Received: 24 June 2017 | Revised: 28 March 2018 | Accepted: 30 March 2018

DOI: 10.1002/qj.3308

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed

Alicia R. Karspeck[1,4] | Gokhan Danabasoglu[1] | Jeffrey Anderson[1] | Svetlana Karol[2] | Nancy Collins[1] | Mariana Vertenstein[1] | Kevin Raeder[1] | Tim Hoar[1] | Richard Neale[1] | Jim Edwards[1] | Anthony Craig[3]

[1]National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, Boulder, Colorado, USA
[2]Cooperative Institute for Research in Environmental Sciences (CIRES), Boulder, Colorado, USA
[3]External contractor to NCAR, USA
[4]Jupiter Intelligence, Boulder, Colorado, USA

Correspondence
Alicia R. Karspeck c/o G. Danabasoglu, NCAR 1850 Table Mesa Drive, Boulder CO 80305, USA.
E-mail: alicia.karspeck@jupiterintel.com

This paper presents a description of the CESM/DART ensemble coupled data assimilation (DA) system based on the Community Earth System Model (CESM) and the Data Assimilation Research Testbed (DART) assimilation software. The CESM/DART should be viewed as a flexible system to support the DA needs of the CESM research community and not as a static reanalysis product.

In this implementation of the CESM/DART, conventional *in situ* observations of the ocean and atmosphere are assimilated into the respective component models of the CESM using a 30-member ensemble adjustment Kalman filter (EAKF). CESM/DART is run in a "weakly coupled" configuration wherein observations native to each climate system component only directly impact the state vector for that component. Information is passed between components indirectly through the short-term coupled model forecasts that provide the EAKF background ensemble. This system leverages previous ensemble DA development for the Community Atmosphere Model and Parallel Ocean Program models using the DART EAKF. The CESM/DART project is a step towards providing increasingly useful DA capabilities for the CESM research community.

Results are presented for our prototype 12-year reanalysis, run from 1970 to mid 1982. Multiple lines of evidence demonstrate that the system is capable of constraining the CESM coupled model to simulate the historical variability of the climate system in the well-observed Northern Hemisphere. A collection of monthly average variables, climate mode indices, observation diagnostics and snapshots of synoptic weather in the ocean and atmosphere are compared to established datasets, showing especially good agreement in the Northern Hemisphere. A discussion of the CESM/DART as a modular, community facility and the benefits and challenges associated with this vision is also included.

KEYWORDS

Community Earth System Model, coupled data assimilation, coupled reanalysis, Data Assimilation Research Testbed, ensemble assimilation, ensemble Kalman filter, modular data assimilation

1 | INTRODUCTION

Since the mid 1990s, when the development and production of historical climate reanalyses became an established enterprise, most major modelling and operational forecasting centres have conducted their ocean and atmosphere reanalyses as complementary, but essentially separate, activities. For example, at the National Centers for Environmental Prediction (NCEP), the NCEP/NCAR reanalysis project (NCEP-R1; Kalnay *et al.*, 1996) and NCEP/DOE reanalysis (NCEP-R2; Kistler *et al.*, 2001; Kanamitsu *et al.*, 2002) used gridded sea surface temperature (SST) products as their marine boundary conditions. The Global Ocean Data Assimilation System (GODAS; Behringer *et al.*, 1998; Behringer and Xue, 2004) system, also developed at NCEP, was produced in a separate effort, leveraging the surface fields of the NCEP atmospheric reanalyses. Until recently, the European Centre for Medium-Range Weather Forecasting (ECMWF) had a similar separation of efforts. Their series of ERA products (ERA-15, ERA-Interim, ERA-40; Uppala *et al.*, 2005) were all produced using prescribed SST boundary conditions, while their ocean reanalyses (ORA v.1 to v.4) have been forced with the ERA atmospheric surface fields (e.g. Mogensen *et al.*, 2012; Zuo *et al.*, 2015). In both the NCEP and the ECMWF products, the analyses of the ocean and atmosphere were essentially performed independently, with data assimilation (DA) practices that were designed to reflect the dominant time- and space-scales and observational sources for either the ocean or atmosphere systems.

In the last decade, however, there has been increasing interest in the development of *coupled* ocean–atmosphere DA systems which can be used for generating coupled reanalysis products and for initializing near-term coupled climate predictions. Broadly categorized, the anticipated benefits of performing DA in a coupled configuration may include

(a) better representation and constraint of coupled phenomena, near-surface processes, and air–sea fluxes;
(b) more straightforward use of interface observations (e.g. skin SST and surface wind stress); and
(c) representation of the joint uncertainty in the coupled ocean–atmosphere analysis.

There is also evidence from a limited set of studies that coupled assimilation may reduce forecast "initialization shock", potentially leading to improved forecast performance on time-scales of weeks to decades (e.g. Zhang *et al.*, 2007; Balmaseda *et al.*, 2009; Zhang, 2011; Mulholland *et al.*, 2015).

Recognizing the potential benefits of coupled assimilation and initialization, research and operational centres have made rapid progress in the development of coupled assimilation systems in the last decade. The National Oceanic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory (NOAA GFDL) was the first centre to produce a coupled reanalysis for initializing seasonal forecasts. Their ensemble coupled DA (ECDA; Zhang *et al.*, 2007; Chang *et al.*, 2013) product was based on the ensemble adjustment Kalman filter (EAKF; Anderson, 2001; 2003). They mitigated the costs and complexities associated with atmospheric assimilation by ingesting the NCEP-R1 and R2 gridded products into their coupled model – focusing their development efforts on the assimilation of altimetry and subsurface ocean hydrography into their ocean model. Experimental systems have been developed at the Japan Agency for Marine–Earth Science and Technology (JAMSTEC) and the UK Met Office. JAMSTEC has developed a coupled ocean–atmosphere 4D variational (4D-Var) system for initializing seasonal forecasts and conducted experimental simulations over a three-year period. They used ocean initial conditions and adjustment factors in the surface turbulent flux parametrizations as control variables for their nine-month window 4D-Var (Sugiura *et al.*, 2008). The UK Met Office has designed a cycled coupled assimilation system to support seamless prediction (hourly to seasonal). It adjusts the coupled state using updates from separate 3D-Var ocean and 4D-Var atmosphere systems (Lea *et al.*, 2015). It has been prototyped in only short (one year) experiments.

At the time of writing, only two coupled ocean–atmosphere DA efforts have progressed to the stage where multi-decade state-of-the art reanalysis products for both ocean and atmosphere have been generated. NCEP has produced the Climate Forecast System Reanalysis (CFSR) from 1979 to the present (Saha *et al.*, 2010). The CFSR coupled system uses NCEP's seasonal coupled forecasting model and cycled updates are computed and implemented with their separate, pre-existing assimilation systems for the atmosphere and the ocean. The ECMWF has produced 10 ensembles of a coupled reanalysis from 1901 to 2010 based on the Coupled ECWMF ReAnalysis system (CERA; Laloyaux *et al.*, 2016a). Like the CFSR system, the CERA system was built by bringing together the existing ECMWF DA systems for the ocean and atmosphere components. It is based on a 24 h window incremental 4D-Var method, wherein the separate ocean and atmosphere variational systems are applied to the innovations in the inner loops, and the outer-loop forecasts are performed with the coupled model. In this way, observational information propagates across the ocean–atmosphere interface through the outer-loop forecasts.

These developments have established the historical context for efforts that are progressing at the National Center for Atmospheric Research (NCAR) to develop DA capabilities for the Community Earth System Model (CESM). The purpose of this paper is to present an overview of the coupled global DA system (referred to as CESM/DART) developed at NCAR. The CESM/DART is a framework for ensemble DA that can be used for constraining coupled atmosphere–ocean–ice–land simulations of the CESM. CESM/DART uses an ensemble Kalman filter algorithm implemented in the Data Assimilation Research Testbed software (DART; Anderson and Collins, 2007; Anderson *et al.*,

2009). Like the CESM, DART is an open-source facility, developed and maintained by NCAR scientists.

The CESM (and its predecessor model, the Community Climate System Model) has a long history of use as a research tool for the investigation of Earth System processes and Earth System response to changes in atmospheric composition. Its successful use in those arenas have, more recently, led to its use in experimental initialized prediction applications. Examples of this include two CESM-based contributions to the National Multi-Model Ensemble database for seasonal prediction (NMME; Kirtman *et al.*, 2014)[1], initialized global prediction on multi-year to multi-decade time horizons (Yeager *et al.*, 2012; Meehl *et al.*, 2014; Karspeck *et al.*, 2015), multi-year prediction with eddy-resolving ocean models (B. Kirtman, personal communication, 2017), Arctic sea ice prediction (Blanchard-Wigglesworth *et al.*, 2017) and short-term prediction of tropical cyclone tracks (Zarzycki and Jablonowski, 2015). Notably, these applications either require the initialization of model components other than the atmosphere, or in the case of cyclone prediction, may realize some benefit from doing so. While only one of these existing efforts have used CESM-based DA as the means of initializing the model (Karspeck *et al.*, 2015), every initialized prediction effort using CESM models *could* be supported with a native DA and initialization capability. The CESM/DART project is a step towards providing increasingly useful DA capabilities for the CESM research community.

Like most of the coupled assimilation systems currently being used or developed, the CESM/DART project leverages previously developed DA capabilities for the ocean (Karspeck *et al.*, 2013) and atmosphere (Raeder *et al.*, 2012) components of the CESM. One of the attractive attributes of the DART framework for coupled DA is that these single-component assimilation systems are based on the same algorithm, software implementation, format standards for storing and processing observational data, and routines for interfacing with the CESM. In the prototype experiment discussed here, data are only assimilated directly into the ocean and atmosphere, however DA capabilities using DART have also been developed for the land (Zhang *et al.*, 2014) and sea-ice components (Y.-F. Zhang and C. Bitz, personal communication, 2017) of the CESM. The basic framework presented here is extensible to incorporate these other major components.

In addition to a description of the CESM/DART framework, we also present a selection of results from the 12-year prototype coupled ocean–atmosphere assimilation experiment which we performed over the period January 1970 to March 1982. Results from this experiment are compared to existing data products to evaluate both interannual variability as well as synoptic weather patterns. We also show performance metrics relative to conventional *in situ* observations of the ocean and the atmosphere. The results presented here are not an exhaustive evaluation of the system performance, but rather a demonstration that the CESM/DART system can constrain both interannual and synoptic historical variability. This paper is meant to highlight the initial success of this effort, discuss some of the challenges, and open the door for further improvements and developments of the CESM/DART system.

[1]The Community Climate System Model version 3, used in the NMME, is a predecessor model to the CESM

2 | A COUPLED ASSIMILATION SYSTEM FOR THE CESM

2.1 | The CESM model

The model used for assimilation in this experiment is the global coupled configuration of the CESM version 1.1 (CESM1; Hurrell *et al.*, 2013) with active atmosphere, ocean, land and sea-ice components. The atmosphere component is the Community Atmospheric Model version 5 (CAM5; Neale *et al.*, 2012) with a finite volume dynamical core, the ocean model is the Parallel Ocean Program version 2 (POP2; Danabasoglu *et al.*, 2012), the sea-ice component is the Community Ice Code version 4 (CICE4; Holland *et al.*, 2012) and the land model is the Community Land Model version 4 (CLM4; Lawrence *et al.*, 2011). All components were run with a nominal 1° horizontal resolution. The configuration used here is effectively the "workhorse" coupled version of the CESM. For all relevant scientific purposes, this CESM configuration is equivalent to the model version used to produce NCAR contributions to the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor *et al.*, 2012) and the CESM Large Ensemble Project (Kay *et al.*, 2015). One difference, however, is that the model coupling frequency between the ocean and atmospheric components is increased from the CESM1 default of once per day to once every 6 h. As will be described in the following section, the assimilation update is applied to the atmospheric state at 6 h intervals. Increasing the coupling frequency ensures that the ocean state is impacted by the atmosphere updates through the model coupling. Although the model may have detectable differences in behaviour within the atmospheric and oceanic boundary layers due to changes in the coupling frequency, in this version of CESM (where the diurnal cycle in SST is parametrized within the model coupler) there are no obvious physical implications related to this choice.

For the purpose of ensemble DA, the CESM is run in a "multi-instance" configuration (Figure 1). In this configuration, the traditional spoke-and-hub CESM modular design, wherein a single coupler manages the communication of fluxes between components, is extended to multiple instances of each component. Each instance is uniquely matched across the components. For example, CAM instance number 1 exchanges fluxes exclusively with POP instance number 1. In this way, the solutions from multi-instance CESM will be bit-for-bit identical to an ensemble of traditional CESM

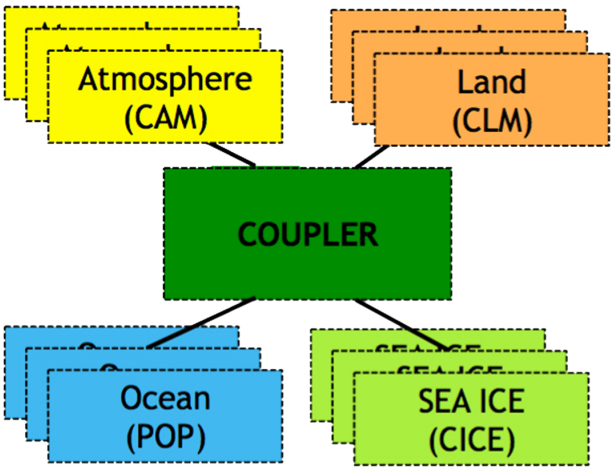

FIGURE 1 Schematic diagram of the CESM coupled model in the multi-instance configuration. Each model component can have multiple instances, but a single coupler manages the communication of all fluxes between components [Colour figure can be viewed at wileyonlinelibrary.com]

simulations. The primary benefits of the multi-instance configuration are ease of implementation and ease of ensemble management. The disadvantages of this configuration are that (a) the use of a single coupler breaks what could otherwise be a parallel ensemble simulation process and (b) the failure of any instance (whether for dynamical or computational reasons) will lead to the termination of all the instances.[2]

2.2 | The ensemble DA system and coupled framework

Most modern systems used for global ocean or atmosphere DA fall into the categories of either variational systems or ensemble Kalman filters. Although very different in their implementation, both of these methods can be understood from a Bayesian perspective: given prior knowledge of the distribution of the multivariate model state (called the "background"), the goal is to find a posterior distribution for the multivariate model state (called the "analysis") as the conditional probability of the model state given the observations. Under commonly made assumptions of linear models and observation operators and Gaussian observational error distributions, optimal solutions can be obtained by either of these methods by consideration of only the mean and the covariances of the prior distribution and observational errors.[3]

The DA scheme used in both the ocean and the atmosphere is an ensemble adjustment Kalman filter (EAKF; Anderson, 2001) as implemented in the DART software. As in all Kalman filters, the prior covariance can be non-stationary because it is evolved through time by the model dynamics. This is one of the strengths of Kalman filters as applied to complex geophysical systems. The central idea is that the relevant features of the distribution can be estimated from the statistics of the ensemble members. The EAKF is a deterministic variant of the ensemble Kalman filter (which was introduced within the literature by Evensen 1994). It falls within the general class of ensemble Kalman square-root filters (Tippett *et al.*, 2003), which also includes the ensemble transform Kalman filter (Bishop *et al.*, 2001). This class of filters transforms the prior state ensembles such that the posterior ensembles have a mean and covariance that is consistent with the full Kalman filter analysis for a sufficiently large ensemble.

In Anderson (2003), the EAKF update is described in a conceptually straightforward manner as a sequence of scalar computations that involves (a) an adjustment to the ensemble in the scalar space of a single observation, and (b) the subsequent update of each variable in the model state vector via linear regression based on the ensemble statistics. The assumption of uncorrelated observational errors is central to the validity of this sequential algorithm. Anderson and Collins (2007) describe the parallel implementation of this algorithm, and can be referenced by readers interested in many of the technical details of the DART software.

In this CESM/DART experiment, observations of the ocean and atmosphere (described in the next section) are assimilated into the atmosphere and ocean components of the coupled CESM model in a *multi-component* or *weakly coupled* framework. In a weakly coupled approach, the analysis increments applied to the state variables of a given component are based only on observational innovations local to that component. Thus, updates to the ocean and atmosphere states can be computed through completely independent instantiations of the DART system. However, between updates, the ensemble of simulations is advanced by a coupled model, exchanging heat, freshwater, and momentum fluxes across the component interfaces.

Thirty instances of the CESM coupled model are advanced in 6 h forecasts; the model program comes to a complete halt at the end of each forecast and writes an ensemble of files containing the state vectors of the POP and CAM components. At 0000, 0600, 1200, and 1800 UTC, a CAM instantiation of the DART program (CAM/DART) reads in the ensemble of atmospheric states and uses atmospheric observations collated across a centred 6 h window to perform the state update. The prognostic atmospheric state vector that is updated by the assimilation consists of temperature, specific humidity, vector winds, and surface pressure at each model grid point. All relevant details of the CAM implementation of the 30-member DART EAKF are described in Raeder *et al.* (2012). At 0000 UTC ocean observations collated across a centred 24 h window are additionally assimilated into the ocean component using a POP instantiation of the DART program (POP/DART). The prognostic ocean state vector that is updated by the assimilation consists of temperature,

[2] As of the time of writing, implementation of a fully parallel multiple-coupler CESM ensemble configuration is planned for the CESM version 2 release.
[3] These assumptions are always violated in problems of geophysical DA, but they remain useful approximations, and necessary in many cases for formulating a computationally tractable problem.

