid point by labelling each of the possible lines with what we shall call a *color* uniquely matching one of the non-null elements of the Galois (finite) fields $GF(4)$ or $GF(8)$, where these elements are mapped periodically onto the two- or three-dimensional lattice containing the grid generators. Thus, in two dimensions where there are three colors, the filtering over the entire physical grid is performed in three sweeps; in three dimensions, seven sweeps are performed, and it is guaranteed then that there is no interference between mutually oblique line filters, however much the choice of these lines changes geographically. The fact that the Galois fields' multiplicative operations each form a cyclic group is also exploited to simplify and accelerate the algorithms used in the iterative search for a target aspect tensor's correct triad/hexad and, in the 3D case, where the set of paths passing between adjoining hexads is not simply connected, the elements of $GF(8)$ allow each hexad to be consistently oriented.

In more sophisticated *blended* versions of the algorithm for finding the smoothing intensities, or *weights*, for a given aspect tensor we can alleviate the problem of numerically spurious roughness that sometimes accompanies the transition from one triad/hexad to the next by replacing the weights implied by the single target aspect tensor in the basic algorithm by a linear accumulation of weights that emerges from an integration of those basic weights over an averaging kernel specified as a ball with convenient radial generalized distribution of *mass* centered on the target aspect tensor. The radius of the support of this kernel depends continuously on its position within the aspect cone, and the averaging effected by integration of neighboring aspect tensors accumulated in proportion to the kernel's mass implies the accumulation of the corresponding line filter weights since, in the appropriate local metric, these weights vary linearly within each triad/hexad. In general this blending of piecewise-linear contributions from the local tight cluster of triads/hexads that are overlapped by the kernel leads to a successful elimination of roughness in the modelled covariance, but it necessarily involves a larger set of line filters at each physical grid point, and therefore a larger set of distinguishing colors than in the respective basic triad/hexad algorithms in two/three physical dimensions if numerical conflicts are still to be avoided. A sufficient degree of blending requires the larger palettes of 4 or 13 colors respectively that emerge naturally from the central collineations of pairs of the non-null elements of $GF(9)$ or $GF(27)$ mapped onto the lattices containing the grid generators in two or three physical dimensions.

The representation of covariances by these efficient approximations of diffusion involve a fundamental ambiguity relating to how the amount of dispersion is partitioned into a contribution coming from the diffusivity tensor itself and a contribution coming from the metric of an implied Riemannian geometry in which the diffusivity tensor is separately represented. We can choose to resolve this ambiguity in the most natural way by declaring the diffusivity tensor itself to be the

identity, so that the entire contribution to the filter's dispersion is controlled by the choice of metric (which thereby effectively *becomes* the aspect tensor).

A major advantage of this convention is that an estimate for the point-wise amplitude, or *variance* in the difficult case of an inhomogeneous diffusion is provided as the modulation of the standard Gaussian amplitude by a *parametrix function*, $A(x)$, expressed as an asymptotic expansion in the curvature and its derivatives only [4], [2], [5]. For example, to a second order in an expansion parameter and for general $n \geq 2$:

$$A(x) = 1 + \frac{R}{12} + \frac{1}{1440}\left(12\nabla^2 R + 5R^2 - 2R_{ij}R^{ij} + 2R_{ijkl}R^{ijkl}\right).$$

For $n \leq 3$ the Riemann-Christoffel curvature is not independent of the Ricci, so we gain the simplification:

$$A(x) = 1 + \frac{R}{12} + \frac{1}{480}\left(4\nabla^2 R + R^2 + 2R_{ij}R^{ij}\right),$$

and for $n = 2$, where the Gaussian curvature $\kappa = R/2$, we gain the further simplification,

$$A(x) = 1 + \frac{\kappa}{6} + \frac{1}{60}\left(\nabla^2\kappa + \kappa^2\right).$$

Being asymptotic, we are not surprised that this expansion, implemented directly to achieve our amplitude approximation, exhibits a problematic divergence reminiscent of that seen in applications of the Gram-Charlier expansion of classical Statistics, by which a nearly-Gaussian probability density is approximately recovered by the simplest expansion in its successive diagnosed cumulants [3]. In the statistical context, the alternative Edgeworth expansion, though less intuitive and appearing to contain, at each order, extra terms that seem unnecessary, is actually a much more robust approximation in most practical cases. Thus, we conclude with the analogous challenge: what, in the context of estimating the local parametrix function, $A(x)$, is the *robust* expansion of it that we can consider to correspond to the 'Edgeworth-type' expansion for this Riemannian-space diffusion model?

References

[1] J. C. Derber and A. Rosati, *A global ocean data assimilation system.* J. Phys. Ocean. **19** (1989), 1333–1347.

[2] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen.* Teubner, Leipzig and Berlin, (1912) 282 pp.

[3] M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol I, 4th Edition.* Macmillan, New York (1977).

[4] E. E. Levi, *Sulle equazioni lineari totalmente ellittiche alle derivate parziale.* Rend. Circ. Mat. Palermo **24** (1907), 275–317.

