Regularized shadowing-based data assimilation method for imperfect models and its comparison to the weak constraint 4DVar method

Bart de Leeuw* Svetlana Dubinkina†

October 17, 2018

Abstract

We consider a data assimilation problem for imperfect models. We propose a novel shadowing-based data assimilation method that takes model error into account following the Levenberg-Marquardt regularization approach. We illuminate how the proposed shadowing-based method is related to the weak constraint 4DVar method both analytically and numerically. We demonstrate that the shadowing-based method respects the distribution of the data mismatch, while the weak constraint 4DVar does not, which becomes even more pronounced with fewer observations. Moreover, sparse observations give weaker influence on unobserved variables for the shadowing-based method than for the weak constraint 4DVar.

Key words: model error; data assimilation; shadowing; weak constraint; variational data assimilation

1 Introduction

A data assimilation method [8] combines a solution of a physical system model with measurement data to obtain an improved estimate for the state of a physical system. In this paper we study problems for which the system model is discrete in both space and time and contaminated by model errors

$$x_{n+1} = F_n(x_n) + Q_n, \quad x_n \in \mathbb{R}^m, \quad n = 0, \dots, N-1, \tag{1}$$

where $F_n : \mathbb{R}^m \to \mathbb{R}^m$ and Q_n is some unknown model error drawn from a Gaussian distribution with zero mean and covariance matrix C_{m}. We assume F_n to be $\mathcal{C}^3$ for all n. Let the sequence $\boldsymbol{X} := \{\mathcal{X}_0, \dots \mathcal{X}_N\}$ be a distinguished orbit of (1), referred to as the true solution of the model, and presumed to be unknown. Suppose we are given a sequence of noisy observations $\boldsymbol{y} := \{y_0, \dots y_N\}$ related to $\boldsymbol{X}$ via

$$y_n = H(\mathcal{X}_n) + \xi_n, \qquad y_n \in \mathbb{R}^d, \quad n = 0, \dots, N, \tag{2}$$

*Centrum Wiskunde & Informatica, PO Box 94079, 1090 GB Amsterdam, Netherlands, b.m.de.leeuw@cwi.nl.
†Centrum Wiskunde & Informatica, PO Box 94079, 1090 GB Amsterdam, Netherlands, s.dubinkina@cwi.nl.

where $H : \mathbb{R}^m \to \mathbb{R}^d$, $d \leq m$, is the observation operator, and the noise variables ξ_n are drawn from a normal distribution with zero mean and known observational error covariance matrix C_o. The goal of data assimilation is to find $\boldsymbol{u} = \{u_0, u_1, \ldots, u_N\}$, $u_n \in \mathbb{R}^m$, such that the differences $\|y_n - H(u_n)\|$ and $\|u_{n+1} - F_n(u_n)\|$, $n = 0, \ldots, N$ are small in an appropriately defined sense.

The data assimilation problem may be solved through algorithms known as variational data assimilation methods, see [1] for a recent review of operational data assimilation. Those methods are based on minimization of cost functions. We assume that N is big enough so that the contribution of a so-called background cost function is negligible. Denote $\|\boldsymbol{v}\|_M := \sqrt{\boldsymbol{v}^T M^{-1} \boldsymbol{v}}$ and as a convenient abuse of notation denote C_o (C_m) as a block diagonal matrix with $N+1$ identical blocks equal to the covariance matrix C_o (C_m). Then we define the observation cost function

$$J_\mathrm{o} = \frac{1}{2}\|H(\boldsymbol{u}) - \boldsymbol{y}\|^2_{C_\mathrm{o}}, \tag{3}$$

and the model cost function

$$J_\mathrm{m} = \frac{1}{2}\|G(\boldsymbol{u})\|^2_{C_\mathrm{m}}, \tag{4}$$

where $G(\boldsymbol{u})$ is the mismatch functional defined as

$$G(\boldsymbol{u}) = \begin{pmatrix} G_0(\boldsymbol{u}) \\ G_1(\boldsymbol{u}) \\ \vdots \\ G_{N-1}(\boldsymbol{u}) \end{pmatrix}, \qquad G_n(\boldsymbol{u}) = u_{n+1} - F_n(u_n), \quad n = 0, \ldots, N-1. \tag{5}$$

Then a so-called strong constraint 4DVar problem [16] minimizes the observation cost functions (3) under the constraint that a model orbit should be satisfied, $G(\boldsymbol{u}) = 0$. Thus the strong constraint 4DVar is only applicable for the models without model errors. When model error is present, one should consider a so-called weak constraint 4DVar problem [17], which in the state space formulation minimizes the sum of the observation (3) and model (4) cost functions.

The data assimilation problem may also be solved using shadowing-based methods. A shadowing data assimilation method known as pseudo-orbit data assimilation (PDA) [3] solves a problem of minimizing $\frac{1}{2}\|G(\boldsymbol{u})\|^2_I$. In order to stay close to observations, PDA is initialized at observations and the minimization is approximately solved using a fixed number of gradient descent steps. The PDA methods have also been applied to the weak constraint problem with the most recent approach [4] to still apply gradient descent to $\frac{1}{2}\|G(\boldsymbol{u})\|^2_I$, but only taking a limited number of steps using a stopping criterion.

