AN INTRODUCTION TO DATA ASSIMILATION

AMIT APTE

Abstract. This talk will introduce the audience to the main features of the problem of data assimilation, give some of the mathematical formulations of this problem, and present a specific example of application of these ideas in the context of Burgers' equation.

The availability of ever increasing amounts of observational data in most fields of sciences, in particular in earth sciences, and the exponentially increasing computing resources have together lead to completely new approaches to resolving many of the questions in these sciences, and indeed to formulation of new questions that could not be asked or answered without the use of these data or the computations. In the context of earth sciences, the temporal as well as spatial variability is an important and essential feature of data about the oceans and the atmosphere, capturing the inherent dynamical, multiscale, chaotic nature of the systems being observed. This has led to development of mathematical methods that blend such data with computational models of the atmospheric and oceanic dynamics - in a process called data assimilation - with the aim of providing accurate state estimates and uncertainties associated with these estimates.

This expository talk (and this short article) aims to introduce the audience (and the reader) to the main ideas behind the problem of data assimilation, specifically in the context of earth sciences. I will begin by giving a brief, but not a complete or exhaustive, historical overview of the problem of numerical weather prediction, mainly to emphasize the necessity for data assimilation. This discussion will lead to a definition of this problem. In section 2, I will introduce the "least squares" or variational approach, relating it to the Kalman filter and the full-fledged Bayesian approach. In the final section, I will introduce the Burgers' equation for which I will present some results and ongoing research on variational and Bayesian approaches to data assimilation.

1. Data assimilation in the context of earth sciences

In order to highlight the need for incorporating observational data into numerical models of the atmosphere and the oceans, I will give a brief historical account of numerical weather prediction, using it to drive towards a definition of data assimilation that I find most convenient to keep in mind. Extensive descriptions of this history are available from various sources such as [14, 15, 12, 16] and the account below is far from complete, serving only to set the context for data assimilation.

It was Vilhelm Bjerknes who was the first one to develop, in 1904, the idea of predicting weather using the hydrodynamic and thermodynamic equations for the atmosphere. Note that this was around 90 years after Laplace discussed[13] the concept of what is now commonly known as *Laplace's deamon*: "Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes." This is one of the first but still most complete statements of determinism, ironically in an essay on probabilities! Bjerknes' program was also formulated (i) around 80-60 years after the formulation of Navier-Stokes equations for viscous flow and the equations of thermodynamics, (ii) around the same time as Poincaré's explorations of chaotic motion in celestial dynamics, but (iii) at least 60 years before the implications of chaos came to be appreciated widely by the scientific community.

The first actual attempt at executing Bjerknes' idea was made by Lewis Fry Richardson about a decade later. He basically attempted to solve the partial differential equations for the atmosphere by dividing Europe into 200km by 200km blocks. The actual calculation for a six hour forecast took him weeks of calculation

Date: April 2, 2014.

by hand! And at the end of it all, his prediction for the change in pressure in six hours period was larger by around 100 times than the actually observed change in pressure. Some of Richardson's comments that are of relevance to us were as follows:[17]

(1) "It is claimed that the above [predictions] form a fairly correct deduction from a somewhat unnatural initial distribution" – I would interpret this as pointing out the need for accurate and smooth initial condition for a good weather forecast.
(2) "... 64,000 computers would be needed to race the weather for the whole globe. ... Imagine a large hall like a theatre ... The walls are painted to form a map of the globe. ... A myriad computers are at work upon the weather of the part of the map where each sits ... The work is coordinated by an official of higher rank ..." – essentially he was foreseeing the use of supercomputers and parallelization of computational tasks in attempting to solve this problem.

There is also another reason for the failure of Richardson's attempt: the sparsity and non-smoothness of the observational data available to him. When his experiment was reproduced[15] with smoothed version of the same data, the prediction was quite accurate. In the context of data assimilation, the so called "background" or the prior that we will discuss in detail later in section 2 provides a systematic way of achieving this "smoothing."

