Bayesian Nonlinear Assimilation of Eulerian and Lagrangian Coastal Flow Data

Dr. Pierre F.J. Lermusiaux
Department of Mechanical Engineering
Center for Ocean Science and Engineering
Massachusetts Institute of Technology; 5-207B
77 Mass. Avenue
Cambridge, MA 02139-4307
phone: (617) 324-5172 fax: (617) 324-3451 email: pierrel@mit.edu

Grant Number: N00014-14-1-0725 FY15 report
05/01/2014 - 04/30/2017
http://mseas.mit.edu/

LONG-TERM GOALS

The long-term goal is to: *Develop and apply theory, schemes and computational systems for rigorous Bayesian nonlinear assimilation of Eulerian and Lagrangian coastal flow data, fully exploiting nonlinear governing equations and mutual information structures inherent to coastal ocean dynamical systems and optimally inferring multiscale coastal ocean fields for quantitative scientific studies and efficient naval operations.*

OBJECTIVES

The specific objectives for the three-year project are to:

- Implement and further develop our DO equations and numerical schemes for predicting the pdfs of nonlinear multiscale coastal ocean fields, both in Eulerian and Lagrangian forms.
- Further develop and implement our GMM-DO schemes for robust Bayesian nonlinear estimation of coastal ocean fields by assimilation of Eulerian and Lagrangian flow data.
- Apply our DO and GMM-DO schemes, as well as their theoretical extensions, numerical schemes and distributed implementation, in idealized-to-realistic coastal dynamics conditions and coastal flow observing system simulation experiments.
- Evaluate results using information-theoretic metrics, explain multiscale dynamics and interactions, quantify coastal flow observation requirements, and complete computational analyses
- Collaborate and transfer data, expertise, approaches, algorithms and software to NRL and other colleagues. Utilize and leverage the MIT Naval Officer education program.

APPROACH

A motivation is to exploit the information provided by coastal platforms (drifters, floats, gliders, AUVs or HF-radars) so as to best augment the limited resolution and accuracy of satellite data in coastal regions and to determine coastal sampling needs for successful Bayesian field estimation in diverse coastal regimes. For optimal scientific studies and naval operations, data assimilation in these regions should combine the information contained in measurements with that predicted by models in accord with the known complexities of the probability density functions (pdfs). To do so, the notion of covariance functions should be replaced by information-theoretic concepts such as mutual information functions among state variables or other relevant attributes. Such rigorous estimation based on governing equations and information theory is one of the critical motivation for our effort. Our aim is not to shy away from the known nonlinearities and nonstationary heterogeneous statistics, but to fully exploit these known information structures and correspondences, for robust and accurate Bayesian nonlinear estimation.

We are implementing and further developing our Dynamically Orthogonal (DO) numerical schemes for idealized and realistic primitive-equations with a nonlinear-free-surface and apply them to a range of coastal flow dynamics and probabilistic forecasting conditions. DO schemes are researched and DO terms, functions and operations are implemented, mostly by re-use and alteration of existing codes, focusing on: DO advection terms and specific advection schemes, terms for DO modes, diagnostic equations for pressure and density, non-polynomial nonlinearities, boundary and initial conditions for the modes and stochastic coefficients, numerical DO orthonormalization, stochastic forcing terms, implicit-explicit time-stepping schemes, and the integrals in space and statistical averages. The schemes are developed for both hydrostatic and non-hydrostatic dynamics. We also plan to research novel closure schemes for our DO equations, possibly using entropy criteria for optimal adaptive probability tracking or a Mori-Zwanzig formalism, leading to modified adaptive DO equations. We plan to investigate extensions to other norms (energy, information, entropy, etc.) as well as DO equations for other expansions than our dynamic Karhunen–Loève expansion. We plan to also explore DO schemes for predicting probability density functions of Lagrangian trajectories, LCSs or other dynamical system properties, which are needed for direct assimilation of such Lagrangian data.

We are further developing and implementing our GMM-DO assimilation schemes, again for primitive-equations with a nonlinear free-surface. Using guidance from our existing codes, the steps of the GMM-DO schemes including the prior input and posterior output are being made generic and efficient. We also propose to complete theoretical and algorithmic assimilation research, specifically: explore the effects of different DO closures, norms and equations on the assimilation results; utilize and derive efficient schemes for the estimation of mixture parameters for the pdf of large nonlinear systems; and, consider other mixture models and multi-resolution approaches. We plan to apply and study the properties of the GMM-DO filter/smoother, including entropy and information-flows, for the estimation of velocity fields and circulation features. We also plan to explore the extension of our Eulerian GMM-DO schemes to Lagrangian and Eulerian-Lagrangian GMM-DO schemes, so as to assimilate the nonlinear time-integrated information contained in Lagrangian trajectories and observed coherent structures.

We plan to apply the DO and GMM-DO schemes, as well as their theoretical extensions, numerical schemes and distributed implementation, in idealized-to-realistic coastal dynamics conditions and coastal flow observing system simulation experiments. We evaluate results using information-theoretic metrics, explain the multiscale dynamics and interactions we find, quantify the coastal flow observation requirements, and complete computational analyses. We plan to transfer expertise, approaches and algorithms to naval laboratories and centers.

WORK COMPLETED (FY15)

Uncertainty Prediction with DO equations for Eulerian and Lagrangian Fields

New stochastic DO primitive equations for probabilistic ocean predictions: For efficient uncertainty quantification and prediction of Eulerian and Lagrangian fields in the coastal ocean, we first derived the stochastic Dynamically Orthogonal (DO) PDEs for the MSEAS primitive equation ocean model with a nonlinear free surface. We also completed the design of numerical schemes for their space-time integration. In deriving the mean, mode and coefficient equations, we identified and successfully answered a set of theoretical and computational questions, e.g.: (i) What are appropriate state variables and equations on which the DO decomposition should be applied?; (ii) Are there advantages/ disadvantages to a diagnostic treatment of certain state variables?; (iii) Should separate stochastic DO modes/coefficients be used for the internal and external velocity modes (i.e. barotropic vs. baroclinic)? (iv) What are the numerical schemes that efficiently decouple the mean, mode and coefficient for DO primitive equations?; and, (v) How can the stochastic spatial discretization be performed?

Implementation and further development of numerical schemes for DO coastal ocean equations: In developing an algorithm for solving the DO primitive equations, we first focused on developing a semi-implicit temporal discretization algorithm. The motivation for this approach comes from the algorithm used in our modular MSEAS primitive-equation code (Haley and Lermusiaux, 2010). The approach allows for maximum reuse of existing codes. A new complication is that in addition to the coupling between transport and free-surface, the new DO equations have coupled mean, mode and coefficient equations. To numerically decouple the time-integration of these equations, we explored schemes to handle the stochastic transport and free-surface. Specifically, we analytically assessed the effect of the nonlinear free-surface terms in the deterministic primate equation code. We utilized the time integration of deterministic barotropic velocities to understand how to discretize the transport and free-surface terms in the stochastic barotropic velocity DO equation. We then derived a conservative time dependent stochastic spatial discretization with terrain following generalized vertical coordinates.

