

Objective Array Design: Application to the Tropical Indian Ocean

PAVEL SAKOV AND PETER R. OKE

CSIRO Marine and Atmospheric Research, and Wealth from Oceans Flagship Program, Hobart, Tasmania, Australia

(Manuscript received 6 February 2007, in final form 26 June 2007)

ABSTRACT

A simple, versatile, computationally efficient ensemble-based method for objectively designing an observation array is described. The method seeks to compute the observation array that minimizes the analysis error variance, according to Kalman filter theory. While most elements of the method have been described elsewhere, this paper attempts to present a simple, yet comprehensive, recipe for array design based on an ensemble of anomalies that represents the background error covariance. The versatility of the method is demonstrated through a series of applications to the tropical Indian Ocean (TIO). The first application uses model-generated fields of high-pass-filtered mixed layer depth to design an array to monitor intraseasonal variability. The second uses gridded observations of sea level anomaly to design an array to monitor intraseasonal-to-interannual variability. For both applications, the objectively designed arrays are compared to an array that will soon be implemented under the auspices of the Climate Variability and Predictability–Global Ocean Observing System (CLIVAR–GOOS) Indian Ocean Panel (CG-IOP). The authors conclude that the CG-IOP array produces results that compare well to the objectively designed arrays for intraseasonal variability, and observations to the east and northeast of the TIO and south of India are most important for resolving intraseasonal variability. The authors also find that observations near 9°S, where seasonal Rossby waves dominate, are important for observing seasonal-to-interannual variability. The described method for objective array design can be applied to a wide range of geophysical applications where time series of gridded modeled or observed fields are available.

1. Introduction

The use of models and data assimilation theory to aid in the design of atmospheric and oceanic observing systems has gained momentum over recent decades (e.g., Kållberg 1984; McIntosh 1987; Barth and Wunsch 1990; Kuo et al. 1998; Hackert et al. 1998; Bishop et al. 2001; Hirschi et al. 2003; Schiller et al. 2004; Oke and Schiller 2007b). The applications described in this study are motivated by the proposal of the Climate Variability and Predictability–Global Ocean Observing System (CLIVAR–GOOS) Indian Ocean Panel (CG-IOP; CLIVAR et al. 2006) to implement a surface mooring array in the tropical Indian Ocean (TIO) (CG-IOP array) that is similar to the Tropical Ocean and Global Atmosphere–Tropical Atmosphere Ocean (TOGA–TAO) array in the tropical Pacific Ocean. This initiative has resulted in a series of model-based studies that

have attempted to assess the CG-IOP array and identify possible improvements. Results from these studies are varied. For example, Vecchi and Harrison (2007) presented results from a series of observing system simulation experiments (OSSEs) using a high-resolution ocean model and assessed the ability of an integrated observing system, including Argo observations, XBT lines, and the CG-IOP array to monitor intraseasonal and interannual variability. Ballabrera-Poy et al. (2007) used a reduced-order Kalman filter to objectively determine an array for mapping sea surface height and sea surface temperature in the TIO. Oke and Schiller (2007a) use an EOF-based approach to assess the CG-IOP array's ability to monitor intraseasonal and interannual variability. Vecchi and Harrison (2007) conclude that in conjunction with the integrated observing system, the CG-IOP array should be capable of resolving intraseasonal and interannual variability. Both Ballabrera-Poy et al. (2007) and Oke and Schiller (2007a) argue that the CG-IOP array may oversample the region within a few degrees of the equator. These studies also suggest that key regions for monitoring seasonal-to-interannual variability are south of 8°S, at

Corresponding author address: Peter R. Oke, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, 7001 TAS, Australia.
E-mail: peter.oke@csiro.au

about 4° – 5° from the equator and along the coast of Indonesia. These regions correspond to the locations of the maximum amplitude of seasonal Rossby waves (Masumoto and Meyers 1998; Schouten et al. 2002), equatorial Rossby waves, and strong Indian Ocean dipole events (Murtugudde et al. 2000), respectively.

In this study, we describe a simple, versatile, computationally efficient ensemble-based method for objectively designing an observation array, and consider two applications of the method to the TIO. The first application uses model-generated fields of high-pass-filtered mixed layer depth to design an array to monitor intraseasonal variability. The second application uses gridded observations of sea level anomaly to design an array to monitor intraseasonal-to-interannual variability.

Ensemble-based methods for optimal array design have been used intensively in the last few years (e.g., Bishop et al. 2001). These methods are based on ensemble square root filter theory (e.g., Tippett et al. 2003) and allow one to handle large systems in cases when explicit manipulation of the background error covariance matrix is not possible or not feasible. Most of the studies on the ensemble-based optimal array design consider the problem of *targeted observations*, or adaptive sampling, aimed at improving the model's forecast at a given time (e.g., Bishop et al. 2001; Langland 2005; Khare and Anderson 2006). The problem of targeted observations implies the use of an ensemble of model states representing the background error covariance at a particular time, which in itself often poses a formidable task. Consequently, the advantages of the ensemble-based optimal array design methods may not have been fully exploited by the wider atmospheric and oceanographic scientific communities.

In this study, we consider a problem of optimal array design based on an ensemble of system states that is formed from output from a long model run or from time series of observation-based gridded fields. Because such an ensemble represents the time-averaged statistics of the system, the described method is most suitable for designing a stationary array that is intended to be deployed over a long period of time. Furthermore, the method does not take advantage of possible time correlations because it treats the ensemble as a nonordered collection of system states that adequately approximates the background error covariance.

The main steps in the objective design of an observation array, as described in this manuscript, are represented schematically in Fig. 1. Suppose an array of p observations is to be designed. The method starts with the construction of a representative ensemble. The first optimal observation is then calculated, and the en-

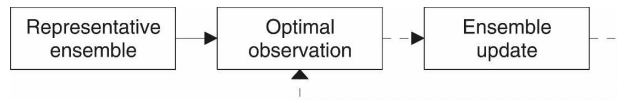


FIG. 1. Schematic diagram depicting the serial calculation of an optimal observation array. The dashed arrows represent optional calculations, depending on the number of observations that are to be included in the array.

semble is updated so that the influence of this observation is reflected (i.e., the background error variance of the ensemble is reduced). The next optimal observation is then calculated, the ensemble is updated again, and so on, until all p observations are calculated.

Equation (3.3) of Berliner et al. (1999) provides a means for identifying the optimal observation for the simplest case where the observation is of the same field that is to be analyzed by the assimilation system. In this paper, we express this formula in terms of ensemble data assimilation, so that an ensemble of anomalies can be used instead of the background error covariance matrix. We also extend their formulation to facilitate the design of multisensor observation networks. An example of a multisensor observation network is a mooring at a certain location that measures more than one variable and/or has multiple instruments at different depths. The ensemble update is conducted using the ensemble square root filter, so that the updated, or transformed, ensemble reflects the reduced background error variance, given information from the optimal observation, or observations, identified in step 2 above. The computational cost of the described method is shown to be linear with respect to the system state dimension, which makes it scalable for application to large systems. Further, by augmenting the state vector and/or rescaling its components, one can design an observing array that best constrains an arbitrary function of the state vector. We demonstrate the versatility of the method through a series of applications that are intended to resolve different time scales of variability, and that use ensembles from different sources (i.e., model and observation based).

The paper is organized as follows: a description of the method is presented in section 2, a description of the models and observations used here is presented in section 3, the results are presented in section 4, and a discussion and our conclusions are in section 5.

2. Method

Suppose we have a discrete system that is characterized by a state vector $\mathbf{x}^{n \times 1}$ and let us assume that the uncertainty of \mathbf{x} is characterized by a background error covariance matrix \mathbf{P}^b . Suppose that after assimilation of a set of observations, the uncertainty of the analyzed

system is characterized by the analysis error covariance matrix \mathbf{P}^a . In this case, the optimal set of observations characterized by the observation matrix \mathbf{H}^{opt} may be defined as a set that minimizes some norm of \mathbf{P}^a :

$$\mathbf{H}^{\text{opt}} = \arg \min_{\{\mathbf{H}\}} \|\mathbf{P}^a\|, \quad (1)$$

where $\{\mathbf{H}\}$ denotes the set of all possible observations.