Over decades, several new advances have been made in addressing the problem of weather prediction. The most notable of these are: (i) the realization by Charney and others of the importance of quasi-geostrophy and the subsequent development of models based on these ideas; (ii) the use of electronic computers, pioneered by von Neumann in the 1950s, leading to ever increasing resolution for these models; (iii) improved observations, including explosion in the number of observations from satellites; and last but the most relevant to us (iv) improved mathematical methods for incorporating these observations into the models – this is the domain of data assimilation.

One of the fundamental characteristics of the atmospheric and oceanic dynamics is its chaotic nature, which is manifested in the sensitivity to the initial conditions and puts severe restrictions on the predictability of these systems. Chaotic systems have been studies extensively for a little over a hundred years, beginning with Poincaré's work on Hamiltonian chaos, and continuing with the work of Cartwright and Littlewood, Ed Lorenz's famous study of the "Lorenz system" and many, many others. One important implication of the presence of chaotic solutions is that for predicting the state of systems such as the atmosphere, it is necessary to continually use observational data in order to "correct" the model state and bring it closer to reality.

1.1. **Incomplete models; Inaccurate, insufficient data.** Models of the atmosphere and the oceans, or I would argue, of *any* physical system in general, are necessarily incomplete, in the sense that they do not represent all the physical processes in the system, even though the models are based on sound and well-tested physical theories of nature. For example, it is well known that processes associated with convection and cloud formation are not captured well by models, and even with the highest foreseeable resolution, there will be features of these processes that cannot be captured. Arguably, such inadequacies also apply to models of even "simple, low-dimensional" chaotic systems such as nonlinear electronic circuits, but I will not delve into this point further.

Additionally, even if we have an atmospheric model which is "perfect" (without precisely defining what I mean by "perfect"), if we do not use observational data, we will only be able to make qualitative and quantitative statements about "generic" states on the chaotic attractor of the model, or statistical statements about averages of quantities. Without observational data, we will be unable to make predictions of states of the atmosphere, or compare the time series of a specific variable such as temperature with observational time series. In this sense, data provide a crucial link to reality. More specifically, we hope that the data will help us make statements about one specific trajectory of the atmosphere, out of the infinitely many trajectories that comprise the chaotic attractor of the system. It is important to keep in mind that it is entirely unclear whether a single trajectory, or even a large ensemble of trajectories, of an incomplete and imperfect model of a chaotic system can provide useful predictive information about the real state of such a system, but again I will not dwell on this point further.

The problem of how the data can be used quantitatively is far from trivial since the data themselves are (i) inaccurate, in the sense that they contain inherent measurement errors and noise, and (ii) insufficient,