Another shadowing-based approach to data assimilation is introduced in [2] and named here Newton shadowing. It is concerned with finding a root of the mismatch functional $G(\boldsymbol{u})$ rather than a minimum of a cost function, but with observations being an initial guess. It originates from numerical analyses to impose bounds on numerical errors approximations of pseudo-orbits of dynamical systems [5]. Newton shadowing is, however, limited to models without model errors. In this paper, we develop a shadowing-based data assimilation method for models with model errors. We adopt a regularization approach to find a solution of an imperfect model constrained by observations, namely an iterative regularizing Levenberg-Marquardt approach [6]. This method has been applied to data assimilation problems [7], though to stably minimize (3) under a strong constraint. The "classical" iterative regularizing Levenberg-Marquardt approach starts at a solution of a perfect model and aims at finding a solution of an optimization problem close to observations. This approach depends on a regularization parameter and a stopping criterion. According to Morozov's discrepancy principle, the regularization parameter can be determined uniquely. For noise-free observations and under some regularity conditions, the convergence theory guarantees that an iterative solution converges to a model orbit satisfying the observations exactly, when the number of iterations goes to infinity. For noisy observations, the stopping criterion chosen according to the discrepancy principle guarantees that at a finite iteration the distance between an iterative solution and the minimum is smaller that the distance between the initial guess and the minimum. In the present study, we start at observations and aim at finding a solution of an imperfect model. Determining the regularization parameter uniquely and imposing an appropriate stoping criterion, we ensure the correctly distributed data mismatch. We call this approach weak constraint shadowing and compare it to the weak constraint 4DVar both analytically and numerically.

The paper is organized as follows: we first recall the Newton shadowing data assimilation method in Section 2, then introduce the weak constraint shadowing data assimilation in Section 3. In Section 4 imperfect models are described and in Section 5 results of the numerical experiments with the imperfect models are discussed. Conclusions are drawn in the last section.

2 Newton shadowing data assimilation

In this section we describe Newton shadowing for strong constraint data assimilation developed in [2]. Newton shadowing finds model orbits close to observations by employing Newton's method for root searching of the mismatch functional $G(\boldsymbol{u})$ initialized at observations. Suppose $\boldsymbol{u}$ is an ε-orbit in a neighborhood of a hyperbolic set for F, namely $\|G_n(\boldsymbol{u})\| < \varepsilon$, $n = 0, \dots, N-1$, where $\|\cdot\|$ is a norm in $\mathbb{R}^m$. The shadowing lemma (e.g. Theorem 18.1.2 of [11]) states that, for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\boldsymbol{u}$ is δ-shadowed by an orbit of F, i.e. there exists an orbit $\boldsymbol{x}$ satisfying $G(\boldsymbol{x}) = 0$ such that $\|u_n - x_n\| < \delta$ for all $n = 0, \dots, N$. For example, let F be the exact time-τ flow map of an autonomous ODE $\dot{x} = f(x)$. If the components of $\boldsymbol{u}$ are the iterates of a numerical integrator with local truncation error bounded by ε, then these define an ε-orbit of F. Shadowing refinement [5] employs the pseudo-orbit as an initial guess for $G(\boldsymbol{u}) = 0$ and, as opposed to proving the existence of a nearby zero of G, iteratively refines the pseudo-orbit to obtain an improved approximation of a true solution. The inverse problem to shadowing is to determine an optimal initial condition u_0 for a numerical integration, such that the numerical iterates $\boldsymbol{u}$ δ-shadow a desired orbit of $\dot{x} = f(x)$.

Denoting by k the index of the Newton's iteration, we have at $k = 0$ $\boldsymbol{u} = \boldsymbol{y}$ and we seek an update $\boldsymbol{\delta}^{(k)}$ by approximately solving

$$G(\boldsymbol{u}^{(k)} + \boldsymbol{\delta}^{(k)}) = 0. \tag{6}$$

We then update using $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{\delta}^{(k)}$. The solution to (6) is approximated by iterating

$$G'(\boldsymbol{u}^{(k)})\boldsymbol{\delta}^{(k)} = -G(\boldsymbol{u}^{(k)}), \quad \boldsymbol{u}^{(k+1)} := \boldsymbol{u}^{(k)} + \boldsymbol{\delta}^{(k)} \tag{7}$$

to convergence. We solve each Newton's step using the right pseudo-inverse of G', i.e.

$$\boldsymbol{\delta}^{(k)} = -G'(\boldsymbol{u}^{(k)})^{\dagger} G(\boldsymbol{u}^{(k)}) = -G'^T(G'\,G'^T)^{-1}G.$$

The function $G(\boldsymbol{u})$ has a zero for every orbit of the model. The Jacobian of G has a $m(N-1) \times mN$ block structure:

$$G'(\boldsymbol{u}) = \begin{bmatrix} -F_0'(u_0) & I & & & \\ & -F_1'(u_1) & I & & \\ & & \ddots & \ddots & \\ & & & -F_{N-1}'(u_{N-1}) & I \end{bmatrix}.$$

We remark that through the use of the pseudo-inverse, for every Newton step, $\boldsymbol{\delta}^{(k)}$ is the minimum 2-norm solution to (7).

3 Weak constraint shadowing data assimilation

Before we introduce weak constraint shadowing, we lift the Newton shadowing assumption of the observation operator being the identity. We introduce completed observations, where the existing observations are completed with long time "climatological" averages. Then the observation covariance matrix C_{o} is also completed by using the covariance of the completed observations.