This approach can be easily implemented using the Kalman filter theory, according to which after assimilating p observations characterized by the observation matrix $\mathbf{H}^{p \times n}$ and error covariance matrix $\mathbf{R}^{p \times p}$ the analysis error covariance is given by

$$\mathbf{P}^a = [\mathbf{I} - \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}] \mathbf{P}^b, \quad (2)$$

where \mathbf{I} is the identity matrix and the superscript T denotes a matrix transpose. We note that the solution given by the Kalman filter is a minimum variance estimation; it does not require the statistics of the model errors and observation errors to be Gaussian; however, the minimum variance estimate coincides with the maximum likelihood estimate if these distributions are Gaussian (e.g., Jazwinski 1970, p. 157). Because the Kalman filter minimizes the trace of the analysis error covariance \mathbf{P}^a , we should assume for consistency that it is the trace of the system error covariance that is used as a norm of the uncertainty in the system state, $\|\mathbf{P}^a\| = \text{trace}(\mathbf{P}^a)$.

Using the relation $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$, the solution of (1)–(2) may be written as

$$\mathbf{H}^{\text{opt}} = \arg \max_{\{\mathbf{H}\}} \text{trace} \frac{\mathbf{H} \mathbf{P}^b (\mathbf{H} \mathbf{P}^b)^T}{\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}_H}, \quad (3)$$

where \mathbf{R}_H is the observation error covariance associated with the observation matrix \mathbf{H} . Calculation of (3) involves inversion of a $p \times p$ matrix, which is computationally inexpensive for small p .

The solution (3) may include multiple observations and multiple sensor elements (e.g., a mooring at a certain location that measures more than one variable and/or has multiple instruments at different depths). In a more simple case when only one measurement of a single element \mathbf{x}_k of the state vector is possible at a time, \mathbf{H} becomes a row vector $\mathbf{H} = \mathbf{h}(k)$ such that $\mathbf{h}_i(k) = \{0, i \neq k; 1, i = k\}$, and \mathbf{R} becomes a scalar, $R = r(k)$. In this case, (3) simplifies to

$$k = \arg \max_{i=1, \dots, n} \left[\frac{1}{\mathbf{P}_{ii}^b + r(i)} \sum_{j=1}^n (\mathbf{P}_{ij}^b)^2 \right]. \quad (4)$$

This solution is equivalent to Eq. (3.3) of Berliner et al. (1999). We note two important features of this solution.

First, the location of the optimal observation is given not by the maximum variance \mathbf{P}_{kk} of the observed element \mathbf{x}_k , but rather by a certain balance between its variance and its covariance with other elements \mathbf{P}_{ki} , $i \neq k$. Second, the location of the optimal observation depends on the value of the observation error variance $r(k)$.

Although solution (3) may include multiple independent observations (corresponding to different rows in \mathbf{H}), the number of possible combinations of locations of the array elements in this case increases as a power of the system dimension. In practice that means the set of possible observations \mathbf{H} becomes so large that cycling through its elements in (3) is too computationally expensive even for systems of moderate size and an array of just two elements. This case of “parallel” optimization should be distinguished from the case of multiple sensor elements, when locations of the individual sensor elements are mutually dependent, and the computational cost remains linear in respect with the system dimension.

After finding the location of the first optimal observation from (3) or (4), and calculating \mathbf{P}^a by (2), one may use the analyzed covariance \mathbf{P}^a to find the location of the second optimal observation and so on (Fig. 1). As discussed above, this serial approach is the only feasible option for most applications. The serial assimilation and parallel assimilation of observations with uncorrelated errors result in the same analysis error covariance (Bierman 1977), so that the serial approach can be used to objectively determine an observation array of a given size (Bishop et al. 2001). Similarly, it is common practice in applications of the ensemble Kalman filter for observations to be assimilated either in small batches or one at a time (e.g., Houtekamer and Mitchell 1998; Whitaker and Hamill 2002). For completeness, we include a proof of this equivalence in the appendix. However, we note that this serial approach does not guarantee, nor does it generally yield, the array that corresponds to the global minimum of $\text{trace}(\mathbf{P}^a)$ (F. De Hoog 2007, personal communication). Hereafter, we therefore refer to the arrays designed using this method as objectively designed arrays, rather than optimal arrays.

In practice, the dimension of the state vector n can be very large, making it impossible to explicitly manipulate the $n \times n$ covariance matrices. Instead, one may store and manipulate the covariance matrix implicitly via a representative ensemble \mathbf{A} of the system state anomalies, $\mathbf{A}^{n \times m} = [\delta \mathbf{x}^{(1)}, \dots, \delta \mathbf{x}^{(m)}]$, where $\delta \mathbf{x}^{(i)} = \mathbf{x}^{(i)} - \bar{\mathbf{x}}$, m is the ensemble size and the overbar denotes the ensemble average. In this case the error covariance \mathbf{P} associated with the ensemble \mathbf{A} is given by

$$\mathbf{P} = \frac{1}{m-1} \mathbf{A} \mathbf{A}^T, \quad (5)$$

so that (3) and (4) for the location of the optimal observation transform to

$$\mathbf{H}^{\text{opt}} = \arg \max_{\{\mathbf{H}\}} \text{trace} \frac{\mathbf{H} \mathbf{A} (\mathbf{A}^T \mathbf{A}) (\mathbf{H} \mathbf{A})^T}{\mathbf{H} \mathbf{A} (\mathbf{H} \mathbf{A})^T + (m-1) \mathbf{R}_H} \quad (6)$$

and

$$k = \arg \max_{i=1, \dots, n} \frac{\mathbf{A}_i (\mathbf{A}^T \mathbf{A}) \mathbf{A}_i^T}{\mathbf{A}_i \mathbf{A}_i^T + (m-1) r(i)}, \quad (7)$$

respectively, where $\mathbf{A}_i = \mathbf{h}(i) \mathbf{A}$ is the i th row of \mathbf{A} . To maximize the computational effectiveness, one should precalculate the product $\mathbf{A}^T \mathbf{A}$ before cycling over the possible locations, in which case the computational cost of calculating (6) or (7) becomes linear with respect to the system state dimension n and is $O(nm^2)$.

Using this ensemble-based approach, after finding an optimal observation, one needs to update the background ensemble $\mathbf{A}^b \rightarrow \mathbf{A}^a$ in such a way that the analysis covariance calculated by (5) matches that given by (2) (Fig. 1). There are a number of formally equivalent solutions for such an ensemble update that are readily provided by different flavors of the ensemble square root Kalman filter (e.g., Tippett et al. 2003). The best choice among these solutions depends on the dimensions n , m , and p of the application. In the context of this paper, when typically $n \gg m$ and $p = 1$ (for a single element array), a computationally effective choice is the solution given by the ensemble transform Kalman filter (Bishop et al. 2001):

$$\mathbf{A}^a = \mathbf{A}^b \mathbf{T} \quad \text{and} \quad (8)$$

$$\mathbf{T} = \left[\mathbf{I} + \frac{1}{m-1} (\mathbf{H} \mathbf{A}^b)^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A}^b \right]^{-1/2}, \quad (9)$$

in which a relatively expensive operation of calculating the inverse square root of a matrix is performed in the ensemble space of dimension m , rather than in the state space of dimension n . This calculation itself can be conducted by using an eigenvalue (or a singular value) decomposition of the matrix in square brackets in (9). Because this matrix is symmetric, it can be decomposed as

$$\mathbf{I} + \frac{1}{m-1} (\mathbf{H} \mathbf{A}^b)^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A}^b = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T,$$

where \mathbf{U} is an orthonormal matrix ($\mathbf{U} \mathbf{U}^T = \mathbf{I}$) and $\mathbf{\Lambda}$ is a diagonal matrix. Subsequently, the inverse square root can be calculated as

$$\mathbf{T} = \mathbf{U} \mathbf{\Lambda}^{-1/2} \mathbf{U}^T.$$

The computational cost of this step is defined by that of the eigenvalue decomposition during the calculation of the transformation matrix \mathbf{T} , which is $O(m^3)$, and by that of the ensemble transformation $\mathbf{A}^b \mathbf{T}$, which is $O(nm^2)$. Because in most applications $m \ll n$, the cost of calculating the ensemble transform \mathbf{T} is typically much less than that of calculating the optimal observation and performing the ensemble transform itself.

