A Multi-Resolution Global Ocean Analysis and Forecast System based on Markov Random Fields

Ashwanth Srinivasan*

Tendral LLC., Miami, Florida

Mohamed Iskandarani

Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Virginia Key, Florida

Toshio M. Chin

Jet Propulsion Laboratory, California Institute of Technology, Pasedena, California

Eric P. Chassignet

Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida

**Corresponding author*: Ashwanth Srinivasan,a.srinivasan@tendral.com

ABSTRACT

In a multi-resolution assimilation methodology based on hierarchical Markov Random Fields, standard Gauss Markov random fields for background error at different resolutions levels are linked together stochastically to form a coherent multi-resolution background error model. This error model captures both global and local correlations and is combined with a multi-grid inspired algorithm for efficient and consistent analysis across a range of scales. Several experiments illustrating the method are presented using a model of the Intra-American Seas at resolutions ranging from 1/4^o to 1/32^o. These experiments show: 1) that the multi-resolution analysis framework accounts for differences in scales and precision of the component models and observations and allows better depictions/control of both global and local features in the analysis, 2) representativity issues can be addressed by assimilating data of different resolutions simply as measurements at different levels in the hierarchical structure, and 3) the information exchange among the component models provides a means to enforce consistency between the model solutions at different resolutions. Following this regional evaluation, the methodology is implemented in a global-scale ocean analysis system consisting of a 1/4° resolution global model and several of 1/16° regional models embedded within. Results from a one-year hindcast with this system are presented and show that the system is comparable to the leading current generation operational ocean analysis and forecast systems.

1. Introduction

In this paper, we introduce a multi-resolution data assimilation method for ocean forecasting and analysis. The method is designed to handle nested modeling configurations wherein fine-scale regional models are embedded in coarser scale basin or global models. The combination of nesting and multi-resolution analysis allows extraction of information from observations at multiple scales and to pass information consistently between nested grids. The end result is a system that can reconcile models and observations at different resolutions and produce consistent estimates across a number of scales in a computationally efficient manner.

Ocean reanalysis and forecasts are routinely used for many practical applications ranging from scientific research, offshore engineering design, environmental impact assessments, operational weather windows and many others. An important goal of these analysis and forecasts is to simultaneously and accurately reconstruct the ocean state spanning multiple interacting spatio-temporal scales. Studies suggest that horizontal resolutions better than 3 km at mid-latitudes are required to better represent narrow meandering currents, fronts, eddies (Hurlburt and Hogan 2000) and to simulate oceanic variability consistent with observational estimates (Chassignet and Xu 2017, 2021). However, deploying global models at these resolutions is extremely expensive computationally at present, and, for many multi-scale applications, it is standard practice to nest fine-scale regional models for areas of interest within a coarser global or basin scale model. It is relatively straightforward to extend this process and build a global scale multi-resolution system by nesting several regional scale models within a coarse global model. Assimilation of available measurements is a necessary component for hindcasting or forecasting applications. The simplest approach to assimilating data into a nested system is to treat each model independently and estimate its state based on available data. Alternatively, a more integrative approach can be considered where the entire set of resolutions are viewed as a scale decomposition of the multi-scale error process and use corrections obtained for any one scale (model) to guide the estimates for all other scales (models) in the system[1]. Furthermore, as we will see, this scale view of the model system provides a means to fuse measurements of varying resolutions and to produce estimates consistent across multiple scales. Here, we follow the integrative approach to assimilate data into a global scale nested modeling system.

Our approach to multi-resolution data assimilation is based on modeling the error process as Gauss Markov Random Fields (GMRF) at multiple resolutions. GMRFs are Gaussian random vectors defined over a set of discrete locations and are equivalent to undirected graphical models in which nodes represent random variables and edges capture conditional dependencies among the variables. They are both theoretically rich and practically versatile leading to their extensive use in statistical inference problems across many disciplines such as image processing, spatial statistics, economics, epidemiology and others (Hernandez-Lemus 2021; Kindermann and Snell 1980). In ocean data assimilation applications, MRFs are used to model the background error process by defining the error vector in terms of a generalized (non-causal) autoregression using a small set of strictly local (in space) neighbors typically on the underlying ocean model grid (Chin et al. 1999; Srinivasan et al. 2022). This results in a sparse graph in which each node in the graph is connected to only a few other nodes. Such sparse graphical structures lead to sparsity in the information matrix (the inverse of the covariance matrix) with concomitant advantages of parsimonious parameterization and efficient inference. The sparse information matrix generally has a dense (nearly full) covariance matrix as its inverse implying long-range spatial dependencies. However, local MRF models are often poor at capturing longer-range correlations, and even when they do, it results in the information matrix being ill-conditioned. In practice strictly local

[1]Here we use the term 'scale' to refer to the spatial discretization process rather than to an intrinsic property of the original random field

parameterizations are seen to be adequate for models with resolutions on the order of 10 km or more (Chin et al. 1999; Srinivasan et al. 2022), but, unsurprisingly, MRF models defined on a fine scale model grids (less than 10 km resolution) do not adequately capture longer-range correlations. One can certainly try to remedy the situation by using a larger neighborhood (denser graph) to capture long-range interactions, but this defeats the intent of exploiting sparsity for efficient algorithms. To overcome this grid-specificity and to capture a range of scales, MRF's can be defined on a hierarchy of grids for a multi-resolution analysis proceeding from the coarse scales to finer scales, successively filling in the details in a manner akin to the classic successive corrections schemes. Such an approach allows one to adequately model the multiscale correlations and to realize computational economy (e.g. Nychka et al. (2015)).

There is a good deal of flexibility in implementing a multi-scale data assimilation system. The hierarchy of grids or the multiple scales may represent physically meaningful quantities with measurements acquired at different scales or alternatively the coarser scales may be hidden variables without measurements introduced solely for efficient analysis. Further, the analysis scales may be disjoint, with estimates in coarser scales used to simplify analysis in the finer scales or they may be linked together into a coherent statistical model, with either deterministic or stochastic interactions between scales. In nested modeling systems, where the parent and child grids are run together, the existing grid hierarchy naturally lends itself to such a multi resolution analysis. One can define a GMRF for each of the ocean model grids and link them stochastically to form a consistent multi-resolution error model across scales. By placing the coarsest model at the top of the hierarchy and finer resolution models at levels below, a pyramid type of multi-level graph structure is realized. The information matrix associated with this multi-level graph can be represented as a sum of sparse "in-scale" information matrices defined on the grids at each level, representing linkages within each level, and a second sparse "cross-scale" information matrix representing statistical links between the levels. This setup can then be combined with a multi-grid type of two-pass estimation procedure for efficient analysis at multiple scales. Starting at the finest level, an analysis is computed at each level using the "in-scale" information matrix and passed upward using the "cross-scale" information matrix in succession until the coarsest level is reached. A subsequent downward sweep then completes a smoothing pass at each level and accounts for longer-range correlations in the finer scale models by capturing such behavior at coarser resolutions. There are several features of the multi-resolution analysis that are worth pointing out. The multi-resolution analysis accounts for differences in scales and precision of the component models and observations simultaneously in an integrated manner. Furthermore, since analyses are available at multiple resolutions, the multi-resolution analysis can be used to address the trade-off between regularity (smoothness) in the coarse resolution analysis and the locality (geometric details) of the finer resolutions. Another, useful aspect is the possibility of assimilating data or fusing information of drastically different resolutions simply as measurements at different levels in the hierarchical structure and thus addressing issue of representativity. Finally, since the analysis at each level is guided by analysis at other levels there is a two-way information exchange among the component models not only at the boundaries but also in the model interior unlike the standard two-way nested configurations where information exchange is only at the boundaries.

The multi-resolution analysis technique has been in use successfully for the past several years in an ocean forecasting system consisting of a suite of nested ocean model configurations including a global 1/4° resolution model, two 1/16° models for the Atlantic and Indian Oceans and several 1/32° resolution regional models implemented in response to unique needs of the offshore industry. The purpose of this paper is to shed some light on the theoretical underpinnings of the system and to present an evaluation of the methodology using a regional model with resolutions ranging from 1/4° to 1/32° and a simplified global system with a 1/4° global model and several 1/16° models nested within. The multi-resolution version of the GMRF assimilation and its evaluation presented herein builds on the evaluation of a GMRF based coarse resolution global system presented in Srinivasan et al. (2022). Although, this is a somewhat esoteric and particular implementation, we believe that this multi-resolution technique has a broader utility in addressing the emerging needs for nested fine resolution forecasts and reanalysis.

The layout of this article is as follows. Section 2 describes the multi-resolution estimation approach. We provide two simple examples in Section 3 that illustrate the implementation and performance of the multi-resolution approach. In section 4, we introduce a global scale ocean system implemented using this methodology and present some hindcast results. We then conclude the paper with a discussion on the system and its possible extensions.

2. Multi-Resolution GMRF Data Assimilation

a. GMRF Models

A GMRF, **x**, is an N-dimensional Gaussian random vector typically defined over a set of discrete locations such as a grid or lattice and has a Markov property: an element x_i at a grid location i is conditionally independent from all other elements given the values of neighbors of i. The conditional distribution of x_i given the values of its neighbors is

$$x_i \sim \mathcal{N}\left(\sum_{j\in\delta_i} \alpha_{ij} x_j \ , \ \beta_i^2\right) \tag{1}$$

where δ_i is the set of neighbors for location i (conventionally, the neighbor set δ_i does not include i.) Modeling with a GMRF consists of specifying the neighborhood system δ_i with its interaction parameters α_{ij} and the error variances β_i^2. It can be shown (Rue and Held 2005) that Equation (1) holds if and only if the joint distribution of $\mathbf{x}$ is Gaussian and the information matrix $\mathbf{L}$ is given by:

$$\mathbf{L}_{ij} = \begin{cases} \frac{1}{\beta_i^2} & \text{if } j = i \\ \alpha_{ij} & \text{if } j \in \delta_i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The conditional distribution of x_i in Equation (1) has the form of a Gaussian linear regression. The cardinality of the neighborhood δ_i determines the order of this regression and in turn the sparsity of the information matrix $\mathbf{L}$ whose diagonal elements are the conditional precisions $\frac{1}{\beta_j^2}$ and off-diagonal elements are the regression parameters α_{ij}. It is common to summarize the conditional independence relations with an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Here $\mathcal{V}$ is a set of nodes or vertices and $\mathcal{E}$ are edges connecting an unordered pair of vertices, i and j, if and only if $\mathbf{L}_{ij} \neq 0$. Therefore, sparsity of the information matrix $\mathbf{L}$ is related to the independence properties of the GMRF, $\mathbf{L}_{ij} = 0$ implies that x_i and x_j independent. Further, when dealing with GMRF models it is convenient to specify the mean $\mu \equiv \mathbb{E}[x]$ and covariance $P \equiv \mathbb{E}[(x-\mu)(x-\mu)^T]$ of Gaussian densities with its equivalent information form $\mu = \mathbf{L}^{-1}\mathbf{z}$ and $P = \mathbf{L}^{-1}$ where $\mathbf{z}$ is known as the information or potential vector. For our data assimilation application, the primary motivation for modeling the background error process as a GMRF is the ability of a small set of regression parameters and of small neighborhoods similar to Laplacian or biharmonic stencils usually, to encode the correlations spanning a range of distances and represent the background error covariance matrix of $\mathbf{x}$ in a numerically efficient, sparse form (Chin et al. 1999; Srinivasan et al. 2022).

b. Hierarchical GMRF Models

Since GMRF models are defined by specifying local interactions on discrete grids, the background error correlation scales that these models can represent are determined by the grid resolution. In particular, long range correlations might not be well represented when progressively finer resolution grids are used. To overcome this limitation GMRF models can be defined at multiple resolutions going from coarse to fine to represent a range of correlation scales. Links, either stochastic or deterministic, between the GMRF models at different resolutions can then be used to transfer longer range correlations captured by coarse resolution GMRF models to the finer grids. This procedure is general and can be used with both stand-alone models and nested models. We consider the general problem first and then map the procedure to a set of nested grids. Consider a multi-resolution random field at M levels of resolution with the finest scale denoted as 1 and the coarsest scale as M. Each scale m, $(1 < m < M)$, has a parent at $m-1$ and a child at $m+1$. The ith random variable at scale m is denoted as x_i^m and the collection of random variable at one scale is denoted by x^m. This arrangement can be thought of as a hierarchical pyramidal structure with coarser grids above the finer scale models (Figure 1). Such a model representing M scales from fine to coarse can be represented by a product of conditional distributions

$$\pi\left(x^1 \ldots x^M\right) = \pi\left(x^1\right) \prod_{n=2}^{M} \pi\left(x^m | x^{m-1}\right) \tag{3}$$

where $\pi(x^m)$, $m = 1 \ldots M$, are MRF models at each scale, and the terms $\pi(x^m|x^{m-1})$ represent the statistical interactions between different scales. For example, we have for a hierarchical model of two different scales:

$$\pi(x^1, x^2) = \pi(x^1)\,\pi(x^2|x^1) = \pi(x^2)\,\pi(x^1|x^2) \tag{4}$$

from Equation 4, the distribution of the fine scale random field indexed by 1 given a coarse scale field indexed by 2 is given by

$$\pi(x^1|x^2) = \frac{\pi(x^1, x^2)}{\pi(x^2)} = \frac{\pi(x^1)\pi(x^2|x^1)}{\pi(x^2)} \propto \pi(x^1)\pi(x^2|x^1). \tag{5}$$

