Data Assimilation: New Challenges in Random and Stochastic Dynamical Systems

Sebastian Reich and Andrew M Stuart

This article appeared in the October and November 2015 issues of SIAM News as "Data assimilation in numerical weather prediction" and "Applying Mathematics to Data Assimilation Methods".

The seamless integration of large data sets into sophisticated computational models provides one of the central challenges for the mathematical sciences in the 21st century. When the computational model is based on dynamical systems, and the data set is time ordered, the process of combining models and data is called **data assimilation.** The assimilation of data into computational models serves a wide spectrum of purposes ranging from model calibration and model comparison, all the way to the validation of novel model design principles.

Historically the rise of numerical weather prediction (NWP) in the 1950s played a major role in germinating the field of data assimilation. The computational models employed immediately demanded algorithms for determining initial model states from available observations. Such a task falls naturally within the realm of ill-posed inverse problems [5] with the important caveat that Tikhonov-type regularizations have to be consistent with the underlying model dynamics [2]; indeed it was discovered that forecast skill could be dramatically improved by explicitly including the NWP models into the data assimilation cycle [26]. The data assimilation technique associated with this viewpoint is still widely used in operational weather forecasting and, collectively, the methods go under the synonym of **4DVAR** (standing for four dimensions – three space plus time – and a cost functional to be minimized). The 4DVAR methodology fits into the framework of Tikhonov regularized inverse problems where the regularization term on the initial condition is balanced by a term reflecting faithful reproduction of the model dynamics.

A second class of algorithms widely used by the NWP community are the **Kalman filter** type methods coming out of the control community [12, 13]. This work of Kalman has been enormously influential, constituting an early systematic development of a methodology to combine model and data for dynamical systems; it is applicable to linear problems subject to additive Gaussian noise. An early suggestion to use the Kalman filter in the solution of linear PDEs arising in the atmospheric sciences is [7]. Early extensions of the classic Kalman filter to nonlinear systems include the extended Kalman filter [11]. However computational expense, together with the strong nonlinearity of atmosphere-ocean dynamics, prevented an operational implementation of the extended Kalman filter. Instead operational weather centers implemented a much simplified version of the Kalman update equations through cycling the **3DVAR** methodology [18, 21] (here the data is incorporated sequentially at each fixed time, so that the optimization is over three space dimensions); structurally this cycled 3DVAR looks like an extended Kalman filter type update, but with a fixed covariance structure to weight the model reliabiity versus that of the data. Steadily increasing computer power eventually allowed for an extension of this cycling approach by combining it with **ensemble prediction**, which became prevalent in the NWP community in the 1980s [14]. In this approach, rather than making a single best weather forecast, an ensemble of forecasts is made and their variability is used to weight the reliability of the model in comparison with the reliability of the data. Current ensemble-based data assimilation methodologies rely on linear regression in order to combine forecast uncertainties and observations and are collectively termed **ensemble Kalman filters (EnKFs)** [6]. Employing EnKFs in operational data assimilation systems has led to improved forecast skill compared to the simplified 3DVAR approach (see Figure 1). Whilst EnKFs thus provide an elegant extension of the classic Kalman filter to the highly large-scale non-Gaussian and nonlinear NWP models in use in the present day, the underlying linear regression ansatz also places limitations on their ability to predict, for example, extreme meteorological events.

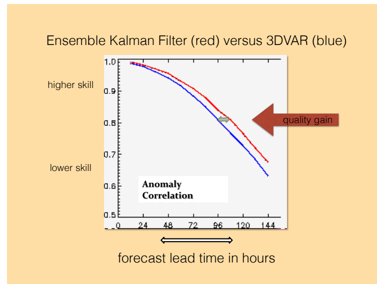

Figure 1. Improvement of forecast skills for temperature in the southern hemisphere through the use of an ensemble Kalman filter data assimilation system at the German Meteorological Service (DWD).

Current research activities in data assimilation for NWP focus on expanding the range of observational systems (see Figure 2) and on merging 4DVAR with ensemble prediction systems on the one hand, and sequential Monte Carlo (SMC, see below) methods with EnKFs on the other. We will discuss some of these developments in more detail below. Practical challenges for such extensions arise, for example, from the relatively small affordable ensemble sizes (on the order of a hundred) and the presence of spatially and temporally correlated data and model errors which cannot be easily represented by standard stochastic processes.