DO Equations for Stochastic LCSs and Lagrangian Variables: The repelling and attracting LCSs we estimated in several two-dimensional flows are related to the stable and unstable manifolds, respectively. These LCSs provide different yet complementary information on the flow fields to that provided by the DO modes. The present DO modes are instantaneous in time, but integrate dynamical information from the past and allow to predict dominant variability in the future. As a first step towards the merging LCS and DO information, direct comparisons were made of the LCS as computed by FTLE (finite time Lyapunov exponents) to mutual information fields and DO modes for different two-dimensional flow configurations.

Bayesian Nonlinear Data Assimilation: Eulerian and Lagrangian Filtering/Smoothing

Bayesian GMM-DO Smoothing: During FY15, we developed the GMM–DO smoother, a scheme for Bayesian inference backward in time (Lolla, 2016, Lolla and Lermusiaux, 2015-sub). It is applicable to high–dimensional stochastic fields governed by general nonlinear dynamics. The smoother carries out

Bayesian inference both forward (filtering) and backward (smoothing) in time, while retaining the non–Gaussian structure of all the state variables. It uses the stochastic DO PDEs and their time–evolving stochastic subspace to predict the prior probabilities. The forward and backward Bayesian inferences are then analytically carried out in the dominant DO subspace, after fitting semi-parametric Gaussian Mixture Models (GMMs) to the DO subspace realizations. We assessed the performance of the GMM–DO smoother using idealized fluid and ocean flows governed by nonlinear and noisy dynamics.

*Implementation of the GMM-DO Filter for Coastal Ocean Fields***:** In FY15, we further developed and applied our Bayesian data assimilation framework to two coastal ocean test cases. The first one is a biogeochemical-physical ocean model. We started with biogeochemical dynamics modeled by advection-diffusion-reaction equations for three tracers representing nutrient, phytoplankton and zooplankton. A technique of parameter estimation in a dynamics-based Bayesian data assimilation using state augmentation was developed. This technique allows us to estimate the states and unknown parameters jointly. One parameter that was jointly estimated is the Ivlev grazing constant for zooplankton. The second test case is a two-dimensional turbulent bottom gravity current. In this test case, saltier water flows down a slope and multiscale flow features develop, in part due to the Kelvin-Helmholtz instability and anisotropic diffusion. This multiscale density driven flow is governed by Navier-Stokes equations with the Boussinesq buoyancy approximation. The results further illustrated the capability of the GMM-DO filter in jointly inferring model parameters and high-dimensional state variables, under realistic multiscale flows, by utilizing sparse observational data.

Further Numerical Development and Improvements for DO and GMM-DO.
In FY15, we also further investigated and improved some numerical aspects of the DO equations and the GMM-DO filter. For the GMM-DO filter, we introduced the isotropic-position normalization to make the Monte Carlo samples have zero mean and an identity covariance matrix via an affine transformation. The fitted GMM means and covariance matrices can be transformed back by the inverse affine transformation. It turned out that this enhances the performance of the GMM-DO filter, at least in the case when the number of modes is small. We also compared the results of GMM-DO filtering with different observation variances and began to investigate how the observation variance can affect the performance. For DO, we implemented and compared several ways to handle non-polynomial forcing terms, including using 1st or 2nd order Taylor series approximation, quadratic least square fitting, and interval-wise linear/quadratic fitting.

RESULTS (FY15)

Uncertainty Prediction with DO equations for Eulerian and Lagrangian Fields

New stochastic DO primitive equations for probabilistic ocean predictions: We showed that for efficient DO equations and for maximum reuse of our modular MSEAS primitive-equation codes, it is best to apply the DO decomposition to an augmented computational state vector comprising of the 3-d finite-volume integrated velocities, temperature and salinity fields, and 2-d free-surface field. This implies a diagnostic treatment of 2-d barotropic velocities, and 3-d density and pressure fields. These variables then have a DO decomposition with diagnostic equations for their respective mean, modes and coefficients. The improved understanding from the tests on the nonlinear free-surface equation in our deterministic code provided critical inputs to discretize and approximate the stochastic versions of the terms in this equation. Overall, the result are new discrete stochastic DO primitive equations for predicting the pdfs of the Eulerian and Lagrangian fields in coastal oceans. These equations are 2^{nd} order in time and space, having conservative finite volumes with a structured grid that evolves with the mean nonlinear free-surface.

DO Equations for Stochastic LCSs and Lagrangian Variables: Comparisons of deterministic FTLE fields to mutual information fields and DO modes are presented in Fig. AA-CC. The ridges of maxima in the FTLE field are indicators of repelling LCSs in forward time and of attracting LCSs in backward time. For a reversible swirling flow (Fig. 1), the FTLE fields are compared to the mutual information field integrated starting from a point in the lower right of the domain to the present time (Fig. 1c). This mutual information field correlates with the attractors of the backward FTLE (Fig. 1b), e.g. the mutual information contours spiral in the same sense as the backward FTLE. In the next two examples, the FTLE fields are compared directly to the DO modes. In the sudden expansion example (Fig. 2), the FTLE field mix information from the DO mean and modes. In the bottom gravity current example (Fig. 3), the DO modes highlight the uncertainty in the leading edge and in the structures of the billows. The FTLE ridges also highlight these regions along with the strong flow at the bottom (where the uncertainty is low).

Bayesian Nonlinear Data Assimilation: Eulerian and Lagrangian Filtering/Smoothing

Bayesian GMM-DO Smoothing: We compared the GMM–DO smoother to Gaussian smoothers, such as the ESSE smoother and Ensemble Kalman Smoother. We also validated the GMM–DO smoother using the example of a passive tracer transported by an analytical flow–field, wherein the exact smoothed variables can be determined numerically by reversing the flow starting from the GMM-DO solution at the final simulation time. We also studied the advantages of non-Gaussian smoothing for a barotropic flow exiting a strait.

