# Inclusion of sea-surface temperature variation in the U.S. Navy ensemble-transform global ensemble prediction system

J. G. McLay,[1] M. K. Flatau,[1] C. A. Reynolds,[1] J. Cummings,[1] T. Hogan,[1] and P. J. Flatau[1]

Received 29 September 2011; revised 30 August 2012; accepted 30 August 2012; published 13 October 2012.

[1] The local ensemble transform (ET) analysis perturbation scheme is adapted to generate perturbations to both atmospheric variables and sea-surface temperature (SST). The adapted local ET scheme is used in conjunction with a prognostic model of SST diurnal variation and the Navy Operational Global Atmospheric Prediction System (NOGAPS) global spectral model to generate a medium-range forecast ensemble. When compared to a control ensemble, the new forecast ensemble with SST variation exhibits notable differences in various physical properties including the spatial patterns of surface fluxes, outgoing longwave radiation (OLR), cloud radiative forcing, near-surface air temperature and wind speed, and 24-h accumulated precipitation. The structure of the daily cycle of precipitation also is substantially changed, generally exhibiting a more realistic midday peak of precipitation. Diagnostics of ensemble performance indicate that the inclusion of SST variation is very favorable to forecasts in the Tropics. The forecast ensemble with SST variation outscores the control ensemble in the Tropics across a broad set of metrics and variables. The SST variation has much less impact in the Midlatitudes. Further comparison shows that SST diurnal variation and the SST analysis perturbations are each individually beneficial to the forecast from an overall standpoint. The SST analysis perturbations have broader benefit in the Tropics than the SST diurnal variation, and inclusion of the SST analysis perturbations together with the SST diurnal variation is essential to realize the greatest gains in forecast performance.

**Citation:** McLay, J. G., 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, doi:10.1029/2011JD016937.

## 1. Introduction

[2] The uppermost layers of the ocean are of particular interest for their regulation of various atmospheric processes. Ocean-atmosphere interaction occurs via the top 0.1–1.0 mm of the ocean, referred to as the "skin layer" [*Kawai and Wada*, 2007]. The skin layer and the underlying one to three meters of the ocean exhibit diurnal variation that is typically in the range 0–1 K, but that can reach up to several degrees in areas such as the Pacific warm pool in conditions of low wind speed and clear skies [e.g., *Flament et al.*, 1994; *Ward*, 2006; *Kawai and Wada*, 2007; *Bellenger and Duvel*, 2009]. There is evidence that this diurnal variation is an important determinant of the surface latent and sensible heat fluxes that modulate the atmospheric planetary boundary layer [e.g., *Clayson and Chen*, 2002; *Kawai and Wada*, 2007]. Also, the diurnal variation appears to affect precipitation rate [*Brunke et al.*, 2008] and local atmospheric circulations that are driven by land-sea thermal contrasts, such as the sea breeze [*Kawai et al.*, 2006]. Furthermore, the diurnal variation may play an important role in the Madden-Julian Oscillation (MJO). Some studies have demonstrated that numerical weather prediction (NWP) simulation of the propagation of the MJO from the Indian Ocean region is improved by inclusion of diurnal variability in the NWP model [*Vitart et al.*, 2007; *Woolnough et al.*, 2007; *Takaya et al.*, 2010a]. Also, it has been hypothesized that diurnal variation acts as a trigger mechanism for shallow convection over the Indian Ocean, which proceeds to moisten and precondition the lower troposphere for the convectively active phase of the MJO [*Slingo et al.*, 2003]. Recent observational evidence seems to support this hypothesis [*Bellenger et al.*, 2010]. SST diurnal variation might help to trigger convection over the western Pacific Ocean as well [*Parsons et al.*, 2000].

[3] The mounting evidence that the upper ocean diurnal cycle influences important weather phenomena suggests the importance of incorporating upper ocean variability into global NWP systems. Efforts in this regard could be worthwhile in spite of the fact that upper ocean variability is already an inherent part of many operational seasonal and climate prediction systems [e.g., *Arribas et al.*, 2011]. This is because these seasonal and climate prediction systems are not optimized for weather prediction. For instance, they typically use a daily or longer update cycle, as opposed to

[1]Naval Research Laboratory, Monterey, California, USA.

Corresponding author: J. McLay, Naval Research Laboratory, 7 Grace Hopper Ave., Stop 2, Monterey, CA 93943, USA. (justin.mclay@nrlmry.navy.mil)

the standard six-hourly update cycle of operational NWP systems. Also, they typically use a "lagged-start" ensemble methodology that provides no theoretical constraint on the sampling of initial-state uncertainty. To date, two NWP centers have made documented efforts to incorporate upper ocean variability into a global NWP system. The European Centre for Medium-Range Weather Forecasting (ECMWF) coupled the K-Profile Parameterization (KPP) one-dimensional ocean mixed-layer model of *Large et al.* [1994] with the ECMWF Integrated Forecasting System (IFS) and tested the modified IFS in conjunction with the ECMWF 51-member ensemble prediction system (EPS) [*Takaya et al.*, 2010a]. Diagnostics for the medium-range forecast interval (0–16 d lead time) indicated that the inclusion of the KPP model improved SST forecast skill at the 10 d lead time and beyond, as well as MJO propagation and Indian monsoon rainfall. A brief comparison also was made between the diurnal SST amplitude (DSA) produced by the KPP model and the DSA produced by the skin-layer SST scheme of *Takaya et al.* [2010b]. Notwithstanding its relative simplicity, the skin-layer scheme appeared to have a slight advantage in simulating the DSA. One nuance of the tests with the EPS is that the ensemble generation neglected SST initial-state uncertainty, such that all NWP integrations were begun from the same prescribed SST field. Nevertheless, research is ongoing in regard to using SST analysis perturbations generated by the ensembles of data assimilation (EDA) technique [*Buizza et al.*, 2010]. Meanwhile, the Met Office of the United Kingdom has introduced random, spatially correlated SST analysis perturbations within the local ensemble transform Kalman filter (ETKF) of its operational Met Office Global and Regional Ensemble Prediction System (MOGREPS). These random SST analysis perturbations provide a beneficial increase in the ensemble spread of surface temperature and wind [*Tennant and Beare*, 2011; *Tennant*, 2012]. In addition to the efforts at ECMWF and the Met Office, *Kunii and Miyoshi* [2012] have recently examined SST initial-state uncertainty in an ensemble Kalman filter (EKF).

[4] The results of *Takaya et al.* [2010a] and *Tennant and Beare* [2011] indicate that global NWP ensembles that take into account upper ocean variability can produce better forecasts than baseline ensembles that lack upper ocean variability. Given these promising results, the present study investigates the incorporation of a variant of the *Takaya et al.* [2010b] diurnal SST scheme into the NWP component of the U.S. Navy's global EPS. In addition, this study investigates the adaptation of the local ensemble transform (ET) ensemble generation scheme of *McLay et al.* [2010] to generate SST analysis perturbations. Hence, the EPS design for the present study is distinct from those of *Takaya et al.* [2010a] and *Tennant and Beare* [2011] in that it takes account of both SST diurnal variation and SST initial-state uncertainty. The new EPS is used to produce a 32-member, 336 h lead time global forecast ensemble and the distributions of SST and near-surface atmospheric variables from this ensemble are compared with those of a control ensemble generated without SST analysis perturbations and with a prescribed SST field that is fixed throughout integration. The importance of the SST analysis perturbations is also investigated. The paper is organized as follows. Section 2 describes the data and methodology including the atmospheric and SST analyses, the global NWP model, the diurnal SST scheme, the ET analysis perturbations, and the forecast ensemble configurations. Results and conclusions are presented in sections 3 and 4, respectively.

## 2. Data and Methodology

### 2.1. Analyses, Numerical Model, and Verification Data

[5] Global atmospheric analyses were obtained from the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System (NAVDAS), a three-dimensional variational assimilation system [*Daley and Barker*, 2001a, 2001b].

[6] SST analyses were obtained from a two-dimensional version of the Navy Coupled Ocean Atmosphere Data Assimilation System (NCODA) [*Cummings*, 2005]. The NAVDAS analyses were generated in a 6 h cycle while the NCODA analyses were generated in a 12 h cycle.

[7] Numerical integrations were carried out using the Navy Operational Global Atmospheric Prediction System (NOGAPS), a primitive-equation, fully parameterized spectral model [*Peng et al.*, 2004]. The experiments employ NOGAPS at horizontal resolution T119 and with 30 vertical levels, the same resolution as used for the Fleet Numerical Meteorology and Oceanography Center (FNMOC) operational forecast ensemble at the time of the study. Integrations to the 336 h lead time were completed for each 00 UTC analysis time in the period of 1 May 2007 to 30 June 2007.

[8] All forecasts in section 3.2.2 were verified against the NOGAPS/NCODA analysis with the exception of 24-h accumulated precipitation, which was verified against Tropical Rainfall Measuring Mission (TRMM) satellite-derived daily rainfall estimates [*Simpson et al.*, 1988]. The daily cycle of precipitation (section 3.1.2) was verified against TRMM three-hourly rain-rate estimates (mm $h^{-1}$) that were converted to 3 h accumulated precipitation values. Note that the TRMM daily rainfall estimates and the TRMM three-hourly rain-rate estimates are two different TRMM products. The DSA (section 3.1.3) was verified against observations of DSA derived from Geostationary Operational Environmental Satellite-East (GOES-East) and GOES-West raw satellite observations of SST.

### 2.2. Prognostic Model of SST

[9] *Zeng and Beljaars* [2005] use the one-dimensional heat transfer equation to derive a model that predicts SST diurnal variation within two layers: a thin, cool skin layer, and a 3 m-deep sub-skin layer. The physical processes represented include shortwave and longwave radiative flux, evaporation, molecular thermal conduction, and wind-driven turbulent diffusion determined through Monin-Obhukov similarity theory. Effects related to wave-breaking, the Langmuir circulation, precipitation, and buoyancy-driven mixing processes are neglected. *Takaya et al.* [2010b] refine the *Zeng and Beljaars* [2005] model by including a different stability function in the definition of the wind-driven turbulent diffusion coefficient, and by including mixing effects related to the Langmuir circulation during stable conditions. The diurnal model of SST that was incorporated into NOGAPS for the present study is

that of *Takaya et al.* [2010b], but excluding the explicit Langmuir circulation and the cool skin layer. The explicit Langmuir circulation is neglected because a wave model was not readily available for use. However, Langmuir circulation effects are implicitly modeled by multiplying the friction velocity by a factor of 1.4 to obtain a slight enhancement of surface stress. This enhancement was validated against Global Drifter Program drifting-buoy SST observations in the Tropics (not shown). The cool skin layer is neglected for a combination of two reasons. The first is that the amplitude of nighttime cooling from the skin layer typically makes only a small contribution to the total diurnal SST amplitude [see *Gentemann et al.*, 2003, Figure 1]. The second reason is to facilitate forecast verification. Note, specifically, that forecasts of the skin layer SST can only be verified against certain satellite data (e.g., the Ocean Pathfinder data of *Gentemann et al.* [2003]), whereas forecasts of the sub-skin layer SST can be verified against the NCODA SST analyses. The NCODA SST analyses represent the warm sub-skin layer by design, because NCODA assimilates satellite radiances by regressing them against drifting-buoy observations which measure the sub-skin layer (J. Cummings, personal communication, 2012). In terms of the diurnal SST model, neglect of the cool skin layer prevents the SST from assuming values that are cooler than a base SST, where the base SST is defined to be the analyzed SST at the time of forecast initialization. The diurnal warming in the sub-skin layer is not affected. Finally, note that the SST is updated at every time step (specifically, every 600 s) during the NOGAPS integrations.

### 2.3. Generation of Analysis Perturbations

[10] The ensemble size was chosen to be 32, commensurate with typical ensemble sizes at operational centers. With the exception of the control forecast ensemble, each ensemble member's initial conditions included perturbations to SST and to atmospheric wind, temperature, specific humidity, and terrain pressure. The perturbations to wind and temperature were calculated for all vertical levels of the numerical model, while the perturbations to specific humidity were calculated for those vertical levels between the surface and roughly 300 hPa.

[11] For the control forecast ensemble, each ensemble member's SST field was held constant throughout the forecast integration and was prescribed to be the NCODA SST analysis for the given ensemble's initialization date.

[12] Perturbations to the NAVDAS and NCODA analyses were generated in a single, unified process using a nine-banded local formulation of the ensemble transform (ET) method. The fundamental ET equation set is

$$\mathbf{Z}_a = \mathbf{Z}_f\mathbf{T} : \mathbf{Z}_a^T(\mathbf{P}_a)^{-1}\mathbf{Z}_a = N\mathbf{I}, \tag{1}$$

where $N$ is the size of the NWP model state vector (here including both atmospheric and SST elements), $K$ is the ensemble size, $\mathbf{Z}_f$ ($\mathbf{Z}_a$) is an $N \times K$ matrix of forecast (analysis) perturbations, $\mathbf{T}$ is a $K \times K$ matrix of weighting coefficients, $\mathbf{P}_a$ is a $N \times N$ matrix of analysis error covariance estimates, and $\mathbf{I}$ is the $K \times K$ identity matrix [*McLay et al.*, 2008]. The constraint $\mathbf{Z}_a^T(\mathbf{P}_a)^{-1}\mathbf{Z}_a = N\mathbf{I}$ imparts a magnitude to the perturbations that is consistent in a domain-average sense with estimates of analysis error covariance and ensures that the perturbations are quasi-orthogonal. For all experiments presented in this paper, $\mathbf{P}_a$ is diagonal and its respective atmospheric and SST elements are obtained directly from NAVDAS and NCODA. Both NAVDAS and NCODA estimate the elements of $\mathbf{P}_a$ using a method based on the block diagonal preconditioner of the conjugate gradient minimization algorithm and Choleski decomposition. The method is detailed in *Daley and Barker* [2001a] and briefly summarized in *Daley and Barker* [2001b] and *Cummings and Smedstad* [2012]. In general, $\mathbf{P}_a$ is not diagonal, but the operational implementations of the ET at the National Center for Environmental Prediction (NCEP) and at the Fleet Numerical Meteorology and Oceanography Center (FNMOC) both employ diagonal $\mathbf{P}_a$ to reduce the computational expense of forecast-ensemble generation.

[13] *McLay et al.* [2010] realize improved performance from the ET through a nine-banded local formulation. In this formulation, the global NWP domain is partitioned into nine non-overlapping latitude bands and perturbation generation proceeds in three steps.