In Equation 5, $\pi(x^1)$ is the joint in-scale spatial distribution of x^1 the fine scale process represented by a GMRF, and $\pi(x^2|x^1)$ represents the links between the scales. If we consider models where only successive scales are linked then the information matrix $\mathbf{L}$ associated with this hierarchical/pyramidal structure can be represented by a sparse matrix (Choi et al. 2010). As an example, for model with 4 levels, the information matrix is:

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} & & \\ \mathbf{L}_{21} & \mathbf{L}_{22} & \mathbf{L}_{23} & \\ & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{L}_{34} \\ & & \mathbf{L}_{43} & \mathbf{L}_{44} \end{pmatrix} \tag{6}$$

Here $\mathbf{L}_{mm}$ is the in-scale matrix and $\mathbf{L}_{m,m+1} = \mathbf{L}^T_{m+1,m}$ are matrices connecting the different scales. We model both in-scale joint spatial distribution and the scale to scale transitions with sparse Markov structures as described below.

c. GMRF models for in-scale conditional distribution

Two models are commonly used to represent the information matrix or the inverse of the background error covariance matrix in ocean data assimilation applications (Chin et al. 1999; Srinivasan et al. 2022). The first one is a first order random walk or the thin-membrane model where each node is modeled close to its cardinal neighbors (Laplacian stencil). Denoting the set of neighbors of $x^m(i)$ as δ_i, the conditional distribution can be written as:

$$\pi(x^m) \propto \exp\left(-\sum_{i=1}^{N^m} \sum_{j\in\delta_i} \alpha_{ij}^m (x_i^m - x_j^m)^2\right) \propto \exp\left(-\mathbf{x}^{mT}\mathbf{L}_{mm}^{tm}\mathbf{x}^m\right) \tag{7}$$

where N^m is the total number of grid points or nodes at level m, α_{ij} is a parameter related to error variances and in this model controls how strongly the model penalizes the gradient by minimizing the differences between the neighbors. $\mathbf{L}_{mm}^{tm}$ is the information matrix at scale m with the superscript tm indicating the thin-membrane model and $\mathbf{L}_{ij} = 0$ for $i \notin \delta_i$, $\mathbf{L}_{ij} = -\alpha_{ij}^m$ for $i \in \delta_i$ and $\mathbf{L}_{ii} = \alpha_{ii}^m |\delta_i| = \frac{1}{(\beta_i^m)^2}$ where β_i^m is the spatially varying background error standard deviation at each scale. The elements of $\mathbf{L}$ in this case is a scaled version of the negative Laplacian of the underlying continuous field.

A second order random walk model where each node is modeled to be close to the average of its neighbors, also referred to as the thin-plate model, is similarly defined:

$$\pi(x^m) \propto \exp\left(-\sum_{i=1}^{N^m} \sum_{j \in \delta_i} \alpha_{ij}^m (x_i^m - \frac{1}{|\delta_i|} \sum_{j \in \delta_i} x_j^m)^2\right) \propto \exp\left(-\mathbf{x}^{mT} \mathbf{L}_{mm}^{tp} \mathbf{x}^m\right) \tag{8}$$

where, as before N^m is the total number of grid points or nodes at level m, α_{ij} is again related related to error variances but in this second order model it controls how strongly curvature is penalized and $\mathbf{L}_{mm}^{tp}$ is the associated information matrix at scale m and superscript tp indicates the thin plate model and as before and $\mathbf{L}_{ij} = 0$ for $i \notin \delta_i$, $\mathbf{L}_{ij} = -\alpha_{ij}^m$ for $i \in \delta_i$ and $\mathbf{L}_{ii} = \alpha_{ii}^m |\delta_i| = \frac{1}{(\beta_i^m)^2}$ where β_i^m is the spatially varying background error standard deviation at each scale. The elements of $\mathbf{L}$ in this case is a scaled version of the negative biharmonic operator of the underlying continuous field.

Therefore, in both cases the error vector at each scale m is defined by a regression using a small set of neighbors at the same scale and the resulting information matrices, $\mathbf{L}_{mm}^{tm}$ and $\mathbf{L}_{mm}^{tp}$ are sparse and banded as they are built using either first or second order numerical difference operators. As mentioned before, the correlation scales implied by these operators are grid-dependent and will be discussed in the subsequent sections.

d. A Stochastic Link Model Between Different Scales

Consider two random fields at different resolutions to be linked so that the components of the coarse field will depend stochastically on the components of the finer grids within the coarser grid. A Gaussian stochastic model for the transformation can then be written as:

$$\mathbf{x}_i^m = F(x_i^{m-1}, i = 1 \ldots x_n^{m-1}) + \gamma_m \tag{9}$$

where F is a linear or non-linear transformation function, n_{m-1} is the number of fine grid nodes that are direct children of the parent grid at scale m and γ_m is Gaussian distributed error that represents the uncertainty in the transformation to the coarse scale. A simple linear averaging model represents the parent node in the hierarchical structure as a coarse representation of its children. Just as was done for the "in-scale" model, we can simply impose the condition that the parent node is close to its children. Denoting the children of x_i^m by δ_{ci}, the link between the parent and child scales is given by:

$$\pi(x^m | x^{m-1}) \propto \exp\left(-\sum_{i=1}^{N_m} \sum_{j \in \delta_{ci}} \gamma_{ij}^m (x_i^m - x_j^{m-1})^2\right) \tag{10}$$

$$\propto \exp\left(-\mathbf{x}^{m-1T} \mathbf{L}_{m,m-1}^T \mathbf{L}_{mm} \mathbf{L}_{m,m-1} \mathbf{x}^{m-1}\right) \tag{11}$$

the parameter γ_{ij}^m determines how severely we penalize the differences between the value at a node at scale m and the value at each of its children at scale $m+1$. $\mathbf{L}_{m,m-1}$ is a $N^m x N^{m-1}$ matrix representing scale to scale transitions and by the above modeling assumption is sparse with entries corresponding to parent-child pairs being $\frac{-1}{|\delta_{ci}|}$ and all others zero, further $\mathbf{L}_{m,m-1} = \mathbf{L}_{m-1,m}^T$.

A potential problem is the loss of Markovianity in the resolution transformation operation Lakshmanan and Derrin (1993). In general, resolution transformation might require approximation of the process at the coarse level by a different Markov random field with potentially different neighborhood. We however retain the same neighborhood structure. Although this might be less than optimal solution, from our experience and that of others (Lee et al. 2000) it does not appear to impact the results significantly. We mention in passing that several approaches to address the loss of Markovianity in resolution transformations have been proposed such as the Covariance Invariance Approximation (Lakshmanan and Derrin 1993), the use of conditional covariance (Choi et al. 2010) or the technique of Krishnamachari and Chellappa (1997) to approximate GMRFs from non-Markov fields. These approaches provide a starting point to address this issue in further developments using this methodology.

e. Correlation scales

To get an idea of the correlation scales implied by the different models, we use a one-dimensional process on a grid of size 128. We first examine the correlation scales implied by the in-scale GMRF models in this grid. The information matrices are nearly singular, so a regularization term ϵI, with $\epsilon = 0.1$ was added to these matrices before inverting them. The correlation plots in Figure 2 show that the thin membrane model is similar to an exponential correlation model. It puts more weight at the central grid point with correlations going to zero at roughly 10 grid points. For the thin-plate model, the initial decrease is not as rapid as the thin-membrane model but overall the correlations go to zero faster than the thin-membrane model. Given that both models do not have significant correlations beyond order of $O(10)$ grid points, it is clear that such models will have problems representing meso and larger-scale correlations in fine-resolution ocean model grids(< 5 km at mid-latitudes) but will be appropriate to capture mesoscale processes in coarser resolution model grids of 10 km or higher grid spacing. The impact of the multi-resolution methodology can be assessed by using progressively coarser grids of $64, 32, 16$ points respectively and transforming the correlations to the fine grid of 128 points.

The multi-resolution approach incorporating progressively coarser grids in the hierarchy is able to capture longer-range correlations as shown in Figure 2. Including a 16 points grid in the hierarchy more than doubles the correlation scales with both type of "in-scale" models while the other coarse grids of 32 and 64 points increase the correlation scales to a lesser extent. This suggests that the grids in the hierarchy can be chosen to match the desired correlation scales. Further, the variance related parameters α and γ in Equations 7, 8, and 10 can be specified as required to obtain a target covariance for the fine grid models. For example, we can decrease α_m and γ_m by a factor of 4 as we move from a finer scale to its parent since the spatial distance between a pair of neighboring nodes at scale m is twice the corresponding distance at scale $m+1$ and since our formulation involves the squares of differences (Equations 7 and 8).

f. Multi-resolution analysis

Computing optimal estimates for the hierarchical graph representing the multi-resolution model is equivalent to solving a linear system $\mathbf{L}\hat{\mathbf{x}} = \mathbf{z}$ where $\mathbf{L}$ is as in Equation 6 and $\mathbf{z}$ is the information vector (Choi et al. 2010). The $\mathbf{L}$ matrix, by construction, can be decomposed as a sum of in-scale information matrices and scale-to-scale transition matrices corresponding to the hierarchical structure. We can alternate between an in-scale analysis for each resolution and scale-to-scale transitions in a multi-grid like approach for efficiently computing the analysis. For the in-scale analysis, we have, $\mathbf{x}^m$, a collection of Gaussian distributed random variables: $\mathbf{x}^m \sim \mathcal{N}(0, L_m^{-1})$ where $\mathbf{L}_m$ is the in-scale information matrix for the given scale. Given a set of measurements $\mathbf{y} = \mathbf{H}\mathbf{x} + \nu$ where $\mathbf{H}$ is a linearized observation operator and ν is zero-mean Gaussian distributed observation error vector with a diagonal covariance R. The maximum *a posteriori* (MAP) estimate

$\hat{\mathbf{x}}^m$, equivalent to the mean of the posterior distribution is:

$$\hat{\mathbf{x}}^m = \arg\max_{x^m} p(\mathbf{x}^m|\mathbf{y}) = \mathbb{E}[\mathbf{x}^m|\mathbf{y}] = (\mathbf{L}_m + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1})\mathbf{z} \tag{12}$$

where the information vector is $\mathbf{z} = \mathbf{R}^{-1}\mathbf{y}$. We start at the finest resolution, compute an analysis and then proceed on a fine-to-coarse sweep computing level transitions with $\mathbf{L}_{m,m-1}\mathbf{x}_{m-1}$ products. After we reach the coarsest node, a downward coarse-to-fine smoothing step is executed. The analysis at any level, m, at downward pass ,n, can be represented as:

$$\hat{\mathbf{x}}_m^{dn} = \mathbf{L}_{mm}^{-1}(\mathbf{z}_m - \mathbf{L}_{m,m+1}\mathbf{x}_{m+1}^{dn} - \mathbf{L}_{m,m-1}\mathbf{x}_{m-1}^{u(n-1)}) \tag{13}$$

where the superscripts d and u indicate downward and upward passes, respectively. A similar expression is used for the upward pass. Since the different scales are statistically connected to each other, the changes at finer scales affect the nodes at coarser scales and vice-versa. Therefore, we need to perform the upward and downward passes a few times before the iterative inference algorithm converges. Typically, one or two iterations are sufficient for convergence. In situations where the child grids cover non-overlapping portions of the parent grid, as is usually the case in nested models, the analysis can be computed directly with a single upward and downward pass in a scale recursive fashion (Willsky 2002). The first pass moves upward from finer grids, merging analysis from each of the children into the parent grid and performing a second update at each parent node above the finest grids. The second pass starts at the coarsest node and moves downward to progressively finer scales, updating each node with smoothing information from coarser scale nodes. It is a generalization of the Rauch-Tung-Striebel smoother(Rauch et al. 1965) used to estimate the states of time series models when scale takes the role of time. Further, computational saving can be achieved by approximating the scale-to-scale transitions with $\mathbf{x}^m = A(m)\mathbf{x}^{m-1} + \gamma(m)$ and $\mathbf{x}^{m-1} = F(m)\mathbf{x}^m + \gamma(m)$. Where A and F matrices are regular prolongation and restriction operators used in multi-grid applications. The $\gamma(m)$ are zero mean Gaussian random errors in transformation from one resolution to another. The statistics of these errors are derived from model runs at different resolution after using the prolongation or restriction operators.

