Naval Research Laboratory
Washington, DC 20375-5320

NRL/7530/MR—2021/1

Impact of Model Error Techniques on the Forecast Skill of the Navy ESPC Ensemble

WILLIAM CRAWFORD

Atmospheric Dynamics and Prediction Branch
Marine Meteorology Division

April 28, 2021

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
28-04-2021	NRL Memorandum Report	01/21/2020 – 01/20/2021

4. TITLE AND SUBTITLE

Impact of Model Error Techniques on the Forecast Skill of the Navy ESPC Ensemble

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
NISE

6. AUTHOR(S)

William Crawford

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
N2W5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/7530/MR--2021/1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

10. SPONSOR / MONITOR'S ACRONYM(S)

NRL-NISE

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Karle Fellowship

14. ABSTRACT

The research presented here illustrates the impact of several model error techniques on the fidelity of extended-range global coupled ensemble forecasts produced by the Navy ESPC model. The presented methods aim to improve the forecasts by reducing errors in the forecasted state both at initial time and during the model integration. To address error in the initial state, we explore the use of Relaxation To Prior Perturbations (RTPP) which will aim to better capture model uncertainty in the initialization of the ensemble forecast by relaxing the initial state from the analysis (produced by the data assimilation system) back toward the prior (or forecasted state). During the model integration, we explore the use of two methods; 1) Analysis Correction-based Additive Inflation (ACAI) and Stochastic Kinetic Energy Backscatter (SKEB). Both of these methods will act as a representation of stochastic model error intended to increase the divergence of the ensemble; however, in the case of ACAI, there is also an explicit term in the perturbations aimed at reducing systematic errors. On the other hand, we have also found that the SKEB perturbations can act to modify the mean state resulting in improvements to the bias. All three model error techniques present clear improvements to the skill of the forecasts both in the short-range (weeks 1 and 2) and at extended range time scales (weeks 3-6). The Navy ESPC model is used operationally to generate 45-day forecasts, and these research findings present a clear pathway to improve the skill of our global coupled ensemble forecasting system.

15. SUBJECT TERMS

Navy ESPC Ensemble
Model error

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			William Crawford
U	U	U	SAR	25	**19b. TELEPHONE NUMBER** *(include area code)* (972) 336-1490

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

CONTENTS

FIGURES

TABLES

This page intentionally left blank.

EXECUTIVE SUMMARY

The research presented here illustrates the impact of several model error techniques on the fidelity of extended-range global coupled ensemble forecasts produced by the Navy ESPC model. The presented methods aim to improve the forecasts by reducing errors in the forecasted state both at initial time and during the model integration. To address error in the initial state, we explore the use of Relaxation To Prior Perturbations (RTPP) which will aim to better capture model uncertainty in the initialization of the ensemble forecast by relaxing the initial state from the analysis (produced by the data assimilation system) back toward the prior (or forecasted state). During the model integration, we explore the use of two methods; 1) Analysis Correction-based Additive Inflation (ACAI) and Stochastic Kinetic Energy Backscatter (SKEB). Both of these methods will act as a representation of stochastic model error intended to increase the divergence of the ensemble; however, in the case of ACAI, there is also an explicit term in the perturbations aimed at reducing systematic errors. On the other hand, we have also found that the SKEB perturbations can act to modify the mean state resulting in improvements to the bias. All three model error techniques present clear improvements to the skill of the forecasts both in the short-range (weeks 1 and 2) and at extended range time scales (weeks 3-6). The Navy ESPC model is used operationally to generate 45-day forecasts, and these research findings present a clear pathway to improve the skill of our global coupled ensemble forecasting system.

This report presents research conducted by Dr. William Crawford of the Marine Meteorology Division of NRL in Monterey, California. Dr. Crawford would like to thank the Karles Fellowship program for the opportunity to conduct this research.

This page intentionally left blank.

Impact of model error techniques on the forecast skill of the Navy ESPC ensemble

1. INTRODUCTION

At extended-range time scale (~2-6 weeks), where deterministic forecast skill is lost, probabilistic outlooks provided by an ensemble of forecasts become progressively more important for capturing the most likely environmental state. To fully capture the true uncertainty in a forecast, the ensemble system must be formulated to account for stochastic model error to ensure accurate representation of forecast uncertainty both at initial time and during the forecast. Stochastic error can arise from many sources including discretization and parameterization of unresolved process, and without proper representation, the ensuing ensemble forecast will be under-dispersive. Furthermore, at extended forecast lead-times as the model is drawn toward an attractor that may not align with reality, large biases may form in the forecasted state. It is important to also explore model error techniques that are capable of removing these systematic errors.

Presently, the Navy ESPC system uses an ensemble of data assimilations (EDA) with perturbed observations [1] to represent uncertainty in the initial state and has no method to account for uncertainty during the forecast. Recent work at NRL [2] illustrates that the EDA-based system results in an ensemble that is under-dispersive at initial time and does not identify modes of variability conditioned to grow in time. Similar results have been found in the EDA implementation at ECMWF where they find they must combine the EDA-perturbations with fast growing singular vectors to achieve a proper representation of model uncertainty [3].

The atmospheric component of the Navy ESPC model uses the Navy Global Environmental Model [4] run at the T359 horizontal resolution (~37km) with 60 vertical levels. The ocean and cryosphere are encompassed within the Global Ocean Forecasting System (GOFS) 3.1 [5] where the Hybrid Coordinate Model (HYCOM; [6]) and Los Alamos Community Ice model (CICE; [7]) are used to forecast the ocean and sea-ice states respectively. In the operational configuration of the Navy ESPC system, the ocean/ice resolution is set to 1/12°; however, in order to conserve valuable computational resources, many of the presented experiments are run with a 1/4° resolution for the ocean/ice. In both cases; HYCOM is set to run with 41 vertical levels. Two-way coupling between all model components is achieved through the Earth System Modeling Framework (ESMF; [8]) as shown in Figure 1. The data assimilation (DA) systems of the atmosphere (NAVDAS-AR; [9, 10]) and ocean/ice (NCODA; [11]) are weakly coupled in that neither system directly assimilates observations of the other medium. However; observational information will be exchanged between the model components through the 1-hourly forecast coupling. In this way, oceanic observations can influence the background of the subsequent atmospheric DA cycle and vice versa.

The Navy ESPC system has been transition to the Fleet Numerical Meteorology and Oceanography Center (FNMOC) for operational production of 45-day ensemble forecasts. The presented research outlines several model error techniques aimed at improving the forecast skill of the Navy ESPC system as measured through typical forecast skill metrics include bias, root mean squared error and ensemble spread skill. We also investigate impacts on more phenomenological based skill metrics including integrated vapor transport (IVT) important for forecasting atmospheric rivers as well as variables related to forecasting of the Madden-Julian Oscillation (MJO).

The report outlines the model error techniques explored in Section 2 followed by a description of the numerical experiments carried out in Section 3. Section 4 will describe results of each model error technique with conclusions given in Section 5.

Manuscript approved Month 00, 2021.

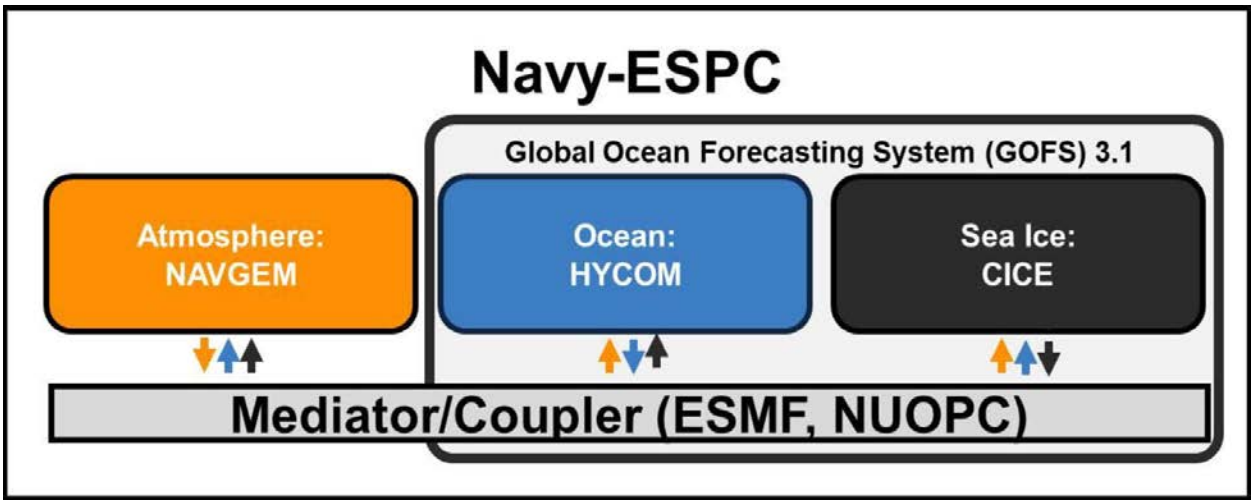

Figure 1 — Illustration of coupling between components of the Navy ESPC model. Reproduced from [2]

