CAFE60v1: A 60-Year Large Ensemble Climate Reanalysis. Part I: System Design, Model Configuration, and Data Assimilation

TERENCE J. O'KANE,^a PAUL A. SANDERY,^a VASSILI KITSIOS,^{b,c} PAVEL SAKOV,^d MATTHEW A. CHAMBERLAIN,^a MARK A. COLLIER,^b RUSSELL FIEDLER,^a THOMAS S. MOORE,^a CHRISTOPHER C. CHAPMAN,^a BERNADETTE M. SLOYAN,^a AND RICHARD J. MATEAR^a

^aCSIRO Oceans and Atmosphere, Hobart, Tasmania, Australia

^bCSIRO Oceans and Atmosphere, Aspendale, Victoria, Australia

^cLaboratory for Turbulence Research in Aerospace and Combustion, Department of Mechanical and Aerospace Engineering Monash University, Clayton, Victoria, Australia

^dBureau of Meteorology, Docklands, Victoria, Australia

(Manuscript received 17 December 2020, in final form 8 March 2021)

ABSTRACT: We detail the system design, model configuration, and data assimilation evaluation for the CSIRO Climate retrospective Analysis and Forecast Ensemble system, version 1 (CAFE60v1). CAFE60v1 has been designed with the intention of simultaneously generating both initial conditions for multiyear climate forecasts and a large ensemble retrospective analysis of the global climate system from 1960 to the present. Strongly coupled data assimilation (SCDA) is implemented via an ensemble transform Kalman filter in order to constrain a general circulation climate model to observations. Satellite (altimetry, sea surface temperature, sea ice concentration) and in situ ocean temperature and salinity profiles are directly assimilated each month, whereas atmospheric observations are subsampled from the JRA-55 atmospheric reanalysis. Strong coupling is implemented via explicit cross-domain covariances between ocean, atmosphere, sea ice, and ocean biogeochemistry. Atmospheric and surface ocean fields are available at daily resolution and monthly resolution for the land, subsurface ocean, and sea ice. The system produces 96 climate trajectories (state estimates) over the most recent six decades as well as a complete data archive of initial conditions, potentially enabling individual forecasts for all members each month over the 60-yr period. The size of the ensemble and application of strongly coupled data assimilation lead to new insights for future reanalyses.

KEYWORDS: Atmosphere; Ocean; Climate prediction; Ensembles; Data assimilation

1. Introduction

Understanding climate variability over recent decades can be undertaken in a number of ways. Observations of the atmosphere, oceans, and sea ice may be gathered, quality controlled, and interpolated through space and time. However, globally heterogeneous observing systems with sufficient spatial and temporal density to provide accurate estimates of the evolution of the major climate teleconnections have only been in existence in the most recent decades. One of the aims of this work is to assess the usefulness of strongly coupled data assimilation (SCDA) where cross-domain covariances act as an additional constraint during periods where one or more domains are poorly observed.

Optimal interpolation (OI) of sparse observational data is a common approach for state estimation (Jazwinski 1970). Application of formal data assimilation (DA) methods can be

employed combining available observations over a specified temporal window and short-term forecasts that provide the so-called background states with information provided in the data void areas. Alternately, OI methods can be also employed whereby static or climatological background covariances are determined from prior independent model simulations (Smith and Murphy 2007; Oke et al. 2013). Data assimilation methods can be regarded as generalized filters that extract the signal from observations and produce analyzed states at required time intervals and with complete coverage at the required resolution (Houtekamer and Zhang 2006; O'Kane and Frederiksen 2008).

Systematic errors with respect to the observed mean climatological state are referred to as model biases. However, this relies on two assumptions: first, that there are sufficient observations to accurately estimate the climatological state over a given period, and second, that any changes in the climatological state will arise largely as a response to changes in the (radiative) forcing. In particular, it is well known that temperature and salinity observations of the subsurface ocean (0–2000 m) have only reached sufficient density since the Argo program became fully established (mid-2000s), enabling ocean general circulation models to be effectively constrained using state-of-the-art data assimilation systems and without recourse to restoring to climatology (O'Kane et al. 2019) or other forms of bias correction (Carton et al. 2018). Additional challenges arise in obtaining sufficient forecast samples for the accurate estimation of flow-dependent background covariances. Where observations assimilated from one domain (e.g., ocean) have

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-20-0974.s1>.

Corresponding author: Terence J. O'Kane, terence.okane@csiro.au

DOI: 10.1175/JCLI-D-20-0974.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

no influence on the state of any other domain (e.g., atmosphere), the procedure is termed weakly coupled DA as described in Zhang et al. (2007). In SCDA, in addition to covariances specific to each domain of the Earth system, cross-domain covariances are also estimated, thus representing a technical and computational challenge.

While there are a number of recent global atmospheric (Kalnay et al. 1996; Compo et al. 2011; Hersbach et al. 2019; Gelaro et al. 2017; Kobayashi et al. 2015) and ocean (Smith and Murphy 2007; Penny et al. 2015; Forget et al. 2015; Carton et al. 2018; Oke et al. 2013; Balmaseda et al. 2013) reanalyses, there are still only a relatively few coupled ocean–atmosphere reanalysis products (Laloyaux et al. 2018; Saha et al. 2010) available. Both the CERA-20C (Laloyaux et al. 2018) (1901–2010) and NCEP-CFSR (Saha et al. 2010) (1979–present) coupled reanalyses have sophisticated atmospheric data assimilation schemes with assimilation windows of a few hours and weak coupling (no cross-domain covariance) to the ocean. Of the two, only the CERA-20C reanalysis provides an ensemble of state estimates (10 members) with which to account for observational and model errors and to provide some estimate of the uncertainties. As noted previously (Laloyaux et al. 2018), climate state estimation is a tremendous scientific challenge due to the sparse atmospheric observing system prior to the radiosonde (pre-1930s) and satellite (pre-1970s) eras (Stickler et al. 2014). The situation is even more problematic in the ocean given the paucity of routine subsurface observations prior to the Argo era.¹

In addition to state estimation, reanalyses may also provide initial conditions for prediction. By definition, the goal of near-term climate prediction (NTCP) (Kushnir et al. 2019) is to produce a skillful forecast of the evolving climate system over 1–10 years. In common with seasonal forecasts, accurate prediction of the internally generated climatic variability requires that the system be initialized close to the observed climatic state. The initial model state can be imposed; however, more sophisticated approaches occur where the observations are assimilated to construct an analyzed state that then serves as the forecast (forward) model's initial condition. As near-term predictions extend well beyond the time scales where one might reasonably expect prediction skill from initial conditions and into the time scales where external forcing begins to dominate, NTCP systems are required to use both the present and projected anthropogenic radiative forcing in much the same way as climate projections do. While reanalyses are generally conducted to provide reconstructions of the past climate, recent advances toward operational near-term climate prediction have shown that observation-based initialization of coupled general circulation model (CGCM) predictions of the last half-century can lead to significant enhancement of predictive skill for particular variables on time scales from a year to a decade (Smith et al. 2007, 2013; Doblas-Reyes et al. 2013; Saha et al. 2014; Kushnir et al. 2019; Smith et al. 2020).

Here we describe the CSIRO Climate retrospective Analysis and Forecast Ensemble System, version 1 (CAFE60v1). CAFE60v1 was developed to fulfil three key purposes: 1) to provide a large ensemble of state estimates of sufficient number and appropriate spatiotemporal resolution that the evolving probability distribution conditioned on the available observations over the recent six decades might be estimated; 2) to provide a large ensemble of balanced initial conditions for retrospective forecasts (hindcasts) and initialized forecasts specifically focused beyond weather and seasonal time scales for the near term-climate over the coming decade; and 3) to explore the utility of strongly coupled data assimilation to better constrain data void areas in one domain through cross-domain covariances with a better observed domain. As it has been established that assimilation of ocean observations leads to better balanced initial conditions for seasonal forecasts (Alves et al. 2004), it is expected that strongly coupled DA has the further potential to produce analyzed states with improved balance for climate forecast initial conditions.