3.1 Levenberg-Marquardt regularization

It is our aim to use an iterative regularization method to modify the Newton shadowing into a weak constraint shadowing data assimilation method. We propose using the Levenberg-Marquardt iteration

$$\boldsymbol{\delta}^{(k)} = -C_{\mathrm{o}}G'^T(\boldsymbol{u}^{(k)})\left(G'(\boldsymbol{u}^{(k)})C_{\mathrm{o}}G'^T(\boldsymbol{u}^{(k)}) + \alpha^{(k)}C_{\mathrm{m}}\right)^{-1}G(\boldsymbol{u}^{(k)}), \qquad \boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{\delta}^{(k)}, \tag{8}$$

for $\alpha^{(k)} > 0$. Under some regularity conditions and algorithms for choosing $\alpha^{(k)}$, convergence to a model orbit can be proven as $k \to \infty$ [6]. We remark that if $C_{\mathrm{o}} = I$ and we choose $\alpha^{(k)} = 0$, for all k, then (8) reduces to the Newton shadowing (7). If $C_{\mathrm{m}} = C_{\mathrm{o}} = I$ and we choose $\alpha^{(k)} \to \infty$, for all k, then (8) reduces to the gradient descent algorithm of [10].

Dropping the trajectory dependency and iteration index for notational convenience, we rewrite (8) into a minimization problem. Assuming C_{m} and C_{o} are invertible, we have in exact arithmetic

$$\begin{aligned}
\boldsymbol{\delta} &= -C_{\mathrm{o}}G'^T(G'C_{\mathrm{o}}G'^T + \alpha C_{\mathrm{m}})^{-1}G \\
&\equiv -(G'^TC_{\mathrm{m}}^{-1}G' + \alpha C_{\mathrm{o}}^{-1})^{-1}(G'^TC_{\mathrm{m}}^{-1}G' + \alpha C_{\mathrm{o}}^{-1})C_{\mathrm{o}}G'^T(G'C_{\mathrm{o}}G'^T + \alpha C_{\mathrm{m}})^{-1}G \\
&\equiv -(G'^TC_{\mathrm{m}}^{-1}G' + \alpha C_{\mathrm{o}}^{-1})^{-1}(G'^TC_{\mathrm{m}}^{-1}G'C_{\mathrm{o}}G'^T + \alpha G'^T)(G'C_{\mathrm{o}}G'^T + \alpha C_{\mathrm{m}})^{-1}G \\
&\equiv -(G'^TC_{\mathrm{m}}^{-1}G' + \alpha C_{\mathrm{o}}^{-1})^{-1}G'^TC_{\mathrm{m}}^{-1}(G'C_{\mathrm{o}}G'^T + \alpha C_{\mathrm{m}})(G'C_{\mathrm{o}}G'^T + \alpha C_{\mathrm{m}})^{-1}G \\
&\equiv -(G'^TC_{\mathrm{m}}^{-1}G' + \alpha C_{\mathrm{o}}^{-1})^{-1}G'^TC_{\mathrm{m}}^{-1}G,
\end{aligned}$$

which implies that

$$G'^T C_\mathrm{m}^{-1} \left(G'\boldsymbol{\delta} + G\right) + \alpha C_\mathrm{o}\boldsymbol{\delta} = 0. \tag{9}$$

Introducing back the index notation and trajectory dependency, the solution $\boldsymbol{\delta}^{(k)}$ to (9) is the minimizer of

$$\frac{1}{2}\|G'\left(\boldsymbol{u}^{(k)}\right)\boldsymbol{\delta}^{(k)} + G\left(\boldsymbol{u}^{(k)}\right)\|^2_{C_\mathrm{m}} + \frac{\alpha^{(k)}}{2}\|\boldsymbol{\delta}^{(k)}\|^2_{C_\mathrm{o}} \quad \text{and} \quad \boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{\delta}^{(k)}. \tag{10}$$

At the first iteration the weak constraint shadowing is identical to the weak constraint 4DVar, when initialized at the full observations. As the iteration proceeds it, however, becomes distinct since (10) does not stay fixed throughout the iteration.

3.2 Parameters choice

The regularization parameter $\alpha^{(k)}$ can be determined uniquely by imposing that for some $0 < \rho < 1$ $\alpha^{(k)}$ is the smallest non-negative scalar satisfying

$$\rho^{-1}\|\delta\left(\alpha^{(k)}\right)\|_{C_\mathrm{o}} \leq \sqrt{Nd} - \|H(\boldsymbol{u}^{(k)}) - \boldsymbol{y}\|_{C_\mathrm{o}}, \tag{11}$$

where we made the dependency of the update step δ on the parameter $\alpha^{(k)}$ explicit. In practice we fix ρ and search for $\alpha^{(k)}$ in a sequence $\{0, 2^\nu\}$ for $\nu \in \mathbb{N}$ by computing $\rho^{-1}\|\delta\left(\alpha^{(k)}\right)\|_{C_\mathrm{o}}$ and accepting the update if (11) is satisfied. At the next iteration $(k+1)$ we do not start from zero but rather from a previous $\alpha^{(k)}$.

For the stopping criterion we require that the distance between analysis and observations remains bounded. Denoting the principal square root of the observational precision by $C_\mathrm{o}^{-\frac{1}{2}}$, $C_\mathrm{o}^{-\frac{1}{2}}(H(\boldsymbol{X}) - \boldsymbol{y})$ is distributed according to a standard normal distribution. In particular, $\mathbb{E}(\|H(\boldsymbol{X}) - \boldsymbol{y}\|^2_{C_\mathrm{o}})/Nd = 1$ and when the number of observations is large enough we may assume $\|H(\boldsymbol{X}) - \boldsymbol{y}\|^2_{C_\mathrm{o}})/Nd \approx 1$ with high probability. Thus we stop the algorithm at the minimum k for which $\|H(\boldsymbol{u}^{(k)}) - \boldsymbol{y}\|^2_{C_\mathrm{o}})/Nd > r$ for a predefined parameter r close to 1.