Note that the calculation of a constrained array, as in section 4a, requires multiple ensemble updates for each subset of observations, which may substantially increase the computational cost. For example, assume that we have to find an optimal set of six observations within a two-dimensional gridded field of size 50×50 , constrained to be located at the same column of the grid (e.g., the same longitude). We may proceed by finding six optimal observations within each column and choosing the set with the minimal trace of the analyzed covariance. In this case, finding the best set will require $6 \times 50 = 300$ ensemble updates compared to just 6 ensemble updates in the case of independent, or unstructured, observations. It will also require 300 calculations of the optimal location, but each of these calculations will be restricted to cycling through one column (longitude) of the grid and therefore will become about 50 times less expensive than in the nonconstrained case.

Equations (7) and (8) form a theoretical basis for the ensemble-based method of objectively designing observation arrays in which the ensemble is used to factorize the system covariance by means of (5). The two main assumptions for this method are that the error covariance of the system is correctly represented by the ensemble of state vector anomalies, and that observations are assimilated in an optimal way. The statistical properties of the ensemble should depend on the application to which the observing system is being designed. For example, suppose an ensemble Kalman filter system is used for data assimilation. In this case, the background field is the ensemble mean and the ensemble should be formed by the anomalies about the ensemble mean at a given time. Alternatively, suppose the observation array is designed for a monitoring system that will produce gridded fields of the observed quantity, using climatology as the background field (e.g., Smith and Meyers 1996). In this case, the ensemble may be a time series of gridded anomalies generated either from a long model run or from observations. In the remainder of this paper, we restrict our attention to the latter case, where it is appropriate to use a stationary ensemble.

As noted above, the application of a Kalman filter does not necessarily imply that the system is Gaussian.

Similarly, there are no restrictions on the statistics of the initial ensemble of anomalies. Its distribution does not have to be Gaussian or even represent the distribution of the model perturbations; also, it does not have to be zero centered. The only requirement to the initial ensemble is to factorize the background error covariance by means of (5).

The described method can be easily generalized to calculate observations targeted for the maximum reduction of variance in an arbitrary quantity \mathbf{q} (e.g., volume transport, rate of overturning, water mass properties) related to the system state,

$$\mathbf{q} = \mathbf{Q}(\mathbf{x}). \quad (10)$$

This can be achieved by augmenting the state vector to include both \mathbf{x} and \mathbf{q} ,

$$\mathbf{x}_{\text{aug}} = \begin{pmatrix} \varepsilon \mathbf{x} \\ \mathbf{q} \end{pmatrix}, \quad (11)$$

and scaling down \mathbf{x} by setting ε so small that the trace of the covariance matrix of the new ensemble will be dominated by the contribution from \mathbf{q} . Alternatively, instead of scaling, one can replace the product $\mathbf{A}^T \mathbf{A}$ in (6) or (7) by $\mathbf{A}_q^T \mathbf{A}_q$, where \mathbf{A}_q is the ensemble formed only by the anomalies of the elements of interest of the augmented state vector: $\mathbf{A}_q = [\delta \mathbf{q}^{(1)}, \dots, \delta \mathbf{q}^{(m)}]$, $\delta \mathbf{q} = \mathbf{q} - \bar{\mathbf{q}}$.

Equation (8) may be used to calculate the analyzed anomalies for any (not necessarily optimal) set of observations. Together with (5), it provides a basis for an inexpensive comparison of the performance of different observation arrays without performing OSSEs.

In this paper, the performance of objectively derived arrays is compared with the performance of the CG-IOP mooring array for the TIO. The CG-IOP array is structured, with a number of mooring lines at various longitudes across the TIO to facilitate routine servicing. We therefore constrain our objectively designed arrays to have a structure that is similar, although not identical to the CG-IOP array. Namely, we allow 33 locations only to be structured in five meridionally aligned lines of six moorings, one line of three moorings, and to be located within 12° of the equator. To find the “optimal” locations for the first mooring line, we calculate the optimal locations for six sites for all possible lines and choose the line that provides the maximal reduction of trace(\mathbf{P}^a) and so on.

3. Data

We present results for applications to the TIO using model-generated ensembles of high-pass-filtered mixed

layer depth to represent intraseasonal variability (representing periods of 3–100 days) and an ensemble of observation-based gridded sea level anomalies (GSLAs) representing intraseasonal-to-interannual variability. The details of the models and observations are outlined below.

a. Model configurations

The reliability of an objectively designed array from a model-based ensemble depends on the representativeness of the ensemble. If the background error covariance represented in the ensemble is unrealistic, then the objectively designed array will be flawed. One way to assess the robustness of an objectively designed array is to derive multiple arrays from multiple ensembles generated by different models or different model simulations. For this study, we derive and evaluate an objectively designed array using output from three different models, with different resolution, different forcing, and integrated for different time periods. All three models are global configurations of different versions of the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model (Pacanowski 1995; Griffies et al. 2004), with the hybrid mixed layer model described by Chen et al. (1994). Specifically, the models used here are the Australian Community Ocean Model (ACOM), versions 2 and 3 (ACOM2 and ACOM3), and the Ocean Forecasting Australian Model (OFAM). The details of these models are summarized below and in Table 1.

1) ACOM2

Details of the configuration of ACOM2 are described by Schiller and Godfrey (2003). ACOM2 is based on version 2 of the Modular Ocean Model (MOM2; Pacanowski 1995), has constant zonal resolution of 2° , enhanced meridional resolution of $1/2^\circ$ within 8° latitude of the equator that gradually increases to $3/2^\circ$ toward the poles, and 25 vertical levels. ACOM2 is initially spun up for 20 yr using climatological forcing and is subsequently integrated for 1982–94 using 3-day-averaged wind stress from a blend of National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) fields (Kalnay et al. 1996) and the Florida State University climatology (Legler et al. 1989). Surface heat and freshwater fluxes are obtained by coupling ACOM2 to the atmospheric boundary layer model (ABLM) described by Kleeman and Power (1995).

2) ACOM3

Details of the configuration of ACOM3 are described by Schiller (2004). ACOM3 is based on version

TABLE 1. Summary of model configurations. FC refers to a flux correction.

	ACOM2	ACOM3	OFAM
Model code	MOM2	MOM3	MOM4
Zonal resolution	2°	0.5°	0.1°–2°
Meridional resolution	0.5°–1.5°	0.33°	0.1°–2°
No. of vertical levels	25	33	47
Wind forcing	NCEP–NCAR + FSU	ERS-1/2	ERA-40
Nonshortwave heat flux	ABLM + FC	ABLM + FC	ERA-40 + FC
Shortwave heat flux	As above	OLR + NCEP	ERA-40
Freshwater flux	As above	Monthly analyses	Levitus
Simulated period	1982–94	1992–2000	1992–2004

3 of MOM (MOM3), has a zonal resolution of $1/2^\circ$, meridional resolution of $1/3^\circ$, and 33 vertical levels. The model is initially spun up for 8 yr using climatological forcing and is subsequently integrated for 1992–2000 using daily *European Remote Sensing Satellite-1/2* (ERS-1/2) scatterometer winds, surface heat fluxes (latent, sensible, and longwave) from the ABLM of Kleeman and Power (1995), solar shortwave radiation derived from observed outgoing longwave radiation within 15° of the equator and NCEP elsewhere, and precipitation from Xie and Arkin (1997).

3) OFAM

Details of the configuration of OFAM are described by Oke et al. (2005; 2008). OFAM is based on MOM4 and has zonal resolution of $1/10^\circ$ between 90°E and 180° that gradually increases to 2° between 60°W and 0° , meridional resolution of $1/10^\circ$ south of 25°S that gradually increases to 2° north of 45°N , and 47 vertical levels. The model is initially spun up for 3 yr with climatological forcing and is subsequently integrated for 1992–2004 using 6-hourly wind, heat, and freshwater fluxes from the 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40; Kållberg et al. 2004).

4) MODEL VALIDATION

Results from ACOM2 and ACOM3 have been used to explore intraseasonal variability (Schiller and Godfrey 2003) and upper ocean dynamics (Schiller et al. 1998). These studies have undertaken extensive comparison with observations, including comparisons of time series of subsurface currents in the central Indian Ocean (Schiller and Godfrey 2003) and validation of the model's surface heat fluxes (Schiller et al. 1998) that represent an assessment of both the ocean's upper ocean dynamics and the ABLM. Additionally, Schiller (1999) presented a series of comparisons between modeled and observed subsurface temperature along frequently repeated XBT lines in the Indian Ocean. The

extensive model validation referred to above was typically very favorable. This demonstrates that these models generally provide a realistic representation of the intraseasonal variability in the TIO.