[14] 1. Nine different $[K \times K]$ weighting matrices $\mathbf{T}_j(j = 1,.., 9)$ are obtained, one for each of the nine latitude bands. Define $n_j$, $\mathbf{Z}_f^j$, and $\mathbf{P}_a^j$ to be the number of elements of the state vector within band $j$, the $[n_j \times K]$ forecast perturbation matrix for band $j$, and the $[n_j \times n_j]$ diagonal analysis-error covariance matrix for band $j$, respectively. Then $\mathbf{T}_j$ is obtained by solving the standard ET eigen-equation

$$\left[\frac{1}{n_j}\mathbf{Z}_f^{jT}\left(\mathbf{P}_a^j\right)^{-1}\mathbf{Z}_f^j\right]\mathbf{C} = \mathbf{C}\Lambda \tag{2}$$

where $\mathbf{T}_j = \mathbf{C}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{C}^T$ (note that solution of this eigen-equation requires special treatment of a zero-valued eigenvalue, as detailed in *McLay et al.* [2008]). Once the $\mathbf{T}_j$ are obtained, they are "centered" in their respective latitude bands. The centers of the $\mathbf{T}_j$ are analogous to the "localization centres" of *Bowler et al.* [2009].

[15] 2. Let $M$ equal the number of Gaussian latitudes of the NWP model. Then for each Gaussian latitude $m(m = 1,.., M)$, a $[K \times K]$ weighting matrix $\mathbf{T}^m_{\text{Gaussian}}$ is obtained through linear interpolation of the centered weighting matrices $\mathbf{T}_j(j = 1,.., 9)$.

[16] 3. Define $N_m$ to be the number of elements of the state vector within the vertical slice of domain through Gaussian latitude $m$. Then for each Gaussian latitude $m(m = 1,.., M)$, the ensemble transform equation

$$\mathbf{Z}_a^m = \mathbf{Z}_f^m\mathbf{T}^m_{\text{Gaussian}} \tag{3}$$

is solved, where $\mathbf{Z}_f^m(\mathbf{Z}_a^m)$ is the $[N_m \times K]$ forecast (analysis) perturbation matrix for Gaussian latitude $m$.

### 2.4. Forecast-Ensemble Configurations

[17] Four different forecast ensembles were generated for this study.

[18] 1. *Control*: An ensemble generated with no SST variation during the numerical integration. Each ensemble member's SST field is held constant throughout the integration and is prescribed to be the NCODA SST analysis for the given ensemble's initialization date.

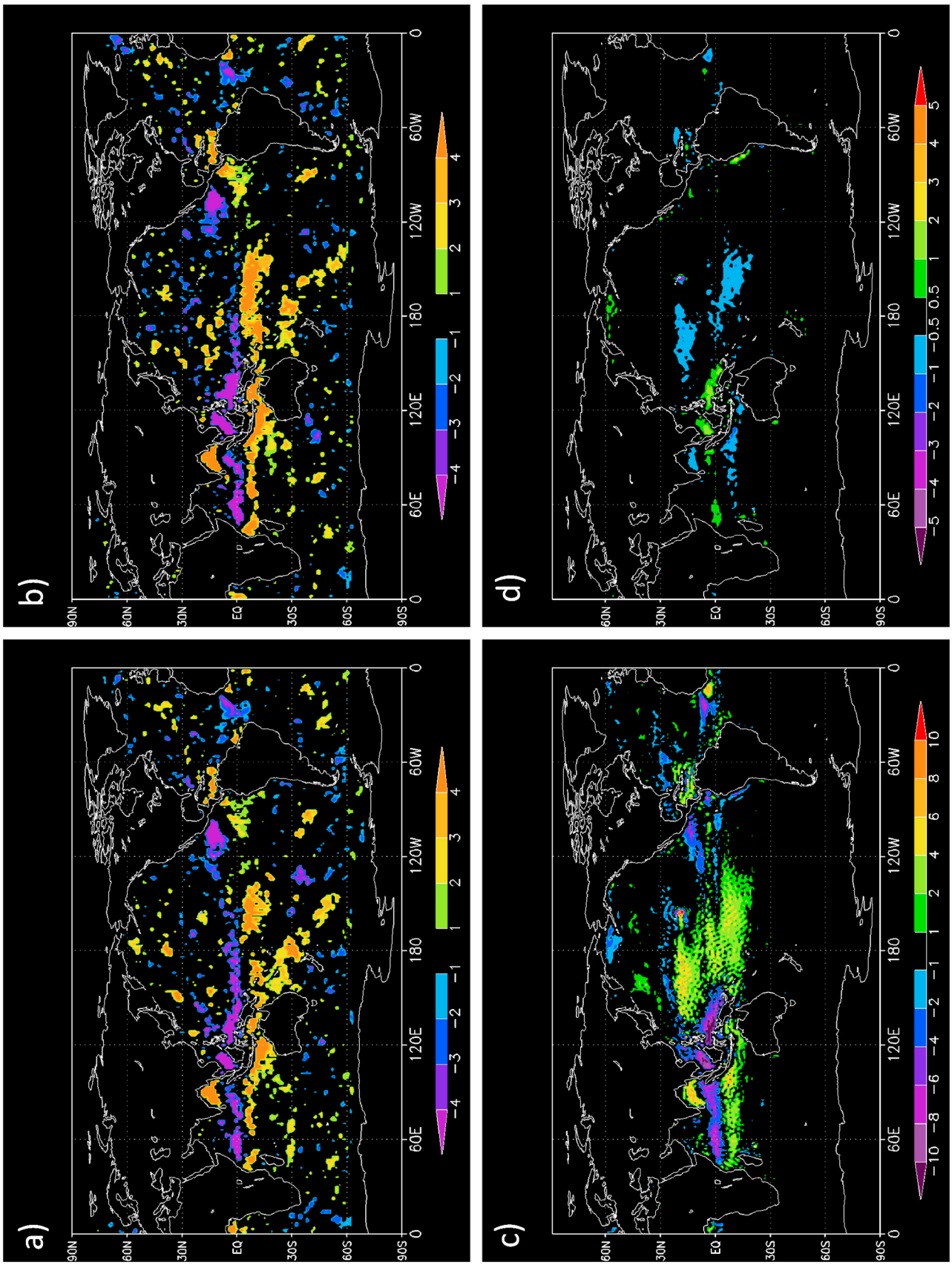
a)
b)
c)
d)
90N
60N
30N
EQ
30S
60S
90S
0
60W
120W
180
120E
60E
0
−4 −3 −2 −1 1 2 3 4
−5 −4 −3 −2 −1 −0.5 0.5 1 2 3 4 5
−10 −8 −6 −4 −2 −1 1 2 4 6 8 10

Figure 1

[19] 2. *SSTidf*: An ensemble generated with both diurnal forcing of SST and SST analysis perturbations (idf stands for initial-perturbations and diurnal forcing).

[20] 3. *SSTdf*: An ensemble generated with diurnal forcing of SST but no SST analysis perturbations. At analysis time, the SST field of each ensemble member is identical to that of the NCODA SST analysis for the given ensemble's initialization date. Note that although there are no SST analysis perturbations the diurnal SST forcing is still capable of affecting the atmospheric analysis perturbations by virtue of its interaction with the atmospheric forecast perturbations during the 6 h cycling interval.

[21] 4. *SSTi*: An ensemble generated with SST analysis perturbations but no diurnal forcing of SST. The analysis perturbations are taken to be those of forecast ensemble SSTidf, to ensure that the perturbations are balanced. Recognize that each ensemble member's SST field is different from the other members', but the SST field is held constant throughout the integration.

## 3. Results

### 3.1. Diagnostics of Physical Properties

[22] Forecast-ensemble SSTidf was compared to the control ensemble in terms of various physical variables including 24-h accumulated precipitation, surface latent- and sensible-heat flux, outgoing longwave radiation (OLR), cloud radiative forcing, 2-m air temperature, and 10-m wind speed to highlight differences to these variables that are brought about by inclusion of the diurnal SST model and the SST analysis perturbations. Also, ensemble SSTidf was compared to observations in terms of its daily cycle of precipitation in the Tropics and its DSA conditioned on wind speed.

#### 3.1.1. Differences in Physical Variables

[23] All differences described here are average differences over the test period. Also, the statistical significance of the average difference at each grid point was tested using a method that accounts for serial dependence of data [*Wilks*, 2006, pp. 143–145] and only differences that meet a 95% significance threshold are shown. Additionally, only over-ocean differences are discussed, except in the case of 24-h accumulated precipitation. For interpretation purposes, note that the T + 240 h (T + 252 h) forecast is valid at 00 UTC (12 UTC).

[24] 1. *Surface latent-heat flux*: Figure 1a (Figure 1b) shows the difference between the ensemble-mean surface latent-heat flux at T + 240 h (T + 252 h) in SSTidf and that in the control. One can infer that during daylight hours (e.g., over the Pacific Ocean at T + 240 h) the latent-heat flux in SSTidf is greater over the ocean in both the Tropics and the NH Midlatitudes, as might be expected. During nighttime hours (e.g., over the Atlantic Ocean at T + 240 h), the latent-heat flux in SSTidf tends to be smaller over the ocean in the Tropics, but not generally in the NH Midlatitudes. The latent-heat flux in SSTidf is greater in the equatorial and southern Indian Ocean at both day and night.

[25] 2. *Surface sensible-heat flux*: Figure 1c (Figure 1d) shows the difference between the ensemble-mean surface sensible-heat flux at T + 240 h (T + 252 h) in SSTidf and that in the control. As with the latent-heat flux, the sensible-heat flux in SSTidf during daylight hours is typically greater over the ocean in both the Tropics and the NH Midlatitudes. At night, the sensible-heat flux in SSTidf is smaller over the ocean in the Tropics, the NH Subtropics, and the higher-latitude North Pacific Ocean, the sensible-heat flux is not significantly different over most of the eastern Indian Ocean at either day or night.

[26] 3. *OLR*: Figure 2a (Figure 2b) shows the difference between the ensemble-mean OLR at T + 240 h (T + 252 h) in SSTidf and that in the control. At T + 240 h, the OLR in SSTidf is substantially smaller (greater) to the immediate north (south) of the Equator throughout the Indian Ocean and the western and central Pacific Ocean. In the eastern Pacific Ocean, the OLR is greater over the Equator and smaller in the Subtropics. Meanwhile, the OLR is smaller over the tropical Atlantic Ocean and greater over the Bay of Bengal. The same general patterns are seen at T + 252 h. The roughly cross-equatorial change in OLR seen in the Indian and western Pacific Oceans suggests that in SSTidf there is a strengthened Inter-Tropical Convergence Zone (ITCZ) in these regions, with deeper convection (i.e., colder convective cloud tops) to the north of the Equator where the ITCZ is located in NH summer.

[27] 4. *Cloud radiative forcing*: Figure 2c (Figure 2d) shows the difference between the ensemble-mean shortwave (longwave) cloud radiative forcing [SCRF (LCRF)] in SSTidf and that in the control. The cloud radiative forcing is defined to be the difference between the total-sky and the clear-sky net radiative flux at the surface. The difference in SCRF generally bears a direct relationship to the difference in OLR (Figure 2a), with notable exceptions in the equatorial central and eastern Pacific Ocean. The difference in LCRF is roughly anti-correlated to the difference in SCRF, although both the extent and magnitude of the significant differences in LCRF are smaller (Figure 2d). From a global-mean perspective, the ensemble-mean SCRF in the control (SSTidf) is −43.1 (−43.0) $Wm^{-2}$, and the ensemble-mean LCRF in the control (SSTidf) is +34.1 (+34.0) $Wm^{-2}$. Hence, the SST variation induces a slight reduction in global-mean magnitude of both the SCRF and LCRF. This reduction is statistically significant.

[28] 5. *Accumulated precipitation over a 24-h interval*: Figure 3 shows the difference between the ensemble-mean precipitation accumulated over the 24-h forecast interval ending at T + 240 h in the SSTidf ensemble and that in the control. Ensemble SSTidf produces more precipitation over the ocean in certain parts of the Tropics, including the West Pacific warm pool region, the western Indian Ocean, and the eastern Pacific Ocean near Central America. Conversely, SSTidf produces less precipitation over some land masses, including the Maritime Continent and Central America.

**Figure 1.** (a) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean surface latent-heat flux ($Wm^{-2}$) at T + 240 h. Colored shading indicates statistically significant average differences. (b) As for Figure 1a but at T + 252 h. (c) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean surface sensible-heat flux ($Wm^{-2}$) at T + 240 h. (d) As for Figure 1c but at T + 252 h.

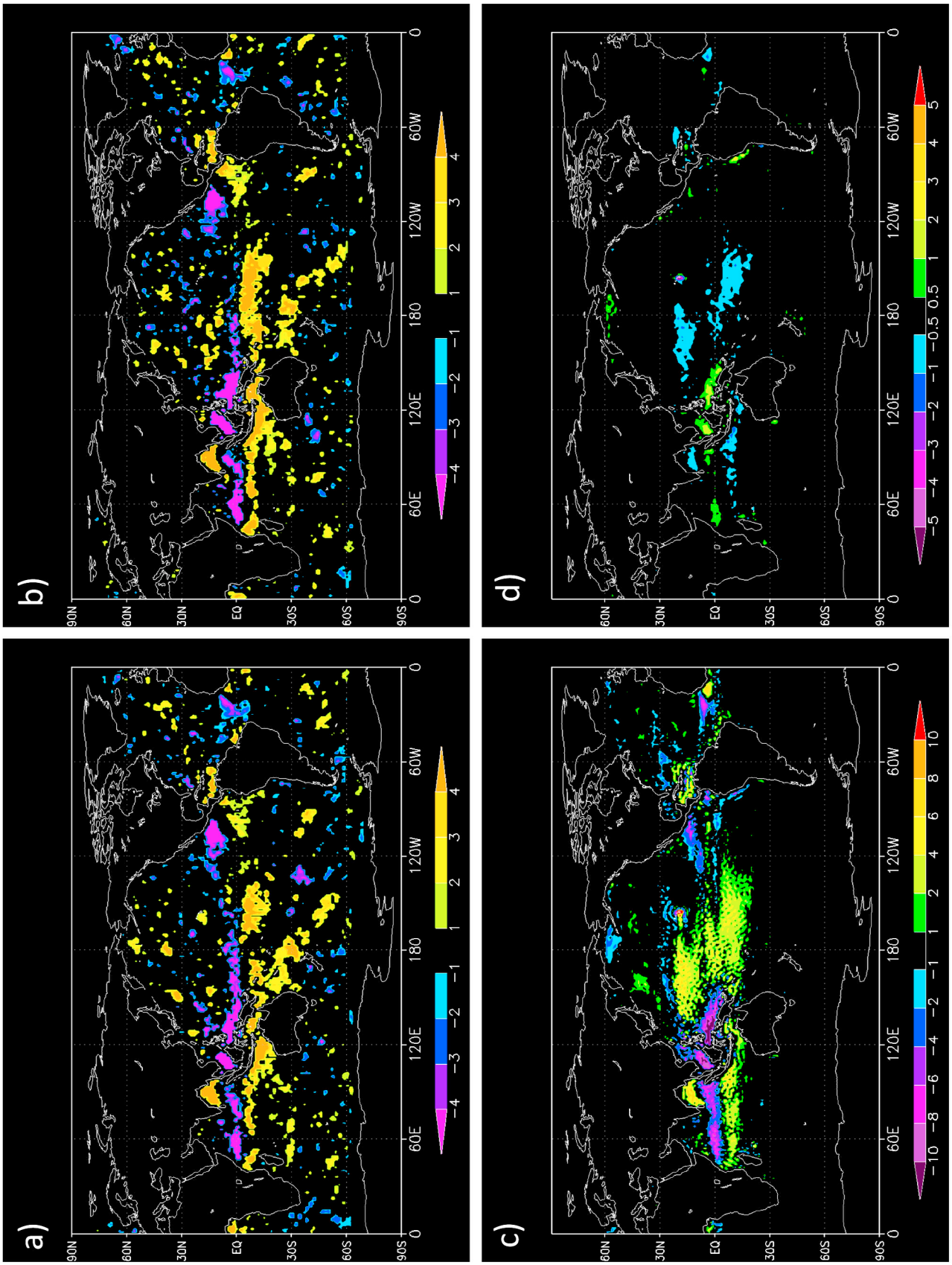
a)
b)
c)
d)
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
−4 −3 −2 −1 1 2 3 4
−4 −3 −2 −1 1 2 3 4
−10 −8 −6 −4 −2 −1 1 2 4 6 8 10
−5 −4 −3 −2 −1 −0.5 0.5 1 2 3 4 5