3.3 Comparison to the weak constraint 4DVar

We compare weak shadowing to weak constraint 4DVar following [9]. The maximum likelihood principle assumes that both model mismatches $\boldsymbol{G}(\boldsymbol{u})$ and the observation mismatches $H(\boldsymbol{u}) - \boldsymbol{y}$ are independent Gaussian variables. The weak constraint 4DVar derived based on the maximum likelihood principle provides, however, a solution such that the model mismatches $\boldsymbol{G}(\boldsymbol{u})$ depend on the observation mismatches $H(\boldsymbol{u}) - \boldsymbol{y}$ as

$$H'^T(\boldsymbol{u})C_\mathrm{o}^{-1}\left(H(\boldsymbol{u}) - \boldsymbol{y}\right) = -G'^T(\boldsymbol{u})C_\mathrm{m}^{-1}\boldsymbol{G}(\boldsymbol{u}). \tag{12}$$

Therefore we compare the weak shadowing to the weak constraint 4DVar in terms of distributions of $C_\mathrm{o}^{-1/2}\left(H(\boldsymbol{u}) - \boldsymbol{y}\right)$ and of $C_\mathrm{m}^{-1/2}\boldsymbol{G}(\boldsymbol{u})$ for normally distributed $C_\mathrm{o}^{-1/2}\left(H(\boldsymbol{X}) - \boldsymbol{y}\right)$ and $C_\mathrm{m}^{-1/2}\boldsymbol{G}(\boldsymbol{X})$ at the true trajectory $\boldsymbol{X}$.

4 Imperfect Models

We assume that a perfect model is given by equation (1) with $Q_n = \sqrt{C_\mathrm{m}}\eta_n$, where η_n is drawn form the normal distribution. The imperfect model is given by equation (1) with $Q_n = 0$. The imperfect and perfect models are related to each other as the Euler and the Euler-Maruyama discretisation of an ODE or an SDE with Brownian motion, respectively.

4.1 The double-well model

The stochastic double-well model has been used to test variational data assimilation in [15]. It is described by

$$x_{n+1} = x_n + \tau\left(x_n(1 - x_n^2)\right) + \sqrt{\tau}\sigma_\mathrm{m}\eta_n, \qquad (13)$$

where we choose the time step of the numerical discretization $\tau = 0.05$ and the model error $C_\mathrm{m} = \tau\sigma_\mathrm{m}^2$ with scalar σ_m. It has two stable equilibria at $x = \pm 1$ and one unstable equilibrium at $x = 0$. For sufficiently small stochastic noise, the model (13) stays near one of the stable equilibria for most of the time. Over long times, however, transitions between two stable equilibria occur. In the absence of the stochastic noise, transitions do not occur. This means that no orbit of the imperfect model is able to shadow the perfect model over long times.

4.2 The Lorenz 63 model

A stochastic version of the Lorenz 63 (L63) model [12] has been used in a data assimilation context in [14]. It is given by

$$\begin{aligned}
x_{n+1}^1 &= x_n^1 + \tau 10(x_n^2 - x_n^1) + \sqrt{\tau}\sigma_\mathrm{m}\eta_n^1,\\
x_{n+1}^2 &= x_n^2 + \tau(28x_n^1 - x_n^2 - x_n^1 x_n^3) + \sqrt{\tau}\sigma_\mathrm{m}\eta_n^2,\\
x_{n+1}^3 &= x_n^3 + \tau(x_n^1 x_n^2 - 8/3x_n^3) + \sqrt{\tau}\sigma_\mathrm{m}\eta_n^3,
\end{aligned}$$

where we use the standard parameter values, time step $\tau = 0.005$ and scalar σ_m.

4.3 The Lorenz 96 model

We also consider a stochastic version of the Lorenz 96 (L96) model [13] given by

$$x_{n+1}^l = x_n^l + \tau(-x_n^{l-2}x_n^{l-1} + x_n^{l-1}x_n^{l+1} - x_n^l + 8) + \sqrt{C_\mathrm{m}}\eta_n^l, \qquad l = 1, \ldots, 15, \quad (14)$$

where we use the standard parameter value for the forcing, time step $\tau = 0.005$, dimension size 15, and spatially correlated C_m

$$C_\mathrm{m} = \tau\sigma_\mathrm{m}^2 \begin{pmatrix} 0.5 & 0.25 & 0 & \cdots & 0 & 0.25 \\ 0.25 & 0.5 & 0.25 & 0 & \cdots & 0 \\ 0 & 0.25 & 0.5 & 0.25 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0.25 & 0.5 & 0.25 \\ 0.25 & 0 & \cdots & 0 & 0.25 & 0.5 \end{pmatrix} \qquad (15)$$

with scalar σ_m.

Method	Iterations	J_o/Nd	J_m/Nm	$2\left(J_\mathrm{o}+J_\mathrm{m}\right)/N(d+m)$
NA Shadowing	2 ± 0	0.516 ± 0.008	0.050 ± 0.002	0.565 ± 0.009
Shadowing	6.8 ± 0.6	0.492 ± 0.001	0.062 ± 0.005	0.554 ± 0.005
W4DVar	4.3 ± 0.5	0.365 ± 0.006	0.133 ± 0.003	0.499 ± 0.008

Table 1: Results for the stochastic double well model averaged over 100 experiments and with standard deviations. The cost functions J_o and J_m are defined in equations (3) and (4) respectively. NA Shadowing stands for Non-Adaptive Shadowing.