Implementation of the GMM-DO Filter for Coastal Ocean Fields: We tested the GMM-DO data assimilation in an ecosystem simulation in an idealized flow over a seamount. Fig. 4 illustrates the estimation of the state and value of a parameter (Ivlev grazing constant) for the three component NPZ model. Observations for only concentration of nutrients are made at all integer times between t=5 to t=18 at the 9 locations marked by white circles in the top-left panel of Fig. 4. The pdf for the unknown parameter is shown in the middle panel in the right column, along with the true value (3.6). The pdf remains essentially constant after time T=18. The evolution of the RMSE of the state (N, P and Z field) and parameter estimates are shown in the upper right panel. The decreasing values show that along with the parameter, we jointly estimate the states. Fig. 5 shows several spatial covariance plots for N, P and Z before and after assimilation at T=5. Such plots can help us identify where to make observations to reduce the uncertainty of a particular variable field. The plots show that the magnitude of the covariance is reduced significantly (Note the different color scales) after assimilation.

We also evaluated the GMM-DO filter for a two-dimensional turbulent bottom gravity current. In this dynamics, saltier water flows down a slope due to gravity and forms multiscale flow structures. Fig. 6 shows the DO stochastic simulation results without data assimilation (top panels) and the comparison of RMSE time history for runs with and without data assimilation (bottom panels). These results illustrate how the GMM-DO filter improves the prediction despite the multiscale turbulent features that develop and the strong non-Gaussian character of the statistics. Fig. 7 illustrates how the GMM-DO filter works in more detail. Salinity is observed at the locations marked by white dots in Fig. 7, once per minute from T=10 min. The bottom plot demonstrates the effectiveness of the GMM fitting to the non-Gaussian probability distributions, while the top one illustrates how the GMM-DO filter improves the prediction over the whole domain, even though it only uses very sparse point measurements.

Further Numerical Development and Improvements for DO and GMM-DO

Fig. 8 compares the evolution of the RMSEs and standard deviations of two runs using the GMM-DO filter with and without the isotropic-position normalization. We can see that the GMM-DO filter with the isotropic-position normalization outperforms the one without, for all three RMSEs. Moreover, the RMSE evolution curve in the former case is much smoother than that in the latter, which indicates an improvement in robustness. Fig. 9 compares the time histories of the RMSEs and standard deviations of three runs using GMM-DO filter with different observation variances (Note that this variance is the one used in the likelihood model in the filter. It is different from the observation noise variance, which is used to generate observation data from the truth and is unknown). The RMSE has two contributions: one from the discrepancy between the estimate mean and the truth, and the other from the variance of the estimate. Here we can see a tradeoff between reducing these two contributions. When the observation variance increases, we trust the observation data less, and thus more uncertainty (see the standard deviation curves) remains after assimilation. However, this also avoids overreacting to the observations and leads to smaller differences between the estimate mean and the truth, because the observations contain errors and the stochastic model involves some approximations to the space of state uncertainty. This indicates that there is an optimal choice of observation variance for minimizing RMSEs. An approximate optimal choice can be obtained by either jointly inferring the state variables and the observation variance, or by an analytical derivation with some simplifications. Investigation on this tradeoff is still undergoing.

IMPACT/APPLICATIONS

Our research has direct impact on efficient velocity predictions and data assimilation in coastal regions for naval operations, undersea surveillance, homeland security and coastal management. The research is relevant to naval interests in improving and implementing new theories and methods for Bayesian nonlinear assimilation of velocity information in coastal regions. It improves coastal ocean forecasting and acoustic performance forecasting, especially equation-based uncertainty predictions and nonlinear data assimilation to reduce flow uncertainties in coastal areas. Our integration of research areas in several disciplines such as ocean modeling, ocean observing and prediction systems, stochastic modeling, uncertainty quantification, information theory, estimation theory, dynamical system theory, Lagrangian Coherent Structures and distributed computing is directly applicable to naval needs of this new century.

TRANSITIONS AND COLLABORATIONS

We collaborate with colleagues to develop, demonstrate and transfer ideas and approaches. We transfer our results, test beds and codes to NRL-Stennis, e.g. metrics to evaluate pdfs, see (Lermusiaux et al., 2015). We build our software community http://mseas.mit.edu/software, so as to provide software to interested Physical Oceanography PIs and related groups. We plan to apply our work to idealized test cases but also to realistic ocean fields and to real-time sea exercises of opportunity. Each of these activities can lead to additional transfer of knowledge and expertise. We presented results at various conferences and to local groups. We leveraged our existing computational facilities and encouraged collaborations with national and international partners. To provide efficient education, we leverage the

MIT Naval officer education program so as to continue to attract METOC officers and practitioners, either for focused shorter visits (e.g. in the summer), or for Master's or PhD degrees.

RELATED PROJECTS

Our project on Active Transfer Learning for Ocean Modeling (N00014-11-1-0337) benefits from the test cases we develop for the present study.

STUDENT SUPPORTED: This project supported one graduate student. A summer visiting student from India, A. Gupta, contributed to the project for a second summer. Mr. Matt Swezey - LT USN (Navy support) is working on his SM with our group on coupled ocean physics and acoustics uncertainty forecasts for subsurface counter-detection in coastal regions.

PUBLICATIONS

Lermusiaux, P.F.J., Haley, P.J. Jr., Gawarkiewicz, G.G. and Jen, S., 2015. *Evaluation of Multiscale Ocean Probabilistic Forecasts: Quantifying, Predicting and Exploiting Uncertainty*. To be submitted to the Journal of Ocean Dynamics.

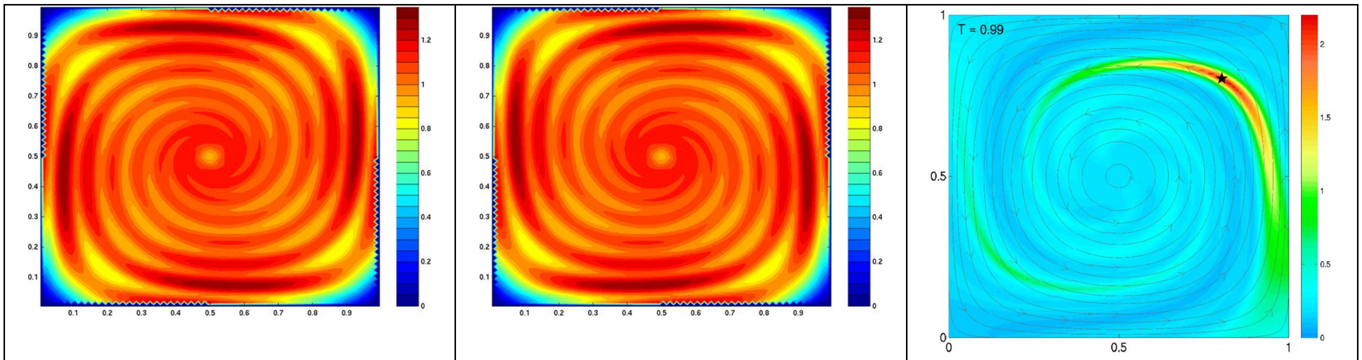
Lolla, T. and P.F.J. Lermusiaux, 2015. *Gaussian-Mixture Model – Dynamically Orthogonal Smoothing for Continuous Stochastic Dynamical Systems*. Monthly Weather Review. Submitted.