**Figure 2**

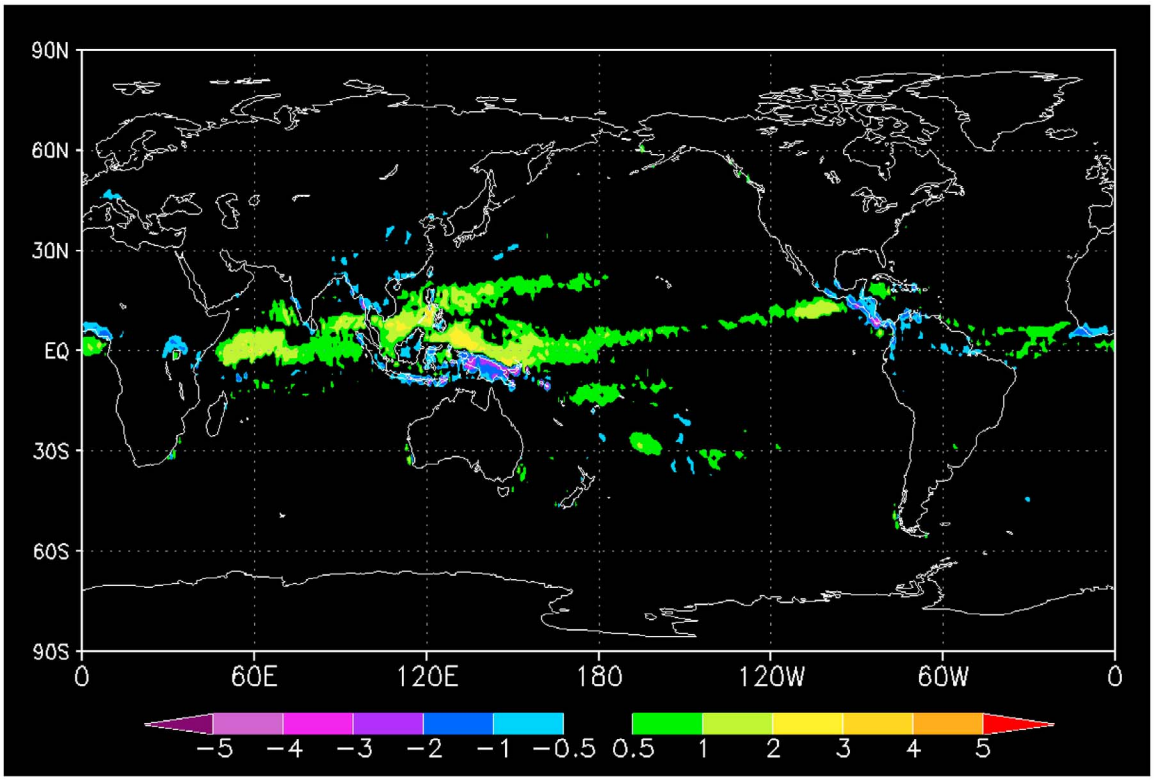


**Figure 3.** Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean precipitation (mm) accumulated over the 24-h forecast interval ending at T + 240 h. Colored shading indicates statistically significant average differences.

[29] 6. *Air temperature at 2 m*: Figure 4a (Figure 4b) shows the difference between the ensemble-mean 2-m air temperature at T + 240 h (T + 252 h) in SSTidf and that in the control. Figure 4a shows that at T + 240 h the 2-m air temperature in SSTidf is greater over a number of regions, most notably the subtropical eastern Pacific Ocean, but also the midlatitude eastern Pacific Ocean, the Gulf of Mexico, the western Atlantic Ocean, the Indian Ocean, and the West Pacific warm pool. Figure 4b shows that at T + 252 h the 2-m air temperature in SSTidf is substantially greater over the West Pacific warm pool, and also is greater over most of the Indian Ocean north of the Equator. These difference patterns suggest that the diurnal warming of SST enhances the warm part of the daily cycle of 2-m air temperature, as would be expected.

[30] 7. *Wind speed at 10 m*: Figure 4c (Figure 4d) shows the difference between the ensemble-mean 10-m wind speed at T + 240 h (T + 252 h) in SSTidf and that in the control. At both T + 240 h and T + 252 h, the 10-m wind speed in SSTidf is substantially higher over the Indian Ocean and West Pacific warm pool, with the maximum difference $\approx$0.5 $ms^{-1}$. In a small number of places the wind speed is lower, notably the Bay of Bengal, the eastern Pacific south of Panama, and in the vicinity of the Gulf of Mexico.

[31] The differences noted above in the physical variables' ensemble-mean properties might arise for two reasons. One is the fact that a non-zero mean can emerge within the set of ensemble perturbations as the perturbations are integrated, even though the mean of the perturbations is zero by design at the outset of integration. The emergence of a non-zero mean is a consequence of integration with a nonlinear model, and can occur for both the atmospheric perturbations and the SST perturbations. The other possible reason for the differences in the ensemble-mean properties is the configuration of the model of SST diurnal variation. Recall that this model superposes the diurnal variation on a base temperature which in the present study is taken to be the SST analysis for the initialization date of the forecast. Also, recall that

**Figure 2.** (a) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean OLR ($Wm^{-2}$) at T + 240 h. Colored shading indicates statistically significant average differences. (b) As for Figure 2a but at T + 252 h. (c) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean SCRF ($Wm^{-2}$) averaged over the 336 h forecast interval. Colored shading indicates statistically significant average differences. (d) As for Figure 2c but for LCRF.

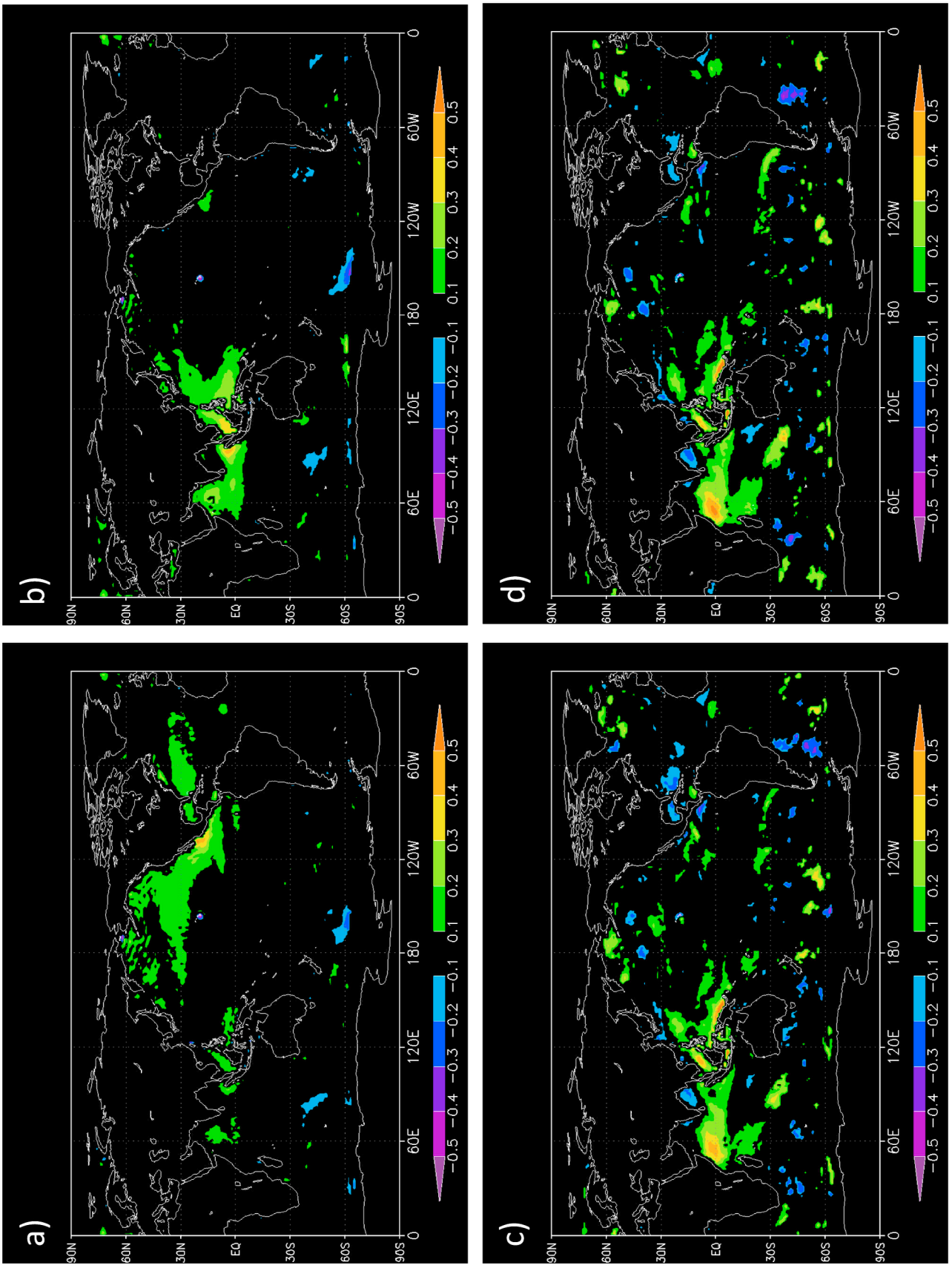
a)
b)
c)
d)
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
−0.5 −0.4 −0.3 −0.2 −0.1 0.1 0.2 0.3 0.4 0.5

Figure 4

the SST cannot cool below the base temperature. It follows that a forecast with SST diurnal variation will have a slightly warmer SST, from a spatially and temporally averaged perspective, than that same forecast would if only persisted SST were used. Thus, inclusion of the SST diurnal variation enhances not just the variability but also the mean of the NWP model's bottom boundary thermal forcing. Quantitatively, the SST diurnal variation causes the daily average SST to increase by an amount typically between 0.10°C and 0.25°C in those tropical regions with the greatest DSA, with the largest increases being observed in the South China Sea region. The daily average SST in the midlatitude North Pacific and North Atlantic Oceans increases by an amount typically between 0.05°C and 0.15°C. To investigate if either the nonlinear evolution of the ensemble perturbations or the SST diurnal variation is the dominant source of the differences noted above, comparisons were made between the control and SSTi ensembles and between the control and SSTdf ensembles in terms of ensemble-mean surface latent and sensible heat flux (not shown). In general, the SSTi and SSTdf ensembles exhibited quantitatively comparable differences relative to the control. This suggests that both the nonlinear evolution of the ensemble perturbations and the SST diurnal variation make important contributions to the differences.

#### 3.1.2. Diurnal Cycle of Precipitation in the Tropics

[32] As shown above, the inclusion of diurnal SST variation and SST analysis perturbations discernably changes the forecasts of variables such as latent-heat flux that play a role in governing convective processes. Given these changes, it is possible that the timing and structure of the forecast daily cycle of precipitation are correspondingly altered. To investigate this possibility, here a composite daily cycle of precipitation conditional upon wind speed regime is constructed for forecast SSTidf and compared to a corresponding composite for the control forecast and to a composite of the observed daily cycle as derived from TRMM satellite data (described above in section 2.1).

[33] The calculation of the composite forecast daily cycle of precipitation proceeds as follows. First, a longitude band of 10° width is chosen (e.g., 0°E to 10°E, 10°E to 20°E, .., 350°E to 360°E). Then, the compositing region is defined as the intersection of this longitude band with the 20°S to 20°N tropical latitude band. The use of 10° longitude bands ensures that all grid points within the compositing region are less than 1 h apart in local time. Next, a wind speed regime, a 24-h forecast interval (e.g., T + 240 h to T + 264 h), and a particular grid point (on a 1° × 1° latitude-longitude grid) within the compositing region are chosen. Then, for each 00 UTC forecast initialization in the test period and each ensemble member, a check is made to ensure that the ensemble member's 10-m wind speed forecast at the chosen grid point meets the specified wind speed criterion (e.g., 0–3 $ms^{-1}$) at every three-hourly point of the forecast interval (e.g., T + 240 h, T + 243 h, T + 246 h, ..). If the wind speed meets the criterion, then the ensemble member's sequence of three-hourly accumulated precipitation for the 24-h forecast interval is added to the composite. When the sequence of precipitation is added to the composite, the sequence is normalized by its maximum value. This normalization ensures that sequences with light precipitation get the same weight in the composite as sequences with heavy precipitation. The preceding process is repeated for all 00 UTC forecast initializations, all ensemble members, and all grid points within the compositing region. Once the composite sequence is obtained, it is normalized by its maximum value. This final normalization removes differences in amplitude between one composite sequence and another that arise because of the different number of sequences that comprise each composite.

[34] Calculation of the composite observed daily cycle of precipitation proceeds in much the same way as for the composite forecast daily cycle. First, the TRMM data are interpolated to the same 1° × 1° latitude-longitude grid as the forecast data. Then, a wind speed regime, a date, and a particular grid point within the compositing region are chosen. A check is made to ensure that the analyzed 10-m wind speed at the grid point meets the specified wind speed criterion at all four daily analysis times 00, 06, 12, and 18 UTC. If the wind speed meets the criterion, then the TRMM sequence of three-hourly accumulated precipitation for the date is added to the composite. The TRMM sequences used for the composite are normalized in the same fashion as the sequences used for the composite forecast daily cycle.