[5] R. J. Purser, *Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part II: Riemannian geometry and the generic parametrix expansion method.* NOAA/NCEP Office Note **457** (2008) 48 pp.

[6] A. Weaver and P. Courtier, *Correlation modelling on the sphere using a generalized diffusion equation.* Quart. J. Roy. Meteor. Soc., **127** (2001) 1815–1846.

Parameter Identification and Optimal Control of a Chemotaxis Problem

Arnd Rösch
(joint work with Hendrik Feldhordt, Michael Winkler)

Chemotaxis describes a directed movement of cells caused by the concentration gradient of a chemical substance. We consider the following simplified model:

$$\begin{aligned} u_t &= \Delta u - \nabla \cdot \{f(u)\nabla v\} \\ v_t &= \Delta v - v + u \\ \partial_n u &= 0 \qquad \partial_n v = g \\ u(0) &= u_0 \qquad v(0) = v_0 \end{aligned}$$

In this quasilinear system u describes the concentration of bacteria and v stands for the concentration of a chemoattractant. The quasilinearity of the system is caused by the cross diffusion term $-\nabla \cdot \{f(u)\nabla v\}$.

Several publications in the last years discussed a possible blow up behavior of the solution for specific choices of the function f and for $g \equiv 0$. We are interested in controlling the chemotaxis system via the function g and in identifying the function f.

Theorem 0.4. *Let $f \in C^1(\mathbb{R})$ be a globally bounded and globally Lipschitz function of class C^1 and $g \in L^\infty((0,T); L^p(\partial\Omega))$. Then the chemotaxis system has a unique solution. Moreover, the a priori estimate*

$$\begin{aligned} \|u\|_{C^0([0,T];C^0(\overline{\Omega}))} &+ \|v\|_{L^\infty((0,T);W^{1,p}(\Omega))} \\ &\leq C\left(\|u_0\|_{C^0(\overline{\Omega})} + \|v_0\|_{W^{1,p}(\Omega)} + \|g\|_{L^\infty((0,T);L^p(\partial\Omega))}\right) \end{aligned}$$

is valid.

The proof uses semigroup theory and combines techniques from Horstmann and Winkler [1] and Tröltzsch [2].

Let us consider the following minimization problem:

$$\min J(g;(u,v)) := \frac{\mu}{2}\|u - u_Q\|^2_{L^2(Q)} + \frac{\nu}{2}\|v - v_Q\|^2_{L^2(Q)} + \frac{\lambda}{2}\|g\|^2_{L^2(\Sigma)}$$

with $\mu, \nu \geq 0$ and $\lambda > 0$ over the set

$$G_{ad} := \{g \in L^\infty((0,T); L^p(\partial\Omega)), g_a \leq g \leq g_b\}.$$

Theorem 0.5. *The minimization problem posesses at least one optimal control $\overline{g} \in G_{ad}$ with corresponding state $(\overline{u}, \overline{v})$.*

One can show that the mapping $g \mapsto (u, v)$ is Fréchet differentiable. Consequently the following optimality system is obtained:

$$\begin{aligned} u_t &= \Delta u - \nabla \cdot f(u)\nabla v \\ v_t &= \Delta v - v + u \\ \partial_n u &= 0 \qquad \partial_n v = g \\ u(0) &= u_0 \qquad v(0) = v_0 \end{aligned}$$

$$\begin{aligned} -p_t - \Delta p &= f'(u)\nabla v \nabla p + q + \mu(u - u_Q) \\ -q_t - \Delta q + q &= -\nabla \cdot \{f(u)\nabla p\} + \nu(v - v_Q) \\ \partial_n p + f'(u)pg &= 0 \\ \partial_n q - f(u)\partial_n p &= 0 \\ p(T) &= 0 \qquad q(T) = 0 \end{aligned}$$

$$\langle -f(u)p + q + \lambda\overline{g}, g - \overline{g}\rangle_{L^2(\Sigma)} \geq 0 \qquad \forall g \in G_{ad}.$$

Similar results can be derived for a modified system. The modified boundary condition

$$\partial_n u - f(u)\partial_n v = 0$$

is motivated by a no flux condition for u.

Another interesting minimization problem is the parameter identification for the function f. We minimize the functional

$$\min J(f; (u, v)) := \frac{\mu}{2} \|u - u_Q\|^2_{L^2(Q)} + \frac{\nu}{2} \|v - v_Q\|^2_{L^2(Q)} + R(f)$$

with $\mu, \nu \geq 0$ over $f \in F_{ad}$ where F_{ad} is a closed, convex, and bounded set in $C^{1,\nu}(\mathbb{R})$. Here, $R(f)$ is a suitably chosen regularization term. For simplicity we set $g \equiv 0$.

Let us mention that this parameter identification problem has a non-standard form since the parameter f depends on the solution u of the chemotaxis problem. This structure leads to specific difficulties in analysis, optimization and numerics. Nevertheless, one can derive an adjoint equation

$$\begin{aligned} -p_t - \Delta p &= f'(u)\nabla v \nabla p + q + \mu(u - u_Q) \\ -q_t - \Delta q + q &= -\nabla \cdot \{f(u)\nabla p\} + \nu(v - v_Q) \\ \partial_n p &= 0 \qquad \partial_n q = 0 \\ p(T) &= 0 \qquad q(T) = 0. \end{aligned}$$

The adjoint equation is one part of the optimality system. Another part of the system is a variational inequality depending on the choice of the regularization term $R(f)$.