To date, the validation of OFAM has focused on the Australian region for a 13-yr reanalysis experiment, known as the Bluelink Reanalysis (BRAN), that was based on OFAM (Oke et al. 2005). Oke et al. (2005) showed that BRAN provides a good synoptic representation of the eddy field around Australia, and that the subsurface temperature variability along the equatorial Pacific Ocean is in good agreement with observations from the TOGA-TAO array in both BRAN and a free run of OFAM. More recently, Oke et al. (2008) presented a comprehensive evaluation of the latest BRAN experiment, showing that both BRAN and OFAM compare well to observations.

b. Remapping

To simplify our analysis, we remap all fields onto the ACOM2 grid. This is achieved by averaging the original fields onto the ACOM2 grid, using a boxcar filter that has a footprint that spans each model grid cell. This remapping has the effect of eliminating the details of the fields that have scales that are not resolvable on the ACOM2 grid. All gridded products therefore become more or less compatible and can be compared directly in the experiments that follow. This remapping also means that the range of possible locations included in the computation of optimal locations, by (7) or (6), are the same for all applications presented here.

One limitation here, for ACOM3, is that the ACOM3 fields are only available on a $1^\circ \times 1^\circ$ grid that has been subsampled (not averaged) from the original grid. As a result, the ACOM3 fields are noisier than they would be if we could remap them by averaging from the original grid.

c. Observations

Weekly maps of GSLA, on a $1/3^\circ$ Mercator grid, are used here to represent intraseasonal-to-interannual

variability and to demonstrate that the method described in section 2 can be applied as easily to gridded observations, as it can to model-based fields. The GSLA maps used here are produced by Aviso, are freely available using OpenDAP (opendap.aviso.oceanobs.com/thredds/dodsC/), and are generated by combining along-track sea level anomaly observations from all altimetric missions [Ocean Topography Experiment (TOPEX)/Poseidon, *ERS-1/2*, *Jason-1*, and *Envisat*] using optimal interpolation (OI). The details of the OI mapping are described by Ducet et al. (2000) and La Traon et al. (1998). Briefly, the length scales used for the OI range from 100 km in the zonal and meridional directions at 60°N–60°S to 250 (350) km in the meridional (zonal) direction at the equator.

4. Results

a. Model-based array design

We objectively compute an observation array using an ensemble of intraseasonal mixed layer depth (IMLD) from ACOM2, ACOM3, and OFAM and subsample and average fields from ACOM3 and OFAM onto the ACOM2 grid, as described in section 3b. The mixed layer is defined here as the depth over which the potential density decreases by 0.1 kg m^{-3} compared to the surface density and has been high-pass filtered to retain only variability on time scales of less than 100 days. The IMLD has been shown to provide a good representation of intraseasonal variability (e.g., Shinoda and Hendon 1998). We apply the method described in section 2 to objectively design an observation array for the ACOM2, ACOM3, and OFAM ensemble of IMLD and calculate the structured observation array as described in section 2. As mentioned above, to be consistent with the configuration of the CG-IOP array, the mooring locations are restricted to within 12° of the equator, while the system state is defined to include all grid points in the TIO north of 15°S. We use an ensemble of 1463 (12 yr), 976 (8 yr), and 1460 (12 yr) members of three daily IMLD fields from ACOM2, ACOM3, and OFAM, respectively.

The variance of IMLD for all three models is shown in Fig. 2. These variance maps show good qualitative agreement, with the maximum variance in the eastern equatorial Indian Ocean. This region has been identified by observations to be a site where intraseasonal oscillations have their largest amplitude (Webster et al. 2002; Sengupta et al. 2001). The variance in OFAM is about twice as large as in ACOM2 and ACOM3 owing to the finer horizontal resolution, rendering a richer representation of equatorial Kelvin waves and other small-scale phenomena.

Taking into account the likely vertical resolution of the array, and subgrid-scale and daily variability, we assume that the observed IMLD has an error variance of 25 m^2 , and that observation errors are not correlated in space. The method allows us to not only calculate the observation array objectively, but also to rank the individual sites or mooring lines of the CG-IOP array in terms of their impact on $\text{trace}(\mathbf{P}^a)$. We compute an array for each model and calculate the corresponding final theoretical analysis error variance [diagonal elements of \mathbf{P}^a from (2)], including the CG-IOP array, for the ensembles from all three models. This allows us to cross check the likely performance of each array using statistics from all models. Figure 2 shows maps of the theoretical analysis error variance for each experiment along with the locations and rankings of each array. We have also ranked the locations from the CG-IOP array using each model's ensemble (Fig. 2, second row). We note that each of the derived arrays have several mooring lines in common. The rankings of the locations all agree that the best mooring line is between 90° and 95°E (locations denoted 1–6 for all four arrays), and that the mooring line south of India is also very important (locations 7–12 for the CG-IOP, ACOM2, and ACOM3, and 13–18 for OFAM). These regions are where the IMLD have the largest variance.

The basin-averaged theoretical analysis error variance of IMLD and its percent reduction (given by the observed variance minus the final analysis error variance, divided by the observed variance, multiplied by 100) are presented in Table 2 for each array. This shows that when the ACOM2 ensemble is used to assess the arrays, the ACOM2 array performs best, as we expect, and the CG-IOP array performs worst. Similarly, when the ACOM3 (OFAM) ensemble is used to assess the relative performance of each array, the ACOM3 (OFAM) array performs best, and the CG-IOP array performs worst (Table 2). Again, this is as we expect, indicating that the arrays derived from the models are indeed an improvement on the CG-IOP array, at least in theory. In practice, consideration of the maps of the theoretical analysis error variance in Fig. 2 and the statistics in Table 2 suggest that we should not expect a very significant difference in the ability of the CG-IOP array or any of the objectively designed arrays to monitor variability in IMLD. Indeed, for IMLD, all arrays considered here only reduce the error variance by 19%–32%. This reduction is quite small and is due to the relatively short decorrelation length scales of IMLD. We conclude that the CG-IOP array is probably as good as any array that we can objectively design for resolving IMLD. However, based on our experiments, it is not clear that a mooring array, with only 33 moor-