The paper is structured in the following way. The model configuration and forcing is described in section 2, followed by the description of the coupled data assimilation methodology and observations assimilated (sections 3 and 4 respectively). Evaluation of the assimilation system is described in section 5. Comments on the constructed initial conditions for forecasts (section 6) and future development (section 7) follow.

2. Model

The CAFE60v1 climate model configuration is based on the Geophysical Fluid Dynamics Laboratory's (GFDL) Climate Model 2.1 (CM2.1) (Delworth et al. 2006). The particular configuration used in the CAFE60v1 system has been described previously (O'Kane et al. 2019; Sandery et al. 2020) and is only briefly described here. The ocean model configuration uses the 1° ocean grid as described by Bi et al. (2013). The ocean model is coupled to the land, atmosphere, and sea ice components from CM2.1, namely, Land Model 2 (LM2), Atmospheric Model 2 (AM2), and Sea Ice Simulator (SIS) respectively.

The nominal resolution of the CAFE60v1 ocean component (Modular Ocean Model, version 4.1) is 1°, with extra latitudinal resolution in the tropics, 0.33° at the equator, with extra horizontal resolution in the Southern Ocean, corresponding to 0.25° at 75°S. There are 50 vertical levels, with a 10-m resolution in the upper ocean, increasing to ~300 m at depth. The grid is tripolar over the Arctic, north of 65°N, to avoid the North Pole singularity. Subgrid processes for the CAFE60v1 ocean component are adopted from CM2.1, including neutral physics (Redi diffusivity and Gent–McWilliams skew diffusion; Griffies 2009), Brian–Lewis vertical mixing profile, a Lagrangian friction scheme, and a *K*-profile parameterization for the mixed layer calculation. Biases in the mode water structure and deep open ocean convection are reduced by restoring the ocean temperature and salinity below 2000 m to climatology based on World Ocean Atlas observations with a 1-yr time scale. Deep restoring further accounts for poor data coverage at depth. The

¹ See <http://www.argo.ucsd.edu/>.

sea ice model (SIS) is on the ocean grid and uses five ice thickness categories.

The atmospheric model (AM2) has a resolution of 2° in latitude and 2.5° longitude, and 24 hybrid (sigma-pressure or terrain-following pressure) vertical levels. The land model LM2 is on the same horizontal grid as the atmosphere. Concentrations of atmospheric aerosols and radiative gases, and land cover are based on the radiative forcing used over the historical period 1960–2006 and are as for the GFDL CM2.1 submissions to the Coupled Model Intercomparison Project (CMIP), phases 3 and 5 (Zhang et al. 2017), and supplemented with RCP4.5 forcing data supplied by GFDL. The data were extended beyond 2006 using RCP4.5 forcings for all the major radiative gases (i.e., CO_2 , CH_4 , N_2O , etc.) and aerosols. The only fields that change at different dates are volcanic sulfate aerosols, stratospheric O_3 , and ocean CO_2 used to estimate ocean carbon. Volcanic emissions post 2000 were based on a “neutral” year. Ocean CO_2 was based on data used in the ACCESS-ESM1 CMIP6 contribution. One specific point of difference is that we have included the use of spatially heterogeneous stratospheric ozone forcing from the CMIP6 archive which was verified over various assimilation cycles to produce lower RMS errors than zonal mean observed O_3 .

3. Data assimilation methodology

By choosing to implement SCDA with the ensemble Kalman filter (EnKF) framework, we are required to produce a sufficiently large background ensemble of model states such that the cross-domain covariances might be estimated with some fidelity. Previous investigation by Sandery et al. (2020) indicated that on the order of 100 members were required. The motivation to sample the climate probability distribution function and produce balanced initial conditions for multiyear forecasts combined with available computational resources were further key factors in configuring the CAFE60v1 data assimilation system.

A severe challenge for the CAFE60v1 reanalysis is the representation of weather variability and in particular the synoptic midlatitude troposphere. The basic premise we have employed in CAFE60v1 is that monthly mean atmospheric increments consistent with surface temperature gradients and the ocean state can be used to constrain the large scales of the atmosphere, for example the jets and cells (e.g., Hadley, Walker, etc.), and that the statistics of the daily synoptic features during any given month will be generated by the model dynamics consistent with the adjusted large-scale structures. While one should not expect to capture the details of the weather on any particular day correctly, it is expected that we capture the trends and low-frequency variability of the various teleconnection modes and statistics of the occurrence and frequency of persistent coherent synoptic features with a degree of fidelity. Earlier variants of the CAFE data assimilation system, where only the large-scale structures of the atmosphere were constrained by cross-domain covariances from ocean observations (O'Kane et al. 2019; Sandery et al. 2020), have demonstrated that this is a reasonable assumption. Next we present a complete description of the specific SCDA system used to generate the CAFE60v1 reanalysis.

Ensemble Kalman filter

Data assimilation (DA) for large numerical prediction problems in the ocean and atmosphere requires calculating the background error covariances. In the ensemble Kalman filter formalism this can be achieved either in terms of static covariances estimated as anomalies or deviations from the climatological mean calculated from a long (detrended) control simulation, the so-called ensemble optimal interpolation (EnOI) (Evensen 2003; Oke et al. 2013; O'Kane et al. 2014, 2019; Castruccio et al. 2020). Alternately, the time-evolving background covariance can be estimated via propagating an ensemble of model states and constructing the forecast (background) error covariance in terms of deviations from the first moment or ensemble mean at a particular time (Bishop et al. 2001; Sakov and Sandery 2015; Sandery et al. 2020). This comes at the additional computational cost of running an ensemble of forward models and is the approach taken here.

We apply the ensemble transform Kalman filter (ETKF) methodology (Bishop et al. 2001) to a system with model dynamics given by Φ and considering an n -dimensional state vector $\mathbf{x}$ at time step t , and a p -dimensional observation vector $\mathbf{d}$ (here n and $p \in \mathbb{N}$ natural numbers); the analysis and forecast fields are defined as $\mathbf{x}_i^a$ and $\mathbf{x}_i^f$, where $i = 1, 2, \dots, k$ runs over the entire ensemble (i.e., $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_k]$), and angle brackets $\langle \rangle$ indicate the ensemble mean.

The ETKF propagates an ensemble of k anomalies from which the first and second moments of $\mathbf{x}$ can be calculated recursively. The background and analysis covariances are defined in terms of the anomalies $\mathbf{z}_i$ for $i = 1, 2, \dots, k$ as

$$\mathbf{Z}^f = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^f, \mathbf{z}_2^f, \dots, \mathbf{z}_k^f], \quad (1a)$$

$$\mathbf{Z}^a = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^a, \mathbf{z}_2^a, \dots, \mathbf{z}_k^a], \quad (1b)$$

and where the state vectors are $\mathbf{z}_i^f = \mathbf{x}_i^f - \langle \mathbf{x}^f \rangle$ and $\mathbf{z}_i^a = \mathbf{x}_i^a - \langle \mathbf{x}^a \rangle$, which are n -dimensional in model space. The ETKF acts to choose appropriate initial forecast anomalies consistent with error covariance update equations within the vector subspace of ensemble anomalies [Eq. (2a)] formed as $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$ and $\mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}$ and where the covariance update is given by

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{D})^{-1} \mathbf{H} \mathbf{P}^f = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}^f, \quad (2a)$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{D})^{-1}, \quad (2b)$$

where $\mathbf{K}$ is the $n \times p$ Kalman gain, $\mathbf{P}$ is the positive definite state covariance error matrix, $\mathbf{I}$ is the identity matrix, $\mathbf{D}$ is the measurement error covariance, and $\mathbf{H}$ is the $p \times n$ linearized observational operator mapping forecast grid point values onto observational points. The innovation is defined as $[\mathbf{d}_i - \mathbf{H}(\langle \mathbf{x}^f \rangle)]$ where $\mathbf{d}_i$ are measurements in observation space.

