What is the Point of Data Assimilation when the Model(s) is Wrong?

Hailiang Du and Leonard A. Smith

Centre for the Analysis of Time Series, London School of Economics

Email: *h.l.du@lse.ac.uk*

Abstract

The traditional aim of data assimilation is a point close to "Truth". Ed Lorenz noted this assumes the model-state space and the system state space (if such a thing exists) are similar enough that these two points are "subtractable," providing meaningful distance. The modern aim is an ensemble near "Truth". Yet the mathematical sophistication of modern methods is shown to obscure their lack of internal coherence when the model is wrong.

Introduction

Data assimilation is a formal method of "nowcasting": melding observations and a model into an quantitative estimate of the current state of the system "now." But if the model is wrong, how could any point in the model state space serve as a target for nowcasting? The "Truth" is not out there, in model land. Even if it were, how would we quantify our approximation given that the observations are uncertain? This second question is addressed first, and ensemble nowcasting methods are developed in the Perfect Model Scenario. A new approach is discussed, one which appears more robust when the fact that the model is wrong is admitted. Technical difficulties that arise in this real-world case are noted; a fundamental connection between the assimilation method and the aim of the forecast may be inescapable, if the forecast is to be useful.

Problem Statement

The problem of nowcasting can be cast as forming an ensemble estimate of the current model state given a model and observations. The problem is first tackled within the fictional perfect model scenario, where the mathematical structure of the system is known. The aim of nowcast here is to capture "Truth". The challenge is then raised in the imperfect model scenario (reality) where the model is wrong. The model state space and the system state space are almost certainly different. In general, the property of the projection operator between these two space is unknown, and one may question whether or not "Truth" exists. The goal here is reset to produce relevant nowcast that is consistent with the observations and an informative estimate of the model error to improve the future forecast and model.

Gradient Descent Nowcasting in PMS

Even with a perfect nonlinear model, noisy observations prevent us from identifying the true state of the system precisely. An ensemble algorithm is desired to account the initial condition uncertainty. Such algorithm may generate an ensemble directly or from perturbations of a *reference trajectory*. The Gradient Descent (GD) method [1] (described below) is adopted to locate such reference trajectory.

Methodology

Let the dimension of our model state space be m and the number of observed states used in the assimilation be n; the sequence space is an $m \times n$ dimensional space in which a single point can be thought of as a particular series of n states $\mathbf{u}_i, i = -n+1, .., 0$ where $\mathbf{u}_i$ is a m dimensional vector. Define a *pseudo-orbit* to be a sequence of model states that at each step differ from trajectories of the model, i.e., $\mathbf{u}_{i+1} \neq F(\mathbf{u}_i)$ where F is the mathematical form of the model. The observations themselves projected into the model state space define a pseudo-orbit which, with probability one, will not be a trajectory, call this initial Analysis, A_0. A GD algorithm is applied to minimise the mismatch cost function of a pseudo-orbit $\mathbf{u}_i$:

$$C(\mathbf{u}) = \sum \mid F(\mathbf{u}_i) - \mathbf{u}_{i+1} \mid^2 . \qquad (1)$$

The mismatch cost function has no local minima other than on the trajectory manifold for which $C(\mathbf{u}) = 0$ (Variational method like 4DVAR has exponentially increasing number of local minima as the assimilation window gets longer [4]). In practice, the GD algorithm initialised with A_o and run for a finite time; only a pseudo-orbit is obtained. The model trajectory defined by the middle component of the pseudo-orbit defines the reference trajectory. An ensemble nowcast is then formed by sampling the local space around reference trajectory according the likelihood given the observations.

Ensemble Kalman Filter vs GD

The GD nowcasting results are compared with the results produced by the well-established Ensemble Kalman Filter[6]. Fig 1 represents four nowcast results examples in a 2-D state space. In all panels, the ensemble, produced by GD method is not only closer to the true state but also reflects the structure of the model's attractor as the ensemble members lie near the model attractor. The EnKF ensemble has its own distinctive structure, sometimes (bottom two panels) systematically off the attractor and without covering the true state.

To quantitatively measure the difference between these two methods, we first translate the nowcast ensemble into a predictive distribution function by standard kernel dressing [7], which is then evaluated using the "log p score" (Ignorance Score [8,9]). Comparisons are made in both lower dimensional Ikeda Map and higher dimensional Lorenz96 flow. An even higher dimensional example is noted in Fig 3. In Table 1, the ensemble generated by GD nowcast outperforms the one generated by EnKF in both experiments. Relative Ignorance between two methods is found to be 1.4 bits in Ikeda experiment and 0.7 bits in Lorenz96 experiment; that is as GD places on average, 160% (and 60%) more probability mass on the verification than the EnKF.

