Assimilation of Low-Peaking Satellite Observations Using the Coupled Interface Framework

SERGEY FROLOV,^a WILLIAM CAMPBELL, BENJAMIN RUSTON, AND CRAIG H. BISHOP^b

Naval Research Laboratory, Marine Meteorology Division, Monterey, California

DAVID KUHL

Remote Sensing Division, Naval Research Laboratory, Washington, D.C.

MARIA FLATAU AND JUSTIN MCLAY

Naval Research Laboratory, Marine Meteorology Division, Monterey, California

(Manuscript received 1 February 2019, in final form 12 November 2019)

ABSTRACT

Coupled data assimilation (DA) provides a consistent framework for assimilating satellite observations that are sensitive to several components of the Earth system. In this paper, we focus on low-peaking infrared satellite channels that are sensitive to the lower atmosphere and Earth surface temperature (EST) over both ocean and land. Our atmospheric hybrid-4DVAR system [the Navy Global Environmental Model (NAVGEM)] is extended to include the following: 1) variability in the sea surface temperature (both diurnal variability and climatological perturbations to the ensemble members), 2) the coupled Jacobians of the radiative transfer model for the infrared sensors, and 3) the coupled covariances between the EST and the atmosphere. Our coupling approach is found to improve forecast accuracy and to provide corrections to the EST that are in balance with the atmospheric analysis. The largest impact of the coupling is found on near-surface atmospheric temperature and humidity in the tropics, but the impact extends all the way to the stratosphere. The role of each coupling element on the performance of the global atmospheric circulation model is investigated. Inclusion of variability in the sea surface temperature has the strongest positive impact on the forecast quality. Additional inclusion of the coupled Jacobian and ensemble-based coupled covariances led to further improvements in scores and to modification of the corrections to the ocean boundary layer. Coupled DA had significant impact on latent and sensible heat fluxes over land, locations of western boundary currents, and along the ice edge.

1. Introduction

Development of coupled data assimilation (DA) methods that better initialize coupled Earth system models (ESMs) is essential to the improvement of coupled model forecasts. Better initial conditions improve short- to medium-range forecasts and better expose model biases thus enabling more rapid model improvement. One component of such a system is the assimilation of satellite observations sensitive to more than one component of the ESM, such as lowpeaking infrared and microwave channels that are simultaneously sensitive to the lower atmospheric temperature, moisture, and the surface. Traditionally, such radiances were not assimilated, or were assimilated with very large observation errors, because of the uncertainty of the Earth surface temperature (EST)¹ and emissivity in both microwave and infrared frequencies.

^a Current affiliation: Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado.

^bCurrent affiliation: University of Melbourne, Melbourne, Victoria, Australia.

Corresponding author: Sergey Frolov, sergey.frolov@noaa.gov

 $^{^1\,\}mbox{We}$ define EST as the union of land, water, and ice surface temperature.

^{© 2020} American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

These uncertainties are particularly large over land and ice, surfaces that can be highly inhomogeneous (English 2008) within the footprint of a satellite observation. Misspecification of Earth surface temperature and emissivity in the radiative transfer model can result in aliasing poor knowledge of the Earth surface into erroneous initial condition for the lower atmosphere.

Assimilating low-peaking channels can be further divided, based on the level of uncertainty in surface temperature and emissivity, into assimilation over the ocean, ice, and land surface. Over the ocean, the surface temperature is reasonably well constrained (with errors in the SST analysis used in this study on the order of 0.2–0.7 K), with even better knowledge of the surface emissivity (Geer 2019; Prigent et al. 2017). Over land, both the surface temperature and the surface emissivity are poorly known in the microwave and the infrared (Pavelin and Candy 2014; Karbou et al. 2005; English 2008). Over ice, developing reliable ice surface temperature retrievals has been challenging due to uncertainty of the infrared cloud clearing algorithm (Liu et al. 2010) and uncertainties in microwave ice emissivity due to the presence of subgridscale ice leads (Mathew et al. 2008).

In this article, we focus on extending assimilation capabilities for low-peaking infrared channels over the ocean surface, where most of the uncertainty can be attributed to uncertainty in the ocean skin temperature (the first few microns of the ocean surface). Following Akella et al. (2017), we used the model of McLay et al. (2012) that incorporates the impact of the diurnal warming. Our model is a modification of the one-dimensional heat transfer model for the ocean warm layer and skin sea surface temperature (SST) (Zeng and Beljaars 2005; Takaya et al. 2010). In addition to the diurnal heating, we account for uncertainties in mesoscale oceanic features by introducing climatological SST perturbations in the short-term ensemble forecast used by the hybrid DA. In the atmospheric DA system, we extend the control vector by adding the two-dimensional field of EST to the standard fields of atmospheric temperature, velocity, humidity, and surface pressure. We also utilize the Jacobian information from the radiative transfer model that partitions the satellite brightness temperatures into atmospheric and surface components.

However, unlike Akella et al. (2017) and Derber and Li (2018), we also specify coupled ensemble covariances between the atmospheric variables and the EST. These covariances enable us to compute the EST increments not only over the ocean (where we assimilate low-peaking satellite channels) but also over land and ice, where, through ensemble cross correlations, we estimate the EST that is in balance with the atmospheric correction generated by the assimilation of routine atmosphere-only observations. It should be noted, however, that while these corrections to the EST are in balance with the atmospheric observations, they might not verify well with independent measurements of the EST because of strong biases in the land and ice temperature models and because of large uncertainties in the surface emissivity.

The aforementioned coupled DA approach follows the spirit of the interface solver method developed in Frolov et al. (2016) that proposes to incrementally extend existing DA solvers by adding progressively more information from the interfaces between coupled fluids. This is in contrast to the ambition of an exhaustive solver implementation of the strongly coupled DA, where the cross covariance are prescribed and are used to estimate the entire state of the ESM—from the top of the atmosphere to the bottom of the ocean (Sluka et al. 2016; Penny et al. 2017).

2. Methods

a. Baseline system

This study used Navy Global Environmental Model (NAVGEM; Hogan et al. 2014)—a semi-Lagrangian/semiimplicit integration of the hydrostatic dynamical equations of the atmosphere, the first law of thermodynamics, and conservation of moisture and ozone. The resolution of the NAVGEM model used in this study was T425 triangular truncation (about 31 km horizontal resolution at the equator) and a 60 level hybrid sigmapressure coordinate (top at 0.04 hPa, or about 65 km).

The NAVGEM DA solver, based on the accelerated representer (AR) method (Xu et al. 2005; Rosmond and Xu 2006), was a strong-constraint hybrid-4DVAR system (Kuhl et al. 2013). The flow-dependent aspect of the hybrid covariance is computed from an 80-member low-resolution (T119L60) ensemble generated using the ensemble transform (ET) method (McLay et al. 2008). (See Fig. 1 for the detailed description of how the 4DVAR and the ET systems interact.) On average, approximately three million observations were assimilated during each 6-h assimilation cycle. (The complete list of satellite observation used in this paper is listed in Table A1.) In the next section we expand on how EST perturbations are introduced to obtain an ensemble of coupled forecasts that are used to generate flow-dependent error covariance in the hybrid-4DVAR DA system.