2. APPROACH

2.1 Relaxation To Prior Perturbations (RTPP)

In the Navy ESPC model, the atmospheric data assimilation system is run every 6 hours at 00, 06, 12 and 18UTC. During the data assimilation process, observations are combined with forecast information to arrive at the most accurate initial state. In the EDA framework of the Navy ESPC model, the assimilated observations are perturbed so that each ensemble member will have a slightly different initial (or analysis) state. However, it has been shown that the analysis uncertainty provided by the perturbed observations method does not adequately capture the true uncertainty (i.e. the ensemble is under-dispersive). Regardless of any uncertainty growth in the model over the prior 6-hour period of the forecast, the DA system will invariably draw the ensemble members back toward the observations, effectively reducing the ensemble variance at initial time. The method of relaxation to prior perturbations (RTPP; [12]) acts to combat the reduction of the ensemble variance by relaxing the ensemble analyses back toward the prior (or forecasted) state after the DA has run. By drawing the ensemble initial states back toward the prior forecast, RTPP allows the ensemble to retain some forecasted variance that has developed over the last 6-hours of the forecast.

The new ensemble perturbations are computed according to equation (1) where the f and a superscripts represent prior and analysis states respectively. The i subscript represents individual ensemble member states and the e subscript with an overbar represents an ensemble mean state.

$$\delta x_i^{RTPP} = \alpha \left[x_i^f - \overline{x_e^f} \right] + (1-\alpha)\left[x_i^a - \overline{x_e^a} \right] \qquad (1)$$

The RTPP perturbation for any ensemble member i (δx_i^{RTPP}) is then a weighted sum of its deviation from the mean forecast state ($\overline{x_e^f}$) and mean analysis state ($\overline{x_e^a}$). The relaxation parameter (α) determines the level of weighting between the forecast and analysis perturbations.

Given that the mostly likely estimate of the true initial state is represented by the ensemble mean analysis ($\overline{x_e^a}$), the ensemble is re-centered on this state. The ensemble of initial conditions (x_i') will then be given by

$$x_i' = \overline{x_e^a} + \delta x_i^{RTPP} \qquad (2).$$

2.2 Analysis Correction-based Additive Inflation (ACAI)

In addition to initial condition uncertainty, the ensemble system must be formulated to account for forecast uncertainty and systematic errors. One method explored as part of this research has been tested by several other operational forecasting centers around the world, most notably by the UK Meteorological Office [13, 14] and uses analysis corrections from the data assimilation (DA) system as a representation of model error. The DA system routinely generates analysis corrections to be applied to the forecast state to bring it closer to the observed state. We can therefore anticipate that the average of these analysis corrections can tell us something about the biases existent in our model and be used during the model forecast to reduce bias. Furthermore, in ensemble forecasting systems it is desirable to accurately capture the uncertainty in the ensemble mean by matching the ensemble spread to the error in the ensemble mean relative to some estimate of the true state (commonly referred to as spread-skill). A common problem in many ensemble forecasting systems is for the ensemble of forecasts to diverge too slowly, and therefore not accurately capture the uncertainty of future forecast states. To increase the divergence of our ensemble forecasts, we combined the mean analysis correction with a randomly sampled analysis correction from the archive as an additional representation of stochastic model error. We refer to this method as Analysis Correction-based Additive Inflation (ACAI; [15]).

2.2.1 Perturbation calculation

The ACAI perturbations draw upon an archive of analysis corrections which can be expressed as

$$\delta x_a = x_a(t) - M\big(x_a(t - \Delta t_a)\big) \qquad (3),$$

where x_a represents an analysis state valid at time t and $M\big(x_a(t - \Delta t_a)\big)$ represents a forecast state produced by advancing the analysis state valid at time $t - \Delta t_a$ by the non-linear model M. The analysis correction is then the difference between the analysis state and the forecasted state.

The ACAI perturbations are represented by (4) as a combination of a mean analysis correction and a randomly sampled analysis correction:

$$\delta x_i = \frac{1}{N_s}\sum_{k=1}^{N_s} \delta x_k^a + \beta\left[\delta x_{r_i}^a - \frac{1}{N_e}\sum_{j=1}^{N_e} \delta x_{r_j}^a\right] \qquad (4).$$

The first term on the right hand side represents an average of the N_s analysis corrections (δx_k^a) produced over some extended time period. The mean correction term is aimed at reducing bias in the forecast. The second term on the RHS is meant as an additional representation of stochastic model error and is aimed at increasing divergence in the ensemble. For each ensemble member, i, a random analysis correction is sampled from the same set of corrections used to produce the mean term (i.e. 1st term on RHS). Before adding to the mean correction, the mean of the N_e randomly sampled analysis corrections is subtracted to reduce any contamination to the mean correction. An optional scaling term for the random perturbations is given by β and is intended to control the overall impact of the random portion of the

ACAI perturbations. The ACAI perturbations are computed at each model grid point for surface pressure, temperature, humidity, zonal wind speed, and meridional wind speed.
The ACAI perturbations are divided by the number of time steps per 6-hour period of the forecast (T) and added as a tendency during the forecast integration as

$$\frac{dx_i}{dt} = f(x_i) + \frac{\delta x_i}{T} \qquad (5),$$

where $f(x_i)$ is the tendency term of the prognostic equation. A new ACAI perturbation for each ensemble member is computed for each 6-hour period of the forecast.

2.2.2 Mean analysis corrections

As stated above, the mean analysis corrections are a representation of the model bias since they illustrate how the data assimilation is adjusting the forecasted state on average. However; it has been found that the structure of the mean corrections vary considerably based on the time of day. Figure 2 shows the mean analysis corrections to surface level pressure and temperature at 00, 06, 12 and 18UTC. In the case of surface pressure, the primary adjustment is in the meridional pressure gradient with a decrease in pressure poleward of ~30° N/S and an increase in pressure in the tropics. However, a secondary signal indicates that the area of maximum adjustment to surface pressure migrates from east to west with the passage of a day (00Z➔18Z). Similarly, the positive adjustment to surface temperature over the Pacific Ocean/Australia at 00Z migrates westward to Eurasian and subsequently to the Americas by 18Z.

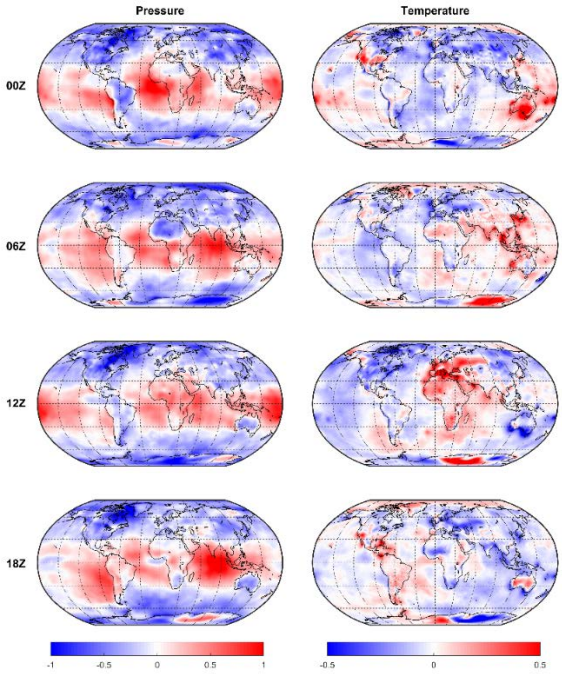

Figure 2 — Seasonal (3-month) average analysis corrections to surface pressure (left) and temperature (right) according to time-of-day (00, 06, 12 and 18Z).

For this reason, the perturbations computed in (4) only use analysis corrections from the archive that coincide with the time-of-day to which the correction will be applied.