We calculate the matrix of ensemble perturbations in normalized observation space, that is,

$$\mathbf{E} = (\mathbf{D}^{-1/2} \mathbf{H} \mathbf{Z}^f)^T (\mathbf{D}^{-1/2} \mathbf{H} \mathbf{Z}^f), \quad (3)$$

where $\mathbf{H}$ maps from $\mathbf{Z}^f$ into observable space and $\mathbf{D}^{1/2}$ does the normalization.

We are next required to find the eigenvectors $\mathbf{C}$ and eigenvalues Γ of Eq. (3). The transform matrix $\mathbf{T}$ is now defined in terms of the $k \times (k - 1)$ matrix of nonzero eigenvalues such that

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{1/2} \mathbf{C}^T, \quad (4)$$

where the Γ (nonzero eigenvalues) is $(k - 1) \times (k - 1)$ and $\mathbf{C}$ is $k \times (k - 1)$. Equation (4) corresponds to the transform matrix in spherical simplex form as described by Wang et al. (2004). The ETKF generates new analysis perturbations as

$$\mathbf{z}^a = \mathbf{z}^f \mathbf{T}, \quad (5)$$

where the transform matrix $\mathbf{T}$ determines how the perturbations from different members are mixed. We directly initialize the full state vector to the analysis

$$\langle \mathbf{x}^a \rangle = \langle \mathbf{x}^f \rangle + \mathbf{K}[\mathbf{d} - \mathbf{H}(\langle \mathbf{x}^f \rangle)], \quad (6a)$$

$$\mathbf{x}_i^a = \langle \mathbf{x}^a \rangle + \sqrt{k - 1} \mathbf{z}^f \mathbf{T}, \quad (6b)$$

whereby model biases are reduced by constraining the model to be close to observations and where the analysis fields are calculated for each realization i . All state variables, regardless of domain, are initialized to their analyzed fields without application of incremental analysis updating.

The EnKF-C software (Sakov 2018) is used to apply the ETKF. EnKF-C is designed to handle model states defined on multiple grids, thus making it particularly suitable for strongly coupled systems. For each model grid EnKF-C calculates arrays of local ensemble transforms defined on a horizontal subgrid with specified stride, and using linear interpolation to obtain the transform on a given grid node. Inflation is employed in order to avoid possible systematic underestimation of analysis error covariances leading to filter divergence. Applying spatially uniform ensemble inflation can include areas where no local observations are present and hence where no assimilation is conducted, leading to the gradual injection of energy into the model and a corresponding deterioration in the performance of the DA system over time. Even in the presence of local observations, where there occurs a lack of correlation between particular state elements updated with the same transforms, the ensemble spread for some elements may hardly reduce after assimilation, yet the ensemble anomalies will be inflated. To avoid this behavior, EnKF-C caps the inflation by a specified amount, which means that for a given model element it is limited by the magnitude of the reduction of ensemble spread in the analysis: $\lambda = \min(\lambda_0, \sigma_f/\sigma_a)$ (Sakov 2018, section 2.8.1 therein), where σ_f is the forecast ensemble spread, and σ_a is the analysis ensemble spread. The term λ_0 is the specified inflation coefficient; here $\lambda_0 = 1.05$.

CAFE60v1 uses a calendar month data assimilation cycle length and an ensemble size of 96 members. The state vector contains ocean temperature, salinity, sea level anomaly, ocean velocity, ocean biogeochemistry (e.g., dissolved inorganic carbon, alkalinity, phosphate, oxygen), sea ice concentration thickness categories, atmospheric pressure, temperature, humidity, and winds. The model sea level anomaly (SLA) is the

difference between model sea surface height and mean sea level from a 100-yr control run (i.e., the model mean dynamic topography).

All assimilated analysis gridded products—e.g., JRA-55 (Kobayashi et al. 2015), HadISST (Rayner et al. 2003), OISST (Huang et al. 2020), and Ocean and Sea Ice Satellite Application Facility (OSISAF) (Laverne et al. 2019)—are preprocessed to calendar month averages and assimilated synchronously. For all heterogeneous observations a calendar month window is applied to calculate superobservations. In practice, this means that the increment $\mathbf{K}[\mathbf{d} - \mathbf{H}(\langle \mathbf{x}^f \rangle)]$ in Eq. (6a) is calculated as $\mathbf{K}[\bar{\mathbf{d}} - \mathbf{H}(\langle \bar{\mathbf{x}}^f \rangle)]$, where $\langle \bar{\mathbf{x}}^f \rangle$ is the ensemble average monthly mean forecast, and $\bar{\mathbf{d}}$ represents the monthly mean observations.

In the usual way, spurious long-range correlations in the background are removed via horizontal localization function (Gaspari and Cohn 1999) with support radius of 1000 km for all ocean and atmospheric observations. A reduced length scale of 250 km is used for sea ice temperature and concentration observations. Localization radii were determined via experimentation and on the basis of the number of superobservations assimilated per cycle. Specifically the assimilation uses monthly mean superobservations and monthly mean background states such that the localization length scales between superobservations and background covariances are consistent. When assimilating dense gridded atmospheric data from JRA-55, localization length scales of between 500 and 1000 km were determined to give similar forecast innovation deviations. As we assimilate all observations in parallel rather than sequentially, choices of larger localization radii required additional thinning of the observations in order to avoid exceeding memory limits. Localization removes spurious long-range correlations present in the ensemble estimated background error covariance, and has also been shown to improve the consistency between ensemble spread and error. Disadvantages to horizontal covariance localization have been reported with regard to weather prediction, where localization can remove information due to large-scale flow dependent inhomogeneities, creating potential imbalances (Kepert 2009; Žagar et al. 2010); however, the large length scales used in CAFE60v1 mitigate this effect. When only sparsely distributed observational data that are heterogeneous in both space and time are available, such as is the case for the ocean in the early period of the 1960s and prior to Argo, the specified radius of influence can often lead to little observational influence on remote but data-sparse regions. Smith and Murphy (2007) demonstrated that global covariances sampled from a model simulation (i.e., pseudo-observations extracted from the model data at the locations of real observations) can be substituted, giving an effective global in-filling that has proven to be of utility; however, this has not been tested here. No vertical localization is employed in CAFE60v1.

Prior to 1992, sea surface temperature (SST) is bias corrected using the method of Evensen (2003, section 4.2.2 therein) (see also O'Kane et al. 2019; Sandery et al. 2020). To estimate the SST bias field, an ensemble of bias fields is initialized to independently identically distributed random spatially uniform values and the observed SST is then assumed to be the

sum of the model SST and unknown bias. This field is then updated similar to other model fields and the resulting ensemble of biases is evolved using a first-order autoregressive [AR(1)] function: $x_b^f = \lambda x_b^a + (1 - \lambda^2)^{1/2} \sigma$. Here λ is a forgetting factor, $0 < \lambda < 1$, and $\sigma \sim \sigma_0 N(0, 1)$, where σ_0 is the standard deviation of the bias state variable x_b for SST $\lambda = 0.95$, and $\sigma = 0.8$. Further, the SST bias influences subsurface temperature in the mixed layer, and therefore affects innovations from subsurface observations. Bias correction in this form can only detect and correct systematic errors in the initial conditions. Due to the relatively long assimilation window (calendar month), model SST bias was found to grow appreciably with lead time in particular regions and very specifically the South Pacific Ocean during the austral summer. Application of weak (90-day time scale) restoring to an SST climatology constructed from HadISST was applied to the forward model post 1992 to ameliorate the biases replacing the method of Evensen (2003).