The ocean SST and the ice concentrations in the baseline system were provided by the Navy Coupled Ocean Data Assimilation system (NCODA; Cummings and Smedstad 2014). It should be noted that unlike in (Cummings and Smedstad 2014), no active ocean model was used to generate a prior ocean state; instead,

FEBRUARY 2020



FIG. 1. Block diagram of the developed system. Modifications to the baseline system are shown in blocks with orange border. New elements are shown in orange blocks. Blocks with no change are shown in blue. Legend gives preview of modifications, with full details presented in the text of the paper.

persistence was used to propagate SST 12 h into the future. The NCODA SST analysis is similar to other SST analyses (Reynolds et al. 2007), but with a more extensive set of satellite observations: see (Cummings and Smedstad 2014) for the complete list. We also want to further clarify that this work is distinct from the work of Cummings and Peak (2014) that also attempted to account for assimilation of sea surface temperature radiances.

b. Modifications to the baseline system

1) OVERVIEW

An overview of the enhanced system is presented in Fig. 1. Modifications to the baseline system are shown in blocks with orange border, and they include the following: addition of the diurnal SST model to NAVGEM, inclusion of the climatological SST perturbations in the ensemble, and modifications to the hybrid-4DVAR solver. The details of these modifications are described below.

2) SST PERTURBATIONS IN THE ENSEMBLE SYSTEM

To emulate the distribution of the mesoscale ocean features that one might expect from an ensemble of fully coupled models, we perturbed the SST in the low-resolution (T119) ensemble [see Fig. 1 and Kuhl et al. (2013) for a description of how the high-resolution control and the low-resolution ensemble interact]. These SST perturbations were generated by sampling

random perturbations from a 20-yr-long archive of SST anomalies computed from the ERA-Interim reanalysis (Dee et al. 2011). The anomalies were computed by differencing the 20-yr archive from a seasonal SST mean. The random anomalies where then selected within 7 days from the month-day of the experimental date (e.g., for each date there were 14 days \times 20 years = 280 possible samples to draw from). The random perturbations were then scaled to enforce the average standard deviation of 0.4K. We chose this magnitude as an optimistic value for the average RMS error of the NCODA SST analysis. After the experiments were conducted, the actual SST analysis errors were evaluated to be closer to 0.7 K (not shown). We do not expect that this slightly lower magnitude of the enforced SST spread would lead to significantly different conclusions from this paper. We decided to implement this simple scaling algorithm instead of the original coupled ET approach of McLay et al. (2012) because coupled ET has significantly more atmospheric grid points than SST and, hence, the rotation and scaling of the SST fields in the original approach was dominated by the number of the atmospheric grid points. As a result, the original coupled ET struggled to preserve the desired SST ensemble variance.

To simulate slowly evolving ocean perturbations, we enforced temporal autocorrelations between climatological perturbations using the following AR(1) model:

$$\operatorname{sst}_{\operatorname{red}}(k) = 0.75 \times \operatorname{sst}_{\operatorname{red}}(k-1) + \Delta \operatorname{sst}_{\operatorname{white}}(k), \quad (1)$$

where $\operatorname{sst}_{\operatorname{red}}(k)$ is the snapshot of the SST perturbations smoothed in-time with the AR(1) model for each 6-h DA cycle (k), 0.75 is the AR(1) coefficient that was selected to represent typical weather decorrelation time scales (~5 days), and $\Delta \operatorname{sst}_{\operatorname{white}}(k)$ is the random draw from the SST anomaly archive [i.e., $\Delta \operatorname{sst}_{\operatorname{white}}(k)$ are correlated in space but not in time]. Perturbations were applied only in the open ocean areas with ice concentrations less than 50%. It is important to note that the random perturbations $\operatorname{sst}_{\operatorname{red}}(k)$ were generated before we scaled the magnitude of the SST perturbations to enforce the average global magnitude of the SST error (see Fig. 1). Hence, we didn't need to worry that the red noise process will generate perturbations with standard deviations different than the magnitude of the sst_{white}(k).

3) DIURNAL SURFACE MODEL

To add diurnal variation to our SST analysis, we used the diurnal SST model of (McLay et al. 2012). The diurnal SST model was applied to both the high-resolution forecast and to each individual ensemble member used in the computation of the hybrid error covariance. Our diurnal SST model (McLay et al. 2012) was based on Takaya et al. (2010), which represents the following surface layer processes: shortwave and longwave radiative flux, evaporation, molecular thermal conduction, and wind-driven turbulent diffusion determined through Monin-Obhukov similarity theory. Unlike the original model of Takaya et al. (2010), our implementation did not represent impact of Langmuir circulation directly. Instead the friction velocity was multiplied by a factor of 1.4 to obtain a slight enhancement of surface stress. Unlike in (McLay et al. 2012), the cool skin correction was active in our experiments following the original publication of Zeng and Beljaars (2005). The diurnal SST model was active between 60°N and 60°S, where we expect the magnitude of the diurnal signal to be the greatest because of the favorable sun inclination angle.

It is important to note that in the original paper, McLay et al. (2012) tested the impact of the diurnal SST model in the context of the extended-range ensemble prediction system, where an ensemble of 10-day forecasts was initialized using the ET technique centered on an external control analysis. In McLay et al. (2012), the diurnal SST model did *not* impact the centering analysis. In contrast, this paper exercises the diurnal SST model in a deterministic, cycling forecast system, so that diurnal SST *will* impact the first guess of the high-resolution (T425) forecast, and will improve the near-surface spread of the low-resolution (T119) ensemble system.

4) CHANGES TO THE HYBRID-4DVAR SYSTEM

Three changes were introduced to the atmosphereonly hybrid-4DVAR equation:

$$\mathbf{x}^{a} = \mathbf{x}^{f} + \mathbf{P}_{0}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{M}\mathbf{P}_{0}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}[y - \mathbf{H}(\mathbf{x}^{f})], \quad (2)$$

where \mathbf{x}^{f} and \mathbf{x}^{a} are the model forecast and analysis states, y are the observations, $\mathbf{H}(\mathbf{x}^{f})$ is the first guess (observation of the nonlinear forecast performed using a nonlinear observation operator), \mathbf{R} is the (potentially correlated) observation error covariance, \mathbf{P}_{0} is the initial time hybrid covariance, \mathbf{M} and \mathbf{M}^{T} are the tangent linear and the adjoint of the forecast model, and \mathbf{H} and \mathbf{H}^{T} are the tangent linear and the adjoint of the nonlinear observational operator. Following common terminology, we also refer to the linear operator \mathbf{H} as the Jacobian operator.

The first change to Eq. (2) was to extend the analysis vector \mathbf{x}^{a} to include \mathbf{x}^{EST} —the Earth surface temperature state:

$$\mathbf{x}^{\text{coupled}} = \begin{bmatrix} \mathbf{x}^{\text{atm}} \\ \mathbf{x}^{\text{EST}} \end{bmatrix}.$$
 (3)

We should clarify that the analysis vector in our reference system (the CONTROL experiment) did not contain \mathbf{x}^{EST} as a sink variable. However, the nonlinear operator $\mathbf{H}(\mathbf{x}^f)$ always depended on the EST values in all of our experiments, where the background values of the EST were provided to the nonlinear radiative transfer model as a parameter. When diurnal model was disabled, the EST over ocean was equivalent to the external SST analysis from NCODA as described in section 2a. When diurnal model was enabled, the EST over ocean was equivalent to the NCODA analysis with the addition of the simulated diurnal signal.