[35] Figure 5 shows the composite daily cycle of precipitation for the SSTidf ensemble, the control ensemble, and the TRMM precipitation observations for six different compositing regions. The composite forecast daily cycles are for the interval from T + 120 h to T + 144 h. Also, all composites are for the low-wind speed (0–3 $ms^{-1}$) regime, since one would expect the largest signal from the diurnal SST forcing in this regime. Considering the western and eastern Indian Ocean compositing regions (Figures 5a and 5b), one sees a prominent midday precipitation peak in the TRMM observations. There is little to no indication of this midday peak in the control ensemble, whereas there is a pronounced midday peak in the SSTidf ensemble, albeit three hours too late in both regions. All three composite cycles also show an early morning precipitation peak. However, in the case of the SSTidf ensemble, this peak is somewhat diminished in relative prominence as compared to the midday peak. The timing of the early morning peak for the western Indian Ocean (Figure 5a) is three hours too late in both the control and the SSTidf ensembles. For the South China Sea compositing region (Figure 5c), the representation of the midday precipitation peak in the SSTidf ensemble is

**Figure 4.** (a) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean 2-m air temperature (°C) at T + 240 h. Colored shading indicates statistically significant average differences. (b) As for Figure 4a but at T + 252 h. (c) Average difference between the SSTidf ensemble and the control ensemble in terms of the ensemble-mean 10-m wind speed ($ms^{-1}$) at T + 240 h. (d) As for Figure 4c but at T + 252 h.

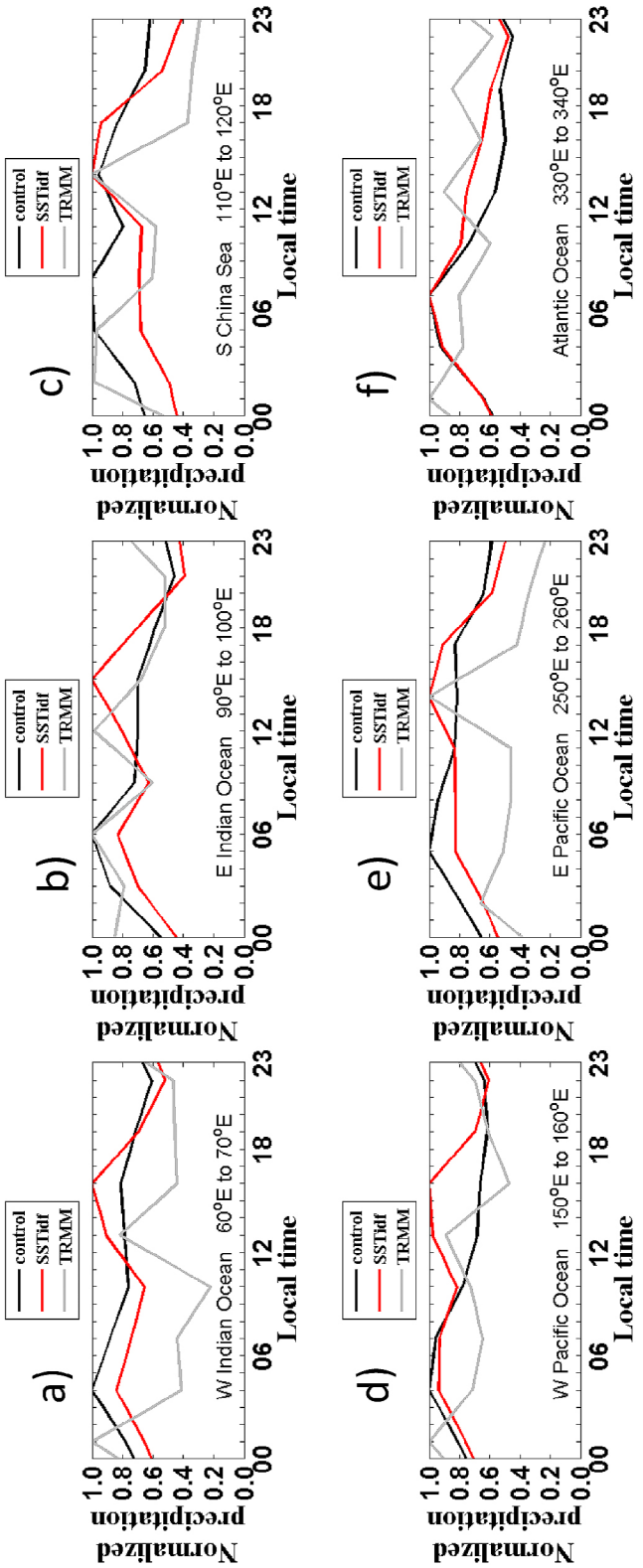


**Figure 5.** Composite daily cycle of tropical over-ocean precipitation for the 0–3 $ms^{-1}$ wind speed regime for the control ensemble (black line), the SSTidf ensemble (red line), and the TRMM precipitation observations (grey line) for six different compositing regions. (a) Western Indian Ocean [60°E to 70°E, 20°S to 20°N], (b) eastern Indian Ocean [90°E to 100°E, 20°S to 20°N], (c) South China Sea [110°E to 120°E, 20°S to 20°N], (d) western Pacific Ocean [150°E to 160°E, 20°S to 20°N], (e) eastern Pacific Ocean [250°E to 260°E, 20°S to 20°N], (f) Atlantic Ocean [330°E to 340°E, 20°S to 20°N].

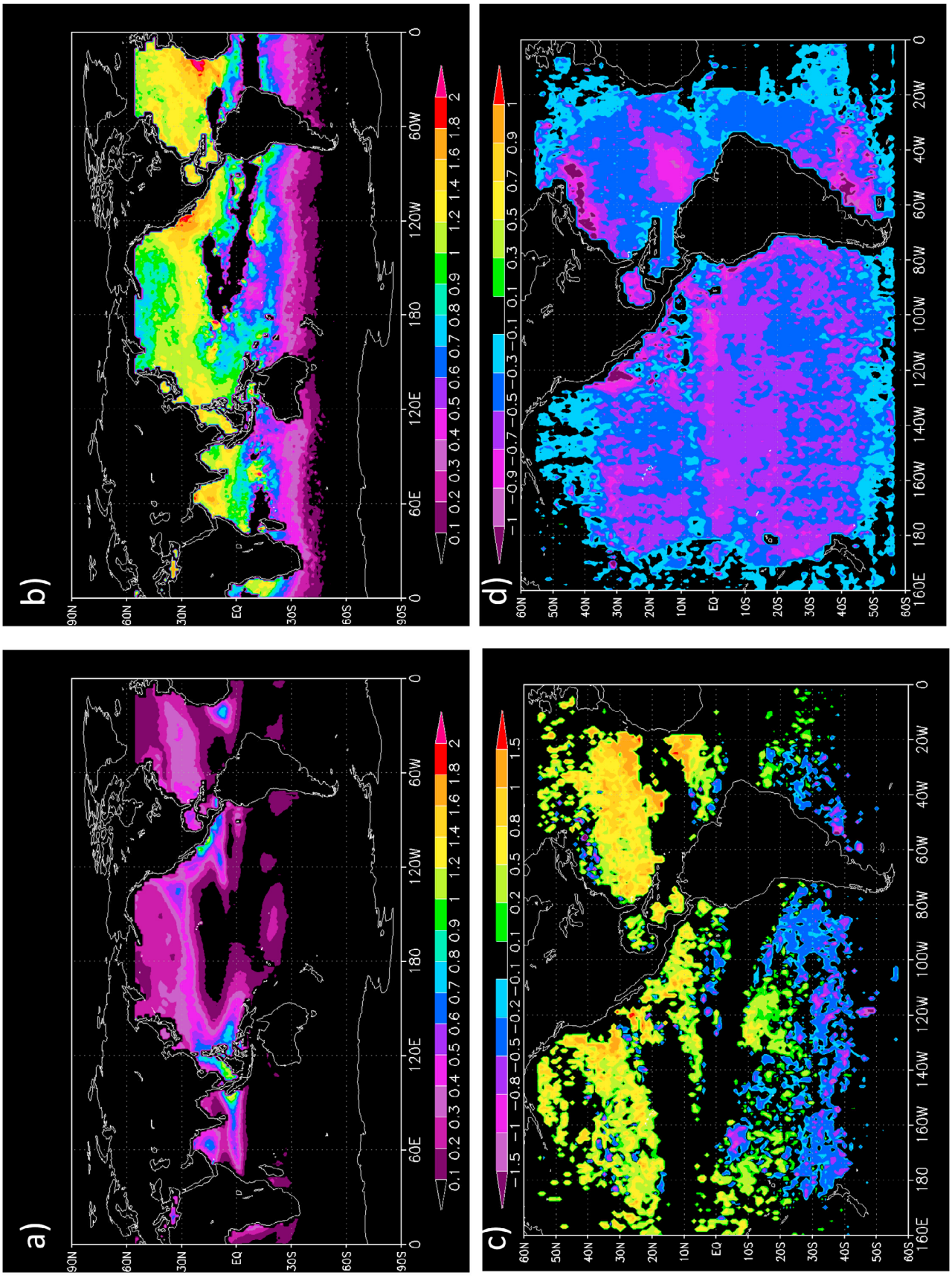
a)
b)
c)
d)
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.2 1.4 1.6 1.8 2
−1 −0.9 −0.7 −0.5 −0.3 −0.1 0.1 0.3 0.5 0.7 0.9 1
−1.5 −1 −0.8 −0.5 −0.2 −0.1 0.1 0.2 0.5 0.8 1 1.5
60N
50N
40N
30N
20N
10N
EQ
10S
20S
30S
40S
50S
60S
160E
180
160W
140W
120W
100W
80W
60W
40W
20W
0

Figure 6

not much different than that in the control. However, the relative prominence of the early morning peak is again diminished in the SSTidf ensemble. The timing of the early morning peak is three hours too late in both the control and SSTidf ensembles. Considering the western Pacific Ocean compositing region (Figure 5d), one sees again that the control ensemble fails to represent the midday precipitation peak whereas the SSTidf ensemble is able to capture it. The early morning peak is slightly diminished in prominence in the SSTidf ensemble, but to much less of an extent than for the Indian Ocean or South China Sea regions. The timing of the early morning peak is unchanged in the SSTidf ensemble as compared to the control. For the eastern Pacific Ocean compositing region (Figure 5e), the afternoon precipitation peak is more prominent in the SSTidf ensemble as compared to the control while, conversely, the early morning peak in the SSTidf ensemble is less pronounced. Thus, the SSTidf ensemble better represents both the afternoon and early morning peaks of the daily cycle, unlike the cases for the Indian Ocean and western Pacific Ocean compositing regions. Nonetheless, the timing of the early morning peak remains three hours too late in the SSTidf ensemble. Last, considering the Atlantic Ocean compositing region (Figure 5f), one finds not much difference between the daily cycles of the SSTidf and control ensembles. The timing of the early morning peak is six hours too late in both ensembles, and both ensembles fail to capture the early afternoon and late evening peaks. Even so, the SSTidf ensemble is arguably more realistic than the control because it suggests the occurrence of more precipitation during the afternoon and evening timeframe. The improvement noted above in the midday precipitation peak is consistent with the notion that the SST diurnal variation with its attendant higher peak daily SST and higher latent and sensible heat fluxes enhances the NWP model's ability to trigger atmospheric convection. On the other hand, the SST diurnal variation's suppression of the early morning precipitation peak in certain regions such as the South China Sea is not so readily interpreted, and will require further work to understand.

[36] Composite daily cycles of precipitation also were separately produced for those cases where the daily accumulated precipitation amount was below median, above median, below third quartile, and above third quartile (not shown). The median and third quartile were calculated for each respective data set and region, e.g., regionally varying medians and third quartiles were calculated specific to each of the control, SSTidf, and TRMM data sets. All of these composites yield the same general results as above, with minor case to case variation. Additionally, the composite daily cycle of precipitation for each of these cases was replicated without the normalization of each sequence in the composite by its maximum value. These composites again yield the same general results as above.

#### 3.1.3. Diurnal SST Amplitude Conditioned on Wind Speed

[37] The diurnal cycle that is observed in skin-layer SST is a sensitive function of wind speed, since strong winds will induce mixing that acts to reduce thermal stratification in the upper ocean [*Gentemann et al.*, 2003; *Kawai and Wada*, 2007]. Existing literature suggests that SST forecasts tend to have some difficulty in representing this function [e.g., *Takaya et al.*, 2010a]. Thus, here composites of the DSA for ensemble SSTidf are constructed both with and without conditioning upon wind speed. The resulting composites are then compared to composite observations of DSA derived from raw satellite data. For the conditioning upon wind speed, three different regimes of daily average wind speed are defined: Low wind speed (0–3 $ms^{-1}$), intermediate wind speed (3–5 $ms^{-1}$) , and high wind speed (>5 $ms^{-1}$). The DSA for a given location and 24-h interval is calculated as the maximum value of SST at that location within the 24-h interval minus the minimum value of SST at that location within the 24-h interval.

[38] The calculation of composite forecast DSA proceeds as follows. A wind speed regime, a 24-h forecast interval (e.g., T + 120 h to T + 144 h), and a particular grid point on a $1^\circ \times 1^\circ$ latitude-longitude grid are chosen. Then, for each 00 UTC forecast initialization in the test period and each ensemble member, a check is made to ensure that the average over the forecast interval of the ensemble member's 10-m wind speed meets the specified wind speed criterion (0–3 $ms^{-1}$ or >5 $ms^{-1}$). If the wind speed meets the criterion, the DSA is calculated and collected. This process is repeated for all 00 UTC forecast initializations and all ensemble members, and the resulting collection of DSA values are averaged to produce the composite at the chosen grid point.

[39] Figure 6a shows the unconditioned composite forecast DSA for ensemble SSTidf for the interval from T + 120 h to T + 144 h. The unconditioned forecast DSA exhibits a maximum in the eastern Indian Ocean that is on average 1.2–1.4°C. Slightly weaker maxima of ≈1°C are seen in the West Pacific warm pool and along the west coast of Mexico. Unconditioned forecast DSA greater than 0.2°C is common in many other parts of the NH.

[40] The forecast DSA conditional upon the low wind speed regime for the interval from T + 120 h to T + 144 h (Figure 6b) is uniformly larger than the unconditioned forecast DSA, as expected. Maximum conditional DSA is 1.8°C or more in places in the Subtropics and lower Midlatitudes, including most prominently the lower-latitude portions of the eastern North Pacific and North Atlantic Oceans. Values greater than 1°C are common elsewhere in the NH.

**Figure 6.** (a) Composite forecast DSA (°C) for the T + 120 h to T + 144 h interval from the SSTidf ensemble, unconditional upon wind speed regime. (b) As for Figure 6a but conditional upon the 0–3 $ms^{-1}$ wind speed regime. Black shading denotes regions with no realizations of forecast DSA for this wind regime. (c) Difference between the composite forecast DSA for the T + 108 h to T + 132 h forecast interval and the composite observed DSA for the low wind speed regime (0–3 $ms^{-1}$). Black shading denotes regions with no realizations of forecast DSA and/or no observations of DSA for this wind regime. (d) As for Figure 6c but for the high wind speed regime (>5 $ms^{-1}$).