THESES

Lolla, T., 2015. *Path Planning and Adaptive Sampling in the Coastal Ocean*. Ph.D. Thesis, Massachusetts Institute of Technology, Dept. of Mechanical Engineering, Feb. 2016.

Lin J., 2016. *Bayesian Learning for Multiscale Ocean Flows*, SM Thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, February 2016. Expected.

Other publications are in preparation. Additional presentations and other publications are available from http://mseas.mit.edu/. Other specific figures are available upon request.



(a) (b) (c)

Figure 1: Comparing LCS to mutual information for reversible swirl flow: (a) forward FTLE (b) backward FTLE (c) mutual information integrated starting from a point in the lower right of the domain to the present time. The mutual information field is correlated with the attractors of the backward FTLE (b), as shown by the mutual information contours spiraling in the same sense as the backward FTLE.

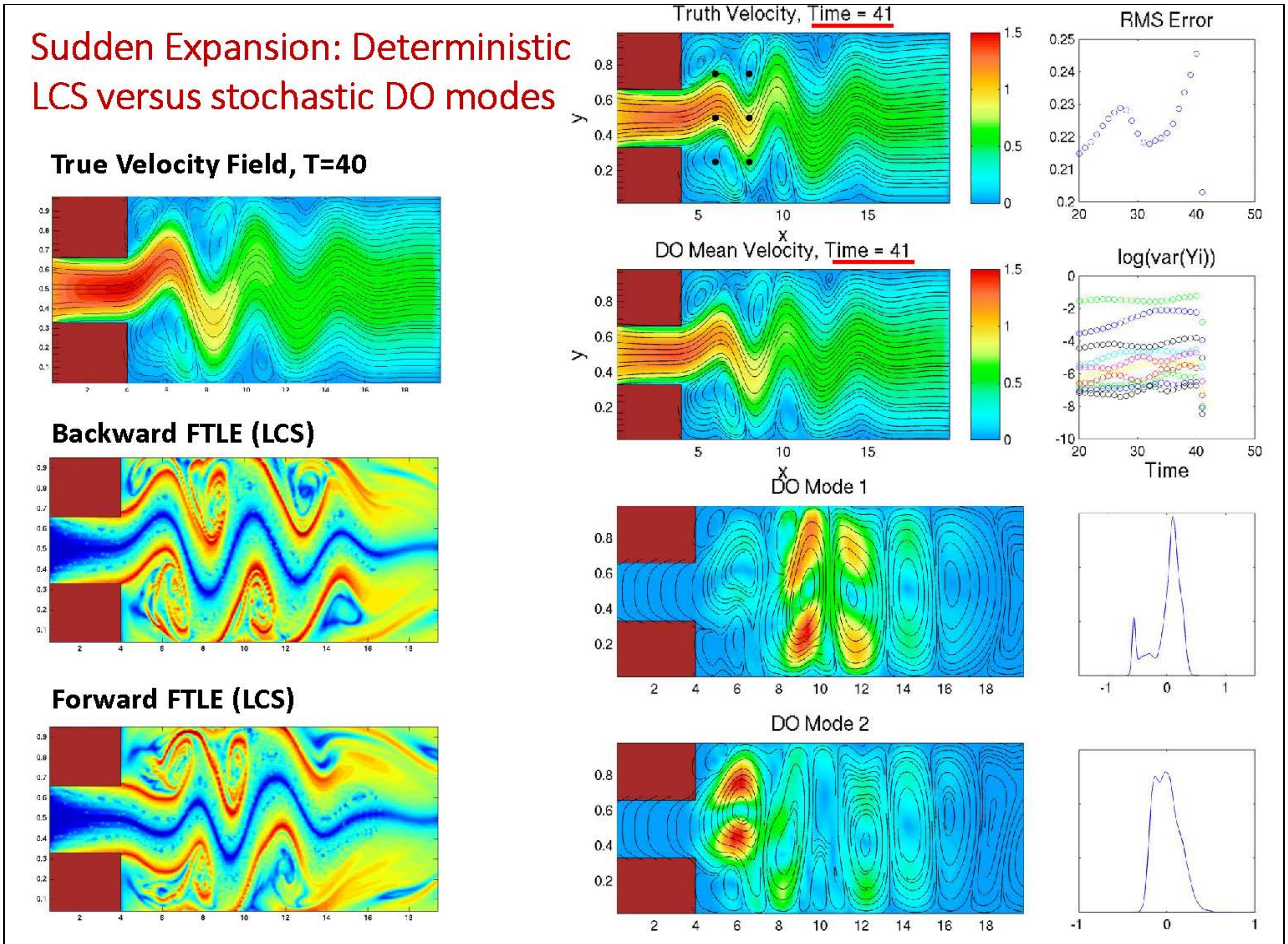


***Figure 2: Comparing LCS with DO modes in the sudden expansion flow. The FTLE fields mix information from the DO mean and modes*.**