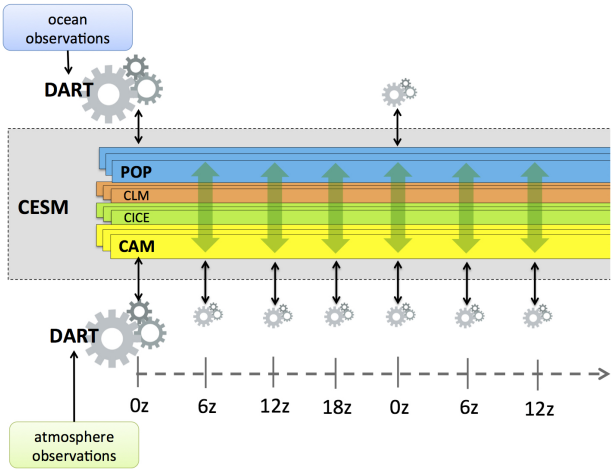

FIGURE 2 Schematic of the DART system interacting with the CESM in the "weakly coupled" configuration. In this configuration, POP and CAM couple every 6 h (CAM and CICE also couple every 6 h and CAM and CLM couple every 0.5 h). The DART assimilation updates the CAM state with atmospheric observations every 6 h (at 0000, 0600,1200,1800 UTC) with binned observations within ±3 h windows. Ocean observations are assimilated using a separate DART system every 24 h (at 0000 UTC), with binned observations within ±12 h windows. The CICE and CLM models are integrated with the CESM and, while their state vectors are never directly updated by the assimilation, they receive information indirectly through the model coupling [Colour figure can be viewed at wileyonlinelibrary.com]

salinity, vector current, and sea level height at each model grid point. The 30-member DART EAKF system as applied to the POP model has been documented in Karspeck *et al.* (2013). As mentioned before, no updates to the land and sea-ice components of the CESM are made – they receive information only indirectly through the model integration. A schematic representation of this configuration is shown in Figure 2. After the POP/DART and CAM/DART programs update the ensembles of ocean and atmosphere states, CESM restarts, reads in the updated state vectors, and advances the next set of 6 h forecasts. Note that for both CAM/DART and POP/DART, ensemble prior estimates of all observations in a window are computed as if the observations were taken at the analysis time.

In this implementation, both the CAM/DART and the POP/DART systems employ the spatially and temporally varying adaptive inflation algorithm of Anderson (2009) as well as covariance localization (in the horizontal and vertical) based on the Gaspari and Cohn (1999) distance-dependent functional form (Anderson and Collins, 2007). The covariance inflation is applied to the prior ensembles at the 6 h assimilation intervals in the atmosphere and at the 24 h assimilation intervals in the ocean. In this implementation of the adaptive inflation algorithms, no deflation is permitted. Table 1 details the relevant parameters of the ocean and atmosphere filters. Note that no formal tuning was performed to arrive at these filter parameters; parameters were set to values that are standard in the community or were observed in previous experience to yield acceptable filter performance. For example, our experiences with the POP/DART system suggested that the adaptive inflation algorithm could lead to regional pockets of very high inflation in the ocean that could excite numerical instabilities in the ocean model. Thus an imposition of a maximum inflation of 20% was set in the ocean. There was no evidence that the CAM/DART system needed a conservative limit on the adaptive inflation algorithm.

The 30 ensemble members used to initialize the filter were taken from a single date in the CESM Large Ensemble (LE) simulation (these can be interpreted as samples from a climatological distribution). Biases in the ocean temperature and salinity fields of the LE were estimated as the difference between the ensemble monthly mean in January 1970 and the climatological January values from the World Ocean Atlas (Locarnini *et al.*, 2013a; Zweng *et al.*, 2013b). This bias was then subtracted from the ocean temperature and salinity from each member of the LE.

2.3 | *In situ* observations of the ocean and atmosphere

Table 2 lists all the ocean and atmosphere observations that are assimilated into the CESM/DART system in this

TABLE 1 Details of the CESM/DART EAKF for the ocean and atmosphere components

	Atmosphere assimilation	Ocean assimilation
Data assimilation method	CAM/DART EAKF	POP/DART EAKF
No. in the ensemble	30	30
Covariance inflation	Multiplicative	Multiplicative
	Adaptive (Anderson, 2009)	Adaptive (Anderson, 2009)
	Applied to prior every 6 h	Applied to prior every 24 h
	$\lambda_{min;max} = 1; 10$	$\lambda_{min;max} = 1; 1.2$
Localization function	Gaspari and Cohn (1999)	Gaspari and Cohn (1999)
Horizontal half-width	0.2 radians (~ 1000 km)	0.2 radians
Vertical half-width	200 mb	750 m
Observational error variance	NCEP prepbufr reports	Karspeck (2016)
	(assumed uncorrelated)	(assumed uncorrelated)

The minimum and maximum multiplicative inflation factors are given by $\lambda_{min;max}$.

TABLE 2 List of observations used in the CESM/DART experiment

Ocean observation type	Ocean variable	Assimilation time (UTC)	Window
CTD/STD/mooring/bottle	*In situ* temperature	0000	24 h
	In situ salinity	0000	24 h
XBT/MBT/drifters	*In situ* temperature	0000	24 h
Atmospheric observation type	**Atmospheric variable**		
Radiosondes/dropsondes	Upper-air temperature	0000/0600/1200/1800	6 h
	Upper-air wind	0000/0600/1200/1800	6 h
	Upper-air humidity	Not assimilated	—
Aircraft reports/ACARS	Upper-air temperature	0000/0600/1200/1800	6 h
	Upper-air wind	0000/0600/1200/1800	6 h
Satellite drift winds	Upper-air wind	0000/0600/1200/1800	6 h

experiment, the time of day that the observations are assimilated, and the window over which they are aggregated. The vast majority of atmospheric data are from radiosonde and aircraft reports of temperature and winds. The influence of data above 100 mb was curtailed following findings by Raeder *et al.* (2012) that assimilation into the lower stratosphere in CAM/DART led to degraded performance. Following from Karspeck *et al.* (2013), *in situ* ocean observations are processed from the 2009 World Ocean Database (Johnson *et al.*, 2009). In the 1970s, the ocean is primarily observed through temperature measurements from XBT profiles and bottle samples from ocean stations. Figure 3 shows locations of the atmosphere and ocean observations over representative summer and winter months during the 1970s. As is evident in these maps, the network of both oceanic and atmospheric observations is exceedingly sparse over the Southern Hemisphere oceans in this decade, especially poleward of 30°S. The footprint of the static radiosonde network over land is apparent, providing relatively even coverage over the Northern Hemisphere land. This network is augmented by intermittent observations over the open ocean, especially in the North Atlantic and North Pacific basins, and along intercontinental routes connecting major cities. As in all decades prior to the Argo era, oceanic observations tend to be clustered near the basin boundaries and along shipping routes, leaving vast areas of the global ocean completely unobserved.

There are a number of key observational sources for constraining the coupled ocean–atmosphere system that were available in the 1970s but are not included in this experiment. These include

(a) marine and land-based *in situ* surface pressure, air temperature, and winds;
(b) radiosonde humidity; and
(c) surface skin temperature and upper-air profiles from infrared satellite radiometers.

Other data sources that are available in later decades include sea surface height from along-track altimetry, *in situ* hydrography from the Argo profiling floats, surface winds from radar scatterometers, atmospheric soundings from GPS radio occultation and satellite microwave radiometers, and satellite cloud-drift winds. Given code for a forward operator within the DART system that maps from the ocean or atmosphere model state vector to the measured data, the CESM/DART can potentially ingest any of these data types. For observational sources that are closely linked to the model state variables (such as retrieved sea level height or *in situ* surface pressure), the forward operator can be as simple as an interpolation from the model prognostic state vector. Implementation

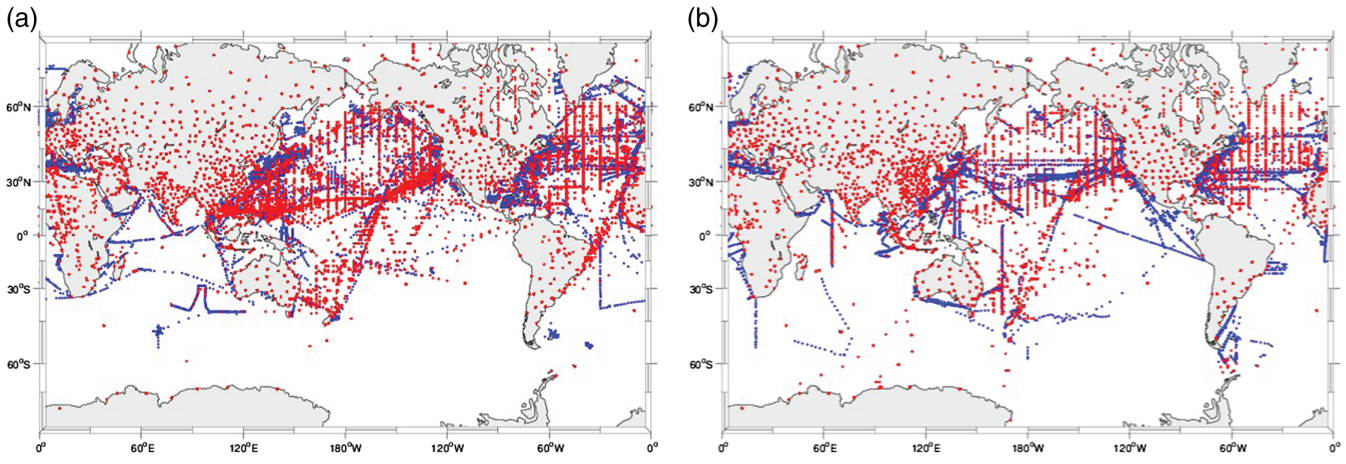

FIGURE 3 Location of oceanic (blue) and atmospheric (red) observations at any height or depth during the months of (a) July 1971 and (b) January 1975

of forward operators for radiances or radio occultation data is far more complex. For example, a forward operator for radiances requires the use of a numerical radiative transfer model and careful accounting for biases and nonlinearities in this mapping. DART currently includes forward operators for radio occultation observations and can use any forward operators available as part of the NCEP GSI, including radiances. Future developments within the CESM/DART will involve the progressive addition of these key historical data sources.

3 | RESULTS FROM A 12-YEAR EXPERIMENT

A prototype coupled reanalysis has been run from January 1970 to March 1982, following the framework outlined above. The original plan was to extend the analysis to the near-present, however computing- and time-resource limits resulted in a substantial reduction in the scope of the experiment. In the course of this experiment, a great deal has been learned about the computational bottlenecks in the CESM/DART system (see Discussion). The assimilation experiment which was completed should be considered a "proof-of-concept" for coupled ocean–atmosphere assimilation with the CESM model.

As this is the first multi-year coupled DA experiment using the CESM model, there is no obvious benchmark against which to measure the performance of the system. A free-running coupled integration of the CESM, initialized from one of the ensemble members (hereafter "NoAssim"), provides one such reference point. We also contextualize the performance of the system in terms of other well-regarded climate reanalysis products. We show multiple lines of evidence that the system is able to reasonably constrain both the historical synoptic and monthly scale variability of the atmosphere and upper ocean. This is not an exhaustive evaluation of this reanalysis, but we show results that are generally representative of the global performance of the system.

The validation emphasis is necessarily placed on atmospheric variables, as these are far more plentiful in the 1970s than ocean observations. The deep ocean is essentially unconstrained in this experiment and, not surprisingly, exhibits decadal-scale drifts in temperature and salinity that are characteristic of ocean simulations that are not in equilibrium with the atmosphere. Results show that, as expected, the simulation is of higher quality in the Northern Hemisphere (relative to the Southern) where the vast majority of the observational data are located.

We remind the reader that only a limited set of the observational sources available in the 1970s and early 1980s was used and that no formal tuning for the coupled assimilation system was performed. Thus, there is every reason to believe that, with the inclusion of additional observational platforms and dedicated tuning efforts, the performance of the CESM/DART system would be improved.

3.1 | Observation-space diagnostics

The DART software provides utilities for comparing model solutions that have been mapped to the observed data type at the measurement location directly to the observations. (These utilities draw on the same forward operators that are used in the assimilation scheme). These *observation-space diagnostics* are a common means of assessing the performance of operational atmospheric reanalysis products because they provide an unambiguous means of comparing to the raw observations. Comparisons can be performed for both the short-term background forecasts and also the updated analysis, allowing error reductions due to the assimilation of all the available observations to be monitored. Figure 4a) shows time series of the ensemble mean root mean square (rms) error and ensemble spread (expressed as a standard deviation around the ensemble mean) of temperature at the 500 mb pressure level for all radiosonde observations globally within weekly windows. Note that inflation has been applied prior to the computation of these diagnostics. As expected, the rms errors and ensemble spread following the update (thin lines) are always lower than for the background forecast (thick lines). Forecast errors decrease during the first 5 years of the assimilation, evidence of a gradual improvement in the quality of the atmospheric state. Interestingly, this multi-year decrease in atmospheric error is longer than would be expected from an atmosphere-only assimilation system (which we might expect to equilibrate on the order of weeks). This decline in rms error is not coincident with an increase in the total number of observations ingested. In fact, Figure 4b shows that there is a slight decrease in the number of observations assimilated over this time. While we cannot say definitely that this behaviour results from coupling with the ocean, one hypothesis is that the slow-time-scale improvement in the atmospheric state stems from a multi-year equilibration of the upper-ocean and atmosphere systems. Note that the time evolution of the observation-space diagnostics are qualitatively similar for temperature at other pressure levels and for the horizontal vector wind components (not shown).

From Figure 4, we also see that there is a clear correspondence between the time evolution of the rms error and the ensemble spread – suggesting that the ensemble spread is providing some indication of the expected forecast error.

Figure 5 shows the time-average of the radiosonde temperature rms error from January 1976 (after skill has stabilized) to March 1982 at pressure levels through the troposphere. Because radiosonde temperature is commonly used in observational diagnostics for atmospheric reanalysis, we are able to put the CESM/DART performance in the context of three current-generation reanalyses with published radiosonde rms error metrics – the Japanese 55-year ReAnalysis (JRA-55; figure 8 of Kobayashi *et al.*, 2015), CFSR (figure 3 of Saha *et al.*, 2010) and ERA-40 (figure 8 of Uppala *et al.*, 2005). For JRA-55 and ERA-40, we show errors from the decade of the 1970s, and for CFSR (which is only available

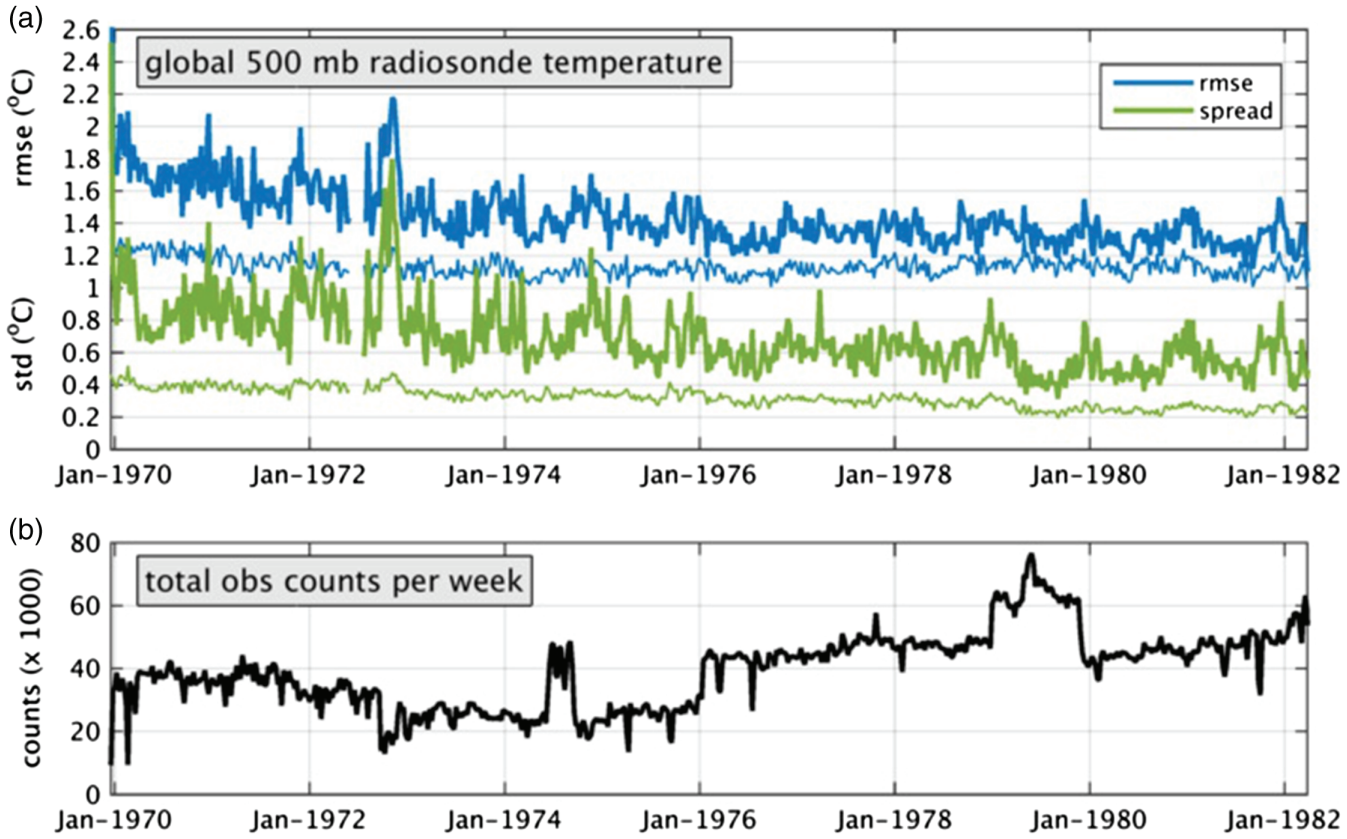

FIGURE 4 (a) Root mean square error (blue) and ensemble spread (standard deviation; green) of 500 mb radiosonde temperatures aggregated over 1 week periods over the global set of observations at 0000 UTC. Bold lines denote the 6 h forecast priors, and thin lines the posteriors. (b) Total number of observations ingested by the atmospheric assimilation each week

since 1979) we show error levels from the early 1980s. Our 6 h background skill is roughly comparable to 9 h background forecasts from CFSR and ERA-40 and roughly 20% less skilful than the JRA-55 published estimates. It is important to note that these are only approximate comparisons – a formal comparison would require that we assess against identical sets of data and at identical forecast leads. (Quality control procedures that determine whether a observation is valid for verification will be different for each of these systems. Note that we validate against all the radiosonde data, even those data that are excluded from the assimilation.) To have error levels that are slightly worse than, but in the same general range as, these operational products is a very promising result; not only do the JRA-55, CFSR, and ERA-40 products ingest more data sources, they are also very mature systems that have benefited from decades of development. For both radiosonde and zonal wind, we also outperform 24 h persistence forecasts (from our posterior analysis) at nearly all levels in these global statistics. A notable exception is the atmospheric low-level winds, where 6 h forecasts do not beat 24 h persistence forecasts. We believe this is due to the rapid development of systematic biases in the low-level atmosphere winds in CAM5 (J. Bacmeister, personal communication, 2017).