2.3 Stochastic Kinetic Energy Backscatter (SKEB)

The Stochastic Kinetic Energy Backscatter (SKEB) method [16] was also tested as a method for representation of stochastic model error. SKEB was first introduced in the stand-alone NAVGEM atmospheric model [17] and has been retained as an option for the atmospheric component of the fully coupled Navy ESPC model. The implementation of SKEB applied here uses a moisture convergence mask to determine areas in which to make perturbations to the rotational component of the wind (vorticity). The magnitude of perturbations to zonal wind speed are comparable between ACAI are SKEB (not shown); however, ACAI also adds perturbations to temperature, surface pressure and humidity resulting in a greater perturbation to the system overall.

3. EXPERIMENTS

The Navy ESPC system is formulated to run the data assimilation cycle every 6 (24) hours for the atmosphere (ocean/ice) to ingest relevant observations and produce an updated analysis. Presently, extended range forecasts are initialized every Wednesday at 12UTC and integrated out to 45 days. Given the model error techniques described above are aimed at addressing both initial and forecast uncertainty, we will conduct a series of experiments using both the cycling and extended range forecasting systems.

Table 1 — List of numerical experiments

Experiment name	# of ensemble members	Time period	Model error technique	Forecast length
Cycling experiment				
EDA_{ctrl}	5	July 2017	POBS	6 hours
RTPP/SKEB	5	July 2017	POBS+RTPP+SKEB	6 hours
Extended-range forecasts				
ER_{ctrl}	7	Feb. 2017 - Jan. 2018	POBS	45 days
$ACAI_{SA}$	7	Feb. 2017 - Jan. 2018	POBS+ACAI (w/ static archive)	45 days
$ACAI_{MA}$	7	Feb. 2017 - Oct. 2017	POBS+ACAI (w/ moving archive)	45 days
SKEB	5	Feb. 2017 - Jan. 2018	POBS+SKEB	30 days

3.1 Cycling Experiments

To investigate the impact of RTPP on the baseline EDA system, a 1-month cycling experiment was run over the July 2017 time period with RTPP perturbations generated during each 6-hour update cycle of the atmospheric DA system and using $\alpha = 0.725$ (c.f. Eq. 1). A list of experiments is given in Table 1. Since RTPP is formulated to retain some of the forecast variance generated during the prior 6 hours, the short-range forecasts were also run with SKEB to allow for a further increase of the initial condition uncertainty and also uses the method of perturbed observations during the assimilation procedure. The RTPP/SKEB experiment will be compared to a baseline EDA experiment (EDA_{ctrl}; Table 1) run over the same period of time. The EDA_{ctrl} experiment also uses the method of perturbed observations to generate the ensemble, but does not use any other model error techniques.

3.2 Extended-range forecasts

A series of extended-range experiments were run using initial conditions from the EDA_{ctrl} cycling experiment describe in Section 3.1. All of the experiments were initialized from the same initial state, and therefore, any differences in forecast skill are due only to modifications to the forecast during the model integration. The control experiment (ER_{ctrl}; Table 1) to which all other forecasts will be compared was taken from the recent Validation Test Report issued to FNMOC as an illustration of the Initial Operational Capability of the Navy ESPC system [18]. In each of the extended-range experiments, ensemble forecasts are issued once per week on Wednesday at 12UTC. The time period of the forecasts is from 01 February 2017 to 24 January 2018 resulting in 52 extended range forecasts.

Two experiments testing the ACAI method in the Navy ESPC system were also run in overlapping periods with the EDA_{ctrl} experiment. The two ACAI-based experiments differ in their formulation of the analysis correction archive they draw upon as well as the period of time used to for sampling. The $ACAI_{SA}$ experiment (Table 1) uses a static archive of analysis corrections from the year 2011 when deriving the perturbations computed using Eq. (4). The average correction is computed using a 3-month average centered on the month of the forecast, and the randomly sampled correction will be drawn from the same 3-month period. In the $ACAI_{MA}$ experiment, the analysis corrections from 60-days prior to the forecast initialization date are used to derive the mean analysis correction and the random sample. The $ACAI_{MA}$ experiment is intended to explore methods to optimize the ACAI method for operations. By using a moving archive of analysis corrections, the $ACAI_{MA}$ method alleviates the dependence on maintaining an archive of analysis corrections, which can become particularly challenging when making large upgrades to the DA or forecast system. Additionally, the moving archive may be able to provide more relevant information about the model biases than using analysis corrections from an independent year (as in the $ACAI_{SA}$ experiment). The time period of the $ACAI_{SA}$ experiment mirrors that of the ER_{ctrl} experiment; however, due to computation constraints the $ACAI_{MA}$ experiment only runs from February to October 2017. In all ACAI-based experiments, there is no scaling of the random portion of the perturbations (i.e. $\beta = 1$).

A third extended range experiment was run using SKEB as a representation of model error. Again due to computational constraints, this experiment was run using only 5 ensemble members and the forecasts were only integrated out to 30-days and will be compared against the EDA_{ctrl} and $ACAI_{SA}$.

4. RESULTS

4.1 RTPP

RTPP has been implemented in the Navy ESPC model to address deficient spread in the ensemble at initial time. The short forecasts of each 6-hour cycle of the DA system is designed to produce the background state for the subsequent cycle. Each of these 6-hour short forecasts has been run with SKEB turned on to generate divergence in the model over this short period of time. RTPP then relaxes the analysis state back toward the prior state which will contain a more accurate representation of the analysis uncertainty. As the system cycles, more and more ensemble divergence will be retained by the system, allowing the initial states to increase their estimates of uncertainty. Figure 3 illustrates the impact of cycling the system with a combination of RTPP and SKEB by showing the variance ratio at initial time as a function of initialization date. The variance ratio is a measure of the spread-skill of the ensemble represented by the ratio of the squared ensemble variance to the bias corrected mean squared error (c.f. Appendix). In the VARR diagrams, a value of 1 indicates a perfect match between the ensemble variance and squared error of the ensemble mean.

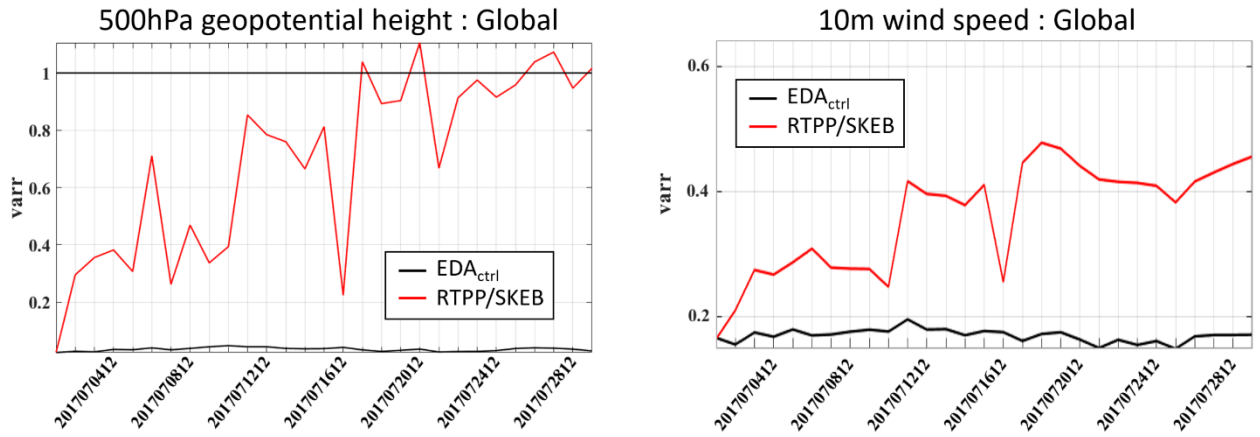

Figure 3 — Comparison of the globally averaged variance ratio (VARR) as a function of initialization date for 500hPa geopotential height (left) and 10m wind speed (right) for the EDA$_{ctrl}$ and RTPP/SKEB experiments.

As expected, at the beginning of the experiment the globally averaged variance ratios in 500hPa geopotential height and 10m wind speed are identical for the EDA$_{ctrl}$ and RTPP/SKEB experiment. However, as the system cycles the experiments diverge with RTPP/SKEB experiment progressively approaching the ideal value of 1. In both panels, the initial values of VARR are far below the ideal value, and in the case of the EDA$_{ctrl}$ experiment, remain very close to the deficient value throughout the entire experiment. In the case of 500hPa geopotential height, by the end of the 1-month time period the variance ratio has increased dramatically and is very close to 1 by the end of the experiment. While not as responsive as 500hPa geopotential height, the 10m wind speed variance ratio has also increase significantly from the EDA$_{ctrl}$ experiment.