Sea ice assimilation is applied via the five thickness categories in the SIS ice model, which are strongly coupled to the ocean via the cross-domain covariances. Total sea ice area fraction or concentration is calculated by summing the thickness categories and is used as the background concentration. To ensure values at all times remain within physical bounds (i.e., between 0 and 1) the analyzed thickness categories are normalized by the analyzed concentration. To avoid representativeness error due to large uncertainties near the ice edge we apply a prescribed ice-concentration error estimate of the form $\sigma^2 = 0.01 + (0.5 - ||0.5 - c||)^2$ following Sakov et al. (2012). This approach targets regions where the largest uncertainties in observed ice concentrations c occur (i.e., in regions where cover is around 50%).

The ensemble of atmospheric models is on hybrid sigma-pressure levels whose vertical grid positions vary from member to member according to their respective surface pressure fields. To assimilate the atmospheric data from the JRA-55 reanalysis into the respective forward models, we found it necessary to regrid all ensemble members onto a common grid to calculate the covariances. The common vertical grid used was defined using a reference daily mean surface pressure from JRA-55 corresponding to analysis time. Initialization of the atmospheric ensemble states was done by regridding the analyzed fields back to the hybrid sigma-pressure levels of the atmospheric model using the analyzed surface pressure for each member.

The CAFE60v1 system, and more specifically EnKF-C, adaptively moderates the impact of observations of various types using two parameters, the so-called R and K factors (Sakov and Sandery 2017). The K factor, here equal to 1, limits the impact of individual observations (i.e., those with large innovations likely to be inconsistent with the state of the DA system) by smoothly increasing observation error variance depending on the projected increment such that the resulting increment to generate the analysis does not exceed the estimated state error times K . A further benefit of the method is that it minimizes the innovations by modifying the distribution of observation error to better match the distribution of model error, thereby increasing the gain and the observation impact. In EnKF-C, R factors are defined for each observation type and

represent scaling coefficients for the corresponding observation error variances. Increasing the R factor decreases the impact of particular observation types. (Sandery et al. 2020).

CAFE60v1 uses SCDA where, in addition to the projection of oceanic, atmospheric, and sea ice observations onto their respective domains, specific cross-domain covariances are included. In particular, oceanic observations are projected onto the atmospheric domain via the atmosphere–ocean cross-covariances and onto the sea ice domain via the sea ice–ocean cross-domain covariances. Sea ice temperatures are projected back onto the ocean and atmosphere; however, projection of atmospheric observations (O’Kane et al. 2019; Sandery et al. 2020) onto the ocean domain was found to degrade the assimilation and is not used in CAFE60v1. We do not assimilate any ocean biogeochemistry (OBGC) observations; rather, the cross-domain covariance from the available ocean observations projected onto the OBGC background is used to apply an increment to the OBGC background state vector as a constraint. No land observations are assimilated nor are there any explicit cross-domain covariances relevant to the land model.

4. Observations

Table 1 describes assimilated observations. Before being assimilated, the observations are converted into superobservations by combining all observations of a given type and specific product falling within model grid cells over the time window, with known error estimates, into one superobservation with a smaller error estimate. The superobservation location, values, and error estimates are based on a weighted average using inverse error variances of the original observations. Assigned observation errors are also given in Table 1. As stated earlier, CAFE60v1 does not include assimilation of any land observations.

a. Altimetry

Altimetric SLA is assimilated from September 1991 onward. SLA is provided by the Radar Altimeter Database System (RADS) (Naeije et al. 2000), which includes tide, mean dynamic topography (MDT), and inverse barometer corrections. SLA observations are limited to water depths greater than 200 m due to the small signal-to-noise ratio on the shelf and coastal regions. Given that CAFE60v1 ingests SLA superobservations based on monthly mean averages, these observations are weighted (large R factor; see Table 1) such that the main influence is to constrain the global sea level rather than local features.

b. Sea surface temperature

Prior to the satellite era, SST is assimilated from gridded HadISST over the period January 1960 through February 2004. Once available, remotely sensed SST is assimilated from a range of products: AVHRR from February 2006, AMSR-E from June 2002, AMSR-2 from July 2014, VIIRS from June 2015, and WindSat (Gaiser et al. 2004) between March 2003 and June 2009. For a brief period between January 1988 and August 1990, OISST and OSISAF data were assimilated but led to increased bias and forecast RMS errors in air temperature

TABLE 1. Observation type and error estimates; [C] refers to concentration; daggers (†) indicate error estimates provided by the vendor.

Obs type	Abbreviation	Dataset	Spatial distribution	Error estimate	<i>R</i> factor
Sea surface temperature	SST	HadISST, OISST, AVHRR, AMSR-E, AMSR-2, VIIRS, WindSat	Point	0.5 K, 0.5 K, †, 0.5 K, †, 0.25 K, †	8
Sea level anomaly	SLA	RADS	Track	†	64
In situ ocean temperature	TEM ^a	Argo, XBT, CTD, TAO, PIRATA	Profiles	0.5 K	8
In situ ocean salinity	SAL ^a	Argo, CTD, TAO, PIRATA	Profiles	0.075 psu	8
Sea ice concentration	SIC	HadISIC, OSISAF	Gridded	0.1 [C], 0.1 [C]	8
Sea ice temperature	SIT	HadISIT	Gridded	0.1 K	8
Zonal wind	ARU	JRA-55	Gridded	1 m s ⁻¹	8
Meridional wind	ARV	JRA-55	Gridded	1 m s ⁻¹	8
Air temperature	ART	JRA-55	Gridded	1 K	8
Specific humidity	ARH	JRA-55	Gridded	0.05 kg kg ⁻¹	8

^a Low confidence CORA5.0 in situ data TEM2 and SAL2 have twice the error and 4 times the *R* factor to the high confidence data listed in the table.

and specific humidity over time. Our conclusion was that although no dramatic impacts of SST assimilation statistics are evident, there must emerge a systematic bias when both HadISST and OISST are assimilated that is inconsistent with the JRA-55 reanalysis. Therefore the use of OISST was discontinued in favor of continued assimilation of HadISST up to the satellite era.

c. In situ temperature and salinity

Ocean in situ observational data are spatially, vertically, and temporally heterogeneous over the 60-yr period (Fig. 1). For the period between 1960–2000 the data are dominated by shallow (<100 m) coastal temperature observations in the Northern Hemisphere, as well as expendable bathythermograph (XBT) measurements, which provide temperature profiles to around 800-m depth along a set of frequently occupied transects. Since 2000 the ocean in situ data coverage has improved for both temperature and salinity with increased and more homogeneous global distribution with increased vertical profile depths (to

2000 m). However, there are still few observations below 2000 m and a significant fraction of the global ocean with fewer than 10 observations per horizontal grid box per year.

