



Article An Efficient Method for Nested High-Resolution Ocean Modelling Incorporating a Data Assimilation Technique

Georgy I. Shapiro ^{1,*} and Jose M. Gonzalez-Ondina²

- ¹ School of Biological and Marine Sciences, University of Plymouth, Plymouth PL4 8AA, UK
- ² University of Plymouth Enterprise Ltd., Plymouth PL4 8AA, UK; jose.ondina@plymouth.ac.uk
- * Correspondence: gshapiro@plymouth.ac.uk

Abstract: A simple and computationally efficient method is presented for creating a high-resolution regional (child) model nested within a coarse-resolution, good-quality data-assimilating (parent) model. The method, named Nesting with Downscaling and Data Assimilation (NDA), reduces bias and root mean square errors (RMSE) of the child model and does not allow the child model to drift from reality. Usually coarser resolution models, e.g., global scale, are used to provide boundary conditions for the nested child model. The basic idea of the NDA method is to use a complete 3D set of output data from the parent model using a process which is similar to data assimilation of observations into an ocean model. In this way, the child model is physically aware of observations via the parent model. The method allows for avoiding a complex process of assimilating the same observations which were already assimilated into the parent model. The NDA method is illustrated in several simple 2D synthetic cases where the true solution is known. The NDA method reduces the child model bias to the same level as in the parent model and reduces the RMSE, typically by a factor of two to five, occasionally more.

Keywords: ocean modelling; data assimilation; oceanography



Citation: Shapiro, G.I.; Gonzalez-Ondina, J.M. An Efficient Method for Nested High-Resolution Ocean Modelling Incorporating a Data Assimilation Technique. *J. Mar. Sci. Eng.* **2022**, *10*, 432. https://

Academic Editor: Marcello Di Risio

doi.org/10.3390/jmse10030432

Received: 16 February 2022 Accepted: 11 March 2022 Published: 16 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Fine-resolution ocean modelling is becoming a ubiquitous practice to resolve important mesoscale and submesoscale features such as eddies, fronts, boundary currents, and localised upwellings, which play important roles in ocean dynamics (e.g., [1-4]). These localised models can be run by relatively small groups due to the availability of good-quality ocean models such as ROMS or NEMO to the wider oceanographic community [5,6]. Local fine-resolution models require initial and boundary conditions that can be obtained from good-quality but coarser resolution (parent) models run by major ocean modelling centres such as Mercator Ocean International (France) or Met Office (UK) via Copernicus Marine Service [7]. However, if only the boundary data are used for running the nested model (hereafter called the child model), the bulk of the data produced the parent model within the child model domain is not used for improvement of the child model. Therefore, most of the parent outputs which are constantly validated and kept on track by assimilating observations are lost for the child model. Philosophically, it seems promising to use more data from the good-quality parent to improve the child model. The method designed in this study and named Nesting with Downscaling and Data Assimilation (NDA) is focussed on extending the link between the parent and child models.

Because of inevitable approximations in the equations, numerical schemes, parameterisation, and uncertainties in input data, ocean models tend to drift from reality. A process called data assimilation (hereafter DA) is often regarded as a way of keeping a model 'on the tracks' by constantly correcting it with fresh observations [8–10]. Many modern DA methods originated from the methods of objective analysis [9,11]. While the DA problem can be formulated precisely, finding a solution to it is challenging [12–14]. Complex software systems have been developed for operational data assimilation, such as in [15–17].

A critical element of modern variational DA is the estimate of the model and observations error covariance matrices (ECMs). The true state of the ocean is unknown; therefore, ECMs are estimated by some proxy [18]. Various methods use different assumptions and simplifications in order to make the process feasible, e.g., [19,20]. Despite data assimilation being currently ubiquitous in geosciences, it has so far remained a topic mostly reserved to experts [21,22]

The purpose of this paper is to develop a simple and computationally efficient method of implementing DA for a high-resolution child model that is nested into a data-assimilating parent model. The method is mainly intended for smaller academic and operational centres that do not have resources of the same scale as the major ocean forecasting institutions. An example is a regional forecasting group that relies on downloading the outputs from global/basin-scale models available from a reputable ocean forecasting service such as EU CMEMS but also prefers to tune their local high-resolution models.

The basic idea suggested in this study is to assimilate data from a good-quality but coarser resolution parent model instead of using observations. The parent model is assumed to be data-assimilating itself, ensuring that the observations, however indirectly, will not allow the fine-resolution child model to deviate significantly from the true state of the sea. In other words, the coarse model can be regarded as a mean for incorporating a sparse number of observations into a fine-resolution model in a way that honours the laws of physics.