3. Illustrative Applications

In this section we illustrate the above methodology using two examples. The first is a simple 1d identical twin experiment which is then followed by a example with a set of realistic nested models for a regional domain.

a. Illustration with an Order One Auto-Regressive ($AR1$) Process

We consider a stationary time series given by the standard first-order difference equation $\mathbf{x}_{n+1} = \alpha\mathbf{x}_n + \omega$ with the value of α taken as 0.9 (Figure 3). The variance of this process is normalized to 1 so that the variance of ω, a white noise process, is $(1-\alpha^2)$. The sample is generated over a grid of $N = 128$ locations. The measurements of $\mathbf{x}_n$ are sampled using the measurement process $\mathbf{y}_n = \mathbf{x}_n + \nu_n$ where the variance in ν is R, so that the SNR is $R^{1/2}$. To represent patchy high resolution observations we sample once every 2 points in the first and last 32 points from the sample and once every 4 points in the middle to represent low resolution observations. We consider three cases to reconstruct the sample path from the measurements, a standard analysis with short correlation scales of 2 points, a large correlation scale of 6 points and then a multi-resolution case where we implement the algorithm of the previous section. For the standard cases we use the squared exponential correlation model, while for the multi resolution version we define a GMRF on a fine grid at the sample resolution and a coarse grid at one fourth resolution of the fine grid or 32 points. As seen from Figure 3, the standard analysis with short decorrelation scale fits the observations better in regions with high density of measurements while performance falls in the middle where only coarse resolution observations are available. The analysis with longer decorrelation scale produces a much smoother reconstruction which is better in the middle than in the edges where high resolution observations are not used advantageously. In contrast to either of these approaches, the multi-resolution analysis results in significantly better reconstruction than either of the standard analysis. Here, the coarse data used in the example of Figure 3 are of 4-point averages of the fine-level process $\mathbf{x}$ corrupted with noise. As expected, the results from the standard analysis cases of short and long decorrelation scales show that the shorter decorrelation scale imposes a strong locality while a longer decorrelation scales serves to provide a measure of regularity or smoothness to the analysis. An analysis method that provides a optimal combination of these properties is likely to be better and this is what is seen in the multi-resolution analysis. The interpolation in the fine-grid preserves locality while information from the coarse data provides smoothing information that removes offset errors in regions where noisy fine scale data are available. The results also suggest that representativity errors in observations can be mitigated by the two pass algorithm. Observations can be assimilated into the model grid that has similar resolution as the data and the two pass algorithm can then be used to spread the increments to other models.

b. Nested models of the Intra-American Seas (IAS)

We extend the analysis of the previous subsection to a multi-resolution modeling system of the IAS region consisting of $1/4°$, $1/16°$ and $1/32°$ resolution models. As before, we examine the impact of different correlation scales implicitly defined by GMRF parameterizations defined on each model grid (referred here as single scale) and then compare it with the multi-resolution analysis. For these experiments, the HYCOM model is configured for the region between 99^oW to 56^oW and 7^oN to 32^oN as shown in the top left panel of Figure 4. Bathymetry for the models were derived from the ETOPO dataset, the $1/32°$ bathymetry was first generated by smoothing the raw data two times and then further smoothed and sub-sampled to the $1/16°$ and the $1/4°$ grid to make bathymetry consistent between the models. The surface boundary conditions were derived from the ERA 5 reanalysis, while lateral boundary conditions were derived from the Mercator operational model and provided every 3 days. Further model configuration details are provided in section 4.

The model runs were initialized with the Mercator model state on 2017/12/01 and assimilated SLA data from all available altimeter platforms for 60 days(Figure 4). Additionally, SST and ARGO temperature and salinity profiles were also assimilated for this time period but we focus on the SSH field since the impact of the different correlation scales are manifested more visibly in the SSH reconstructions. Several experiments were run to examine the impact of single scale and the multi-resolution analysis (Table 1). For the single scale analysis, the models were run independently first with each model assimilating data with an equally weighted combination of thin membrane and thin plate GMRF models. This combination produces better results than either model by themselves. In particular, the bulls-eye type of artifacts in the analysis when using the thin membrane model is mitigated while simultaneously increasing the correlation scales associated with the thin plate model. This reason for this is clear from the correlation decays in Figure 2. These runs were then followed by a second run where the models were linked by the multi-resolution analysis and each using the weighted single scale combination described above. For these experiments, simple block averaging and interpolation were used to transition between the nested models.

We compare snapshots of SSH (ADT) from AVISO for 20180131 (day 60 after initialization) with model derived SSH (Figure 4 top panel) and time series of RMS innovations(Figure 6) computed with respect to along track data for each experiment. In general there is a good qualitative agreement between the mesoscale features in all the model runs and AVISO product in the Caribbean Sea and the Atlantic portion of the domain. The main differences between the model runs are in the depiction of the Loop Current (LC) and the associated cyclonic frontal eddies, two in the eastern flank and two in the western flank of the LC as seen in the AVISO product. The snapshots from the $1/4°$ (second row in Figure 4), for both single scale and multi-resolution analysis, depict the LC and the cyclonic eddies almost identically. The major difference between these model runs and the AVISO product is the weak cyclonic circulation in the western flank of the LC instead of the two well defined smaller cyclonic eddies visible in the AVISO product. The correlation scales implied by the GMRF defined on the $1/4°$ grid is adequate for SSH reconstruction similar to AVISO maps. In contrast to the $1/4°$ model, there are significant differences between the runs with single scale and multi resolution analysis in the case of $1/16°$ and $1/32°$ models (last two rows in Figure 4). The single scale runs depict a smaller separated eddy to the north of the main LC, particularly so in the case of the $1/32°$ model. However, both runs with the $1/16°$ depict smaller scale cyclonic features in the western flank and on top of the LC. The multi-resolution analysis results in a better qualitative match of the $1/16°$ and $1/32°$ models with the AVISO product. The smoothing pass from the $1/4°$ grid to the $1/16°$ and $1/32°$ grids adds larger scale information and better reconstructs the LC and the cyclonic features in the eastern flank of the LC bringing it closer to both AVISO depiction. The filtering step has added some small scale features to the $1/4°$ particularly evident in the western flank of the LC and the Caribbean Sea. The increments for SLA and temperature for the single scale and multi-resolution analysis (Figure 5) clearly show the smoothing effects in the multi-resolution analysis. Innovations are the lowest for the multi-resolution runs for both grids among all experiments (Figure 6). Overall the multi-resolution analysis produces both qualitatively and quantitatively consistent results across scales.

To further examine the impact of the multi-resolution analysis on transferring information across scales a second set of experiments were done with subsampled and localized observations. To generate the subsampled data one out every 40 points along each altimeter track was retained (Figure

7, top left panel) while the localized observations were generated by restricting the SLA data to the 89-83W and 23-29N to the LC portion of the of the IAS domain (Figure 7, top right panel). For this experiment, only the 1/4° and 1/16° were used and both models were first run independently assimilating subsampled SLA and localized SLA (and other observations) with single scale GMRF parameterization and these are compared to the multi-resolution run assimilating the subsampled data only in the 1/4° grid and the localized patchy data only into the 1/16° grid.

Snapshots of the model SSH from both models assimilating the subsampled data (middle row of Figure 7) show that the data are too coarse to reconstruct the SSH filed adequately as compared to the AVISO SSH snapshot (upper right panel Figure 7). Both the LC and the frontal eddies are not well represented in these runs. As can be expected from the results of the previous section, the single scale 1/4° grid does a better reconstruction with lower RMS innovations than the single scale 1/16° grid but RMS innovations (Figure 8) are significantly higher than the full data case(Figure 6). In the case of localized observations, the SSH reconstructions by both models in regions where data is available is not accurate compared to the full data case and the errors are even larger for regions with no data. However, the multi-resolution run assimilating local observations in the 1/16° grid and the subsampled data in the 1/4° grid is able to reconstruct the SSH field accurately at both resolutions. The LC and frontal eddies are in the right locations in both models and RMS errors (Figure 8) are significantly reduced compared to the single model subsampled and localized data cases. The filtering step of the multi-resolution analysis brings smaller scale information from the high-resolution observations and is essential for reconstruction the LC and the frontal eddies in the coarse grid. The smoothing step on the other hand adds larger scale information to the high-resolution details available in the 1/16° . The net result again is a consistent estimate across both models. The upshot of the above analysis is that the multi-resolution analysis is a practical way to address the regularity vs locality trade off that is inherent in any data assimilative system.

Coarse (fine) resolution observations can be assimilated into coarse (fine) models and the two pass update scheme can then be used to distribute the impact of the observations consistently to all the models in the nested system.

4. Global Multi-Resolution Data Assimilation System

We now extend the multi-resolution analysis to a global scale system consisting of a coarse resolution model with several high resolution nest for areas of interest (Figure 9). One of our goals in implementing the multi-resolution analysis system is to simultaneously address the issues related to correction of large-scale biases, better representation of interannual/seasonal variability and better control of mesoscale activity. Therefore, we implement models at two resolutions, a global $1/4°$ model is used to control larger scale errors and biases and to reconstruct the large scale variability while $1/16°$ models are used to depict and constrain mesoscale activity over four regions in the Atlantic and Indian Ocean namely, the Intra American Seas (IAS), East Africa (EFA), Brazil Region (BRS) and South Africa (TSA). In this setup, we have a two level pyramid with the global model at the top level and the four regional models at the lower level level of the pyramid.

a. Model configuration and implementation details

At all levels, we use the HYbrid Coordinate Ocean Model (HYCOM, http://hycom.org) code, a circulation model with Lagrangian vertical coordinates that is widely used by the oceanographic community (Bleck 2002; Chassignet et al. 2003). Our implementation of HYCOM is similar to configurations used in other HYCOM based operation centers such as NRL and NCEP (Chassignet et al. 2009). The model is configured with 32 hybrid layers with potential densities referenced to 2000db. The model bathymetry is a combination of ETOPO1 and GEBCO products with local corrections in the Indian Ocean, Gulf of Mexico and Brazil current regions. Surface atmospheric forcing is derived from the ECMWF Reanalysis v5 (ERA5) dataset (Hersbach et al. 2018) and consists of three hourly fields of air temperature and specific humidity at 2 m, surface net downward and long and shortwave radiation, precipitation and 10 m wind components. The atmospheric radiative fluxes are scaled using CERES energy balanced and filled gridded product (Kato et al. 2018). Monthly climatological river discharge is used to specify a virtual salinity flux to include the effects of river inflow. A combination of Laplacian and bi-harmonic mixing is used for horizontal momentum diffusion while a bi-harmonic formulation is used for horizontal thickness diffusion. These are specified with diffusion velocities of magnitude 0.003 m/s for the Laplacian term and 0.02 m/s for the bi-harmonic mixing terms, respectively. The K-Profile Parameterization (KPP, Large et al. (1994)) is used for vertical mixing with default values. Finally, a simple thermodynamic energy-loan model is used for heat balance in regions with ice in the global model. The models at each level of the tree are one way nested within the coarse model at a level above.

Remotely sensed sea level anomalies (SLA) and SST as well as in-situ temperature/salinity (T/S) profiles from the ARGO program are the backbone of the system and thus are systematically assimilated in all nested levels (Figure 10). Along track SLA data of 7 km nominal resolution from six altimeters, Jason-3, Cryosat, Sentinel-3a and Sentinel-3b, Altika and HY-2B are sampled from a 7 day window (±3 days) centered on the analysis data are pooled together and assimilated daily (data coverage for a typical day from these altimeters is shown in Figure 9). A reference Mean Dynamic Topography (MDT) based on Centre National d'Etudes Spatiales–Collecte Localisation Satellites 18 MDT (Mulet et al. 2021) is added to the anomalies to convert the anomaly fields into the Sea surface Height (SSH) fields. For the altimeters used in this experiment, the data provider, CLS, suggests instrument errors ranging from 2 cm to 4 cm. However, we used a constant 7 cm error for the altimeter data; this is slightly on the higher side than instrument accuracy but compensates for unknown errors in MDT in a crude way for this demonstration experiment. For

SST, the system assimilates both remotely sensed data and in-situ data with observation errors provided by the data producers. For ARGO temperature and salinity profiles an error of 0.1° and 0.05 were specified. The observation errors are inflated to give more weight to the observations closest to the analysis day with a Gaussian weighting scheme which smoothly sets the errors after 10 days to climatological levels.

In our implementation of the two-step algorithm, the two regional scale models are run first and data is assimilated to get a corrected state at these locations. This is then block averaged and merged into the coarser model grid with added uncertainty. This merged product is the prior for the coarser model and data is assimilated now in this coarser grid implicitly with larger correlation scale. The corrections estimated at large scales at the top of the tree are then transmitted back down the tree by simple interpolation.

b. Evaluation of the multi-resolution system

In this section, we present the results from an hindcast for the year 2018 using the multi-resolution approach detailed above. We check for consistency, quality and accuracy of the analysis by examining innovation statistics (rather than data residuals) since innovations represent errors before assimilation and thus can be considered as ccomparisons with quasi-independent data. Since the outputs from the system are available at multiple resolution it is important to choose the correct resolution to evaluate the results. At present, several gridded products derived from observations are available at $1/4°$ degree resolution and these products can be naturally compared with the global model results. Such comparisons of the global model were presented in Srinivasan et al. (2022), therefore, we mainly focus on the results from $1/16°$ models in regions of intense mesoscale activity.

Snapshots of SSH depictions from ARMOR3D, a multi observation global analysis product (Guinehut et al. 2012; Mulet et al. 2012), and the models for a typical day are shown in Figure 11. In all cases, the mesoscale field depicted by the model is qualitatively nearly identical to the ARMOR3D analysis. Both small and larger cyclonic and anti-cyclonic features are well depicted in all four regions. SLA innovations time series are shown for the four regions in the upper panel of Figure 12. With the exception of the South African domain, the SLA innovation levels in the assimilated product are close to the specified error levels and well below the observation standard deviation (Figure 12). However, in the Agulhas region, the innovations while still lower compared to the observed standard deviation, is higher than the specified error suggesting room for improvement.

The system assimilates in-situ and satellite derived SST products and similar to SLA, SST is well constrained over the four domains with a slightly higher innovations for the Agulhas region (Figure 12). Here the SST innovations are on the order of 1.0° but these errors are well below the observed standard deviation of SST in these regions.

Vertical profiles innovation statistics for temperature and salinity for the four regions are shown in Figure 13. As before, the maximal innovations are seen in the Agulhas region with RMS values of 1.5° for temperature and 0.3 for salinity. In general, the regions considered here are areas of vigorous mesoscale activity and error levels are on the higher side of commonly presented values. However, time series of globally averaged error is about 0.5° for temperature and 0.1 for salinity in line with the specified observation error(Figure 14). Globally, innovations averaged over the ($0-2000$ m) water column decrease with the depth and do not exceed 0.2° C for temperature and 0.03 for salinity. Between 0 and 500 m, departures from in situ observations rarely exceed 1° C and 0.2 with the exceptions of high variability regions such as the Agulhas region.