Prior to 1 January 2016 we assimilate in situ temperature and salinity from the Coriolis Ocean Dataset for Reanalysis (CORA5) (Szekely et al. 2019). CORA5 contains 16 data groups including Argo, XBT, conductivity–temperature–depth (CTD), XCTD, moorings, marine mammals, and drifting buoys. After 1 January 2016 all available duplicate checked global delayed mode in situ ocean observation from the Australian Bureau of Meteorology MMT database were assimilated and used for verification. MMT is sourced from the WMO GTS, CORIOLIS, and USGODAE Data Assembly Centers (DACs) and has much of the same delayed mode data while also providing access to real-time in situ observations. CORA5 data reported to be less accurate, such as XBT, marine mammals, and all transmitted GTS data not received at the DACs, were assigned higher errors to reduce their impact on assimilation. Both databases include Argo profiles (Roemmich et al. 2009), TAO, PIRATA,

FIG. 1. The spatial distribution of ocean in situ observational data for the periods (left) 1960–2000 and (center) 2000–18. (right) Ratio of temperature and salinity observations by depth.

and RAMA moored arrays, and ship-based CTD and XBT profiles. The CORA5.0 dataset low-confidence (here denoted TEM2 and SAL2) and high-confidence (here denoted TEM and SAL) in situ observations, errors and associated R factors are described in Table 1. The low-confidence in situ data have a doubling of the error estimates relative to the high-confidence data and accordingly the R factors used in CAFE60v1 are a factor of 4 times larger.

d. Sea ice concentration and freezing point temperature

Sea ice concentration data over the period January 1960 through December 2017 are sourced from HadISST Sea Ice Concentration (HadISIC) (Rayner et al. 2003). Post January 2018 sea ice concentration data are sourced from the 25-km resolution climate data record EUMETSAT OSISAF global sea ice concentration reprocessing dataset 1978–2015 (v1.2, 2015) by the Norwegian and Danish Meteorological Institutes. This uses passive microwave data from SMMR, SSM/I, and SSMIS sensors.

e. Atmospheric reanalysis data

For the atmosphere, daily mean zonal and meridional wind, temperature, humidity, and surface pressure from the JRA-55 atmospheric reanalysis are assimilated with no direct assimilation of atmospheric observations carried out. Sandery et al. (2020) discuss the possible degradation caused by assimilation of an atmospheric reanalysis product due to both systematic and random errors inherent in a reanalyzed data product; however, any systematic biases present in JRA-55 may be in part offset by judiciously ascribing observation errors to the gridded atmospheric data in order to avoid overfitting.

5. Assimilation results

In the following, global mean error statistics are calculated at the end of each calendar month and for all assimilated observation types. Specifically, we define forecast innovation mean absolute deviation (MAD) as $(1/k)\sum_{i=1}^k |\mathbf{d} - H(\mathbf{x}_i^f)|$ and where forecast innovation bias refers to the mean magnitude of the forecast innovation. All error statistics are with respect to monthly mean model background, analysis, and superobservations. The number of observations indicated is in terms of the monthly mean superobservations; the number of raw observations ingested each cycle is many times larger.

a. Ocean

As previously discussed, the ocean observing network is both spatially and temporally highly heterogeneous. Prior to Argo, satellite SST, and altimetry, large parts of the ocean domain were not sufficiently well observed to constrain the model such that, from a data assimilation perspective, the ocean is effectively unobserved. This is evident in Fig. 2 in the number of surface and in situ observations available for assimilation. Prior to the availability of satellite SST data, gridded HadISST data are assimilated, and for a brief period between January 1988 and August 1990 OISST. Throughout this period analysis biases are very small averaging between ± 0.2 K, spread is

about 0.55°C , and forecast MAD is typically less than 1 K. As altimetry, then satellite SST and Argo come online (see also Fig. 1), there is a significant adjustment of the upper ocean structure and in particular the thermocline. This effect is most clearly reflected in the statistics for TEM and SAL post 2005 with the halving of forecast MAD and the near elimination of analysis biases and the corresponding reduction in spread as the system becomes constrained.

Altimetry rapidly comes online from June 1992 quickly adjusting the mean sea level with systematic biases of less than 5 cm and forecast MAD averaging 8.5 cm. The unconstrained sea level background ensemble was initially biased and was rapidly corrected once altimetry was assimilated. Error statistics for the low confidence CORA5.0 in situ temperature TEM2 and salinity SAL2 observations are also shown. Despite these observations having reduced impact weights (i.e., larger R factors), these data are clearly of some utility with little or no evidence that they degrade overall biases or forecast MAD in any way.

To understand both the spatial distribution of observations and forecast errors, we consider a particular assimilation cycle in September 2018. Figure 3 illustrates the ensemble mean forecast, the observed state, the ensemble means of the background and analysis innovations errors (i.e., differences between analysis, forecast, and observations in observation space), and the background ensemble spread for SST in observation space where the number of superobservations ($=171\,550$) is entirely due to satellite data. Here the ensemble mean forecast and observations look similar; however, the ensemble mean forecast innovation reveals errors of up to $\pm 1^\circ\text{C}$ in the tropical Pacific and both the North Pacific and North Atlantic. The ensemble mean analysis innovation reveals a significant reduction in amplitude and spatial heterogeneity in the differences between the analyzed and observed SST. Spread in the background ensemble reveals the regions of large variance to be associated with the Kuroshio, the Gulf Stream, and the eastern equatorial Pacific in particular. Given the relative lack of spread outside the aforementioned regions, it is quite surprising that the analysis errors are so substantially reduced relative to the background innovation.

A similar analysis for SLA, temperature, and salinity in September 2018 is shown in Figs. 4 and 5. Here we show the observed state, ensemble mean forecast, and analysis innovations and background ensemble spread for sea level anomaly, temperature (0–200-m depth), and salinity (0–200-m depth) binned on a uniform $3^\circ \times 3^\circ$ grid. The largest ensemble-mean innovations in SLA are evident about the Antarctic sea ice zone and also obvious in the ensemble mean forecast and background ensemble spread. Over the rest of the globe SLA analysis innovations are generally reduced and sufficient to constrain the global mean sea level. Errors in ensemble mean forecast salinity are largest in the western equatorial Pacific but, despite some few specific large values, are mostly homogeneously distributed. Temperature innovations are largest and most coherent in the equatorial Pacific with values up to $\pm 1^\circ\text{C}$ in the equatorial Pacific, indicative of displacement errors in the thermocline. While the amplitude of these errors

FIG. 2. Forecast mean absolute deviation (red) and bias (blue), ensemble spread (orange), and number of observations assimilated (black) for SST, SLA, in situ temperature, and salinity. Auxiliary observations refer to the CORA5.0 TEM2 and SAL2 observations.

is reduced in the ensemble mean analysis innovation, they retain their spatial distribution indicating persistent differences in the analyses and observed equatorial Pacific thermocline.

Ensemble mean depth-averaged (0–500 m) and time-averaged (2005–17) forecast and analysis innovations for temperature (Figs. 6a,b) and salinity (Figs. 6c,d) reveal regions of systematic model biases. In particular, forecast temperature biases are mostly located where the currents are strongest (i.e., the equatorial Pacific and Atlantic) and where boundary currents such as Kuroshio and the Gulf Stream separate from the coast. Other major regions of forecast temperature bias occur in the eddying regions of the ocean, for example the Antarctic Circumpolar Current (ACC) and the Brazil–Malvinas Confluence. These errors are in part due to the lack of fine-scale structure due to coarse model resolution, which further contributes to errors (i.e., displacement of the time mean currents). An additional major region of bias occurs in the form of a too warm western Pacific warm pool. The analysis temperature innovations show a significant reduction in upper ocean biases; however, lack of model resolution and ensemble spread in the aforementioned regions where significant observed variance occurs limits the

impact observations can have in the assimilation. Salinity innovations have largest magnitudes in regions, broadly similar to those of temperature; however, the region about the Indonesian Seas is generally too fresh, indicating a significant density bias.

b. Atmosphere

We next consider global forecast MAD and bias, ensemble spread, and number of observations assimilated for atmospheric air temperature, zonal and meridional wind, and specific humidity (Fig. S1 in the online supplemental material). Errors are with respect to JRA-55 gridded data and averaged over the entire volume of the atmosphere. Air temperature (ART) and air specific humidity (ARH) have generally similar behaviors. Apart from a period between 1992 and 1994, mean biases in ART are consistently around 0.5°C or less; ensemble spread is $\approx 1.5^{\circ}\text{C}$ with variation due to the seasonal cycle evident but otherwise generally uniform; ART MAD is consistently between 2 and 4 K. For ARH, mean biases are $0.5^{\circ}\text{g kg}^{-1}$ or less, ensemble spread is $\approx 0.5^{\circ}\text{g kg}^{-1}$, and MAD between 0.5 and $1.5^{\circ}\text{g kg}^{-1}$. The 3-yr period 1992–94 sees the rapid introduction of altimetric data and a corresponding adjustment of the global sea level. The rapid introduction of SLA

SST (C) September 2018 #sobs=171550

FIG. 3. Ensemble mean forecast and observed state, ensemble mean forecast and analysis innovations, and background ensemble spread for SST in September 2018 in observation space.

data has had a transitive impact on both the ARH analysis bias and forecast MAD. The impacted analyzed states for specific humidity during this period, particularly in the tropics, are consequently reflected in ART via the cross-domain covariances. Post-1994 global sea level has been adjusted and ARH biases systematically reduced whereas ART biases return to prealtimetry values. That said, the reduced ARH biases do not appear to significantly improve forecast MAD over the pre-1992 values.