The second change to Eq. (2) was to include coupling through the observation operator between infrared radiances and the ocean surface:

$$y^{\text{radiance}} = \mathbf{H}^{\text{crtm}-\text{jacobian}} \mathbf{x}^{\text{coupled}} = [\mathbf{J}^{\text{atm}} \mathbf{J}^{\text{sst}}] \begin{bmatrix} \mathbf{x}^{\text{atm}} \\ \mathbf{x}^{\text{sst}} \end{bmatrix}, \quad (4)$$

where J^{atm} and J^{sst} are the Jacobians of the radiative transfer model provided by the Community Radiative Transfer Model v 2.2.3 (CRTM; JCSDA 2018). In the baseline system (CONTROL experiment in Table 1), \mathbf{x}^{sst} was not included in the state vector, which was mathematically equivalent to setting J^{sst} to zero. The observation operator in Eq. (4) was only coupled for low-peaking infrared satellite channels which were assimilated over the surface of the ocean (\mathbf{x}^{sst}).

The last change was to include coupled initial time covariance:

$$\mathbf{P}^{\text{coupled}} = \begin{bmatrix} \mathbf{P}^{\text{AA}} & \mathbf{P}^{\text{AE}} \\ \mathbf{P}^{\text{EA}} & \mathbf{P}^{\text{EE}} \end{bmatrix} \approx \alpha^{\text{static}} \begin{bmatrix} \mathbf{P}^{\text{AA}} & 0 \\ 0 & \sigma^2_{\text{EST}} \mathbf{I} \end{bmatrix} + \alpha^{\text{ens}} \begin{bmatrix} \mathbf{P}^{\text{AA}}_{\text{ens}} & \mathbf{P}^{\text{AE}}_{\text{ens}} \\ \mathbf{P}^{\text{EA}}_{\text{ens}} & \mathbf{P}^{\text{EE}}_{\text{ens}} \end{bmatrix},$$
(5)

where superscripts AA, EE, and AE denote cross covariances between atmosphere-atmosphere, Earth surface-Earth surface, and atmosphere-Earth surface states, respectively. The subscript static corresponds to the static (climatological) covariance and ens to the localized ensemble covariance. We used the hybrid coefficient of $\alpha^{\text{static}} = \alpha^{\text{ens}} = 0.5$. We did not specify cross covariances for the static and we used a trivial identity matrix (I) to specify covariance of errors in the Earth surface. Our weak justification for the last assumption was based on the very coarse resolution of the atmospheric ensemble (approximately 100 km). We assumed that at such long scales, horizontal errors in the Earth surface temperature can be assumed largely decorrelated (e.g., our stand-alone SST analysis system assumes that decorrelation scales exceed 100 km only in

Diurnal Diurnal Experiment SST in SST in Coupled Coupled name high-res ensemble Jacobian covariance CONTROL Y DSST_HR_ONLY Y DSST Y CJAC Y Y Y CPB0 Y Y Y COUPLED_DA Y Y Y Y

TABLE 1. Configuration of the experiments.

the vicinity of the equator). We also expect that at these scales, the vertical processes will dominate, as evidenced by robust vertical cross correlations (\mathbf{P}_{ens}^{AE}) (Fig. 2). We also assumed a conservative EST forecast error σ_{EST} of 0.4 K, which was possibly too low over land and ice.

To localize the ensemble correlation between the EST and the atmosphere, we used the same vertical and horizontal correlation matrix as used in the atmospheric hybrid-4DVAR to localize the information between the lowest level of the atmosphere and the rest of the vertical column. Following Fig. 1 from Kuhl et al. (2013), the ensemble correlations between the EST and the atmospheric variables were localized to zero at the level of 500 hPa and were localized to 0.5 at approximately 800 hPa. Horizontal localization scales where about 2000 km.

We approximated the coupled tangent linear model using

$$\mathbf{M}^{\text{coupled}} = \begin{bmatrix} \mathbf{M}^{\text{AA}} & \mathbf{M}^{\text{AE}} \\ \mathbf{M}^{\text{EA}} & \mathbf{M}^{\text{EE}} \end{bmatrix} \approx \begin{bmatrix} \mathbf{M}^{\text{AA}} & 0 \\ 0 & \mathbf{I} \end{bmatrix}.$$
(6)

That is, we assumed persistence as the tangent linear model for the evolution of the diurnal SST and the land surface model. [We may be able to remove this approximation in future work by specifying the tangent linear model for the surface processes and cross terms using the ensemble tangent linear models of Bishop et al. (2017) and Frolov and Bishop (2016).]

c. Experimental strategy

To evaluate the impact of the changes introduced in this paper, we performed a series of 2-month-long (1 June–1 August 2016) cycling DA experiments (CONTROL, DSST_HR_ONLY, DSST, CJAC, CPB0, and COUPLED_DA) summarized in Table 1. The series of experiments was designed to add progressively increasing level of coupling in the designed system. We started by introducing the diurnal SST model in the high-resolution forecast (DSST_HR_ONLY) and the ensemble (DSST), followed by adding coupling in the 4DVAR solver (CJAC and CPB0), and culminated with the run that combine all previous changes (COUPLED_DA).



FIG. 2. Statistics of the Earth surface and the lower atmosphere temperatures in run DSST. (a) Average standard deviation of EST, (b) average standard deviation of 2-m atmospheric temperature, and (c) average correlations between EST and 2-m temperature. Standard deviations and correlations were first computed from 80 members of the T119 ensemble for each analysis window and then averaged over all analysis windows from 1 Jun to 1 Aug 2016.

In addition to the cycling run with the DA cadence of 6 hours, 5-day forecasts were performed daily at 0000 and 1200 UTC. We had 247 (total) 6-h deterministic forecasts that were used to compute background fit-to-observations statistics and 123 (total) 5-day deterministic forecasts that were compared against ECMWF analysis.

d. Performance metrics

1) SKILL EVALUATION AGAINST ECMWF ANALYSIS

To evaluate the skill of the modified system, we compared the 0–5-day forecast skill against European Centre for Medium-Range Forecasts (ECMWF) real-time analysis from the TIGGE archive.

Our primary metric was improvement in the root-mean-square error (RMSE) between the

CONTROL forecast and each of the experimental forecasts. We computed this metric $d(r, z, \tau)$ as

follows for a set of regions (r), forecast length (τ), and height (z):

$$d(r, z, \tau) = \frac{100}{N_t} \sum_{l} \left[\frac{\sqrt{\frac{1}{N_r} \sum_{\{i,j\} \in r} \{ [x_{\text{cntrl}}(i,j,z,t,\tau) - x_{\text{verif}}(i,j,z,t,\tau)]^2 \}} - \sqrt{\frac{1}{N_r} \sum_{\{i,j\} \in r} \{ [x_{\text{exp}}(i,j,z,t,\tau) - x_{\text{verif}}(i,j,z,t,\tau)]^2 \}}}{\sqrt{\frac{1}{N_r} \sum_{\{i,j\} \in r} \{ [x_{\text{cntrl}}(i,j,t,z) - x_{\text{verif}}(i,j,t,z)]^2 \}}} \right]},$$
(7)

where x (i, j, z, t, τ) is the field of interest (temperature, vector winds, precipitable water, or geopotential height); *i*, *j*, *z*, *t*, τ are indices of the longitude, latitude, vertical level, initial time of the forecast and the forecast lead time respectively; r are the verifying regions [NH $(20^{\circ}-80^{\circ}N)$, tropics $(20^{\circ}S-20^{\circ}N)$, SH $(20^{\circ}-80^{\circ}S)$]; N_r is the number of points in the verifying region; and N_t is the number of verifying forecasts. Note that $d(r, z, \tau)$ is positive whenever the forecast error measure $\sqrt{(1/N_r)\sum_{\{i,j\}\in r} \{[x_{\exp}(i,j,z,t,\tau) - x_{\operatorname{verif}}(i,j,z,t,\tau)]^2\}}$ associated with the experimental setup is smaller than that associated with the baseline control experiment. Negative values of $d(r, z, \tau)$ indicate that the experimental configuration has actually degraded forecast skill. Also note that $d(r, z, \tau)$ is a "percentage-improvement" type measure so that, for example, if the experimental configuration had zero error by exactly matching the verification then $d(r, z, \tau) = 100$. The averaged metric $d(r, z, \tau)$ is used in Fig. 3 with positive (negative) change colored in red (blue). All colored boxes in Fig. 3 are statistically significant at the 95% level. White boxes are approximately neutral and not significantly different.