While in principle several existing DA methods could be adapted to assimilate data from the parent into the child model, this study uses the new NDA method. This method combines, from time to time, the outputs from the child model forecast with the forecast from the data-assimilating parent model. This combination creates an improved representation of the ocean state called the analysis, which is used, together with boundary conditions from the parent model, to generate the next high-resolution forecast.

The paper is arranged as follows. The NDA algorithm is described in the Data and Methods section. This section also describes the synthetic idealised case where the true solution is known and is used to illustrate the method. The Results section presents three configurations of synthetic cases for demonstrating the performance of the NDA method when the parent model is 'perfect', i.e., its only error is the representativeness error [23] due to insufficient resolution. The three configurations are a front, a single eddy, and multiple eddies. Each of these examples was tested for different sizes of the field structures and error parameters of the child model. The section also presents results when the parent model is not 'perfect' but produces some spatially correlated and uncorrelated errors. The strengths and limitations of the NDA method are discussed in the final section of the paper in comparison with the commonly used combination of Hollingsworth–Lönnberg and variational DA (VAR) methods [24,25].

2. Materials and Methods

This section presents the algorithm used for Nesting with Downscaling and Data Assimilation and describes a number of simple synthetic illustrative examples which are processed using this method in the subsequent section. The NDA method consists of two steps. The first step uses stochastic–deterministic downscaling [26] of the parent model forecast onto the child model grid. The second step combines, using the same philosophy of variational data assimilation, this downscaled forecast with the forecast generated by the child model.

In general, the algorithm detailed below is applicable when the vectors representing the condition of the ocean are composed of a number of state variables (e.g., temperature, salinity, pollution concentration, etc.) and are discretised in a 3D domain. For example, the vector of true state of the ocean x can take the form $x = \{x_i\} = \{T_1, \ldots, T_N, S_1, \ldots, S_N, C_1, \ldots, C_N\}$ where T_i, S_i, C_i are the temperature, salinity, and concentration of a chemical, and i = 1, ... N are indices for the *N* nodes in the child model domain. In the case when *x* represents multiple variables, the method can take into account cross-correlations between different state variables. When the state vectors represent a 3D field, the covariances along the vertical axis are also taken into account. In a simpler approach, the method can be applied to only one variable at a time and layer-by-layer. The method itself is not restricted by an assumption of isotropy; however, this assumption is used in the illustrative examples. For illustration purposes, the simple synthetic cases are all 2D and are related to one state variable. In these illustrative examples, a layer-by-layer data assimilation is presented as the simplest and most practical variant.

2.1. The NDA Algorithm

The proposed algorithm is designed for assimilating data from an ocean model of a coarser resolution parent model into a finer resolution child model. The algorithm should be used at each time point when the analysis state for the child model is generated as an initial state for the next forecast.

A common approach to data assimilation into an ocean (or atmospheric) model is based on a variational principle, i.e., on minimising the cost function *J* given in Equation (1):

$$J(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^b\right)^T \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{x}^b\right) + \left(\mathbf{y} - H(\mathbf{x})\right)^T \mathbf{R}^{-1} \left(\mathbf{y} - H(\mathbf{x})\right),$$
(1)

where *x* is the (unknown) vector which estimates the true state of the ocean, x^b is the model forecast before data assimilation, *y* is the vector of observed values, *H* is the operator which projects data from the model grid onto the locations of observations, and **B** and **R** are the forecast and observations error covariance matrices, respectively. The superscripts 'T' and '-1' denote the transposed and inverted matrices, respectively (see [19,22,23,25] for more detail).

The value of $x = x^a$ that minimises the cost function *J* is called the 'analysis' and it is closest in an RMS (root mean square) sense to the true state x^t [23]. This approach works well for a relatively small number of data. When the number of observations and model grid nodes is large, the matrices used in Equation (1) become very large and their inversion causes significant computational problems. For example, Bouttier and Courtier [23] state that 'except in analysis problems of very small dimension (like one-dimensional retrievals), it is impossible to compute exactly the least-squares analysis.' Therefore, a number of approximate methods are used to calculate the analysis state, and even this process is computationally expensive.