5 Numerical experiments

We use a spin-up of 5 time units for a true trajectory to reside on the attractor. We perform 100 numerical experiments with different truth and observation realizations in order to check the robustness of the results. We initialize the weak constraint 4DVar with (completed) observations, unless specified otherwise. The minimization of the cost function of the weak constraint 4DVar is done by a Matlab built-in Levenberg-Marquardt algorithm and stopping when the relative change in the cost function compared to the initial value is less then 10^{-6}. To find an adaptive $\alpha^{(k)}$ for the weak constraint shadowing we fix $\rho = 0.8$ and $r = 0.99$. We also check the performance of the weak constraint shadowing when $\alpha^{(k)} \equiv 1$, for all k, as this decreases the number of iterations.

5.1 Stochastic double well

For numerical experiments with the stochastic double well model we use $N = 4000$. We choose $\sigma_\mathrm{m} = 1$ and take observations at each time step with $C_\mathrm{o} = 0.16$. The histograms of the observation and model mismatches are shown in figure 1 for both the weak constraint shadowing and the weak constraint 4DVar. We see that the weak constraint 4DVar overfits both distance to the observations, though only slightly, and the model mismatch; while the weak constraint shadowing is overfitting only the model mismatch. In table 1, we show the mean and standard deviation over 100 experiments for the number of iterations, the observation cost function (3), the model cost function (4), and the total cost function. Results are given for the weak constraint 4DVar, the weak constraint shadowing with adaptive α and the weak constraint shadowing with fixed $\alpha \equiv 1$. We observe that adaptive shadowing is outperforming the weak constraint 4DVar but demands more iterations. By fixing $\alpha \equiv 1$ the shadowing results get only slightly worse but the number of iteration decreases considerably.

5.2 Stochastic Lorenz 63

For numerical experiments with the stochastic L63 model we use $N = 2000$ and observe only the x^1 coordinate every time step with observation error σ_o^2. The completed observations are obtained by averaging x^2 and x^3 of the deterministic L63 model over 2×10^7 time steps, $(\bar{x}^2\ \bar{x}^3)^T = (0.1015\ 24.3515)^T$. The observation covariance matrix is completed by the covariances of the long-time

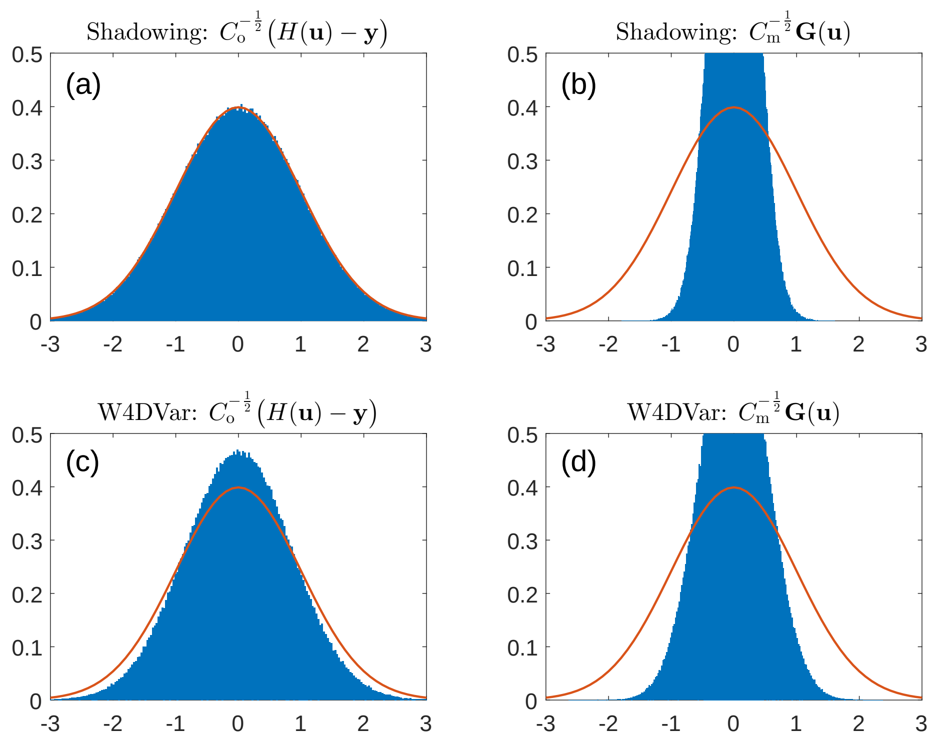

Figure 1: Histogram of the normalized data (left) and model (right) mismatches from the shadowing method (top) and the weak constraint 4DVar (bottom), for the stochastic double well model. In red we plot the standard normal distribution for reference.