The analyzed zonal (ARU) and meridional (ARV) winds show small but generally positive mean biases typically $<0.5 \text{ m s}^{-1}$, with forecast MAD of around 1.5 and 2 m s^{-1} respectively with ensemble spread typically 0.2 m s^{-1} below the MAD values. Obviously, when comparing daily analyzed states to daily JRA-55 fields, many of the midlatitude synoptic-scale features would be out of phase after a month and errors would be saturated. Instead, we compare monthly mean analyzed CAFE60v1 states to monthly mean gridded JRA-55 data with quite low MAD and bias values evident for the atmospheric winds, reflecting consistent observed and analyzed large-scale features in the troposphere. Noticeably, errors in ARU and ARV are only weakly impacted by the larger errors in ART and ARH during 1992–94.

Again we examine the spatial distribution of observations and forecast errors for September 2018. Around 2.4×10^4 superobservations are assimilated each cycle varying only with the number of days in each calendar month (Fig. S1). In Figs. 7 and 8 we compare ensemble mean forecast, observed state, ensemble mean forecast, and analysis innovations and background ensemble spread for air temperature and meridional velocity, vertically averaged through the troposphere (from the surface to 200 hPa) for September 2018 based on the assimilation of 242 417 superobservations. The ensemble mean forecast innovation represents the difference between the ensemble and time average of a 30-day forecast started on 1 September 2019 and the time mean September “observed” state from JRA-55. Clearly evident in the forecast innovations are the large errors in the midlatitude troposphere associated with synoptic-scale features.

The large-scale background errors in ART of up to $\pm 3^\circ\text{C}$ are substantially reduced in the analysis. In addition to reduced ARV analysis errors of up to $\pm 3 \text{ m s}^{-1}$, the analyzed innovation errors are made largely homogeneous, thereby removing the opportunity for rapid error growth arising from disturbances in particular spatially localized regions of instability. For ART, innovation errors evident in the mean forecast are effectively absent in the analysis apart from some regions near land in the tropics (e.g., Brazil). Here the lack of ensemble spread in both the ART and ARV covariances arises in the tropics as a consequence of reduced model resolution. This makes addressing sources of error, particularly in the major convective regions, a challenge.

Time-mean ART forecast innovations between 850 hPa and the surface and averaged over the entire reanalysis 1960–2019 (Fig. 9) reveal the largest surface biases as an excessive warming in the regions over Brazil and the Amazon basin, the Arctic, and parts of Antarctica. Cool biases are present over the elevated regions of the Northern Hemisphere midlatitudes. These biases are significantly reduced in

September 2018 CAFE-60

FIG. 4. Ensemble mean forecast and observed state for sea level anomaly, temperature (0–200-m depth), and salinity (0–200-m depth) for September 2018 binned on a uniform $3^\circ \times 3^\circ$ grid.

the corresponding analysis innovations however are still substantial over South America. As expected, greatest bias reductions occur in regions where the background ensemble spread is largest.

Overall, the error statistics indicate the assimilation has constrained the large scales of the analyzed atmosphere to observations. The efficacy of this approach will be demonstrated later in direct comparison to independent observations

September 2018 CAFE-60

FIG. 5. Ensemble mean forecast innovation, analysis innovation, and background ensemble spread for sea level anomaly, temperature (0–200-m depth), and salinity (0–200-m depth) for September 2018. The data are binned on a uniform $3^\circ \times 3^\circ$ grid in observation space, hence the gray areas where observations were absent.

FIG. 6. Ensemble mean, time averaged over 2005–17 and depth averaged (0–500 m) forecast and analysis innovations for (a),(b) temperature and (c),(d) salinity binned on a uniform $3^\circ \times 3^\circ$ grid.

of the major atmospheric teleconnections across a range of products.

c. Sea ice

As described earlier, assimilated observations of the sea ice can impact the ocean but not the atmosphere via the explicit

choice of cross-domain covariances. Observations are assimilated as monthly mean superobservations with their error statistics described in Fig. S2. Analyzed biases in sea ice concentration in both hemispheres are quite small relative to the large biases in summer sea ice freezing point temperatures, which range between -2° and -4°C , indicating far too little sea

FIG. 7. Ensemble mean forecast and observed state for air temperature and meridional velocity vertically averaged through the troposphere (from the surface to 200 hPa) and binned onto a uniform $3^\circ \times 3^\circ$ grid for September 2018.

FIG. 8. Ensemble mean forecast innovation, analysis innovation, and background ensemble spread for air temperature and meridional velocity vertically averaged through the troposphere (from the surface to 200 hPa) and binned onto a uniform $3^\circ \times 3^\circ$ grid for September 2018.

ice relative to observations. Antarctic sea ice concentration and freezing point temperature forecast MAD reach their maximums during summer (0.4 and 2 K) and nearly double those values for Arctic sea ice during the boreal summer. These results point not only to significant biases in the simulated summer sea ice but also poor ocean–sea ice cross covariances due to model biases producing too warm under ice ocean temperatures at this time of year.

Examination of ensemble spread in sea ice concentration around Antarctica during September 2018 (see Fig. 10) shows variance concentrated at the edge of the sea ice zone. Comparison of the ensemble mean forecast to the observed state and ensemble mean forecast innovation shows the largest errors occur where the spread is large but also in the region near 90°E . The analysis innovations show only modest improvements on the largely very good forecast state with the black contour indicating the maximum climatological extent for September calculated using the 15% concentration threshold over the full OSISAF record. The sea ice forecasts over a 1-month lead time are highly sensitivity to model biases due

to the rapid adjustment of sea ice concentration to both the fast variations of the atmosphere and subsurface ocean temperatures.

d. Ocean biogeochemistry

Quantifying the annual to multiyear variability in the sea–air flux of CO_2 is a challenging problem given the sparse observations available (Gloeger et al. 2020). Furthermore, in the absence of any explicit representation of ocean eddies in our model, the fast OBGC processes (i.e., phytoplankton, production of nutrients and their transport, etc.) cannot be well represented as they are intrinsically tied to the eddying regions of the ocean. Slower processes, including the subduction of carbon and alkalinity, are more closely tied to the slower oceanic dynamics but insufficient observational data exist to constrain the modeled OBGC processes over time. With these constraints in mind, CAFE60v1 does not explicitly assimilate OBGC observations, for example ocean color (i.e., chlorophyll A) (Dunstan et al. 2018); instead, we have taken the approach of using the cross-domain covariances from the ocean observations

FIG. 9. Ensemble mean air temperature forecast and analysis innovations and background ensemble spread for air temperature vertically averaged from the surface to 850 hPa summed over the entire reanalysis period and binned on a uniform $3^\circ \times 3^\circ$ grid.

to project onto the OBGC state, which are then used to form increments necessary to generate the analyzed OBGC states. This approach is applied to constrain one of the dominant processes (i.e., ocean dynamics) responsible for modulating annual to multiyear variability in the OBGC fields.

