Reviewer #1

- *This manuscript contains a description of the results of a series of experiments in which 4DVAR was used to assimilate simulated data from a nonlinear system corresponding to the damped driven pendulum in a chaotic parameter regime. This example differs from most other examples in the literature of data assimilation in strongly nonlinear systems in that it is a non-autonomous system, unlike, say, Lorenz (63). With a few reservations, the example is fairly well worked out. The application of the χ^2 test is particularly noteworthy. The basic results are worth publishing in some form.*

  We thank the reviewer for noting the care we took in this analysis, especially the χ^2 posterior test, and for noting that the key results are worth publishing. We address some of your reservations in this point-by-point response.

- *The authors never state their model system explicitly. It is not (1). The system with which they are actually working differs from (1) in that it has a white noise term with variance S_f (see (5)) added to the right hand side. The distinction is not trivial. I assume that the reference solution in their twin experiments is the stochastic system with the stochastic term set to zero. The effect of adding the unknown stochastic term is to increase the number of degrees of freedom in the control space from two, i.e., the initial conditions in the purely deterministic problem, to the number of time steps taken by the numerical method, which is potentially infinite.*

  Thanks to the reviewer for spotting the error in defining the external forcing term. The equation is deterministic and is now stated explicitly in equation (1). The weight matrix S_f in equation (5) is selected in the state estimation process to limit the difference of the improved guess from the first guess, similar to the formulation in Bennett (2002). The revised text reads as follows.

  > The motion of the forced pendulum is described by the deterministic equation (*Baker and Gollub*, 1990),
  >
  > $$\frac{d^2\theta}{dt^2} + \frac{1}{q}\frac{d\theta}{dt} + \frac{g}{l}\sin\theta = f(t), \tag{1}$$
  >
  > where θ is the displacement angle from vertical, q is a damping coefficient, g is gravitational acceleration, l is the pendulum length, and $f(t)$ is an external forcing term. In turn, the external forcing has a first guess and a perturbation, $f(t) = f_0 + \delta f(t)$, where the first-guess is set to periodic forcing, $f_0(t) = b\cos(\omega_d t)$.

- *The general level of discussion in this manuscript might have been marginally acceptable twenty years ago, when implications of applications of techniques from the engineering world were still being explored, but most of the manuscript is far below the current state of the art.*

  It is clear that the motivation and goals of the manuscript need to be made more explicit. The Lagrange multiplier method is popular in oceanography due to automatic adjoint model compilers and strategies to reduce computer memory consumption. Much time and effort has been spent to develop this technique in real-world scenarios, yet it is unclear whether this method should be applied to eddy-resolving models and how long the time window should be. For the Lagrange multiplier method to be successful in state-of-the-art ocean models, two major issues need to be addressed: (1) the high dimensionality of the forward model and estimation problem, and (2) the nonlinearity of ocean models at increasingly fine resolution. Issue (1) has been overcome by groups such as the ECCO Consortium. Here we focus on (2). It is true that issues may arise by the combined effect of (1) and (2), but first we attempt to isolate the effect of nonlinearity.

  With this problem in mind, it is logical to find a numerical model that can be thoroughly understood and one that is highly nonlinear. It is not the goal of this manuscript to use a state-of-the-art numerical model. We believe that these expectations should be set at the outset, so we include the following in the Introduction.

> Because the effect of nonlinearity is seen as the major roadblock for application of the Lagrange multipler method, we isolate this effect by choosing a model that is highly nonlinear but low-dimensional: the forced, chaotic pendulum (Section 2). Toy models are worth revisiting because the dynamics are comparatively simple to understand, and they have strongly influenced when the Lagrange multiplier method has been deployed to realistic ocean problems. We will show that previous toy models have sometimes been misinterpreted.

We now also emphasize upfront that the development of a new state-of-the-art data assimilation technique is not the goal of this work, either. Instead, we wish to evaluate the current use of the Lagrange multiplier method. Now, the Introduction makes this explicit.