References

[1] D. Horstmann and M. Winkler, *Boundedness vs blow-up in a chemotaxis system*, Journal of differential equations **215** (2005), 52–107.

[2] F. Tröltzsch, *On convergence of semidiscrete Ritz-Galerkin schemes applied to the boundary control of parabolic equations with nonlinear boundary condition*, ZAMM **72** (1992), 291–301.

Reconstruction of matrix parameters from noisy measurements

Michael Hinze

(joint work with Klaus Deckelnick)

For an open bounded domain $\Omega \subset \mathbb{R}^n$ with boundary Γ we consider the reconstruction of the diffusion matrix $A \in L^\infty(\Omega, \mathbb{R}^{n,n})$ in the ellipitc

$$(PDE) \quad -\operatorname{div}(A\nabla y) = g \text{ in } \Omega,\ y = 0 \text{ on } \Gamma$$

from noisy measurements $z \in Z$. Here, $g \in H^{-1}(\Omega)$ is given and fixed, and Z is $L^2(\Omega)$ or $H^1(\Omega)$.

For given $A \in \mathcal{M}$, $g \in H^{-1}(\Omega)$ let $T(A,g)$ denote the unique weak solution to (PDE). For the reconstrunction of A we consider the minimization problem

$$(P) \quad \min_{A\in\mathcal{M}} \frac{1}{2}\|y - z\|_Z^2 \text{ s.t. } y = T(A,g),$$

where for $0 < a < b < \infty$

$$\mathcal{M} := \{A \in L^\infty(\Omega)^{n,n} \mid A(x) \in K \text{ a.e. in } \Omega\},$$

with

$$K := \{A \in \mathcal{S}_n \mid a \le \lambda_i(A) \le b, i = 1, \dots, n\}.$$

Here, $\mathcal{S}_n$ denotes the set of all symmetric $n \times n$ matrices endowed with the inner product $A \cdot B = \operatorname{trace}(AB)$, and $\lambda_1(A), \dots, \lambda_n(A)$ denote the eigenvalues of A. This optimal control problem is not convex, and admits nonstandard control constraints.

Due to a theorem of Murat and Tartar [5] $\mathcal{M}$ is H–compact, i.e. every sequence $(A_k)_{k\in\mathbb{N}}$ in $\mathcal{M}$ contains a subsequence $(A_{k'})_{k'\in\mathbb{N}}$ converging to an element $A \in \mathcal{M}$ in the sense that for every $g \in H^{-1}(\Omega)$

$$T(A_{k'}, g) \rightharpoonup T(A,g) \text{ in } H_0^1(\Omega) \text{ and } A_{k'}\nabla T(A_{k'}, g) \rightharpoonup A\nabla T(A,g) \text{ in } L^2(\Omega)^n.$$

$(A_{k'})_{k'\in\mathcal{M}}$ is then said to be H–convergent to A ($A_{k'} \overset{H}{\to} A$). This directly delivers

Theorem 0.1. *(P) admits a solution $A \in \mathcal{M}$ with corresponding state $y = T(A,g)$.*

For $\gamma > 0$ the Tikhonov regularization of problem (P) is

$$(P)_\gamma \quad \min_{A\in\mathcal{M}} J_\gamma(y,A) := \frac{1}{2}\|y - z\|_Z^2 + \frac{\gamma}{2}\|A\|_{L^{2(n,n)}}^2 \text{ s.t. } y = T(A,g).$$

Then $(P)_\gamma$ admits a solution, which follows from the fact that $A_k \overset{H}{\to} A$ and $A_k \overset{*}{\rightharpoonup} A_0$ in $L^\infty(\Omega, \mathbb{R}^{n,n})$ imply $A(x) \leq A_0(x)$ a.e. in Ω, and

$$\|A\|^2 \leq \|A_0\|^2 \leq \liminf_{k\to\infty} \|A_k\|^2,$$

see also [5].

Let $(a \otimes b)_{kl} := \frac{1}{2}(a_k b_l + a_l b_k), k, l = 1, \ldots, n$ and let $A \in \mathcal{M}$ be a solution of $(P)_\gamma$. From [2] we have for every $\lambda > 0$ the optimality condition

$$A(x) = P_K\left(A(x) - \lambda\left(\gamma A(x) - \nabla y(x) \otimes \nabla p(x)\right)\right) \text{ a.e. in } \Omega,$$

where

$$P_K(A) = S^t \mathrm{diag}\left(P_{[a,b]}(\lambda_1(A)), \ldots, P_{[a,b]}(\lambda_n(A))\right) S,$$

if $A = S^t diag(\lambda_1(A), \ldots, \lambda_n(A))S$.

