

# 「日本沿海予測可能性実験」のための海洋モデル大規模計算: JCOPE-Tを用いた、データ同化と予測の不確実性評価のためのアンサンブル生成

課題責任者

宮澤 泰正 海洋研究開発機構 付加価値情報創生部門 アプリケーションラボ

著者

宮澤 泰正\*<sup>1</sup>, Varlamov M. Sergey\*<sup>1</sup>, Max Yaremchuk\*<sup>2</sup>

\*1 海洋研究開発機構 付加価値情報創生部門 アプリケーションラボ

\*2 Naval Research Laboratory

地球シミュレータ上で、海洋循環モデルによるアンサンブル生成システムを開発し、4次元変分法データ同化と予測の不確実性推定を行った。4次元変分法のアルゴリズムは、予測誤差と海洋力学発展の情報を組み合わせてアンサンブルメンバーを生成し、アンサンブルを用いて観測とシミュレーションの差を表わすコスト関数を最小化する手法である。ここではさらに、モデルのふるまいに依存した誤差特性を表現するため、コスト関数における背景誤差共分散行列を、アンサンブルメンバーから計算した共分散に置きかえる変更を試みた。

**キーワード:** 海洋循環モデル, データ同化, a4dVar, アンサンブル, 北西太平洋

## 1. はじめに

日本南岸の黒潮流路変動を予測するスキルを向上させるために、アンサンブルベースの四次元変分法同化システムをES4上で開発した。これまで、三次元変分法で初期化した渦解像海洋循環モデルを用いて予測実験を繰り返して来たが、三次元変分法は基本的に想定する対象期間内における同化データの時間依存性を無視している[1]。2020年3月から10月において、黒潮大蛇行の先端から冷水渦が頻繁に切離された。この期間に四次元変分法を適用することにより、衛星海面高度偏差(SSHA)データによって表現される時間変動の情報を活用しスキルを改善する可能性を調べる。ここで用いた手法(a4dVar)は、海洋循環モデルランのアンサンブル生成に基づいているため、実用的な予測能力を確立するには、対象となる海洋循環モデルJCOPE-Tの効率的な計算が必要である。ここでは、ES4上におけるJCOPE-Tの計算効率向上と、これを用いたa4darシステムの実装について報告する。

## 2. ES4上でのJCOPE-T計算

アンサンブル同化システムの開発には、北西太平洋地域(10.5-62N, 108-180E)を対象とする外洋域JCOPE-T[2]による領域海洋循環モデルを適用した。水平解像度は1/12度(約9km)であり、鉛直シグマレベル数は46とした。このバージョンでは水平粘性・拡散についてはスマゴリンスキー型の係数(定数0.1)をかけた倍調和演算子で表現し、係数の最低値は $10^4 \text{m}^4/\text{s}^2$ とした。

昨年度、前世代の地球シミュレータ(ES3)において、JCOPE-Tコード内のインライン展開とdoループ順序の変更を施した[2]。これにより北西太平洋モデルの計算効率は約37%向上した(表1)。今年度、ES3からES4に切り替えるだけで効率が向上し、「Mpi\_Barrier」の関数呼び出しの挿入などの追加変更により効率がさらに向上した(表1)。アンサン

ブルランには最新バージョンのコード(1.30Tflops)を使用した。

表1. JCOPE-T計算コードの計算効率

	Default (Tflops sec/d ay)	Optimized (Tflops sec/d ay)
ES3 (32PE)	0.13   149	0.21   84
ES4 (64PE)	0.76   50	1.30   30

## 3. アジョイントフリー四次元変分法同化システム

ES4上で計算効率化したJCOPE-T北西太平洋モデルを対象として、アジョイントモデルを使わない(アジョイントフリー)四次元変分法[3]を用いてアンサンブル同化システムを構築した(JCOPE-a4dVar)。JCOPE-a4dVarは、マルチスケール3次元同化(ms3dVar)[1]によって作成した初期推定値から、10日の同化期間に対し22個のアンサンブルメンバーを生成する。

a4dVarとms3dVarはどちらも、軌道上で計測されたSSHAと海面水温解析データ(MGDSST)[4]を同化している。ms3dVarは、高解像度の衛星海面水温(ひまわり8号)と現場水温-塩分プロファイルデータ(GTSPP)を同化している。a4dVarアルゴリズムは、主として同化期間内で計測されたSSHAから海洋変動の時間依存情報をとりだすように設計されている[3]。同化期間の長さは、主要な衛星であるJasonシリーズ星の回帰周期である10日とした。すべてのSSHAデータは同化期間内で日毎に同化される。MGDSSTは初期値の日付のデータのみを同化した。SSHAとMGDSSTを同化する際の観測誤差はそれぞれ0.1m、2度Cとした。