2) BACKGROUND FIT TO SATELLITE OBSERVATIONS

In addition to traditional long-forecast error statistics described in section 2d(1), we also implemented background fit-to-observations statistics for the 6-h short forecasts used in the DA cycle (similar to observationminus-background statistics in Akella et al. 2017, see their Fig. 12). It has been shown that improvements in the background fit-to-observation statistics track improvements in the long-forecast skill well (A. Geer 2017, personal communication). In addition, it only requires about two weeks of cycling to establish if a proposed system change has a positive, negative, or neutral impact using the background fit-to-observation statistics. In contrast, it takes two to four month of long forecasts to establish similar improvement or degradation with the traditional long forecast metrics. Finally, unlike the external analysis, background fit-to-observation statistics compares the forecast against observations and, hence, is not sensitive to biases present in the external analysis.

In this paper, we evaluated the background fit-toobservations statistics only for radiance channels used in the assimilation (see Table A1 for the complete list of channels and categories used for verification). Because of the uncertainty in the radiance bias correction, we only computed statistics for the standard deviation of the innovations (i.e., our statistics is insensitive to changes in the bias of the innovations). Details of our background fit-to-observation calculations are presented in appendix B.

3. Results

a. Correlations between atmospheric temperatures and the EST

Figure 2 examines patterns of variability and correlation between the EST and the 2-m atmospheric temperature. Figure 2a shows that the largest spread of the EST was over land (about 1 K on average); in particular, the largest spread was over deserts (about 1.5 K) and the lowest spread in the Amazon basin (0.3 K). This spread over the land was driven by the fact that each low-resolution ensemble member was coupled to its own land surface model; hence, ensemble perturbations in the atmospheric temperatures and precipitation were resulting in correlated land surface temperature perturbations. Over the oceans, the highest spread was in the summer midlatitudes (about 0.5–0.7 K), in the eastern tropical Pacific (about 0.5 K), and along the boundary currents (about 0.6–1.0K), such as the Antarctic Circumpolar Current (ACC) and the Gulf Stream. The spread patterns of the 2-m temperature followed the EST spread closely, with an exception of winter midlatitudes (SH for our study period where the spread in the 2-m temperature was approximately equal to the spread in the summer hemisphere) and the Antarctic ice edge (about 2K). Correlation





FIG. 3. Improved (degraded) RMSE skill score [Eq. (7)] for CONTROL vs COUPLED_DA run in red (blue) compared against the ECMWF analysis. Colored boxes indicate change at 95% significance level. White boxes are approximately neutral and not significantly different. Tropics are defined from 20°S to 20°N and the Northern and Southern Hemispheres between 20° and 80° latitude. The availability of vertical levels for verification was determined by the availability of verifying data on the TIGGE archive.

between 2-m temperature and EST follows previously documented patterns in the CERA system (Laloyaux et al. 2018), where the higher correlations are associated with the regions of shallower mixed-layer depth, such as tropical east Pacific and summer midlatitudes.

It is remarkable that the system described in this work can generate correlation and SST variance patterns similar to the fully coupled forecast model used in the CERA system (Laloyaux et al. 2018) but, in our case, without a dynamically resolved ocean mixed layer depth. We attribute this to our use of time-correlated climatological SST perturbations, which introduce mesoscale variability in the ensemble of SST analysis. Additionally, while our diurnal model did not have a resolved ocean mixed layer depth, it did respond to atmospheric processes that correlated with shallowing of the mixed layer, such as lower winds and higher solar insulation.

We should note, however, that while the patterns of correlations in Fig. 2 resemble correlation patterns in a coarse (1) CERA system, they diverge from the early results that we obtained in the high-resolution coupled ensemble ($1/3^{\circ}$ atmosphere and $1/12^{\circ}$ ocean).

	Variable	Vertical	Spectrum	# ch.	DSST_H	R_ONLY	DS	ST	CJ	۹C	CPB	0	COUPLE	D_DA
					Score	WLT	Score	WLT	Score	WLT	Score	WLT	Score	WLT
1	all	all	all	474	0.41	т	0.47	w	0.48	w	0.50	w	0.51	w
2	all	all	IR	348	0.37	Т	0.43	Т	0.43	W	0.47	W	0.48	W
3	all	all	MW	126	0.52	W	0.60	W	0.60	W	0.60	W	0.60	W
4	Other	all	MW	12	1.00	W	1.00	W	1.00	W	1.00	W	1.00	W
5	Ozone	all	IR	3	0.67	W	0.67	W	0.67	W	0.67	W	0.67	W
6	Surface	all	IR	65	0.48	W	0.34	Т	0.32	W	0.37	W	0.45	W
7	Surface	all	MW	6	1.00	W	1.00	W	1.00	W	1.00	W	1.00	W
8	Temperature	strat	IR	127	0.43	Т	0.50	W	0.57	W	0.64	W	0.56	W
9	Temperature	strat	MW	32	-0.25	Т	0.16	Т	0.00	Т	-0.16	Т	-0.13	Т
10	Temperature	trop	IR	118	0.18	Т	0.25	Т	0.20	Т	0.24	Т	0.31	Т
11	Temperature	trop	MW	47	0.68	W	0.66	W	0.68	W	0.72	W	0.70	W
12	Water vapor	trop	IR	51	0.73	W	0.80	W	0.84	W	0.82	W	0.88	W
13	Water vapor	trop	MW	29	0.79	W	0.76	W	0.90	W	0.97	W	0.97	W

FIG. 4. Background fit-to-observations scores for experiments in Table 1. Scores are computed for assimilated channels only. Number of channels corresponds to a unique combination of platform, sensor, and channel identification and will not necessary replicate number of channels in Table A1. W, L, and T stand for "win," "loss," and "tie" as specified in appendix B. Highlighted in red is the best score in this category.

Our preliminary analysis shows that ocean and atmosphere have significant coupling (correlation) between near-surface atmospheric and ocean temperatures in the regions of strong ocean fronts such as the Gulf Stream and the ACC [see e.g., slide 39 in Frolov et al. (2018a)]. In contrast, Fig. 2 shows depressed correlations in these frontal regions. This last result suggests that an eddy-resolving ocean model is required to properly capture atmosphere–ocean exchanges along the mesoscale ocean fronts.

b. Impact of coupled DA on the forecast

1) LONG FORECAST ERROR DIAGNOSTICS

Comparisons of the control forecast (CONTROL) and the forecast initialized using coupled DA (COUPLED_DA) against ECMWF analysis showed that COUPLED_DA forecasts were significantly better than CONTROL forecasts for a wide range of metrics and forecast lead times (Fig. 3). The RMSE scores improved the most (greater than 3%) for the tropical geopotential above 700 hPa and for the boundary layer and 200 hPa temperatures in the tropics. Modest improvements (1%– 2%) were also detected for humidity in the tropics and the SH. The impact of coupled DA was not localized to the lower atmosphere and extends throughout the atmospheric column (see e.g., temperature scores up to day 3 that were improved throughout the troposphere). Closer investigation (not shown) showed that the positive impact was also apparent in the bias score metrics.

