Review of npg-2017-44 “The Onsager–Machlup functional for data assimilation"

General remark

For an ordinary reader who is not already familiar with the Onsager–Machlup (OM) functional, Girsanov formula and Radon–Nikodym derivative, the formulations and related derivations presented in this paper are too sketch to follow. My specific comments are given below.

Specific Comments

1. The first sentence in the abstract is confusing or inaccurate, because one does not have to resort to the more advanced and more difficult OM functional, as long as the stochastic differential equation (SDE) is to be solved in a time-discretized form (rather than time-continuous form) for data assimilation application. Nonlinear SDEs can rarely be solved analytically in time-continuous form. Although the OM functional is useful and important for rigorous theoretical considerations when the time-continuous limit is applied to a time-discretized form of quadratic cost function, the time-continuous limit can be derived formally (or intuitively) without considering the OM functional, as shown in (5.36)-(5.40) on pages 155-156 of Jazwinski (1970: *Stochastic Processes and Filtering Theory*). Since the SDEs are actually solved in time-continuous forms in this paper (as well as in most data assimilation studies), the importance and utility of the OM functional for real data assimilation appear to be overstated in this paper.

2. The noise intensity σ in (1) is a constant rather than a function of x_t. In this case, as explained at the end of section 4.6 (pages 119-120) of Jazwinski (1970), the associated Ito integral and Stratonovich integral are identical. Thus, as a stochastic differential equation, (1) can be viewed either as an Ito equation or its equivalent Stratonovich equation. The authors may need to clarify this point.

3. Eq. (13) is derived from (A2) by the scaling for a smooth curve, but the scaling x_n - x_{n-1} = O(δ_t) is not explained in Appendix A.

4. It appears that (21) is derived from (13) and (19) with ϕ changed into x, but it is not clear why ϕ can be changed into x.

5. It is not shown and unclear how (22) is derived.

6. Due to the questions in above comments 3-4, it is not clear whether the four schemes considered in section 2 all converge to the same time-continuous limit. If the answer is yes, then the differences between the numerical results obtained from the four schemes for each example in section 3 are caused the differences in discretization, and these differences should diminish as δ_t approaches to zero. To verify this numerically, the authors need to show for each example that the differences between the numerical results obtained from the four schemes become increasingly small as δ_t decreases toward zero.

7. Section 6.3.2 of Law et al. (2015) is cited for the derivation of divergence term in Appendix B. I checked but found that there are only 5 chapters in Law et al. (2015).

8. The formulation on the line above (B7) appears to be for $\delta\mu_o/\delta x$ or $d\mu_o/dx$ [that is, the

variation of μ_o with respect to variation of $x(t)$ for $0 \leq t \leq T$] rather than for the Wiener measure μ_o itself. Similarly, μ should be $\delta\mu/\delta x$ or $d\mu/dx$ on the left-hand side of (B7) and (B8). Correct?

9. As a reader, I would like to see the detailed step-by-step derivations (with adequate interpretations) of (B11)-(B14).