The presented RTPP experiment used a relaxation value (α) of 0.725. This value was selected based on prior research [19] which shows that the ensemble root mean squared error (RMSE) and ensemble are highly sensitive to the relaxation parameter with both RMSE increasing drastically beyond a particular threshold. Figure 4 shows that using a value of 0.725 for relaxation keep the RMSE values of the RTPP/SKEB experiment close to those of the baseline system. Further testing has indicated that increasing the relaxation even to 0.8 greatly increases the RMSE of the RTPP/SKEB experiment.

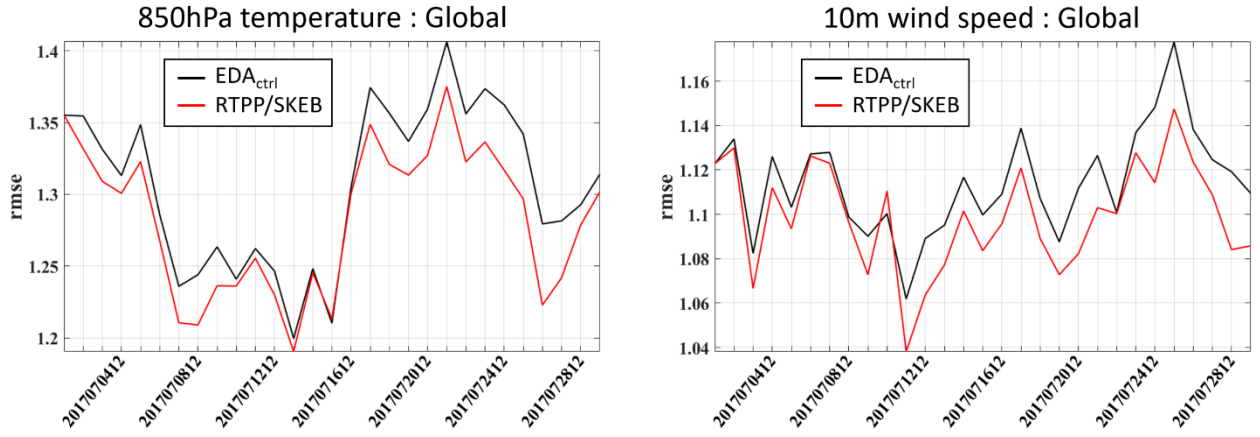

Figure 4 — Comparison of the globally averaged root mean squared error (RMSE) as a function of initialization date for 850hPa temperature (left) and 10m wind speed (right) for the EDA$_{ctrl}$ and RTPP/SKEB experiments.

4.2 ACAI

4.2.1 Forecast skill

Here we will examine improvements to Navy ESPC extended range forecasts by including the ACAI method to account for model error in the atmospheric component of the Navy ESPC model. Comparisons are made between two ACAI-based experiments ($ACAI_{SA}$ and $ACAI_{MA}$; Table 1) and a control experiment (ER_{ctrl}; Table 1). The two ACAI-based experiments presented differ in the formulation of the analysis correction archive used to derive the perturbations defined by Eq. (4) as described in Section 3.2. Forecast skill performance is analyzed using the variables, metrics and regions described in the Appendix. Figure 5 shows scorecards as a function of forecast lead-time for the $ACAI_{SA}$ experiment. The scorecards depict improvement or degradation relative to the ER_{ctrl} experiment based on verification against ECMWF analyses.

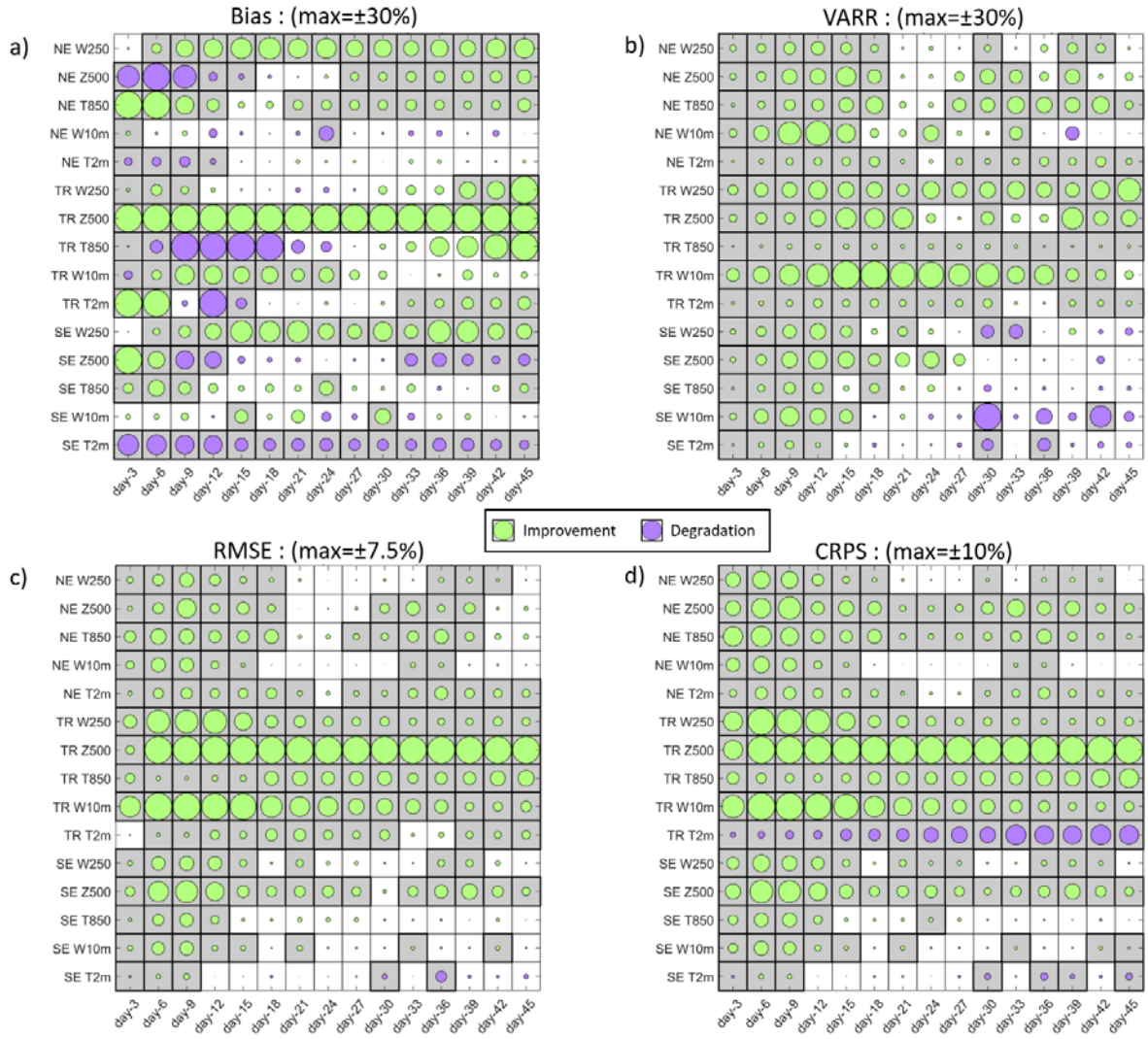

Figure 5 — Scorecards showing percent change in (a) bias, (b) VARR, (c) RMSE and (d) CRPS for the $ACAI_{SA}$ experiment comparted to ER_{ctrl}. Green (purple) circles represent improvement (degradation). Grey shading and bold outline represents statistical significance at the 95% level. Maximum percentage change represented by circle size is indicated in the panel titles. Scores for individual variables (see Appendix) shown on the vertical axis with forecast lead-time on the horizontal axis. All scores computed against ECMWF analyses.

Figure 5a shows the percent change in bias between the two experiments computed as:

$$100 * \frac{|experimental\ bias| - |control\ bias|}{|control\ bias|}$$

where $|\cdot|$ represents the absolute magnitude of the bias. Overall, the bias is considerably improved by the implementation of ACAI. In the extra-tropics, the improvements are most pronounced in tropospheric temperatures (T850) and jet level wind speeds (W250), while in the tropics the most significant impacts are on 10m wind speeds (W10m) and 500hPa geopotential height (Z500). In the case of Z500, a decrease in bias of more than 30% is shown for all lead-times out to 45-days.