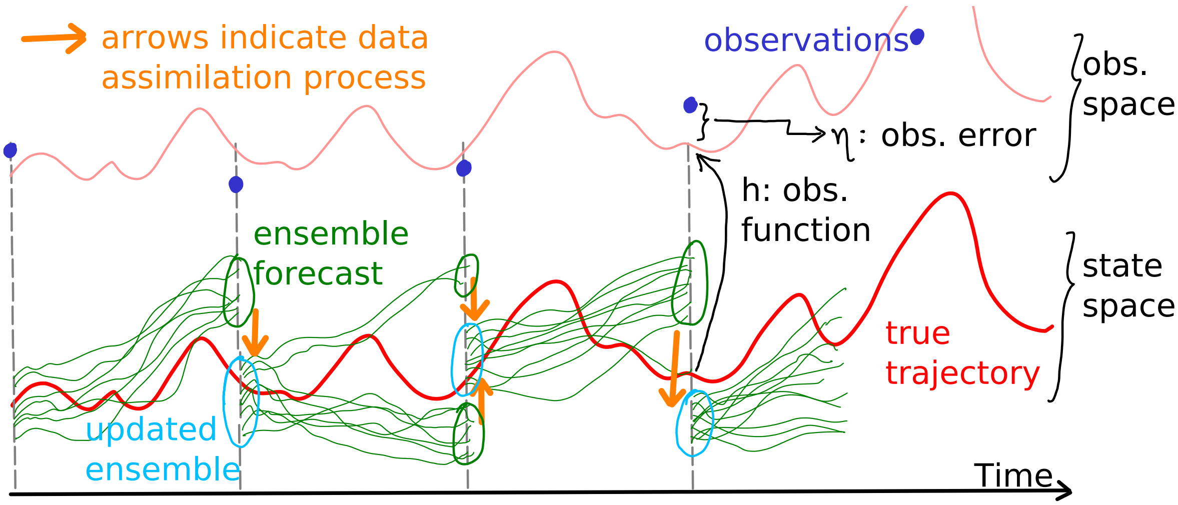

FIGURE 1. A schematic representation of data assimilation process

in the sense that they do not completely specify the state of the system by themselves. For example, most atmospheric and oceanic data are not uniformly distributed in space or time, and are quite sparse compared to the resolution of the numerical models being used. Additionally, there are inaccuracies in the relation of these observational data with the model of the system, which I will refer to as imperfections in the observational model. Thus the problem of determining the state of the atmosphere from the data alone is an under-determined, ill-posed inverse problem.

1.2. **A definition of data assimilation.** The above discussion has hopefully made it clear that models and data must be brought together in order to attack the problems such as weather prediction. This is precisely the main idea behind data assimilation. Thus we could define data assimilation to be the art of

(1) *optimally incorporating*
(2) *partial and noisy observational data* of a
(3) *chaotic and thus necessarily nonlinear (and most often, complex, multiscale) dynamical system*, with an
(4) *imperfect model of the system dynamics and of the observations*, in order to obtain
(5) *predictions of and the associated uncertainty for* the state of the system, repetitively in time.

A caricature of this definition is captured in a schematic shown in Fig. 1. To recap, a few of the characteristics of data assimilation problem are as follows:

(1) We have good physical theories to describe the systems but the models based on these theories are incomplete and imperfect.
(2) The systems are described by nonlinear, chaotic, but usually deterministic dynamical models.
(3) The dynamics contain interactions across multiple temporal and spatial scales, and occur on very high or infinite dimensional state space.
(4) The observational data of the system are
    (a) very high dimensional,
    (b) noisy and inaccurate,
    (c) partial and usually sparse, and
    (d) discrete in time.

I will now discuss some of the mathematical formulations that are used frequently to resolve this problem. It will become clear that each of these formulations takes into account only some, but not all, of the above characteristics. Thus formulating an overarching mathematical framework for data assimilation that addresses all the features above is a formidable and interesting modelling problem.

2. Least squares approach, Kalman filter, and the Bayesian formulation

Consider the problem of estimating n quantities $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ using m noisy observations $y = (y_1, \dots, y_m) \in \mathbb{R}^m$, which are related to the unknowns x through *observation function* (or matrix, in case of linear relation): $h(\text{or } H) : \mathbb{R}^n \to \mathbb{R}^m$. I.e. in the absence of noise, the observations $\hat{y}$ of unknowns x would be $y = h(x)(\text{or } = Hx)$. We assume that $m < n$ (and typically $m \ll n$). Thus even in the linear case, H will not be an invertible matrix.

For example, x could be the velocity field on a model grid whereas y could be the velocity measurement at locations which may or may not be on the same grid. In this case, H will be the matrix for projection or interpolation between model grid and measurement grid.

The "least squares" formulation of this problem is to find the minimum x^a of the following *cost function*:

$$\tilde{J}(x) = \frac{1}{2}\|y - Hx\|_R^2 \,, \tag{1}$$

where the norm $\|z\|_R^2 = z^T R^{-1} z$ simply gives different weights to different observations. Of course, since H is not invertible, we immediately see that $\nabla \tilde{J} = -H^T R^{-1}(y - Hx^a) = 0$ cannot be solved for x^a, we need to "regularize" this problem. This can be achieved by, e.g., modifying the cost function to

$$J(x) = \tilde{J}(x) + \frac{1}{2}\|x - x^b\|_{P^b}^2 \tag{2}$$

where x^b is some "background" (or *a priori*) information about x and the norm $\|\cdot\|_{P^b}$ again gives different weights to different components of this background state x^b. This cost function is minimized by solving $\nabla J = (H^T R^{-1} H + (P^b)^{-1})x^a - (H^T R^{-1} y + (P^b)^{-1} x^b) = 0$ and the solution is given by

$$x^a = x^b + \left[H^T R^{-1} H + (P^b)^{-1}\right]^{-1} H^T R^{-1}(y - Hx^b) \,. \tag{3}$$