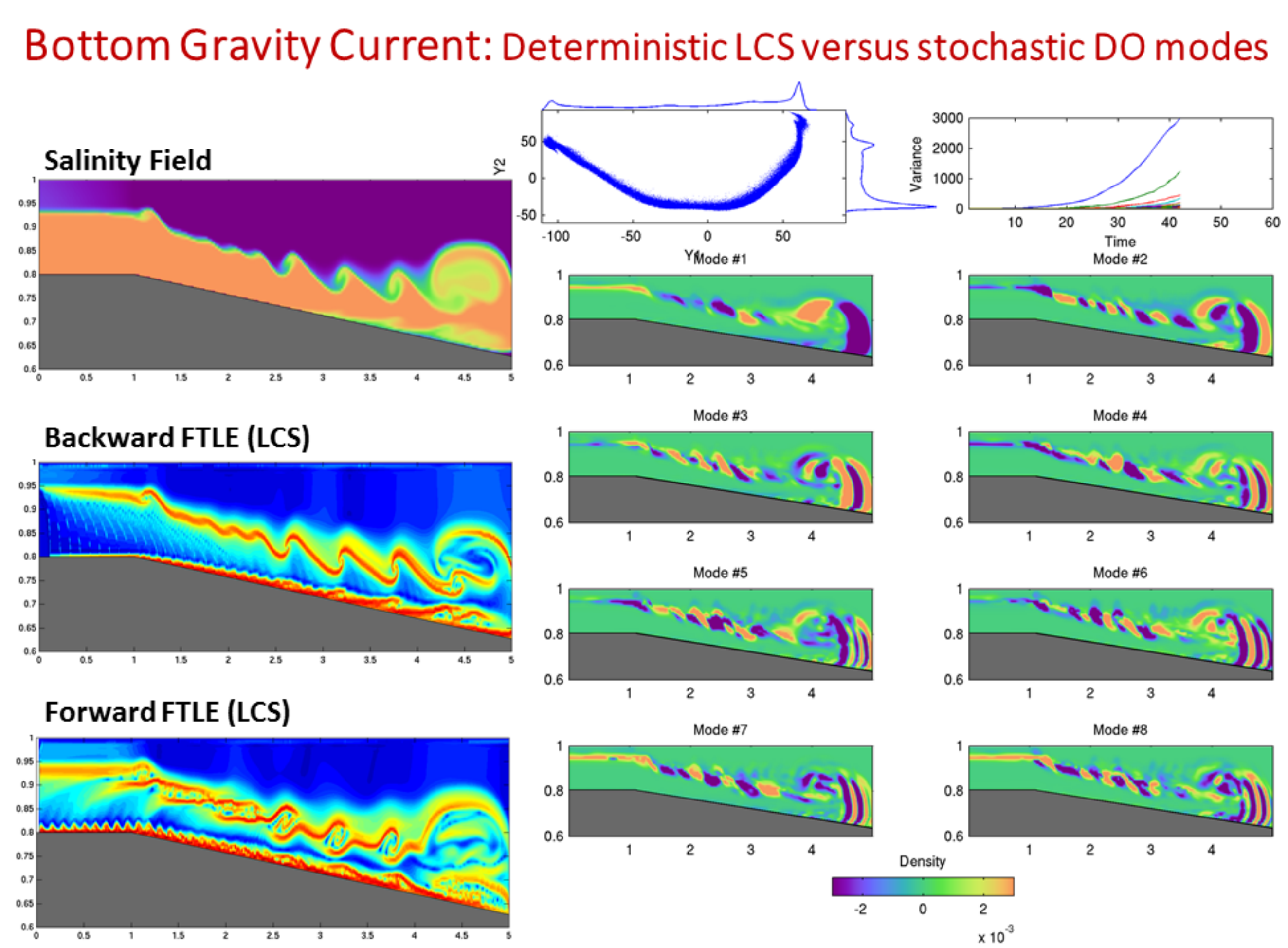


Figure 3: Comparing LCS with DO modes for a bottom gravity current. The DO modes highlight the uncertainty in the leading edge and in the structures of the billows. The FTLE ridges also highlight these regions along with the strong flow along the bottom (where the uncertainty is low).

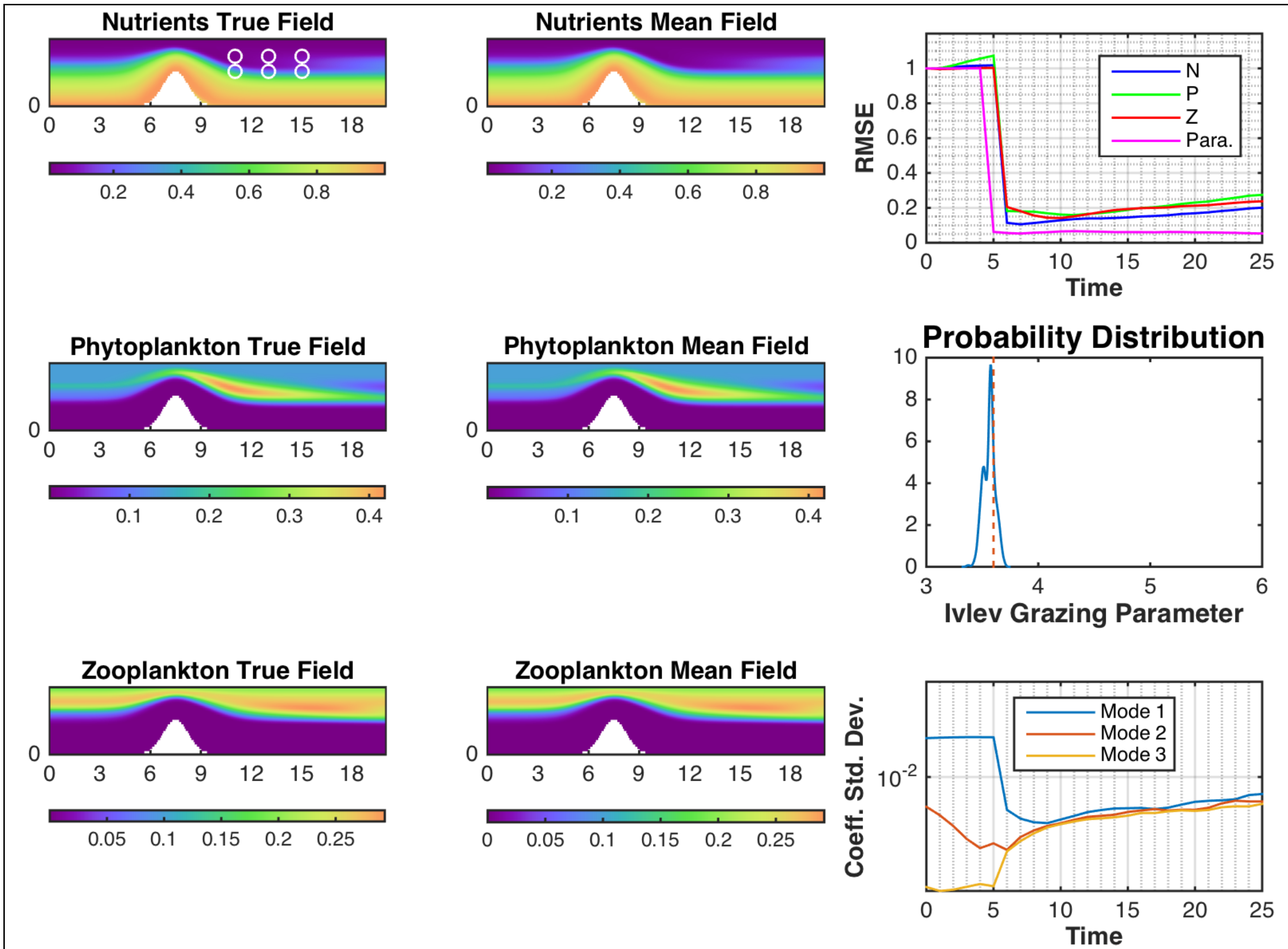


Figure 4: Parameter estimation results for NPZ (Nutrients, Phytoplankton and Zooplankton) biological model at T=25. All colored contours are for different biogeochemical tracers as mentioned above each such plot. White circles mark observation locations in the top-leftmost plot. The plots in the leftmost column shows the true solution, while the plots in the middle column show the Bayesian DO-mean state. The top-rightmost plot show the variation of RMS error for the 3 states and parameter with time. The middle plot in rightmost column presents the probability distribution for the Ivlev grazing parameter, with true value of 3.6 marked by red dashed. A very tall and tight peak can be noticed at the true value. The bottom-rightmost plot shows the variation of standard deviation of the three highest-energy modes with time.