アンサンブルメンバーは、同化期間中のモデル力学とモデルとデータの残差の情報を組み合わせ、コスト関数最小化の各反復ステップで生成される[5]。初期推定値に対する修正は、各反復ステップでアンサンブルメンバーの摂動の加重和を加えることで繰り返し行う。コスト関数減少率はおおむね10回の反復後には小さくなるため、反復回数は各同化期間に対し10回に固定した。

同化のスキルは、同化期間経過後のSSHA、海面漂流ブイ、水温・塩分プロファイル、および潮位計の潮位偏差など、同化プロセスと無関係(独立)な観測データによって検証している。海上保安庁が日毎に推定している黒潮流路位置も検証に使っている。a4dVarシステムのパラメーターは、独立データによる検証によって定めた。

a4dVarシステムの最も重要なパラメーターは、同化プロセスにおけるシミュレーション誤差を規定する背景誤差共分散(BEC)の表現に含まれている。a4dVarシステムにおけるBECは以下のように表される[3]。

$$BEC = C_s L V_t C_s(S) V_t L \quad (1)$$

$C_s$ はコスト関数におけるシミュレーション誤差の観測誤差に対する相対的な大きさを規定する無次元定数である。 $L$ は地衡流関係を記述する行列演算子、 $V_t$ は各制御変数(水温、塩分、流速、水位)のシミュレーション誤差の大きさの三次元分布を記述する対角行列、 $C_s(S)$ はスケールパラメーターで特徴づけられる指数関数型の水平相関を表現する拡散演算子である。上付きの $t$ は行列の転置を示す。

$V_t$ は  $V_t = \min(STD_T, MAX_T)$ と表わされる。 $STD_T$ は海洋再解析データJCOPE2Mから計算した各変数の標準偏差(期間1993-2012)であり、 $MAX_T$ は最大値指定である。各変数の最大値指定は試行的な実験を行ったうえで、 $MAX_{temperature}$ ,  $MAX_{salinity}$ ,  $MAX_{velocities}$ ,  $MAX_{sea\ level}$ はそれぞれ、0.5 (deg.C), 0.05(psu), 0.1 (m/s), 0.05 (m)とした。水平空間スケール  $S$ , と無次元定数  $C_B$ はそれぞれ 4格子点間隔(約36km)、0.4とした。

SSHA観測の同化には、対応するモデルSSHAを求めるための期間平均SSHが必要であり、1993年から2012年までの海洋再解析データJCOPE2Mの時間平均SSHを使用した。これは、COPERNICUSによって提供されている観測SSHAの平均参照期間と同じである。さらに、NCEP/ NCA Rデータの風で駆動されるJCOPE2MとNCEP-CFSデータの風で駆動されるJCOPE-Tの海面高度の領域平均的な違いを考慮して、元のSSHから20cmを差し引き平均SSHを修正している。また、絶対値0.8mを超えるSSHA観測データは、同化した場合黒潮続流前線付近の水温・塩分構造が異常になる場合があるので、同化プロセスから除外した。

#### 4. ES4におけるa4dVar実験

2020年3月8日から10月18日までの期間、8日間隔でa4dVar実験を実施した。初期条件の最初の推定値は、同じ期間の3dVar同化プロダクトから与えた。コスト関数最小化の比は初期推定値(3dVar)の場合に比べ0.6から0.7の範囲で推移し、3dVar初期推定値から開始したフォワードランでのSSHA観測の残差を同化期間内でさらに縮減できることを示す。同化していない観測データを使った検証により、同化期間(10日間)および

それに続くハインドキャスト期間(60日間)におけるスキルが3dVar初期推定値の場合に比べて向上することがわかった。

アンサンブルの情報をさらに活用するため、拡散演算子 $C_d(s)$ (‘Diffusion BEC’)をアンサンブルから計算した共分散 $C_e$ (‘Ensemble BEC’)に置きかえて背景誤差共分散(BEC)を変更することによるスキル改善の可能性についても検証を進めている。アンサンブル共分散の計算は、ES4上の85個のスカラータイプCPUを使用して、各鉛直層(4つの3次元21zレベル(シグマレベルではない)と1つの2次元変数1rレベル)の各変数に対して並列に行っている。