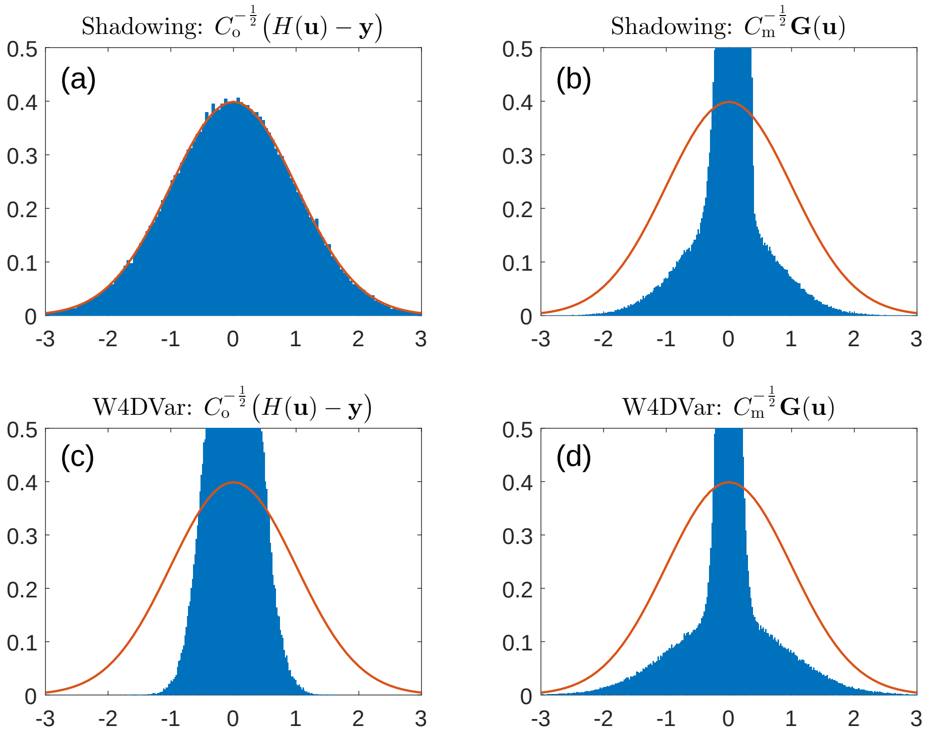

Figure 2: Histogram of the normalized data (left) and model (right) mismatches from the shadowing method (top) and the weak constraint 4DVar (bottom), for partially observed stochastic L63 model. In red we plot the standard normal distribution for reference.

trajectories of x^2 and x^3 and is

$$C_{\mathrm{o}} := \begin{pmatrix} \sigma_{\mathrm{o}}^2 & 0 & 0 \\ 0 & 82.9135 & 0.3134 \\ 0 & 0.3134 & 67.2204 \end{pmatrix},$$

assuming there is no temporal correlation and no correlation between non-observed states and the observation. We choose $\sigma_{\mathrm{o}}^2 = 0.05$ and $\tau\sigma_{\mathrm{m}}^2 = 0.6$. In figure 2, we see that the shadowing method ensures the distances to observations are distributed approximately correctly, while the model mismatch is overfitted. For the weak constraint 4DVar both distance to observations and model mismatch are overfitted. Moreover, the model mismatch is not Gaussian distributed for either method.

Comparing the methods in table 2, we observe that the weak constraint 4DVar overfits the total mismatch as well. Shadowing takes only one iteration more on average compared to the weak constraint 4DVar. Fixing α to one does not decrease the number of iterations substantially, though increases the overfit, which could be improved by tuning r. We performed the same experiments but with observations every 10 time steps by completing the observations every unobservable time step. As we did not see qualitative differences compared to observing every time step, we omit the results here.

Method	Iterations	J_o/Nd	J_m/Nm	$2(J_o + J_m)/N(d+m)$
NA Shadowing ($r = 0.9$)	3.7 ± 0.5	0.54 ± 0.08	0.10 ± 0.01	0.41 ± 0.02
NA Shadowing ($r = 0.99$)	4 ± 0	0.6 ± 0.02	0.09 ± 0.003	0.43 ± 0.01
Shadowing	6.2 ± 0.6	0.494 ± 0.002	0.101 ± 0.005	0.398 ± 0.007
W4DVar	5.1 ± 0.3	0.064 ± 0.002	0.145 ± 0.005	0.249 ± 0.008

Table 2: Results for the partially observed stochastic L63 model averaged over 100 experiments and with standard deviations. The cost functions J_o and J_m are defined in equations (3) and (4) respectively. NA Shadowing stands for Non-Adaptive Shadowing with different values of r in brackets.

Method	Iterations	$10J_o/Nd$	J_m/Nm	$2(J_o + J_m)/N(0.1d+m)$
NA Shadowing ($r = 0.9$)	3.2 ± 0.5	0.50 ± 0.06	0.03 ± 0.01	0.09 ± 0.03
NA Shadowing ($r = 0.99$)	3.7 ± 0.6	0.59 ± 0.06	0.03 ± 0.01	0.08 ± 0.02
Shadowing	6.5 ± 0.7	0.498 ± 0.002	0.03 ± 0.01	0.08 ± 0.02
W4DVar (Bg)	55 ± 29	0.017 ± 0.002	0.014 ± 0.003	0.028 ± 0.005
W4DVar (Obs)	49 ± 20	0.017 ± 0.002	0.011 ± 0.002	0.023 ± 0.003

Table 3: Results for the partially observed stochastic L96 model with observations every 10 steps, averaged over 100 experiments and with standard deviations. The cost functions J_o and J_m are defined in equations (3) and (4) respectively. We adjust the cost function normalization, namely $Nd \mapsto 0.1Nd$ to take into account the sparsity in time of the observations. NA Shadowing stands for Non-Adaptive Shadowing with different values of r in brackets. W4DVar (Bg) stands for initialization at background and W4DVar (Obs) stands for initialization at observations.

5.3 Stochastic Lorenz 96

For numerical experiments with the stochastic L96 model we use $N = 1000$ and observe x^1, x^6, and x^{11} coordinates every 10th time step with observation error $\sigma_o = 0.01$. The competed observations are averages of 2×10^6 time steps of the deterministic L96 model. For the model error we choose $\sigma_m = \sqrt{20}$. In figure 3, we see that the shadowing method provides the correct distribution for the data mismatch, while the weak constraint 4DVar drastically underestimates its variance.