Root mean square departures from XBT and bottle (ocean station data) observations of ocean temperature in the upper 250 m are shown in Figure 6. As expected in this aggregated measure, the prior error always exceeds the posterior error. The reduction in error from the EAKF update is typically less than 10% – which can be understood as a consequence of the relatively large error of representation in a non-eddy-resolving ocean model relative to the resolved model variability in the upper ocean (Karspeck, 2016). In the Tropics, where eddy activity is less dominant, the update decreases rms error by up to 30% (not shown). However, in part because of the scarcity of ocean observations and the inhomogeneity of the observing network used for verification, it is not evident from this diagnostic that the ocean state is becoming more faithful through time. The dominant signal is an apparent seasonal cycle in the rms error. Our investigations suggest that this is actually an artifact of systematic seasonal changes in the location of *in situ* observations. In boreal summer and into autumn, ships taking XBT measurements were more likely to traverse the ocean at higher northern latitudes, where the ocean has larger systematic biases in the upper-ocean temperature (Figure 7). These high latitudes are less frequently sampled in the winter and early spring. In the next section, we show some evidence that the ocean solution is becoming progressively more consistent with gridded ocean products (note the evolution of pattern correlations in Figure 12). We would expect to see a greater impact of the update step and enhanced overall performance

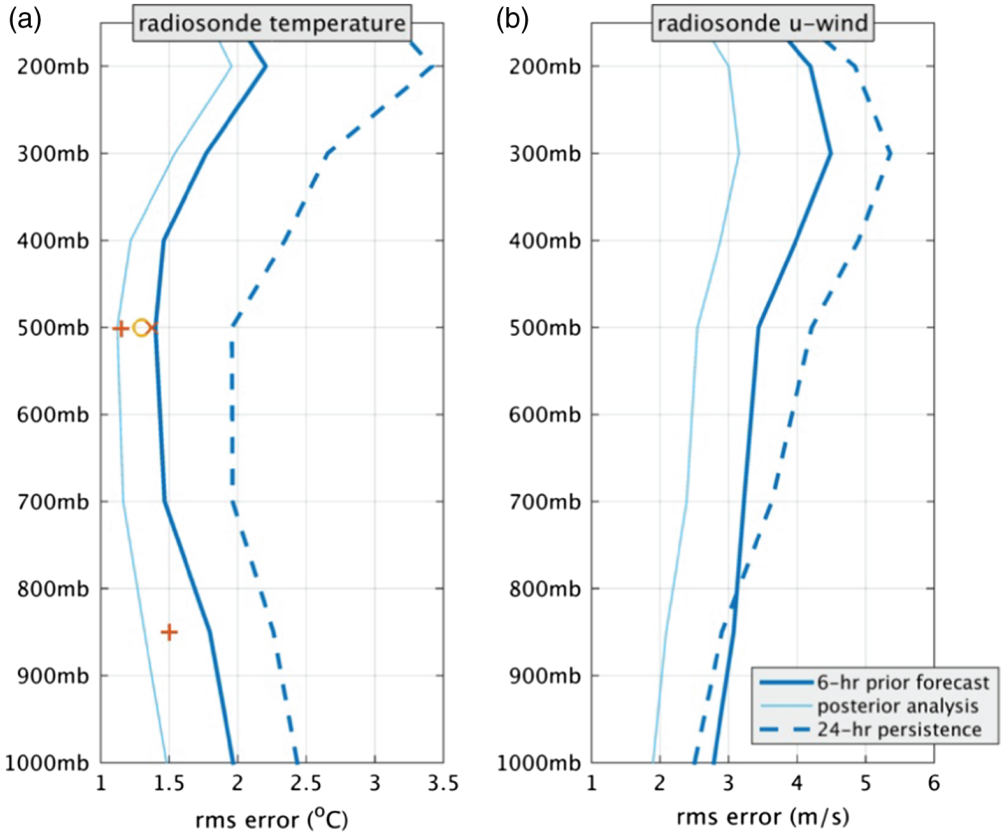

FIGURE 5 Time average of root mean square error of the global set of (a) radiosonde temperature and (b) zonal wind as a function of atmospheric pressure level for January 1976 to March 1982. Thick blue is the 6 h forecast prior, thin blue is the posterior analysis. Dashed lines indicate the performance of a 24 h persistence forecast from the posterior mean. For comparison, the red markers indicate existing published estimates of performance from three operational products: JRA-55 (+), CFSR (x), ERA-40 (o)

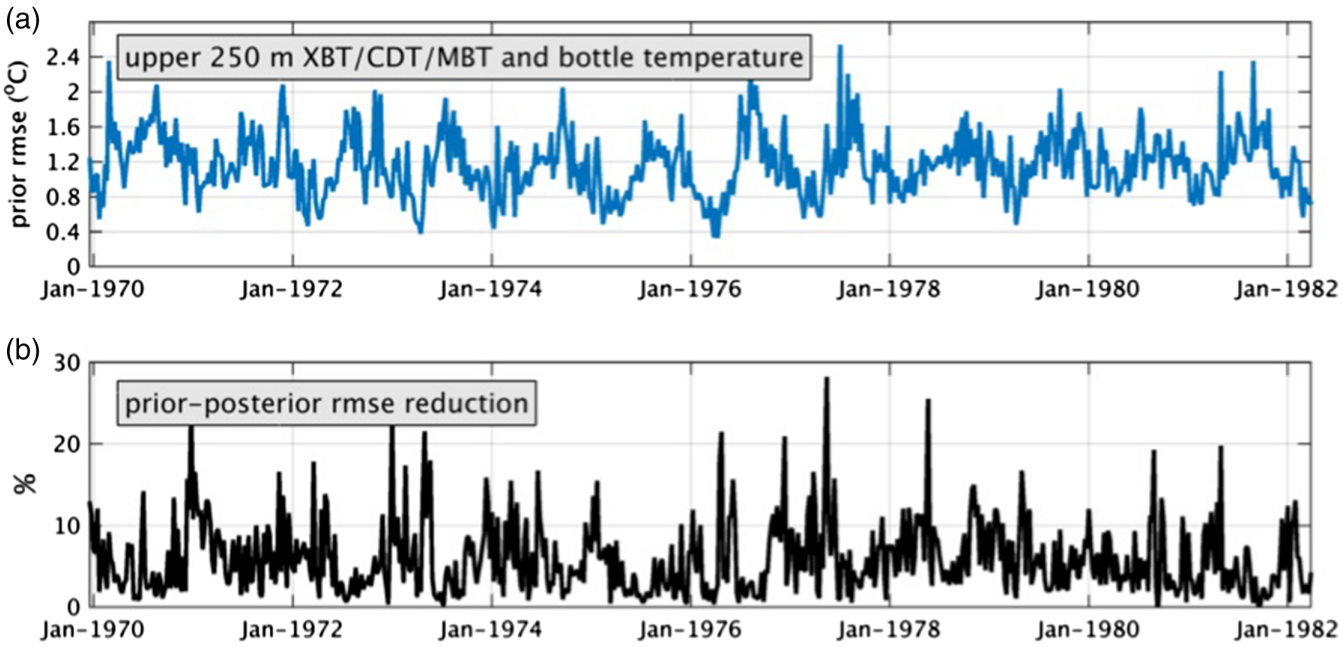

FIGURE 6 (a) Prior root mean square error of ocean temperature for the global set of *in situ* observation platforms aggregated over upper 250 m and over 1 week intervals. (b) The difference between the prior and the posterior rmse expressed as a percent of the prior rmse [Colour figure can be viewed at wileyonlinelibrary.com]

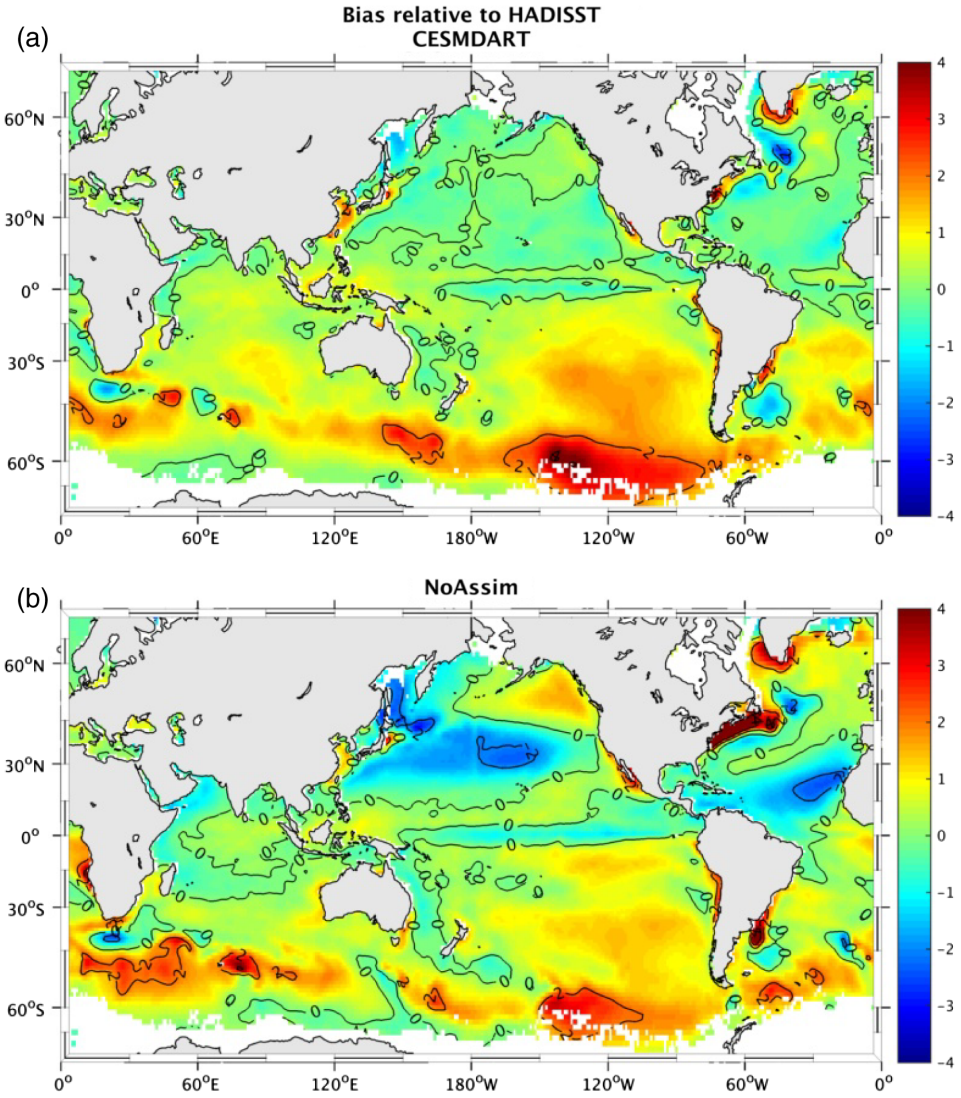

FIGURE 7 (a) SST difference (°C) – time-mean CESM/DART minus the HADISST dataset – over the 1970–1982 period. (b) is as (a) but for the NoAssim control integration minus the HADISST dataset. Note that the HADISST product is not used in the CESM/DART simulation. [Colour figure can be viewed at wileyonlinelibrary.com]

in tests with the more extensive modern observing network (including Argo floats and satellite altimetry). Indeed, Karspeck *et al.* (2013) show a similar seasonal cycle in the upper ocean prior rms error which disappears as Argo comes online.

3.2 | Comparison of monthly averages to global gridded products

Here we show comparisons of the CESM/DART simulation to a collection of monthly average global gridded products including SST from the Hadley Centre Global Sea Ice and Sea Surface Temperature dataset (HADISST; Rayner *et al.*, 2003), zonal wind-stress and sea level pressure from the CERA 20th century coupled reanalysis (CERA20C; Laloyaux *et al.*, 2016b) and upper-ocean temperature from the EN4 dataset (Good *et al.*, 2013). We choose these products to reflect a set of variables that are commonly assessed in global climate models, reflecting large-scale patterns of climate mean state and variability, and for which there are observed data over the decade of assimilation. We also compare the atmosphere temperature, zonal wind, humidity and precipitation to the ERA-40 and JRA-55 reanalyses products. Note that for HADISST, EN4, ERA-40, and JRA-55 there is overlap in the *in situ* ocean and/or atmosphere data ingested in this CESM/DART experiment, however none of these gridded products was used directly in the production. (This point is especially important in regard to the SST products. Both the CERA20C and CFSR products use gridded SST products to stabilize their simulations – we do not.) The CERA20C atmosphere is constrained only by surface pressure and marine surface winds and these data sources are *not* ingested in the CESM/DART experiment. In every case, the CESM/DART simulation results are based on monthly (and ensemble) averages of 6 h forecasts and are interpolated from the CESM native grid to the product grid for comparison. Climatological means are computed for each dataset as the average over the 12-year period of consideration.

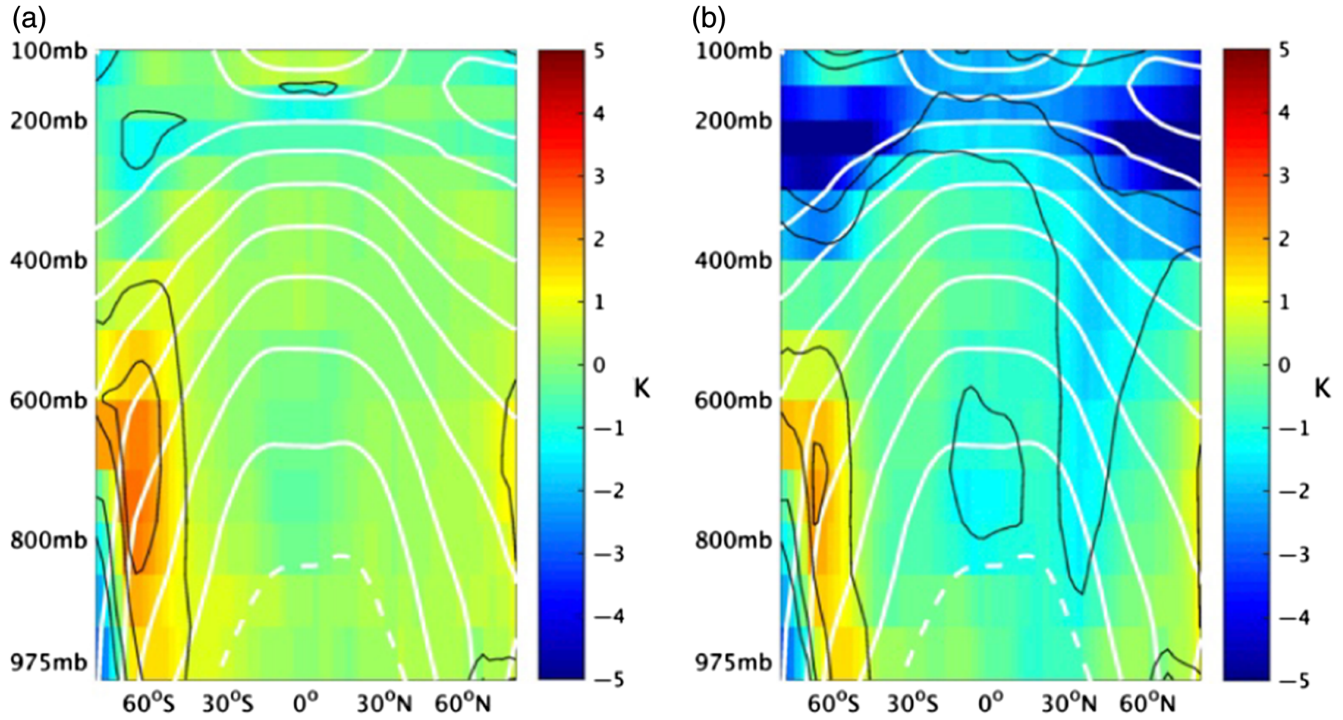

FIGURE 8 Time- and zonal-mean atmospheric temperature difference over 1970–1982. (a) CESM/DART minus ERA-40, and (b) the NoAssim control integration minus ERA-40. Black contours are differences at −2, −1, 1, and 2 K. White contours show the CESM/DART mean field, with the bold dashed line denoting 290 K and contour intervals of 10 K [Colour figure can be viewed at wileyonlinelibrary.com]

3.2.1 | Mean state

In general, the mean state of the CESM/DART in the Northern Hemisphere is more realistic than a free integration (NoAssim) of the CESM when compared to all datasets mentioned above. (The unconstrained NoAssim coupled simulation will be used as a reference throughout this section. Note that NoAssim is not a sequence of short-term forecasts; it is a 12-year coupled integration of CESM.) As three illustrations of the improvement in mean state, we show the time-averaged SST (Figure 7) and the zonal- and time-averaged temperature and zonal wind in the atmosphere (Figures 8 and 9, respectively). The assimilation significantly reduces biases in the Northern Hemisphere SST. Prominent biases throughout the North Pacific, in the tropical Atlantic, and along the path of the Gulf Stream are typical of the CESM model (figure 6 of Danabasoglu *et al.*, 2012). In contrast, consistent with the lack of observational constraints in the Southern Hemisphere, there is not significant improvement in the Southern Hemisphere SST. In fact, the mean state of the central South Pacific SST is slightly worse relative to the HADISST.