We hypothesize that improved tropical temperatures near 200 hPa were due to improved representation of tropical convection, which connects surface perturbations with the tropopause. This hypothesis is supported



FIG. 5. Average increments for the lowest atmospheric level for the CONTROL experiment: (left) temperature increment and (right) specific humidity.



FIG. 6. The 6-h temperature forecast bias averaged from 15 Jun to 1 Aug 2016. (a) First guess bias, (b) analysis bias, (c) ratio of the experiment bias to control bias (forecast), and (d) ratio of the experiment bias to control bias (analysis). Lower (higher) than 1 magnitudes in (c) and (d) indicate improvements (degradation) in the bias of one of the experiments compared to the control.

by McLay et al. (2012) where introduction of the diurnal SST submodel enhanced deep convection along the intertropical convergence zone.

Some degradation of the forecast scores was also apparent (Fig. 3). Specifically, the tropical geopotential and temperature scores (between 500 and 200 hPa) were degraded beyond day 3 of forecast. This degradation is likely due to known biases in NAVGEM tendency terms. NAVGEM is overactive and tends

to have a strong negative geopotential tendency (R. Langland 2019, personal communication). As a result, NAVGEM has been historically tuned to produce a biased analysis in favor of a more accurate 5-day forecast score. In other words, an improved temperature analysis in this study (as shown by Fig. 3) will lead to degraded midtroposphere temperature scores at day 5. Such mixed impacts on forecast scores are common for almost any changes to the forecast system



FIG. 7. Difference between the average analysis increments for the CONTROL experiment and the (a),(b) DSST experiment; (c),(d) CJAC experiment; and (e),(f) COUPLED_DA experiment. (left) Lowest level of temperature and (right) lowest level of specific humidity.

and, in this case, they are clearly outweighed by the overwhelming positive impacts for other variables and lead times.

2) SHORT-FORECAST ERRORS AND ATTRIBUTION OF FORECAST IMPACT

Figure 4 summarizes background fit-to-observation statistics for a sequence of runs with progressively increasing level of coupling. The grand score across all channels (line 1) indicates that introduction of diurnal SST model in the high-resolution model was responsible for the majority of the positive impact; however, using our conservative metric, the introduction of the diurnal SST along was still marked as a "tie." Introduction of the diurnal SST had the biggest positive impact on water vapor and surface channels in both infrared (IR) and microwave (MW), which is consistent with the long-forecast score statistics in Fig. 3. Diurnal SST also improved the fit to tropospheric sensitive channels in the MW. Progressive addition of coupling characteristics [introduction of diurnal SST model in the ensemble (DSST), introduction of coupled Jacobians (CJAC), and introduction of coupled covariance (COUPLED_DA)] resulted in progressive improvements to the grand score (line 1, Fig. 4). Specifically, introduction of the diurnal SST model in the ensemble (cf. DSST_HR_ONLY and DSST) made statistically significant changes to temperature soundings in the IR and MW (lines 2 and 3) and, as a result, flipped the grand score in line 1 from a tie to a "win." The same change however, resulted in the degradation of the IR surface scores (line 6).

Introduction of the coupled Jacobian and coupled covariance further improved the grand score (cf. columns DSST against COUPLED_DA). The biggest improvements that can be attributed to coupled DA were in surface IR (line 6), tropospheric temperatures (lines 10 and 11), and in the water vapor channels (lines 12 and 13). Improvements in the surface IR channels



FIG. 8. Zonal averages of the average difference between temperature analysis increments for CONTROL and (a) DSST_HR_ONLY experiment, (b) DSST experiment, (c) CJAC experiment, and (d) COUPLED_DA experiment.

almost completely reversed degradation of scores due to introduction of diurnal SST model in the ensemble (cf. line 6 in columns DSST_HR_ONLY, DSST, and COUPLED_DA).

3) COMPARISON OF ANALYSIS INCREMENTS

To further understand the impact of coupled observation operators and coupled initial time covariances on the analysis, we compared average analysis fields from the CONTROL, DSST, CJAC, and COUPLED_DA experiments. On average, the analysis increment in the CONTROL experiment Fig. 5 was cooling the air above the ocean (indicating warm bias over the ocean) and warming the air above the continents (indicating cold bias over land). For specific humidity, the analysis was adding moisture over most of the ocean (dry bias) and removing moisture over land (wet bias). The pattern of warm bias in the troposphere (and the corresponding cool correction) is consistent with known NAVGEM biases as illustrated by the average first guess and analysis errors measured against radiosondes (Fig. 6).

Introduction of the diurnal SST further increased the strength of the average temperature correction that cooled the warm bias over the oceans (Fig. 7a). This correction extends through a large part of the troposphere (Fig. 8a). No coherent change in the humidity increment emerged from introduction of the diurnal SST model.

Introduction of the coupled DA altered the structure of the average near-surface temperature correction over the oceans (Figs. 7e and 8d) by introducing a warm correction in the boundary layer while slightly strengthening the cold correction aloft. These near-surface changes can be attributed to the impact of both coupled Jacobians and coupled covariances (Figs. 8c,d).

In COUPLED_DA, the largest EST increments (in both the average magnitude and the standard deviation, Fig. 9) were over land (positive) and off the west coast of Americas (negative). This pattern of EST increments is consistent with the pattern of cold average corrections to the atmospheric temperatures in Figs. 5–8. We attribute these large signals to 1) systematic errors in our



FIG. 9. Statistics of EST increments for the COUPLED_DA experiment. (a) Mean EST increment and (b) standard deviation of the EST increment.

land model (the land was too cold) and 2) known warm bias in SST quality control procedure that erroneously rejects cold water pixels in the upwelling filaments in presence of low stratus clouds (C. Barron, December 2018, personal communication). The EST increments were close to zero over ice because very few atmospheric observations were available to constrain the lower atmosphere and, through cross covariances, the ice surface temperatures. EST increments were also close to zero over the Amazon, possibly because of the extensive tropical forest cover that isolated EST from the fluctuations in the tropospheric temperatures.

The variance of the EST increments was largest over land (Fig. 9b)-indicating a strong diurnal cycle in the land surface temperatures (see discussion below). Over the ocean, the variance of the corrections was largest in the boundary current regions (Kuroshio, Gulf Stream, and the Agulhas current retroflection region) and the west coast of the Americas. Large SST increments can be expected in the boundary current regions because of large SST gradients, difficulties with the cloud clearing algorithms, and advection of the ocean mesoscale features. Finally, examination of the average EST increments binned by the time of the day (Fig. 10) revealed strong diurnal patterns of the average EST increment. We found that the strongest corrections over land were in the afternoon, suggesting that our land surface was too cold during the sun-lit hours. In contrast, the bias over the ocean upwelling regions was positive and strongest during the night time, suggesting that the warm bias in the SST quality control procedure was greater at night. We also found that coastal upwelling regions were always too warm (cold increment on average), and the equatorial Pacific region was too warm at night (cold increment) and too cold during the daytime (warm increment).

c. Impact on heat fluxes

Surface fluxes play important role in the fidelity of the coupled model and, as we will see here, are very sensitive to changes introduced by coupled DA. Because turbulent fluxes are diagnostic quantities computed using the bulk formula, coupled DA changes these fluxes through changes to the ESTs and to the lower atmospheric temperature, humidity, and wind velocity. We choose not to evaluate estimated fluxes against external fluxes retrievals because of high uncertainty in the satelliteretrieved fluxes. Instead, we rely on traditional observations described in section 3b to evaluate the fidelity of our atmospheric model forecast. In this section our goal is to evaluate how coupled DA changes the statistics of the turbulent surface fluxes that provide flux of temperature and moisture into the lower atmosphere.