While the bias has been decreased on average, there are a few areas of significant degradation (increase) in the model bias as well. Most notable are the degradations in northern extra-tropical geopotential height (NE Z500), tropical tropospheric temperatures (TR T850) and southern extra-tropical 2m air temperature (SE T2m). Figure 6 shows the true bias for NE Z500 and TR T850 for both the ER_{ctrl} and $ACAI_{SA}$ experiments. In the case of NE Z500, the average bias in both experiments is positive at the beginning of the forecast, but quickly decreases and passes through zero in the first few days and remain negative until the end of the forecast. Application of the ACAI perturbations appears to hasten the negative trend in bias resulting in degraded bias scores in the first half of the forecast (c.f. Figure 5a). On the other hand, as the forecast progresses, ACAI is able to slow the negative trend resulting in improved bias scores in NE Z500 by the end of the forecast. A somewhat similar response is seen in TR T850 (Figure 6) which shows that ACAI is able to reduce the negative trend in bias in the ER_{ctrl} experiment beginning on day-2 of the forecast. However, because the negative trend in the ET_{ctrl} is more subtle than in NE Z500, the bias in TR T850 appears drastically degraded at the beginning of the forecast (particularly days-9, 12, 15 and 18), but results in improved scores in the second half of the forecast (c.f. Figure 5a).

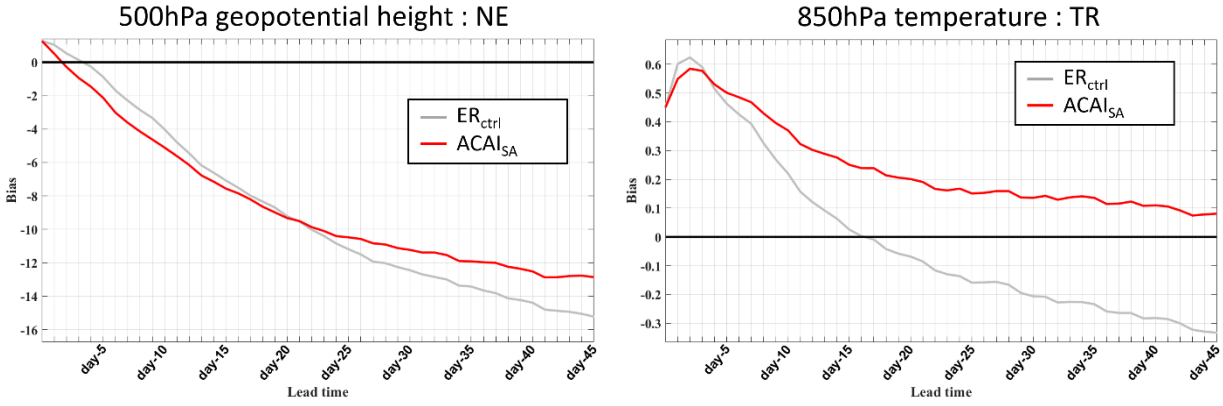

Figure 6 — Comparison of bias as a function of forecast lead-time for northern extra-tropical 500hPa geopotential height (left) and tropical 850hPa temperature (right) for the ER_{ctrl} and $ACAI_{SA}$ experiments.

One of the primary motivations for implementing ACAI in the Navy ESPC model was to improve the spread-skill of the ensemble forecasts which is measured here by the ratio of the ensemble variance to the bias corrected ensemble mean squared error (or VARR), and Figure 5b shows the scorecard results for this metric. The scorecard shows that ACAI is able to improve the spread-skill as early as 3-days into the forecast, but is most impactful in the first half of the forecast. The ACAI perturbations continue to be impactful in the 2nd half of the forecast in the northern extra-tropics and tropics, but begin to decrease in impact in the southern extra-tropics. In some cases, ACAI appears to even degrade the spread-skill in the

southern extra-tropics; however, many of these are shown to not be statistically significant. The largest and most consistent positive impacts are seen in the tropical regions, with 500hPa geopotential height and wind speeds seeing the largest improvements. In particular, the spread-skill in 10m wind speed is improved by ~20-30% for the entire 45-day forecast.

Figure 5c and 5d show the RMSE and CRPS scorecards of the $ACAI_{SA}$ experiment respectively. In both cases, the largest improvements are see in the tropical regions with the largest impact being on Z500, W10m and W250. In all three variables, the largest improvements are on the order of 7.5% and 10% for RMSE and CRPS respectively. The only degradation in RMSE is in SE T2m and only two of these are indicated as statistically significant. CRPS, however, shows a large degradation in TR T2m from the outset of the forecast, and even increases in magnitude with forecast lead-time. It has been found that the degradation in TR T2m is dominated by a regional degradation focused in the eastern tropical Pacific. Variability in the eastern tropical Pacific is dominated by fluctuations driven by the El Nino/Southern Oscillation (ENSO) a dominate mode of variability in the global ocean and atmosphere. Given the ACAI perturbations in the $ACAI_{SA}$ experiment are drawn from an independent year (2011), it is anticipated that some of this degradation may be due to an incorrect estimate of the model bias by the mean analysis correction.

Contrary to the static analysis correction archive of the $ACAI_{SA}$ experiment, the $ACAI_{MA}$ experiment uses analysis corrections from the 60-days prior to the forecast initialization date to compute the perturbations in Eq. 4. While it can be anticipated that $ACAI_{MA}$ perturbations will have a more reliable estimate of the current model bias, the $ACAI_{SA}$ has the advantage of additional information provided from a 3-month centered mean used when computing the perturbations. Figure 7 shows scorecard summations for the $ACAI_{SA}$ and $ACAI_{MA}$ experiments. The two experiments are competitive for the first ~2-weeks of the forecast, however; after this point, $ACAI_{SA}$ is slightly superior to the $ACAI_{MA}$.

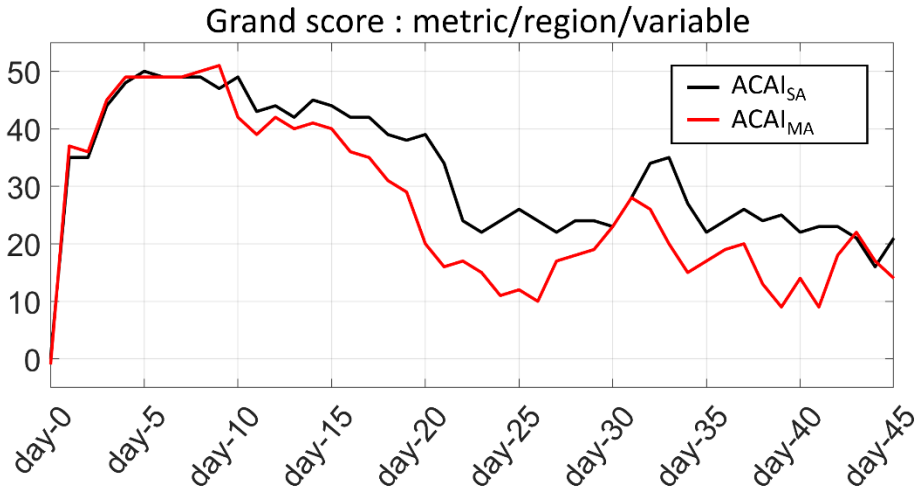

Figure 7 — Summation across all scorecard combinations of variable (W250, Z500, T850, W10m, T2m), region (NE, T2, SE) and metric (Bias, VARR, RMSE. CRPS) as a function of forecast lead-time for the $ACAI_{SA}$ and $ACAI_{MA}$ experiments. Scores are computed again ECMWF analyses, and positive (negative) values indicate improved (degraded) performance relative to the ER_{ctrl} experiment.

While the $ACAI_{MA}$ is not competitive with the $ACAI_{SA}$ experiment beyond the week-2, there are localized benefits from using a moving archive of analysis corrections. Figure 8 illustrates that the $ACAI_{MA}$ experiment drastically reduces the degradation of CRPS in 2m air temperature in the tropical east Pacific which is primarily responsible for the degradation of TR T2m shown in Figure 5d. However, as can be seen the $ACAI_{MA}$ experiment is not able to improve upon the performance of the baseline system.