The zonal- and time-average temperature in the atmosphere and its bias relative to the ERA-40 are shown in Figure 8. The cold polar tropopause bias (equivalent to a tropopause that is too high), which has been documented in older versions of the CAM (e.g. Neale *et al.*, 2013) as well as CAM5[4], is virtually eliminated in CESM/DART in the Northern Hemisphere, and reduced to near-negligible levels in the Southern Hemisphere. Consistent with a tropopause that is too high, the zonal-mean wind bias in the NoAssim shows polar jets that are also too high (Figure 9b). Jets are also too far south in both hemispheres. These biases in NoAssim are robust to the choice of atmospheric reanalysis product used for comparison. Again, these biases are eliminated in the CESM/DART simulation. Surface subtropical easterlies are also weakened in CESM/DART (Figure 9a) – another improvement to a well-documented bias in CAM.

However, the NoAssim warm bias in the middle troposphere in the Southern Hemisphere is intensified with assimilation, resulting in a weakened midlatitude meridional temperature gradient that is balanced (through the thermal wind relation) by a large-scale reduction in Southern Hemisphere middle-troposphere westerlies. Although the reduction in Southern Hemisphere *surface* westerlies appears to be an improvement (Figure 10), it is associated with the spurious weakening of the westerlies above the boundary layer. A more detailed investigation of the mean-state degradation in the upper levels of the atmosphere while surface winds are improved in the Southern Hemisphere is beyond the scope of this paper. However, it is worth noting that spurious circulation features can arise when model biases are corrected in only selective regions of the globe. This is always a possibility when observational networks are inhomogeneous on hemispheric scales.

[4]Online diagnostics from CAM5 in the context of the CESM Large Ensemble project can be viewed at http://www.cesm.ucar.edu/experiments/cesm1.0/; accessed 5 September 2018.

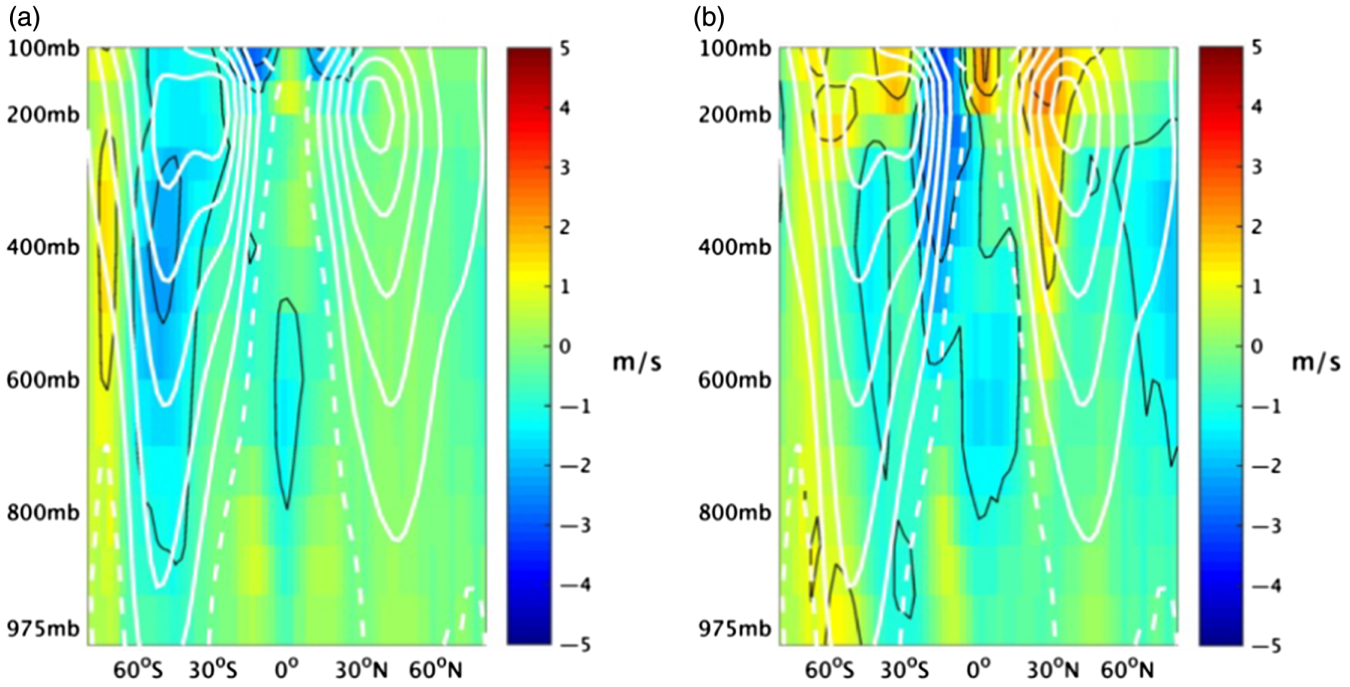

FIGURE 9 As Figure 8, but for atmospheric zonal wind. Black contours are differences drawn at −2, −1, 1, and 2 m s^{-1}. White contours are the CESM/DART mean field, with the bold dashed line denoting 0 m s^{-1} and contour intervals of 5 m s^{-1} [Colour figure can be viewed at wileyonlinelibrary.com]

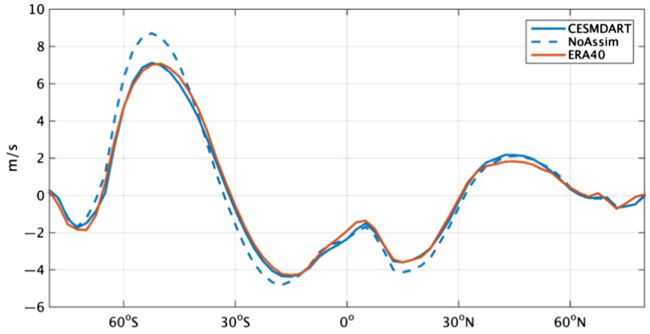

FIGURE 10 Time- and zonal-mean atmospheric zonal wind at 1000 mb for CESM/DART (blue), ERA-40 (red), and NoAssim (dashed)

3.2.2 | Variability of monthly averages

In this section, we focus on the variability around the climatological mean state because it is a more stringent criterion for a DA system and reflects the underlying goals of the data-constraint endeavour. Here, monthly means are used to highlight performance on seasonal to interannual time-scales, showing that there is also reasonable agreement in a collection of synoptic weather patterns and events.

The Taylor diagrams in Figure 11 provide summary plots of the performance of the CESM/DART monthly mean anomalies of atmospheric temperature and zonal wind compared to the NoAssim control experiment and the JRA-55 dataset. Anomaly correlation (along the radial axis) and anomaly standard deviation (along the abscissa) are measured relative to the ERA-40 dataset. The normalized standard deviation is measured relative to the ERA-40 standard deviation. Each point on the diagram corresponds to a randomly sampled grid point from the full horizontal and vertical atmospheric grid, and the correlation and normalized standard deviation are computed along the time dimension[5,6]. The clustering of the CESM/DART points on the Taylor diagram (blue) toward high correlations and standard deviations of unity indicates agreement with ERA-40 in both the magnitude and evolution of the historical variability. The clustering of the NoAssim points (yellow) around correlations of zero, but standard deviation of 1 is expected from an unconstrained simulation with realistic variability characteristics. The JRA-55 points (in red) are provided to give the reader context – for atmospheric temperature and zonal wind, the level of agreement between CESM/DART and ERA-40 tends to be only slightly less than the agreement between JRA-55 and ERA-40. This can be seen more quantitatively in the insets of Figure 11, which show distributions of the *differences* between the correlation of the CESM/DART and ERA-40 at each grid point and JRA-55 and ERA-40 at the *same* grid point. Correlation differences are nearly always negative, indicating that at most points, JRA-55 and ERA-40 fields will be in greater agreement than CESM/DART and ERA-40, albeit by only a small amount. In the Northern Hemisphere, for both temperature and zonal wind, at 95% of geographic grid locations, CESM/DART agrees with ERA-40 at correlation levels that are comparable to JRA-55 within ± 0.1. In the Southern Hemisphere only 50% of points meet this criterion. The enhanced agreement between JRA-55 and ERA-40 is likely due to their ingestion

[5]Note that this is not a typical representation of model performance on a Taylor diagram; Taylor diagrams often plot the correlation and standard deviation computed over the space dimension.
[6]We sample 5,000 points and statistics are stable at this level.

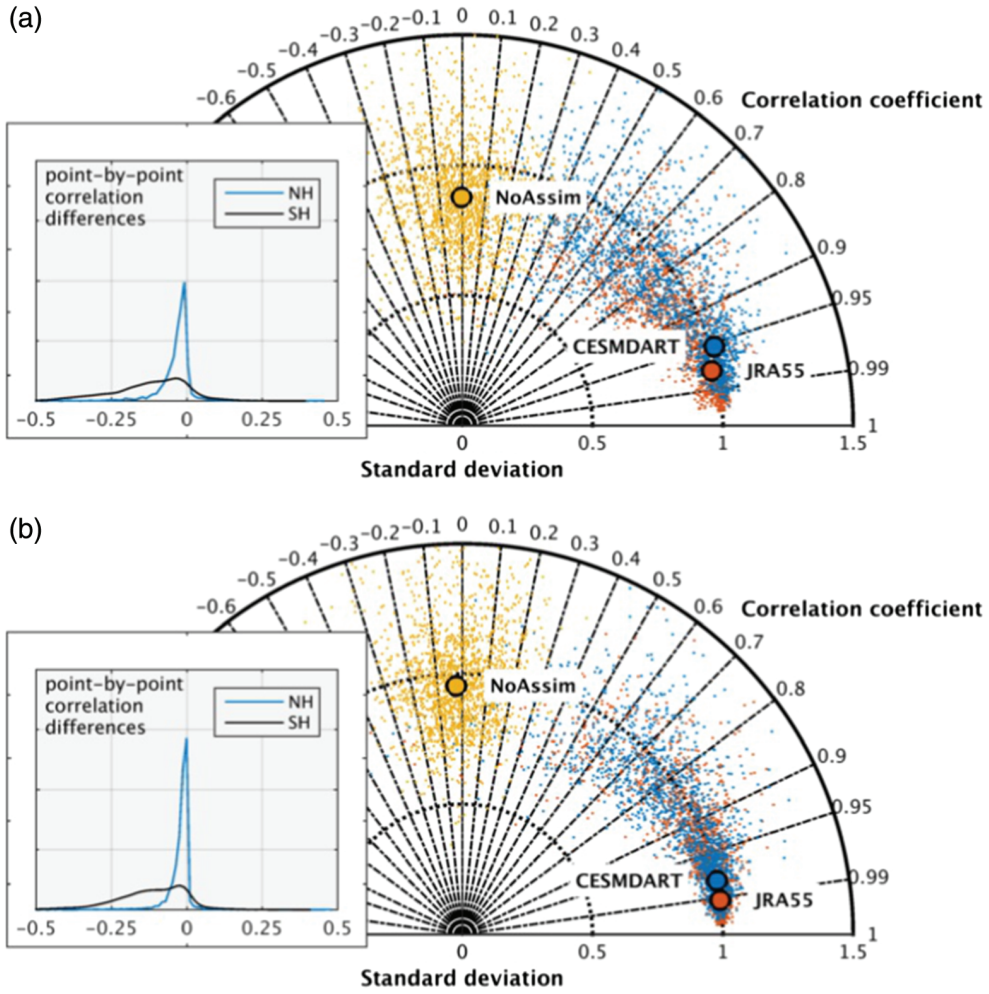

FIGURE 11 Taylor diagrams showing CESM/DART (blue), NoAssim (yellow), and JRA-55 (red) monthly mean anomalies of (a) temperature and (b) zonal wind from 1970 to 1981 referenced to ERA-40. All simulations have been interpolated to the ERA-40 grid prior to analysis. Each point on the plot corresponds to a grid point in horizontal and vertical space. In this way, the correlation and standard deviations represent time (not space) characteristics of the simulations. The most densely populated area of the Taylor diagram (essentially the mode of the joint distribution of points) is indicated by the bold dots. The insets show distributions of the *difference* between the correlation of the CESM/DART and ERA-40 at each grid point and JRA-55 and ERA-40 at the same grid point. These are aggregated by hemisphere

of more shared data sources in the 1970s – including surface pressure measurements, radiosonde humidity, and (probably most significantly) satellite radiances from infrared sounders. None of these data sources are included in CESM/DART. Measurements from the infrared sounder are especially important for constraint in the Southern Hemisphere; the fact that they are not used in the CESM/DART is likely a factor in the reduced Southern Hemisphere consistency.

At most locations in the Northern Hemisphere, the pointwise time correlations between SST, upper-ocean heat content, sea level pressure (SLP), and zonal wind stress datasets and the corresponding CESM/DART fields exceed 0.7 (Figures 12 and 13). In all cases the NoAssim correlations with observations are shown for reference. No formal significance testing is performed, as the NoAssim reference correlations can essentially be interpreted as a sample from a null-hypothesis of "no shared internal variability." Consistent with the Taylor diagram comparisons to ERA-40 shown in Figure 11, the maps in Figures 12 and 13 show lower correlation values in the Tropics and Southern Hemisphere, reflecting the limited *in situ* data for both constraint and verification. The time evolution of pattern correlations are also included, naturally showing greater agreement in the Northern Hemisphere and always exceeding the NoAssim reference. For SLP and wind stress there is a clear seasonal cycle, with exceptionally high pattern correlations in the Northern Hemisphere winters (routinely exceeding 0.95). Compared to SLP and zonal wind, both SST and upper-ocean heat content contain a significantly larger fraction of their total variability at smaller horizontal scales, and for these variables we also show the global pattern correlations based on the leading 20 global empirical orthogonal functions (resolving approximately 80% of the total variability) to highlight the coherence of large-scale patterns. Consistent with these global correlation maps, climate indices that reflect major modes of interannual climate variability, including the North Atlantic Oscillation (NAO) index, El Niño Southern Oscillation indices (Niño3 SST and the Southern Oscillation

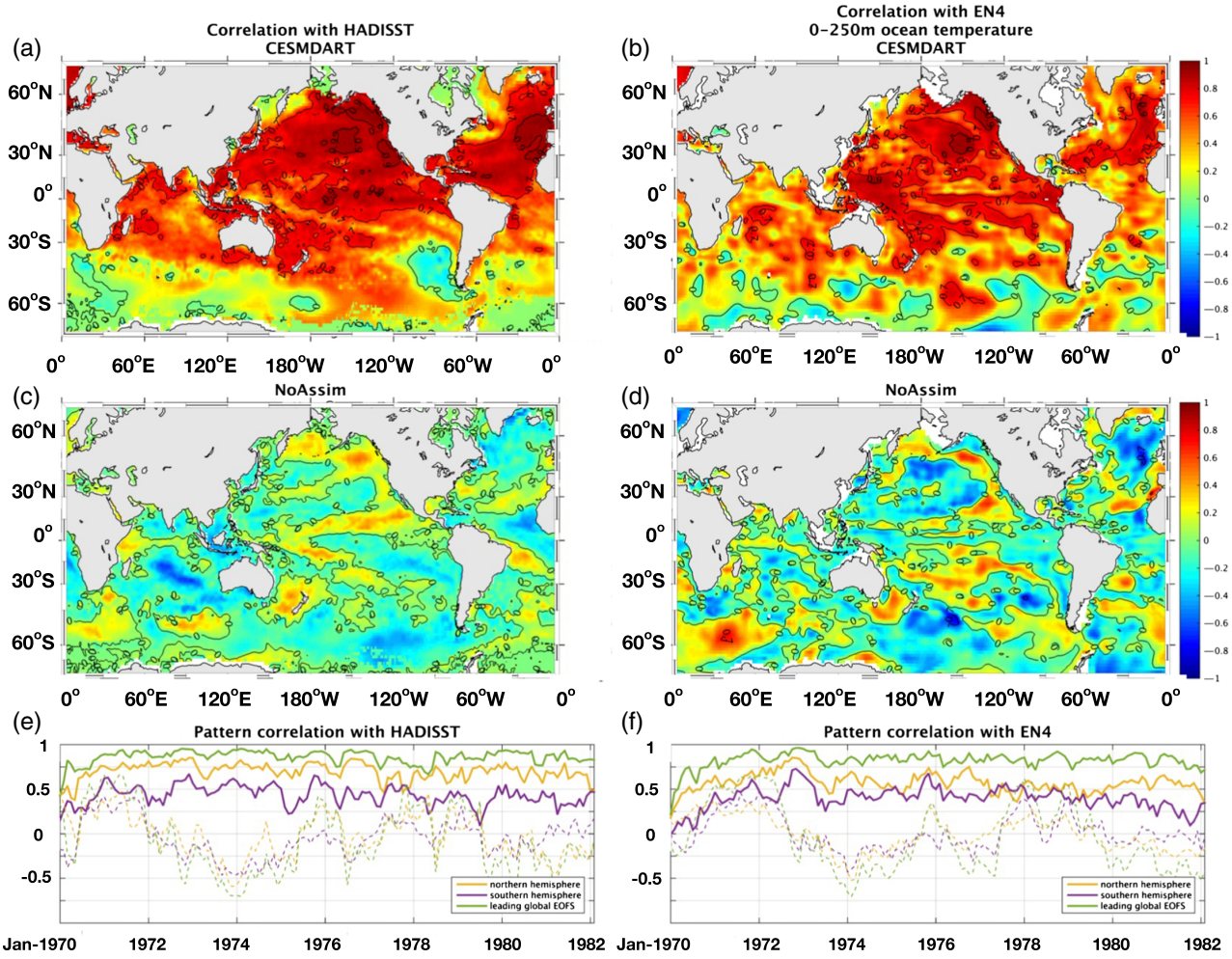

FIGURE 12 (a, c) Pointwise anomaly correlations with HADISST of (a) CESM/DART and (c) the NoAssim monthly average SST. (e) shows pattern correlations with HADISST in the Northern and Southern Hemispheres for CESM/DART (solid lines) and NoAssim (dashed lines). Pattern correlations retaining only the leading 20 global EOFs are also shown. (b, d, f) are as (a, c, e) but showing correlations with the EN4 data product [Colour figure can be viewed at wileyonlinelibrary.com]

Index, SOI) and the North Pacific Index (NPI), are also in close agreement with existing products (Figure 14).