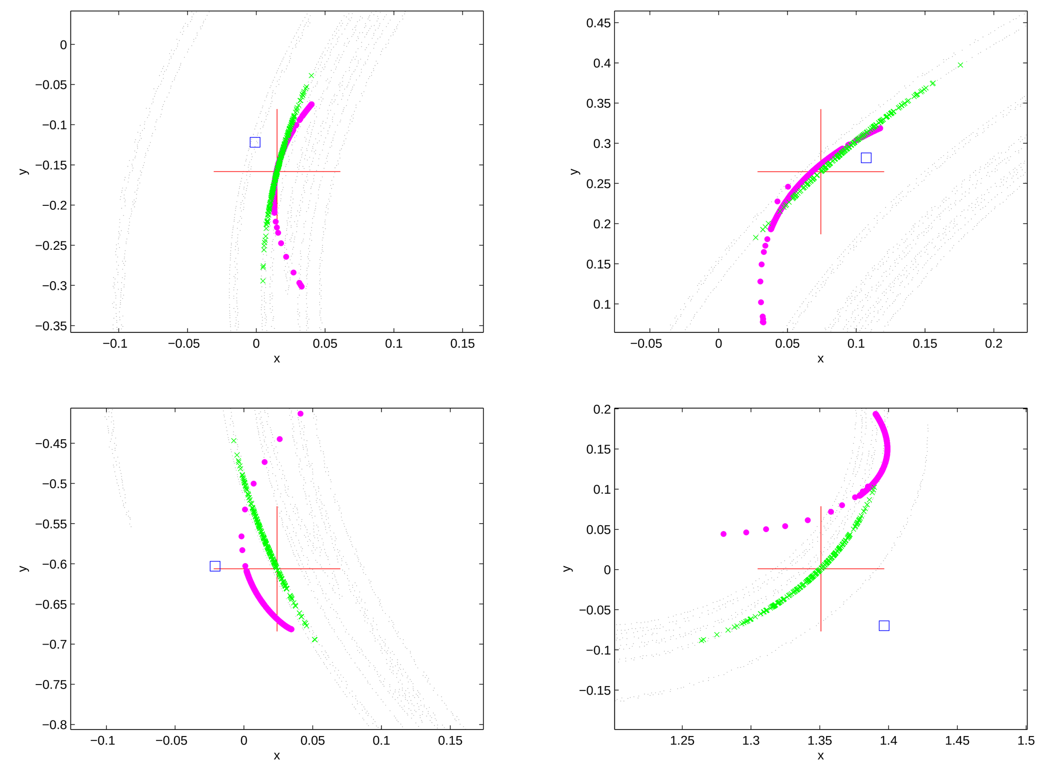

Fig 1: Ensemble results from both EnKF and GD nowcast for the Ikeda Map. The true state of the system is centred in the picture located by the plus sign; the square is the corresponding observation; the background dots indicate samples from the Ikeda Map attractor. The EnKF ensemble is depicted by 512 purple dots. The GD nowcast ensemble is depicted by 512 green crosses. Each panel is an example of one nowcast.

Systems	Ignorance		Lower		Upper	
	EnKF	GD	EnKF	GD	EnKF	GD
Ikeda	-3.21	-4.67	-3.28	-4.75	-3.13	-4.60
Lorenz96	-3.72	-4.44	-3.78	-4.49	-3.66	-4.38

Table 1: Ignorance score of nowcast ensemble for Ikeda Map and Lorenz96 System, the noise model is $N(0,1)$ and $N(0,0.05)$ respectively. Ensemble generated by GD method and Ensemble Kalman Filter are compared. Lower and Upper are the 90 percent bootstrap re-sampling bounds of Ignorance score, and the statistics are calculated based on 2048 assimilations and 512 bootstrap samples are used to calculate the re-sampling bounds.

GD Nowcasting in Reality

All models are wrong. In such Imperfect Model Scenarios, the model state and system state are distinct entities. No trajectory of the model is consistent with an infinite series of observations, there are pseudo-orbits, however, that are consistent with observations and these can be used to estimate the model state. To locate relevant pseudo-orbits, the GD algorithm is applied with a stopping criteria. An informative estimate to the model error is also produced.