This study is related to assimilation of large numbers of data (which take the place of observations used in standard DA methods) from one model to another. For this purpose, an alternative cost function J_S is proposed, namely

$$J_{S}(\mathbf{x}') = \left(\mathbf{x}' - \mathbf{x}'^{b}\right)^{T} \mathbf{B}^{-1} \left(\mathbf{x}' - \mathbf{x}'^{b}\right) + \left(S'(\mathbf{y}) - \mathbf{x}'\right)^{T} \mathbf{R}^{-1} \left(S'(\mathbf{y}) - \mathbf{x}'\right),$$
(2)

where the primed variables denote deviations of the respective variable from their statistical mean, which is defined in probability theory as the average of a theoretically infinite number of measurements. It is denoted by the angle brackets in the following equations:

$$x' = x - \langle x \rangle x'^{b} = x^{b} - \langle x^{b} \rangle S'(y) = S(y) - \langle S(y) \rangle$$

Here, y is the vector of data from the parent model (instead of observations) and S is the operator that projects (downscales) the data from the parent model onto the fine grid of the child model.

The best estimate x' for the true value is denoted x'^a and represents the fluctuation of the analysis vector around its statistical mean. It is obtained by minimising the cost function J_S given in Equation (2), as follows:

$$\nabla I_S = 0.$$

Taking into account that matrices **B** and **R** are symmetric so that $\mathbf{B}^T = \mathbf{B}$ and $\mathbf{R}^T = \mathbf{R}$, the following equation is obtained:

$$\nabla J_{S}(\mathbf{x}'^{a}) = (\mathbf{x}'^{a} - \mathbf{x}'^{b})^{T} (\mathbf{B}^{-1} + (\mathbf{B}^{-1})^{T}) - (S'(\mathbf{y}) - \mathbf{x}'^{a})^{T} (\mathbf{R}^{-1} + (\mathbf{R}^{-1})^{T}) = 2(\mathbf{x}'^{a} - \mathbf{x}'^{b})^{T} \mathbf{B}^{-1} - 2(S'(\mathbf{y}) - \mathbf{x}'^{a})^{T} \mathbf{R}^{-1} = 0,$$
(3)

Taking the transpose of Equation (3) and dividing by two gives

$$\mathbf{B}^{-1}\left(\mathbf{x}^{\prime a}-\mathbf{x}^{\prime b}\right)=\mathbf{R}^{-1}\left(S^{\prime}(\mathbf{y})-\mathbf{x}^{\prime a}\right).$$
(4)

Let us introduce the error correlation matrices C_B and C_R for the child and downscaled parent models:

$$\mathbf{B} = \mathbf{V}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}, \quad \mathbf{R} = \mathbf{V}_{\mathrm{R}}\mathbf{C}_{\mathrm{R}}$$

where V_B , V_R are diagonal matrices containing the respective error variances at each fine grid node. The inverted diagonal matrices are also diagonal. Equation (4) then becomes

$$\mathbf{C}_{\mathrm{B}}^{-1}\mathbf{V}_{\mathrm{B}}^{-1}\left({\mathbf{x}'}^{a}-{\mathbf{x}'}^{b}\right)=\mathbf{C}_{\mathrm{R}}^{-1}\mathbf{V}_{\mathrm{R}}^{-1}\left(S'({\mathbf{y}})-{\mathbf{x}'}^{a}\right).$$
(5)

Most assimilation systems assume **R** to be diagonal (no spatial observation error correlations) so that its inverse can be calculated quickly and easily. This approach is obviously not appropriate in the case where the 'observation errors' coming from the parent model have significant spatial correlations and the NDA method does not make this assumption. Correlation matrices for both the child and downscaled parent models relate to the same resolution and the same area of the ocean; hence, it is reasonable to use the same correlation function, e.g., a Gaussian of a certain length scale and therefore the same correlation matrix for both models, i.e., $C_B = C_R = C$. After premultiplying Equation (5) by **C** and rearranging the terms:

$$\left(\mathbf{V}_{\mathrm{B}}^{-1} + \mathbf{V}_{\mathrm{R}}^{-1}\right) {\mathbf{x}'}^{a} = \mathbf{V}_{\mathrm{B}}^{-1} {\mathbf{x}'}^{b} + \mathbf{V}_{\mathrm{R}}^{-1} S'({\mathbf{y}})$$
(6)

or

$$\mathbf{x}^{\prime a} = \left(\mathbf{V}_{\rm B}^{-1} + \mathbf{V}_{\rm R}^{-1}\right)^{-1} (\mathbf{V}_{\rm B}^{-1} \mathbf{x}^{\prime b} + \mathbf{V}_{\rm R}^{-1} S^{\prime}(\mathbf{y})).$$
(7)