For the discretization of $(P)_\gamma$ $(\gamma \geq 0)$ we use variational discretization [4] and consider

$$(P_h)_\gamma \quad \min_{A\in\mathcal{M}} \frac{1}{2}\|y_h - z\|_Z^2 + \frac{\gamma}{2}\|A\|^2_{L^{2(n,n)}} \text{ s.t. } y_h = T_h(A, g),$$

where T_h denotes the $c(1)$ FE discretization of T. Let us note that the matrix parameter is not discretized in this problem. In [2] we prove that $(P_h)_\gamma$ for every $h > 0$ admits a solution $A_h \in \mathcal{M}$, where we use the following concept of Hd–convergence adapted from Eymard/Galouët [1]; let $(A_h)_{h>0}$ be a sequence in $\mathcal{M}$. Then there exists a subsequence $(A_{h'})_{h'>0}$ and $A \in \mathcal{M}$ such that for every $g \in H^{-1}(\Omega)$

$$T_{h'}(A_{h'}, g) \rightharpoonup T(A, g) \text{ in } H_0^1(\Omega) \text{ and } A_{h'}\nabla T_{h'}(A_{h'}, g) \rightharpoonup A\nabla T(A, g) \text{ in } L^2(\Omega)^n.$$

This means $(A_{h'})_{h'\in\mathcal{M}}$ Hd–converges to A, i.e. $A_{h'} \overset{Hd}{\to} A$. Moreover, any solution A_h of $(P_h)_\gamma$ satisfies

$$A_h(x) = P_K\left(A_h(x) - \lambda\left(\gamma A_h(x) + \nabla y_h(x) \otimes \nabla p_h(x)\right)\right) \text{ a.e. in } \Omega.$$

Convergence: Given z, g, let us suppose that

$$S := \{A \in \mathcal{M} \mid z = T(A, g)\} \neq \emptyset$$

and set $\|A^*\|_{L^2} = \min_{A\in S} \|A\|_{L^2}$. Let us consider

$$(P_\delta^h) \quad \min_{A\in\mathcal{M}} J_\delta^h(A) = \frac{1}{2}\|y_h - z_\delta\|_Z^2 + \frac{\gamma}{2}\|A\|^2_{L^2} \text{ s.t. } y_h = T_h(A, g_\delta).$$

In [2] we prove

Theorem 0.2. *Let $A_\delta^h \in \mathcal{M}$ be a solution of (P_δ^h) and suppose that $\|z_\delta - z\|_{L^2} \leq \delta$, $\|g_\delta - g\|_{H^{-1}} \leq \delta$. Furthermore, with*

$$\rho_h := \inf_{v_h\in X_H} \|z - v_h\|_{H^1},$$

let us suppose that $\gamma \to 0$, $\frac{\delta}{\sqrt{\gamma}} \to 0$, and $\frac{\rho_h}{\sqrt{\gamma}} \to 0$ as $h, \delta \to 0$. Then

$$\|A_\delta^h - A^*\|_{L^2} \to 0 \text{ as } \delta, h \to 0.$$

Moreover, in [3] we under further assumptions are able to derive an error bound. For this purpose let us assume that $\Omega \subset \mathbb{R}^2$ is a bounded polygonal, convex domain. Furthermore, let us assume that with some $p > 2$

(A1) $z = T(A, g) \in H_0^1(\Omega) \cap W^{2,p}(\Omega)$, and that

(A2) there exists $\psi \in H_0^1(\Omega) \cap W^{2,p}(\Omega)$ such that

$$A(x) = P_K\left(\nabla z \otimes \nabla \psi\right) \text{ a.e. in } \Omega \text{ (source condition).}$$

Then we have

Theorem 0.3. *Theorem. Let $A_\delta^h \in \mathcal{M}$ be a solution of (P_δ^h) and suppose that $\|z_\delta - z\|_{L^2} \leq \delta$, $\|g_\delta - g\|_{H^{-1}} \leq \delta$. If $h > \sqrt{\delta}$ and $\gamma \sim h^2$, then*

$$\|A_\delta^h - A\|_{L^2} \leq Ch, \text{ and } \|y_h - z\|_{L^2} + h\|\nabla(y_h - z)\|_{L^2} \leq Ch^2,$$

where $y_h = T_h(A_\delta^h, g_\delta)$.

The proofs of these theorems are technical and rely on the fact that a solution to (P_γ) is feasible as comparison matrix for the discrete problems (P_δ^h), and that a matrix A satisfying (A2) represents the norm-minimal element in $S = \{A \in \mathcal{M} \mid z = T(A, g)\}$.

References

[1] Eymard, R., Gallouët, T.: *H–convergence and numerical schemes for elliptic problems*, Siam J. Numer. Anal. **41** (2003), 539–562.

[2] K. Deckelnick and M. Hinze, *Identification of matrix parameters in elliptic PDEs*, Control and Cybernetics **40** (2011), 957-969.

[3] K. Deckelnick and M. Hinze, *Convergence and error analysis of a numerical method for the identification of matrix parameters in elliptic PDEs.*, Inverse Problems **28** (2012), doi:10.1088/0266-5611/28/11/115015.

[4] M. Hinze, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, J. Comput. Optim. Appl. **30** (2005), 45-63.