Figures 11c and 12c show that on average, the mean latent and sensible heat fluxes changed the most over the ocean (greater by as much as 5%). Average changes in the latent heat flux where greatest over the tropics indicating decreased evaporation in the COUPLED_DA, which is consistent with cooling of the SST in the COUPLED_DA run. Figures 11d and 12d show that the standard deviations of the latent and sensible heat fluxes changed the most over land (as high as 1/3 of the average value) indicating that estimated changes in the EST were having a significant impact on the daily variations in the heat and moisture fluxes over the land surface. Over the ocean, daily changes in the EST increments led to largest changes in the latent heat flux along the ITCZ and along the boundary currents. Daily changes in the sensible heat flux were greatest along the boundary currents and along the Antarctic ice edge.

d. Comparisons with buoy data

To further understand the structure of the coupled increments, we compared estimated SST increments



FIG. 10. Diurnal evolution of the average COUPLED_DA EST increment. Panels show increments centered at (a) 0000, (b) 0600, (c) 1200, and (d) 1800 UTC.

with two buoys in the California current system, where the average magnitude of the coupled increments was large. Figure 13 showed that the NCODA SST boundary condition had very little diurnal signal (the blue line in Fig. 13 is very smooth compared to observations). This smoothness could be expected from the definition of the SST produced by the NCODA analysis. In contrast, EST-corrected SST from experiment COUPLED_DA (red line in Fig. 13) had a significant diurnal cycle that was comparable to the magnitude of the diurnal cycle at the



FIG. 11. Latent heat fluxes (average over 24 h and over the duration of the experiments). (a) Average flux for the CONTROL experiment. (b) Standard deviation of the flux in the CONTROL experiment. (c) Average difference in flux between CONTROL and COUPLED_DA. (d) Standard deviation of the differences in flux between CONTROL and COUPLED_DA.



FIG. 12. As in Fig. 10, but for the sensible heat flux.

location of the buoy sensor. Frequently, the magnitude of the EST diurnal signal was even higher than the observed at the location of the buoy sensor (e.g., between 26 June and 9 July for buoy 46059). We attribute this to the decrease of the magnitude of the diurnal warming from the surface to the submerged location of the sensor (approximately one meter depth).

For the COUPLED_DA experiment, mean error statistics of the EST-corrected analysis showed significant improvement at the location of the in-shore buoy 46047, reducing mean error from 0.89 to 0.58 K. We attribute this to the fact that buoy 46047 is located in the area of the largest mean EST correction over the ocean. For the CJAC experiment, there was no significant change in mean error statistics.

4. Summary and conclusions

This paper demonstrated the first implementation of coupled covariance modeling in a hybrid-4DVAR system with an operational model. In addition to coupled covariance modeling, observation operators for low-peaking infrared channels also used the coupled Jacobians of the radiative transfer model, similar to (Akella et al. 2017; Derber and Li 2018). These developments were made possible by applying principles of the interface solver design outlined in Frolov et al. (2016). We showed that introduction of the EST in the atmospheric DA solver had positive impact on atmospheric forecast scores, including reduction of the geopotential height errors (up to 50 hPa) and humidity errors in the boundary layer.

Examination of analysis increments revealed patterns of model bias in our atmospheric model that was coupled to the land model and to the diurnal SST model:

- 1) Our land model had strong daytime biases (land surface was too cold during the daytime (Fig. 10).
- 2) Both Equatorial and coastal upwelling regions had strong SST biases (our SST analysis was too warm at night in the upwelling areas). We attribute them to the known warm bias in the quality control procedures for SST analysis in presence of strong upwelling filaments and low stratus clouds (upwelling waters have been aggressively screened as clouds). These bias issues have been since fixed in the operational system.
- 3) Differences in the heat flux forecasts, showed that coupled DA altered latent and sensible heat fluxes over both land and ocean, hence, changing supply of heat and moisture into the lower atmosphere.
- 4) Comparisons with buoy data suggested that EST corrections from the COUPLED_DA run introduced both the diurnal variation in the SST forcing and reduced the mean error in the SST.

We conclude by highlighting some limitations of our work that we plan to address in future research:

- We used uncorrelated climatological background error covariance for EST. Specifying correlated errors for EST will improve smoothness of the EST increments.
- Our software implementation did not allow us to tune localization functions between EST and the atmospheric variables. It would be beneficial to specify more complex cross-fluid localization



FIG. 13. (a),(b) Comparisons of the buoy SST observations (at 1 m depth) in black with analysis of the SST from NCODA analysis (blue) and NCODA analysis incremented by the coupled EST increment (red). Locations of the buoy observations are shown in (c) for National Data Buoy Center buoy 46059 [time series in (a)] and panel (d) for National Data Buoy Center buoy 46047 [time series in (b)]. Background colors in (c) and (d) are average EST increments from experiment COUPLED_DA (similar to Fig. 9a). Statistics of the mean error (ME) and the root-mean-square error (RMSE) are noted on top of (a) and (b).

functions, for example as used in Frolov et al. (2016) or Laloyaux et al. (2018).

- 3) We did not include surface emissivity as a part of the hybrid-4DVAR control vector. We plan to include surface emissivity in our future work to improve estimation of EST over land, ice, and snowcovered areas following (Karbou et al. 2005; Pavelin and Candy 2014; Mathew et al. 2008).
- 4) We plan to further extend the state vector to include soil moisture variables.
- 5) We used trivial TLM and ADJ of the dynamical model for the evolution of the EST and for dynamical coupling between EST and the atmosphere. It would be ideal to specify an actual TLM and ADJ model based on simplified physics (Storto et al. 2018) or based on the local ensemble tangent linear approach of (Allen et al. 2017; Bishop et al. 2017; Frolov and Bishop 2016; Frolov et al. 2018b).
- 6) We used climatological perturbations to the SST. We plan to use flow-dependent SST perturbations once