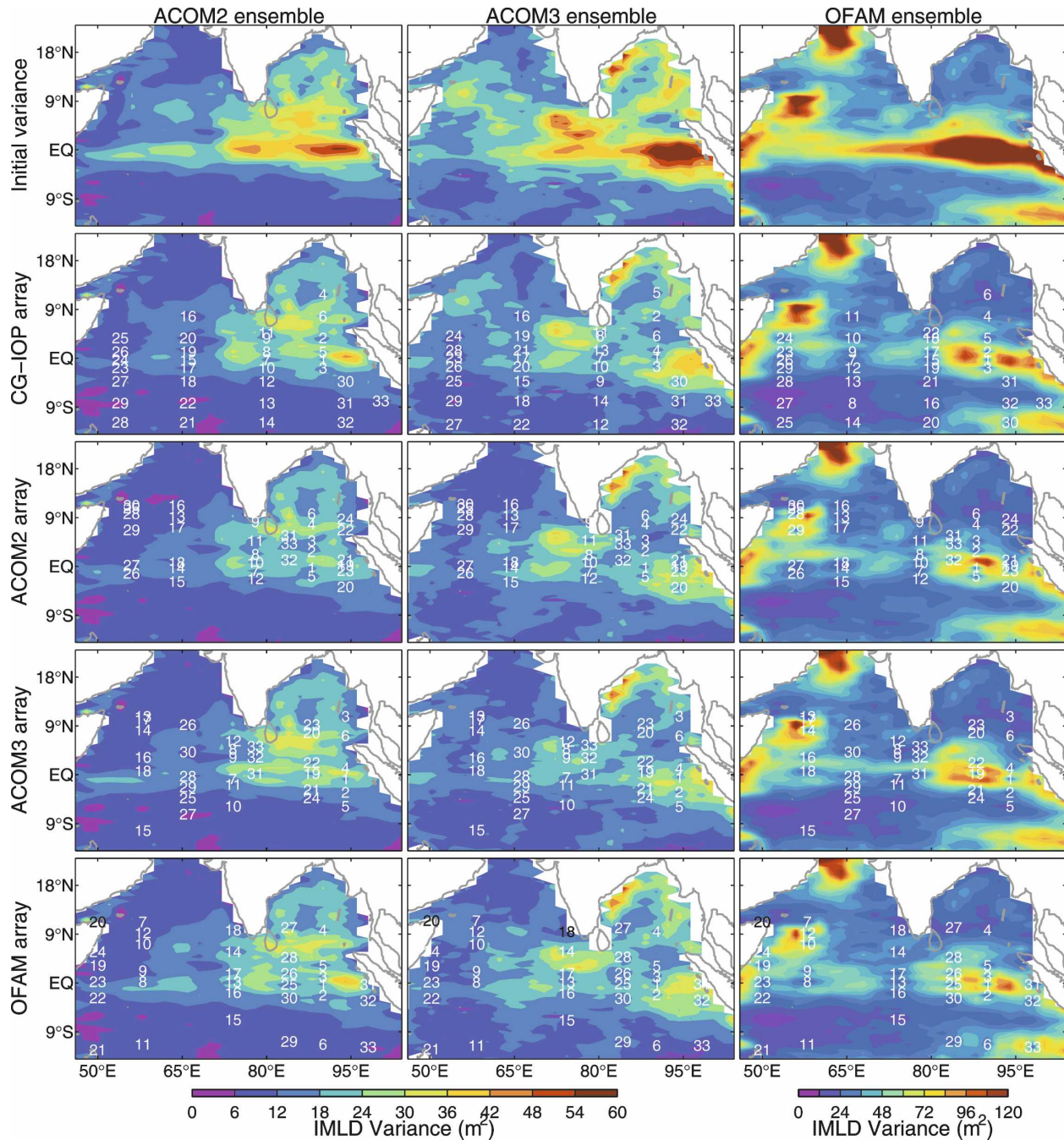


FIG. 2. (top) The variance of the IMLD in (left) ACOM2, (middle) ACOM3, and (right) OFAM, and the theoretical analysis error variance for each model using (top to bottom) the CG-IOP-, ACOM2-, ACOM3-, and OFAM-derived arrays, as labeled to the left of each row. The numbers in each panel denote the mooring locations and the ranking of each location (i.e., the locations marked “1” are the best location).

ings in the TIO, in isolation will resolve intraseasonal variability with the accuracy that may be needed for some applications. This may only be achieved using an integrated observing system, as Vecchi and Harrison (2007) suggest.

In addition to the experiments described above, we have also applied the objective array design method to several different model variables (e.g., upper ocean heat content over the top 400 m). However, we found that the models used here give quite a different repre-

TABLE 2. The basin-averaged observed variance and theoretical analysis error variance of IMLD (m^2) for different arrays, and the percent reduction in parentheses.

	Ensemble		
	ACOM2	ACOM3	OFAM
Observed variance	15.2	22.7	49.3
CG-IOP array	12.3 (19%)	16.7 (26%)	36.6 (26%)
ACOM2 array	11.5 (24%)	16.2 (29%)	36.6 (28%)
ACOM3 array	11.9 (22%)	15.4 (32%)	35.8 (27%)
OFAM array	12.1 (20%)	16.3 (28%)	33.6 (32%)

sensation of these variables. It is not clear whether these differences are due to inadequacies in the models, due to the models simulating different time periods, or that the differences in the model configurations (e.g., resolution) fundamentally change the model's ability to represent various phenomena (e.g., equatorial waveguide). Regardless of the reasons for the discrepancies, this indicates that we should not base an optimal array design on these particular variables. This highlights one of the limitations of our method; that is, the reliability of the objectively derived array directly depends on the representativeness of the ensemble used. Simply stated, the model-based array design is only appropriate if the model's representation of the variability is realistic. This suggests that model-based array design should probably always be performed with either a number of different models or with a very well-validated model.

We also found that the upper ocean heat content often has smaller correlation radii compared with the IMLD. We believe that this may be a result of correlation between the temperature in the upper 400 m and the thermocline depth, so that the contribution to the heat content from lowering (or rising) of the temperature is compensated to some degree by the contribution from lowering (or rising) of the thermocline. This possibly indicates that the upper ocean heat content may not be appropriate for characterizing the intraseasonal variability in TIO than IMLD.

We have also performed an experiment with a superensemble, where the anomalies from all three models are combined. We find that the results closely resemble the results from the OFAM experiments (not shown). This is because the variance in OFAM is about twice as large as the variance in ACOM2 and ACOM3. The ensemble members derived from the OFAM ensemble therefore dominate this superensemble.

b. Observation-based array design

We objectively compute an observation array using a 676-member ensemble of GSLA (i.e., anomalies from a

long-term mean) that are derived from altimetric observations as described in section 3c (i.e., 13 yr of weekly Aviso maps). The GSLA are unfiltered and provide a realistic representation of intraseasonal-to-interannual variability. We understand that the CG-IOP mooring array for the TIO is not intended to independently monitor variability on these time scales. Indeed other components of the integrated observing system, such as altimetry and Argo, are more suitable for resolving this variability. However, to demonstrate the versatility of the method, we apply it here to objectively design an observation array based on these maps of GSLA. A map of the variance of GSLA is shown in Fig. 3a. It can be seen that there is strong variability on seasonal to interannual time scales in the boundary currents in the TIO and across the entire basin at around 8° – 12° S where seasonal Rossby waves are known to be prevalent (Masumoto and Meyers 1998; Schouten et al. 2002).

In contrast to the previous example, we present both a structured array and an unstructured array, where the array need not conform to some predefined design. We restrict the possible locations of the observations to the TIO ($\pm 12^{\circ}$). We assume observations have an error variance of 9 cm^2 and that the observation errors are uncorrelated in space. The results are shown in Fig. 3. The basin-averaged variance of the GSLA in the TIO is shown in Table 3, along with the basin average of the theoretical analysis error variance and its percent reduction. These statistics show that all three arrays resolve a significant percentage of the observed signal. Table 3 also indicates that the objectively designed arrays outperform the CG-IOP array and that the unstructured array outperforms the structured array, as we expect.

While the objectively designed arrays are arguably an improvement on the CG-IOP array for resolving interannual variability, the increase in the percent variance explained is only 7%–9% over the CG-IOP array. This suggests that, although the CG-IOP array is not designed to resolve interannual variability in isolation from the rest of the global ocean observing system, it is still likely to do quite a good job.

Both the structured and unstructured arrays include many observations in the seasonal Rossby wave band between 6° and 12° S. Interestingly, they both include many observations off Somalia and relatively few observations off Indonesia, while both regions have relatively large variance (Fig. 3a). These results are explained by the correlation maps in Fig. 4, showing the correlations between sea level from a reference location (denoted by the star) and sea level across the entire TIO. Figure 4a shows that sea level offshore of Indo-