> Rather than developing a new state-of-the-art data assimilation technique, we proceed by taking the existing Lagrange multipler method and developing diagnostics regarding when and why it succeeds or fails, as evaluated by the ability to fit observations. Relative to the initialization problem, the prospects for a successful state estimate are shown to be improved in the boundary control problem, even if one uses a highly nonlinear model such as the forced, chaotic pendulum (Section 3).

The typical criterion for successful state estimation has been the stability to initial perturbation. In the manuscript, we provide a counterexample showing that state estimation can be successful for an unstable system when it is controllable. This is one novel result we are reporting, and the previous works suggested by the reviewers have not already made this point, nor do they appear to contradict it.

- *Studies of chaotic systems forced by white noise have appeared in a number of places in the literature. One example can be found in a paper by Tziperman from the early 90s.*

  We now mention Tziperman's work on chaotic systems as a motivating factor in using our toy model.

  > Toy models are worth revisiting because the dynamics are comparatively simple to understand, the nonlinear coupling to periodic forcing has been shown to be important in atmosphere-ocean dynamics (e.g., *Tziperman et al.*, 1994), and these models have strongly influenced when the Lagrange multiplier method has been deployed to realistic ocean problems.

- *There is nothing novel about writing the 4DVAR cost function in terms of a Lagrange multiplier. The use of Lagrange multipliers in variational formulations of estimation and control problems has been in the engineering textbooks since the 70s, and ap- peared in the early work of Thacker in the ocean modeling literature. In the present context, in which the task is to estimate an unknown stochastic forcing function, the Lagrange multiplier formulation is valid, but the same Euler-Lagrange equations result from equivalent cost function formulations without Lagrange multipliers, see, e.g., the text by Kalnay or either of the books by Bennett, as well as many of the reviews in the literature.*

  The reviewer's point that the Lagrange multiplier method is just another way to express a minimization problem is in line with the point of our Section 4.1. In that section, we suggest that the criteria for successful use of the Kalman Filter/Smoother, which also minimizes the same cost function, is the same as that for the Lagrange multiplier method. In particular, Sec. 4.1 states the following.

  > Our results suggest that the equivalence of the Kalman filter/smoother and Lagrange multiplier method may be extended to nonlinear problems, thus explaining why the chaotic estimation problem may be solved by the Lagrange multiplier method.

- *The authors should note that the estimation problem is the dual of the control problem. General questions of linear controllability and observability are dealt with in engineering textbooks. This topic has been well worked out in the context of models of the ocean and atmosphere in the work of S. E. Cohn in the late 80s and early 90s. The question of nonlinear observability is very complex. There was a book by Casti on the subject published some time ago.*

We thank the reviewer for the excellent suggestions for further references. In addition, we now point out the duality of the estimation and the control problem. As pointed out by the reviewer, observability and controllability conditions for nonlinear state estimation are difficult problems. Results for certain nonlinear systems are found in the book by *Casti* (1985) (which evolves from the review paper of *Casti* (1982)). For general nonlinear estimation problems, the down-gradient-based iterative optimization is likely one of the best methods, and we have shown the relevance of controllability to the iterative process in our example. We include the following new paragraph in the Discussion.

> To recover the true trajectory of a system, observability is also important, as the estimation problem is the dual of the control problem (*Fukumori et al.*, 1993; *Marchal*, 2014). For the linear problem, *Cohn and Dee* (1988) showed that completely observability implies asymptotic stability of the Kalman filter/smoother. Defining observability and controllability conditions for nonlinear state estimation problems is difficult *Casti* (1985). In practice, the important criterion is ability to solve equation (10). Strictly speaking, the solution criteria will therefore depend upon both the controllability matrix, **C**, and the observational matrix, **E**, which combines the issues of observability and controllability. Here, we suggest the operational definition that a system is effectively controllable when the solution to (10), generalized to multiple observations, exists.