Here $I := (y - Hx^b)$ is called the *innovation* vector and the prefactor of innovation

$$K = \left[H^T R^{-1} H + (P^b)^{-1}\right]^{-1} H^T R^{-1} = P^b H^T \left(HP^b H^T + R\right)^{-1} \tag{4}$$

is called the "Kalman gain matrix." Eq. (3) may also be rewritten as $x^a = x^b + K\left(y - Hx^b\right)$ to make it clear that the *analysis* x^a is a linear combination of the *background* x_b and the innovation I. (Note that the final equality in Eq. (4) is obtained by using the Sherman-Morrison-Woodbury identity.)

This calculation can be extended to the case when the unknown x is the initial condition of a dynamical system whereas the observations are spread over time. We will now see that such a setup leads to the Kalman filter.

2.1. Observations over time - Kalman filter.

Consider a *linear* dynamical model, $x_{i+1} = m(x_i) = Mx_i$ on $\mathbb{R}^n$. The initial condition x_0^+ is unknown and we wish to determine it based on some observations $y_i = h_i(x_i^+) = H_i x_i^+$ (plus noise) for $i = 1, \dots, N$. Here x_i^+ is the trajectory starting with the initial condition x_0^+, i.e., $x_i^+ = M^i x_0^+$. In parallel with our approach above, we can now consider a cost function

$$J(x_0) = \sum_{i=1}^{N} \frac{1}{2}\|y_i - H_i M^i x_0\|_{R_i}^2 + \frac{1}{2}\|x_0 - x^b\|_{P^b}^2 \,, \tag{5}$$

where the last term is the "background" term that regularizes the minimization problem. This is a quadratic function of x_0 and thus has a unique minimum which can be obtained by setting $\nabla_{x_0} J = 0$. But we will now consider the following iterative way of calculating this minimum, defined by a two stage process:

1. The "forecast" step gives the dynamical evolution of the state from step $k-1$ to next step k:

$$x_k^f = Mx_{k-1}^a \qquad \text{and} \qquad P_k^f = MP_{k-1}^a M^T \,. \tag{6}$$

2. The "update" step gives the "analysis" x_k^a as a linear combination of the observation y_k at step k with the forecast x_k^f:

$$x_k^a = x_k^f + K_k I_k \qquad \text{and} \qquad P_k^a = (I - K_k H_k) P_k^f \,, \tag{7}$$

   where $I_k = y_k - H_k x_k^f$ is the innovation and $K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}$ is the Kalman gain.

This iteration is begun with $x_0^a = x_0^b$ and $P_0^a = P_0^b$ at $k = 1$ and continues until $k = N$. At the end of these N steps, we get the final "analysis" given by (x_N^a, P_N^a). This two-step process is known as the Kalman filter.

The main relation of the result of this two-step process to the least squares problem of minimization of the cost function is as follows: Suppose x_0^m is the minimum of the cost function $J(x_0)$ from Eq. (5), and P_0^m be its Hessian. Then the dynamical evolution of this minimum and the Hessian from step 0 to step N given by the following equations,

$$x_N^m = M^N x_0^m \qquad \text{and} \qquad P_N^m = M^N P_0^m (M^N)^T \,, \tag{8}$$

is exactly the same as the analysis of the Kalman filter: $x_N^m = x_N^a$ and $P_N^m = P_N^a$. Thus, Kalman filter provides a way of solving the minimization problem. It is important to note that this equivalence of Kalman filter and the variational approach holds *only* in the case when the dynamical model is linear *and* the observations depend linearly on the state.

In practical problems in earth sciences, there are two main reasons why the Kalman filter is not usable. Firstly, the size of the system n is quite large, usually $n = 10^6$ or more. In these cases, it is impossible to solve the above equations for Kalman filter, since they involve the $n \times n$ covariance matrices P_k^f which are impossible to store and manipulate for such large system sizes. The second equally serious reason is that these systems are non-linear and chaotic as we saw in the previous section. Thus the above equations need to be modified appropriately.