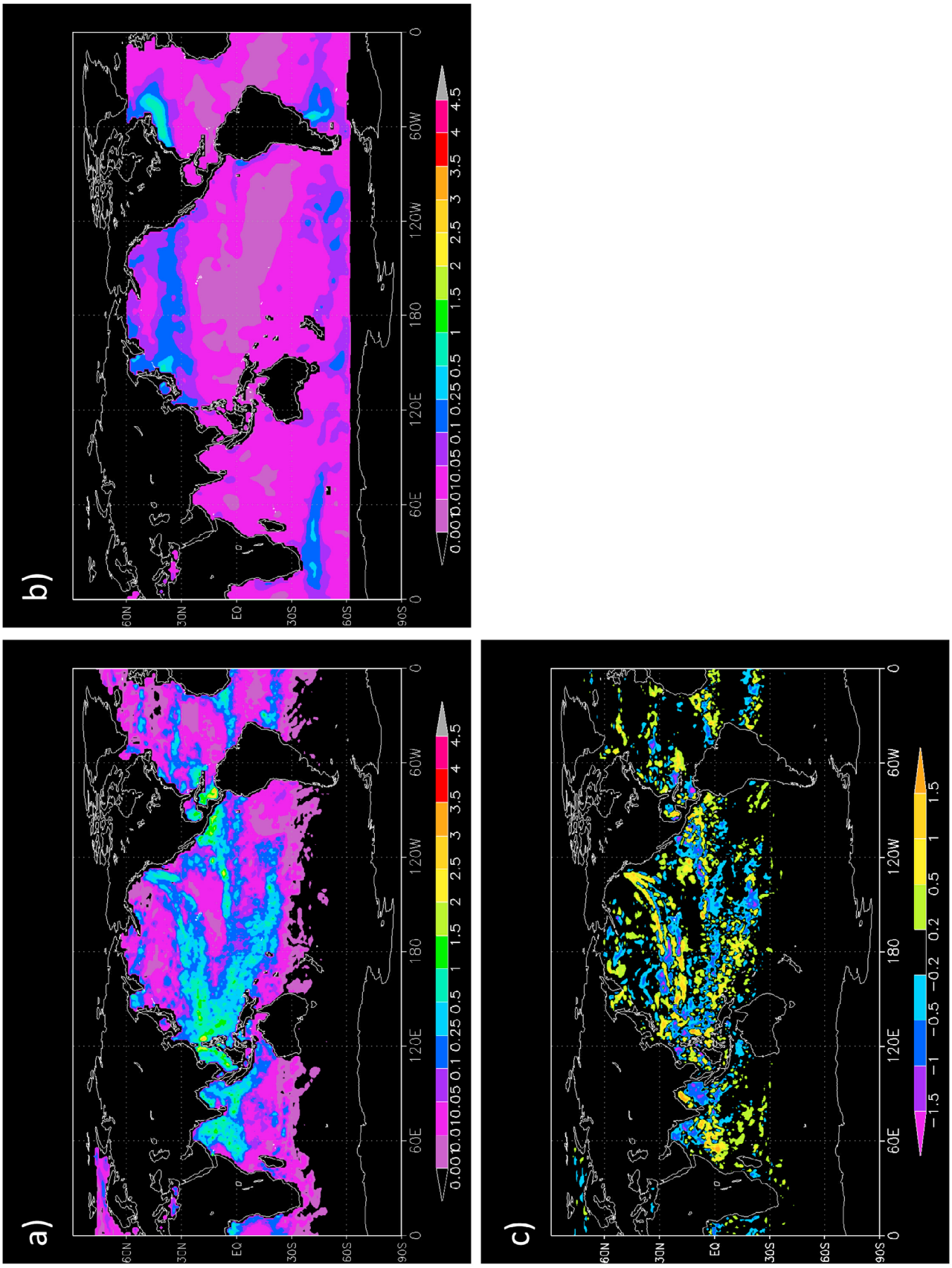
a)
b)
c)
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0.00 0.01 0.05 0.1 0.25 0.5 1 1.5 2 2.5 3 3.5 4 4.5
−1.5 −1 −0.5 −0.2 0.2 0.5 1 1.5

Figure 7

[41] To validate the forecast DSA, observations of DSA were derived from GOES-East and GOES-West raw satellite observations of SST and composited conditional upon analyzed wind speed regime. The GOES raw satellite observations of SST were obtained from the NCODA data assimilation system for each date in the test period. Since the footprint of the GOES observations encompasses the region roughly from 180°W to 20°W, the calculation of composite observed DSA was restricted to this region. Calculation of the composite observed DSA proceeds as follows. For a given date, all the satellite SST observations within a 24-h interval ending at 12 UTC of that date are interpolated to a $1° \times 1°$ latitude-longitude grid using the nearest-neighbor method. Next, a wind speed regime and a particular grid point are chosen, and a check is made to ensure that the average over the 24-h interval of the analyzed 10-m wind speed at the grid point meets the specified wind speed criterion. If the wind speed meets the criterion, the DSA is calculated from the interpolated satellite SST observations. This process is repeated for all dates, and the resulting collection of DSA observations are averaged to produce the composite at the chosen grid point.

[42] Figures 6c and 6d show the difference between the composite forecast DSA for the T + 108 h to T + 132 h forecast interval and the composite observed DSA for the low wind speed and high wind speed regimes, respectively. These difference fields are equivalent to the average over the test period of the error in the ensemble-mean forecast DSA for the given wind speed regime. From Figure 6c, it is apparent that in the low wind speed regime the forecast DSA is too large as compared to observations in most parts of the NH extratropical Pacific and Atlantic Oceans. The only notable exception is in the gulf stream current off the northeast coast of the United States. The most prominent excesses of the NH extratropical forecast DSA are found in the Atlantic Ocean, where difference values of 0.5–1.0°C are common. It is not clear what the cause of these excesses is, although one can speculate that it is probably due to excessive solar insolation arising from insufficient clouds, or to insufficient wind-driven mixing of the ocean surface.

[43] For the Tropics in Figure 6c, the forecast DSA also tends to be too large, particularly near the West African coast. However, in the equatorial East Pacific Ocean the forecast DSA is too small. For the SH, the forecast DSA tends to be too large in the Subtropics and too small in the Midlatitudes.

[44] Considering the DSA differences for the high wind speed regime (Figure 6d), it is apparent that the forecast DSA is too small over the entire compositing region. The most notable differences are found in the gulf stream current, in the eastern Pacific Ocean along the southwest coast of North America, and along the Equator in the Pacific Ocean.

[45] Validation of the forecast DSA was also independently performed against observations of DSA derived from Global Drifter Program drifting-buoy observations, with results similar to those above (not shown). Thus, the above results point to the need for future work to fine-tune the forecast DSA. They also raise the possibility that such fine-tuning could yield forecast performance gains beyond those discussed in section 3.2.2.

## 3.2. Diagnostics of Ensemble Properties

### 3.2.1. Ensemble SST Variance

[46] The sample variance of the ensemble SST is important as a measure of the size and spatial structure of the uncertainty in the SST forecast. This section examines the nature of the ensemble SST variance both at the forecast initial time and after the SST has been evolved by the NWP model.

#### 3.2.1.1. Initial-Time Ensemble SST Variance

[47] The T + 000 h ensemble SST variance is theoretically constrained by the NCODA SST analysis error variance, as per equation (1). Figures 7a and 7b respectively show the T + 000 h ensemble SST variance and the NCODA SST analysis error variance for 00 UTC 31 May 2007. From Figure 7a it's apparent that the ensemble has substantial initial-time SST variance throughout the Tropics and Subtropics, as well as certain parts of the NH Midlatitudes. Variance is largest in the West Pacific warm pool region near the Philippines and in the waters surrounding Central America. Comparing the ensemble SST analysis perturbation variance (Figure 7a) with the NCODA SST analysis error variance (Figure 7b), the most obvious relationship is that the ensemble variance is too large in most areas of the Tropics. On the other hand, the peak values of ensemble variance in the NH Midlatitudes are the same as those for the NCODA error variance. The overly large tropical T + 000 h ensemble variance is a known property of the local ET [see *McLay et al.*, 2010, Figure 5c] and follows on theoretical grounds from the large difference between the ensemble size $K$ and the rank of the matrices $\mathbf{P}_a^j$. This excess variance can be mitigated in the local ET method by tuning (i.e., downscaling) the relevant elements of $\mathbf{P}_a^j$, but the series of experiments required for such tuning were considered beyond the scope of this study. It is worth noting that such tuning may be a simple way to obtain performance gains over and above those discussed in this study.

[48] Further illustration of the initial-time ensemble SST is provided by Figure 7c, which shows the SST analysis perturbation of member 12 of the ensemble initialized 00 UTC 31 May 2007. Most evident in Figure 7c is the abundant smaller-scale structure of the analysis perturbation. Also notable is the long perturbation feature that stretches southwestward from Canada to the Central Pacific Ocean. This perturbation feature marks a region of relatively low 10-m wind speed on the equatorward edge of the midlatitude atmospheric frontal zone (not shown), and is also coincident with relatively large NCODA analysis error variance (Figure 7b). The feature highlights the fact that the SST

**Figure 7.** (a) Ensemble SST analysis perturbation variance ($[°C]^2$) for the 00 UTC 31 May 2007 initialization of the SSTidf ensemble. (b) NCODA SST analysis error variance ($[°C]^2$) for 00 UTC 31 May 2007. (c) SST analysis perturbation (°C) of member 12 of the SSTidf ensemble initialized 00 UTC 31 May 2007.

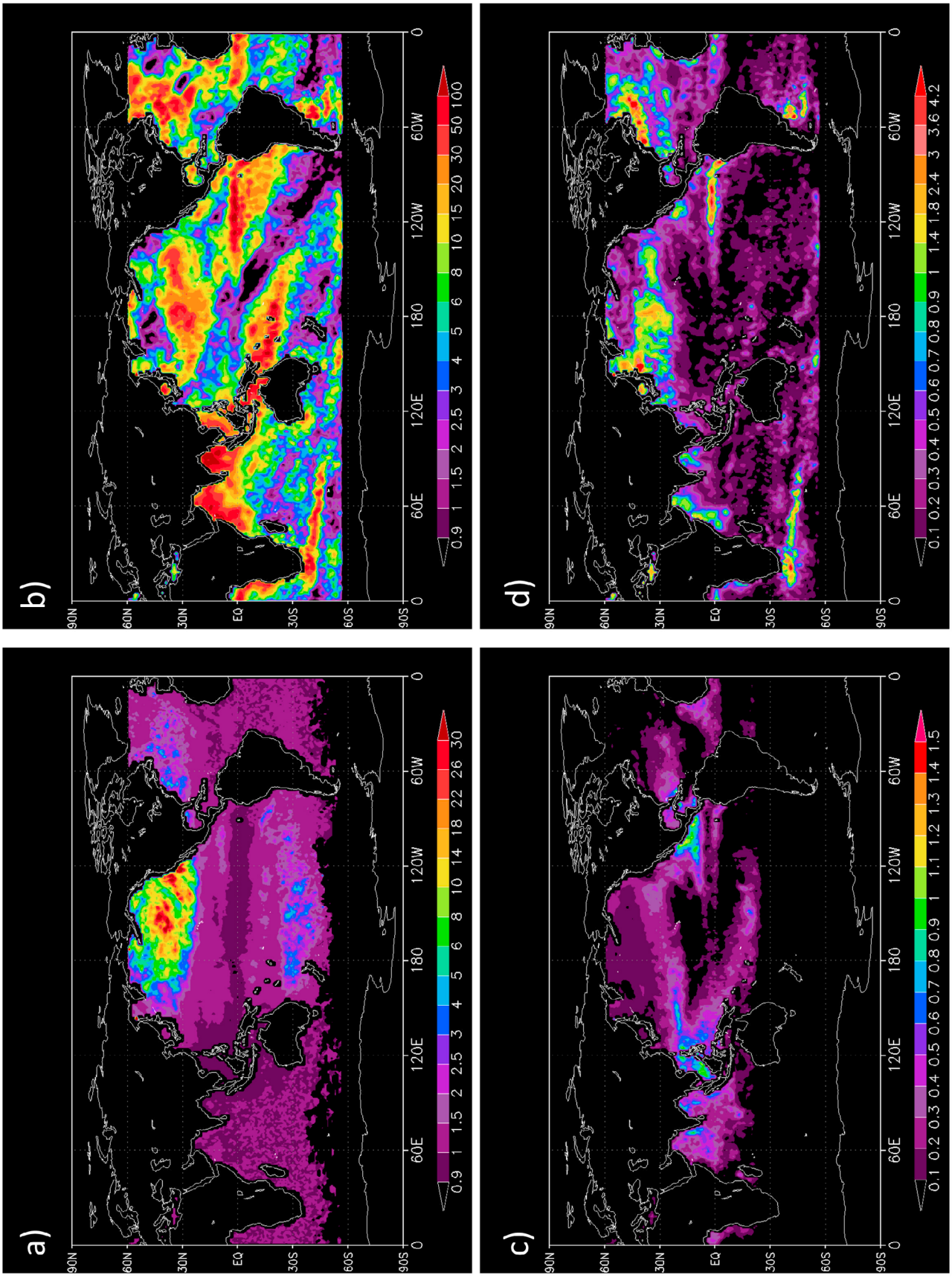
a)
b)
c)
d)
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0.9 1 1.5 2 2.5 3 4 5 6 8 10 14 18 22 26 30
0.9 1 1.5 2 2.5 3 4 5 6 8 10 15 20 30 50 100
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.4 1.8 2.4 3 3.6 4.2

Figure 8

perturbations from the ET method have structural relationships with the atmospheric flow.

#### 3.2.1.2. Evolved Ensemble SST Variance

[49] When a nonlinear, chaotic system is evolved from imperfect initial conditions, the forecast uncertainty as measured by the variance is expected to display bounded growth over some sufficiently long evolution period. The mechanisms of ensemble SST variance growth for the present system are quite limited, deriving only from the interaction between winds, surface fluxes, radiative fluxes, and SST within the model of SST diurnal variation. Nevertheless, some ensemble SST variance growth still arises, as Figure 8a helps to illustrate. Figure 8a specifically shows the amplification factor of the ensemble SST variance over the interval T + 000 h to T + 240 h, averaged over all 00 UTC forecast-ensemble initializations. The amplification factor of the ensemble SST variance is defined as the ratio between the variance at T + 240 h and the variance at T + 000 h. One notes some variance growth across most of the ocean basins, although the most prominent growth is clearly confined to the central and eastern North Pacific Ocean. The growth that occurs in the Tropics is weak and is generally described by an amplification factor of 1 to 1.5. Additionally, there is weak decay of variance (proportionally less than 10%) in the tropical Pacific Ocean in the immediate vicinity of the Equator, as well as in the West Pacific warm pool region. For comparison, Figure 8b presents the amplification factor of the SST forecast error variance over the interval T + 000 h to T + 240 h, where the amplification factor of the SST forecast error variance is defined as the ratio between the SST forecast error variance at T + 240 h and the average NCODA SST analysis error variance. The SST forecast error variance is estimated as the variance of a sample of errors in the ensemble-mean SST forecast. The sample of errors is composed by first calculating the forecast error for each 00 UTC initialization of the forecast ensemble in the period 01 May 2007 to 30 June 2007 and then pooling together the forecast errors from all initializations. Comparing Figures 8a and 8b one finds that the ensemble's amplification factor approaches that of the forecast error in the central and eastern North Pacific Ocean, but that elsewhere the ensemble's amplification factor is considerably smaller. The difference between the two is particularly pronounced throughout most of the Tropics. Some large differences are to be expected given the limited mechanisms of ensemble SST variance growth for the present system. However, it's also possible that NCODA underestimates the SST analysis error variance in certain regions, thereby inflating the forecast error amplification factor.

