Estimating physical parameter of stochastic dynamical system using an Ensemble^2-Expectation-Maximization algorithm

Yin Yang[1]*, Etienne Mémin[2]

[1] Irstea, UR OPAALE, F-35044 Rennes Cedex, France

[2] Inria, Fluminance group, Campus universitaire de Beaulieu, F-35042 Rennes Cedex, France

* correspondent author : yin.yang@irstea.fr

Estimating parameters of geophysical dynamic model is an important task in Data Assimilation (DA) technique that serves the purposes of forecast/reanalysis. Recovering the exact parameters helps compensate for errors due to approximations made in the analytical constitution steps of the models. Among all the approximations made, an important sources of errors comes from a lack of physical information on complex small-scale processes. Consequently, the actions of those unresolved process on the large-scale (resolved) components must be properly modeled. Recently, [1] proposed to introduce a stochastic dynamics to model the unresolved small-scale components. Due to a rigorous random representation of geophysical flow dynamics, in this approach, recovering the subgrid-scale errors boils down to the calibration of a variance tensor representing the fast-changing fluctuation associated to turbulence. Nowadays observation snapshots are reaching a resolution that is much finer than the model grid resolution. Therefore one major incentive of this work is to retrieve as accurate as possible such subgrid parameters from high resolution observations.

In the past, most parameter estimation strategies followed the augmented state techniques. One such augmented state approach based on EnVar method as shown in [2] has been applied to this state/subgrid parameter estimation problem. The authors found that although the estimated state can be quite accurate, the estimated parameter field tends to be much less accurate needing to be further retuned by adhoc inflation techniques to avoid model blow-up and/or ensemble collapse.

In this work, we investigated an ensemble formulation of Expectation-Maximization (EM) algorithm to overcome such shortcomings associated to augmented-state-based techniques. The EM algorithm ([3]) offers an elegant and flexible solution for estimating separately the state and the model parameter that has been used extensively in many research domains, such as machine learning or system identification. But studies on integrating EM algorithm into data assimilation framework have only appeared recently. [4, 5], among others, proposed applying the EM algorithm on re-estimation of a low-order model error covariance terms associated to a reduced order noisy dynamics. Estimating such error covariance is viable but is also largely limited by the dimension of the parameters because estimating correlated model error covariances is intractable for high dimension problem.

In our approach, we proposed a novel ensemble formulation of the Maximization step that allows the optimal physical parameters to be directly estimated using iterative solution methods for linear system. We have also shown that, subject to some simplifications, the covariance update scheme in [4] is an approximation of our proposed scheme. More importantly, our proposed scheme identifies the physical parameter explicitly, which greatly simplifies the conception and implementation process of applying the EM algorithm to other parameter estimation problems, where the EM algorithm is traditionally to be thought as hardly applicable because the closed form solution does not exist.

Finally, the proposed method is applied to estimate the stochastic physical parameters associated with stochastic shallow water system using a twin experiment as proof-of-concept. The twin experiment consists in two cases : the first case corresponds to a typical state and parameter estimation problem where the true stochastic parameter distribution is known, the second case corresponds to finding optimum parameterizations according to the stochastic dynamics that represent the resolution-induced error. In both cases, our proposed method is shown to be able to yield considerably more accurate parameters than those obtained from augmented state technique. In addition, in terms of prediction capacity, our proposed method leads to smaller generalization error and better forecast compared to augmented state technique. Our proposed EM formulation is a promising approach readily applicable to many other non-linear dynamical model identification problems.

Références

[1] Mémin, E. (2014). Fluid flow dynamics under location uncertainty. *Geophysical and Astrophysical Fluid Dynamics* 108(2), 119–146.

[2] Yang, Y., Mémin, E. (2017). High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4DEnVar. *Tellus A : Dynamic Meteorology and Oceanography* 69(1), 1–15.

[3] McLachlan, G., Krishnan, T. (2007) *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics, John Wiley & Sons : Hoboken, NJ, USA.

[4] Dreano, D., Tandeo, P., Pulido M., Ait El Fquih B., Chonavel T., Hoteit I., A. (2017). Estimating model-error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximization algorithm. *Quarterly Journal of the Royal Meteorological Society* 143(705), 1877–1885.

[5] Pulido M., Tandeo P., Bocquet M., Carrassi A., Lucini M. (2017). Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods. *Tellus A : Dynamic Meteorology and Oceanography* 70(1), 1442099.