表2は、Ensemble BECの実行に必要な計算資源が、Diffusion BECの実行と比較してCPU計算部分において増加することを示している。EnsembleとDiffusionの両方の実行で、実際の経過時間は1同化期間(10日間)に対しおおむね4~5時間となった。経過時間の変動については、適用したBECの種別よりも、ジョブ割り当て時におけるES4の全体ユーザー数による影響が大きかった。

表2. a4dVarの1同化期間実行に要した計算資源量

BEC type	Vector (Resource Set Hours (RSH))	Scalar (RSH)
Diffusion BEC	150	1.5
Ensemble BEC	150	22

1同化期間実行(Ensemble BEC)に要するベクトルとスカラータイプの資源使用量は、本課題に割り当てられた資源量(ベクトル50,000、スカラータイプ15,000 RSH)のそれぞれ0.3%と0.15%であった。これにより、a4dVarシステムの運用がES4上で問題なく行えると判断した。現在、実験結果をまとめ学術論文として公表するべく準備中である。

#### 謝辞

地球情報科学技術センター(CEIST)支援グループの皆様のご支援に心より感謝申し上げます。本報告に記載したコード最適化は、同支援グループの皆様へ実施していただきました。

#### 文献

- [1] Miyazawa, Y. et al., “Assimilation of high resolution sea surface temperature data into an operational nowcast/forecast system around Japan using a multi scale three dimensional variational scheme”, Ocean Dynamics 67, 713-728, (2017)
- [2] Varlamov, S. M. and Miyazawa, Y., “High-performance computing of ocean models for Japan

Coastal Ocean Predictability Experiment; A parallelized Sigma-coordinate ocean model JCOPE-T”, Annual Report of the Earth Simulator April 2020 -February 2021, II-3, 4-6, (2021)

[3] Miyazawa, Y. et al., “Applying the adjoint-free 4dVar

assimilation to modeling the Kuroshio south of Japan”, Ocean Dynamics 70, 1129-1149, (2020)

[4] Kurihara Y. et al., “Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and in-situ observations” (in Japanese), Weather Bull 73, s1-s18, (2006)

[5] Yaremchuk, M. et al., “A hybrid approach to generating search subspaces in dynamically constrained 4-dimensional data assimilation”, Ocean Modelling 117, 41-51, (2017)

# High-performance Computing of Ocean Models for Japan Coastal Ocean Predictability Experiment: Ensemble generation for data assimilation and estimation of prediction uncertainty using JCOPE-T

## Project Representative

Yasumasa Miyazawa, Application Laboratory, Research Institute for Value Added Information Generation Japan Agency for Marine Earth Science and Technology

## Authors

Yasumasa Miyazawa <sup>\*1</sup>, Sergey M. Varlamov <sup>\*1</sup>, Max Yaremchuk <sup>\*2</sup>

<sup>\*1</sup> Application Laboratory, Research Institute for Value Added Information Generation, Japan Agency for Marine Earth Science and Technology

<sup>\*2</sup> Naval Research Laboratory

We have developed an ensemble generation system based on an ocean circulation model on the Earth Simulator 4 (ES4) for four-dimensional variational (4dVar) data assimilation (DA) and uncertainty estimation in prediction. The adjoint-free 4dVar (a4dVar) algorithm generates ensemble members by hybrid combination of prediction error information and evolution of ocean dynamics. The ensemble members are utilized for minimization of the 4dVar cost function, penalizing the distance between observed and simulated oceanic states. For representing the flow-dependent error characteristics, the background error covariance in the cost function is modified by replacing the usual horizontal diffusion operator with the sample covariance calculated from the ensemble members.

**Keywords :** ocean circulation model, data assimilation, a4dVar, ensemble, Northwest Pacific

## 1. Introduction

To improve skills in predicting the Kuroshio path variations south of Japan, we have developed an ensemble-based 4dVar DA system on ES4. So far we have repeatedly performed the forecast experiments using an eddy resolving ocean circulation model initialized by the 3dVar DA method, which basically ignores time dependence of assimilated data within an assumed time window [1]. For a period from March to October 2020, cold eddies were frequently ejected from the tip of the Kuroshio Large Meander. By applying 4dVar to this period, we investigate the feasibility of utilizing time dependent information represented by the satellite sea surface height anomaly (SSHA) data. Since our method a4dVar is based on ensemble generation of ocean circulation model runs, effective calculation of a targeted ocean circulation model JCOPE-T is required for establishing practical forecast ability. Here we report improvements in numerical efficiency of JCOPE-T and implementation of the a4dar using JCOPE-T on ES4.