Data from surface drifting buoys are not assimilated, and are therefore a source of independent information on the consistency and performance of the system. Using instantaneous velocities at 00Z daily from the GDS data base from NOAA-AOML, we compare the model results with the observed drifter velocities at 15 m depth. The spatial map of error distributions shows the decrease in errors over all three oceans for the hindcast compared to a companion free run (Figure 15). The improvements in velocities range from 5 – 10%. Although these improvements are modest, they are reassuring since the velocity data has not been assimilated. The corrections to the tracer fields and pressure act to improve the unconstrained variables through the model dynamics.

As seen in the results presented in the preceding sections, the system produces analysis that closely match observations and in most cases are within specified error levels. Sea level, upper ocean temperature, upper ocean salinity and near surface velocities match observations to within 8 cm, 0.5° C, 0.2 and 0.20 m/s respectively. The error metrics that we obtain for our system here is similar to the metrics obtained by other global and regional ocean analysis and prediction systems (e.g., Oke et al., 2012, Lellouche et al., 2013; Blockley et. al., 2014).

5. Summary and Discussion

The goal of this paper is to introduce and demonstrate the use of a computationally efficient multi-resolution analysis technique for assimilating data into a global scale system of nested models. In this technique, the error process is modeled as a GMRF at multiple resolutions with statistical links between successive resolutions. This setup can be likened to a pyramid type of structure in which the coarse resolution models are on the top and progressively finer resolution model populating the lower levels. The links between variables both at a particular level and between successive levels are described by sparse Markov structures that allow extremely efficient analysis - the primary advantage of this method. The GMRF used for each resolution (level) implicitly defines a correlation scale for each of the levels, resulting in longer correlation scales for coarser models and progressively shorter correlation scales as the resolution of the models increase. A multi-grid inspired two-pass analysis scheme is used to impose some measure of locality in coarse resolution analysis and regularity to the fine resolution analysis. The two-pass scheme starts at the finest resolution and executes a upscaling/filtering operation successively from fine to the coarse nodes. In this filtering pass, local information from observations is retained at that level and non-local information is transmitted to the coarser resolution/larger scale models. After completing the upward pass for all models, the downward pass then adds large scale increments derived from longer distance correlations captured at coarser models to fine resolution models as smoothing operation. The multi-resolution technique addresses the grid-specificity and local nature of the single scale GMRF analysis and produces analysis that are consistent across all models in the system

In results reported here, comparisons between single scale and multi-resolution analysis showed a modest improvement for the coarser models ($1/4°$ and $1/16°$) not unlike results reported by Muscarella et al. (2014) and Mirouze et al. (2016). Results are significantly improved in the finer grid of $1/32°$ resolution. This is as expected since the single-scale analysis at $1/32°$ and higher resolutions will be highly local and will need smoothing information for adequately depicting mesoscale features. On the other hand, by controlling the effect of the smoothing pass, the finer details such as sub-mesoscale features in the fine scale analysis can be preserved to a greater or lesser extent depending on the application. The multi-resolution technique provides an alternative means to address errors at multiple scales and can be compared with other efforts such as multi-scale three-dimensional variational data assimilation(Li et al. 2015; Muscarella et al. 2014) in which the cost function is split into large and small scale components or other approaches where multiple correlation scales are used in constructing the background covariance (Martin et al. 2007; Mirouze et al. 2016).

The two-way information exchange between the sub components of the nested modeling system presented here has aspects in common with the two way nesting technique introduced by Sheng et al. (2005), in that both the inner and outer models are constrained by each other. The method presented here can be considered as an extension of their nudging approach to a fully data assimilative approach.

The multi-resolution technique was implemented in a year long hindcast experiment with a global scale system of nested models. Sea surface temperature, sea-level anomaly, temperature and salinity profiles are assimilated regularly to constrain the component models. The system is evaluated for consistency with respect to other depictions of the ocean state based entirely on observations. As seen in the results presented in the preceding sections, the system produces analysis that closely match observations and in most cases are within specified error levels. Sea level, upper ocean temperature, upper ocean salinity and near surface velocities match observations to within 8 cm, 0.5° , 0.2, and 0.20 m/s respectively. The error metrics that we obtain for our system here are similar to the metrics obtained by other global and regional ocean analysis and prediction systems (Oke et al. 2013; Lellouche et al. 2013; Blockley et al. 2014; Waters et al. 2015; Martin et al. 2015).

Finally, it is pointed out that the multi-resolution analysis presented here is not limited to a nested modeling system where both coarse and fine resolution models are available, it can also be used with a one way nested model by constructing analysis grids of varying resolutions and then successively iterating from coarse to fine grids. As noted in Moore et al. (2019) the need for analysis and reanalysis with higher horizontal and vertical resolution using regional models will continue to increase either in a stand-alone downscaling mode or embedded within coarser models. Further, responding to emerging observing platforms such as Surface Water and Ocean Tomography(SWOT), high resolution radiometers, and rapid sampling using in-situ probes will require data assimilation into models with resolution $O(1)$ km and finer. We believe that the multi-resolution analysis framework presented here provides a useful starting point to integrate the patchy high resolution observations and conventional observations with models of varying resolution to derive state estimates that are consistent across the models(scales) in a nested system.

ACKNOWLEDGEMENTS: The present work was supported by ONR grant N00014-19-1-2671. M. Iskandarani was also partially supported by NSF-1639722. Contributions by TMC was done as a private venture and not in the author's capacity as an employee of the Jet Propulsion Laboratory, California Institute of Technology.

References

Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnal-cartesian coordinates. *Ocean Modelling*, **4**, 55–88.

Blockley, E. W., and Coauthors, 2014: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geoscientific Model Development*, **7**, 2613–2638, doi:doi:10.5194/gmd-7-2613-2014.

Chassignet, E. P., L. T. Smith, G. R. Halliwell Jr., and R. Bleck, 2003: North Atlantic simulations with the hybrid coordinate ocean model(hycom): Impact of vertical coordinate choice,reference pressure and thermobaricity. *Journal of Physical Oceanography*, **33**, 2504–2526.

Chassignet, E. P., and X. B. Xu, 2017: Impact of horizontal resolution 1/12 to 1/50 on gulf stream separation, penetration, and variability. *Journal of Physical Oceanography*, **47**, 1999–2021, doi:https://doi.org/10.1175/JPO-D-17-0031.1.

Chassignet, E. P., and X. B. Xu, 2021: On the importance of high-resolution in large-scale ocean models. *Advances in Atmospheric Sciences*, **38**, 1621–1634, doi:https://doi.org/10.1007/s00376-021-0385-7.

Chassignet, P., E., and Coauthors, 2009: U.s. godae: Global ocean prediction with the hybrid coordinate ocean model (hycom). *Oceanography*, **22(2)**, 64–75.

Chin, T. M., A. J. Mariano, and E. P. Chassignet, 1999: Spatial regression and multiscale approximations for sequential data assimilation in ocean models. *Journal of Geophysical Research*, **104**, 7991–8014.

Choi, M. J., V. Chandrasekaran, and A. S. Willsky, 2010: Gaussian multiresolution models: Exploiting sparse markov and covariance structure. *IEEE Transactions on Signal Processing*, **58 (3)**, 1012–1024, doi:10.1109/TSP.2009.2036042.

Guinehut, S., A. L. Dhomps, G. Larnicol, and P. Y. Le Traon, 2012: High resolution 3d temperature and salinity fields derived from in situ and satellite observations.

Hernandez-Lemus, E., 2021: Random fields in physics, biology and data science. *Frontiers in Physics*, **9**, doi:10.3389/fphy.2021.641859, URL https://www.frontiersin.org/article/10.3389/fphy.2021.641859.

Hersbach, H., and Coauthors, 2018: Era5 hourly data on single levels from 1979 to present. copernicus climate change service (c3s) climate data store (cds). Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 10.24381/cds.adbb2d47.

Hurlburt, H. E., and P. J. Hogan, 2000: Impact of 1/8° to 1/64° resolution on gulf stream model-data comparisons in basin-scale subtropical atlantic ocean models. *Dynamics of Atmosphere and Oceans*, **32**, 283–329, doi:https://doi.org/10.1016/S0377-0265(00)00050-6.

Kato, S., and Coauthors, 2018: Surface irradiances of edition 4.0 clouds and the earth's radiant energy system (ceres) energy balanced and filled (ebaf) data product. *Journal of Climate*, **31**, 4501–4527, doi:https://doi.org/10.1175/JCLI-D-17-0523.1.

Kindermann, R., and L. Snell, 1980: *Markov random fields and their applications*, Vol. 1. American Mathematical Society, doi:http://dx.doi.org/10.1090/conm/001.

Krishnamachari, S., and R. Chellappa, 1997: Multiresolution gauss-markov random field models for texture segmentation. *EEE Trans Image Process.*, **6(2)**, 251–267, doi:10.1109/83.551696.

Lakshmanan, S., and H. Derrin, 1993: Gaussian markov random fields at multiple resolutions. *Markov Random Fields: Theory and Applications*, R. Chellapa, and A. Jain, Eds., Academic Press, New York, 131–157.

Large, W. G., J. C. McWilliams, and S. C. Doney, 1994: Oceanic vertical mixing: A review and a model with a non-local boundary layer parameterization. *Reviews in Geophysics*, **32**, 363–403.

Lee, S. H., A. Malallah, A. Datta-Gupta, and D. Higdon, 2000: Multiscale Data Integration Using Markov Random Fields. **All Days**, doi:10.2118/63066-MS, URL https://doi.org/10.2118/63066-MS, sPE-63066-MS, https://onepetro.org/SPEATCE/proceedings-pdf/00ATCE/All-00ATCE/SPE-63066-MS/1903565/spe-63066-ms.pdf.

Lellouche, J. M., and Coauthors, 2013: Evaluation of global monitoring and forecasting systems at Mercator Ocean. *Ocean Science*, **9**, 57–81.

Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara, 2015: A multiscale variational data assimilation scheme: Formulation and illustration. *Monthly Weather Review*, **143 (9)**, 3804 – 3822, doi:10.1175/MWR-D-14-00384.1, URL https://journals.ametsoc.org/view/journals/mwre/143/9/mwr-d-14-00384.1.xml.

Martin, M. J., A. Hines, and M. J. Bell, 2007: Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society*, **133**, 59–89.

Martin, M. J., and Coauthors, 2015: Status and future of data assimilation in operational oceanography. *Journal of Operational Oceanography*, **8**, 28–48, doi:doi:10.1080/1755876X.2015.1022055.

Mirouze, I., E. W. Blockley, D. J. Lea, M. J. Martin, and M. J. Bell, 2016: A multiple length scale correlation operator for ocean data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, **68 (1)**, 29 744, doi:10.3402/tellusa.v68.29744, URL https://doi.org/10.3402/tellusa.v68.29744, https://doi.org/10.3402/tellusa.v68.29744.

Moore, A. M., and Coauthors, 2019: Synthesis of Ocean Observations Using Data Assimilation for Operational, Real-Time and Reanalysis Systems: A More Complete Picture of the State of the Ocean. *Frontiers in Marine Science*, **6 (90)**, doi:doi:10.3389/fmars.2019.00090.

Mulet, S., M. H. Rio, A. Mignot, S. Guinehut, and R. Morrow, 2012: A new estimate of the global 3d geostrophic ocean circulation based on satellite data and in-situ measurements. *Deep Sea Research Part II : Topical Studies in Oceanography*, 70–81.

Mulet, S., and Coauthors, 2021: The new CNES-CLS18 Global Mean Dynamic Topography,[preprint]. *Ocean Science Discuss*, **2020-117**, 1927–1932.

Muscarella, P. A., M. Carrier, and H. Ngodock, 2014: An examination of a multi-scale three-dimensional variational data assimilation scheme in the kuroshio extension using the naval coastal ocean model. *Continental Shelf Research*, **73**, 41–48, doi:https://doi.org/10.1016/j.csr.2013.11.009, URL https://www.sciencedirect.com/science/article/pii/S0278434313003749.

Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain, 2015: A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, **24 (2)**, 579–599, doi:10.1080/10618600.2014.914946, URL https://doi.org/10.1080/10618600.2014.914946.

Oke, P. R., and Coauthors, 2013: Towards a dynamically balanced eddy-resolving ocean reanalysis: Bran3. *Ocean Modelling*, **67**, 52–70, doi:https://doi.org/10.1016/j.ocemod.2013.03.008.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, **3 (8)**, 1445–1450, doi:10.2514/3.3166, URL https://doi.org/10.2514/3.3166, https://doi.org/10.2514/3.3166.

Rue, H., and L. Held, 2005: *Gaussian Markovian Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Sheng, J., R. Greatbatch, and X. Zhai, 2005: A new two-way nesting technique for ocean modeling based on the smoothed semi-prognostic method. *Ocean Dynamics*, **55**, 162–177, doi:https://doi.org/10.1007/s10236-005-0005-6.

Srinivasan, A., T. M. Chin, E. P. Chassignet, M. Iskandarani, and N. Groves, 2022: A statistical interpolation code for ocean analysis and forecasting. *Journal of Atmospheric and Oceanic Technology*, **39 (3)**, 367 – 386, doi:10.1175/JTECH-D-21-0033.1, URL https://journals.ametsoc.org/view/journals/atot/39/3/JTECH-D-21-0033.1.xml.