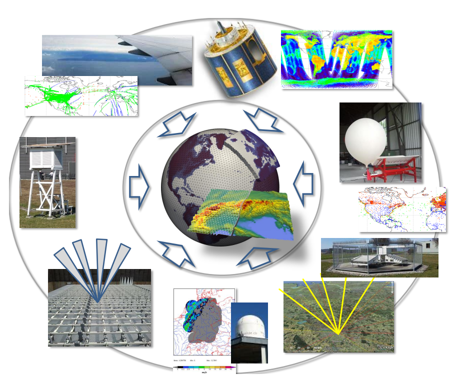

Figure 2. Range of observational systems that deliver data to numerical weather prediction systems.

At the same time, and largely disconnected from NWP, the field of petroleum reservoir simulations has led to the development of data assimilation methods with a stronger focus on combined model state and parameter estimation [20]. With reservoir model parameters, such as permeability, often being hugely uncertain, data assimilation and uncertainty quantification becomes even more challenging for petroleum reservoir engineering.

As the preceding discussion demonstrates, the subject of data assimilation has been driven primarily by practitioners working in the geophysical sciences. However the potential for application in all realms of science and engineering cannot be over-stated. For this reason, the subject is ripe for development by the mathematics community [10]. The primary benefits of mathematizing a discipline of this type are threefold: (i) firstly systematic development leads to clarity about the right questions to ask, and distinguishes between generic algorithmic and mathematical questions, and application-specific ones; (ii) secondly it leads to the possibility of importing algorithmic innovation from the computational mathematics community; (iii) and finally it allows for the exchange of ideas between different application areas, through a common language. Of course, this perspective is not news to most of our readers, but the value of mathematics as the language of science and engineering is always worth re-emphasizing.

A central transferable idea in this article is that, in many areas of applied mathematics, the data model and the mathematical model should be considered *in conjunction*. Thinking about a scientific or engineering problem in this way, from the very start, is certainly a non-traditional way of thinking, but we argue that it is, in many areas, the right viewpoint. In the context of data assimilation we thus consider a combined model for the **signal** with a model for the **observation** process. For expository purposes we consider a discrete time signal $V_J = \{v_\ell\}_{\ell=0}^J$ given by

$$v_{j+1} = \Psi(v_j) + \xi_j.$$

Here the model noise $\{\xi_\ell\}_{\ell=0}^{J-1}$ represents stochastic forcing to a deterministic evolution given by $\Psi(\cdot)$; this stochastic forcing may or may not be included, depending on the setting. The mathematical model for the signal may have many centuries of intellectual development behind it (for example in NWP), or may be the product of more recent applications-driven needs (for example in traffic flow). The level of confidence in the purely deterministic signal model will affect whether or not it is appropriate to include model noise in it. The observations $Y_J = \{y_\ell\}_{\ell=1}^J$ are assumed to be given by

$$y_{j+1} = h(v_{j+1}) + \eta_{j+1}.$$

This will typically model the use of data aquisition instruments, which will of course be application specific (see Figure 2 for NWP examples), but often involving very recent, new, technology. Here the observational noise $\{\eta_\ell\}_{\ell=1}^J$ is almost always present as very few observing instruments are perfect.

We can state two formulations of the data assimilation problem. The first is to find information about v_j given $Y_j = \{y_\ell\}_{\ell=1}^j$, and to update this information sequentially as $j \mapsto j+1$;

this is known as **filtering**. The second is to find information about V_J given Y_J for some given J; this is known as **smoothing**. Computationally smoothing is more demanding than filtering because it operates in a state space of dimension $J+1$ times that of the state space of filtering. While filtering and smoothing lead, theoretically in the fully probabilistic model described below, to the same result at $j=J$, current computational implementations of smoothing in the form of 4DVAR and filtering in the form of EnKFs often demonstrate that smoothing is more informed by the data than is filtering. However EnKFs deliver an estimate for forecast uncertainties and do not require the computation of adjoint operators (and can thus be seen as derivative-free minimization methods). Merging the advantages of 4DVAR with those of EnKFs is currently a very active area of research in NWP.