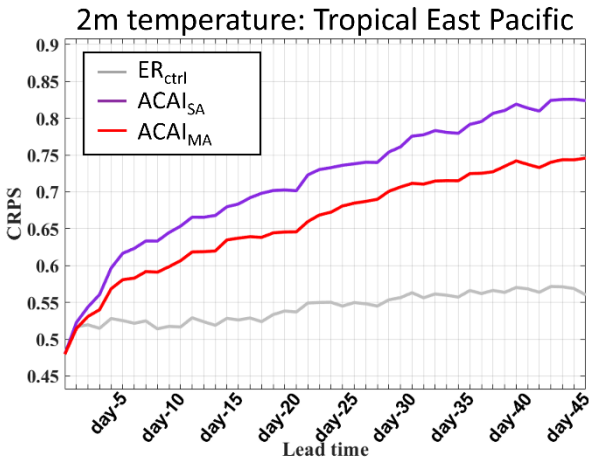

Figure 8 — Comparison of CRPS as a function of forecast lead-time for tropical east Pacific 2m air temperature for the ER_{ctrl}, $ACAI_{SA}$ and $ACAI_{MA}$ experiments.

4.2.2 Integrated vapor transport (IVT)

In addition to the impact of ACAI on the usual forecast skill metrics presented in the previous section, we have also looked at the impact on more unconventional metrics included integrated vapor transport (IVT). IVT is often used as a measure by which to define atmospheric rivers (ARs), which are major source of extra-tropical water vapor transport and precipitation. Figure 9a and 9b show maps of IVT at day-14 in summer-time forecast (JJA) for the baseline ER_{ctrl} and the $ACAI_{SA}$ experiment respectively with large regional difference apparent between the two. Figure 9c shows the difference in the absolute value of the bias in the two experiments, and while it is apparent there are comes localized increases to the bias by ACAI (shown by red colors in Fig. 9c) ACAI is able to reduce many of the larger magnitude biases. The regions of largest impact are the central/western Pacific and central Atlantic/Africa. While Figure 9 focuses on the reduction of bias at day-14, ACAI has also been found to reduce IVT biases throughout the entire 45-day forecast.

Lastly, because IVT is dependent on the state of the winds as well as moisture, and because ACAI will add perturbations to both of these fields, additional research was focused on understanding which component was most responsible of the improvement to the forecasts of IVT. It was found that while ACAI is improving the moisture content of the atmosphere, the reduction in IVT biases is dominated by improvements to the wind speed.

The results presented here are part of a forthcoming publication lead by Dr. Carolyn Reynolds of the Marine Meteorology Division at NRL Monterey [20].

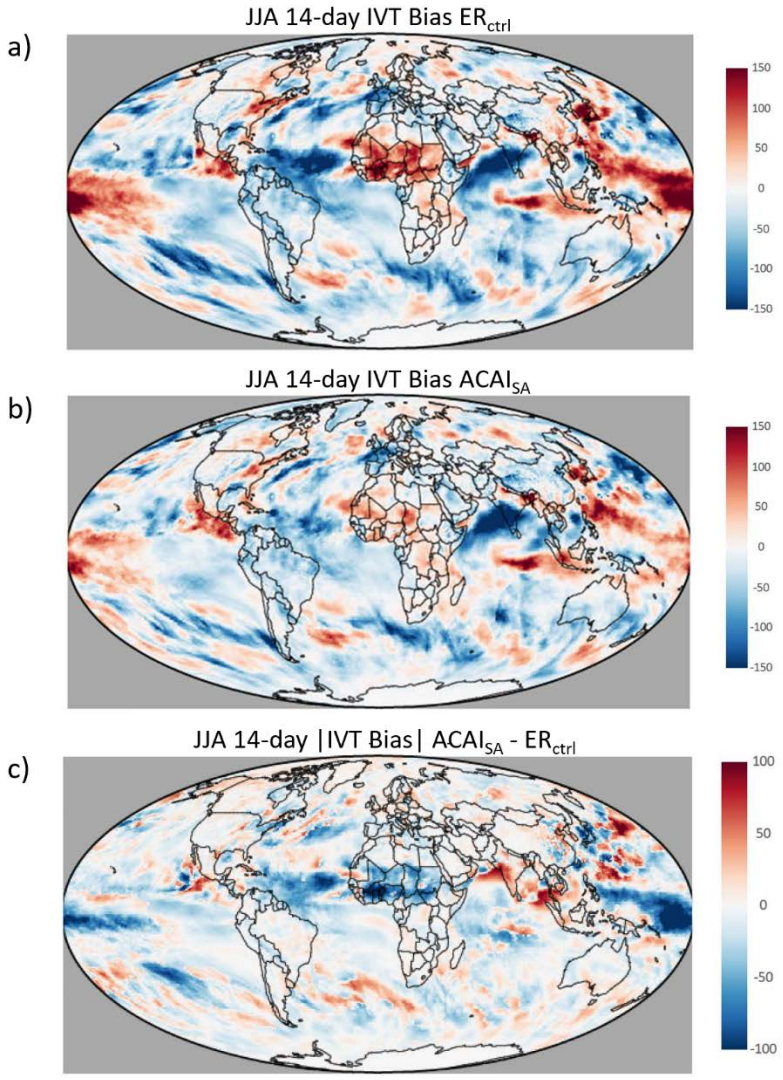

Figure 9 — JJA IVT bias for (a) ER_{ctrl}, (b) $ACAI_{SA}$ and (c) absolute value of the IVT bias for $ACAI_{SA}$ minus the absolute value of the IVT bias for ER_{ctrl} for forecast day 14 (kg m^{-1} s^{-1}). Reproduced from [20].

4.2.3 Madden-Julian Oscillation (MJO)

Lastly, we have also evaluated the impacts on ACAI on variables important for forecasting the Madden-Julian Oscillation (MJO), an area of high priority for the Navy given its connectivity with global weather patterns, particular the development of tropical cyclones. Figure 10b shows differences in mean absolute bias, anomaly correlation and RMSE for four such variables evaluated over several of the environmental domains shown in Figure 10a and overall the results are very positive. The $ACAI_{SA}$ experiment is shown to reduce biases in OLR (except in the northern Indian Ocean), wind shear magnitude and 850hPa vorticity. However, 700hPa relative humidity biases are shown to be increased by ACAI, particularly in the eastern Pacific. As shown in Figure 8, by using a moving archive of analysis corrections, degraded performance in eastern tropical 2m air temperatures CRPS scores can be somewhat alleviated, and a similar impact of the $ACAI_{MA}$ experiment on 700hPa relative humidity might also be expected, though this analysis has not yet been completed.

The anomaly correlation is improved (increased) for all variables in weeks 1-3 with some degradation shown beyond this point. It should be noted that the signal in anomaly correlations in the

later weeks is quite noisy and the degradations are typically shown as not statistically significant. The values of RMSE are generally improved (reduced) with some isolated increases except for 700hPa relative humidity in the eastern Pacific where there is a persistent increase beyond week-3. Again, we expect that using a moving archive of analysis corrections (as in the $ACAI_{MA}$ experiment) will help reduce the degradation to RMSE brought about by the ACAI perturbations.

The results presented here are part of a forthcoming publication lead by Dr. Matthew Janiga of the Marine Meteorology Division at NRL Monterey which will also aim to include the impacts of the $ACAI_{MA}$ experiment.