Waters, J., D. J. Lea, M. J. Martin, D. Storkey, and J. While, 2015: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, **141**, 333–349, doi:doi:10.1002/qj.2388.

Willsky, A. S., 2002: Multiresolution markov model for signal and image processing. *Proceedings of the IEEE*, **90**, 1396–1458.

LIST OF TABLES

TABLE 1. List of experiments discussed in section 3b

S.No	*Experiment*	*Model*	*Observations*	*Assimilation*
1	G04-SS	1/4°	full resolution SLA	Single Scale
2	G04-MR	1/4°	full resolution SLA	multi-resolution
3	G16-SS	1/16°	full resolution SLA	Single Scale
4	G16-MR	1/16°	full resolution SLA	multi-resolution
5	G32-SS	1/16°	full resolution SLA	Single Scale
6	G32-MR	1/16°	full resolution SLA	multi-resolution
7	G04-sub	1/4°	subsampled SLA	single scale
8	G04-Local	1/4°	local SLA	single scale
9	G16-sub	1/4°	subsampled SLA	single scale
10	G16-local	1/4°	local SLA	single scale
11	G04-MR-sub	1/4°	subsampled + local SLA	multi-resolution
12	G16-MR-sub	1/16°	subsampled SLA + local SLA	multi-resolution

LIST OF FIGURES

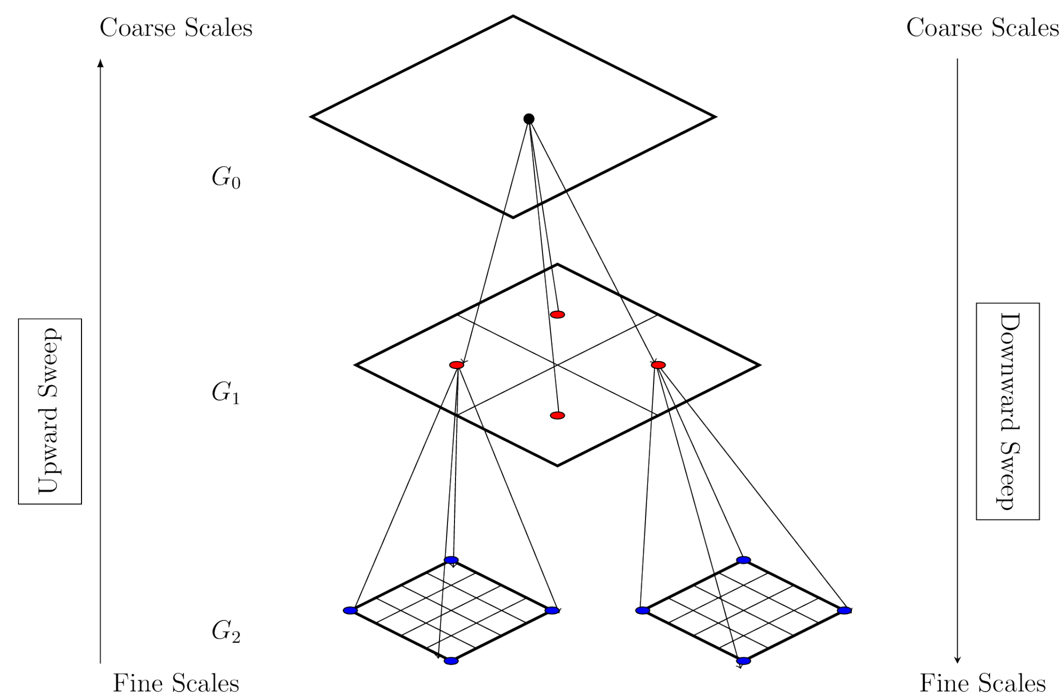


FIG. 1. A set of three nested models grids mapped onto a hierarchical graphical structure. The coarsest resolution, G_0, is at the top of the structure and finer resolutions are at progressively lower levels. Each level has a parent and(or) a child. A parent node in this structure is a coarse representation of its four children.

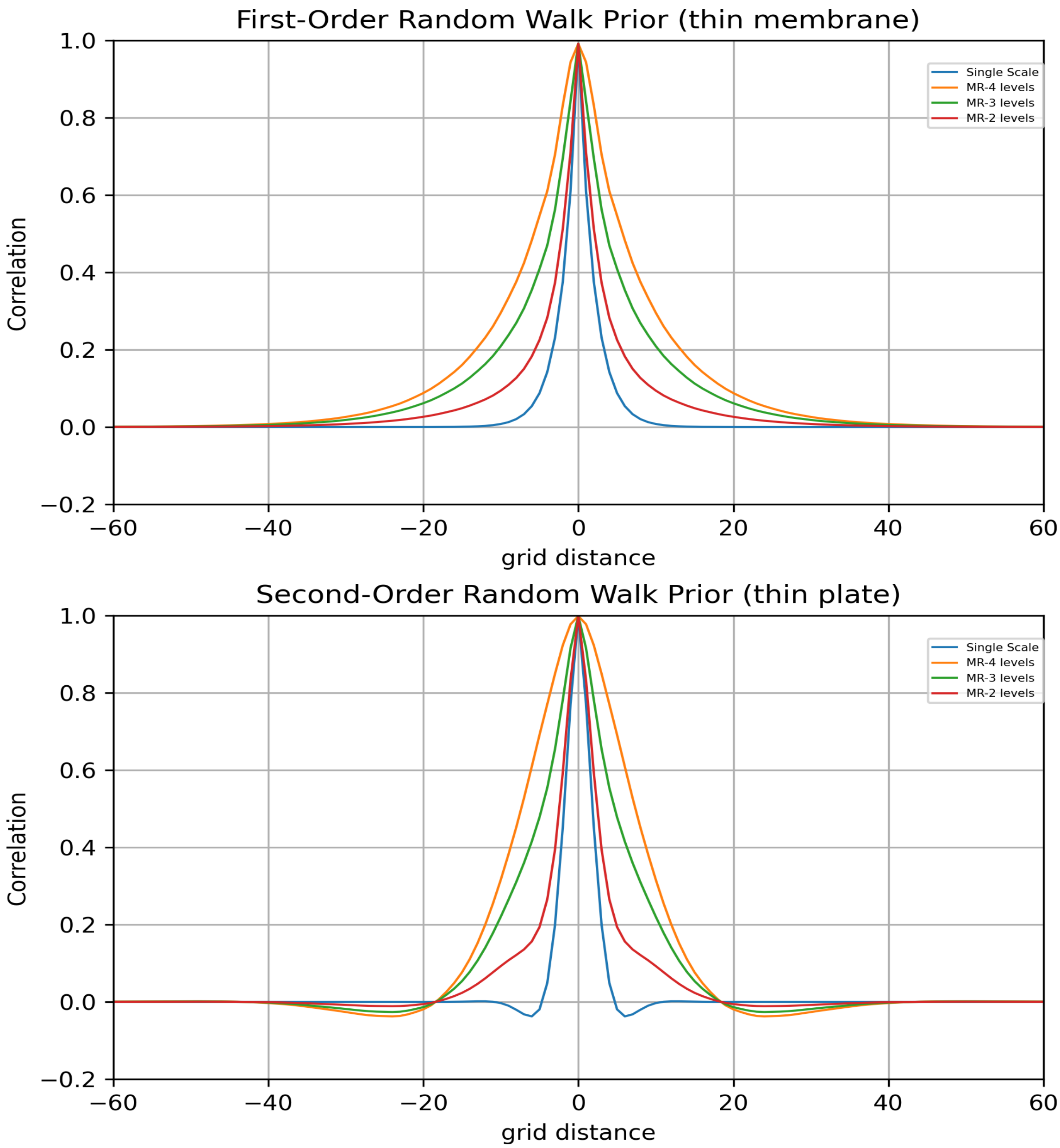


FIG. 2. Correlations from the first order model (top panel) and second order model(bottom panel). The correlations are implicitly defined by the underlying grid and decay rapidly to zero at around 10-15 grid point distances. This implies that these models by themselves will be insufficient to capture the longer-distance correlations particularly when used with fine-resolution models (< 10 km) resolution. The multi-resolution correlations shown in the Figure is able to capture longer-distance correlations at coarser resolutions. Two different combinations of multi-scale correlations are shown for different values of parameter γ in equation 6.

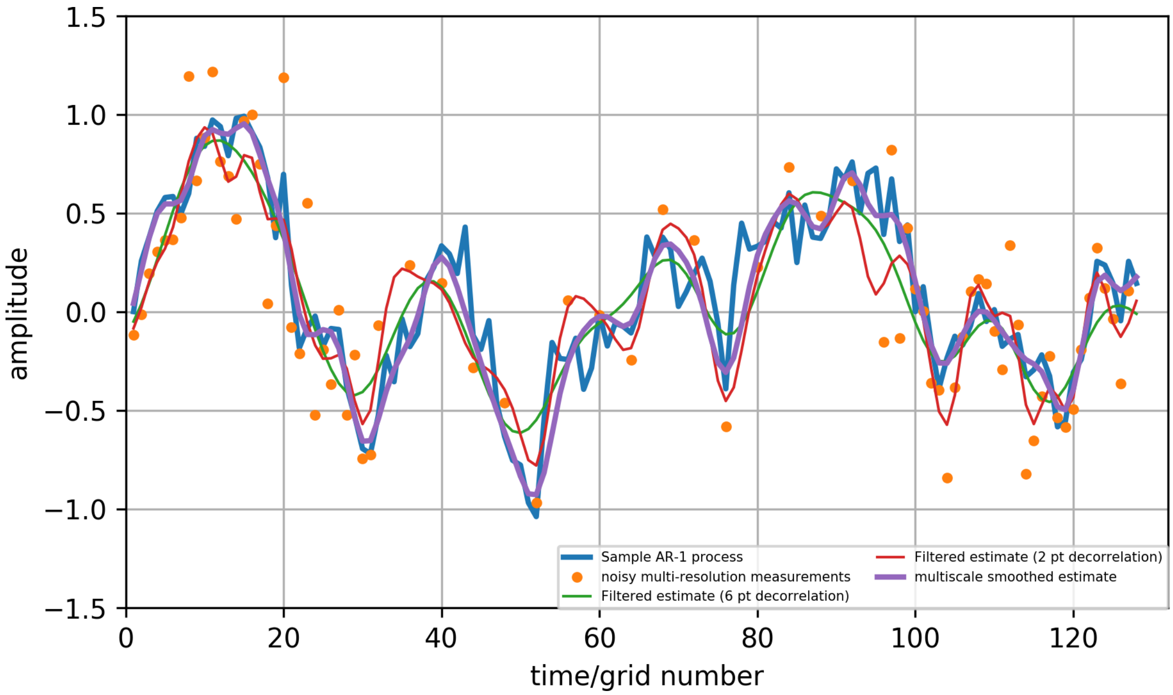


FIG. 3. An $AR-1$ process and its sampling by noisy measurements at different resolutions. The measurements are available at high resolution at the beginning and end of the path and at every fourth sample point at the middle of the sample path. Reconstruction of the process for two correlation scales (short and long) and using the multi-resolution approach are shown.