- *The question of dealing with underdetermination has been discussed extensively in the literature. Solutions to underdetermined problems are not, in general, unique. The problem, in practice, is the fact that minimizing the cost function (5) involves searching a space of corrections that is potentially infinite. The highly irregular reconstructed forcing shown in the bottom panel of figure 4 is most likely one of an enormous number of minimizers of (5). There are almost certainly many others that will minimize the cost function, some smoother, many even more irregular.*

  The reviewer's point about the underdetermined nature of the problem is consistent with our discussion in Sec. 3.3, where we acknowledge that the solution is not unique, but we focus on finding any acceptable fit. We view the problem as having two clear steps. It is a first step to find any solution that acceptably fits the data. Only then can we proceed to investigate the uniqueness of the solution. In real-world situations, the first step may be the only one that is practical.

- *Bennett showed that, in the linear problem, one solution can be found by choosing a correction to the first guess that lies in an N y dimensional space spanned by repre- senter functions, where N y is the number of observations. This solution corresponds to the Moore-Penrose inverse. Arguments as to why that solution should be preferred over others are the stuff of textbooks.*

  It is worth clarifying that the external forcing had been treated as a controllable parameter in many works in the literature. In the Introduction, we now state the following.

  > As has been documented in detail by many authors including the textbook of *Bennett* (1992), the ocean state estimation problem is better described as a time-variable boundary value problem because synoptic atmospheric variability acts as an external forcing on the ocean (Section 2). Given our relatively uncertain knowledge regarding air-sea fluxes, the ocean state estimation is rightfully considered a time-variable boundary value problem where both the initial conditions and boundary conditions must be found. For example, *Bennett* (2002) described an estimation method for the external forcing, initial and boundary conditions that solves the Euler-Lagrange equations for a linear model.

- *Similar practical results can be had without explicit calculation of representers. In practical problems in modeling the ocean and atmosphere, the correction to the forcing function lies in a space of enormous dimension, so it is common to precondition the search for a cost function minimizer. This effectively reduces the dimension of the con- trol space by choosing corrections to be a linear combination of singular vectors of the error covariance matrix. This approach is documented in the work* *of A. Lorenc and O. Talagrand. In the present problem, it might be reasonable to impose nontrivial temporal correlation on the forcing correction, which might have the effect of limiting the spec- trum of the correction and thus ruling out irregular forcing corrections like that shown in figure 4.*

As the reviewer suggests, temporal correlations in the forcing field can be imposed through the use of nondiagonal weighting matrix, $\mathbf{S}_f$, in the cost function. The revised manuscript now describes how temporal correlations have been enforced in our analysis. The following material has been added to Sec. 3.4.

> We investigate the effect of a decrease in the number of controls by redefining the external forcing control perturbation. For N_u forcing controls, we define,
>
> $$f(t) = f_0(t) + \Gamma(t) \begin{pmatrix} \delta f(0) \\ \delta f(T/N_u) \\ \delta f(2T/N_u) \\ \vdots \\ \delta f(T) \end{pmatrix}, \tag{2}$$
>
> where $\Gamma(t)$ is a matrix that performs linear interpolation in time, and $\delta f(t)$ is only defined at N_u control times. This formulation enforces some temporal correlation in the external forcing. Alternatively, this could be accomplished using a nondiagonal weighting matrix, $\mathbf{S}_f$.

- *The authors have a choice. They can simply report on the results of their twin exper- iments on their nonautonomous system and eliminate nearly all of the discussion, or they can go back over twenty or twenty five years of literature and rewrite the discussion to make it a meaningful contribution to the current state of the art.*

We have taken seriously the reviewer's suggestions to place our work in the greater context of the published literature. Major revisions include new paragraphs in the Introduction regarding the motivation and aims, more information about the background of controllability that places our work in a broader context, and a new figure using an imperfect model that recreates a more realistic scenario.