Another important methodological distinction is between **deterministic** and **probabilistic methods**. Deterministic methods for smoothing can be formulated through optimization as attempting to find the model and observational noise sequences which give the best fit to the overall mathematical/data model. This leads to the 4DVAR objective function

$$J(V_J) := \sum_{j=0}^{J-1} \Big(|C^{-\frac{1}{2}}\big(v_{j+1} - \Psi(v_j)\big)|^2 + |\Gamma^{-\frac{1}{2}}\big(y_{j+1} - h(v_{j+1})\big)|^2 \Big)$$

which will typically be augmented with a regularization term for the initial condition, as discussed above. The *covariance matrices* C and Γ weight the relative confidence in the mathematical model and in the data. There are many variants on the above and, in particular the singular limit $C \to 0$, where the model is thought to be noise free ($\xi_j \equiv 0$), and hence optimization is over v_0 only, is widely used. Deterministic methods for filtering can also be expressed in terms of optimization; 3DVAR type methods have the form:

$$v_{j+1} = \operatorname{argmin}_v J_j(v),$$
$$J_j(v) := |C_j^{-\frac{1}{2}}\big(v - \Psi(v_j)\big)|^2 + |\Gamma^{-\frac{1}{2}}\big(y_{j+1} - h(v)\big)|^2.$$

As with smoothing the choice of covariances C_j and Γ leads to a variety of different methods. A key question is whether such methods can reproduce accurate estimates of the true signal, even when initialized incorrectly; a dynamical systems perspective on such problems was established in the paper [8] in the noise free model and data case. EnKF methods employ N copies of the above iterated minimization in parallel, and the covariance C_j is estimated empirically from an ensemble of forecasts. Such methods provide a transition from deterministic to probabilistic data assimilation techniques in that the ensemble information may be used as a surrogate for model uncertainty. Furthermore, such methods also lead to complex interaction between the different ensemble members, and hence to interesting and challenging problems in random dynamical systems [15]: there is a great deal of scope for new research in this area.

More generally, probabilistic filtering methods concern approximation of the sequence of probability measures $\mu_j(\cdot) = \mathbb{P}(v_j \in \cdot | Y_j)$. There are various approaches to this, but the most prevalent for low-dimensional applications are SMC methods which attempt to approximate the probability distributions μ_j by weighted sums of Dirac measures. This can be very hard to do in problems where the state space dimension is large, or where the data is very informative [25, 27]. EnKF methods partially address the need to tackle such problems by employing linear regression during each data assimilation step, but rigorous analysis justifying their accuracy in practical scenarios (fixed, small ensemble size) is very much lacking and very much required. An interesting connection between probabilistic filtering and optimal transportation theory [22, 23] provides an important conceptual foundation for the analysis of these problems.

The smoothing distribution requires study of the probability measure $\mu(\cdot) = \mathbb{P}(V_J|Y_J)$. This measure is on a space of dimension $J+1$ times that of the space where each measure μ_j from filtering lives. As a consequence it can be very difficult to study this probability measure accurately and efficiently. Monte Carlo-Markov chain (MCMC) methods can be used in some cases, but these are primarily for model problems in benchmarking mode [16]; there remains a significant number of challenging questions in numerical analysis and statistics concerning how to make these methods accurate and efficient for high-dimensional applications [4].

Data Assimilation is at a very exciting juncture for mathematical scientists. There are a plethora of applications in which dynamical models are confronted with significant data sets. The question of how to merge the dynamical model with the data in order to either estimate model states or model parameters, or to estimate both, is thus very timely. In addition to the legacy applications in the geophysical sciences [2, 14, 20], for which data assimilation remains key, new areas include traffic flow [28, 9], neuroscience [1], personalized medicine [24] and power grids [3]. Furthermore, the subject has been applications-led to date. The opportunity for mathematical scientists to systematize the field, to develop and import new ideas and algorithms, and to export these into application domains old and new, is a great one. The recent texts [17, 19, 23] provide introductions to the mathematical underpinnings of data assimilation. The field is one which will only grow in importance over the next few decades, and an ideal one for younger researchers in the mathematical sciences.