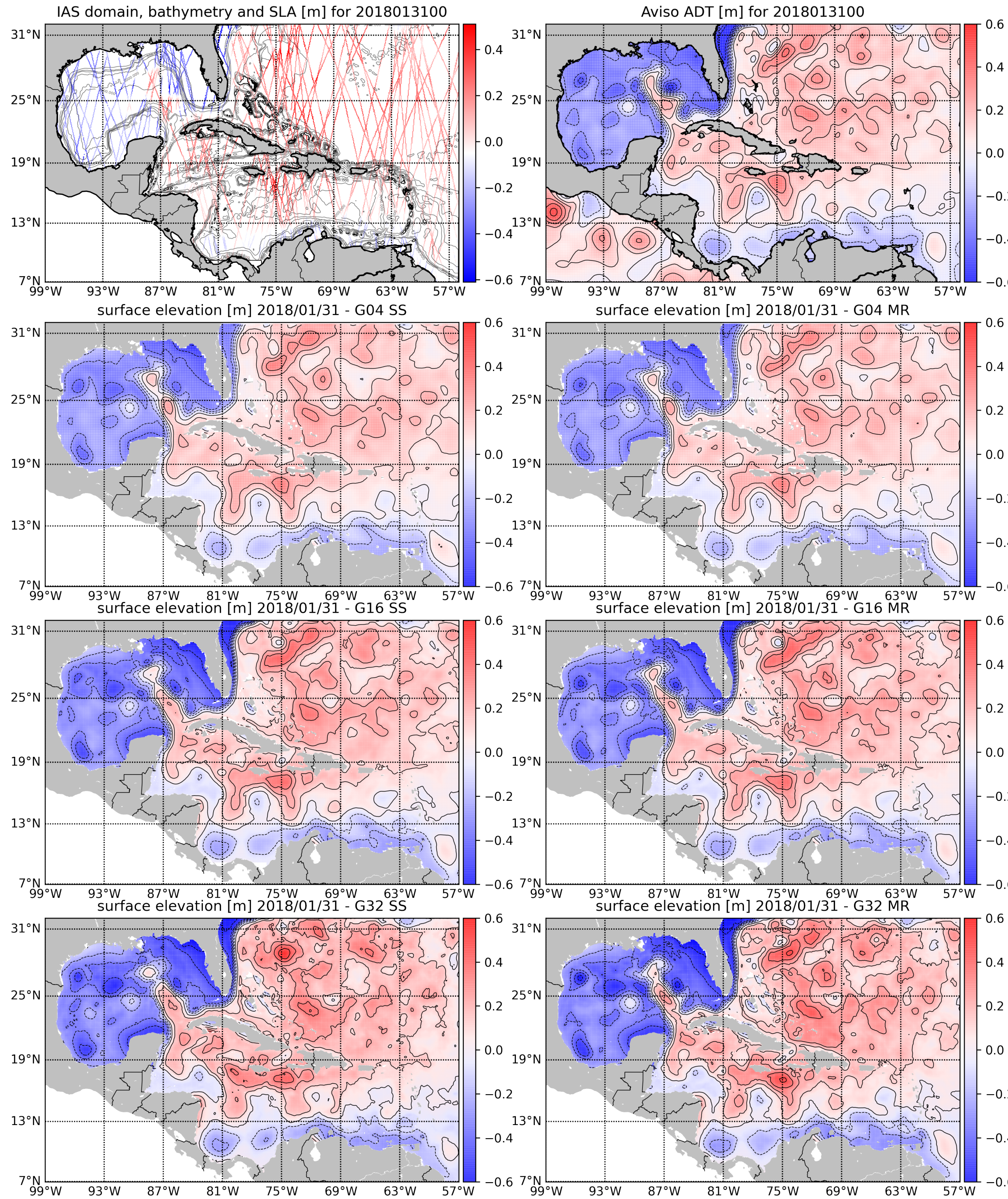
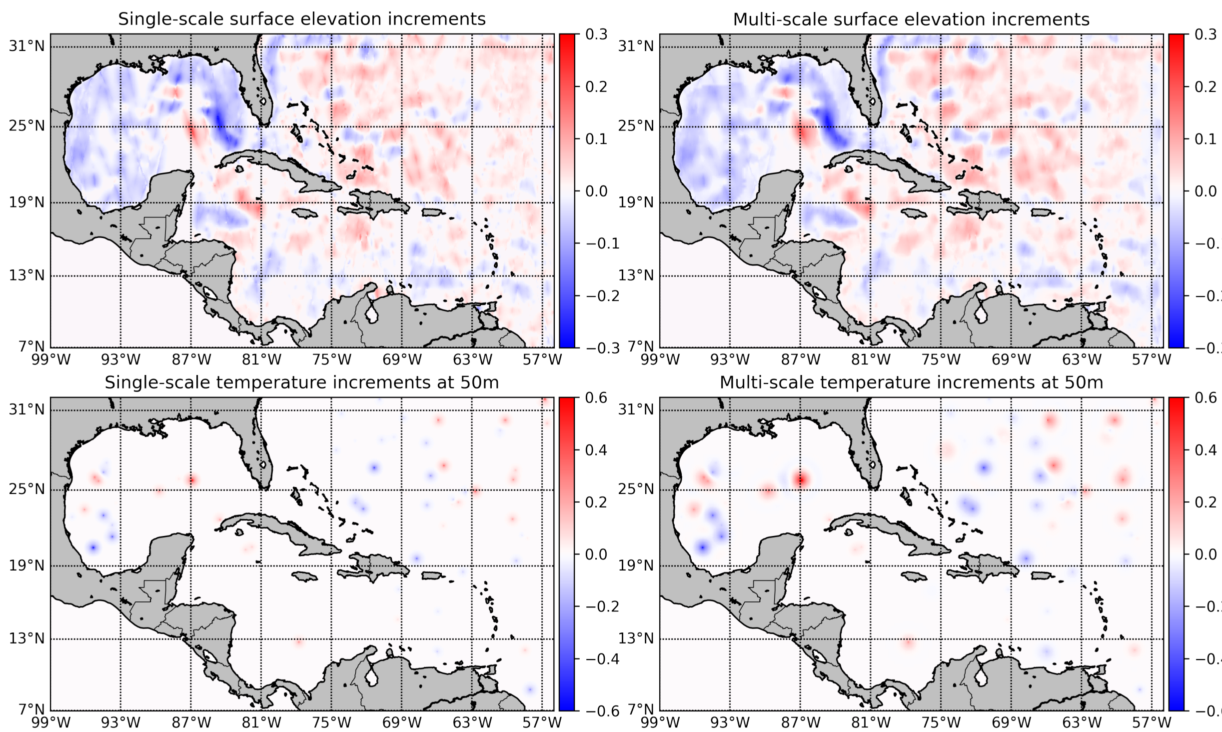
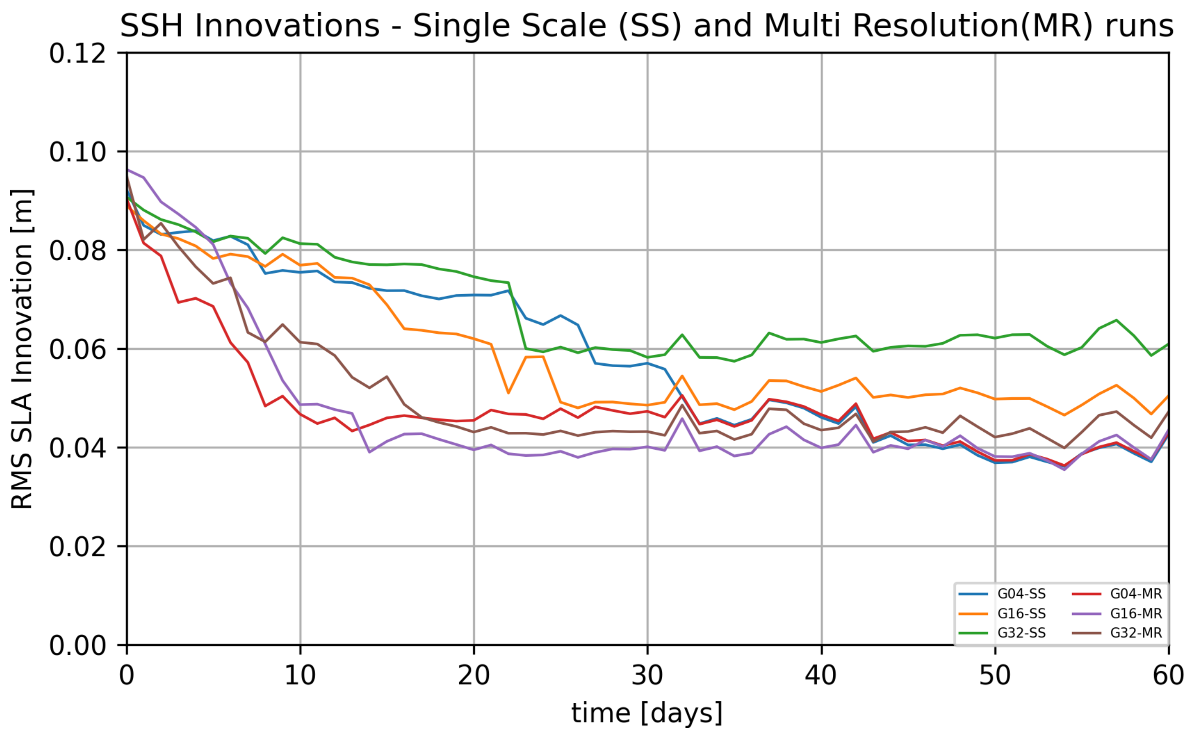


FIG. 4. A sample of the along-track SLA assimilated in the experiments and the SSH (ADT) field for 20180131 used for comparison and verification(upper panel). Snapshots of SSH for the same day from single scale and multi resolution analysis from a $1/4°$, $1/16°$ and $1/32°$ labeled respectively as $G04$, $G16$, and $G32$. The single scale (SS) runs for each model were done with an equally weighted combination of the thin membrane and thin plate models. The multi-resolution (MR) runs are qualitatively consistent in the SSH reconstruction across all resolutions unlike the single scale reconstructions.



766 FIG. 5. Surface height and temperature increments for the $1/16°$ model calculated with single scale and multi
767 resolution analyses for the same initial conditions and observations.



768 FIG. 6. RMS error in SLA from experiments discussed in section 3b and listed in Table 1. In all cases,
769 high resolution along track data over the entire IAS domain were used to calculate the RMS errors. The multi-
770 resolution analysis results in the lowest RMS errors for all models when assimilating high resolution full domain
771 observations

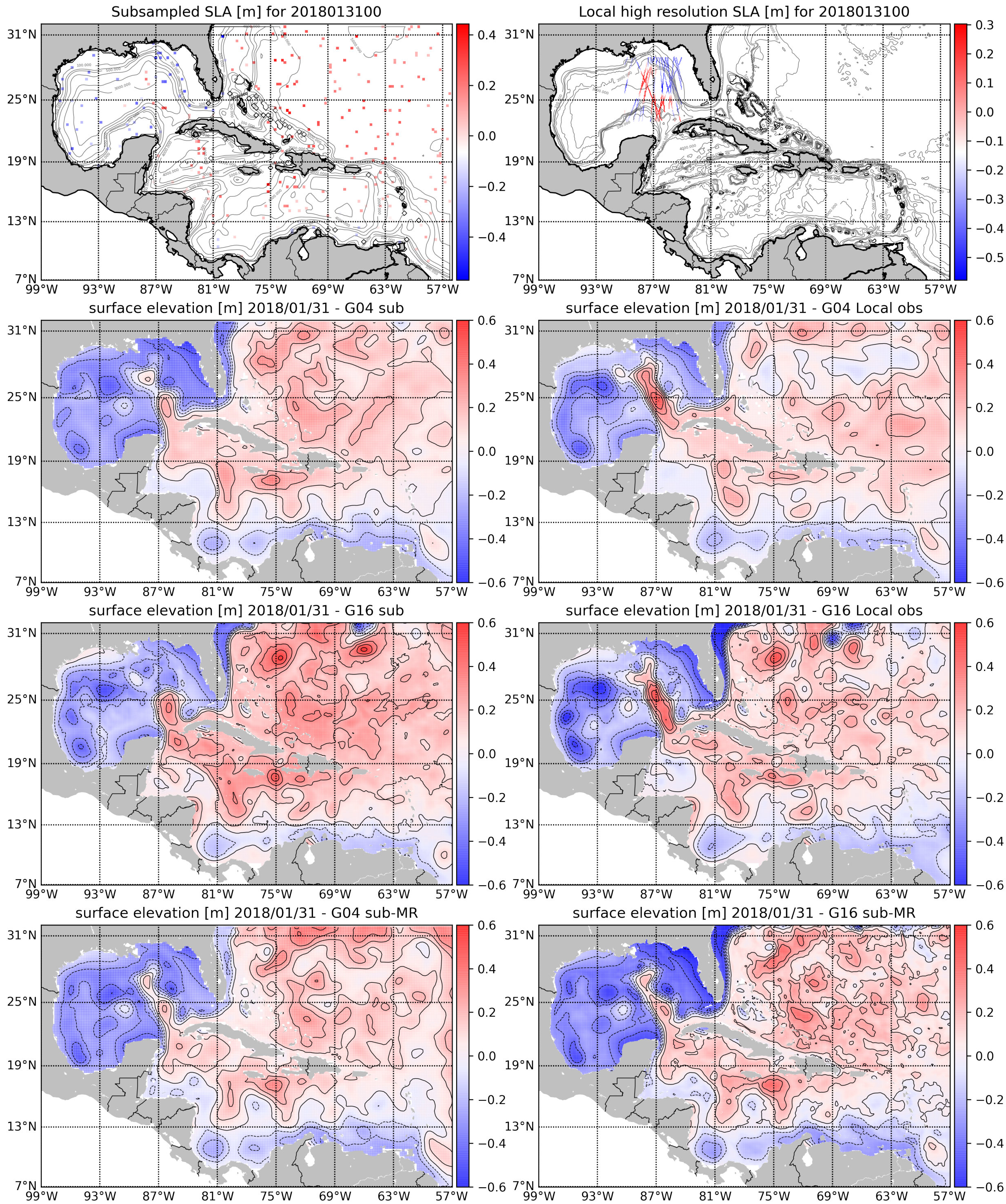


FIG. 7. A sample of the subsampled SLA dataset retaining one out of every 40 along track points and local high resolution observations in the LC vicinity (upper panel). Snapshots from experiements assimilating only subsampled observations and experiments assimilating only high resolution observations (middle panels) and from the multi-resolution analysis (bottom panel). In the latter case, the subsampled observations were assimilated into the 1/4° and high resolution local observations were assimilated into the 1/16° .