I will only mention the two main modifications that address these two issues separately. The extended Kalman filter (EKF) is designed to work with nonlinear systems which are close to being linear. The ensemble Kalman filter (EnKF) is a set of methods designed to work with an ensemble of states, but without the explicit construction of the covariance matrices P_k^f. The EnKF and the variational methods are two of the most commonly used methods in the earth sciences, but there are several theoretical and practical problems that are still being investigated.

Before moving on to describing the variational and Bayesian approaches to data assimilation in the context of Burgers' equation, I will very briefly introduce the Bayesian framework in the next paragraph

2.2. **Bayesian formulation.** Let us consider a deterministic, discrete time dynamical model for $x \in \mathbb{R}^n$:

$$x_{n+1} = m(x_n) \quad \text{or equivalently} \quad x_n = \Phi(x_0, n) \,. \tag{9}$$

The initial condition x_0 is the unknown, but we will assume that we know a *prior* distribution for it, given by a density $\zeta(x_0)$. We will consider the case when noisy observations y_k at time k depend on the state x_k at that time and contain additive noise η_k:

$$y_k = h(x_k) + \eta_k = h(\Phi(x_0, k)) + \eta_k \,, \qquad k = 1, \ldots, N \,. \tag{10}$$

I will only talk about the so-called *smoothing problem*, which is to assimilate all these observations to get an estimate of the initial condition x_0. For this purpose, the observations are concatenated:

$$y = \{y_k\}_{k=1}^N = H(x_0) + \eta, \quad \text{where} \tag{11}$$

$$H(x_0) = \{h(\Phi(x_0, k))\}_{k=1}^N \quad \text{and} \quad \eta = \{\eta_k\}_{k=1}^N \,. \tag{12}$$

We will assume that the noise η has a density. Then this density indeed gives the conditional probability of the observation y given the initial condition x_0, i.e., $p(y|x_0)$. For example, if $y \in R^m$ and $\eta \sim \mathcal{N}(0, \Sigma)$ (Gaussian observational errors), then

$$p(y|x_0) \propto \exp\left(-\hat{J}(x_0, y)\right), \quad \text{where} \quad \hat{J}(x_0, y) = \frac{1}{2} \left\| y - H(x_0) \right\|_\Sigma^2 \,. \tag{13}$$

But we are really interested in the probability density for the initial condition x_0. This will be obtained as *the posterior probability density* as given by the Bayes' rule:

$$p(x_0|y) = \frac{p(y|x_0)p(x_0)}{p(y)} \quad \text{where} \quad p(y) = \int p(y|x_0)p(x_0)dx_0 \quad \text{is a constant.} \tag{14}$$

In the context of data assimilation,

$$p(x_0|y) \propto \zeta(x_0) \exp\left(-\hat{J}(x_0, y)\right) \tag{15}$$

For uncorrelated errors η_k (Σ is block diagonal with blocks R_k), this becomes

$$p(x_0|y) \propto \zeta(x_0) \prod_{k=1}^{K} \exp\left(-\frac{1}{2}\|y_k - h(\Phi(x_0,k))\|^2_{R_k}\right) \tag{16}$$

Now, we see that if the prior is Gaussian, e.g., $\zeta(x_0) \propto \exp(-\|x_0 - x_0^b\|^2_{P^b}/2)$, then the cost function introduced in Eq. (5) is exactly the logarithm of this density, if the dynamics and the observation function are both linear: $m(x) = Mx$ and $h(x) = Hx$. In this case, the minimum of the cost function of Eq. (5) is the *maximum a posteriori estimate*. Thus we see a close relation between the Bayesian framework, the least squares or the variational approach, and the Kalman filter.

All of the material in this section is discussed extensively in many existing reviews on various aspects of these problems. A fairly incomplete list of references is [8, 4, 5, 12, 11, 6] and references therein. In particular, the last two reference contain an excellent introduction to the relation of estimation theory to data assimilation, explaining in detail the various relations between the mean of the posterior distribution (conditional mean), the minimum variance estimator which in the linear case leads to the best unbiased linear estimator (BLUE) and the Kalman filter. There are several other topics that have a direct bearing on the data assimilation problem: stability and convergence of nonlinear filters and particle filtering, observability of nonlinear dynamical systems, probability measures and Bayes' theorem on infinite dimensional spaces that arise from partial differential equation models, etc. But a short introduction such as this certainly fails to provide a reasonable glimpse to these important relations, many of which are currently active topics of research.