Acknowledgements AMS is grateful to EPSRC, ERC, and ONR for financial support that led to the research underpinning this article. SR acknowledges support under the DFG Collaborative Research Center SFB1114: Scaling Cascades in Complex Systems. The authors also thank Roland Potthast (Head of Data Assimilation, German Meteorological Service & Professor for Applied Mathematics, University of Reading) for providing the images in Figures 1 and 2.

References

[1] H. Abarbanel. *Predicting the Future: Completing Models of Observed Complex Systems.* Springer, 2013.

[2] A. Bennett. *Inverse Modeling of the Ocean and Atmosphere.* Cambridge, 2002.

[3] S. Backhaus and M. Chertkov. Getting a grip on the electrical grid. *Physics Today* 66(5): 42–48, 2013.

[4] S.L. Cotter, G.O. Roberts, A.M. Stuart and D. White. MCMC methods for

functions: modifying old algorithms to make them faster. *Statistical Science*, 28:424–446, 2013.

[5] H.K. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, 1996.

[6] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter.* Springer, 2006.

[7] M. Ghil, S.E. Cohn, J. Tavantzis, K. Bube, and E. Isaacson. Application of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil and E. Källén, Eds. Springer, 139–224, 1981.

[8] K. Hayden, E. Olson, and E. Titi. Discrete data assimilation in the Lorenz and 2d Navier-Stokes equations. *Physica D: Nonlinear Phenomena*, 2011.

[9] J.C. Herrera and A.M. Bayen. Incorporating of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B*, 44:460–481, 2010.

[10] K. Ide and C. Jones. Special issue on the mathematics of data assimilation. *Physica D*, 230:vii–viii, 2007.

[11] A.H. Jazwinski. Stochastic Processes and Filtering Theory. Academic Press, 1970.

[12] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[13] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(3):95–108, 1961.

[14] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge, 2003.

[15] D. T. B. Kelly, K. J. H. Law, and A. M. Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.

[16] K. J. H. Law and A. M. Stuart. Evaluating data assimilation algorithms. *Mon. Wea. Rev.*, 140:3757–3782, 2012.

[17] K.J.H.Law, A.M. Stuart and K.C.Zygalakis. *Data Assimilation: A Mathematical Introduction.* Springer, 2015.

[18] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders. The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. R. Met. Soc.*, 126(570):2991–3012, 2000.

[19] A. Majda and J. Harlim. *Filtering Complex Turbulent Systems.* Cambridge University Press, 2012.

[20] D. Oliver, A. Reynolds, and N. Liu. *Inverse Theory for Petroleum Reservoir Characterization and History Matching.* Cambridge Univ Pr, 2008.

[21] D. F. Parrish and J. C. Derber. The national meteorological centers spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, 120(8):1747–1763, 1992.

[22] S. Reich. *A dynamical systems framework for intermittent data assimilation.* *BIT Numerical Mathematics*, 51(1): 235–249, 2011.

[23] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation.* Cambridge University Press, 2015.

[24] M. Sedigh-Sarvestani, D.J. Albers and B.J. Gluckman. *Data assimilation of glucose dynamics for use in the intensive care unit.* Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 5437–5440, 2012.

[25] T. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev..*, 136:4629–4640, 2008.

[26] P. Talagrand, O. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory. *Quart. J. Royal Met. Soc.*, 113:1311–1328, 1987.

[27] P.J. Van Leeuwen, Y. Cheng and S. Reich. *Nonlinear Data Assimilation.* Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Volume 2, Springer-Verlag, 2015.

[28] D.B. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli and A.M. Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 1:1–35, 2010.

Sebastian Reich is Professor of Numerische Mathematik at the Institut für Mathematik at the Universität Potsdam, Germany, and the Department of Mathematics and Statistics, University of Reading, UK.

Andrew Stuart is a Professor of Mathematics at the Mathematics Institute, Warwick University, UK. This article is based, in part, on a lecture delivered by Stuart at the 2015 SIAM Conference on Applications of Dynamical Systems, held in Snowbird, Utah.

Institut für Mathematik, Universität Potsdam, Am Neuen Palais 10, D-14469 Potsdam, Germany
E-mail: sereich@rz.uni-potsdam.de

Mathematics Institute, Warwick University, Coventry CV4 7AL, UK
E-mail: a.m.stuart@warwick.ac.uk