TABL	JE A1. List of assim	ilated channels and ch	annel groupings. If an ir	strument is deployed on more than one platform a union of channel numbers is presented.
Instrument	IR/MW	Category	No. of channels	Channel numbers
AIRS	Hyper-IR	T stratosphere	33	92, 93, 99, 104, 105, 110, 111, 116, 117, 123, 128, 129, 138, 139, 144, 145, 151, 156, 157, 162, 168, 169, 170, 172, 173, 174, 175, 177, 179, 180, 182, 185, 186
		T troposphere	30	LW: 190, 192, 193, 198, 201, 204, 207, 210, 213, 215, 216, 218, 239, 250, 251, 253, 318 SW: 1881, 1882, 1883, 1884, 1897, 1901, 1911, 1917, 1918, 1921, 1923, 1924, 1928
		Water vapor	10	1304, 1401, 1403, 1644, 1674, 1717, 1751, 1780, 1803, 1812
IASI	Hyper-IR	T stratosphere	24	122, 128, 135, 141, 148, 154, 161, 173, 185, 187, 193, 199, 205, 207, 210, 212, 214, 217, 219, 222, 224, 226, 226, 226, 220, 220
		T troposphere	35	LW: 236, 239, 246, 249, 252, 254, 260, 262, 265, 267, 269, 282, 306, 323, 329, 335, 347, 350, 354, 356,
				360, 366, 371, 373, 375, 379, 381, 404, 407, 410, 426, 428, 432, 434
		Water vapor	18	2889, 3098, 3168, 3248, 3281, 3309, 3442, 3444, 3448, 3450, 3491, 3506, 3575, 3577, 3582, 3589, 3653–3661
		Surface	10	515 545 557 566 571 573 646 662 668 756 867 021 1077 1046 1121 1133 1101 1104 1271
CrIS	Hvner-IR	T stratosphere	46	27.37.49.51.53.59.61.63.64.65.67.69.71.73.75.79.80.81.83.85.87.88.89.93.95.96.99.116
			2	124, 126, 132, 133, 136, 137, 138, 142, 143, 144, 145, 154, 155, 157, 158, 159, 160, 161
		T troposphere	86	LW: 208, 234, 236, 238, 239, 242, 246, 264, 266, 268, 279, 283, 285, 305, 311, 332, 342, 389
		4 4		SW: 988, 995, 998, 1000, 1003, 1008, 1009, 1010, 1014, 1017, 1018, 1020, 1022, 1024, 1026, 1029, 1032,
				1034, 1037, 1038, 1041, 1042, 1044, 1046, 1049, 1050, 1053, 1054, 1058, 1060, 1062, 1064, 1066, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056, 1056,
				1069, 10/6, 10//, 1080, 1086, 1091, 1095, 1101, 1109, 1112, 1121, 1128, 1133, 1163, 1172, 1187, 1180 1205 1211 1210 1231 1245 1271 1280 1200 1313 1216 1225 1220 1246 1247 1472
				1109, 1200, 1211, 1219, 1201, 1240, 12/1, 1209, 1200, 1310, 1320, 1329, 1329, 1347, 1473, 1473, 1473, 1474, 1471
		Water vapor	19	718, 720, 722, 728, 735, 742, 753, 762, 780, 874, 882, 890, 898, 906, 907, 908, 914, 937
		Surface	6	427, 441, 464, 473, 482, 490, 501, 519, 529
		Ozone	ę	707, 710, 713
AMSUA	MM	Water vapor	1	
		Other	1	2
		T troposphere	10	3, 4, 5, 6, 9, 5, 6, 7, 8, 9
		T stratosphere	5	10, 11, 12, 13, 14
ATMS	MW	Water vapor	5	18, 19, 20, 21, 24
		Other	1	2
		T troposphere	8	3, 4, 5, 6, 7, 8, 9, 10
		T stratosphere	5	11, 12, 13, 14, 15
SIMSS	MW	T troposphere	9	2, 3, 4, 5, 6, 7
		T stratosphere	4	21, 22, 23, 24
		Water vapor	4	9, 10, 11, 14
		Surface	2	12, 13
		Other	2	15, 16
GMI	MW	Water vapor	ς	5, 12, 13
AMSR2	MW	Water vapor	2	9,10
SHM	MW	Water vapor	3	3, 4, 5
SAPHIR	IR	Water vapor	9	1, 2, 3, 4, 5, 6
SEVIRI	IR	Varies	10	5, 6, 8, 9, 10, 11, 13, 14, 15, 16
Imager	IR	Varies	5	1, 2, 3, 4, 5

this system is implemented as a part of our fully coupled ensemble prediction system.

Acknowledgments. This work was supported by the U.S. Office of Naval Research through the Navy Earth Systems Prediction Capability Project (PE 0603207N). We are grateful for the access to the Department of Defense high performance computing resources that enabled us to conduct this research. We also thank three anonymous reviewers for the most thorough review of our manuscript.

APPENDIX A

List of Assimilated Satellite Radiances

Table A1 provides a summary of all assimilated satellite radiance channels used in the experiments described in this paper.

APPENDIX B

Fit to Observation Statistics

For each unique combination of satellite radiance channel, sensor, and platform (designated as ch#), over a period of a single DA window ($t:t + \Delta t_{DA}$), the fit-toobservation statistics $f_{t:t+\Delta t_{DA}}^{ch\#}$ was computed as follows:

$$f_{t:t+\Delta t_{\mathrm{DA}}}^{\mathrm{ch\#}} = \frac{\mathrm{std}\left[y_{t:t+\Delta t_{\mathrm{DA}}}^{\mathrm{ch\#}} - \mathbf{H}^{\mathrm{ch\#}}\left(\mathbf{x}_{t:t+\Delta t_{\mathrm{DA}}}^{\mathrm{exp}}\right)\right]}{\mathrm{std}\left[y_{t:t+\Delta t_{\mathrm{DA}}}^{\mathrm{ch\#}} - \mathbf{H}^{\mathrm{ch\#}}\left(\mathbf{x}_{t:t+\Delta t_{\mathrm{DA}}}^{\mathrm{control}}\right)\right]},\tag{B1}$$

where $y_{t:t+\Delta t_{DA}}^{ch\#}$ are the observations for a given radiance channel number and the assimilation window, and $\mathbf{H}^{ch\#}(\mathbf{x}_{t:t+\Delta t_{DA}}^{exp})$ are the matching forecasts of the radiance observation. If $f_{t:t+\Delta t_{DA}}^{ch\#}$ is less (greater) than 1.0, the experiment fits the observation better (worse) than the control.

To aggregate the statistics over the length of the experiment, we form a time series of $f_{t:t+\Delta t_{DA}}^{ch\#}$, and evaluate it with a paired difference *t* test on the log-transformed raw ratios of standard deviations. We accounted for serial correlation in time following Wilks (1995), his Eq. (5.13). The confidence intervals (we used 95% significance level) were then back transformed from the log-space to the normal space. If the confidence interval did not include 1.0, then the channel was counted as a "win" ("loss") if the mean $f_{t:t+\Delta t_{DA}}^{ch\#}$ was lower (higher) than 1.0. If the confidence interval included 1.0, then the channel was counted as a "tie."

To compare across runs, we computed the grand fitto-observations score F as following:

$$F = \frac{N_{\rm win} - N_{\rm loss}}{N_{\rm win} + N_{\rm loss} + N_{\rm tie}},\tag{B2}$$

where N_{win} , N_{loss} , and N_{tie} were the counts of wins, losses, and ties, respectively, over predefined groups of channels specified in Table A1. *F* in Eq. (B2) measure the percentage by which the number of statistically significant wins exceeds (or lags) the number of statistically significant losses. Given a common reference, background fit-to-observation score allows intercomparisons of multiple runs. If the *F* score is larger for one run it means that more radiance channels were significantly improved. However, the *F* score, similar to many other score cards used in numerical weather prediction, is insensitive to the absolute magnitude of the improvement. For example, two runs with perfect score of 1.0 might produce significantly different RMSE for the long forecast verification.

In addition to the grand fit-to-observations score F, we also assigned wins, ties, and loses (WLT score) to subgroups of channels as follows:

$$\begin{cases} if (N_{win} > N_{loss} + N_{tie}) \Rightarrow WLT = win \\ elseif (N_{loss} > N_{win} + N_{tie}) \Rightarrow WLT = loss. \\ else \Rightarrow WLT = tie \end{cases}$$
(B3)

It should be noted that the WLT score in Eq. (B3) above is a conservative estimate because it assumes that all tie scores will go against the better if sufficient data were available. A less conservative estimate would be to assume that ties are akin to random coin flips, then one can assume, within 1% margin, that 60% of ties are converted to losses and 40% are converted to wins.