## 2. JCOPE-T calculation on ES4

A regional ocean circulation model based on the JCOPE-T open ocean version [2] covering a Northwestern Pacific region (10.5-62N, 108-180E) was applied for developing the ensemble DA system. The horizontal resolution and number of active vertical layers are 1/12 degree (9km) and 46 sigma-levels, respectively. In this version, horizontal viscosity and diffusivity are represented by biharmonic operators multiplied by a Smagorinsky-type coefficient with a constant parameter of 0.1. To stabilize the calculation, we apply a minimum value of biharmonic diffusion constant of  $10^9 \text{ m}^4/\text{s}^2$ .

In the last fiscal year, the code of JCOPE-T was modified on the previous generation of the Earth Simulator (ES3) by

combining inline expansion of the code and re-allocation of do loop orders [2]. The computational efficiency of the Northwestern Pacific model was improved approximately by 37% (Table 1). In this fiscal year, the efficiency was improved by just switching from ES3 to ES4, and further improved at ES4 by additional modification including insertions of function calls of 'Mpi\_Barrier' (Table 1). We used the latest version of the code with 1.30 Tflops for the ensemble simulations.

Table 1. Computational efficiency of the JCOPE-T code

	Default (Tflops sec/day)	Optimized (Tflops sec/day)
ES3 (32PE)	0.13 149	0.21  84
ES4 (64PE)	0.76  50	1.30  30

## 3. Adjoint-free 4dVar DA system

Based on the JCOPE-T Northwestern Pacific model optimized on ES4, we have constructed an ensemble DA system using the adjoint-free 4dVar method [3] (JCOPE-a4dVar). JCOPE-a4dVar creates 22 ensemble members from the first guess initial condition produced by the multi-scale three dimensional DA method (ms3dVar) [1].

Both a4dVar and ms3dVar assimilate the along-track SSHA and the synthetic merged sea surface temperature data (MGDSST) [4]. In addition, ms3dVar assimilates high-resolution satellite sea surface temperature (Himawari-8) and *in situ* temperature-salinity profile data (GTSP). The a4dVar algorithm is designed for mainly extracting the time-dependent observed information on the oceanic variability from SSHA during the time window [3]. The length of the

time window is 10-day, which is a repeated period of a typical Jason-series satellite altimeter, and all SSHA data within the time window are assimilated daily. MGDSST is assimilated only at the time of the initialization. The observation errors in assimilating SSHA and MGDSST are fixed as 0.1m and 2 deg.C, respectively.

The ensemble members are created at each iteration step of the a4dVar cost function minimization based on the combined information of model dynamics and model-data misfit during the time window [5]. The correction to the first guess initial condition is iteratively made as the weighted sum of the ensemble member perturbations at each iteration step. The iteration counts are fixed as 10 times for a time window, because the reduction of the cost function is generally slowed down after 10 iterations.

The DA skills are validated against independent observations such as SSHA measured after the time window, sea surface drifters, temperature/salinity profiles, and the tide gauge sea level anomaly. The Kuroshio path locations evaluated by the Japan Coast Guard on the daily basis are also used for the validation. Some critical parameters in the a4dVar system are iteratively tuned through the validation using the independent data.

The most critical parameters of the a4dVar system are included in the background error covariance (BEC) prescribing the simulation errors in the DA process. The BEC in the a4dVar system [3] is represented as,

$$BEC = C_B L V_T C_d(S) V_T L^t \quad (1)$$

, where  $C_B$  denotes a non-dimensional scalar constant determining the whole weight of the simulation errors in the cost function,  $L$  denotes an operator describing geostrophic balance,  $V_T$  denotes a diagonal matrix representing a three-dimensional distribution of the simulation error magnitude for each control variable (temperature, salinity, horizontal velocities, and sea level), and  $C_d(S)$  denotes a non-dimensional horizontal diffusion operator, which represents horizontal correlation characterized by an exponential function with a scale parameter  $S$ , and a superscript 't' means a transpose of a matrix.