[50] Figures 8c and 8d show the evolved, T + 240 h ensemble SST variance and the T + 240 h SST forecast error variance, respectively. The ensemble SST variance is averaged over all 00 UTC forecast-ensemble initializations in the period 1 May 2007–30 June 2007. The SST forecast error variance is calculated as described above for Figure 8b. Figure 8c shows that the ensemble maintains SST variance in all of the tropical oceans as well as throughout the North Pacific Ocean, the Gulf of Mexico, and the southern and western parts of the North Atlantic Ocean. The most pronounced variance is in the West Pacific warm pool region and near the west coast of Mexico and Central America. Comparing Figures 8c and 8d, one sees, not surprisingly, that the ensemble variance is much smaller than the forecast error variance over much of the North Pacific Ocean, in the western Indian Ocean, in the vicinity of the Gulf Stream current in the North Atlantic Ocean, and over the Equatorial East Pacific Ocean. On the other hand, the ensemble SST variance is somewhat excessive in the West Pacific warm pool region and in the East Pacific Ocean near Central America.

### 3.2.2. Ensemble Verification Metrics

[51] Figures 9a–9c summarize the performance of the SSTidf forecast ensemble in relation to the control ensemble at T + 120 h, T + 240 h and T + 336 h, respectively, in terms of four metrics including (1) the bias of the ensemble-mean forecast, (2) the root-mean square (RMS) error of the ensemble-mean forecast with bias removed prior to calculation, (3) a variance metric that measures the absolute difference between the forecast ensemble variance and an estimate of the observed forecast error variance, and (4) the Continuous Ranked Probability Score (CRPS). For the variance metric, the observed forecast error variance is estimated as the variance of a sample of errors in the ensemble-mean forecast. The sample of errors is composed by first calculating the forecast error for each 00 UTC initialization of the forecast ensemble in the period 1 May 2007 to 30 June 2007 and then pooling together the forecast errors from all initializations. The CRPS metric is a measure of the integrated difference between the cumulative distribution function (CDF) of the forecast ensemble and the CDF of the verifying analysis [*Wilks*, 2006]. The above four metrics were calculated for five different variables [surface temperature (which includes both land-surface temperature and SST), 2-m air temperature, 500 hPa geopotential height, 10-m wind speed, and 24-h accumulated precipitation] and thirteen different regions. For a given metric, variable, and region, the calculations were carried out for each 00 UTC forecast ensemble initialization in the period 1 May 2007 to 30 June 2007. Then, the difference between the score for SSTidf and the corresponding score for the control was calculated for each initialization. Finally, the differences in score for all initializations were averaged. A statistical significance test of the average difference in score between the SSTidf forecast ensemble and the control ensemble for a given metric, variable, and region was carried out using a method that accounts for serial dependence of data [*Wilks*, 2006, pp. 143–145]. Cases where SSTidf scores significantly better (worse) than the control are colored green (red), and cases where no significant difference in score is found are colored light grey.

**Figure 8.** (a) Average amplification factor of the ensemble SST variance over the interval T + 000 h to T + 240 h. (b) Amplification factor of the SST forecast error variance over the interval T + 000 h to T + 240 h. (c) Average ensemble SST variance ($[^\circ C]^2$) at T + 240 h for ensemble SSTidf. (d) SST forecast error variance ($[^\circ C]^2$) at T + 240 h for ensemble SSTidf.

In the case of 24-h accumulated precipitation, cross-hatching denotes regions for which the verification metrics were not calculated because the TRMM data set is only available for the 50°N latitude band.

[52] Considering Figures 9a–9c together, it is clear that SSTidf scores significantly better than the control in the overwhelming majority of cases for the tropical regions. Particularly notable is the consistent improvement gained by SSTidf for tropical regions in terms of the CRPS and variance metrics. SSTidf also gains a considerable advantage for tropical regions in terms of RMSE for the variables 10-m wind speed and 24-h accumulated precipitation. The improvement gained by SSTidf is not only positive and statistically significant in a large number of cases, it is often nontrivial in terms of magnitude. This is illustrated by Figure 10, which compares the SSTidf and control ensembles in terms of the forecast variance of 24 h accumulated precipitation in the Tropics (Figure 10a) and the CRPS of 10-m wind speed in the tropical Indian Ocean (Figure 10b). The SSTidf ensemble yields gains equivalent to several days or more of forecast skill at longer lead times in both Figures 10a and 10b. For example, in Figure 10b the CRPS of SSTidf at T + 240 h is roughly equivalent to that of the control at T + 168 h.

[53] Considering again Figures 9a–9c, one finds that SSTidf has much less of an impact in the Midlatitudes, and particularly very little impact on midlatitude 500 hPa geopotential height. This result is similar to that of *Takaya et al.* [2010a]. Nonetheless, in those cases where SSTidf does have an impact in the Midlatitudes, the cases with positive impacts outnumber the cases with negative impacts. The overall least impact is to the SH Midlatitudes, which is to be expected given the high wind speeds over the southern oceans. There are only a few cases where SSTidf scores persistently worse than the control, these being namely the bias of surface temperature in the southern Pacific Ocean, the North Atlantic Ocean, the tropical eastern Pacific Ocean, and tropical Atlantic Ocean, and the RMSE of the surface temperature in the tropical Atlantic Ocean. At T + 120 h (Figure 9a), SSTidf exhibits worse RMSE of surface temperature in the tropical latitudes closest to the Equator, but there is no significant effect at the later lead times. SSTidf also exhibits worse RMSE and CRPS for 500 hPa geopotential height over the North Atlantic Ocean at the intermediate lead time of T + 240 h (Figure 9b), but there is no similar indication at the earlier lead time of T + 120 h (Figure 9a) nor at the later time of T + 336 h (Figure 9c). Some of these cases might be remedied through the systematic retuning of other NOGAPS physical parameterizations such as the boundary layer and surface-flux schemes, something that was not attempted for this study. Also, the overall favorable results of Figure 9 suggest that the statistically significant changes to ensemble-mean physical properties that were discussed in section 3.1.1 are in fact of practical significance as well.

[54] The ensemble diagnostics were further stratified by verifying the SSTidf and control forecast ensembles only at ocean grid points (Figure 11a) and only at land grid points (Figure 11b). For brevity only the results for T + 240 h are shown. Also, the land-point diagnostics were restricted to just five of the 13 regions, because the other seven regions encompassed very few land grid points. The ocean-grid point results are largely similar to the full-grid results presented in Figure 9. They show the general superiority of SSTidf in tropical regions, and the lesser impact of SSTidf in the Midlatitudes. However, they reveal that SSTidf degrades the SST bias in the Tropics, mainly due to the forecasts for the tropical eastern Pacific Ocean. The land-grid point results (Figure 11b) also are overall favorable and much the same as the full-grid results in Figure 9. One new detail that emerges is that SSTidf degrades the over-land tropical 2-m air temperature bias and hence the over-land tropical 2-m air temperature CRPS. Nonetheless, the results also show that SSTidf is not significantly worse in terms of any metric for the over-land NH Midlatitudes, in contrast to the over-ocean case.

[55] The test period of the study (1 May 2007 to 30 June 2007) encompassed one complete cycle of the Madden-Julian Oscillation (MJO). As a final diagnostic, this episode of the MJO was evaluated using the *Wheeler and Hendon* [2004] Real Time Multivariate MJO index (RMM). The RMM index provides information on the geographic location and strength of the MJO signal. No meaningful difference was found between the RMM index from the SSTidf ensemble and that from the control ensemble in terms of either phase or amplitude of the MJO signal (not shown).

### 3.3. Role of SST Initial Perturbations

[56] All the results described thus far pertain to forecast SSTidf, where the forecast ensemble is generated using the diurnal SST model in conjunction with SST analysis perturbations produced by the local ET. Still to be determined is

**Figure 9.** (a) Performance of the SSTidf ensemble relative to the control ensemble at T + 120 h. Varying across the left side of the figure are the metrics [BIAS, RMSE, VAR, CRPS] and the variables [surface temperature (sfctmp) (°C), 2-m air temperature (airtmp2m) (°C), 500 hPa geopotential height (geopht500) (m), 10-m wind speed (10 $m_{wind}$) ($ms^{-1}$), 24-h accumulated precipitation (24 h precip) (mm)]. Varying across the top of the figure are the regions [NM (NH Midlatitudes, 30°N–65°N, 0°E–360°E), SM (SH Midlatitudes, 30°S–65°S, 0°E–360°E), SI (southern Indian Ocean, 60°S–30°S, 0°E–120°E), SPAC (South Pacific Ocean, 60°S–30°S, 180°W–90°W), NPAC (North Pacific Ocean, 30°N–60°N, 130°E–120°W), NATL (North Atlantic Ocean, 30°N–70°N, 70°W–0°E), T3 (tropical band 3, 30°S–30°N, 0°E–360°E), T2 (tropical band 2, 20°S–20°N, 0°E–360°E), T1 (tropical band 1, 10°S–10°N, 0°E–360°E), TI (tropical Indian Ocean, 20°S–20°N, 40°E–110°E), TWP (tropical western Pacific Ocean, 20°S–20°N, 110°E–160°W), TEP (tropical eastern Pacific Ocean, 20°S–20°N, 160°W–80°W), TATL (tropical Atlantic Ocean, 20°S–20°N, 70°W–0°E)]. Cases where SSTidf scores significantly better (worse) than the control are colored green (red), and cases where no significant difference in score is found are colored light grey. In the case of 24-h accumulated precipitation, cross-hatching denotes regions for which the verification metrics were not calculated. See text for further details. (b) As for Figure 9a, but at T + 240 h. (c) As for Figure 9a, but at T + 336 h.

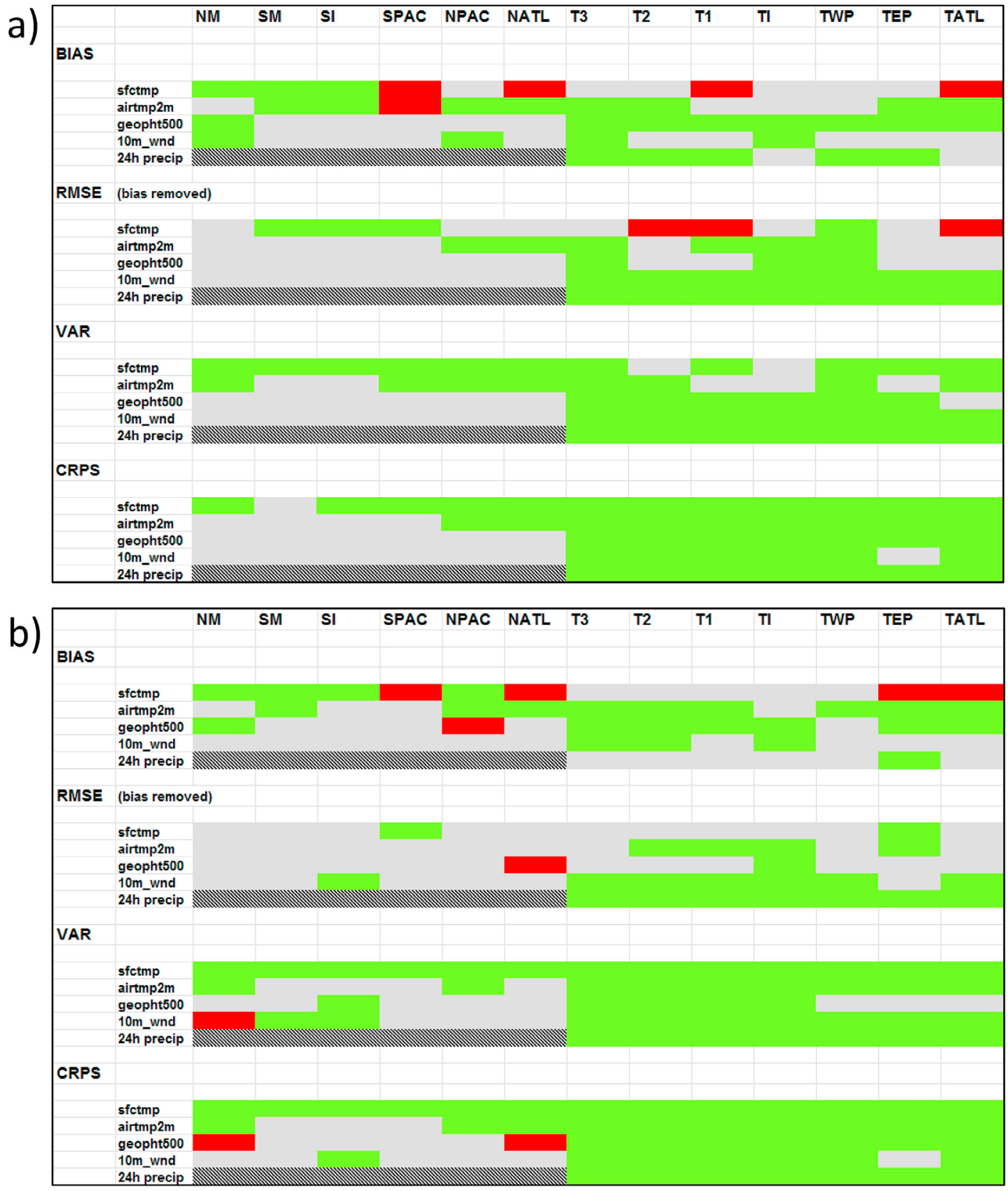
a)
NM SM SI SPAC NPAC NATL T3 T2 T1 TI TWP TEP TATL
BIAS
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
RMSE
(bias removed)
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
VAR
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
CRPS
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
b)
NM SM SI SPAC NPAC NATL T3 T2 T1 TI TWP TEP TATL
BIAS
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
RMSE
(bias removed)
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
VAR
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip
CRPS
sfctmp
airtmp2m
geopht500
10m_wnd
24h precip