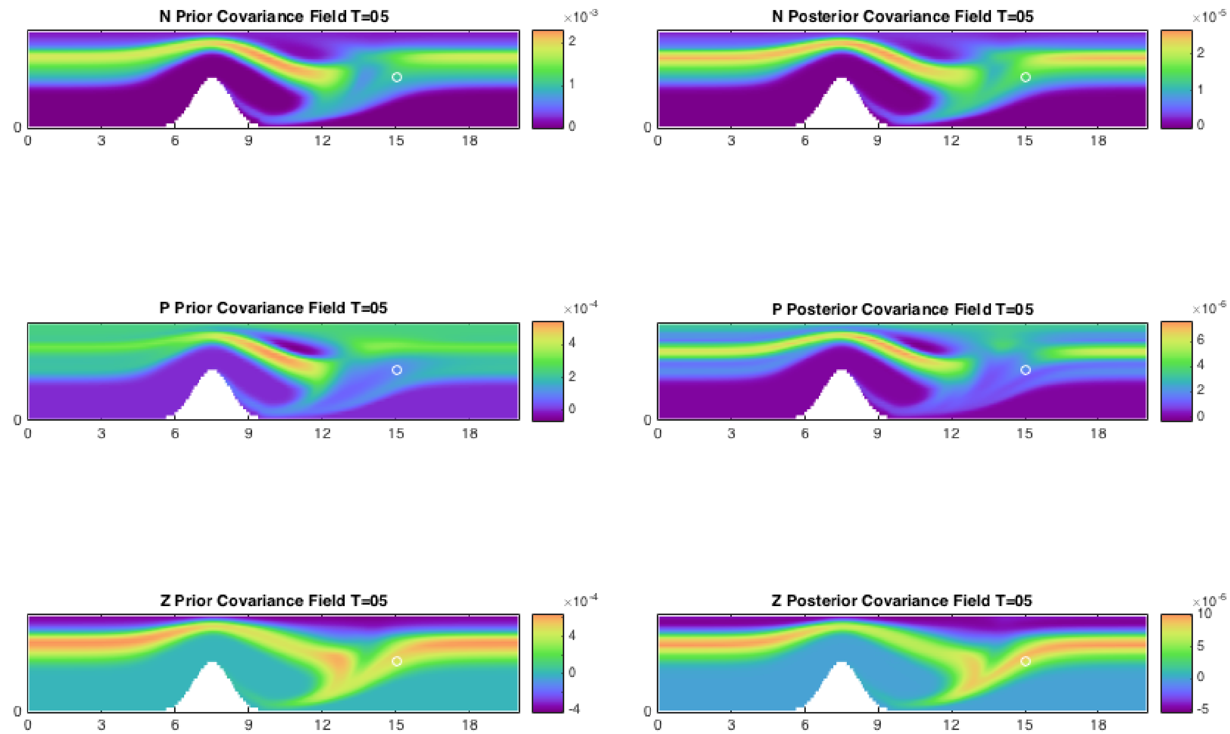


Figure 5: The covariance plots for N, P and Z before and after assimilation at T=5. The colors indicate the magnitude of the covariance between the field (N, P or Z) at the location marked by a white dot and that field elsewhere.