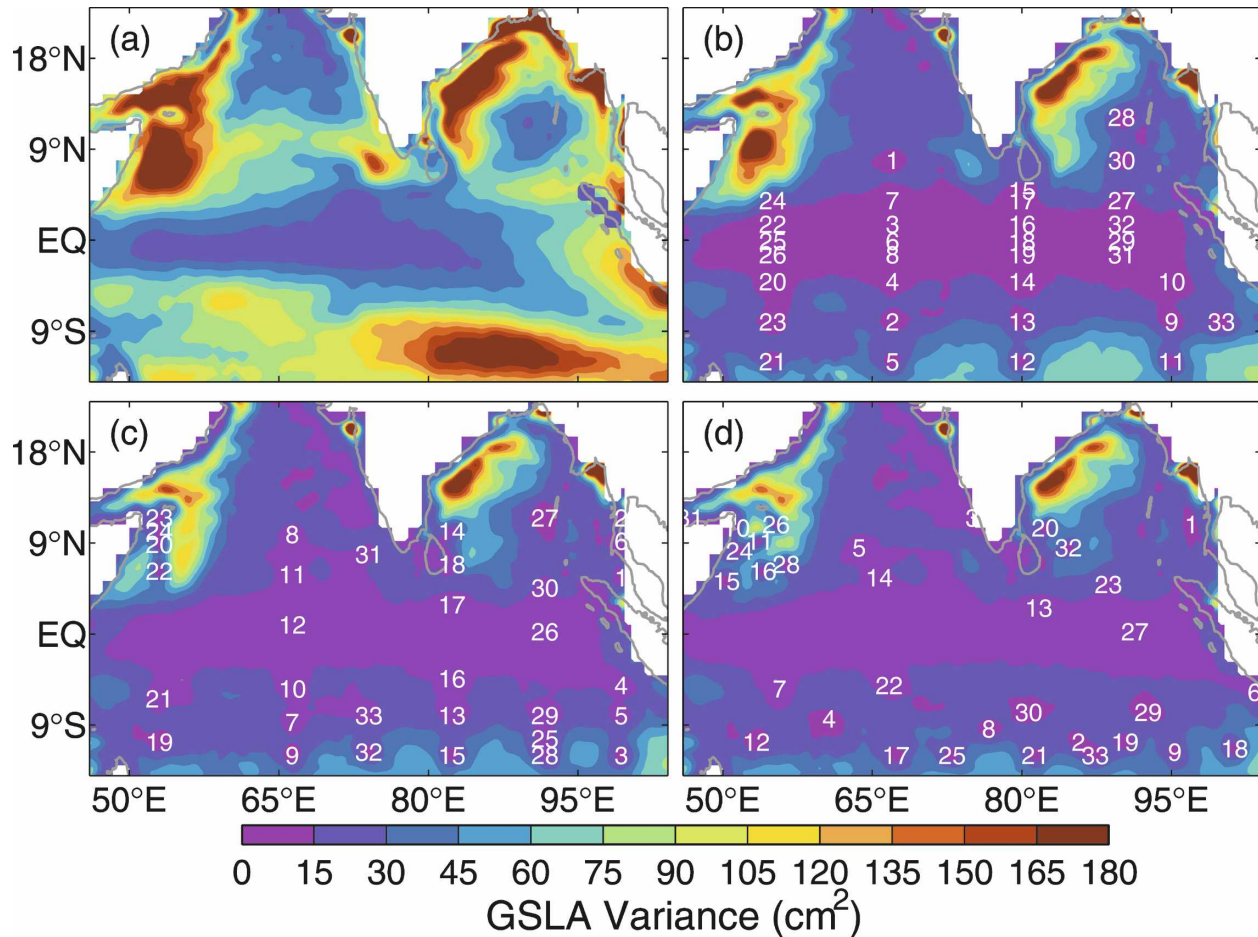


FIG. 3. (a) Variance of the GSLA and the theoretical analysis error variance using (b) the CG-IOP array and an objectively designed (c) structured array and (d) unstructured array. The numbers denote both the mooring locations and the ranking of each location as in Fig. 2.

nesia is well correlated with sea level along the coast and over a very broad region. The spatial structure of the correlation map shows a dipole structure. This structure is observed in several previous studies (Chambers et al. 1999; Feng et al. 2001; Wijffels and Meyers 2004; Rao and Behera 2005). Also, the footprint of the positively correlated region reflects Rossby–Kelvin wave patterns. This explains why even a single observation offshore of Indonesia can reduce the analysis error variance in an extended region including along the coast.

TABLE 3. As in Table 2, but for experiments using GSLA.

	Variance (cm^2)
Observed variance	81.0
CG-IOP array	34.9 (57%)
Unstructured array	27.1 (66%)
Structured array	29.5 (64%)

Figure 4b shows that sea level off Somalia is relatively uncorrelated with sea level across the TIO. The region off Somalia is dominated by mesoscale variability that spawns from the energetic and highly variable boundary currents in this region. While the mesoscale variability in this region is well organized (Schott and McCreary 2001), its variability is apparently somewhat chaotic and is characterized by short decorrelation length scales. This example explains why many observations may be required to reduce the analysis error variance off Somalia.

While we can understand why the method identifies certain regions as optimal locations on dynamical grounds and through analyses like that presented in Fig. 4, the number of observations in the objectively designed arrays in the seasonal Rossby wave band highlights a limitation of the method. This overpopulation of optimal observations near 9°S is partly attributable to the stationary nature of the ensemble used. For ex-

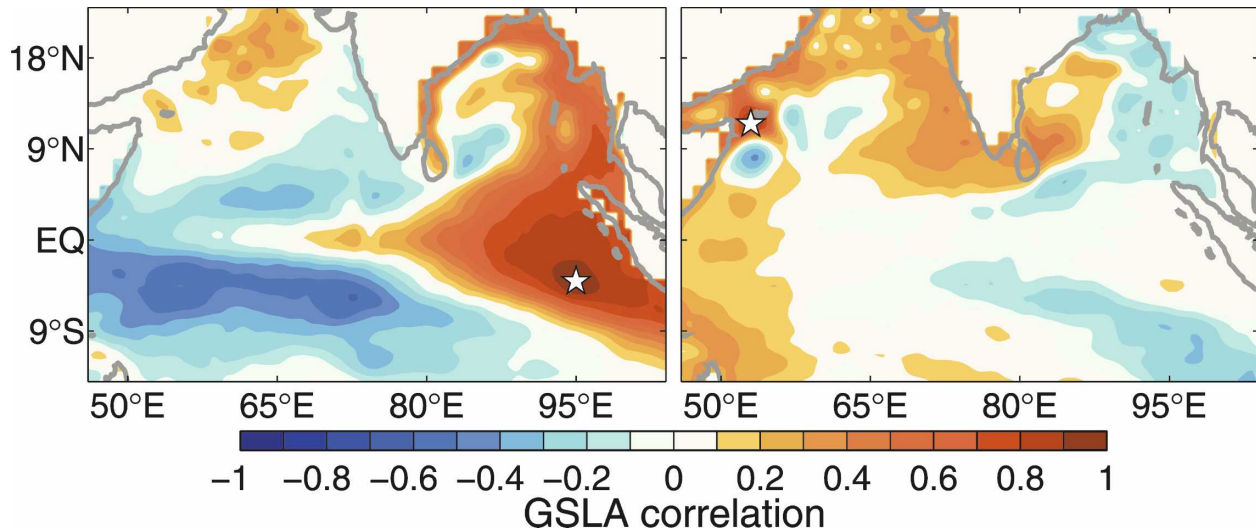


FIG. 4. Correlations between GSLA at a reference location, denoted by the star, and GSLA elsewhere.

ample, suppose observations from the array are assimilated into a dynamical model to improve the forecast. We might reasonably expect the model to accurately propagate a seasonal Rossby wave westward, possibly rendering some of the observations in this latitude band redundant. We expect an adaptive sampling application, where an ensemble Kalman filter is used in conjunction with a dynamic model (e.g., an ocean general circulation model) would identify fewer optimal locations in this latitude band.

By computing the reduction in $\text{trace}(\mathbf{P}^a)$ when each observation is “added” to an observation array, we can quantify the relative value of each observation location. An example of this calculation is presented in Fig. 5, showing both the cumulative percentage reduction of the first n observations and the percentage reduction of each individual observation. We present these statistics for the case of the objectively designed unstructured array using GSLA observations, shown in Fig. 3d. In Fig. 5 we show statistics for arrays that use two different criteria in their objective design. The first, denoted $\text{min trace}(\mathbf{P}^a)$, is the criterion described in section 2. The second, denoted $\text{max diag}(\mathbf{P}^b)$, identifies the “next best” observation (referred to as the optimal observation in Fig. 1) as the location that has the maximum background error variance. We find that the criterion of minimal trace is clearly superior, demonstrating the benefits of including information about the background error covariances, rather than just the background error variances. Indeed, the details of the observation array (not shown) are also quite different for these two criteria. Specifically, the maximum variance criterion yields more observations clustered in areas of large variance and short length scales, namely, in the Somali

Current and off Sri Lanka. Interestingly, despite these differences, the overall reduction of $\text{trace}(\mathbf{P}^a)$ is quite similar for both criteria.

Figure 5b shows that the criterion of maximum background variance, denoted $\text{max diag}(\mathbf{P}^b)$, is clearly sub-optimal. For example, the fifth observation yields a greater reduction in $\text{trace}(\mathbf{P}^a)$ than any of the first four observations. By contrast, the proposed method [$\text{min trace}(\mathbf{P}^a)$], while not optimal, is clearly more consistent in this regard, with each additional observation yielding a smaller reduction in the analysis error variance than any of the observations that are already selected.

For a given application, the type of analysis presented in Fig. 5 might provide insight into the number of observations required to monitor the variability in a given region. If, for example, the first n observations reduce the $\text{trace}(\mathbf{P}^a)$ considerably and all subsequent observations result in only very small further reductions, then it would be clear that only n observations are required. For the example considered here, Fig. 5a shows a quasi-exponential profile in the cumulative percentage reduction of the $\text{trace}(\mathbf{P}^a)$, indicating that there are diminishing returns for each additional observation, but that there is no obvious cutoff beyond which observations yield insignificant return.