[5] Murat, F., Tartar, L.: *H-convergence*, Topics in the mathematical modelling of composite materials, page 21–43, Andrej Charkaev, Robert Kohn Eds., 1997.

Challenges in Atmospheric Chemistry Data Assimilation

Hendrik Elbern

(joint work with Nadine Goris, Jörg Schwinger, Elmar Friese, Lars Nieradzik)

A novel challenge for climate research is the assimilation of trace gases, which aims to optimally exploit remote sensing data from costly space borne sensors. Target models for data assimilation in atmospheric chemistry are chemistry-transport models (CTMs). The need for chemistry data assimilation increased rapidly in recent years due to demands of GMES (Global Monitoring for Environment and Security) activities, and sophisticated scientific field mission analyses. Also advanced weather forecast models are expecting more precise radiative transfer calculations. With atmospheric water in all its physical conditions becoming increasingly important in general data assimilation, meteorological and chemical data assimilation converge in some domains.

The underlying model equation is an advection-diffusion-reaction equation. Omitting aerosol and aqueous phase the differential equation can be written as [2]:

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (\vec{v} c_i) - \nabla \cdot (\rho \mathbf{G} \nabla \frac{c_i}{\rho}) - \sum_{r=1}^{R} \left(k(r) \left(s_i(r_+) - s_i(r_-) \right) \prod_{j=1}^{U} c_j^{s_j(r_-)} \right) = E_i + D_i \tag{1}$$

where c_i is the concentration of species i, $\vec{v}$ is the wind velocity, $s \in \mathbb{N}_0$ is the stoichiometric coefficient, $k(r)$ is the reaction rate of reaction r, either productive (r_+) or destructive (r_-) for species i, U is the number of species in the mechanism, E_i is the emission rate of species i, D_i is the deposition rate of species i, the air density is denoted by ρ, and $\mathbf{G}$ is the symmetric eddy diffusivity tensor. Upon application of the variational calculus, the adjoint formulation of (1) reads

$$-\frac{\partial \delta c_i^*}{\partial t} - \vec{v} \nabla \delta c_i^* - \frac{1}{\rho} \nabla \cdot (\rho \mathbf{K} \nabla \delta c_i^*) + \sum_{r=1}^{R} \left(k(r) \frac{s_i(r_-)}{c_i} \prod_{j=1}^{U} c_j{}^{s_j(r_-)} \sum_{n=1}^{U} \left(s_n(r_+) - s_n(r_-) \right) \delta c_n^* \right) = 0 \tag{2}$$

with δc_i^* the adjoint variable of c_i, while D_i is held fixed.

The presentation detailed open problems of chemisty data assimiltion, which differ from both the meteorological and oceanographic situation

From Data assimilation to Generalized Inversion: Until very recently, optimisation in meteorological data assimilation remains confined to optimal initial value estimation. Here, it is tacitly assumed that a suitably defined product of paucity of knowledge of initial values times their impact on forecast skill is maximal in comparison to model parameters other than initial values. This assumption is not any more valid in atmospheric chemistry, where other parameters like trace gas emission rates and their sinks (uptake) at surfaces are, especially in critical regions, of higher influence. Therefore, the data assimilation problem evolves to a generalized inversion task, where a set of heterogeneous parameters are to be estimated. Special challenges in this context are (i) different charachteristic time scales of the different parameters involved, and (ii) the typical assumption of white noise error characteristics being invalid. A solution based on the 4-dimensional variational assimilation techniques was presented, where the adjoint integration backward in time by (2) allows for the optimisation of emission rates, simultaneously to initial values.

Degree of underdetermination and covariance modelling:
In further contrast to meteorological conditions, the number of variables per grid point is equal or more than $O(100)$ rather than $O(10)$, where only a very small fraction of trace gases is observed, however. In today's "real world applications" the ill posedness of the assimilation problem induced by this condition needs to be solved by proper preconditioning of the forecast error covariance matrix (typically $\mathbf{B} \in \mathbb{R}^7 \times \mathbb{R}^7$), enforcing the calculation of its square root. While purely spatial

univariate covariances can be solved by the diffusion paradigm ([4]), where the covariance matrix is replaced by a diffusion operator providing equivalent statistical properties, this is not the case for multivariate (chemical, that is cross-species) covariances. The calculation of related covariance matrices will be feasible by ensemble simulations. In practice, no chemical consistent estimate across all reactive trace gases is as yet provided in a numerically efficient way. Despite the fact that emission error characteristic are not adequately modelled by white noise, temporal covariances are as yet not included, due to the exceptionally high computational demands. The efficient numerical treatment is still an open problem.

The a posteriori validation of the analyses is essential in estimating the validity of observation and forecast error covariances. The approach devised by [1] is applied to validate a balanced weighting between model skill and observation quality in observation space. This application was presented for a stratospheric dta assimilation application ([3]).

"Unexpected events": wild fires, mineral dust outbreaks, volcanic eruptions: While in "ordinary" cases it is assumed that the difference between the background and analysis does not significantly violate the tangent-linear assumption, this cannot be expected in case of unexpected events. Critical parameters are emission strengths and injection heights, while the emission location is often known. Particle filtering techniques are a candidate solution method.