REFERENCES

- Akella, S., R. Todling, and M. Suarez, 2017: Assimilation for skin SST in the NASA GEOS atmospheric data assimilation system. *Quart.* J. Roy. Meteor. Soc., 143, 1032–1046, https://doi.org/10.1002/qj.2988.
- Allen, D. R., C. H. Bishop, S. Frolov, K. W. Hoppel, D. D. Kuhl, and G. E. Nedoluha, 2017: Hybrid 4DVAR with a local ensemble tangent linear model: Application to the shallow-water model. *Mon. Wea. Rev.*, **145**, 97–116, https://doi.org/10.1175/ MWR-D-16-0184.1.
- Bishop, C. H., S. Frolov, D. R. Allen, D. D. Kuhl, and K. Hoppel, 2017: The local ensemble tangent linear model: An enabler for coupled model 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **143**, 1009– 1020, https://doi.org/10.1002/qj.2986.
- Cummings, J. A., and J. Peak, 2014: Validation test report for the variational assimilation of satellite sea surface temperature radiances. Naval Research Laboratory Tech. Rep. NRL/MR/7320– 14-9520, Naval Research Laboratory, 33 pp., https://apps.dtic.mil/dtic/ tr/fulltext/u2/a608635.pdf.
- —, and O. M. Smedstad, 2014: Ocean data impacts in global HYCOM. J. Atmos. Oceanic Technol., 31, 1771–1791, https:// doi.org/10.1175/JTECH-D-14-00011.1.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/ 10.1002/qj.828.

- Derber, J., and X. Li, 2018: Assimilating SST with an atmospheric DA system. ECMWF Workshop on Sea Surface Temperature and Sea Ice Analysis and Forecast, Reading, United Kingdom, ECMWF, 27 pp., https://www.ecmwf.int/node/17984.
- English, S. J., 2008: The importance of accurate skin temperature in assimilating radiances from satellite sounding instruments. *IEEE Trans. Geosci. Remote Sens.*, 46, 403–408, https:// doi.org/10.1109/TGRS.2007.902413.
- Frolov, S., and C. H. Bishop, 2016: Localized ensemble-based tangent linear models and their use in propagating hybrid error covariance models. *Mon. Wea. Rev.*, **144**, 1383–1405, https://doi.org/10.1175/MWR-D-15-0130.1.
- —, —, T. Holt, J. Cummings, and D. Kuhl, 2016: Facilitating strongly coupled ocean–atmosphere data assimilation with an interface solver. *Mon. Wea. Rev.*, **144**, 3–20, https://doi.org/ 10.1175/MWR-D-15-0041.1.
- —, and Coauthors, 2018a: Comparison of data assimilation coupling strategies for Earth system models. *ECMWF Annual Seminar*, Reading, United Kingdom, ECMWF, 48 pp., https:// www.ecmwf.int/sites/default/files/elibrary/2018/18545-comparisondata-assimilatin-coupling-strategies-earth-system-models.pdf.
- —, D. R. Allen, C. H. Bishop, R. Langland, K. W. Hoppel, and D. D. Kuhl, 2018b: First application of the local ensemble tangent linear model (LETLM) to a realistic model of the global atmosphere. *Mon. Wea. Rev.*, **146**, 2247–2270, https:// doi.org/10.1175/MWR-D-17-0315.1.
- Geer, A. J., 2019: Correlated observation error models for assimilating all-sky infrared radiances. *Atmos. Meas. Tech.*, **12**, 3629–3657, https://doi.org/10.5194/amt-12-3629-2019.
- Hogan, T. F., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, 27, 116–125, https://doi.org/10.5670/ oceanog.2014.73.
- JCSDA, 2018: Community Radiative Transfer Model (CRTM). JCSDA, accessed 26 December 2019, https://www.jcsda.org/ jcsda-project-community-radiative-transfer-model.
- Karbou, F., C. Prigent, L. Eymard, and J. R. Pardo, 2005: Microwave land emissivity calculations using AMSU measurements. *IEEE Trans. Geosci. Remote Sens.*, 43, 948–959, https://doi.org/10.1109/TGRS.2004.837503.
- Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Wea. Rev.*, **141**, 2740–2758, https://doi.org/10.1175/ MWR-D-12-00182.1.
- Laloyaux, P., S. Frolov, B. Menetrier, and M. Bonavita, 2018: Implicit and explicit cross-correlations in coupled data assimilation. *Quart. J. Roy. Meteor. Soc.*, 144, 1851–1863, https:// doi.org/10.1002/qj.3373.
- Liu, Y., S. A. Ackerman, B. C. Maddux, J. R. Key, and R. A. Frey, 2010: Errors in cloud detection over the arctic using a satellite imager and implications for observing feedback mechanisms. *J. Climate*, 23, 1894–1907, https://doi.org/10.1175/2009JCLI3386.1.
- Mathew, N., G. Heygster, C. Melsheimer, and L. Kaleschke, 2008: Surface emissivity of arctic sea ice at AMSU window frequencies.

IEEE Trans. Geosci. Remote Sens., **46**, 2298–2306, https://doi.org/10.1109/TGRS.2008.916630.

- McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. Mon. Wea. Rev., 136, 1093–1108, https://doi.org/10.1175/ 2007MWR2010.1.
- —, M. K. Flatau, C. A. Reynolds, J. Cummings, T. Hogan, and P. J. Flatau, 2012: Inclusion of sea-surface temperature variation in the U.S. Navy ensemble-transform global ensemble prediction system. J. Geophys. Res., 117, D19120, https:// doi.org/10.1029/2011JD016937.
- Pavelin, E. G., and B. Candy, 2014: Assimilation of surfacesensitive infrared radiances over land: Estimation of land surface temperature and emissivity. *Quart. J. Roy. Meteor. Soc.*, 140, 1198–1208, https://doi.org/10.1002/qj.2218.
- Penny, S. G., and Coauthors, 2017: Coupled data assimilation for integrated earth system analysis and prediction: Goals, challenges and recommendations. WWRP-2017-3, WMO/WWRP, 59 pp. https://www.wmo.int/pages/prog/arep/wwrp/new/documents/ Final_WWRP_2017_3_27_July.pdf.
- Prigent, C., F. Aires, D. Wang, S. Fox, and C. Harlow, 2017: Seasurface emissivity parametrization from microwaves to millimetre waves. *Quart. J. Roy. Meteor. Soc.*, 143, 596–605, https:// doi.org/10.1002/qj.2953.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. J. Climate, 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1.
- Rosmond, T., and L. Xu, 2006: Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus*, **58A**, 45– 58, https://doi.org/10.1111/j.1600-0870.2006.00148.x.
- Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi, 2016: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophys. Res. Lett.*, 43, 752–759, https://doi.org/10.1002/2015GL067238.
- Storto, A., M. Martin, B. Deremble, and S. Masina, 2018: Strongly coupled data assimilation experiments with linearized ocean– atmosphere balance relationships. *Mon. Wea. Rev.*, 146, 1233– 1257, https://doi.org/10.1175/MWR-D-17-0222.1.
- Takaya, Y., J. R. Bidlot, A. C. M. Beljaars, and P. A. E. M. Janssen, 2010: Refinements to a prognostic scheme of skin sea surface temperature. J. Geophys. Res., 115, C06009, https://doi.org/ 10.1029/2009JC005985.
- Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction. International Geophysics Series, Vol. 59, Elsevier, 467 pp.
- Xu, L., T. Rosmond, and R. Daley, 2005: Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559, https://doi.org/10.1111/j.1600-0870.2005.00123.x.
- Zeng, X., and A. Beljaars, 2005: A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophys. Res. Lett.*, **32**, L14605, https://doi.org/ 10.1029/2005GL023030.