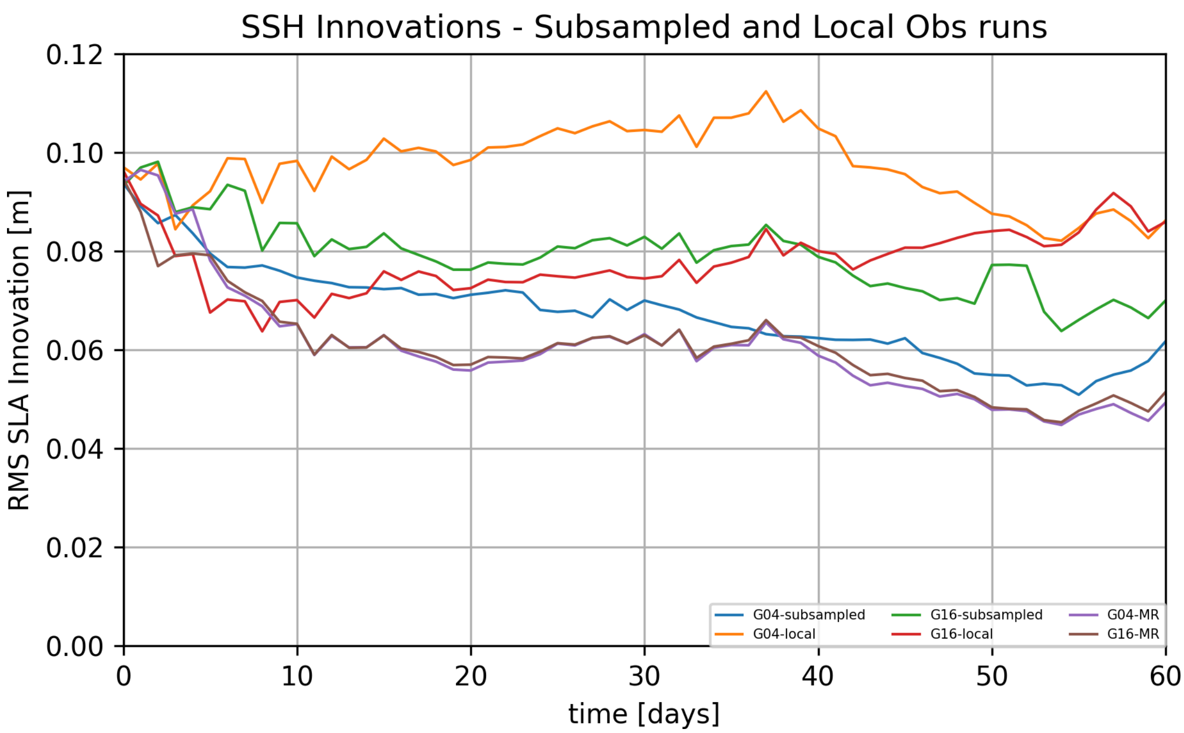


FIG. 8. RMS error in SLA from subsampling and local observations experiments discussed in section 3b and listed in Table 1. In all cases, high resolution along track data over the entire IAS domain were used to calculate the RMS errors. The multi-resolution analysis results in the lowest RMS errors for all models when assimilating subsampled observations in the 1/4° and local observations into the 1/16° model

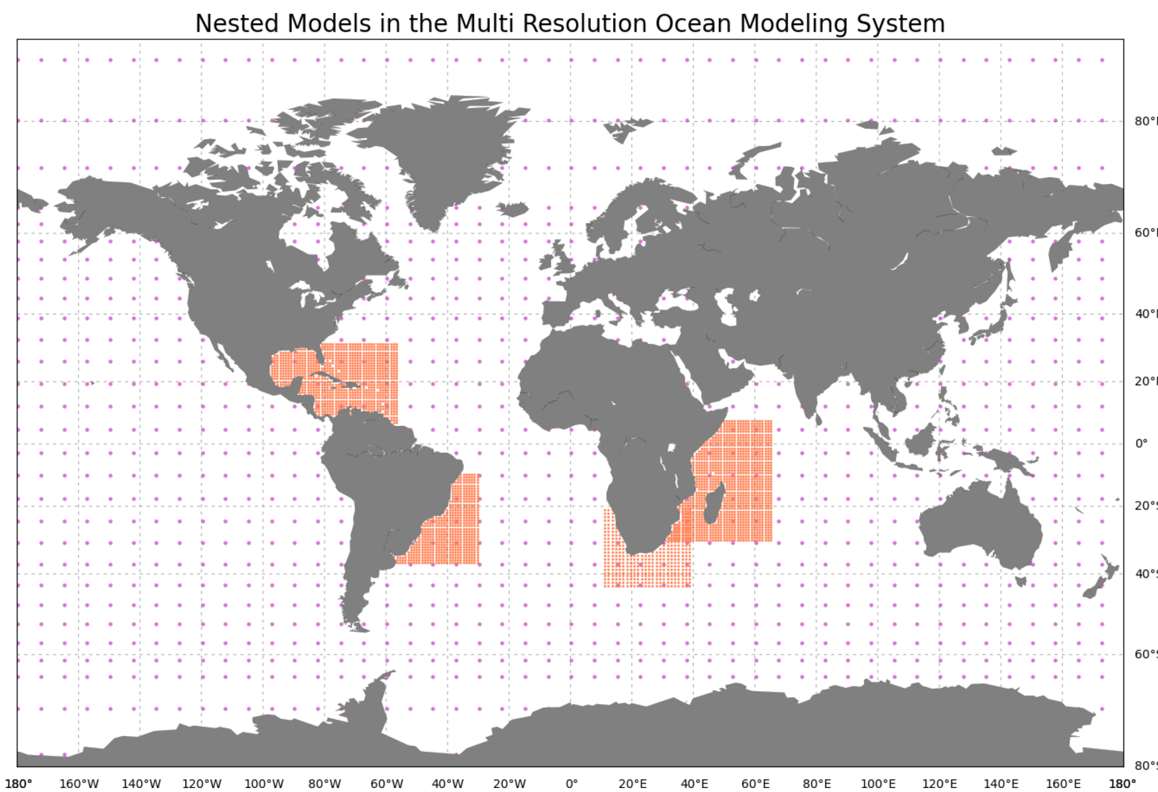


FIG. 9. The models in the multi-resolution system. These include a coarse global (1/4°) grid and four regional scale models of 1/16° for Intra American Seas (IAS), East Africa (EFA), Brazil Region (BRS) and South Africa (TSA)

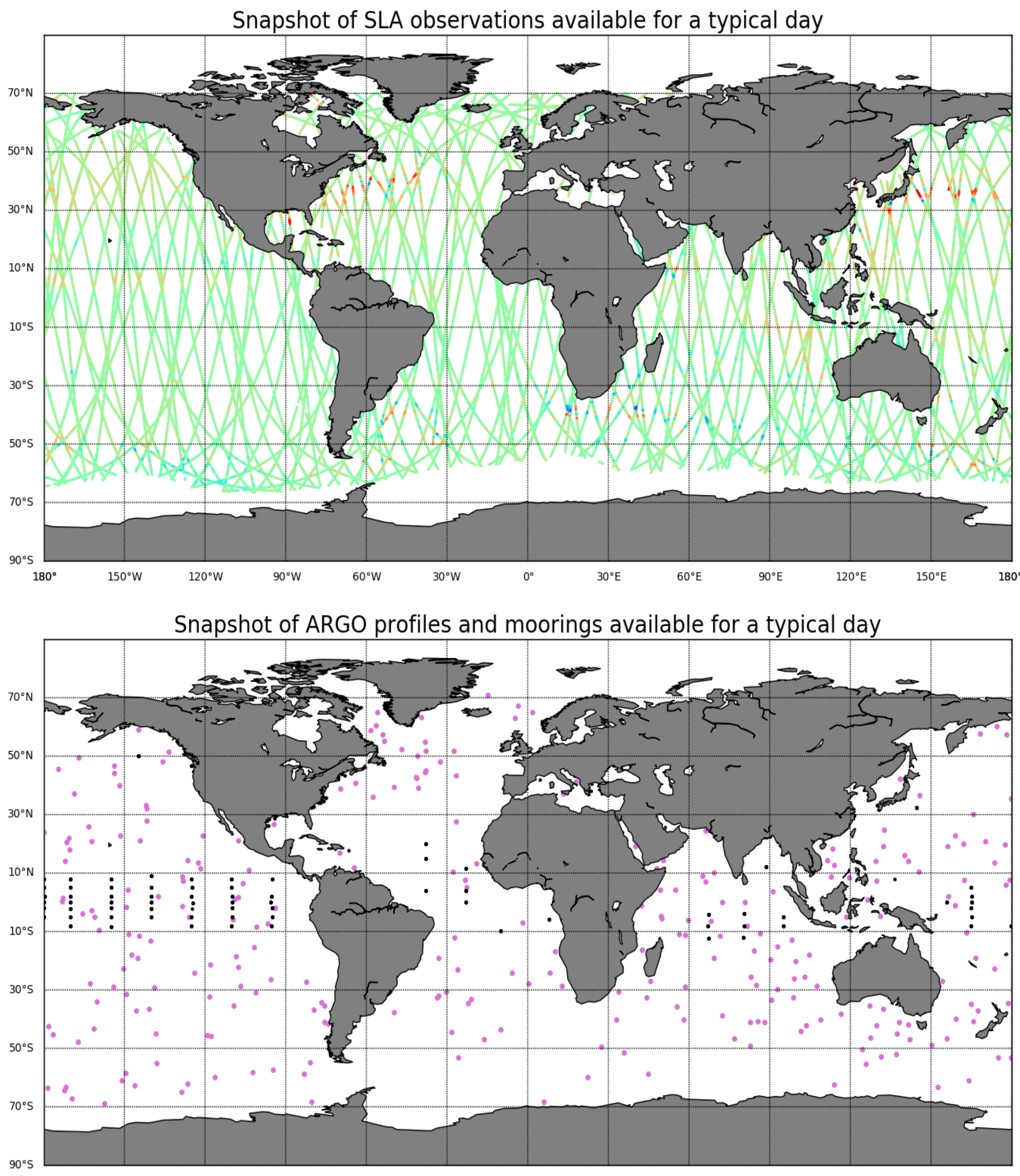
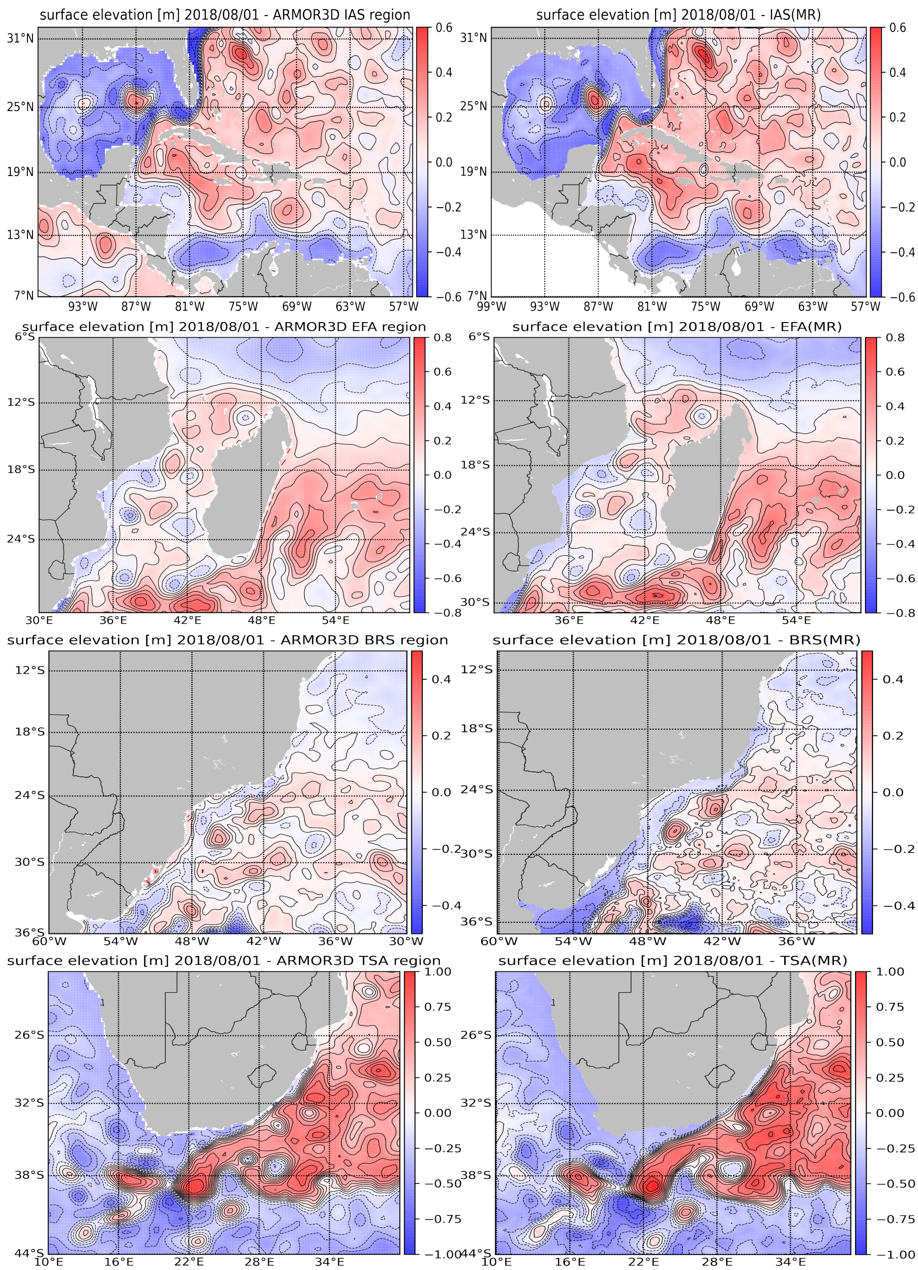
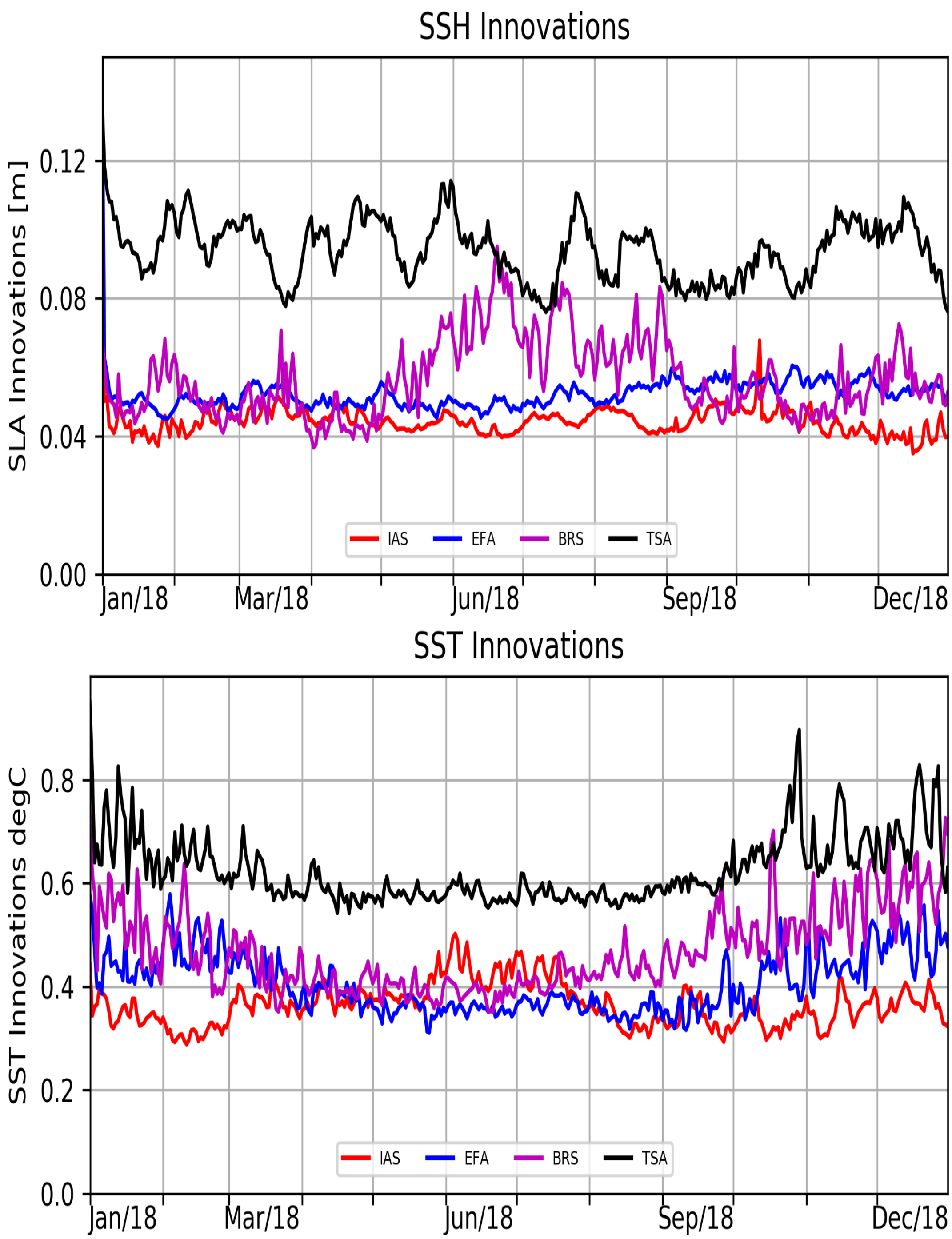