5. Discussion and conclusions

Through a series of experiments using model-based fields, we objectively design an observation array for monitoring intraseasonal variability in the TIO. To assess the robustness of our model-based arrays, we use fields from three different models. The details of the resulting arrays are similar for all models. Specifically, they all give the highest ranking to the same mooring

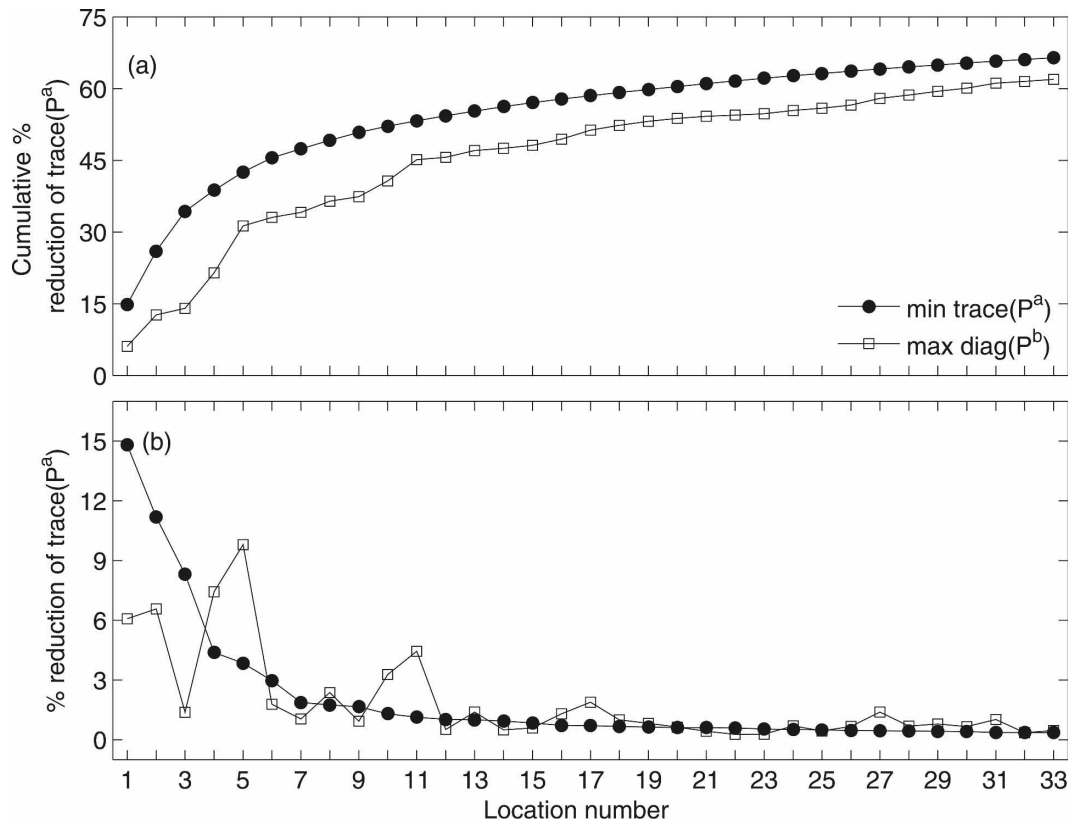


FIG. 5. (a) The cumulative percentage reduction of the trace(P^a) using the first n observation locations and (b) the percentage reduction of the trace(P^a) using the n th observation location for objectively designed arrays that minimize the trace of P^a , as described in section 2, or using a criterion that identifies the locations of maximum background error variance, denoted max diag(P^b).

line in the CG-IOP array at about 90°E . More generally, they all indicate that the key regions to monitor are near the equator to the east, northeast of the TIO, and south of India. Despite the moderately superior performance of the objectively designed arrays, in theory, we conclude that the CG-IOP array is probably as good, in practice, for resolving intraseasonal variability, as any array we can objectively design. This conclusion is a testament to the insight and intuition displayed by the CLIVAR panel in designing the CG-IOP array. We also present results from a series of experiments designed to monitor intraseasonal-to-interannual variability using fields of gridded observations. From these experiments we suggest that the key region to monitor is $6^\circ\text{--}12^\circ\text{S}$, where seasonal Rossby waves are dominant.

To better understand why the method for objective array design “selects” certain regions, we present correlation maps to show the region of influence of different locations. We suggest that this is a very useful tool that could easily be used when an observation array is being designed.

We also demonstrate how our method could be used

to estimate the number of observations required to adequately observe the variability in some region. This involves the computation of the relative value of each observation in an observation array that may also prove to be a useful metric for some practical aspects of observation array design.

In these experiments we use a simple method for objectively designing an observation array using ensemble data assimilation theory. The method can be applied to model-generated fields or to time series of gridded observations and allows one to generate either a structured or unstructured observation array. It is flexible and can be used in a range of different geophysical applications for objective array design.

We underline that having an ensemble that correctly represents the background error covariance is crucial for the applicability of the method. Using ensembles formed from output from a long model run or from observation-based gridded fields implies the ergodicity of the system and does not take advantage of possible time correlations. The application for which the method described here is most suitable is where an en-

semble optimal interpolation (EnOI) system (e.g., Oke et al. 2002; Evensen 2003; Oke et al. 2005; Oke and Schiller 2007b; Oke et al. 2008) is used to assimilate observations from a fixed observation array (e.g., an array of moorings). In this case, the EnOI system should use the same ensemble of anomalies as that used to objectively calculate the observation array. The system should also use the same background field as is used in constructing the ensemble of anomalies. Depending on the application, this background field may be a constant field (e.g., climatology) or a time-varying field (e.g., seasonal cycle). However, we argue that in many other situations, regardless of the analysis system to be used, the nonstationarity of the system's statistics may be not overly important and that it may be appropriate to use the time-averaged statistics, represented by an ensemble that is constructed from realizations of the system over a long period of time. In these cases, the described ensemble-based method represents a simple, objective, versatile, and computationally efficient tool to calculate an observation array.

Acknowledgments. This research is funded by Australia's CSIRO through appropriation funding and by the U.S. Office of Naval Research Ocean Modeling Program through Grants N000140410345 and N000140711054. The altimeter products were produced by Ssalto/Duacs and distributed by Aviso, with support from CNES.

APPENDIX

Equivalence of Serial and Parallel Assimilation

The equivalence of serial and parallel assimilation is demonstrated in a somewhat complicated way by Bierman (1977, p. 72). For completeness here, we present a more straightforward proof of this equivalence.

Suppose that a set O of p observations characterized by the observation matrix $\mathbf{H}^{p \times n}$ and the observation error covariance $\mathbf{R}^{p \times p}$ can be split in two sets O_1 and O_2 containing p_1 and p_2 observations, $p_1 + p_2 = p$, characterized by observation matrices $\mathbf{H}_1^{p_1 \times n}$ and $\mathbf{H}_2^{p_2 \times n}$,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix},$$

and error covariances $\mathbf{R}_1^{p_1 \times p_1}$ and $\mathbf{R}_2^{p_2 \times p_2}$, respectively. Suppose that these two groups of observations have uncorrelated errors, so that \mathbf{R} is block-diagonal:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & 0 \\ 0 & \mathbf{R}_2 \end{pmatrix}.$$

For our purpose, it is convenient to use an alternative expression for the analysis error covariance that is equivalent to (2):

$$\mathbf{P}^a = (\mathbf{P}^{b-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \equiv \mathcal{F}(\mathbf{P}^b, \mathbf{H}, \mathbf{R}).$$

For simplicity, we assume that both the background error covariance matrix and observation error covariance matrix are nonsingular. Then, the equivalence of serial and parallel assimilation means that the assimilation of O should yield the same covariance as the consecutive assimilation of O_1 and O_2 :

$$\mathcal{F}(\mathbf{P}^b, \mathbf{H}, \mathbf{R}) = \mathcal{F}[\mathcal{F}(\mathbf{P}^b, \mathbf{H}_1, \mathbf{R}_1), \mathbf{H}_2, \mathbf{R}_2]$$

or

$$(\mathbf{P}^{b-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = \{[(\mathbf{P}^{b-1} + \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1)^{-1}]^{-1} + \mathbf{H}_2^T \mathbf{R}_2^{-1} \mathbf{H}_2\}^{-1},$$

which simplifies to

$$\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} = \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1 + \mathbf{H}_2^T \mathbf{R}_2^{-1} \mathbf{H}_2,$$

which is the case due to the block-diagonal structure of \mathbf{R}^{-1} :

