Ocean Modelling 131 (2018) 59–70

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Verifying and assessing the performance of the perturbation strategy in polynomial chaos ensemble forecasts of the circulation in the Gulf of Mexico

Shitao Wang[*,a], Guotu Li[d], Mohamed Iskandarani[a], Matthieu Le Hénaff[b,c], Omar M. Knio[d,e]

[a] Rosenstiel School for Marine and Atmospheric Sciences, University of Miami, Miami, FL, USA
[b] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL, USA
[c] NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, FL, USA
[d] Duke University, Durham, NC, USA
[e] King Abdullah University of Science and Technology, Saudi Arabia

ARTICLE INFO

Keywords:
Uncertainty quantification
Polynomial chaos
Relative entropy
Ocean modeling
Data depth

ABSTRACT

We present an analysis of two recent efforts aimed at quantifying the uncertainties in a 30-day HYbrid Coordinate Ocean Model forecast of the circulation in the Gulf of Mexico, with particular emphasis on the separation of Loop Current Eddy Franklin, using Polynomial Chaos methods. The analysis herein explores whether the model perturbations lead to realistic representation of the uncertainty in the Gulf Circulation. Comparisons of model output with Sea Surface Height and current mooring data show that the observational data generally falls within the envelope of the ensemble and that the modal decomposition delivers "realistic" perturbations in the Loop Current region. We use information theory metrics to quantify the information gain and the computational trade-offs between different wind and initial conditions perturbation modes. The relative entropy measures indicate that two modes for initial condition perturbations are enough, in our model configuration, to represent the uncertainty in the Loop Current region; while two modes for wind forcing perturbations are necessary in order to estimate the uncertainty in the coastal zone. The ensemble statistics are then explored using the Polynomial Chaos surrogate and the newly developed contour boxplot methods.

1. Introduction

The 2010 Deepwater Horizon oil spill underscored the need for reliable oceanic and atmospheric forecasts in order to predict the trajectory and evolution of the oil spill. Forecasting systems are, however, inherently uncertain because of uncertainties in, among other things, the input data used to produce these forecasts such as initial conditions, boundary conditions, and subgrid parametrization. Useful forecasts need to quantify the uncertainties in their predictions so that the reliability of the forecast can be assessed.

The present article analyzes the performance of two recent efforts (Iskandarani et al., 2016a; Li et al., 2016) that have relied on Polynomial Chaos (PC) methods (Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002; Le Maître and Knio, 2010; Iskandarani et al., 2016b) to quantify the uncertainties in forecasting the circulation in the Gulf of Mexico stemming from uncertainties in the initial conditions alone (Iskandarani et al., 2016a) or in combination with wind forcing uncertainties (Li et al., 2016). The forecast timeline covers the oil spill period from May 1–30, 2010 and coincides with an extended Loop Current (LC) that threatened to spread the oil along the south Florida coast and, eventually, the Eastern Seaboard of the United States. Fortunately, a LC detachment (LC eddy Franklin) occured and confined the oil to the northern and central parts of the Gulf of Mexico. The uncertainty analysis explores primarily whether the uncertainty in the LC location can be quantified given the uncertainties in the forecast model's data.

The studies in Iskandarani et al. (2016a) and Li et al. (2016) were based on perturbing the model fields (initial conditions and wind forcing) with space-time patterns obtained from an Empirical Orthogonal Function (EOF) decomposition where the amplitudes of these patterns were considered uncertain parameters. The PC formalism was then applied to propagate the uncertainties forward efficiently by: first, running an ensemble of simulations using the HYbrid Coordinate Ocean Model (HYCOM) to sample the uncertain parameter space; second, constructing polynomial-based model-surrogates that accurately represent the changes in model outputs caused by changes in model

* Corresponding author.
E-mail addresses: swang@rsmas.miami.edu (S. Wang), guotu.li@duke.edu (G. Li), miskandarani@rsmas.miami.edu (M. Iskandarani), mlehenaff@rsmas.miami.edu (M. Le Hénaff), omar.knio@duke.edu (O.M. Knio).

https://doi.org/10.1016/j.ocemod.2018.09.002
Received 9 August 2017; Received in revised form 10 July 2018; Accepted 5 September 2018
Available online 05 September 2018
1463-5003/

inputs; and third, using these surrogates to perform a reliable and efficient statistical analysis once the validity of the surrogates was established.

The choices made during the course of the uncertainty analysis in those two articles, and which will be detailed in later sections, have raised a number of issues that we wish to address here concerning the "realism" of the uncertainty analysis, the computational and information trade-offs in choosing different uncertain inputs, and the exploration of the statistical information conveyed by the PC approach. Specifically, in the present study, we 1) assess the performance of the EOF-perturbed PC-ensemble by comparing it to observational data, both at the surface and at depth, to verify whether the measurement data falls within the envelope of the PC ensemble; 2) leverage the ability of PC methods to deliver output Probability Density Function (PDF) to quantify, using information theoretical measures, the uncertainty lost by omitting some uncertain inputs or by limiting their variability. A second aim of this paper is to explore the statistics of the ensemble. In order to obtain the most representative ensemble member and to identify the outliers, contour boxplot (Whitaker et al., 2013), a generalization of the conventional boxplot, is applied to the ensemble. Furthermore, the output PDFs delivered by the PC method are used to explore the non-Gaussian statistics in the vicinity of the LC region.