References

[1] Desroziers, G., L. Berre, B. Chapnik, and P. Poli, *Diagnosis of observation, background and analysis-error statistics in observation space*, Q. J. R. Meteorol. Soc., **131**, (2005), 3385–3396.

[2] Elbern, H., A. Strunk, H. Schmidt and O. Talagrand, *Emission rate and chemical state estimation by 4-dimensional variational inversion*, Atmos. Chem. Phys., **7**, (2007), 1–59.

[3] Schwinger, J., and H. Elbern, *Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space*, J. Geophys. Res., **115**, D18307, (2010), doi:10.1029/2009JD013115.

[4] Weaver, A., and P. Courtier, *Correlation modelling on the sphere using a generalized diffusion equation*, Q. J. R. Meteorol. Soc., **127**, (2001), 1815–1846.

Simple models of convective clouds for data assmilation experiments

George C. Craig

(joint work with George Craig, Michael Würsch, Heiner Lange, Mylène Haslehner)

Assimilation of high-resolution observations of cumulus convection, such as radar reflectivities, is challenging. Convective clouds evolve quickly and nonlinearity becomes significant on timescales that are comparable to the intervals at which observations are available. A common example is the so-called late detection problem. Meteorological radars can only detect relatively large precipitation particles that do not develop until at least a quarter of an hour after the cloud initiates, by which time the air flow in the cloud has become fully developed. A second

challenge is the intermittency of the convective cloud field. The observable precipitation features often occupy a small fraction of the domain, and errors in the background forecast tend to take the form of large spatial displacements or false alarms and missed events. As a result the background error distributions may be highly non-Gaussian.

This talk presents three idealised models designed to capture the nonlinearity and non-Gaussianity of cumulus convection, which may be more suitable for exploring data assimilation algorithms than some of the simple models traditionally used.

A minimal representation of spatial intermittency and the late detection problem can be obtained by allowing clouds to appear randomly on a lattice, then persist for a finite but also random lifetime. The first model is therefore simply a spatial birth-death process, with clouds appearing and disappearing stochastically at specified rates. A full description of the model, and some results of data assimilation experiments using a Local Ensemble Transform Kalman Filter and a simple Sequential Importance Resampling particle filter are given by [1], and will not be repeated here.

The main limitation of the simple stochastic model described in the the prevous paragraph is the use of a single variable which has no spatial correlations. As a result there are no dynamical balances relating the components of the state vector, in contrast to a model based on the equations of fluid motion. The mass and temperature perturbations that cumulus clouds introduce into their environment are communicated in space by radiation of gravity waves. Neighbouring clouds may be initiated or enhanced by uplift, or suppressed by downward motion associated with the waves.

The primary importance of gravity waves suggests a model base around shallow water equations:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} = K\frac{\partial^2 u}{\partial x^2} \tag{1}$$

$$\frac{\partial H}{\partial t} + \frac{\partial (uH)}{\partial x} = K\frac{\partial^2 H}{\partial x^2}, \tag{2}$$

where u is the fluid velocity, and the geopotential $\phi = g(H - H_0)$ is defined as the difference between the fluid depth H and its mean value H_0, multiplied by the gravitational acceleration g. A diffusion term with constant K has been included in both equations.

To provide a representation of cumulus convection, the shallow water equations must be extended to include conditional instability, i.e. positive buoyancy due to latent heat release in ascending, saturated air. This will be accomplished by modifying the geopotential ϕ. The standard definition of ϕ is based on the gradient of the height of the fluid surface H, and provides a momentum forcing away from regions of increased surface height. This will be altered when height exceeds a threshold H_c, representing the level of free convection, with the geopotential replaced by a fixed negative value $-\phi_c$. The gradient of geopotential will thus

force fluid into the region of negative geopotential, increasing the fluid depth there. Once such a "cloud" forms, the width of the negative geopotential region will collapse until limited by diffusion, as occurs for cumulus clouds in simulations with kilometre-scale weather prediction models.

Once a cloud is formed, the fluid level would continue to rise until the height gradient becomes so strong that diffusion prevents further gowth. This is less realistic, because in nature the lifetime of a cumulus cloud in an unsheared environment is limited by the formation of heavy precipitation particles that eventually overcome the postive buoyancy and turn the updraft into a downdraft. This effect will be mimicked by adding a rain water mass mixing ratio r, with units of kg of water per kg of air, to the geopotential. A separate conservation equation for r is then added, with source and sink terms. Rainwater is produced when the fluid level exceeds a threshold value H_r, and is rising (u has positive convergence). The rain production threshold is set higher than the threshold for buoyancy to ensure that rain production is delayed relative to the onset of the cloud circulation. Removal of rain by precipitation is modelled by a simple linear relaxation towards zero.