In general, correlations with ERA-40 precipitation tend to be lower than the other variables considered (Figure 15). This is to be expected, as precipitation is not relatable to atmospheric prognostic variables in a straightforward fashion, but is instead the result of complex interactions involving the convective and microphysics parametrization schemes used in these models. Nonetheless, CESM/DART precipitation correlation still exceeds NoAssim over most of the globe and exceeds 0.7 over much of the United States, Europe, and Australia (where the radiosonde network is most dense).

Comparisons of global monthly variability have been made against other datasets (not shown) and the results are qualitatively similar. These include the Hadley Centre SLP (HADSLP2; Allan and Ansell, 2006), marine surface temperature (HADSST3; Kennedy *et al.,* 2011), vector wind stress from the International Comprehensive Ocean and Atmosphere Dataset (iCOADS; Woodruff *et al.*, 2011), the 20th Century Reanalysis (20CR; Compo *et al.*, 2011) SLP, and land precipitation from the CRU TS dataset (Harris *et al.*, 2014).

3.3 | Atmospheric synoptic variability – a collection of results

Since we assimilate in the atmosphere at 6 h intervals, we are also able to constrain the atmospheric synoptic variability over much of the Northern Hemisphere. The observation-space analytics from section 3.1 gave quantitative measures of the 6 h forecast performance. In this section, we show a collection of further results to give readers a more physical, qualitative sense of agreement with established reanalysis and other data sources.

For a visual comparison at specific dates, Figures 16–18 show CESM/DART SLP and precipitation along with the corresponding ERA-40 variables at 0000 UTC on 3 January 1976, 7 February 1978, and 9 August 1980. For the CESM/DART SLP fields, the 1010 mb pressure level for five of the 30 ensemble members is plotted as a grey contour. These dates are chosen because each of them contains at least

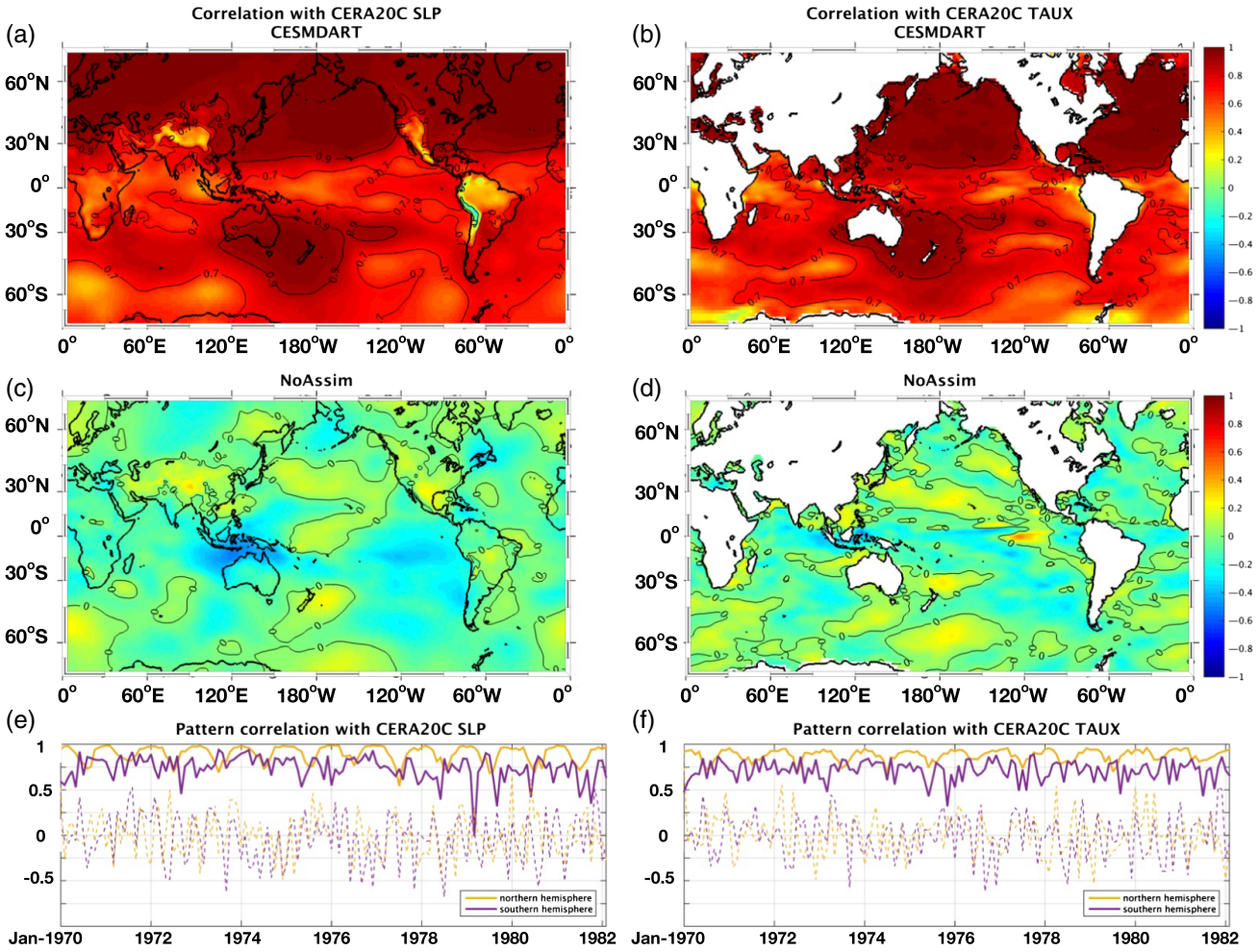

FIGURE 13 As Figure 12, but for (a, c, e) sea level pressure (SLP) and (b, d, f) zonal wind stress, both from CERA20C. Note that the CERA20C atmosphere is constrained only by surface pressure and surface marine winds. These data sources are not ingested in the CESM/DART experiment [Colour figure can be viewed at wileyonlinelibrary.com]

one major weather event for which there is documentary evidence. (The level of visual agreement shown here has been evident for every randomly chosen date we have investigated.) Fields from 3 January 1976 (Figure 16) depict the gale of January 1976, associated with wind damage across Northern Europe and coastal flooding in the British Isles (Shaw *et al.*, 1976). The blizzard of 1978, which brought record-breaking snowfall to the northeast United States (Kocin and Uccellini, 1990) is depicted in (Figure 17), and hurricane *Allen* (at the time, the strongest hurricane ever recorded in the area) as it passed into the Gulf of Mexico (Lawrence and Pelissier, 1981) is shown in Figure 18. Both CESM/DART and ERA-40 show the broad features and timing of each event, but do not have the resolution to simulate the reported maxima in precipitation or minima in SLP. With a hydrostatic, coarse-resolution atmosphere model, we do not expect to simulate the regional details of these events, but their identifiability in CESM/DART provides another piece of evidence that relevant characteristics of synoptic variability are being constrained. More important than any individual weather event, we note that the global patterns of SLP are visually consistent between CESM/DART and ERA-40. Precipitation is less consistent between these products, but the location of most of the major precipitation features in the Northern Hemisphere are co-located in ERA-40 and CESM/DART, with little agreement in the Southern Hemisphere.

A closer inspection of hurricane *Allen* is shown in Figure 19 as it passed through the Gulf of Mexico. The hurricane path closely follows reports by the NOAA National Hurricane Center (the entire reported track is indicated by white dots, with a circle at the specified date). One of the advantages of using a coupled model in the assimilation procedure is that coupled interactions can be simulated. Hurricane *Allen* is followed by a "cool wake," resulting from a combination of turbulent surface heat flux out of the ocean and from mixing induced entrainment of cooler waters from depth. The details of the SST response to hurricane *Allen* are not verified here because archival high-resolution satellite SST is difficult to obtain prior to 1981. However, they represent the expected ocean response to a high-intensity hurricane in the Gulf of Mexico (e.g. Price, 1981). A similar plot is shown of major hurricane *Harvey* as it travels north off the coast of the northeast USA in mid-September 1981 (Figure 20). The Advanced Very High Resolution Radiometer (AVHRR) data

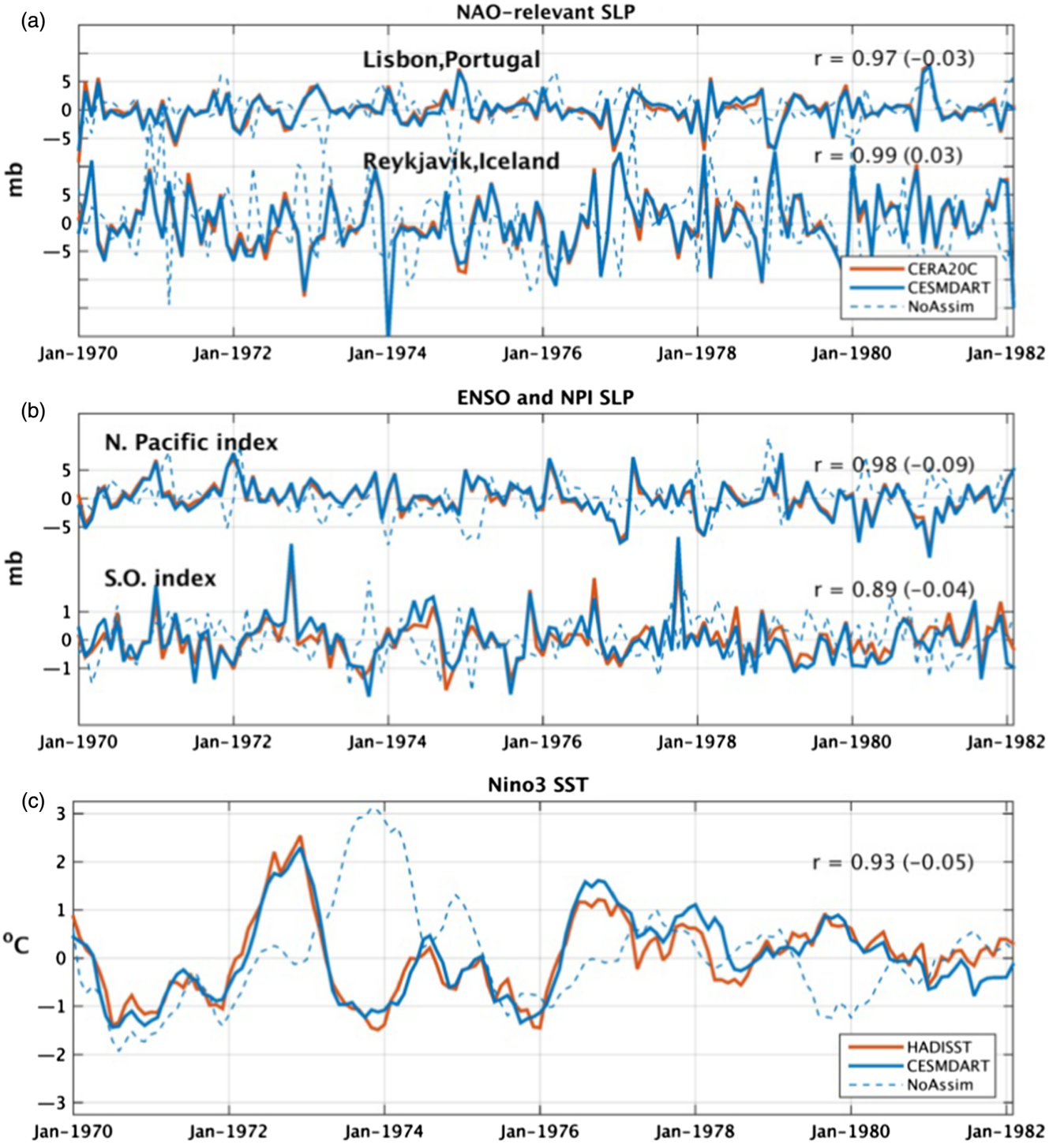

FIGURE 14 A collection of climate-relevant indices from CERA20C and HADISST and ensemble mean CESM/DART simulation. (a) Anomalous SLP at Lisbon and Reykjavik. The difference of standardized SLPs at these locations can be used to indicate the strength of the North Atlantic Oscillation (NAO). (b) The North Pacific Index (defined as the average SLP anomaly from 30 to 65°N and 160°E to 140°W), pressures at Darwin, Tahiti and their difference (the Southern Oscillation Index). (c) The Niño3 index (average of SST from 5°N–5°S, 90–150°W) from the ensemble mean of CESM/DART system and the HADISST dataset. Correlation coefficients between CESM/DART and CERA20C indices are noted, with correlation coefficients between NoAssim and CERA20C indices in parentheses [Colour figure can be viewed at wileyonlinelibrary.com]

are available as a SST analysis during this period (Reynolds *et al.*, 2007), and they are plotted for comparison to the CESM/DART SST. The simulation of hurricane *Harvey* is not as constrained as hurricane *Allen*; the low pressure centre is systematically located west of the reports, and the larger uncertainty is reflected in the disorganization of ensemble members (contours are the 1010 mb isobar for ten ensemble members). Nevertheless, the trajectory is approximately correct. A cool wake in SST is again simulated, and shows qualitatively good agreement with the AVHRR data in terms of the location of the wake. The intensity of the cooling is markedly low (by a factor of 2 or more) likely (in part) due

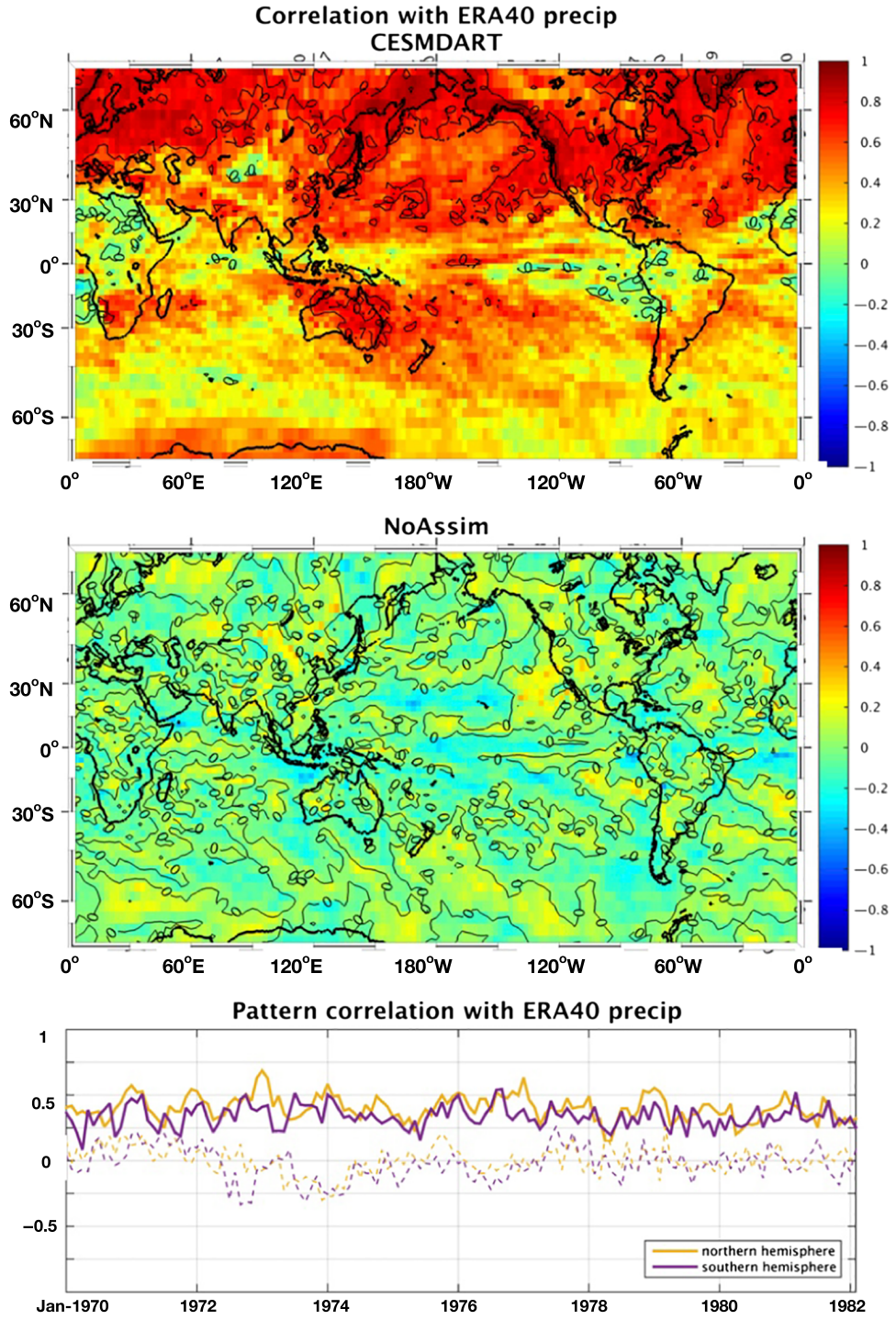

FIGURE 15 As Figure 13, but showing correlation with ERA-40 total (large-scale + convective) precipitation [Colour figure can be viewed at wileyonlinelibrary.com]

to the coarse horizontal and vertical resolution in the ocean model (e.g. Li and Shriver, 2016). The AVHRR data are not ingested into the CESM/DART system in any form, so this feature in the model stems exclusively from the coupled model dynamics.