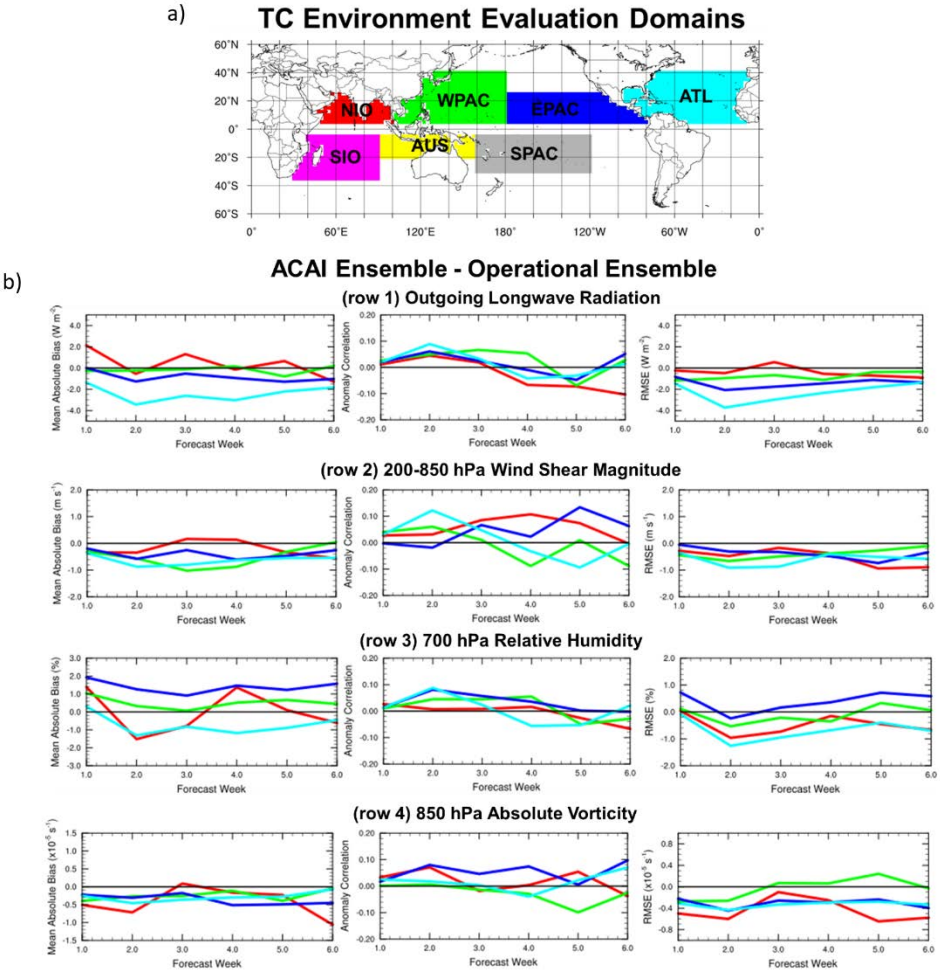

Figure 10 — (a) Large scale environmental domains used to compute metrics relevant for tropical cyclones. (b) Difference in mean absolute bias (left), anomaly correlation (middle), and RMSE (right) between the $ACAI_{SA}$ and ER_{ctrl} experiments for forecasts initialized in JJASON 2017 for (row 1) OLR, (row 2) 200-850 hPa wind shear magnitude, (row 3) 700 hPa relative humidity, and (row 4) 850 hPa absolute vorticity. Line colors correspond to regions shown in (a).

4.3 SKEB

We also tested the Stochastic Kinetic Energy Backscatter (SKEB) method as an additional representation of stochastic error in the Navy ESPC model. Figure 11 shows the variance ratio of the SKEB-based experiment compared to the baseline ER_{ctrl} experiment. As expected, SKEB is able to improve upon the spread-skill of the baseline system. In the case of 500hPa geopotential height, SKEB is able to make large improvements within the first few days of the experiment and remains higher than the

baseline system throughout the 30-day model forecast. For comparison, the $ACAI_{SA}$ experiment is also shown in Figure 11. In both panels, SKEB is shown to outperform ACAI during week-1 of the forecast. The two methods show similar performance during week-2 of the forecast; however, ACAI outperforms SKEB throughout the rest of the forecast. As described in Section 2.3, the implementation of SKEB used here was first introduced in the stand-alone NAVGEM system and was tuned for short to medium-range forecasts (~14 days) and not the extended range scales used here (30 days). In which case, there may be additional gains to be had by re-tuning SKEB for our extended range applications. Furthermore, while SKEB only adds perturbations to the wind, ACAI perturbs temperature, humidity and surface pressure in addition to the wind, leading to larger perturbation to the system overall. This additional perturbation may explain why ACAI is able to outperform SKEB at longer lead-times.

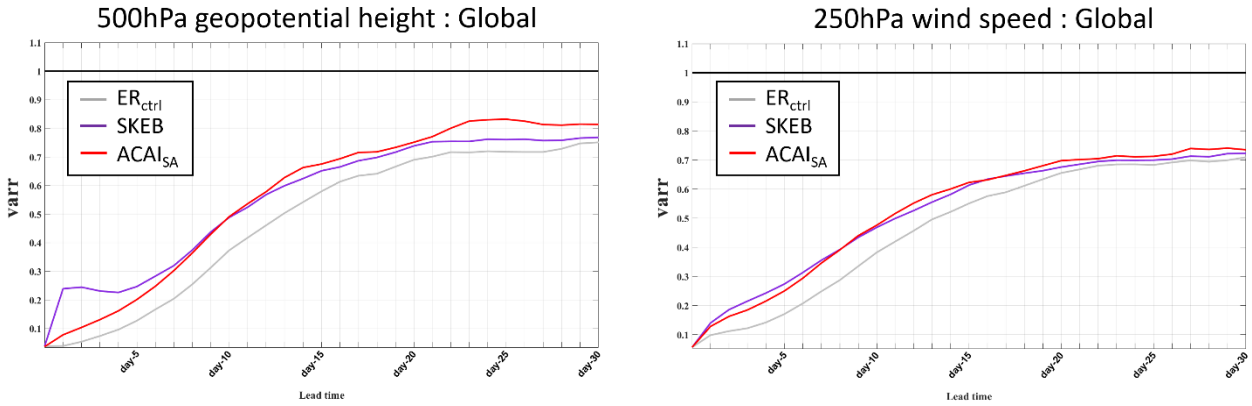

Figure 11 — Comparison of the globally averaged variance ratio (VARR) as a function of forecast lead-time for 850hPa temperature (left) and 10m wind speed (right) for the EDA_{ctrl} and RTPP/SKEB experiments.

While SKEB was originally implemented in Navy ESPC to increase ensemble divergence, we have also found that it can have pronounced impacts on the bias of the system. Figure 12 shows a reduction in tropical 10m wind speed biases for the entire 30-day forecast by including SKEB. These findings are consistent with prior research describing how additive noise of sufficient magnitude can be expected to modify the mean state when applied in non-linear systems [21]. Similarly, it has been found that the stochastic portion of the ACAI perturbations have a pronounced impact on correcting systematic forecast biases [15].

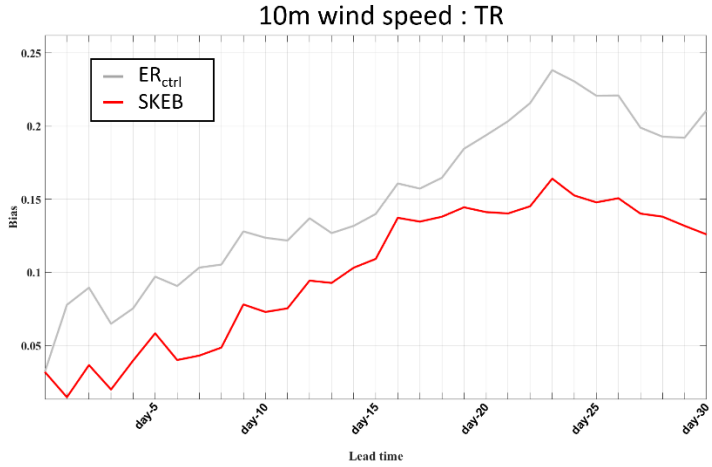

Figure 12 — Comparison of bias as a function of forecast lead-time for tropical 10m wind speed for the ER_{ctrl} and SKEB experiments.

5. CONCLUSIONS

The presented results illustrate several methods to improve the forecast skill of the Navy ESPC ensemble. By applying the RTPP method at the initial time of the ensemble forecast, the ensemble variance ratios are able to increase toward the ideal value of 1 (i.e. a perfect match between ensemble variance and ensemble mean squared error) as the system cycles (Figure 3). It is also clear that the chosen relaxation parameter (α=0.725) is sufficient to keep the root mean squared errors comparable to the baseline system (Fig. 4), which is an important finding given the sensitivity of the forecast skill to this parameter value.

During model integration, application of the ACAI-based perturbations is able to notably improve upon all of the forecast skill metrics with the largest impacts seen in the tropical region particularly in 500hPa geopotential height (Z500), surface level wind speed (W10m) and jet level wind speed (W250) (Figure 5). The northern extra-tropics are also generally improved out to 45-days; however, the response in the southern extra-tropics becomes neutral (and even somewhat degraded) in the second half of the forecast. This is particularly notable in ensemble spread-skill (VARR). We tested two implementation of ACAI where the perturbations are drawn from either a static archive of analysis corrections from an independent year ($ACAI_{SA}$) or from the 60-days prior to the initialization date of the forecast ($ACAI_{MA}$). In aggregate, the two methods are comparable in their improvement to baseline system in the first ~10 days; however, beyond this period the $ACAI_{SA}$ experiment outperforms $ACAI_{MA}$ (Figure 7). On the other hand it was found that the $ACAI_{MA}$ is able to improve upon some of the $ACAI_{SA}$ degradations. This is illustrated in Figure 8 which shows that the degraded values CRPS in 2m air temperature are less degraded relative to the baseline when using the moving archive of analysis corrections.