Having introduced some of the main approaches to the data assimilation problem and their interrelations, I will now talk about a specific application in the case of Burgers' equation.

3. Data assimilation for Burgers' equation

We will work with the viscous Burgers' equation

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial z} = \nu\frac{\partial^2 v}{\partial z^2} \quad \text{with} \quad v(t=0,z) = u(z) \quad \text{and} \quad v(t,0) = 0 = v(t,1) \tag{17}$$

on the domain $(z,t) \in \omega \times (0,T)$ with $\Omega = (0,1)$. This is a nonlinear evolution equation which has unique solution: for $u \in H^1_0(\Omega)$, there exist a unique $v \in L^2(0,T;H^1_0(\Omega)) \cap C(0,T;H^1_0(\Omega))$ and I will indicate this map by $v(t) = \Phi(u,t)$. In fact, using Cole-Hopf transform, the exact solution can be written down, though I will not explicitly use this fact in this lecture.

There are several motivations for studying data assimilation problems for the Burgers' equation model. The dynamical models in earth sciences are based on the partial differential equations (PDE) of fluid dynamics of the air and water. These can be considered as infinite dimensional dynamical systems, of which the Burgers' equation is an example. It is also a nonlinear PDE whose solutions can be written analytically. Even though Burgers' equation does not exhibit the dynamical complexity of the atmospheric or oceanic flows, it acts as a toy model that is amenable to mathematical analysis whose qualitative features are still relevant to the more complicated scenarios in realistic applications.

For data assimilation problem in the context of Burgers' equation, I will now describe some of the known results as well as some of the ongoing research for three types of noisy observations. In all these cases, the problem will be determining the initial condition u given some observations, written as $y = H(u) + \eta$, with H being the observation operator. These three types of problems are as follows.

Observations continuous in space at a specific time. In this case, we will observe v for all $z \in \Omega$ at time T: thus $H(u) = \Phi(u,T)$ and η is a Gaussian measure on the Hilbert space $L^2(\Omega)$ supported on some appropriate subspace, e.g., $H^1_0(\Omega)$. Some of the questions of interest in this case are as follows. Suppose we consider a sequence of problems with observational noise $\eta_n = (1/n)\eta_0$ – we have more accurate measurements as n increases. In the limit of $n \to \infty$, this reduces to a classical inverse problem of determining initial condition u from the solution at time T. The study of qualitative and quantitative behaviour of the posterior distribution for the initial condition u which is conditioned on the observations y, and in particular the limit $n \to \infty$, are some of the open problems, known as the problem of "posterior consistency" in the statistical literature. Some related results in context of linear models such as the heat equations and other related results are contained in, for example, [2, 19, 1, 9, 10]. A general overview of Bayesian approach to inverse problems, including extensive discussion of data assimilation is contained in [18].

Observations continuous in time. Here, we observe some function of v at all times $t \in (0, T)$. Thus at any given time, we assume a continuous map $C : H_0^1(\Omega) \to \mathbb{Z}$ into some Hilbert space $\mathbb{Z}$ (the observation space). Thus, if $Z^d(t)$ are the actual observations, then in terms of this map, the cost function for a variational formulation can be written as

$$J(u) = \frac{1}{2} \int_0^T \left\| C(\Phi(u,t)) - Z^d(t) \right\|_{\mathbb{Z}}^2 \, dt + \frac{1}{2} \left\| u - u^b \right\|_{P^b}^2 . \tag{18}$$

The question of uniqueness of minimum of this cost function has been studied for the cases of small enough observational time horizon T and for large time horizon T in [20] and [7] respectively. The probabilistic formulation of this problem and the study of the corresponding posterior distribution is a possible direction for future research.

Observations which are discrete in space and time. For this case, we take the observations of v at discrete space locations $\{z_i\}$ for $i = 1, \ldots, K$ at discrete times $\{t_j\}$ for $j = 1, \ldots, N$: Here, $H(u) = \{\Phi(u, t_j)(z_i)\}_{i=1,j=1}^{i=K,j=N}$ and η is a probability distribution on $\mathbb{R}^{NK}$. We do not yet have any theoretical results either about the minimum of a cost function in the variational formulation or about the behaviour of the posterior density in the Bayesian framwork, e.g., in the case when the observational noise decreases. Some of the numerical results, possibly indicating presence of multiple minima, are presented in [3].