It is not possible to disentangle observational noise and model error from each other precisely. Applying the GD method with a stopping criteria can, however, yield less inconsistent (or biased) estimates. Let $\mathbf{u}_i - \mathbf{s}_i$ be the implied noise considered as the estimate of the observational noise; and the mismatch $\mathbf{u}_{i+1} - F(\mathbf{u}_i)$ be the imperfection error considered as the estimate of the model error. Fig 2 shows the statistics of pseudo-orbit changes as a function of the number of iterations of Gradient Descent. By comparing the standard deviation of implied noise with that of the real noise model, it appears that at the beginning of the minimisation, the observational noise is underestimated by the implied noise simply because the minimisation algorithm is initialised at the observations. As the minimisation proceeds, the implied noise becomes more consistent with the observational noise and the pseudo-orbit gets closer to the true pseudo-orbit [1]. After more iterations, however, the implied noise tends to overestimate the observational noise and the distance between the pseudo-orbit and the true pseudo-orbit grows. This is due to the model inadequacy. When the imperfection error of the pseudo-orbit becomes smaller than the actual model error, the implied noise compensates for the imperfection error making implied noise too large and yielding a pseudo-orbit inconsistent with the observations.

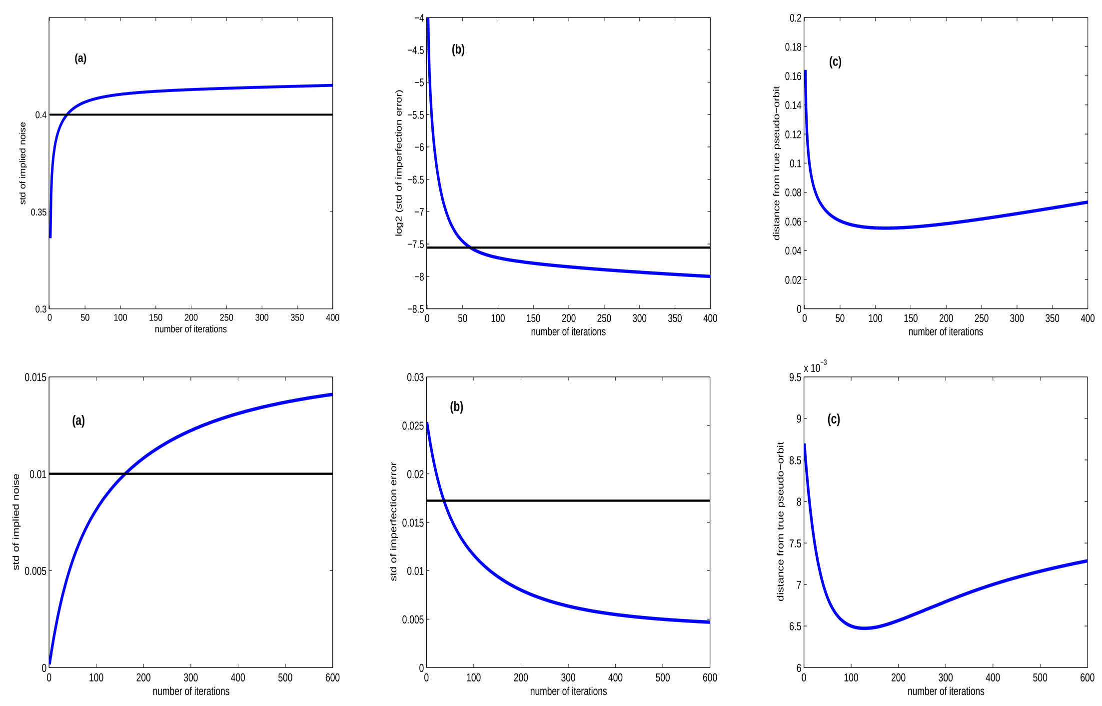

Fig 2: Statistics of the pseudo-orbit as a function of the number of iterations of Gradient Descent iterations for both low dimension Ikeda system-model pair experiment (first row) and higher dimension Lorenz96 system-model pair experiment (bottom row). (a) is the standard deviation of the implied noise (the flat line is the standard deviation of the noise model); (b) is standard deviation of the model imperfection error (the flat line is the sample standard deviation of the model error); (c) is the RMS distance between pseudo-orbit and the true pseudo-orbit.