**Figure 9**

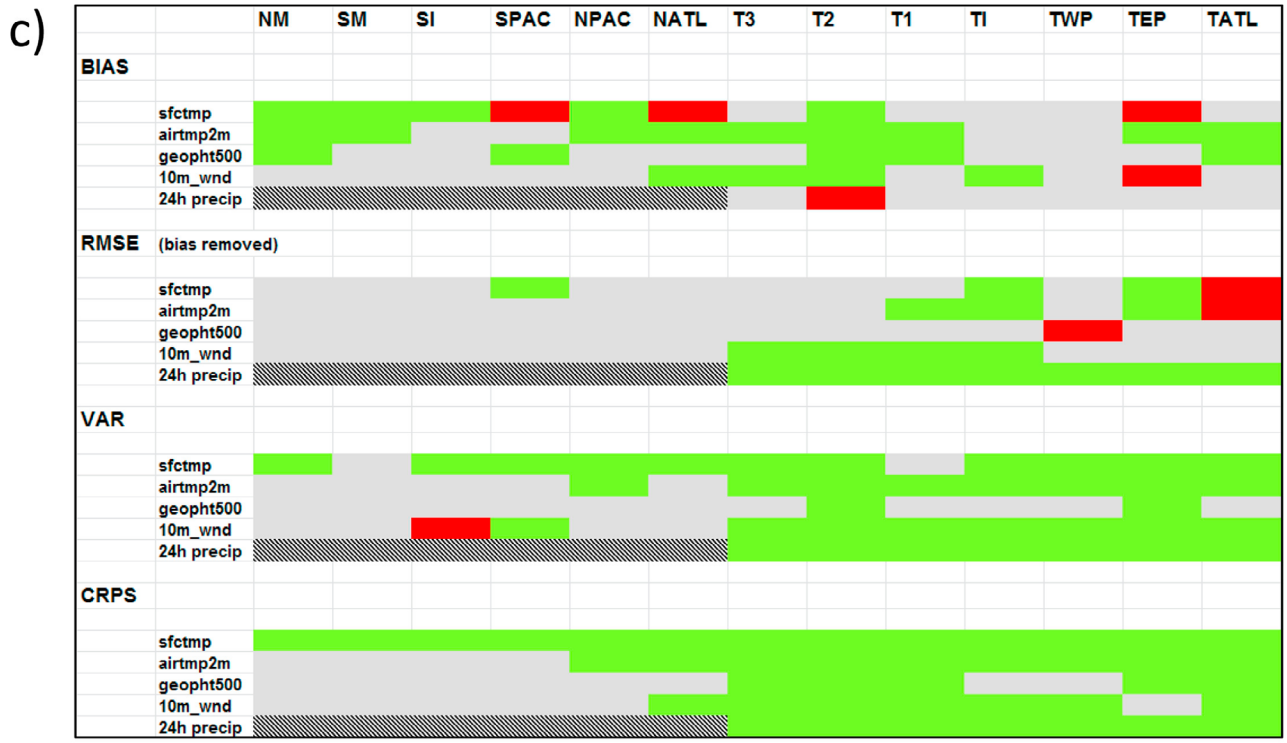


**Figure 9.** (continued)

the effect of the diurnal SST model in isolation from the SST analysis perturbations and, vice versa, the effect of the SST analysis perturbations in isolation from the diurnal SST model. Here, such effects are examined through comparison of the three forecasts SSTdf, SSTi, and SSTidf.

[57] Figure 12a summarizes the performance of the SSTdf forecast ensemble in relation to the control ensemble at T + 240 h for ocean grid points. Overall, one can infer that the diurnal forcing is beneficial to the forecast, since the cases where SSTdf is improved relative to the control outnumber those where it is degraded. However, the number of cases with improvement is not overwhelming, in contrast to the results for SSTidf (Figure 11a).

[58] Figure 12b summarizes the performance of the SSTi forecast ensemble in relation to the control ensemble at T + 240 h for ocean grid points. Considering the results for the tropical regions, it's apparent that the SST analysis perturbations are beneficial to the forecast; SSTi wins in the vast majority of cases, and particularly for 10-m wind speed and 24-h accumulated precipitation. On the other hand, SSTi engenders some adverse effects versus the control in the NH Midlatitudes.

[59] Finally, a comparison is made between the forecasts from SSTidf and SSTdf in Figure 13. From this figure, it is clear that SSTidf is superior to SSTdf for the tropical regions. Thus, inclusion of the SST analysis perturbations provides a clear benefit to forecasts for the Tropics over and above that of diurnal SST variation alone. Inclusion of the SST analysis perturbations is particularly beneficial to tropical 10-m wind speed and tropical 24-h accumulated precipitation. Considering the midlatitude regions in Figure 13, one finds that inclusion of the SST analysis perturbations provides considerably fewer gains. The few cases where SSTidf performs significantly worse are associated with the North Atlantic region and with the bias of SST in the Tropics.

## 4. Conclusions

[60] A modified version of the *Takaya et al.* [2010b] prognostic diurnal SST scheme is incorporated into the NOGAPS global spectral model. Additionally, the local-ET analysis perturbation scheme is adapted to generate analysis perturbations to both atmospheric variables and SST. The NOGAPS model with diurnal SST and the ET analysis perturbations are used in conjunction to generate 336 h lead time, 32-member forecast ensembles for each 00 UTC analysis time in the May–June 2007 period. The resulting forecast ensembles are diagnosed both in terms of various physical properties and in terms of forecast performance. Physical properties that are examined include the nature of surface fluxes, OLR, and near-surface atmospheric variables, the diurnal cycle of precipitation, and the DSA. Performance metrics that are examined include bias and RMS error of the ensemble mean, ensemble variance, and CRPS.

[61] The SST variation is found to introduce statistically significant changes to the ensemble-mean state of all of the physical variables that are analyzed. The changes to the ensemble-mean state likely are introduced by two principal sources, one being the nonlinear evolution of the ensemble

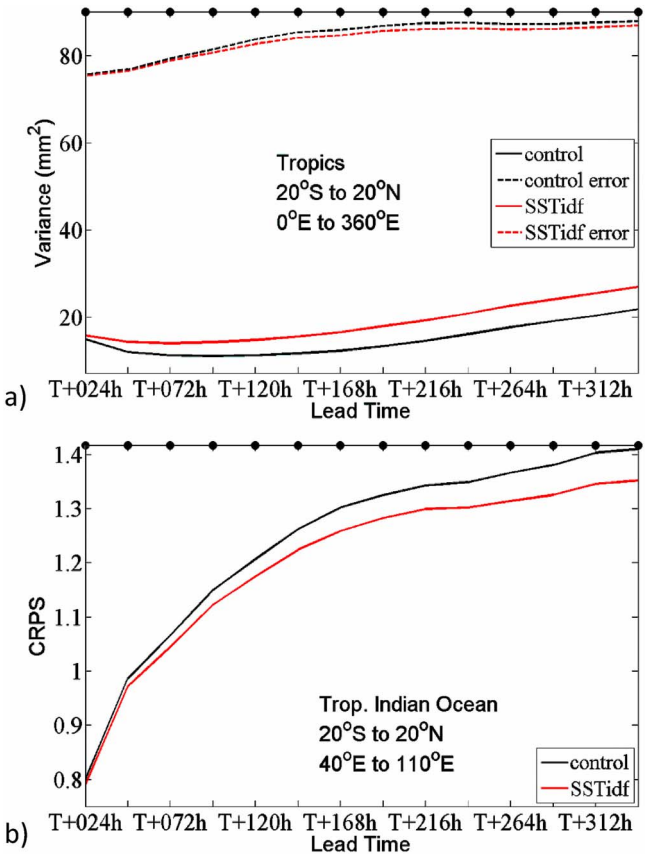


**Figure 10.** (a) Average variance of 24-h accumulated precipitation for the SSTidf ensemble (solid red line) and the control ensemble (solid black line) for the Tropics [0°E to 360°E, 20°S to 20°N]. Observed forecast error variance for the SSTidf (control) ensemble given by the dashed red (black) lines. Filled black (grey) circles indicate SSTidf provides statistically significant (insignificant) improvement relative to the control at the corresponding lead times. (b) CRPS of 10-m wind speed for the SSTidf ensemble (red line) and control ensemble (black line) for the tropical Indian Ocean [40°E to 110°E, 20°S to 20°N]. Meaning of filled circles as for Figure 10a.

perturbations by the NWP model and the other being a slight increase in the daily average SST owing to the chosen configuration of the model of SST diurnal variation. The changes are most prevalent in the tropical regions, but in some cases extend to the higher latitudes of the NH. Also, the SST variation leads to a midday peak in the diurnal cycle of precipitation in low-wind regimes in both the Indian Ocean and western Pacific Ocean. This midday peak is more consistent with observations and is not found in a control ensemble that lacks all SST variation (the control was generated with a prescribed SST field that was held constant throughout integration). Composites of the forecast DSA show a direct relationship with wind speed, as expected, but the forecast DSA is often too large (small) in regimes of low (high) wind speed.

[62] Very favorable results are gained from the SST variation in terms of ensemble performance metrics. The forecasts with SST variation maintain appreciable ensemble SST variance in both the Tropics and the Midlatitudes throughout the 336 h forecast interval. More notably, these forecasts are overall superior in the tropical regions when compared to forecasts from the control ensemble. Statistically significant gains in forecast performance are seen for all the variables that are analyzed, including surface temperature (both land-surface and SST), 2-m air temperature, 500 hPa geopotential height, 10-m wind speed, and 24-h accumulated precipitation. Furthermore, significant gains are seen in all the ensemble metrics, but particularly in the variance metric and the probabilistic CRPS metric. The broad-based gains demonstrate that the SST variation is important to multiple aspects of the forecast probability distribution, including both the mean and the variance. At longer lead times the magnitude of the gains can be meaningful, in some cases exceeding several days' worth of forecast skill. Stratification of the diagnostics using only land grid points or only ocean grid points leads to similar, favorable results. Inclusion of the SST variation has much less of an impact in the Midlatitudes, but even here the SST variation is beneficial on the whole.

[63] A further comparison addressed the individual impact of the diurnal SST variation and the SST analysis perturbations on the performance of the forecast ensemble. Findings indicate that the diurnal SST variation and the SST analysis perturbations each are overall beneficial to the forecast performance in the tropical regions. However, the performance gains from the SST analysis perturbations are much broader than those from the diurnal SST variation. On the other hand, the SST analysis perturbations lead to some adverse effects in the NH midlatitude regions that are not seen when using only diurnal SST variation. Finally, it is shown that the forecasts generated using both the diurnal SST variation and the SST analysis perturbations are superior in the tropical regions as compared to forecasts generated using SST diurnal variation alone. Thus, the SST analysis perturbations are essential to obtain the best forecast performance.

[64] Certain aspects of the forecast ensemble with SST variation stand to be improved through future work. As underscored in section 2, the model of diurnal SST variation used in this study neglects the cool-skin effect of the *Takaya et al.* [2010b] model. This effect should eventually be incorporated for more realism. Also, some issues arise with the SST analysis perturbations. For instance, inclusion of these perturbations causes small but statistically significant degradation of the 500 hPa geopotential height over the North Atlantic Ocean at lead times near T + 240 h. Further, the variance of the SST analysis perturbations is too large in the Tropics. These issues might be addressed through modifications of the analysis perturbation structure and statistics. For instance, if the negative impacts over the North Atlantic Ocean are not due to teleconnections, then they might be mitigated by constraining the SST variation to lie within the 40 N latitude band in the manner of *Takaya et al.* [2010a]. Additional problems, such as a degradation of over-land 2-m air temperature bias, a reduction of the peak early morning precipitation in the Indian Ocean and South China Sea, and the too-large forecast DSA in low wind speed conditions, also need to be addressed. Some of these problems may be mitigated through systematic retuning of the NOGAPS physical parameterizations.

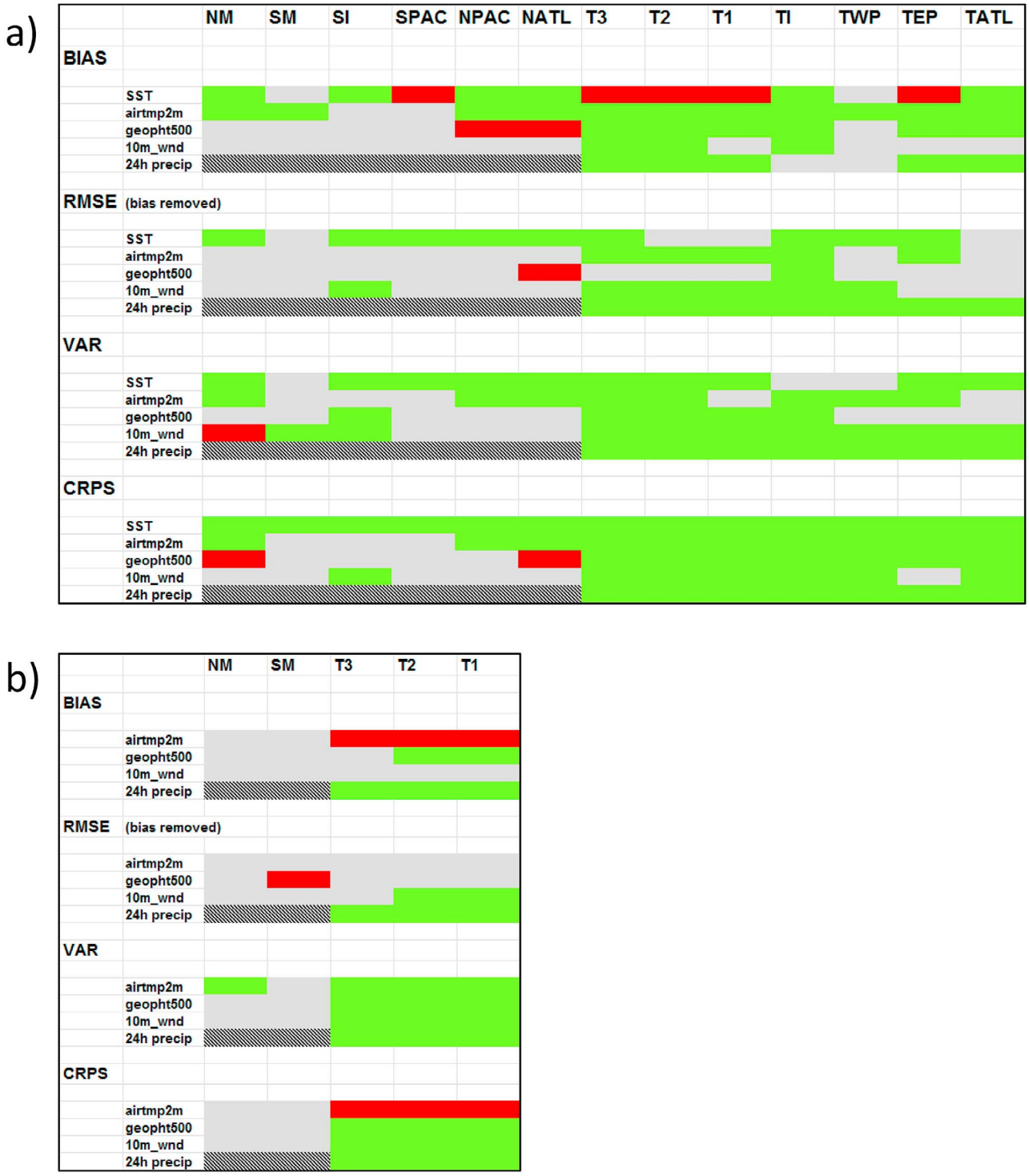


**Figure 11.** (a) Performance of the SSTidf ensemble relative to the control ensemble at T + 240 h for ocean grid points only. Notation as for Figure 9. (b) As for Figure 11a, but for land grid points only.