In table 3, we observe that the weak constraint 4DVar takes an order of magnitude more iterations to converge than the shadowing method, independent of initialization. The weak constraint 4DVar overfits both the data and the model mismatch. Adaptive shadowing, on the other hand, provides the best data mismatch distribution and takes only 6 iterations on average. When fixing α to one and tuning r, an equivalent result can be achieved in even less iterations.

When so little data is given, data assimilation is not expected to give accurate results for unobserved variables. However, their behavior is of importance due to desirable dynamical consistency. Therefore we perform a long-time data assimilation experiment with $N = 10^5$. Computing a distance between the true solution $\boldsymbol{X}$ and a solution $\boldsymbol{u}$, namely $\|\boldsymbol{X} - \boldsymbol{u}\|_I^2/N(m-d)$, a long-time average for the deterministic L96 model gives 18.8, the weak constraint 4DVar 22.7, and the shadowing method 16.5. This mainly reflects that the distance to the orig-

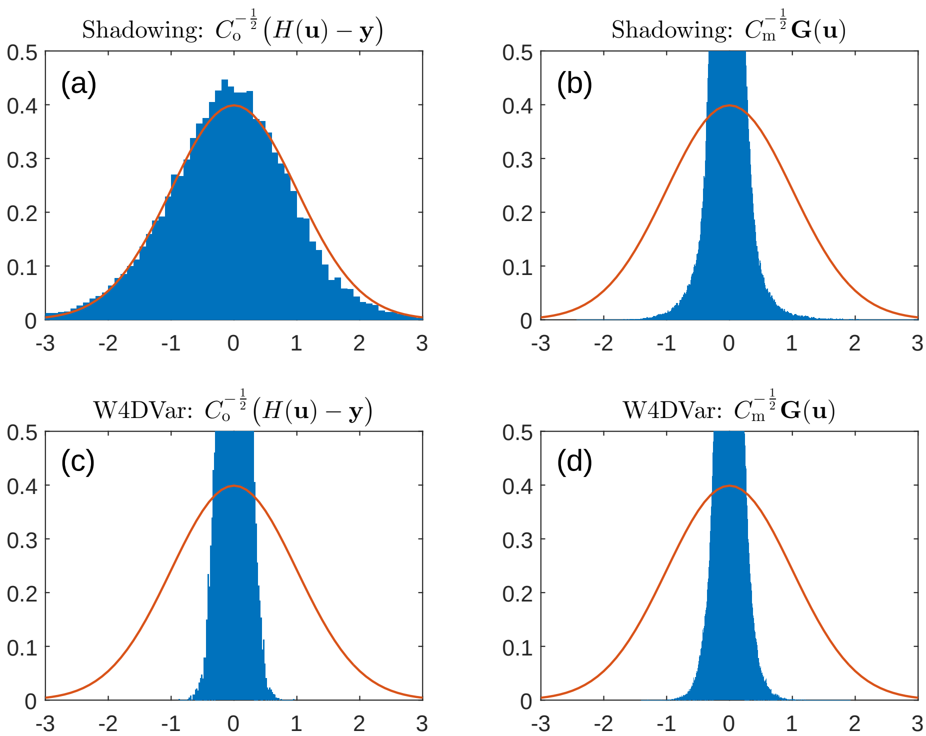

Figure 3: Histogram of the normalized data (left) and model (right) mismatches from the shadowing method (top) and the weak constraint 4DVar (bottom), for partially observed stochastic L96 model. In red we plot the standard normal distribution for reference.

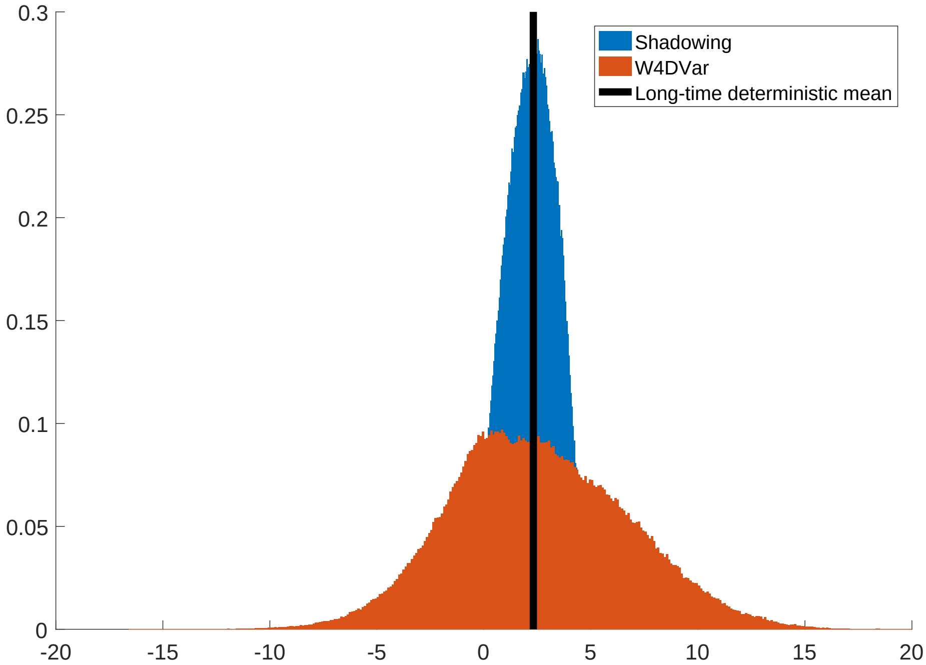

Figure 4: Histogram of all the unobserved variables for a partially observed long trajectory of stochastic L96. The weak constraint 4DVar is in red, and the shadowing method is in blue. The initial guess is in black.