The GD minimisation with intermediate runs produces more consistent pseudo-orbits. Certain criteria need to be defined in advance to decide when to stop. One can for example define the stopping criteria by testing the consistency between implied noise and the noise model or optimized by the forecast performance.

[1] the system trajectory projected onto the model state space

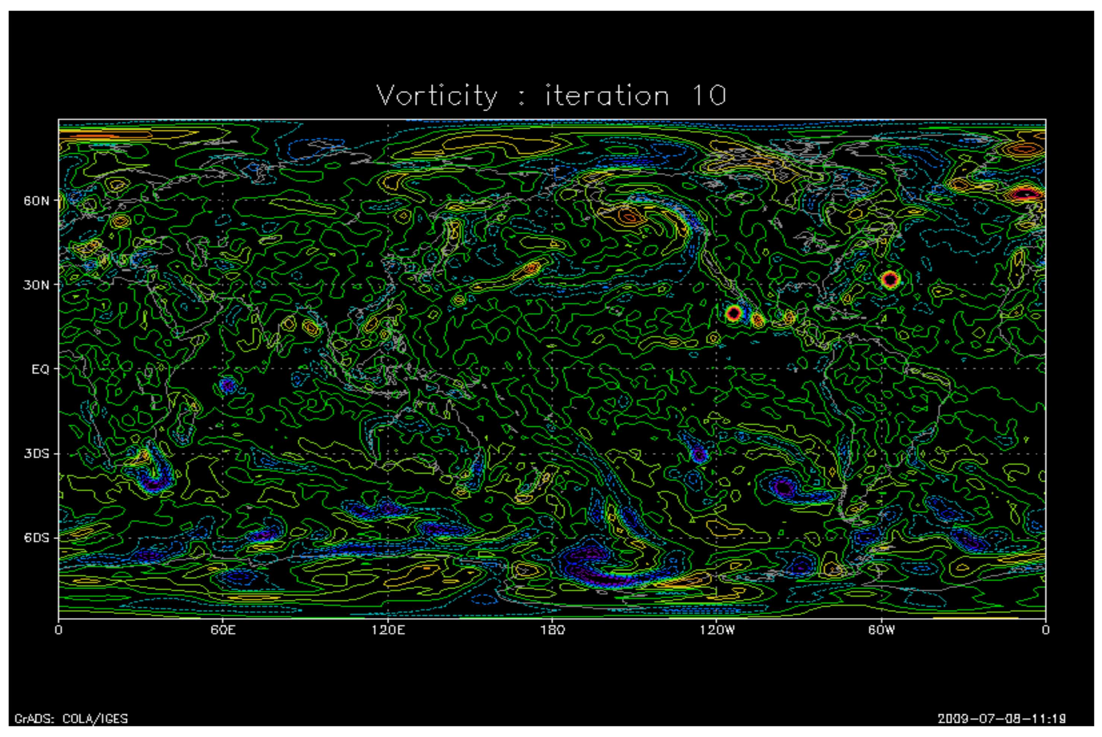

Fig 3: A_0 from a weather model (NOGAP) before GD, showing vorticity.

Using the Imperfection Error

Model error is neither IID nor Gaussian distributed. It is desirable for the estimate of model error to roughly capture spatial correlation and thereby provide useful information to improve the forecast (and model). As the model is wrong, a forecast using the model may be improved by adjustment [10]. For example one can draw at random from the GD imperfection error to produce random adjustment or using analogue methods to provide analogue adjustment. Fig 4 shows six examples of the one step forecast ensemble in the state space. In all examples, forecasts with random adjustment produce ensemble members spread out, as the spatial information is discarded. The first 4 panels present the four cases where the model error is very small, small, moderate and large. Forecasts with analogue adjustment outperform the direct forecast when model error is not negligible. Panel (e) shows an example where the model error is small but the forecast with analogue adjustment is unable to capture the true state. This failure occurs because the model error in this case is overestimated by the imperfection error. Panel (f) shows an opposite example where forecasts with analogue adjustment did not capture the true state because the model error in this case is underestimated by the imperfection error.