The frequency of atmospheric blocking events are another way to assess whether this CESM/DART experiment is demonstrating historically accurate synoptic variability. Atmospheric blocking events are large-scale, persistent atmospheric pressure patterns that act to direct the trajectories of storms and cyclones. Here, blocking events in the Northern Hemisphere are defined as in D'Andrea *et al.* (1998). The frequency of occurrence of blocking events (as a function of latitude) is generated through counting exceedances of a threshold criterion applied to the meridional gradient of the 6 h 500 mb geopotential height fields. Figure 21 shows a comparison of the CESM/DART blocking frequency for eight ensemble members over 6 years from 1976 to 1981. In both the Northern Hemisphere winter and summer seasons, CESM/DART blocking frequency is in excellent agreement

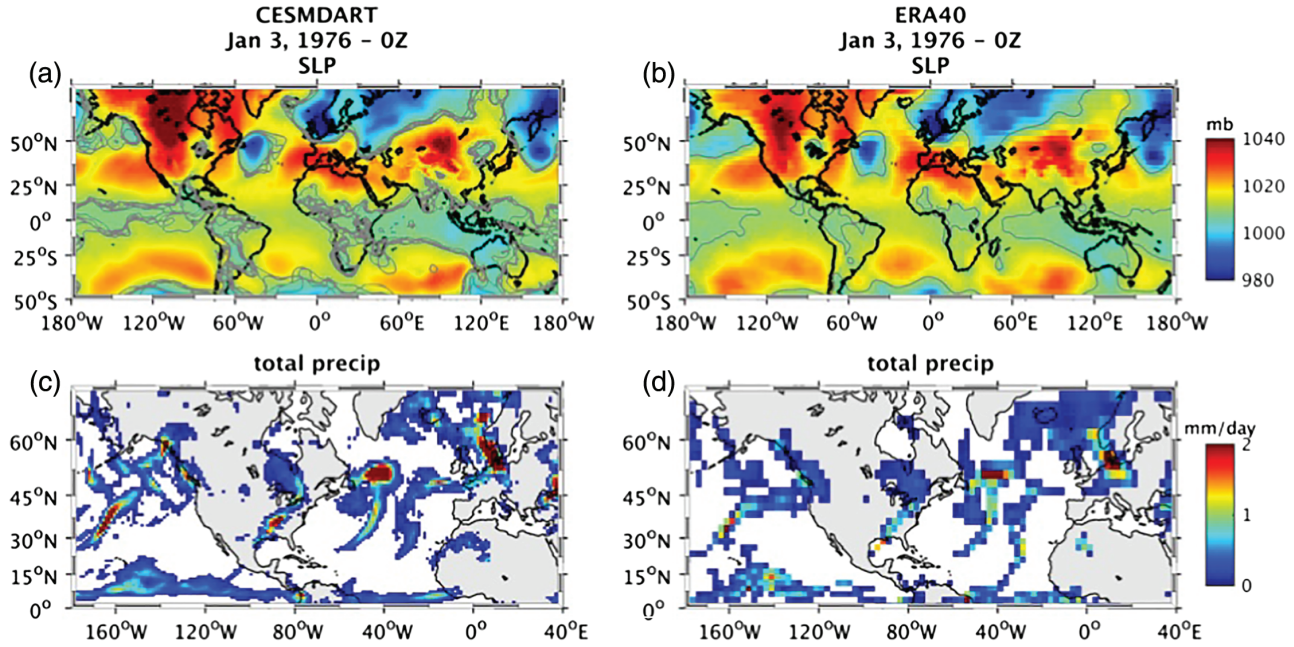

FIGURE 16 Snapshots of 6 h forecasts of (a, b) SLP (colour shading) and (c, d) total (large-scale plus convective) precipitation rate on 3 January 1976 at 0000 UTC from (a, c) CESM/DART and (b, d) ERA-40. In (a, b) grey contours denote 1010 mb (five of the 30 CESM/DART ensemble members are shown). In (c, d) precipitation rates (colour shading) only > 0.05 mm day^{-1} are shown. Note the gale of January 1976 off the coast of northern Europe

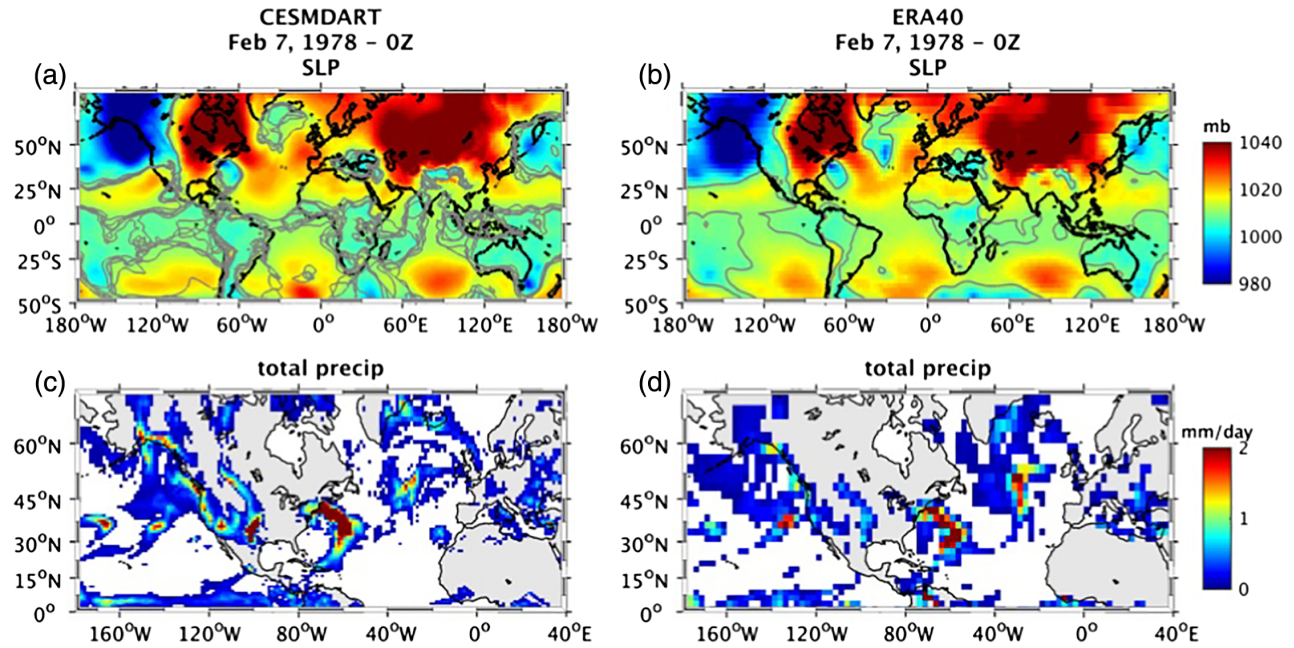

FIGURE 17 As Figure 16, but for 7 February 1978. Note the blizzard of 1978 in the northeast United States [Colour figure can be viewed at wileyonlinelibrary.com]

with the JRA-55 reanalysis datasets. The 20CR dataset is less consistent with both CESM/DART and JRA55 in the Northern Hemisphere summer season. As the JRA-55 uses a more complete set of observational constraints than 20CR in their reanalysis, we believe it represents a more accurate measure of the historical blocking frequency. These results are consistent with the close agreement between CESM/DART and ERA40 SLP shown in the three snapshots in Figures 16–18, but provide a more time-integrative measure of the statistics of synoptic variability.

4 | DISCUSSION

4.1 | Methodological considerations

As a proof of concept, this article has highlighted just one configuration of the CESM/DART system. However, taking a broader perspective, one of the core strengths of this system is its modularity. The DART system was designed at a foundational level to interface with any geophysical model, with diverse observational data sources and with options for a

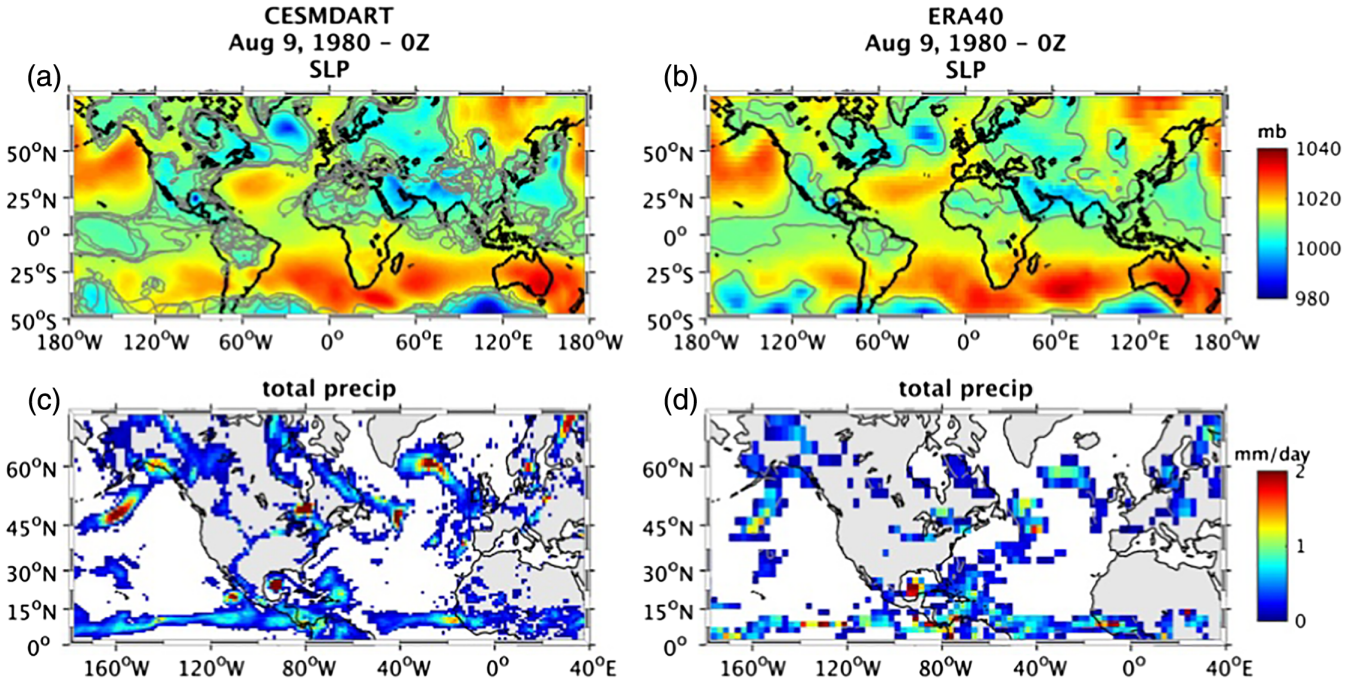

FIGURE 18 As Figures 16 and 17, but for 9 August 1980. Note hurricane *Allen* in the Gulf of Mexico [Colour figure can be viewed at wileyonlinelibrary.com]

variety of ensemble update algorithms. Likewise, the CESM system also has an intrinsically modular design, wherein a diverse set of component configurations can be prototyped with relative ease. The maintenance of this core modularity is one of the key tenets of the CESM/DART system because it encourages novel research applications of the system while still maintaining state-of-the-art modelling and assimilation algorithms and software facilities. Indeed, there is a notably low barrier to entry: numerous researchers have set up and used CESM-component models with DART for research applications with only a few months of dedicated work. Although it is not required that modular systems be purely ensemble-based, there is a natural synergy between ensemble DA systems and modular DA. It is useful as a point of discussion to consider the role that the choice of DA method plays in supporting the tenet of modularity.

Traditionally, variational systems use parametric formulations of a background covariance, with the background mean defined through a model forecast (or ensemble of forecasts). In hybrid variational-ensemble methods, information from ensemble model forecasts is used to augment the background covariance (e.g. Hamill and Snyder, 2000; Buehner, 2005; Penny, 2014; Bonavita *et al.*, 2016), but a parametric background covariance is still necessary. The problem of finding the most likely model state given the observations is then solved through the minimization of a multivariate cost function[7]. Relevant to the discussion of modularity, in traditional variational systems the background covariance is defined by the practitioner – it is not diagnosed purely as an emergent property of the evolving dynamical state. One clear benefit to running a system with prescribed or highly parameterized background covariances is that the system can be designed to behave consistently, efficiently, and with attractive user-defined characteristics. For example, it is common in practice to formulate the variational minimization problem in the ocean and the atmosphere such that physical balances between variables (e.g. mass conservation, geostrophy) are imposed in the update (e.g. Kleist *et al.*, 2009; Mogensen *et al.*, 2012)[8]. The imposition of variable transformations that ensure balance is not just physically desirable, it is also linked to the efficiency of the numerical solvers (Derber and Bouttier, 1999). In contrast, pure ensemble methods assume that nearly all the relevant features of the multivariate distribution can be estimated from the statistics of an ensemble of short-term model forecasts. This potentially allows for a richer and more realistic manifestation of the prior distribution and its non-stationarity. Just as important, it frees the practitioner from the need to develop descriptions of (often complicated) multivariate joint distributions or to have *a priori* intuition about the dominant physical balances of the system. Not only does this support the schema of "modular data assimilation", wherein the same basic DA system can be used for different geophysical systems, it also facilitates the rapid development of DA capabilities for component models that may not have robust disciplinary traditions in DA. This is a key aspect within the context of the CESM.

[7]This is easier said than done. Employing efficient solvers for 3D-Var and 4D-Var in global geophysical systems is a serious endeavour to which a great deal of research and engineering are devoted. Discussions of the iterative solvers at ECMWF and future directions can be found in Fisher *et al.* (2009; 2012) and Tshimanga *et al.* (2008).

[8]Ensemble methods can also be made to respect physical constraints, but this option is not as frequently exercised.

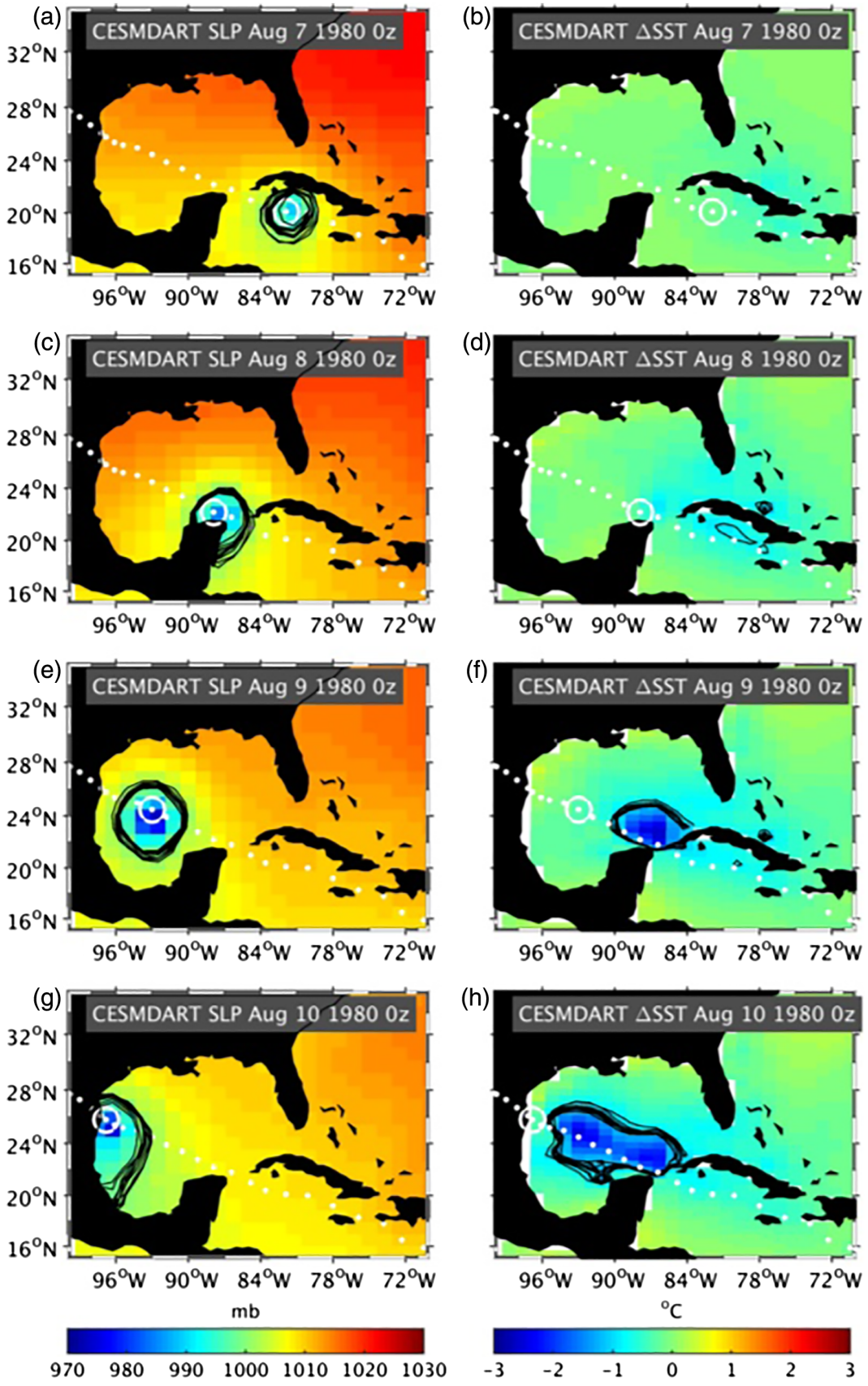

FIGURE 19 Six-hour forecast snapshots of (a, c, e, g) SLP and (b, d, f, h) SST associated with hurricane *Allen* in the Gulf of Mexico at 0000 UTC on (a, b) 7, (c, d) 8, (e, f) 9, and (g, h) 10 August 1980. To highlight the SST cool wake, the difference between SST at each time and that on 6 August 1980 (before the hurricane passed through) are shown. The hurricane track reported by the NOAA National Hurricane Center is shown as white dots and the circles correspond to the plotted date. Black contours show the 1000 mb (−1°C) SLP (SST) values for ten of the 30 ensemble members [Colour figure can be viewed at wileyonlinelibrary.com]