inal "true" state is not a good measure of the performance of data assimilation algorithms for the weak constraint problem. Instead, we study the distribution of unobserved variables. The width of this distribution signifies how far an analysis deviates from the initial guess, which is a long-time average for the deterministic L96 model and is 2.31 in this case. In particular, it can be seen in figure 4 that the weak constraint 4DVar produces any pseudo-trajectory for the unobserved components. This pseudo-trajectory, however, is not particularly close to the truth. For the shadowing method, the unobserved variables do not significantly deviate from the initial background guess, which is shown as a black line. This is related to the stability of the shadowing method with respect to perturbations in the initial condition. Therefore, with little meaningful information on the unobserved variables the shadowing method provides a solution that is both stable with respect to the noise realization and a good reflection of the lack of knowledge.

6 Conclusions

We proposed a data assimilation method for imperfect models. The method is based on combination of numerical shadowing, a weak constraint formulation, and regularization. Numerical shadowing ensures stability with respect to observational noise, the weak constraint formulation introduces model error in the method, and Levenberg-Marquardt regularization prevents overshooting in the estimation. The appropriately chosen regularization parameter together with a data mismatch stopping criterion guarantees that the data mismatch is correctly distributed. We demonstrated that the shadowing method is successful for observations that are sparse both in space and in time.

We compared the proposed weak constraint shadowing-based method to the weak constraint 4DVar both analytically and numerically. We pointed out that they are identical only at the first iteration. Numerical experiments with stochastic models of double well, Lorenz 63, and Lorenz 96 confirmed analytic results that the shadowing method always estimates accurately distributions of the data mismatch. The weak constraint 4DVar, on the contrary, gives poor estimations of the data mismatch distributions, which become more deficient with fewer data. Moreover, unobserved variables only weakly deviate from an initial guess for the shadowing method, while strongly for the weak constraint 4DVar without being particularly close to the true trajectory. This is an advantage of the shadowing method, as sparse observations should not influence model states that are far away from the observation locations.

With respect to the model mismatch distributions, both the shadowing method and the weak constraint 4DVar perform poorly. This could be improved in the shadowing method by applying "classical" Levenberg-Marquardt regularization with a model mismatch stopping criterion. Then one gets the correct model mismatch distribution but at a price of misestimating the data mismatch distribution. As the goal of this study was to be close to accurate observations rather than to erroneous model estimations, this approach is unsuitable.

Future work will consist of further development of the shadowing method to use approximate adjoint models, generalizing the method to an ensemble approximation, and applying it to structurally incorrect models.

7 Acknowledgements

We would like to thank Amos Lawless, Nancy Nichols and Jason Frank for useful discussions. This work is part of the research programme Mathematics of Planet Earth 2014 EW with project number 657.014.001, which is financed by the Netherlands Organisation for Scientific Research (NWO).

References

[1] R. N. Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):607–633, 2017.

[2] B. de Leeuw, S. Dubinkina, J. Frank, A. Steyer, X. Tu, and E. V. Vleck. Projected shadowing-based data assimilation. *SIAM J. Appl. Dyn. Sys.*, 2018. to appear.

[3] H. Du and L. A. Smith. Pseudo-Orbit Data Assimilation. Part I: The Perfect Model Scenario. *Journal of the Atmospheric Sciences*, 71:469–482, 2014.

[4] H. Du and L. A. Smith. Pseudo-orbit data assimilation. Part II: Assimilation with imperfect models. *Journal of the Atmospheric Sciences*, 71(2):483–495, 2014.

[5] C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer. Shadowing of physical trajectories in chaotic dynamics: Containment and refinement. *Phys. Rev. Lett.*, 65:1527–1530, Sep 1990.

[6] M. Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse problems*, 13(1):79, 1997.

[7] M. A. Iglesias and C. Dawson. The regularizing Levenberg–Marquardt scheme for history matching of petroleum reservoirs. *Computational geosciences*, 17(6):1033–1053, 2013.

[8] A. H. Jazwinski. *Stochastic processes and filtering theory.* Mathematics in science and engineering. Academic press, New York, 1970.

[9] K. Judd. Forecasting with imperfect models, dynamically constrained inverse problems, and gradient descent algorithms. *Physica D: Nonlinear Phenomena*, 237(2):216–232, 2008.

[10] K. Judd and L. Smith. Indistinguishable states I. Perfect model scenario. *Physica D*, 151:125–141, 2001.

[11] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 1995.

[12] E. N. Lorenz. Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences*, 20:130–148, Mar. 1963.

[13] E. N. Lorenz. Predictability - a problem partly solved. In T. Palmer and R. Hagedorn, editors, *Proceedings of seminar on Predictability*, volume 1, pages 1–18, Reading, UK, 1996. ECMWF, Cambridge University Press.

[14] R. N. Miller, E. F. Carter, and S. T. Blue. Data assimilation into nonlinear stochastic models. *Tellus A: Dynamic Meteorology and Oceanography*, 51(2):167–194, 1999.

[15] R. N. Miller, M. Ghil, and F. Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056, 1994.

[16] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. *Q.J.R. Meteorol. Soc.*, 113, 1987.

[17] D. Zupanski. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Monthly Weather Review*, 125:2274–2292, 1997.