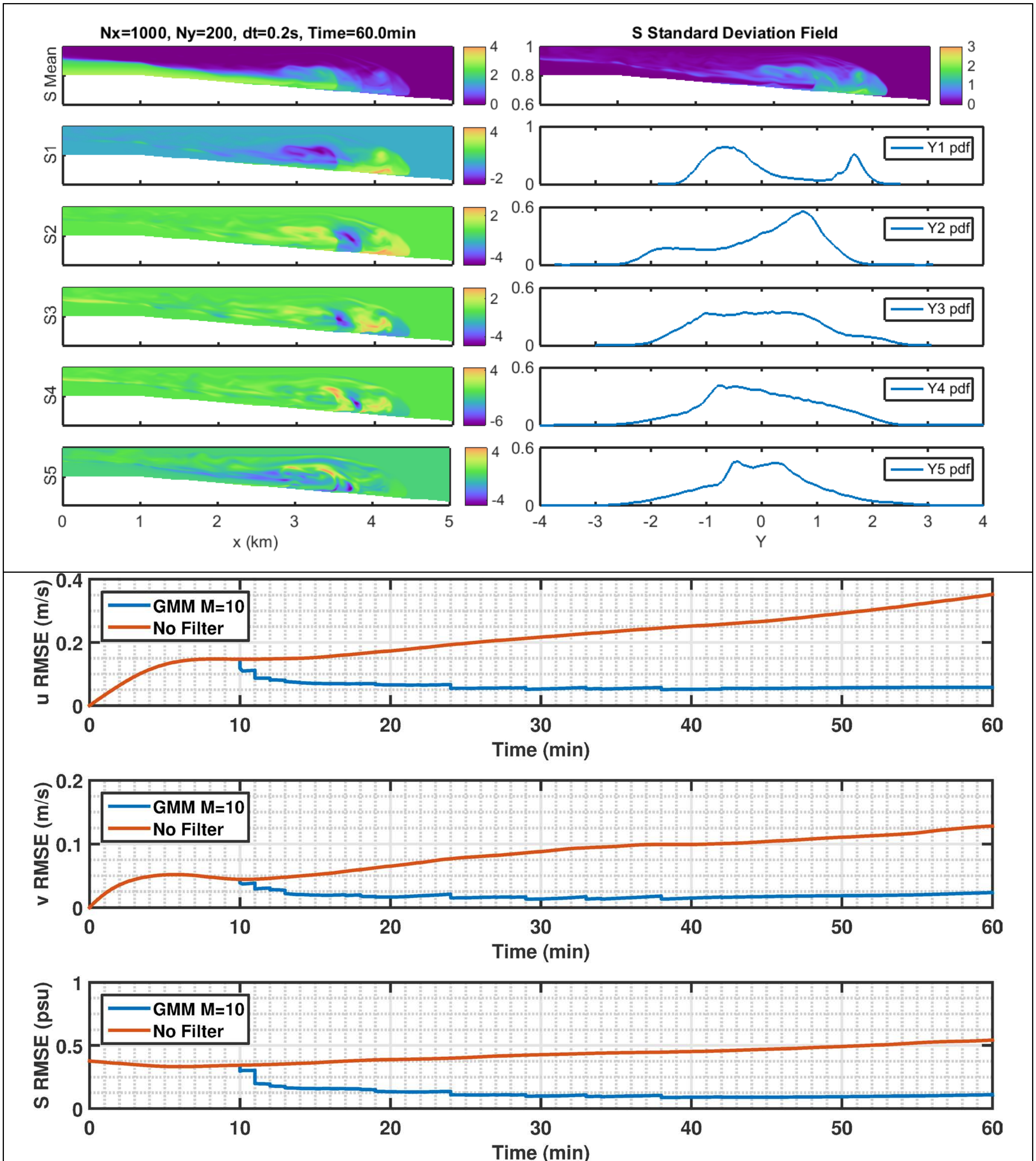


Figure 6: The top panel shows the DO mean, standard deviation and modes of the salinity field along with pdf's of the stochastic coefficients at non-dimensional time T=60, with no assimilation (filtering). The bottom panel shows the time history of RMSEs of u, v and S for a stochastic run without any filtering and a run with the GMM-DO filter using 10 Gaussian components.

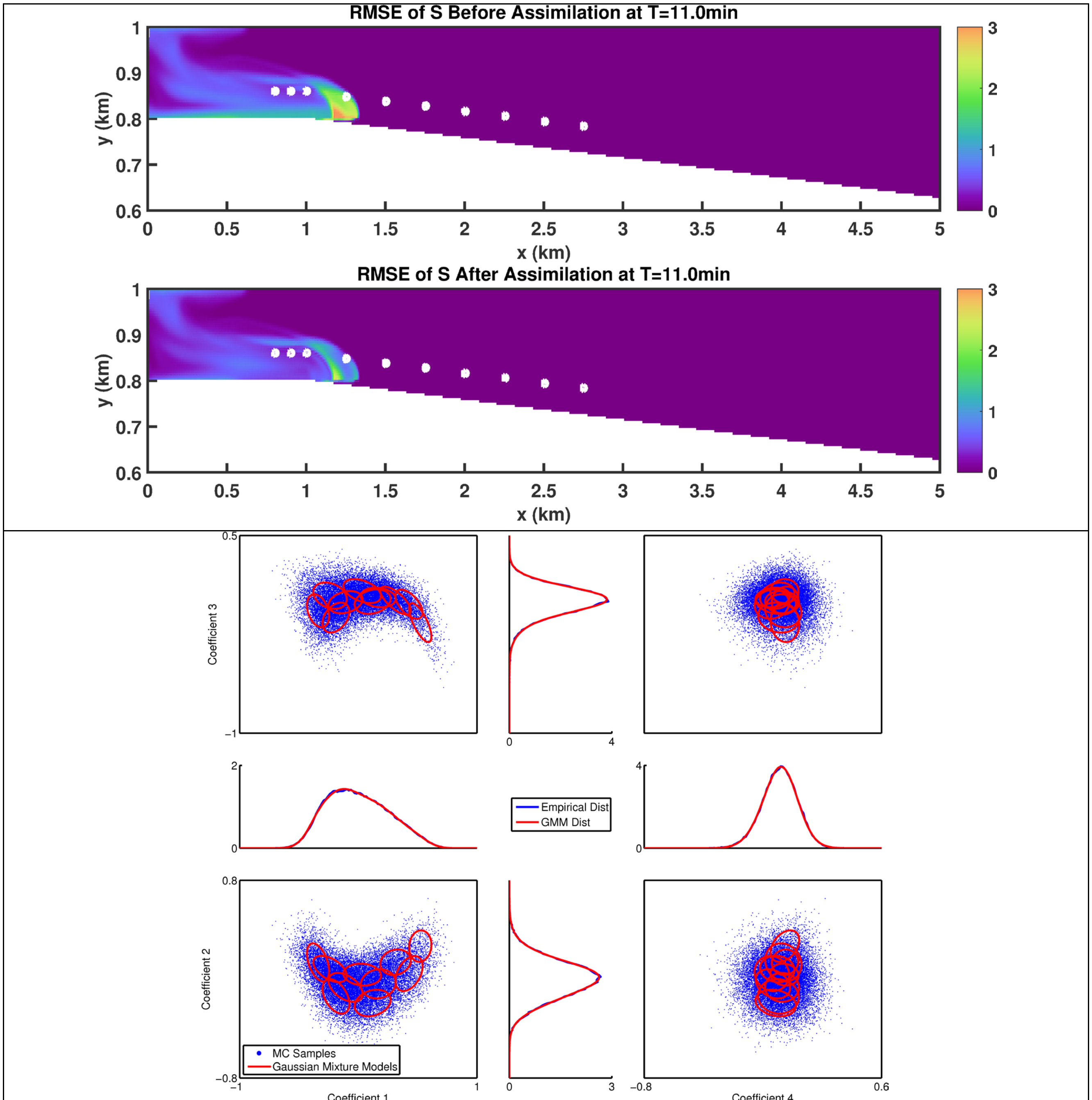


Figure 7: The top panel shows the RMSE of the salinity field before and after data assimilation at T=11 min. The bottom panel shows the empirical and fitted GMM distributions (joint and marginal) of the first four stochastic coefficients at T=10min before data assimilation.