The modified shallow water equations are:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial(\phi + r)}{\partial x} = K\frac{\partial^2 u}{\partial x^2}, \tag{3}$$

where

$$\phi = \begin{cases} -\phi_c, & H > H_c \\ g(H - H_0), & \text{otherwise} \end{cases} \tag{4}$$

The continuity equation is

$$\frac{\partial H}{\partial t} + \frac{\partial(uH)}{\partial x} = K\frac{\partial^2 H}{\partial x^2}, \tag{5}$$

and the equation for rain water is

$$\frac{\partial r}{\partial t} + u\frac{\partial r}{\partial x} = K\frac{\partial^2 r}{\partial x^2} - \alpha r - \begin{cases} \beta\frac{\partial u}{\partial x}, & H > H_r \text{ and } \frac{\partial u}{\partial x} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Note that the equations are written for a single horizontal dimension x, but can be trivially extended to two horizontal dimensions.

Values for the various parameters are chosen to produce realistic space and time scale for the model clouds. The following list suggests values based on their primary effect on the resulting simulations:

- $H = 90$ m is chosen to give a gravity wave speed of 30 m/s, typical for the gravest internal mode in the troposphere,
- $H_c = 90.2$ m and $\phi_c/g = 0.1$ m give a reasonable cloud fraction of a few percent, and a time for the cloud to develop to full height of about 0.5 hr,
- $H_r = 90.3$ m and $\beta = 0.0005$ imply a lag to rain formation of about 15 min, and a cloud lifetime of about 2 hr, and
- $\alpha = 0.0012\ \mathrm{s}^{-1}$ gives persistence of rain and associated negative buoyancy after collapse of the height perturbation.

The value of K is chosen mainly for numerical smoothness and depends on the resolution of the discretized equations.

To produce a cloud field for data assimilation experiments, it is necessary to provide some kind of trigger to initiate clouds by lifting the fluid surface across the threshold H_c. The talk will include results for a baseline experiment where spatially-localised convergent perturbations were applied randomly to u throughout the simulation. Alternatively, clouds could be triggered by including a bottom topography in the shallow water equations that would produce uplift as the fluid flowed over. When the continuous random triggering is applied, realistic interactions between clouds are seen, with strong clouds suppressing weaker neighbours. The rain produced by a cloud outlives the positive height perturbation, leaving a region of negative buoyancy that forces gravity waves that can trigger new clouds. For data assimilation synthetic observations can be produced for rain or wind, or wind conditioned on presence of rain as observed by Doppler radar.

The final level in the heirarchy of idealised convection models proposed here is the use of a full numerical weather prediction mode in a simplified configuration. In particular, assimilation experiments have been carried out using the COSMO model, which is run operationally by the Deutsche Wetterdienst (DWD) with a 2.8 km horizontal resolution, allowing the explicit simulation of cumulus clouds. Ordinarily experiments must be repeated over a large number of events to obtain reliable results about the performance of a data assimilation system, which is computationally expensive. To allow shorter test periods, the model has been configured to used periodic boundary conditions and a uniform land surface. This results in an ensemble of clouds that are statistically identical. Of course this framework does not allow the data assimilation system to be evaluated over the full range of weather conditions that would be encountered in operational use, and eventually tests must be carried out using the full system. However, it is anticipated that a hierarchy of models of increasing realism will enable new ideas to be more efficiently investigated.

References

[1] G.C. Craig, M. Würsch, *The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model*, Quart. J. Roy. Meteor. Soc. DOI: 10.1002/qj.1980.

Lagrangian and en-route data assimilation

Amit Apte

(joint work with Elaine Spiller, Christopher K.R.T. Jones)

Assimilation of data collected by Lagrangian or pseudo-Lagrangian instruments such as drifters, gliders, or floats is of great interest, especially for understanding ocean dynamics. In particular, assimilation of subsurface data such as temperature or salinity is of special importance because of paucity of such data. The main difficulty in using such data is that the positions at which the subsurface

measurements are taken in unknown. Our recent work introduces an observation operator which allows us to express these observations as a function of the state vector of the model at the measurement time. This is possible using the augmented state space introduced by Ide et al. [1, 2] for assimilation of the position data of these Lagrangian instruments. We call this proposed methodology *en-route data assimilation.*[3] Once the subsurface measurements are expressed as function of state space, any existing data assimilation scheme can be used to assimilate these data. We show the efficacy of the en-route data assimilation scheme using Markov Chain Monte Carlo sampling method and also using a particle filter (with resampling). Our numerical results with the linear shallow water (LSW) model[4] show that incorporation of these subsurface data is highly informative for estimating the velocity flow and the height field of the LSW model. We also discuss the particular property of the en-route data assimilation, namely the highly nonlinear nature of the observation operator we propose. In the second half of the talk, I discuss our recent work [5] on understanding the effects of the two most basic features of many dynamical systems: (i) the nonlinear shear, i.e., the differential rotation of trajectories around a stable fixed point, and (ii) the nonlinear differential divergence of trajectories near a saddle or hyperbolic fixed point. These two basic features lead to presence of non-Gaussian distribution functions in the assimilation process, in particular, non-Gaussian prior distributions. We illustrate that this leads to failure of methods such as ensemble Kalman filter. We also discuss two main ways in which these nonlinear effects are less pronounced: (i) by taking a more informative prior distribution, or (ii) by increasing the frequency of observations.