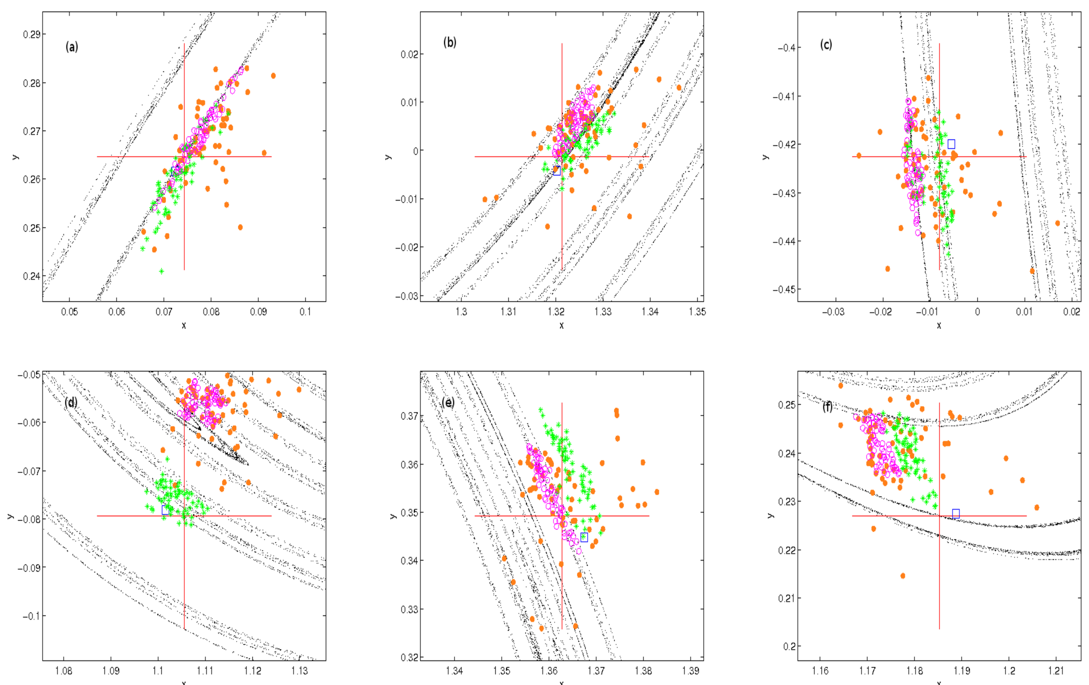

Fig 4: One step forecast ensembles in the state space. Observations are generated by Ikeda Map with IID uniform bounded noise U(0, 0.01). The truncated Ikeda model is used to make forecast. The initial condition ensemble is formed by inverse noise with 64 ensemble members. Four 1-step forecast examples are shown in six panels. In each panel, the background dots indicate samples from the Ikeda Map attractor, the red cross denotes the true state of the system, the blue square indicates the observation, the direct forecast ensemble is depicted by purple circles, the forecast with random adjustment ensemble is depicted by orange dots and the forecast with analogue adjustment ensemble is depicted by cyan stars.

Summary

An ensemble data assimilation method based on GD [1] is introduced. Improved performance comes at the cost of not being a one-step method, which, on the other hand, allows an enhanced balance between the extracting information from the dynamic equations and information in the observations themselves. Outside the perfect model scenario, the GD method requires a stopping criteria to locate relevant pseudo-orbits and informative estimate of the model error.

References

[1] K.Judd and L.A.Smith. Indistinguishable states I: perfect model scenario. *Physica D*, 151:125-141,2001.

[2] A.C. Lorenc, Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194, 1986.

[3] M.L. Berliner, Likelihood and Bayesian Prediction for Chaotic Systems, *J. Am. Stat. Assoc.*, **86**, 938-952, 1991.

[4] R.N. Miller, M. Ghil and F. Gauthiez, Advanced data assimilation in strongly nonlinear dynamical systems. *J.Atmos.Sci.* **51**, 1037-1056, 1994.

[5] H. Du, Combining Statistical Methods with Dynamical Insight to Improve Nonlinear Estimation, PhD Thesis, London School of Economics, 2009.

[6] J.L. Anderson, An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, 129, 2884-2903, 2001.

[7] J. Brocker and L.A. Smith, From ensemble forecasts to predictive distribution functions, submitted to *Tellus*, 2007.

[8] I.J. Good, Rational decisions. *Journal of the Royal Statistical Society*, **XIV**(1) 107-114, 1952.

[9] M.S. Roulston and L.A. Smith, Evaluating probabilistic forecasts using information theory, *Monthly Weather Review*, **130**, 1653-1660, 2002.

[10] D. Orrell, A shadow of a Doubt: Model Error, Uncertainty, and shadowing in Nonlinear Dynamical Systems. PhD Thesis, University of Oxford, 1999.