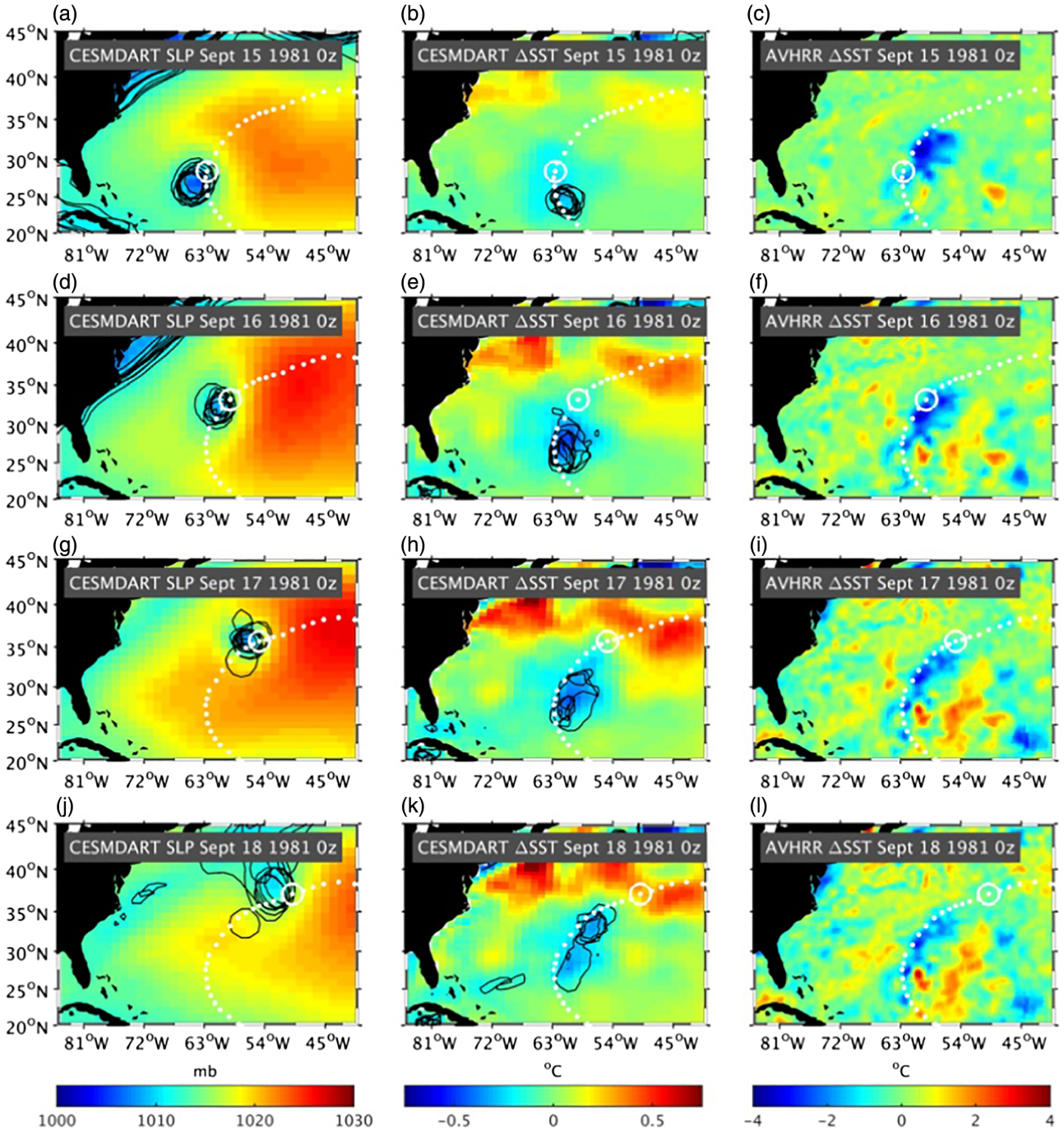

FIGURE 20 As Figure 19, but for major hurricane *Harvey* as it passed northward in the Atlantic from 15 to 18 September 1981. Black contours show the 1010 mb (−0.4°C) SLP (SST) levels for ten of the 30 ensemble members. Also included (c, f, i, l) is SST from the NOAA daily AVHRR-only SST optimal interpolation product (Reynolds *et al.*, 2007). This data source is only available from 1981 and was not assimilated in this experiment. The SST differences are with reference to 14 September 1981 to highlight the cold wake [Colour figure can be viewed at wileyonlinelibrary.com]

Additionally, ensemble DA extends naturally to initialized ensemble prediction, which, as discussed in the Introduction, is garnering increasing interest in the CESM community.

Of course, there are drawbacks to using an ensemble-based approach for global-scale DA that must be balanced against the benefits. The foremost challenge (essentially a "fixed cost" of ensemble DA) is the computational expense of running multiple realizations of global models – typical ensemble sizes for global ocean or atmosphere models range from 10 to 100 members. These ensemble sizes are determined by practical computing limitations and it is widely understood that most global geophysical problems have an effective dimension that is orders of magnitude larger than the available number of samples. (There is evidence that ensemble sizes of 10^3-10^4 can be beneficial for atmospheric assimilation (Miyoshi *et al.*, 2014). Under these conditions, the problem of sampling error in the computation of ensemble-based background statistics cannot simply be ignored. In the context of ensemble Kalman filters, the two most common strategies for dealing with the sampling error

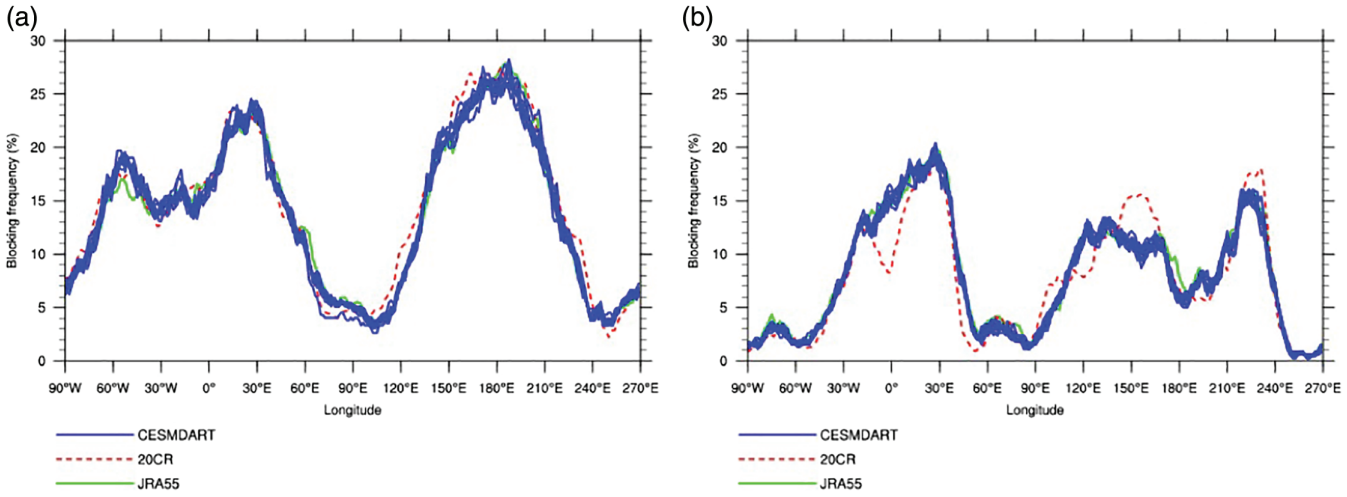

FIGURE 21 Atmospheric blocking frequency for eight of the CESM/DART ensemble members (blue lines) from 1976 to 1981 during the Northern Hemisphere (a) winter (December, January, February) and (b) summer (June, July, August) months. Blocking frequency from JRA-55 (green line) and the 20CR (dashed red line) are included for comparison

problem are (a) "localization" of the ensemble covariance (e.g. Hamill *et al.*, 2001; Furrer and Bengtsson, 2007) wherein the influence of an observational innovation on geophysically distant model-state variables is mitigated by a scaling factor, and (b) artificial "inflation" of the ensemble variance (e.g. Mitchell and Houtekamer, 2000; Anderson, 2009; Whitaker and Hamill, 2012; Houtekamer and Zhang, 2016). In practice, these are heuristic strategies whose effectiveness depends on parameters that are only indirectly associated with physical processes in the model. While more generic than variational systems, it is inadvisable to treat ensemble systems as "black boxes". Discussion of the nuances of localization and inflation are not the focus of this paper, but we mention them here to highlight that realization of the potential benefits of ensemble methods in practice is not trivial.

Further, the admittance of dynamic, sample-based covariances in large-dimensional systems vastly increases the complexity of the DA system; the interactions between the model dynamics and the DA can very often lead to unexpected and non-intuitive system behaviour. It is precisely because the practitioner has only limited control of the solution that understanding spurious behaviour and constructing appropriate remedies can be extremely challenging.

4.2 | Computational considerations in regard to CESM interfacing with DART

There are substantial computational costs associated with naively exploiting the CESM/DART modularity and generic-ism that were brought to light in the process of setting up and running this experiment. With the CESM/DART system, computational performance became a first-order concern. In part, this was because the prototype simulation was intended to be over 50 years in length. Our inability to accomplish a simulation of that length within a reasonable time horizon forced an examination of the computational bottlenecks.

We ran the CESM/DART integration on NCAR's Yellowstone (CISL, 2012) and the DOE-NERSC[9] Edison high-performance computing resources. Using a range of 1350–1800 cores, we had a typical throughput of 10–20 wall-clock minutes per 6 h assimilation cycle. Performance varied significantly between the two machines. Over the 12 years of simulation, we averaged a cost of about 1 million core-hours per simulation year. Inclusive of job-queuing, the frequent need to stop production to archive output to long-term tape storage, computing and disk system failures, and diagnosing dynamical errors induced through the assimilation in any of the 30 ensemble forecast members, we could expect to accomplish 1 year of simulation every 3–6 weeks.

Only a small fraction of the total run-time of the joint CESM/DART system was spent in the DA update (5–10%), and the integration of the ensemble forecasts was only about 25% of the total CESM/DART program time. Where was most of the time spent?

The majority of the time spent by the DART program was in redistributing data in memory such that DART could compute the update to the ensemble members and reassembling the data such that it could be used for re-initializing CESM. It is important to note that this data movement should not strictly be categorized as I/O (which we estimate to be as low as 5% of the total cost of the system). Although CESM and DART pass data through files, there is an intrinsic cost to the data movement from the "state-complete" data-storage strategy of CESM to the "ensemble complete" data-storage strategy of DART and back again that the elimination of file-based transfer will not mitigate. (This is sometimes termed the

[9]US Department of Energy's National Energy Research Scientific Computing Center

transpose cost.) A new release of DART ("Manhattan"[10]), which is available at the time of writing, includes an internal redesign of the way that data are stored in memory. Among other things, it reduces the number of transposes in a DART state-update. More generally, both hardware- and software-based strategies for speeding up inter-executable communication on distributed memory computers with little or no impact on the communicating programs are emerging within the community (e.g. Liao *et al.*, 2017).

The majority of time spent by the CESM was in its start-up initialization phase at the start of each 6 h assimilation cycle. Typical of any model start-up procedure, this includes reading the prognostic state vectors and metadata, distributing the data in memory, and making preparatory computations to begin integration. The expense incurred through the frequent CESM re-initializations (four times per simulation day) is particularly pernicious because there is currently no comparable use-case within the CESM community; runtime initialization is an insignificant cost when integrating for years to centuries, which is more typical of a CESM use-case. The software engineering required to ameliorate the issue is non-trivial and requires a significant rethinking of how CESM and DART should interface. This issue has been present in the POP/DART and CAM/DART efforts, but resolving problems associated with the scientific performance of those systems took precedence over a focus on computational performance.

The need to consider these performance issues is not unique to the CESM/DART project. In any ensemble Kalman filter code designed to accommodate the large state vectors typical of geophysical models, the choice of how to move data between the forecast functionality (the model) and the update functionality (the DA algorithm), as well as how to minimize the overhead of repetitive model start-up costs will be key factors in the overall computational performance.

For systems dedicated to a specific use-case (e.g. ECDA; Zhang *et al.*, 2007) combining the model and DA into a single executable can be a useful strategy for managing these costs. The upfront software engineering cost to architect these system will be larger – but the performance gains can be significant. There is at least one other *modular* ensemble Kalman filter system, also designed to accommodate the codes and state vectors typical of geophysical DA, that has taken the approach of combining the model and DA codes into a single executable. Like DART, the Parallel Data Assimilation Framework (PDAF; Nerger and Hiller, 2013) embraces modularity as a design criterion, and employs a scalable implementation of the ensemble Kalman filter update algorithm. A distinguishing characteristic of the PDAF system is that it is meant to be embedded within the model code as a series of subroutines. In theory this strategy eliminates the need to use files to communicate state vector information between the model and DA update module and the costs associated with restarting the model at each assimilation step. On this basis alone, however, transpose and data movement costs are not eliminated. To our knowledge the PDAF system has never been implemented with the CESM or in a multi-component DA use-case and so performance characteristics and ease of implementation with this system is unknown.

This discussion highlights the significant software design challenges that emerge when trying to balance the genericism of the CESM and DART systems with the desire for a more efficient and scalable joint-system. These are potential factors for many large state vector, complex geophysical models that must interfacing with a separate executable program on high-performance distributed memory computers.

4.3 | Ongoing developments

Regarding the future of CESM/DART development, a key factor is that there is currently no operational mandate for the generation of reanalyses or initialized climate forecasts at NCAR. This gives NCAR scientists and CESM collaborators the freedom to target emerging applications that require DA, research questions relevant to the field of DA science, and other specialized use-cases. Two examples of this are (a) the experimental implementation of a "strongly coupled" DA update, and (b) modification of the POP/DART system to use static ensemble perturbations, essentially creating an ensemble optimal interpolation scheme (e.g. Oke *et al.*, 2002; 2008; Counillon and Bertino, 2009)

In contrast to the weakly coupled system described here, in a strongly coupled system information from observations of each component can impact all the components during the update step. In our CESM/DART framework, this is accomplished as a straightforward extension of the DART state vector to include the prognostic states of both the ocean and the atmosphere. So in this case, our framework has supported a natural research progression – single-component DA systems (POP/DART and CAM/DART) were extended to a weakly coupled system, which is now being extended to a strongly coupled system. The strongly coupled system was prototyped specifically to address open research questions on the value and best practices for global coupled DA (Sluka *et al.*, 2016; Penny and Hamill, 2017). Set-up was done at a modest incremental development cost (several months of expert-person time).

Regarding the project on high-resolution ocean assimilation, this project was designed to support cutting-edge science in initialized high-resolution global coupled-model climate prediction. While the DART developers did not design the system with the use of static perturbations in mind, the overhead for implementing this option was trivial (order weeks of expert-person time). For computationally expensive models like an eddy-resolving global ocean model (1/10° horizontal resolution), where running ensembles is simply not feasible, having software that can easily be configured to support a

[10]https://www.image.ucar.edu/DAReS/DART/Manhattan/documentation/html/Manhattan_release.html

less-sophisticated background distribution is a sensible way to achieve specific science objectives.

Although the results of our 12-year coupled simulation are publicly available (by request to A. Karspeck and G. Danabasoglu), the CESM/DART is best viewed as a system, not as a product. Currently, we envisage that resources will be primarily targeted toward supporting community research through prototyping new system configurations, facilitating the development of more complex capabilities, and aiming to make the system more user-friendly and computationally affordable.

5 | CONCLUSIONS

In this paper we have introduced the first implementation of a coupled global DA system developed with the DART DA software and using the CESM. A 12-year experiment (from 1970 to mid-1982) assimilating conventional *in situ* atmospheric and oceanic observations is described. Multiple lines of evidence are presented to show that this system can represent monthly historical variability of the climate, major modes of climate variability, and synoptic weather patterns in the Northern Hemisphere. Anomaly correlation with well-established gridded products in terms of SST, surface momentum stress, SLP, precipitation and upper-ocean heat content are presented. In all cases the correlations significantly exceed a free run of the coupled model in the Northern Hemisphere. Root mean square errors of 6 h prior forecasts against radiosonde temperature and winds are also presented. On a global average, the fit to atmospheric observations exceeds that of a 24 h persistence forecast at all levels in temperature and at upper levels (outside the atmospheric boundary layer) for vector winds. In terms of the background error in 500 mb radiosonde temperature (a commonly used evaluation metric), the CESM/DART experiment shows performance that is slightly lower but within the range of the published estimates of established reanalysis products. In the atmosphere, we also show that the ensemble spread is a reasonable indicator of the root mean square error – this is a first-order indication of the reliability of the ensemble. Finally, we show a few examples of global and regional synoptic weather features to demonstrate the level of agreement with reanalysis and weather observations.

The results presented here are very promising, as there was no explicit tuning of the coupled system and only a subset of all available observations was used. There is every reason to believe that the system has the potential for greater performance.

We have also presented a brief discussion of some of the computational and resource challenges that were brought to light over the course of this project. Because ultimately CESM will rely on community involvement to advance the research use of CESM/DART, resource usage and ease-of-use issues are currently the most pressing challenges.

As an introductory paper to the CESM/DART system, this paper necessarily leaves the more interesting questions about coupled assimilation unanswered. A few examples are:

- For what class of physical phenomena and applications does coupled assimilation present a quantifiable benefit over the standard single-component approaches to DA?
- In regard to coupled DA frameworks and methodologies, what are the best strategies to account for the abrupt transition in time and space scales across component boundaries?
- Does the value of observational assets change when reconsidered in the context of their ability (or lack thereof) to constrain across boundaries?

Moving forward, the CESM/DART system has the potential to be used in a research context to investigate these and many other applied DA research questions.