In summary, the present article is a follow on to Iskandarani et al. (2016a), Li et al. (2016) and Iskandarani et al. (2016a) identified the two leading EOF modes whose amplitudes represented the uncertainties in the strength of an LC Frontal Eddy; these modes were subsequently used to perturb the initial conditions of a control forecast. Iskandarani et al. (2016a) relied on a 49 member ensemble to build surrogates of model outputs, validated their accuracy and used them for the statistical analysis. Li et al. (2016) expanded the previous study by including additional EOFs modes in the initial conditions perturbations as well as perturbations to the surface wind forcing. Their parameter space was eight-dimensional and required a compressed-sensing based procedure to construct model surrogates using a 798-member ensemble. A variance-based sensitivity analysis showed that uncertainties in the initial conditions dominated the forecast uncertainties in the deep parts of the Gulf of Mexico while wind forcing uncertainties were the dominant contributors on the continental shelves. The present study compares the ensembles simulations to observations to assess whether the EOF perturbations were adequate at representing the uncertainties in the forecast, performs a cost-benefit analysis regarding the enlargement of the uncertain parameter space and perform additional analysis regarding the statistical distribution of sea surface height at the end of the forecast. No additional experiments were performed in the present study.

The layout of this paper is as follows. Section 2 provides a quick overview of the LC dynamics in the Gulf of Mexico, summarizes the experimental setup of the two uncertainty experiments, provides a brief description of the PC methodology and describes the specification of the input uncertainties. Section 3 compares the ensemble results against observational data. The information trade-offs between the different choices of the sources and variability of the input uncertainties are shown in Section 4. Section 5 presents the contour boxplot of the LC edge and the sea surface height (SSH) PDFs. Finally, we conclude with a summary section.

2. Model and ensemble prediction

The Gulf of Mexico, where the Deepwater Horizon oil spill took place, is a suitable test bed for uncertainty studies. It is a well-observed regional sea that presents many dynamical features typical of the deep ocean such as currents and eddying jets. As shown in Fig. 1, the LC is a particularly dominant feature of the circulation in the Gulf of Mexico as it flows from the Yucatan Channel between Mexico and Cuba, to the Straits of Florida between Cuba and the Southeastern U.S. The LC presents a time varying extension, from a retracted path at the south of the basin, to an extended one reaching the edge of the continental shelf in the northeastern Gulf. When it is extended, the LC sheds a large, anticyclonic eddy, called LC Eddy (indicated by the anticyclonic arrow, in black, in the western Gulf), which then drifts westward, and the LC retracts to the south. This shedding sequence often implies temporary detachments of the LC Eddy from the current, before final separation. Small, cyclonic eddies, also called LC Frontal Eddies (shown in white arrows), at the edge of the LC play an active role in necking down and chopping the extended LC, leading to the LC Eddy detachments or separation (Zavala-Hidalgo et al., 2003; Schmitz, 2005; Athié et al., 2012; Le Hénaff et al., 2012a, 2014). The Deepwater Horizon oil spill took place during such a LC Eddy shedding sequence, and the fate of the spilled oil was partly influenced by the LC evolution and its frontal dynamics (Walker et al., 2011). The model setup described below was configured primarily to investigate the uncertainties in this eddy shedding scenario.

2.1. HYCOM setup

The forecast model is the Hybrid Coordinate Ocean Model (HYCOM). The model configuration is the same as GOMl0.04 expt_20.1 run by the Navy Research Laboratory (NRL) for the near-real time system in the period 2003–2010. The details of this configuration can be found on the HYCOM website[1]. The model has a horizontal grid resolution of 1/25 ° and 20 vertical layers. Since the vertical layers in HYCOM are hybrid, their thickness changes at each time step. In this configuration, there are more vertical layers toward the surface, with their depth, in the Eastern Gulf, ranging from 1.5m to about 2700 m in the Eastern Gulf. The computational domain is open along portions of its southern, eastern and northern boundaries, where values are provided by a lower resolution 1/12° North Atlantic HYCOM simulation (Chassignet et al., 2007). This model configuration has been used extensively in the literature, especially in studies of the Deepwater Horizon oil spill (e.g. Mezić et al., 2010; Valentine et al., 2012; Le Hénaff et al., 2012b; Paris et al., 2012). The model is forced by the 27 km resolution Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS) atmospheric products. The initial condition for the model is from the expt_20.1 (McDonald, 2006) near-real time simulation run at NRL, which includes data assimilation. The model assimilates available satellite altimeter observations (along track data altimetry obtained via the NAVOCEANO Altimeter Data Fusion Center), satellite and in situ sea surface temperature (SST) as well as available in situ vertical temperature and salinity profiles from XBTs, ARGO floats and moored buoys. The model is then integrated forward in time without data assimilation, in forecast mode, for 30 days from May 1, 2010 to May 30, 2010.

2.2. PC surrogates

We give a brief overview of PC methods in order to set the stage for the subsequent analysis; more background information can be found in Le Maître and Knio (2010) and Iskandarani et al. (2016b) and references therein. The PC paradigm is based on describing the dependence of a specific model output, say $M(\boldsymbol{x}, t, \boldsymbol{\xi})$ where $\boldsymbol{\xi}$ represents the vector of uncertain inputs and $\boldsymbol{x}$ and t refer to space and time, by a spectral series M_P of the form:

$$M(\boldsymbol{x}, t, \boldsymbol{\xi}) \approx M_P(\boldsymbol{x}, t, \boldsymbol{\xi}) = \sum_{k=0}^{P} \widehat{M}_k(\boldsymbol{x}, t)\Psi_k(\boldsymbol{\xi}) \tag{1}$$

where the $\Psi_m(\boldsymbol{\xi})$ are the user specified multi-dimensional basis functions (usually tensorized orthogonal polynomials from the Askey family (Xiu and Karniadakis, 2002)), and the $\widehat{M}_k(\boldsymbol{x}, t)$ are $(P+1)$ coefficients.

[1] https://hycom.org/data/goml0pt04/expt-20pt1 (last access on July 3rd, 2018).