References

[1] , Kayo Ide, L. Kuznetsov, C.K.R.T. Jones,*Lagrangian data assimilation for point-vortex system*, J. Turbulence **3**, (2002).

[2] , L. Kuznetsov, Kayo Ide, C.K.R.T. Jones, *A Method for Assimilation of Lagrangian Data*, Monthly Weather Review **131** (2003), 2247-2260.

[3] Elaine Spiller, Amit Apte, C.K.R.T. Jones, *Assimilating en route Lagrangian observations*, submitted (2012).

[4] A. Apte, C.K.R.T. Jones, A. M. Stuart, *A Bayesian approach to Lagrangian data assimilation*, Tellus A **60** (2008) 336-347.

[5] Amit Apte, C.K.R.T. Jones, *The impact of nonlinearity in Lagrangian data assimilation*, submitted (2012).

Reporter: Jana de Wiljes

Participants

Dr. Javier Amezcua
Department of Meteorology
University of Reading
Earley Gate
PO Box 243
Reading RG6 6BB
UNITED KINGDOM

Prof. Dr. Amit Apte
Tata Institute of Fundamental Research
Centre for Applicable Mathematics
Post Bag No. 6503, GKVK Post Office
Sharada Nagar, Chikkabommasandra
Bangalore 560 065
INDIA

Prof. Dr. Angelika Bunse-Gerstner
Fachbereich 3
Mathematik und Informatik
Universität Bremen
Bibliothekstr. 1
28359 Bremen
GERMANY

Dr. Alberto Carrassi
IC3 - Institut Català de Ciènces
del Clima
CFU - Clima Forecasting Unit
Career del Doctor Trueta, 203
08005 Barcelona
SPAIN

Yuan Cheng
Institut für Mathematik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
GERMANY

Nawinda Chutsagulprom
Institut für Mathematik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
GERMANY

Prof. Dr. George Craig
Institut für Theoretische
Meteorologie
Universität München
Theresienstraße 39
80333 München
GERMANY

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

Dr. Sylvain Delahaies
Department of Mathematics
University of Surrey
Guildford
Surrey GU2 7XH
UNITED KINGDOM

Jana de Wiljes
Fachbereich Mathematik & Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Dr. Hendrik Elbern
Rheinisches Institut für
Umweltforschung (RIU) an der
Universität Köln
Aachener Straße 209
50931 Köln
GERMANY

Dr. Grit Freiwald
Alfred-Wegener-Institut for
Polar and Marine Research
27515 Bremerhaven
GERMANY

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

Prof. Andreas Griewank
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Prof. Dr. Eldad Haber
Department of Mathematics
University of British Columbia
121-1984 Mathematics Road
Vancouver BC V6T 1Z2
CANADA

Dr. Patrick Heimbach
Department of Earth, Atmospheric
and Planetary Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge MA 02139
UNITED STATES

Prof. Dr. Andreas Hense
Meteorologisches Institut Bonn
Universität Bonn
Auf dem Hügel 20
53121 Bonn
GERMANY

Prof. Dr. Michael Hinze
Fachbereich Mathematik
Universität Hamburg
Bundesstr. 55
20146 Hamburg
GERMANY

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

Dr. Tijana Janjic-Pfander
Hans Ertel Center for Data assimilation
Theresienstr. 39
80333 München
GERMANY

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

Dr. David Kelly
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM

Prof. Dr. Rupert Klein
Fachbereich Mathematik & Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

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

Dr. Kody J.H. Law
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM

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

Dr. Lewis Mitchell
College of Eng. and Mathematics
University of Vermont
101 Votey Building
Burlington, VT 05405-0156
UNITED STATES

Dr. Lars Nerger
Alfred-Wegener-Institut for
Polar and Marine Research
27515 Bremerhaven
GERMANY

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

Dr. Stephen Penny
Department of Atmospheric
and Ocean Science
University of Maryland
College Park, MD 20742-2424
UNITED STATES

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

Prof. Dr. Robert James Purser
NOAA/NCEP
W/NP2 RM 207
WWBG
5200 Auth Rd.
Camp Springs, MD 20746-4304
UNITED STATES

Prof. Dr. Sebastian Reich
Institut für Mathematik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
GERMANY

Prof. Dr. Markus Reiß
Institut für Mathematik
Humboldt-Universität Berlin
Unter den Linden 6
10117 Berlin
GERMANY

Prof. Dr. Arnd Rösch
Fakultät für Mathematik
Universität Duisburg-Essen
Campus Duisburg
Forsthausweg 2
47057 Duisburg
GERMANY

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

Dr. Seoleun Shin
Institut für Mathematik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
GERMANY

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

Dr. Matthias Sommer
Institut für Meteorologie
Universität München
Theresienstraße 37
80333 München
GERMANY

Prof. Dr. Wilhelm Stannat
Fachbereich Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Nikolai Strogies
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

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

Prof. Dr. Edriss S. Titi
Faculty of Mathematics & Computer Sc.
The Weizmann Institute of Science
PO Box 26
Rehovot 76100
ISRAEL

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

Sebastian Vollmer
Mathematics Institute
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
UNITED KINGDOM