ACKNOWLEDGEMENTS

We acknowledge the hard work and dedication of all the scientists and software engineers who contributed to the development of both the CESM and DART projects. Special thanks are extended to Steve Goldhaber, Alice Bertini, and the entire CESM Software Engineering Group for guidance and support in developing interfaces between the CESM and DART system. Thanks as well to John Dennis and Brian Dobbins of the Computational and Information Systems Laboratory (CISL) for their ongoing work investigating bottlenecks in the joint CESM/DART system. Early experiments with coupled CESM/DART configurations were performed by Abhishek Chatterjee (now at NASA GMAO) during his time as a postdoctoral fellow at NCAR – we appreciate his pioneering initiative. AK was funded by the NSF Collaborative Research EaSM2 grant OCE-1243015, and SK (who spent many months shepherding the experiment through the high performance computing systems) was funded by the NOAA Climate Program Office under the Climate Variability and Predictability Program grant NA13OAR4310138. NCAR is sponsored by the National Science Foundation (NSF) and the CESM project is supported by NSF and the Office of Science Biological and Environmental Research of the U.S. Department of Energy (DOE). Computing resources were provided by the Climate Simulation Laboratory at NCAR's CISL, sponsored by the NSF and other agencies, and from the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. Special thanks to scientists in the Climate Change Research Section of NCAR/CGD for facilitating access to NERSC resources through their relationship with the Regional and Global Climate Modeling Program (RGCM) of the U.S. Department

of Energy's Office of Science (BER), cooperative agreement DE-FC02-97ER62402.

REFERENCES

Allan, R. and Ansell, T. (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *Journal of Climate*, 19, 5816–5842.

Anderson, J.L. (2001) An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129, 2884–2902.

Anderson, J.L. (2003) A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131, 634–642.

Anderson, J.L. (2009) Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, 61A, 72–83.

Anderson, J.L. and Collins, N. (2007) Scalable implementations of ensemble filter algorithms for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 24, 1452–1463.

Anderson, J.L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Avellano, A. (2009) The Data Assimilation Research Testbed: a community facility. *Bulletin of the American Meteorological Society*, 90, 1283–1296.

Balmaseda, M.A., Alves, O.J., Arribas, A., Awaji, T., Behringer, D.W., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T. and Stammer, D. (2009) Ocean initialization for seasonal forecasts. *Oceanography*, 22(3), 154–159. https://doi.org/10.5670/oceanog.2009.73.

Behringer, D. and Xue, Y. (2004) Evaluation of the global ocean data assimilation system at NCEP: the Pacific Ocean. In: *Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, AMS 84th Annual Meeting, Seattle, Washington, pp. 11–15

Behringer, D., Ji, M. and Leetmaa, A. (1998) An improved coupled model for ENSO prediction and implications for ocean initialization. Part 1: the ocean data assimilation system. *Monthly Weather Review*, 126, 1013–1021.

Bishop, C.H., Etherton, B.J. and Majumdar, S.J. (2001) Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: theoretical aspects. *Monthly Weather Review*, 129(3), 420–436.

Blanchard-Wigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., Posey, P., Wang, W., Zhang, J., Ardilouze, C., Bitz, C.M., Vernieres, G., Wallcraft, A. and Wang, M. (2017) Multi-model seasonal forecast of Arctic sea-ice: forecast uncertainty at pan-Arctic and regional scales. *Climate Dynamics*, 49, 1399–1410. https://doi.org/10.1007/s00382-016-3388-9.

Bonavita, M., Hólm, E.V., Isaksen, L. and Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142, 4865–4882. https://doi.org/doi:10.1002/qj.2652.

Buehner, M. (2005) Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, 131, 1013–1043.

Chang, Y.S., Zhang, S., Rosati, A., Delworth, T. and Stern, W. (2013) An assessment of the oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dynamics*, 40, 775–803. https://doi.org/10.1007/s00382-008-012-1412-2.

CISL (2012) *IBM iDataPlex System (Climate Simulation Laboratory)*. Boulder, CO: National Center for Atmospheric Research. Available at: http://n2t.net/ark:/85065/d7wd3xhc [Accessed 4th September 2018].

Compo, G., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R., Yin, X., Gleason, B., Vose, R., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley, S.J. (2011) The Twentieth Century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137, 1–28.

Counillon, F. and Bertino, L. (2009) Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. *Tellus A*, 61, 296–308.

Danabasoglu, G., Bates, S., Briegleb, B., Jayne, S.R., Jochum, M., Large, W., Peacock, S. and Yeager, S. (2012) The CCSM4 ocean component. *Journal of Climate*, 25, 1361–1389.

Derber, J. and Bouttier, F. (1999) A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A*, 51(2), 195–221. https://doi.org/10.1034/j.1600-0870.1999.t01-2-00003.x.

Derber, J. and Rosati, A. (1989) A global oceanic data assimilation system. *Journal of Physical Oceanography*, 19, 1333–1347.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99, 10143–10162.

Fisher, M., Nocedal, J., Tremolet, Y. and Wright, S. (2009) Data assimilation in weather forecasting: a case study in PDE-constrained optimization. *Optimization and Engineering*, 10, 409–426. https://doi.org/10.1007/s11081-008-9051-5.

Fisher, M., Tremolet, Y., Auvinen, H., Tan, D. and Poli, P. (2012) *Weak-Constraint and Long-Window 4D-Var*. Technical Memorandum 655. Reading, UK: ECMWF.

Furrer, R. and Bengtsson, T. (2007) Estimation of high-dimensional prior and posterior covariance matrices in filter variants. *Journal of Multivariate Analysis*, 98, 227–255.

Gaspari, G. and Cohn, S. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.

Good, S.A., Martin, M.J. and Rayner, N.A. (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research*, 118, 6704–6716.

Hamill, T.M. and Snyder, C. (2000) A hybrid ensemble Kalman filter–3D variational analysis scheme. *Monthly Weather Review*, 128, 2905–2919.

Hamill, T.M., Whitaker, J. and Snyder, C. (2001) Distance-dependent filtering of background covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129, 2776–2790.

Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *International Journal of Climatology*, 34, 623–642. https://doi.org/10.1002/joc.3711.

Holland, M.M., Bailey, D.A., Briegleb, B.P., Light, B. and Hunke, E. (2012) Improved sea ice shortwave radiation physics in CCSM4: the impact of melt ponds and aerosols on arctic sea ice. *Journal of Climate*, 25(5), 1413–1430. https://doi.org/10.1175/JCLI-D-11-00078.1.

Houtekamer, P.L. and Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1.

Hurrell, J., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J-F, Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M, Bader, D., Collins, W.D., Hack, J.J., Kiehl, J. and Marshall, S. (2013) The Community Earth System Model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1.

Johnson, D., Boyer, T., Garcia, H., Locarnini, R., Baranova, O. and Zweng, M. (2009). In: Levitus, S. (Ed.) *World Ocean Database 2009 Documentation*. Silver Spring, MD: NOAA. NODC Internal Report 20.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–471.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yange, S., Hnilo, J., Fiorino, M. and Potter, G. (2002) NCEP-DOE AMIP II reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83, 1631–1643.

Karspeck, A.R. (2016) An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean model. *Monthly Weather Review*, 144, 1713–1728. https://doi.org/10.1175/MWR-D-14-00336.1.

Karspeck, A.R., Yeager, S., Danabasoglu, G., Hoar, T., Collins, N., Raeder, K., Anderson, J. and Tribbia, J. (2013) An Ensemble Adjustment Kalman filter for the CCSM4 ocean component. *Journal of Climate*, 26, 7392–7413.

Karspeck, A.R., Yeager, S., Danabasoglu, G. and Teng, H. (2015) An evaluation of experimental decadal predictions using CCSM4. *Climate Dynamics*, 44, 907. https://doi.org/10.1007/s00382-014-2212-7.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J.M., Bates, S.C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L. and Vertenstein, M. (2015) The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1.

Kennedy, J.J., Rayner, N.A., Smith, R.O., Parker, D.E. and Saunby, M. (2011) Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research*, 116, D14104. https://doi.org/10.1029/2010JD015220.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L.I., Paolino, D.A., Zhang, Q., van den Dool, H., Saha, S., Pena Mendez, M., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D.G., Tippett, M.K., Barnston, A.G., Li, S., Rosati, A., Schubert, S.D., Rienecker, M., Suarez, M., Li, Z.E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B. and Wood, E.F. (2014) The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95, 585–601.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R. and Fiorino, M. (2001) The NCEP-NCAR 50-year reanalysis: monthly mean CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 82, 247–267.

Kleist, D., Parrish, D., Derber, J., Treadon, R., Errico, R. and Yang, R. (2009) Improving incremental balance in the GSI 3D-Var analysis system. *Monthly Weather Review*, 137, 1046–1060.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93, 5–48. https://doi.org/10.2151/jmsj.2015-001.

Kocin, P. and Uccellini, L. (1990) *Snowstorms Along the Northeastern Coast of the United States: 1955 to 1985*. Meteorological Monographs, No. 44. Boston, MA: American Meteorological Society.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K. and Janssen, P. (2016a) A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142, 65–78.

Laloyaux, P., de Boisseson, E. and Dahlgren, P. (2016b) CERA-20C: an Earth system approach to climate reanalysis. *ECMWF Newsletter*, 150. https://doi.org/10.21957/ffs36birj2.

Lawrence, M. and Pelissier, J. (1981) Atlantic hurricane season of 1980. *Monthly Weather Review*, 109, 1567–1582.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Thornton, P.E., Swenson, S.C., Lawrence, P.J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G.B. and Slater, A.G. (2011) Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, 3, M03001. https://doi.org/10.1029/2011MS00045.

Lea, D., Mirouze, I., Martin, M., King, R., Hines, A., Walters, D. and Thurlow, M. (2015) Assessing a new coupled data assimilation system based on the Met Office Coupled Atmosphere–Land–Ocean–Sea-Ice model. *Monthly Weather Review*, 143, 4678–4694.

Levitus, S. (1982) *Climatological Atlas of the World Ocean*. NOAA Professional Paper 13. Washington, DC: NOAA.

Li, H. and Shriver, R.L. (2016) Effects of ocean grid resolution on tropical cyclone-induced upper ocean responses using a global ocean general circulation model. *Journal of Geophysical Research; Oceans*, 121, 8305–8319. https://doi.org/10.1002/2016JC011951.

Liao, J., Gerofi, B., Lien, G.-Y., Myoshi, T., Nishizawa, S., Tomita, H., Liao, W.-K., Choudhary, A. and Ishikawa, Y. (2017) A flexible I/O arbitration framework for netCDF-based big data processing workflows on high-end supercomputers. *Concurrency and Computation Practice and Experience*, 29(15), 1532–0626. https://doi.org/10.1002/cpe.4161.

Locarnini, R., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Baranova, O.K., Zweng, M.M., Paver, C.R., Reagan, J.R., Johnson, D.R., Hamilton, M. and Seidov, D. (2013a) *World Ocean Atlas 2013, Volume 1: Temperature*. Levitus, S, Mishonov, A (eds), NOAA Atlas NESDIS 73, Washington, DC.

Meehl, G., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A.R., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G.J., Vecchi, G. and Yeager, S. (2014) Decadal climate prediction: an update from the trenches. *Bulletin of the American Meteorological Society*, 95, 243–267.

Mitchell, H. and Houtekamer, P. (2000) An adaptive ensemble Kalman filter. *Monthly Weather Review*, 128, 416-433. https://doi.org/10.1175/1520-0493(2000).

Miyoshi, T., Kondo, K. and Imamura, T. (2014) The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophysical Research Letter*, 41, 5264–5271. https://doi.org/10.1002/2014GL060863.

Mogensen, K., Molteni, M.B.R. and Weaver, A.T. (2012) *The NEMOVAR Ocean Data Assimilation System as implemented in the EWMWF Ocean Analysis System for System 4*. Technical Memorandum 668. Reading: ECMWF.

Mulholland, D., Laloyaux, P., Haines, K. and Balmaseda, M. (2015) Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, 143, 4631–4644.

Neale, R.B., Chen, C.-C., Gettelman, A., Lauritzen, P.H., Park, S., Williamson, D.L., Conley, A.J., Garcia, R., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M., Smith, A.K., Tilmes, S., Vitt, F., Morrison, H., Cameron-Smith, P., Collins, W.D., Iacono, M.J., Easter, R.C., Ghan, S.J., Liu, X., Rasch, P.J. and Taylor, M.A. (2012) *Description of the NCAR Community Atmosphere Model (CAM 5.0)*. Technical Note TN-486+STR. NCAR, Boulder, CO.

Neale, R.B., Richter, J., Park, S., Lauritzen, P.H., Vavrus, S., Rasch, P.J. and Zhang, M. (2013) The mean climate of the Community Atmospheric Model (CAM4) in forced sst and fully coupled experiments. *Journal of Climate*, 26, 5150–5168. https://doi.org/10.1175/JCLI-D-12-00236.1.

Nerger, L. and Hiller, W. (2013) Software for ensemble-based data assimilation systems - implementation strategies and scalability. *Computers and Geosciences*, 55, 110–118. https://doi.org/10.1016/j.cageo.2012.03.026.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D. and Kosro, P.M. (2002) Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research*, 107(C9). https://doi.org/10.1029/2000JC000511.

Oke, P.R., Brassington, G.B., Griffin, D.A. and Schiller, A. (2008) The Bluelink ocean data assimilation system (BODAS). *Ocean Modelling*, 21, 46–70.

Penny, S. (2014) The hybrid local ensemble transform Kalman filter. *Monthly Weather Review*, 142, 2139–2149.

Penny, S. and Hamill, T. (2017) Coupled data assimilation for integrated Earth system analysis and prediction. *Bulletin of the American Meteorological Society*, 98, ES169–ES172. https://doi.org/10.1175/BAMS-D-17-0036.1.

Price, J.F. (1981) Upper ocean response to a hurricane. *Journal of Physical Oceanography*, 11, 153–175.

Raeder, K., Anderson, J.L., Collins, N., Hoar, T.J., Kay, J.E., Lauritzen, P.H. and Pincus, R. (2012) DART/CAM: an ensemble data assimilation system for CESM atmospheric models. *Journal of Climate*, 25, 6304–6317. https://doi.org/10.1175/JCLI--D--11--00395.1.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. and Kaplan, A. (2003) Global analyses of sea surface temperature, sea ice and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14). https://doi.org/10.1029/2002JD002670.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G. (2007) Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20, 5473–5496.

Saha, S., Moorthi, S., Pan, H.L., Wu, X.i., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G. and Goldberg, M. (2010) The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91, 1015–1057. https://doi.org/10.1175/2010BAMS3001.1.

Shaw, M.S., Hopkins, J.S. and Caton, P.G.F. (1976) The gales of 2nd January 1976. *Weather*, 31, 172–183.

Sluka, T., Penny, S., Kalnay, E. and Miyoshi, T. (2016) Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical Research Letters*, 43, 752–759. https://doi.org/10.1002/2015GL067238.

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., Hecht, M., Jayne, S., Jochum, M., Large, W., Lindsay, K., Maltrud, M., Norton, N., Peacock, S., Vertenstein, M. and Yeager, S. (2010) *The Parallel Ocean Program (POP) Reference Manual, Ocean Component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM)*, Technical Report LAUR-10-01853, National Laboratory, Los Alamos, NM. Available at:

http://www.cesm.ucar.edu/models/cesm1.0/pop2/doc/sci/POPRefManual.pdf; accessed 5 September 2018

Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., Igarashi, H. and Iskikawa, Y. (2008) Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variation. *Journal of Geophysical Research*, 103(C10017). https://doi.org/10.1029/2008JC004741.

Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of the CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498. https://doi.org/10.1175/BAMS--D--11--00094.1.

Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M. and Whitaker, J.S. (2003) Ensemble square root filters. *Monthly Weather Review*, 131(7), 1485–1490.

Tshimanga, J., Gratton, S., Weaver, A.T. and Sartenaer, A. (2008) Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 134, 751–769.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosazac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E.V., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.A., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. (2005) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961–3012.

Whitaker, J.S. and Hamill, T.M. (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140, 3078–3089.

Woodruff, S.D., Worley, S.J., Lubker, S.J., Ji, Z., Freeman, J.E., Berry, D.I., Brohan, P., Kent, E.C., Reynolds, R.W., Smith, S.R. and Wilkinson, C. (2011) ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, 31(CLIMAR-III Special Issue), 951–967.

Yeager, S., Karspeck, A.R., Danabasoglu, G., Tribbia, J. and Teng, H. (2012) A decadal prediction case study: late 20th century North Atlantic Ocean heat content. *Journal of Climate*, 25, 5173–5189.

Zarzycki, C. and Jablonowski, C. (2015) Experimental tropical cyclone forecasts using a variable-resolution global model. *Monthly Weather Review*, 143, 4012–4037.

Zhang, S. (2011) A study of impacts of coupled model initial shocks and state-parameter optimization on climate predictions using a simple pycnocline prediction model. *Journal of Climate*, 24, 6210–6226.

Zhang, S., Harrison, M., Rosati, A. and Wittenberg, A. (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, 135, 3541–3564.

Zhang, Y.F., Hoar, T., Yang, Z.L., Anderson, J., Toure, A. and Rodell, M. (2014) Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4. *Journal of Geophysical Research; Atmospheres*, 119, 7091–7103. https://doi.org/10.1002/2013JD021329.

Zuo, H., Balmaseda, M.A. and Mogensen, K. (2015) The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 49, 791–811. https://doi.org/10.1007/s00382-015-2675-1.

Zweng, M.M., Reagan, J.R., Antonov, J.I., Locarnini, R.A., Mishonov, A.V., Boyer, T.P., Garcia, H.E., Baranova, O.K., Johnson, D.R., Seidov, D. and Biddle, M.M. (2013b). In: Levitus, S. and Mishonov, A. (Eds.) *World Ocean Atlas 2013, Volume 2: Salinity*. Washington, DC: NOAA atlas NESDIS 74. 39 pp.

How to cite this article: Karspeck AR, Danabasoglu G, Anderson J, *et al*. A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. *Q J R Meteorol Soc.* 2018;144:2404–2430. https://doi.org/10.1002/qj.3308