$V_T$  is represented as  $V_T = \min(STD_T, MAX_T)$ , where  $STD_T$  is standard deviation of a target variable calculated from an ocean reanalysis product JCOPE2M, and  $MAX_T$  is a maximum estimate. After conducting a number of preliminary experiments, we have determined  $MAX_{temperature}$ ,  $MAX_{salinity}$ ,  $MAX_{velocities}$ , and  $MAX_{sea\ level}$  are 0.5 (deg.C), 0.05(psu), 0.1 (m/s), 0.05 (m), respectively. A horizontal scale parameter  $S$ , and a constant  $C_B$  have been determined as 4 grid size, which approximately corresponds to 36 km, and 0.4, respectively.

The assimilation of the SSHA observation requires the mean SSH for evaluation of the model SSHA. We use the temporal mean SSH of the ocean reanalysis JCOPE2M for a period from 1993 to 2012, which is the same as the mean reference period of the observation SSHA provided by COPERNICUS. We further correct the mean SSH by subtracting 20cm from the original one by considering the possible difference of area mean SSH between JCOPE2M driven by the NCEP/NCAR wind and JCOPE-T driven by NCEP-CFS wind. Also we exclude the SSHA observation data larger than 0.8m from the assimilation process because the fitting to such a large SSHA results in the anomalous

temperature/salinity structure around the Kuroshio Extension front.

#### 4. The a4dVar experiments on ES4

We conducted the a4dVar run for a period from 8 March to 18 October 2020 with an interval of 8-day. The first guess initial conditions were extracted from the 3dVar DA product archive for the same period. The ratio of the cost function minimization generally ranges from 0.6 to 0.7, indicating that there is room to reduce the residuals of the SSHA observations in the forward runs starting from the first guess initial conditions at the exact time of measurement during the time window. The validation of the resulting products using the independent observation data suggests the skill improvements for the time window period (10 - day length) and for the succeeding hindcast period (60 - day length).

To further utilize the information represented in the produced ensemble, we examine possible skill changes by modifying the background error covariance (BEC) through replacing the diffusion operator  $C_d(s)$  ('Diffusion BEC') with the ensemble covariance  $C_e$  ('Ensemble BEC') directly calculated from the ensemble member variables. The calculation of the sample covariance is parallelized for each target variable for each vertical level (4 three-dimensional with 21 z-levels (not sigma levels) and 1 two-dimensional variable with 1-level) using 85 scalar-type CPUs on ES4.

Table 2 indicates that the computational resources required for the Ensemble BEC run increases in the CPU calculation part as compared to the Diffusion BEC run. Note that the actual elapsed time is generally 4-5hour for the 10-day time window in both Ensemble and Diffusion runs, which basically depends on the total user numbers of ES4 at the time of the job allocation rather than the applied BEC types.

Table 2. Resource usages of the 10-day window a4dVar run

BEC type	Vector (Resource Set Hours(RSH))	Scalar (RSH)
Diffusion BEC	150	1.5
Ensemble BEC	150	22

Both vector and scalar resource usages are 0.3% and 0.15% of the requested resources for the present theme: 50,000 and 15,000 (RSHs) respectively. We thus conclude that the operational a4dVar system is reasonably established on ES4. Note that the scientific analyses including the evaluation of the forecast uncertainty using the ensemble runs based on the all a4dVar experiments performed on ES4 will be reported soon in some academic journals (in preparation).

#### Acknowledgement

We deeply appreciate all support by the support staff of the Earth Simulator in the Center for Earth Information Science and Technology (CEIST). In particular, the code optimizations described in the report were actually done by them.

## References

- [1] Miyazawa, Y. et al., “Assimilation of high resolution sea surface temperature data into an operational nowcast/forecast system around Japan using a multi scale three dimensional variational scheme”, *Ocean Dynamics* 67, 713-728, (2017)
- [2] Varlamov, S. M. and Miyazawa, Y., “High-performance computing of ocean models for Japan Coastal Ocean Predictability Experiment; A parallelized Sigma-coordinate ocean model JCOPE-T”, *Annual Report of the Earth Simulator April2020 -February 2021*, II-3, 4-6, (2021)
- [3] Miyazawa, Y. et al., “Applying the adjoint-free 4dVar assimilation to modeling the Kuroshio south of Japan”, *Ocean Dynamics* 70, 1129-1149, (2020)
- [4] Kurihara Y. et al., “Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and in-situ observations” (in Japanese), *Weather Bull* 73, s1-s18, (2006)
- [5] Yaremchuk, M. et al., “A hybrid approach to generating search subspaces in dynamically constrained 4-dimensional data assimilation”, *Ocean Modelling* 117, 41–51, (2017)