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{R}_1^{-1} & 0 \\ 0 & \mathbf{R}_2^{-1} \end{pmatrix}.$$

REFERENCES

- Ballabrera-Poy, J., E. Hackert, R. Murtugudde, and A. J. Busalacchi, 2007: An observing system simulation experiment for an optimal moored instrument array in the tropical Indian Ocean. *J. Climate*, **20**, 3284–3299.
- Barth, N., and C. Wunsch, 1990: Oceanographic experiment design by simulated annealing. *J. Phys. Oceanogr.*, **20**, 1249–1263.
- Berliner, L. M., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.*, **56**, 2536–2552.
- Bierman, G. J., 1977: *Factorization Methods for Discrete Sequential Estimation*. Academic Press, 241 pp.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Chambers, D. P., B. D. Tapley, and R. H. Stewart, 1999: Anomalous warming in the Indian Ocean coincident with El Niño. *J. Geophys. Res.*, **104**, 3035–3047.
- Chen, D., L. M. Rothstein, and A. J. Busalacchi, 1994: A hybrid vertical mixing scheme and its application to tropical ocean models. *J. Phys. Oceanogr.*, **24**, 2156–2179.
- CLIVAR, and Coauthors, 2006: Understanding the role of the Indian Ocean in the climate system—Implementation plan for sustained observations. WCRP Informal Rep. 5/2006, ICPO Publication Series 100, GOOS Rep. 152, 76 pp.
- Ducet, N., P. Y. Le Traon, and G. Reverdin, 2000: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.*, **105**, 19 477–19 498.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- Feng, M., G. Meyers, and S. E. Wijffels, 2001: Interannual upper

- ocean variability in the tropical Indian Ocean. *Geophys. Res. Lett.*, **28**, 4151–4154.
- Griffies, S. M., M. J. Harrison, R. C. Pacanowski, and A. Rosati, 2004: A technical guide to MOM4. NOAA/GFDL Ocean Group Tech. Rep. 5, 371 pp.
- Hackert, E. C., R. N. Miller, and A. J. Busalacchi, 1998: An optimized design for a moored instrument array in the tropical Atlantic Ocean. *J. Geophys. Res.*, **103**, 7491–7509.
- Hirschi, J., J. Baehr, J. Marotzke, J. Stark, S. Cunningham, and J.-O. Beismann, 2003: A monitoring design for the Atlantic meridional overturning circulation. *Geophys. Res. Lett.*, **30**, 1413, doi:10.1029/2002GL016776.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Källberg, P., 1984: Performance of some different FGGE observation subsets for a period in November 1979. *Data Assimilation Systems and Observing System Experiments with Particular Emphasis on FGGE*, Vol. 1, ECMWF, 203–228.
- , A. Simmons, S. Uppala, and M. Fuentes, 2004: The ERA-40 archive. ERA-40 Project Rep. Series 17, 31 pp.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kharné, S. P., and J. L. Anderson, 2006: An examination of ensemble filters based adaptive observation methodologies. *Tellus*, **58A**, 179–195.
- Kleeman, R., and S. B. Power, 1995: A simple atmospheric model of surface heat flux for use in ocean modeling studies. *J. Phys. Oceanogr.*, **25**, 92–105.
- Kuo, Y.-H., X. Zou, and W. Huang, 1998: The impact of global positioning system data on the prediction of an extratropical cyclone: An observing system simulation experiment. *Dyn. Atmos. Oceans*, **27**, 439–470.
- Langland, R. H., 2005: Issues in targeted observations. *Quart. J. Roy. Meteor. Soc.*, **131**, 3409–3425.
- Legler, D. M., I. M. Navon, and J. J. O'Brien, 1989: Objective analysis of pseudostress over the Indian Ocean using a direct-minimization approach. *Mon. Wea. Rev.*, **117**, 709–720.
- Le Traon, P. Y., F. Nadal, and N. Ducet, 1998: An improved mapping method of multisatellite altimeter data. *J. Atmos. Oceanic Technol.*, **15**, 522–534.
- Masumoto, Y., and G. Meyers, 1998: Forced Rossby waves in the southern tropical Indian Ocean. *J. Geophys. Res.*, **103**, 27 589–27 602.
- McIntosh, P. C., 1987: Systematic design of observational arrays. *J. Phys. Oceanogr.*, **17**, 885–902.
- Murtugudde, R., J. P. McCreary Jr., and A. J. Busalacchi, 2000: Oceanic processes associated with anomalous events in the Indian Ocean with relevance to 1997–1998. *J. Geophys. Res.*, **105**, 3295–3306.
- Oke, P. R., and A. Schiller, 2007a: A model-based assessment and design of a tropical Indian Ocean mooring array. *J. Climate*, **20**, 3269–3283.
- , and —, 2007b: Impact of Argo, SST and altimeter data on an eddy-resolving ocean reanalysis. *Geophys. Res. Lett.*, **34**, L19601, doi:10.1029/2007GL031549.
- , J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro, 2002: Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.*, **107**, 3122, doi:10.1029/2000JC000511.
- , A. Schiller, D. A. Griffin, and G. B. Brassington, 2005: An ensemble data assimilation system for an eddy-resolving ocean model of the Australian region. *Quart. J. Roy. Meteor. Soc.*, **131**, 3301–3311.
- , G. B. Brassington, D. A. Griffin, and A. Schiller, 2008: The Bluelink Ocean Data Assimilation System (BODAS). *Ocean Modell.*, **21**, 46–70.
- Pacanowski, R. C., 1995: MOM2 documentation, user's guide and reference manual, version 1.0. GFDL Ocean Group Tech. Rep. 3, 232 pp.
- Rao, S. A., and S. K. Behera, 2005: Subsurface influence on SST in the tropical Indian Ocean: Structure and interannual variability. *Dyn. Atmos. Oceans*, **39**, 103–135.
- Schiller, A., 1999: How well does a coarse-resolution circulation model simulate observed variability in the upper Indian Ocean? *Geophys. Res. Lett.*, **26**, 1485–1488.
- , 2004: Effects of explicit tidal forcing in an OGCM on the water-mass structure and circulation in the Indonesian throughflow region. *Ocean Modell.*, **6**, 31–49.
- , and J. S. Godfrey, 2003: Indian Ocean intraseasonal variability in an ocean general circulation model. *J. Climate*, **16**, 21–39.
- , —, P. C. McIntosh, G. Meyers, and S. E. Wijffels, 1998: Seasonal near-surface dynamics and thermodynamics of the Indian Ocean and Indonesian throughflow in a global ocean general circulation model. *J. Phys. Oceanogr.*, **28**, 2288–2312.
- , S. E. Wijffels, and G. A. Meyers, 2004: Design requirements for an Argo float array in the Indian Ocean inferred from observing system simulation experiments. *J. Atmos. Oceanic Technol.*, **21**, 1598–1620.
- Schott, F. A., and J. P. McCreary, 2001: The monsoon circulation of the Indian Ocean. *Prog. Oceanogr.*, **51**, 1–123.
- Schouten, M. W., W. P. M. de Ruijter, P. J. van Leeuwen, and H. A. Dijkstra, 2002: An oceanic teleconnection between the equatorial and southern Indian Ocean. *Geophys. Res. Lett.*, **29**, 1812, doi:10.1029/2001GL014542.
- Sengupta, D., R. Senan, and B. N. Goswami, 2001: Origin of intraseasonal variability of circulation in the tropical central Indian Ocean. *Geophys. Res. Lett.*, **28**, 1267–1270.
- Shinoda, T., and H. H. Hendon, 1998: Mixed layer modeling of intraseasonal variability in the tropical western Pacific and Indian Oceans. *J. Climate*, **11**, 2668–2685.
- Smith, N. R., and G. Meyers, 1996: An evaluation of expendable bathythermograph and Tropical Atmosphere-Ocean Array data for monitoring tropical ocean variability. *J. Geophys. Res.*, **101**, 28 489–28 501.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Vecchi, G. A., and M. J. Harrison, 2007: An observing system simulation experiment for the Indian Ocean. *J. Climate*, **20**, 3300–3319.
- Webster, P. J., and Coauthors, 2002: The JASMINE pilot study. *Bull. Amer. Meteor. Soc.*, **83**, 1603–1630.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Wijffels, S. E., and G. Meyers, 2004: An intersection of oceanic waveguides: Variability in the Indonesian throughflow region. *J. Phys. Oceanogr.*, **34**, 1232–1253.
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.