To evaluate the OBGC fields in CAFE60v1, we focus on dissolved inorganic carbon, which is separated in our simulation into the total dissolved inorganic carbon (DIC) and anthropogenic dissolved inorganic carbon (ADIC). We do this by including two DIC tracers in the ocean, one called DIC that sees a fixed preindustrial atmospheric CO_2 concentration (280 μatm) and one called ADIC that sees the observed historical atmospheric CO_2 concentration until 2011 and then follows the

RCP2.6 concentrations. From these two tracers, we calculate anthropogenic CO_2 from ADIC minus DIC. We maintain conservation of the anthropogenic CO_2 in the ocean by the following steps:

- 1) subtracting DIC from ADIC and storing the resulting anthropogenic CO_2 field at the end of the free model run cycle (end of the month) before applying data assimilation;
- 2) computing the analysis DIC field using the assimilation of ocean observations which modify DIC through the cross covariances computed from the ensemble; and
- 3) setting the ADIC field to the analysis DIC field plus the anthropogenic CO_2 value determined in step 1 before starting the next cycle.

In CAFE60v1, the DIC and ADIC tracers display the same variability and only differ in how the ocean takes up anthropogenic CO_2 at the surface and how the ocean circulation distributes it within the ocean. Our approach ensures the simulated anthropogenic CO_2 is conserved in the ocean interior and is only modified by the sea–air exchange.

6. Initial conditions

A primary function of the CAFE60v1 reanalysis is as initial conditions for forecasting the near-term climate. As one of the centers contributing to the WMO Lead Centre for Annual-to-Decadal Climate Prediction,² CAFE60v1 forecasts span from 1960 to the present with 10-member 10-yr lead time forecasts available each November to present. Ensemble forecasts of similar size are encompassed in the tier-1 hindcasts of the CMIP6 Decadal Climate Prediction Project (DCPP) (Boer et al. 2016).

Recently, a large ensemble of initialized hindcasts using the Community Earth System Model (CESM) decadal prediction large ensemble (CESM-DPLE) composed of 40-member ensembles initialized each 1 November between 1954 and 2015 with lead times of 122 months has been completed at NCAR (Yeager et al. 2018), employing the model configuration of Kay et al. (2015). In common with many other decadal forecast systems, CESM-DPLE initial conditions for the ocean were obtained from an ocean–sea ice configuration of the model forced at the surface with historical atmospheric state and flux fields obtained by atmospheric reanalysis (see also Leroux et al. 2018).

A significant point of difference between CAFE60v1 and all other systems is the application of strongly coupled data assimilation. Importantly, CAFE60v1 provides multivariate analyzed states each month with a comprehensive set of uncertainty estimates including errors with respect to particular observational products and types. The CAFE60v1 forecasts are initialized using the ensemble members of the coupled reanalysis for which the complete restart files of all 96 member have been archived every month, potentially enabling up to 96 (members) $\times$ 12 (months) $\times$ 60 (years) = 69 120 individual

² See <http://www.wmolc-adcp.org>.

SIC [C] September 2018 CAFE-60 #sobs=12819

FIG. 10. Ensemble mean forecast, observed state, ensemble mean forecast, and analysis innovations and background ensemble spread for Antarctic sea ice concentration in September 2018. The black contour indicates the maximum climatological extent for September calculated using the 15% concentration threshold using the full OSISAF record.

forecasts. CAFE60v1 employs full field initialization from analyzed balanced initial conditions where cross domain covariances are explicitly taken into account. Furthermore, the structures of the 96-member ensemble perturbations are by construction coherent and balanced and capture the time-evolving error modes of the full coupled system. In this regard it is expected that the CAFE60v1 forecast ensemble might diverge less rapidly than systems using random or unbalanced initial perturbations. An analysis of the CAFE60v1 forecast utility is beyond the scope of this paper but is the subject of follow-up studies.

7. Future developments

Current work is toward additional capability in the next iteration of the CAFE system to further develop sea ice assimilation and parameter estimation (Kitsios et al. 2021), and to extend OBGCM DA to include the direct assimilation of chlorophyll A and carbon observations, as well as the assimilation of remotely sensed land surface observations (Babaeiana et al. 2018). With the recent awarding of one of Australia's largest ever competitive computing grants,³ the CAFE60v1 forecast ensemble is currently being extended to 96 members initialized each season over the last decade with lead times of 10 years. Once complete, these additional forecasts will provide an unprecedented resource for analyzing the climate over the recent decade. A complete list of CAFE60v1 variables and diagnostics are listed in Tables S1–S7 in the online supplemental material.

In conclusion, the CAFE60v1 database will add to international efforts to develop a comprehensive data resource for studying internal climate variability and predictability, including the climate response to anthropogenic forcing on multiyear to decadal time scales.

The CAFE60v1 reanalysis will be publicly available for download from June 2021 at the following URL: <http://hdl.handle.net/102.100.100/389002>.

Acknowledgments. The authors were supported by the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) Decadal Climate Forecasting Project (<https://research.csiro.au/dfp>). We gratefully acknowledge the support of National Computing Infrastructure (NCI) Australia and the Pawsey Supercomputing Centre. We appreciate support from Stephen Griffies and Naishali Naik at NOAA GFDL in providing the CM2.1 model codes and radiative forcing data and the large international effort that is required to maintain the diverse observing platforms used in this study. We also thank the editor and two anonymous reviewers for their guidance and thoughtful comments in preparing this manuscript for publication.

REFERENCES

- Alves, O., M. A. Balmaseda, D. Anderson, and T. Stockdale, 2004: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Quart. J. Roy. Meteor. Soc.*, **130**, 647–667, <https://doi.org/10.1256/qj.03.25>.
- Babaeiana, E., M. Sadeghib, T. E. Franzc, S. Jones, and M. Tuller, 2018: Mapping soil moisture with the Optical TRapezoid Model (OPTRAM) based on long-term MODIS observations. *Remote Sens. Environ.*, **211**, 425–440, <https://doi.org/10.1016/j.rse.2018.04.029>.
- Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quart. J. Roy. Meteor. Soc.*, **139**, 1132–1161, <https://doi.org/10.1002/qj.2063>.
- Bi, D., and Coauthors, 2013: ACCESS-OM: The ocean and sea-ice core of the ACCESS coupled model. *Aust. Meteor. Oceanogr. J.*, **63**, 213–232, <https://doi.org/10.22499/2.6301.014>.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, [https://doi.org/10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- Boer, G., and Coauthors, 2016: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3751–3777, <https://doi.org/10.5194/gmd-9-3751-2016>.
- Carton, J. A., G. A. Chepurin, and L. Chen, 2018: SODA3: A new ocean climate reanalysis. *J. Climate*, **31**, 6967–6983, <https://doi.org/10.1175/JCLI-D-18-0149.1>.
- Castruccio, F. S., A. R. Karspeck, G. Danabasoglu, J. Hendricks, T. Hoar, N. Collins, and J. L. Anderson, 2020: An EnOI-based data assimilation system with DART for a high-resolution version of the CESM2 ocean component. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002176, doi:<https://doi.org/10.1029/2020MS002176>.
- Compo, G., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137** (654), 1–28, <https://doi.org/10.1002/qj.776>.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, <https://doi.org/10.1175/JCLI3629.1>.
- Doblas-Reyes, F., and Coauthors, 2013: Initialized near-term regional climate change prediction. *Nat. Commun.*, **4**, 1715, <https://doi.org/10.1038/ncomms2704>.
- Dunstan, P., and Coauthors, 2018: Interactions in global patterns of variation in sea surface temperature and chlorophyll A produce mesoscale unique states. *Sci. Rep.*, **8**, 14624, <https://doi.org/10.1038/s41598-018-33057-y>.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Forget, G., J. M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev.*, **8**, 3071–3104, <https://doi.org/10.5194/gmd-8-3071-2015>.
- Gaiser, P. W., and Coauthors, 2004: The WindSat spaceborne polarimetric microwave radiometer: Sensor description and early orbit performance. *IEEE Trans. Geosci. Remote Sens.*, **42**, 2347–2361, <https://doi.org/10.1109/TGRS.2004.836867>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.