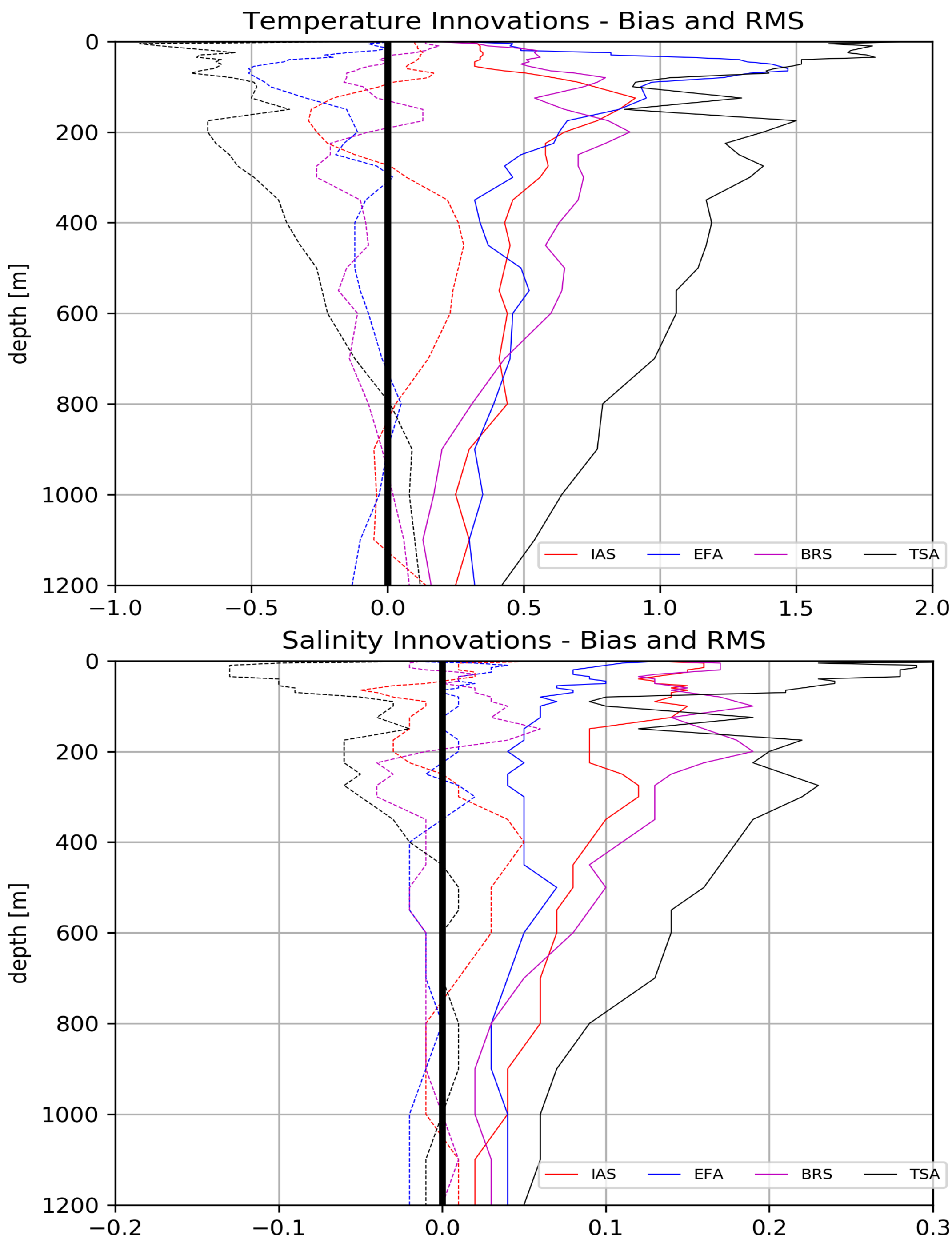
FIG. 10. A sample of the observations assimilated on any given day. Upper panel shows remotely sensed along-track sea level anomalies (SLA) from multiple altimeters. In-situ Temperature/Salinity (T/S) profiles from the ARGO program are shown in the bottom panel in magenta color and mooring locations are shown in black



787 FIG. 11. Snapshots of the surface height fields from ARMOR3D an observation based product for all regions
788 and corresponding surface height fields from the models. As can be seen there is excellent agreement in the
789 location and shape of various mesoscale features in all regions.



790 FIG. 12. Time series of averaged innovations in SLA (top) and SST (bottom). In all cases the innovations are
791 well within the observation standard deviations.



792 FIG. 13. Vertical distribution of averaged innovations in Temperature(top) and Salinity (bottom). The dashed
793 lines are biases and solid lines are RMS innovations

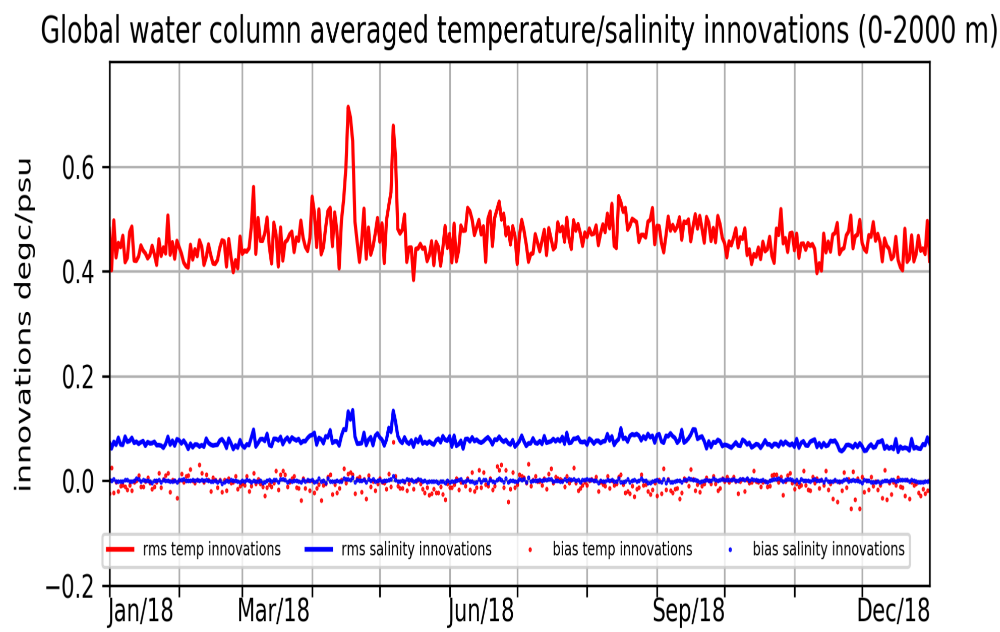


FIG. 14. Globally averaged innovations in temperature and salinity for the top 2000 m. Model forecast and ARGO profiles were interpolated to 42 standard levels (0-2000m) for this comparison

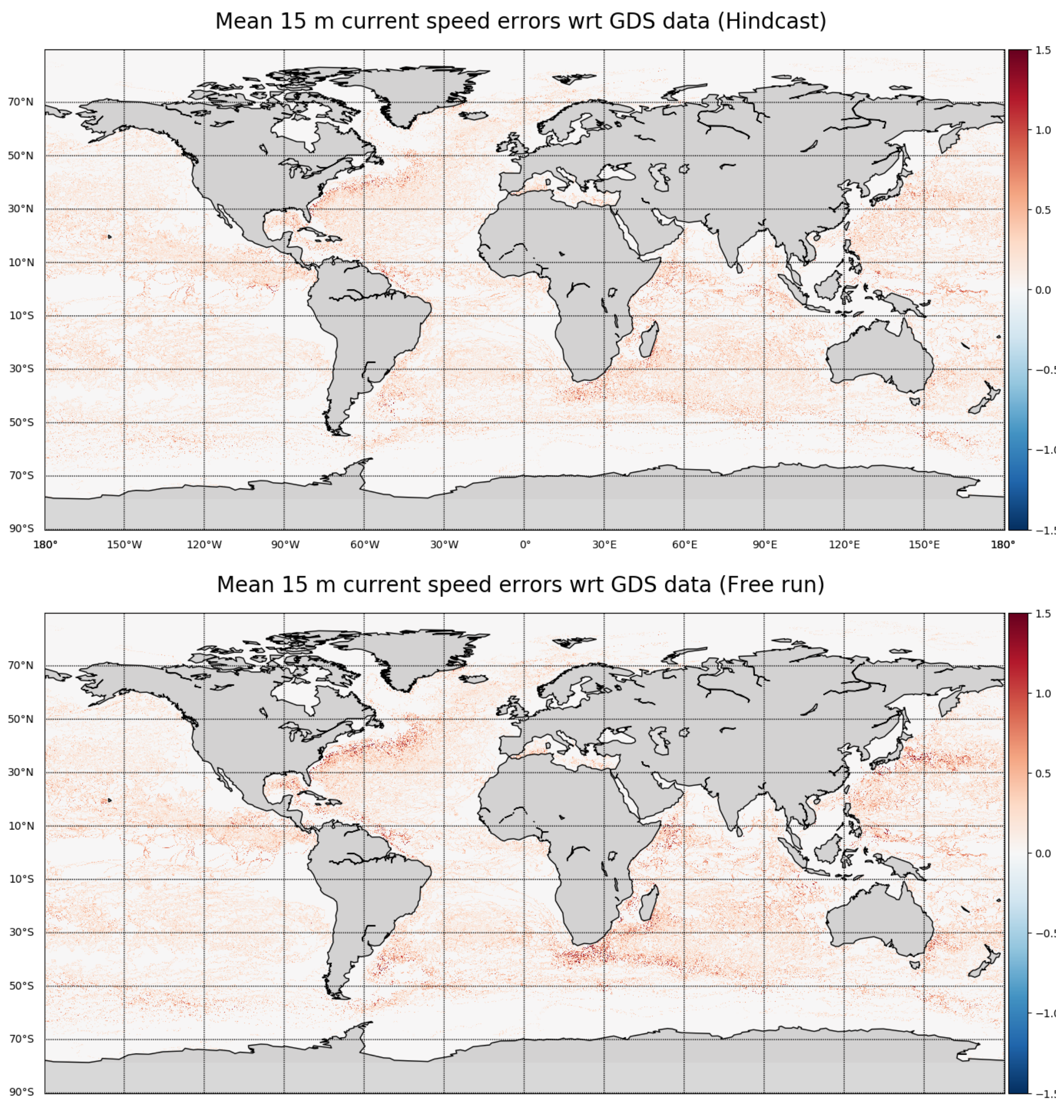


Fig. 15. Spatial map of the errors in current magnitude relative to the GDS drifter dataset. Hindcast is in the top panel and a (control) free run is in the bottom panel.