4. Concluding remarks

The main aim of this lecture was to provide a short introduction to the topic of data assimilation, beginning with attempts to predict weather phenomena using numerical solutions of relevant partial differential equations. I discussed some of the most common approaches including variational methods, the variants of Kalman filter, and the Bayesian framework providing interrelations between these. I also discussed some of these in the context of Burgers' equation which provides an example of a nonlinear partial differential equations where these techniques can be studies in great detail. All along, I tried to point out some of the open questions, emphasizing the interdisciplinary nature of data assimilation research.

Acknowledgments

I would like to thank the referee for comments which improved the exposition. A part of the research in the "data assimilation and dynamical systems" group of ICTS is supported by the EADS International Teaching and Research Chair on "Mathematics Of Complex Systems."

References

1. S. Agapiou, *Aspects of bayesian inverse problems*, Ph.D. thesis, University of Warwick, 2013.
2. S. Agapiou, S. Larsson, and A.M. Stuart, *Posterior contraction rates for the bayesian approach to linear ill-posed inverse problems*, (2012), `http://arxiv.org/abs/1203.5753`.
3. A. Apte, D. Auroux, and Mythily Ramaswamy, *Variational data assimilation for discrete burgers equation*, Electronic J. Diff. Eq. Conference **19** (2010), 15–30.
4. A. Apte, C.K.R.T. Jones, A.M. Stuart, and J. Voss, *Data assimilation: Mathematical and statistical perspectives*, Int. J. Numer. Meth. Fluids **56** (2008), 1033–1046.
5. F. Bouttier and P. Courtier, *Data assimilation concepts and methods*, ECMWF Meteorological Training Course Lecture Series, March 1999.
6. Stephen E. Cohn, *An introduction to estimation theory*, J. Met. Soc. Japan **75** (1997), 257–288.
7. G. Cox, *Large-time uniqueness in a data assimilation problem for Burgers' equation*, (2012), `http://arxiv.org/abs/1207.4782`.
8. Geir Evensen, *Data assimilation: the ensemble Kalman filter*, Springer, 2007.
9. J.N. Franklin, *Well-posed stochastic extensions of ill-posed linear problems*, J. Math. Anal. **31** (1970), 682–716.
10. A. Hofinger and H.K. Pikkarainen, *Convergence rates for linear inverse problems in the presence of an additive normal noise*, Stoch. Anal. Appl. **27** (2009), 240–257.
11. A.H. Jazwinski, *Stochastic processes and filtering theory*, Academic Press, 1970.
12. Eugenia Kalnay, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 2003.
13. Pierre Simon Laplace, *A philosophical essay on probabilities*, John Wiley and Sons, 1902, Translated from sixth French edition by F.W. Truscott and F.L. Emory; `https://archive.org/details/philosophicaless00lapliala`.

14. Edward N. Lorenz, *Reflections on the conception, birth, and childhood of numerical weather prediction*, Annual Rev. Earth Planetary Sci. **34** (2006), 37–45, doi:10.1146/annurev.earth.34.083105.102317.
15. Peter Lynch, *The emergence of numerical weather prediction: Richardson's dream*, Cambridge University Press, 2011.
16. Anders Persson and Federico Grazzini, *User guide to ECMWF forecast products*, Tech. report, ECMWF, 2005.
17. Lewis Fry Richardson, *Weather prediction by numerical process*, Cambridge University Press, 1922.
18. A.M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. **19** (2010), 451559.
19. S.J. Vollmer, *Posterior consistency for Bayesian (elliptic) inverse problems through stability and regression results*, (2013), http://arxiv.org/abs/1302.4101.
20. Luther W. White, *A study of uniqueness for the initialization problem for Burgers' equation*, J. Math. Anal. Appl. **172** (1993), 412.

International Centre for Theoretical Sciences (ICTS-TIFR) Bangalore, India
E-mail address: apte@icts.res.in