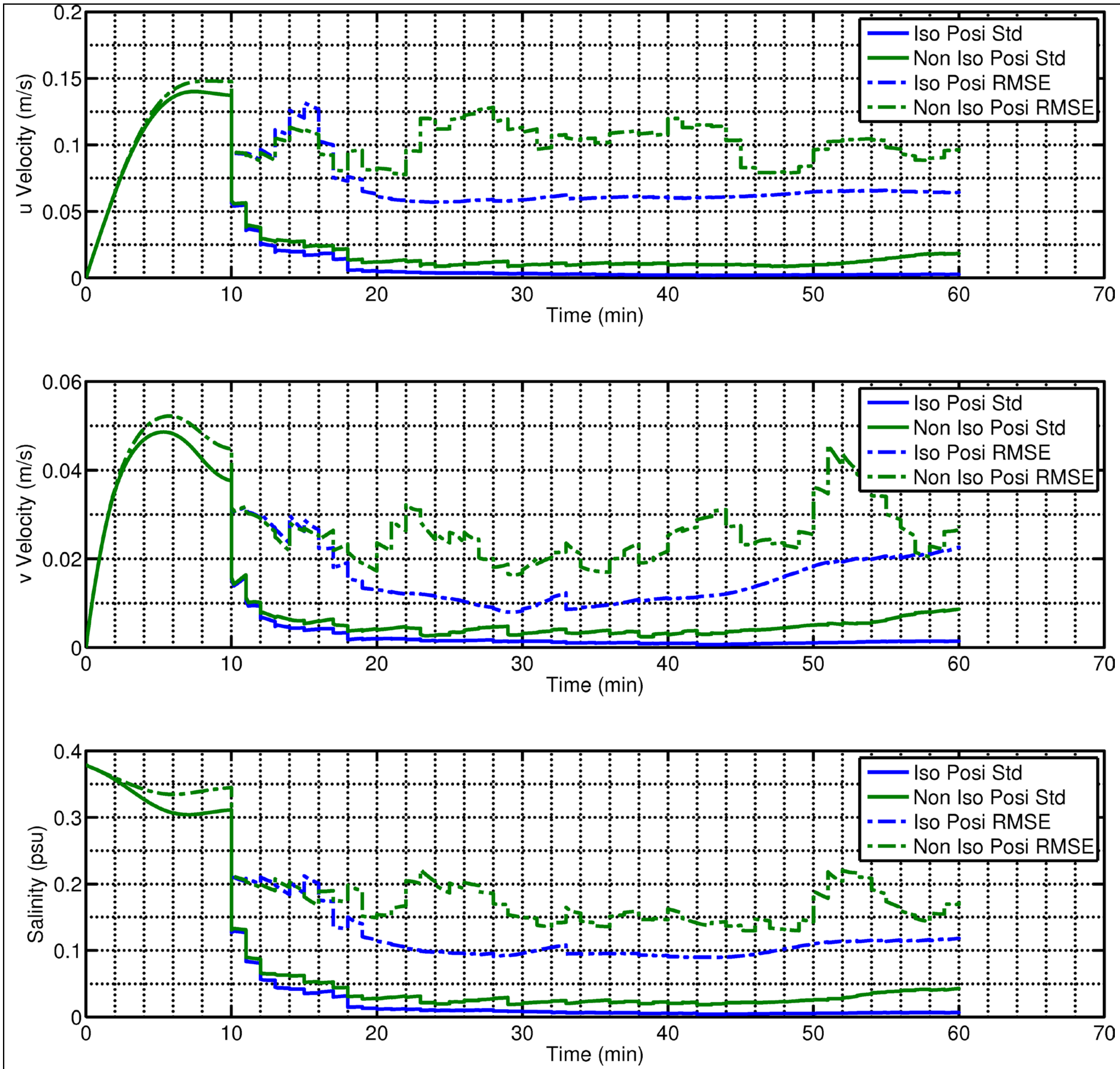


Figure 8: Time history plots of the RMSEs and standard deviations of u, v and salinity for two runs using the GMM-DO filter with and without the isotropic-position normalization.

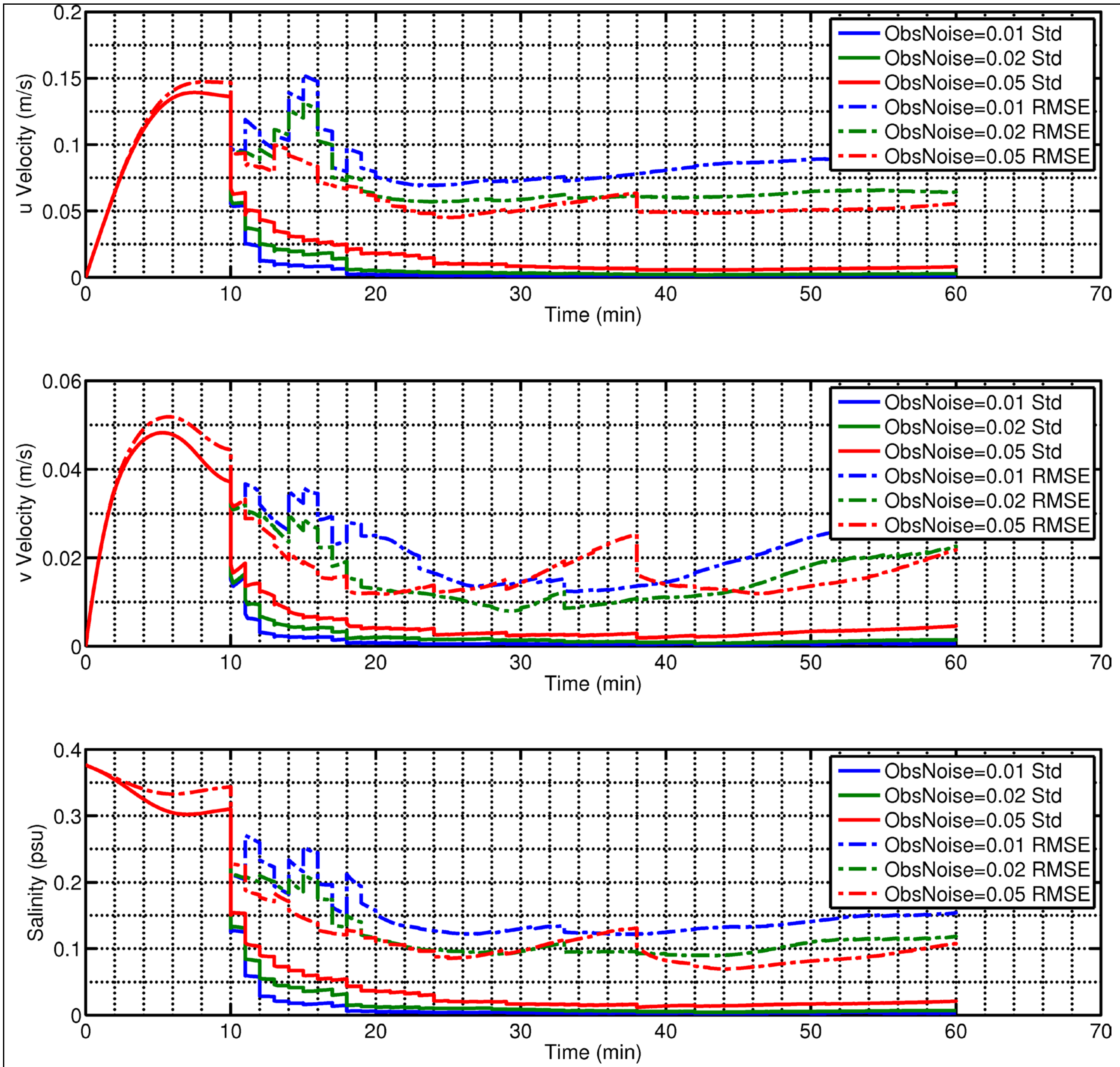


Figure 9: Time history plots of the RMSEs and standard deviations of u, v and salinity for three runs using the GMM-DO filter with observation standard deviation being 0.01, 0.02 and 0.05 times Delta S, respectively.