It was also shown that the $ACAI_{SA}$ experiment is able to improve upon forecasts of integrated vapor transport (Figure 9), which is important for its analogous representation of atmospheric rivers. While ACAI is able to improve upon both the moisture and wind components of the IVT, the most pronounce impact was brought about by correction to the wind fields. Additionally, we find that ACAI also improves several forecast metrics of variables important for forecasting the Madden-Julian Oscillation (Figure 10). Some degradation is seen in the bias and RMSE values of 700hPa relative humidity; however, it is anticipated that the $ACAI_{MA}$ may be able to alleviate some of this degradation which is the topic of ongoing analysis.

Lastly, we find that SKEB is able to improve the spread-skill of the baseline system (ER_{ctrl}) out to 30-days with the largest impacts seen in the first week of the forecast (Figure 11). Comparison against the $ACAI_{SA}$ experiment indicates the SKEB perturbations also provide greater improvement to the spread-skill in the first week than the ACAI-based perturbations and is competitive with ACAI through the 2nd week of the forecast. ACAI does outperform SKEB beyond the week-2 time frame; however, this may be due to the fact that ACAI perturbs additional variables and that the implementation of SKEB used here was tuned for short to medium-range forecasts. Interestingly we find that random perturbations introduced by SKEB can also modify the mean state in a way to reduce the systematic biases (Figure 12).

The presented results comparing the impact of ACAI and SKEB on the forecast skill of the baseline Navy ESPC system will be part of a forthcoming publication by Dr. William Crawford [22].

6. APPENDIX: SCORECARD METICS

Scorecard metrics are computed for three regions including; Northern Extratropics (NE; 20°N-90°N), Tropics (TR; 20°N-20°S) and Southern Extratropics (SE; 20°S-90°S). For each region, we focus on five variables at varying atmospheric heights. These include; 2-meter air temperature (T2m), 10-meter wind speed (V10m); 850hPa air temperature (T850), 500hPa atmospheric heights (Z500) and wind speed at 250hPa (V250).

6.1 Bias and RMSE

1-degree gridded fields are output from the NAVGEM control and experimental model runs, and bias and RMSE are computed at each 24-hour lead time against analyses from the European Center for Medium Range Weather Forecasting (ECMWF) provided by the TIGGE archive. Over a particular region, the bias and RMSE are computed as:

$$Bias = \frac{1}{N_p}\sum_{p=1}^{N_p} \emptyset_p(\bar{x}_p - y_p)$$

$$RMSE = \sqrt{\frac{1}{N_p}\sum_{p=1}^{N_p} \emptyset_p(\bar{x}_p - y_p)^2}$$

where N_p is the number of points in the region, and $\bar{x}_p$ and y_p are the ensemble mean and value of the ECMWF analysis at point p respectively. $\emptyset_p$ is a weighting applied according to the cosine of the latitude at point p. Bias and RMSE at each forecast lead time are averaged across all available forecasts.

6.2 Continuous Ranked Probability Score (CRPS)

The Continuous Ranked Probability Score (CRPS) is measure of the fit between the cumulative distribution functions (CDF) of the forecast and a verifying estimate of truth (here taken as the ECMWF analyses provided by the TIGGE archive). A closer fit between the forecasted CDF and the step function of the analysis CDF results in a lower value of CRPS with 0 indicating a perfect match. CRPS over a particular region is computed as the squared difference in the forecast and analysis CDFs weighted according to latitude and averaged across all available forecasts.

6.3 Ensemble Spread and Variance Ratio (VARR)

Over a particular region, the spread in the ensemble is computed as:

$$\sigma_e = \frac{1}{N_p}\sum_{p=1}^{N_p}\sqrt{\frac{1}{N-1}\sum_{m=1}^{N} \emptyset_p(x_p^m - \bar{x}_p)^2}$$

where N is the number of ensemble members, N_p is the number of points in the region, and x_p^m and $\bar{x}_p$ are the value of ensemble member m and the ensemble mean at point p respectively. $\emptyset_p$ is a weighting applied according to the cosine of the latitude at point p. The variance ratio (VARR) is then the ratio of

the squared ensemble spread (σ_e^2) to the bias corrected mean squared error. Spread and VARR at each forecast lead time are averaged across all available forecasts.

REFERENCES

1. Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., & Mitchell, H. L. (1996). A System Simulation Approach to Ensemble Prediction. *Monthly Weather Review*, *124*(6), 1225–1242. https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.

2. Barton, N., and Coauthors, 2020: The Navy's Earth System Prediction Capability. *Earth and Space Sciences*, accepted for publication.

3. Buizza, R., M. Leutbecher, L. Isaksen & J. Haseler, 2010: Combined use of EDA- and SV-based perturbations in the EPS. *ECMWF Newsletter*, No. 123.

4. Hogan, T., Liu, M., Ridout, J., Peng, M., Whitcomb, T., Ruston, B., … Chang, S. (2014). The Navy Global Environmental Model. *Oceanography*. https://doi.org/10.5670/oceanog.2014.73.

5. Metzger, E. J., and Coauthors, 2014: US Navy Operational Global Ocean and Arctic Ice Prediction Systems. Oceanography, 27, 32-43.

6. Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck, 2003: North Atlantic Simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J Phys Oceanogr*, **33,** 2504-2526.

7. Hunke, E. C., and W. Lipscomb, 2008: CICE: The Los Alamos sea ice model, documentation and software user's manual, version 4.0. *Technical Report*, LA-CC-06-012, Los Alamos National Laboratory, Los Alamos, NM.

8. Theurich, G., and Coauthors, 2016: The Earth System Prediction Suite: Toward a Coordinated U.S. Modeling Capability. *Bulletin of the American Meteorological Society*, **97,** 1229-1247.

9. Xu, L., Rosmond, T., & Daley, R. (2005). Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus, Series A: Dynamic Meteorology and Oceanography*. https://doi.org/10.1111/j.1600-0870.2005.00123.x.

10. Kuhl, D. D., Rosmond, T. E., Bishop, C. H., McLay, J., & Baker, N. L. (2013). Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Monthly Weather Review*. https://doi.org/10.1175/mwr-d-12-00182.1.

11. Cummings, J. A., and O. M. Smedstad, 2013: Variational Data Assimilation for the Global Ocean. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, S. K. Park, and L. Xu, Eds., Springer Berlin Heidelberg, 303-343.

12. Zhang, F., C. Snyder, & J. Sun, 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*. https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

13. Piccolo, C., Cullen, M. J. P., Tennant, W. J., & Semple, A. T. (2019). Comparison of different representations of model error in ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3348.

14. Bowler, N. E., Clayton, A. M., Jardak, M., Lee, E., Lorenc, A. C., Piccolo, C., … Swinbank, R. (2017). Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3004.

15. Crawford, W., S. Frolov, J. McLay, C. A. Reynolds, N. Barton, B. Ruston and C. H. Bishop, 2020: Using analysis corrections to address model error in atmospheric forecasts. *Mon. Wea. Rev*., **148**, 3729-3745.

16. Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. Quart. J. Roy. Meteor. Soc., 612 , 3079–3102.

17. Reynolds, C. A., McLay, J. G., Goerss, J. S., Serra, E. A., Hodyss, D., & Sampson, C. R. (2011). Impact of Resolution and Design on the U.S. Navy Global Ensemble Performance in the Tropics, Monthly Weather Review, **139**, 2145-2155.

18. Barton, N., M. Janiga, J. McLay, C. Reynolds, C. Rowley, P. Hogan, P. Thoppil, 2019: Earth System Prediction Capability (ESPC) Initial Operational Capability (IOC) Ensemble System. *NRL Memorandum Report.* 7532--19-9928.

19. Whitaker, J. S., & Hamill, T. M. (2012). Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*. https://doi.org/10.1175/mwr-d-11-00276.1.

20. Reynolds, C. A., W. Crawford, A. Huang, N. Barton, M. Janiga, J. McLay, M. Flatau, S. Frolov & C. Rowley (2021), Analysis of Integrated Vapor Transport Biases, *In preparation*.

21. Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Cámara, A., Christensen, H. M., … Yano, J. I. (2017). Stochastic parameterization toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-15-00268.1.

22. Crawford, W., J. McLay, C. A. Reynolds, S. Frolov, N. Barton, B. Ruston & S. Camp (2021), Impact of model error techniques on the forecast skill of an extended range global coupled ensemble system, *In preparation*.