³ See <https://www.hpcwire.com/off-the-wire/alcg-largest-computing-grants-in-australian-history-announced>.

- Gloege, L., and Coauthors, 2020: Quantifying errors in observationally-based estimates of ocean carbon sink variability. *Earth and Space Science Open Archive*, **11**, 24, <https://doi.org/10.1002/essoar.10502036.1>.
- Griffies, S. M., 2009: Elements of MOM4p1. GFDL Ocean Group Tech. Rep. 6, 377 pp., http://data1.gfdl.noaa.gov/~arl/pubrel/o/old/doc/mom4p1_guide.pdf.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Houtekamer, P. L., and F. Zhang, 2006: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **134**, 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>.
- Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2021: Improvements of the Daily Optimum Sea Surface Temperature (DOISST) version 2.1. *J. Climate*, **34**, 2923–2939, <https://doi.org/10.1175/JCLI-D-20-0166.1>.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYP>2.0.CO;2).
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Keptert, J. D., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 1157–1176, <https://doi.org/10.1002/qj.443>.
- Kitsios, V., P. Sandery, T. J. O’Kane, and R. Fiedler, 2021: Ensemble Kalman filter parameter estimation of ocean optical properties for reduced biases in a coupled general circulation model. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002252, <https://doi.org/10.1029/2020MS002252>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Kushnir, Y., and Coauthors, 2019: Towards operational predictions of the near-term climate. *Nat. Climate Change*, **9**, 94–101, <https://doi.org/10.1038/s41558-018-0359-7>.
- Laloux, P., and Coauthors, 2018: A coupled reanalysis of the twentieth century. *J. Adv. Model. Earth Syst.*, **10**, 1172–1195, <https://doi.org/10.1029/2018MS001273>.
- Lavergne, T., and Coauthors, 2019: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. *Cryosphere*, **13**, 49–78, <https://doi.org/10.5194/tc-13-49-2019>.
- Leroux, S., T. Penduff, L. Bessieres, J. M. Molines, J.-M. Brankart, G. Serazin, B. Barnier, and L. Terray, 2018: Intrinsic and atmospherically forced variability of the AMOC: Insights from a large-ensemble ocean hindcast. *J. Climate*, **31**, 1183–1203, <https://doi.org/10.1175/JCLI-D-17-0168.1>.
- Naeije, M., E. Schrama, and R. Scharroo, 2000: The radar altimeter database system project RADS. *Proc. IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symp.: Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment*, Vol. 2, 487–490.
- O’Kane, T. J., and J. S. Frederiksen, 2008: Comparison of statistical dynamical, square root and ensemble Kalman filters. *Entropy*, **10**, 684–721, <https://doi.org/10.3390/e10040684>.
- , R. J. Matear, M. A. Chamberlain, and P. R. Oke, 2014: ENSO regimes and the late 1970’s climate shift: The role of synoptic weather and South Pacific Ocean spiciness. *J. Comput. Phys.*, **271**, 19–38, <https://doi.org/10.1016/j.jcp.2013.10.058>.
- , and Coauthors, 2019: Coupled data assimilation and ensemble initialization with application to multiyear ENSO prediction. *J. Climate*, **32**, 997–1024, <https://doi.org/10.1175/JCLI-D-18-0189.1>.
- Oke, P. R., and Coauthors, 2013: Towards a dynamically balanced eddy-resolving ocean reanalysis: BRAN3. *Ocean Modell.*, **67**, 52–70, <https://doi.org/10.1016/j.ocemod.2013.03.008>.
- Penny, S. G., D. W. Behringer, J. A. Carton, and E. Kalnay, 2015: A hybrid global ocean data assimilation system at NCEP. *Mon. Wea. Rev.*, **143**, 4660–4677, <https://doi.org/10.1175/MWR-D-14-00376.1>.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Roemmich, D., and Coauthors, 2009: The Argo Program: Observing the global ocean with profiling floats. *Oceanography*, **22**, 34–43, <https://doi.org/10.5670/oceanog.2009.36>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Sakov, P., 2018: EnKF-C user guide, version 7. 48 pp., <https://arxiv.org/abs/1410.1233>.
- , and P. A. Sandery, 2015: Comparison of EnOI and EnKF regional ocean reanalysis systems. *Ocean Modell.*, **89**, 45–60, <https://doi.org/10.1016/j.ocemod.2015.02.003>.
- , and —, 2017: An adaptive quality control procedure for data assimilation. *Tellus*, **69**, 1318031, <https://doi.org/10.1080/16000870.2017.1318031>.
- , F. Counillon, L. Bertino, K. Lisæter, P. Oke, and A. Korabely, 2012: TOPAZ4: An ocean–sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.*, **8**, 633–656, <https://doi.org/10.5194/os-8-633-2012>.
- Sandery, P. A., T. J. O’Kane, V. Kitsios, and P. Sakov, 2020: Climate model state estimation using variants of EnKF coupled data assimilation. *Mon. Wea. Rev.*, **148**, 2411–2431, <https://doi.org/10.1175/MWR-D-18-0443.1>.
- Smith, D. M., and J. M. Murphy, 2007: An objective ocean temperature and salinity analysis using covariances from a global climate model. *J. Geophys. Res.*, **112**, C02022, <https://doi.org/10.1029/2005JC003172>.
- , S. Cusack, A. Colman, C. Folland, G. Harris, and J. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796–799, <https://doi.org/10.1126/science.1139540>.
- , R. Eade, and H. Pohlmann, 2013: Real-time multi-model decadal climate predictions. *Climate Dyn.*, **41**, 2875–2888, <https://doi.org/10.1007/s00382-012-1600-0>.
- , and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796–800, <https://doi.org/10.1038/s41586-020-2525-0>.
- Stickler, A., and Coauthors, 2014: ERA-CLIM: Historical surface and upper-air data for future reanalyses. *Bull. Amer. Meteor. Soc.*, **95**, 1419–1430, <https://doi.org/10.1175/BAMS-D-13-00147.1>.

- Szekely, T., J. Gourrion, S. Pouliquen, and G. Reverdin, 2019: The CORA 5.2 dataset for global in situ temperature and salinity measurements: Data description and validation. *Ocean Sci.*, **15**, 1601–1614, <https://doi.org/10.5194/os-15-1601-2019>.
- Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, [https://doi.org/10.1175/1520-0493\(2004\)132<1590:WIBAEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2).
- Yeager, S. G., and Coauthors, 2018: Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **99**, 1867–1886, <https://doi.org/10.1175/BAMS-D-17-0098.1>.
- Žagar, N. J., J. Tribbia, J. L. Anderson, K. Raeder, and D. T. Kleist, 2010: Diagnosis of systematic analysis increments by using normal modes. *Quart. J. Roy. Meteor. Soc.*, **136**, 61–76, <https://doi.org/10.1002/qj.533>.
- Zhang, L., T. L. Delworth, X. Yang, R. G. Gudgel, L. Jia, G. A. Vecchi, and F. Zeng, 2017: Estimating decadal predictability for the Southern Ocean using the GFDL CM2.1 model. *J. Climate*, **30**, 5187–5203, <https://doi.org/10.1175/JCLI-D-16-0840.1>.
- Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564, <https://doi.org/10.1175/MWR3466.1>.