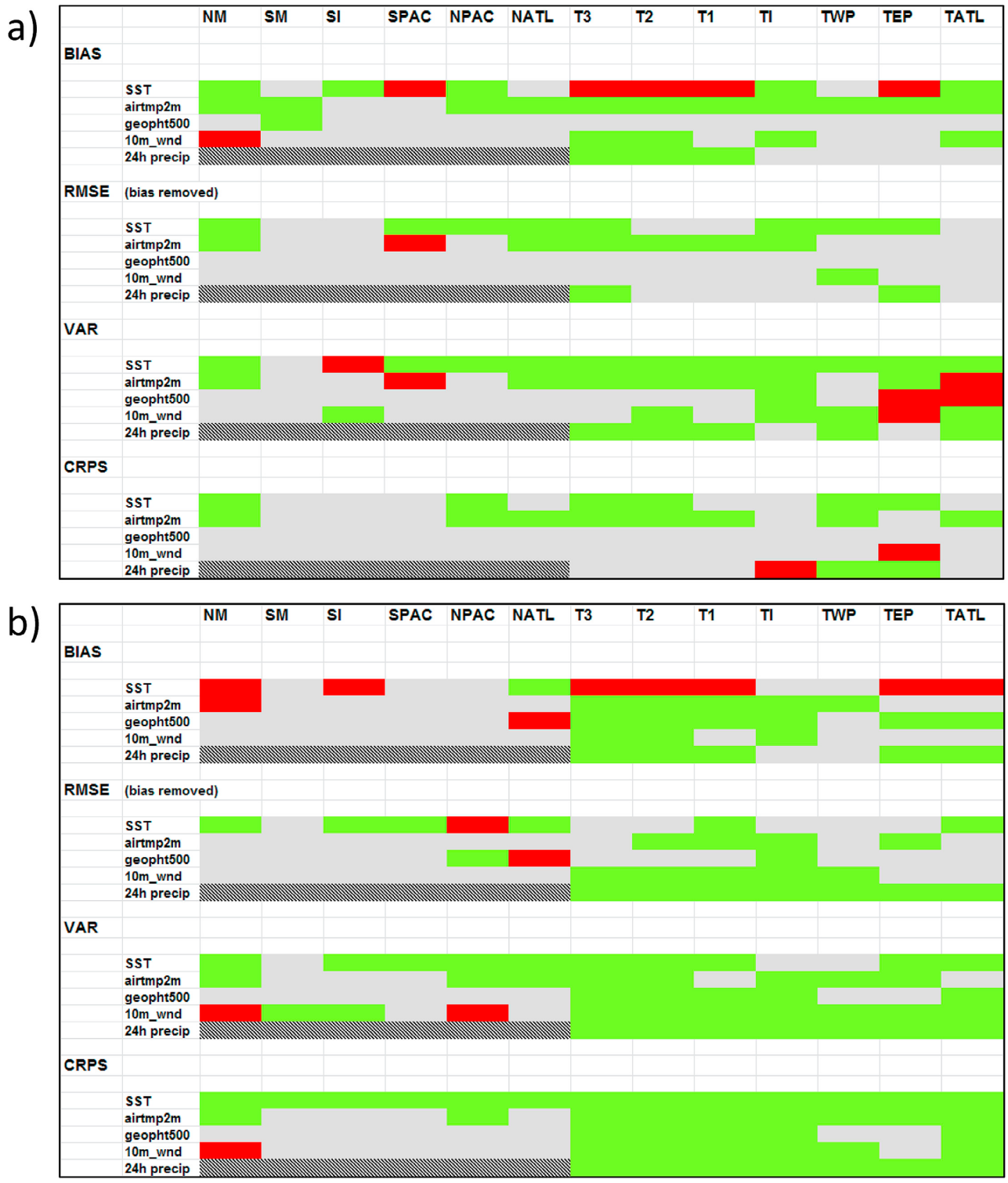


**Figure 12.** (a) Performance of the SSTdf ensemble relative to the control ensemble at T + 240 h for ocean grid points only. Cases where SSTdf scores significantly better (worse) than the control are colored green (red), and cases where no significant difference in score is found are colored light grey. Notation as for Figure 9. (b) Performance of the SSTi ensemble relative to the control ensemble at T + 240 h for ocean grid points only. Cases where SSTi scores significantly better (worse) than the control are colored green (red), and cases where no significant difference in score is found are colored light grey. Notation as for Figure 9.

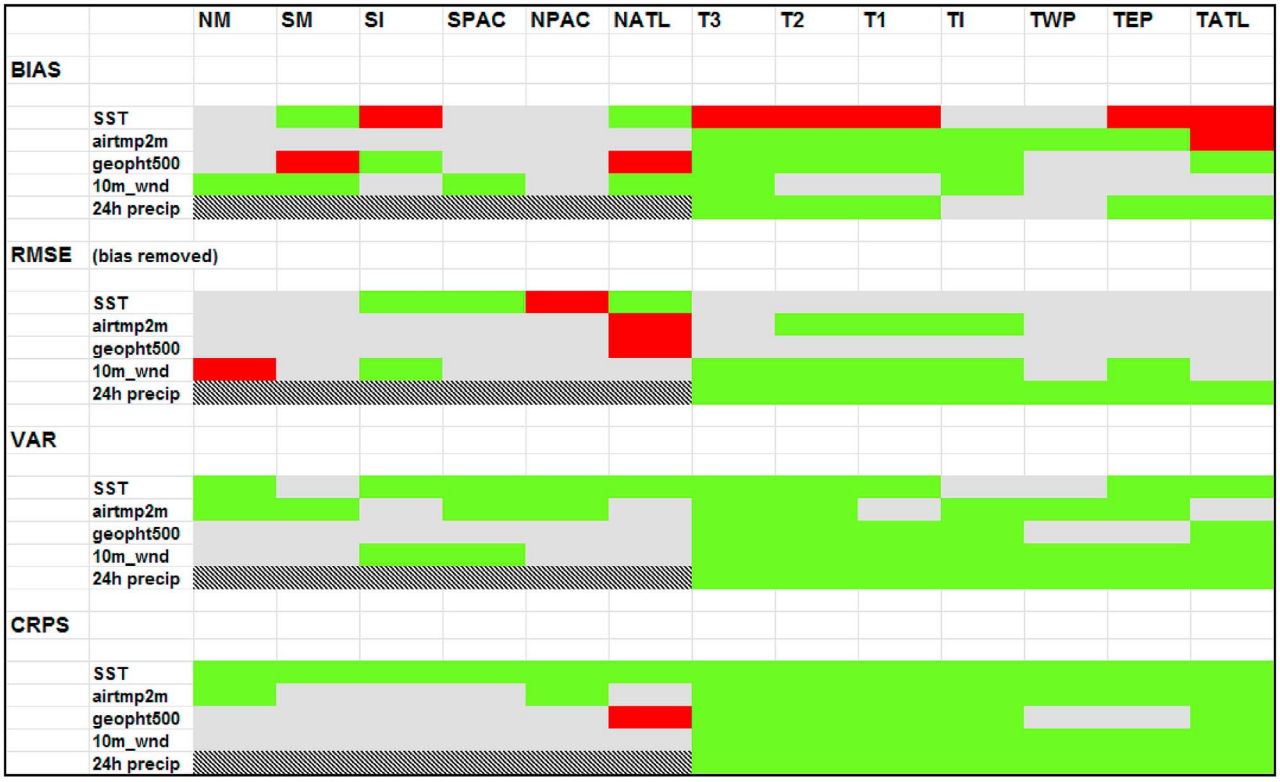


**Figure 13.** Performance of the SSTidf ensemble relative to the SSTdf ensemble at T + 240 h for ocean grid points only. Cases where SSTidf scores significantly better (worse) than SSTdf are colored green (red), and cases where no significant difference in score is found are colored light grey. Notation as for Figure 9.

[65] **Acknowledgments.** The comments of four anonymous reviewers substantially improved this manuscript. This research was supported by the NRL Base Program under program element 0601153 N. The DoD High Performance Computing program at NAVO DSRC provided the computing resources.

## References

Arribas, A., et al. (2011), The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Weather Rev.*, *139*, 1891–1910.

Bellenger, H., and J.-P. Duvel (2009), An analysis of tropical ocean diurnal warm layers, *J. Clim.*, *22*, 3629–3646.

Bellenger, H., Y. N. Takayabu, T. Ushiyama, and K. Yoneyama (2010), Role of diurnal warm layers in the diurnal cycle of convection over the tropical Indian Ocean during MISMO, *Mon. Weather Rev.*, *138*, 2426–2433.

Bowler, N., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts (2009), The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system, *Q. J. R. Meteorol. Soc.*, *135*, 767–776.

Brunke, M. A., X. Zeng, V. Misra, and A. Beljaars (2008), Integration of a prognostic sea surface temperature scheme into weather and climate models, *J. Geophys. Res.*, *113*, D21117, doi:10.1029/2008JD010607.

Buizza, R., R. Hagedorn, L. Isaksen, M. Leutbecher, T. Palmer, M. Steinheimer, and F. Vitart (2010), Recent changes of the ECMWF ensemble prediction system, paper presented at European Meteorological Society Annual Meeting, Eur. Meteorol. Soc., Zurich, Switzerland, 13–17 Sep. [Available at "http://www.emetsoc.org/fileadmin/ems/dokumente/annual-meetings/2010/NWP1-Buizza.pdf".]

Clayson, C. A., and A. Chen (2002), Sensitivity of a coupled single-column model in the tropics to treatment of the interfacial parameterizations, *J. Clim.*, *15*, 1805–1831.

Cummings, J. A. (2005), Operational multivariate ocean data assimilation, *Q. J. R. Meteorol. Soc.*, *131*, 3583–3604

Cummings, J. A., and O. M. Smedstad (2012), Variational data assimilation for the global ocean, in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, Springer, Berlin, in press.

Daley, R., and E. Barker (2001a), NAVDAS Source Book 2001, *Rep. NRL/PU/7530-01-441*, 161 pp., Nav. Res. Lab., Monterey, Calif. [Available online at http://www.nrlmry.navy.mil/sec7531.htm.]

Daley, R., and E. Barker (2001b), NAVDAS: Formulation and diagnostics. *Mon. Weather Rev.*, *129*, 869–883.

Flament, P., J. Firing, M. Sawyer, and C. Trefois (1994), Amplitude and horizontal structure of a large diurnal sea-surface warming event during the Coastal Ocean Dynamics Experiment, *J. Phys. Oceanogr.*, *24*, 124–139.

Gentemann, C. L., C. J. Donlon, A. Stuart-Menteth, and F. J. Wentz (2003), Diurnal signals in satellite sea surface temperature measurements, *Geophys. Res. Lett.*, *30*(3), 1140, doi:10.1029/2002GL016291.

Kawai, Y., and A. Wada (2007), Diurnal sea-surface temperature variation and its impact on the atmosphere and ocean: A review, *J. Oceanogr.*, *63*, 721–744.

Kawai, Y., K. Otsuka, and H. Kawamura (2006), Study on diurnal sea surface warming and a local atmospheric circulation over Mutsu Bay, *J. Meteorol. Soc. Jpn.*, *84*, 725–744.

Kunii, M., and T. Miyoshi (2012), Including uncertainties of sea-surface temperature in an ensemble Kalman filter: A case study of Typhoon Sinlaku (2008), *Weather Forcasting*, doi:10.1175/WAF-D-11-00136.1, in press.

Large, W. G., J. C. McWilliams, and S. C. Doney (1994), Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, *32*, 363–403.

McLay, J., C. H. Bishop, and C. A. Reynolds (2008), Evaluation of the ensemble transform analysis perturbation scheme at NRL, *Mon. Weather Rev.*, *136*, 1093–1108.

McLay, J., C. H. Bishop, and C. A. Reynolds (2010), A local formulation of the ensemble transform (ET) analysis perturbation scheme, *Weather Forecasting*, *25*, 985–993.

Parsons, D. B., K. Yoneyama, and J.-L. Redelsperger (2000), The evolution of the tropical western Pacific atmosphere-ocean system following the arrival of a dry intrusion, *Q. J. R. Meteorol. Soc.*, *126*, 517–548.

Peng, M. S., J. A. Ridout, and T. F. Hogan (2004), Recent modifications of the Emanuel convective scheme in the Navy Operational Global Atmospheric Prediction System, *Mon. Weather Rev.*, *132*, 1254–1268.

Simpson, J., R. F. Adler, and G. R. North (1988), A proposed tropical rainfall measuring mission (TRMM) satellite, *Bull. Am. Meteorol. Soc.*, *69*, 278–295.

Slingo, J. M., P. Inness, R. Neale, S. Woolnough, and G.-Y. Yang (2003), Scale interactions on diurnal to seasonal timescales and their relevance to model systematic errors, *Ann. Geophys.*, *46*, 139–155.

Takaya, Y., F. Vitart, G. Balsamo, M. Balmaseda, M. Leutbecher, and F. Molteni (2010a), Implementation of an ocean mixed layer model in IFS, *ECMWF Tech. Memo. 622*, 34 pp., Eur. Cent. for Med.-Range Weather Forcasts, Reading, U. K.

Takaya, Y., J.-R. Bidlot, A. C. M. Beljaars, and P. A. E. M. Janssen (2010b), Refinements to a prognostic scheme of skin sea surface temperature, *J. Geophys. Res.*, *115*, C06009, doi:10.1029/2009JC005985.

Tennant, W. (2012), Recent developments within MOGREPS at the Met Office, paper presented at 6th North American Ensemble Forecast System Workshop, Meteorol. Serv. of Can., Monterey, Calif., 1–3 May.

Tennant, W., and S. Beare (2011), New schemes to perturb near-surface variables in MOGREPS, paper presented at ECMWF Workshop on Representing Model Uncertainty and Error in Numerical Weather and Climate Prediction Models, Eur. Cent. for Med.-Range Weather Forcasts, Reading, U. K., 20–24 Jun. [Available at "http://www.ecmwf.int/newsevents/meetings/workshops/2011/""Model-uncertainty/presentations/Poster-Tennant.pdf".]

Vitart, F., S. Woolnough, M. A. Balmaseda, and A. M. Tompkins (2007), Monthly forecast of the Madden-Julian oscillation using a coupled GCM, *Mon. Weather Rev.*, *135*, 2700–2715, doi:10.1175/MWR3415.1.

Ward, B. (2006), Near-surface ocean temperature, *J. Geophys. Res.*, *111*, C02004, doi:10.1029/2004JC002689.

Wheeler, M., and H. Hendon (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Weather Rev.*, *132*, 1917–1932.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Academic, New York.

Woolnough, S. J., F. Vitart, and M. A. Balmaseda (2007), The role of the ocean in the Madden-Julian Oscillation: Implications for MJO prediction. *Q. J. R. Meteorol. Soc.*, *133*, 117–128.

Zeng, X., and A. Beljaars (2005), A prognostic scheme of sea surface skin temperature for modeling and data assimilation, *Geophys. Res. Lett.*, *32*, L14605, doi:10.1029/2005GL023030.