Using the commutative properties of diagonal matrices and the following identity

$$\left(\mathbf{V}_{\mathrm{B}}^{-1} + \mathbf{V}_{\mathrm{R}}^{-1}\right)^{-1} = \mathbf{V}_{\mathrm{R}}(\mathbf{V}_{\mathrm{B}} + \mathbf{V}_{\mathrm{R}})^{-1}\mathbf{V}_{\mathrm{B}},\tag{8}$$

the solution for the analysis state given by Equation (7) can be rewritten as

$$\mathbf{x}^{\prime a} = \left(\mathbf{V}_{\rm B}^{-1} + \mathbf{V}_{\rm R}^{-1} \right)^{-1} \left(\mathbf{V}_{\rm B}^{-1} \mathbf{x}^{\prime b} + \mathbf{V}_{\rm R}^{-1} S^{\prime}(\mathbf{y}) \right) = \mathbf{V}_{\rm R} (\mathbf{V}_{\rm B} + \mathbf{V}_{\rm R})^{-1} \mathbf{x}^{\prime b} + \mathbf{V}_{\rm B} (\mathbf{V}_{\rm B} + \mathbf{V}_{\rm R})^{-1} S^{\prime}(\mathbf{y}).$$
(9)

Equation (9) can be interpreted as a zero-dimensional Kalman gain formula applied to the fluctuations of state variables in the parent and child models at each fine grid node independently. The term 'zero-dimensional' reflects the fact that in this case, the matrices

used in the Kalman gain formula have the size 1×1 , i.e., are reduced to a scalar. For a single $x_i^{\prime a}$ element of the state vector x', Equation (9) gives

$$x_i'^a = \frac{V_{Rii}}{V_{Rii} + V_{Bii}} x_i'^b + \frac{V_{Bii}}{V_{Rii} + V_{Bii}} S'(y)_i.$$
 (10)

where $x_i^{\prime b}$ and $S(y)_i^{\prime}$ are the corresponding values of the vectors $x^{\prime b}$ and $S^{\prime}(y)$ at the same grid node *i*.

For the *S* operator, which downscales the fluctuations of the field variable from the coarse to the fine grid, we use the stochastic–deterministic downscaling (SDD) method developed in [26]. The spatial error correlation matrix **C** is not explicitly included in Equations (9) and (10), but it is used at the downscaling stage of data assimilation, hence the error covariances between data at the parent model grid are implicitly present in the operator *S*. This step represents a major difference from standard methodology where the observation errors are usually considered to be uncorrelated (e.g., [23,25]), and therefore all non-diagonal elements in the matrix **C**_R are ignored.

The diagonal values $V_{B,ii}$ and $V_{R,ii}$ represent the variance of forecasting errors in child and parent models, respectively, and in principle they may be computed in any way designed to compute matrix **B** in practical implementations applied for assimilation of observational data, such as the NMC [27] or the Canadian [28]. For consistency, the implementation of the NDA method presented in this paper calculates values $V_{B,ii}$ and $V_{R,ii}$ for the DA step in a similar way as was used for the first step, i.e., SDD downscaling, see [26] for details. The SDD and therefore the implementation of NDA given below requires the knowledge of the ensemble statistical means and variances at each grid node. By the usual assumption of local (i.e., in a small 2D or 3D domain around each grid node) statistical homogeneity, the means $\langle x^b \rangle_i$ and $\langle S(y) \rangle_i$ as well as the variances $V_{B,ii}$ and $V_{R,ii}$ are estimated using the ergodic hypothesis, i.e., by spatial averaging over a small trial domain around the node at the same time point. Generally, the NDA method can be applied for both isotropic and anisotropic situations. In the latter case, the covariances are calculated taking into account directions between the grid nodes in question. The trial domain can be 3D, in which case both horizonal and vertical covariances are taken into account, or 2D so that covariances are calculated layer-by-layer, such as in [29]. In the examples below we use the trial domain in the form of a 2D area on of size $4L \times 4L$, where *L* is the correlation length of $\mathbf{C} = \mathbf{C}_{\mathrm{B}} = \mathbf{C}_{\mathrm{R}}$. As the values $V_{R,ii}$ and $V_{B,ii}$ are computed at each grid node *i* separately, their values vary across the domain. The nodes that are masked out due to coastline or bathymetry, are excluded from calculation of averages and variances within the trial domain.

Equation (10) performs data assimilation on the fluctuations; however, the statistical means for the parent and child model are, in a general case, different. The NDA method should be applied to a good-quality parent model that has been debiased during its own data assimilation cycle and therefore its statistical mean represents the true value as closely as possible. Therefore, it is reasonable to replace the (potentially biased) statistical mean from the child model with the statistical mean of the parent model.

This gives the final equation for calculation of the analysis state:

$$\mathbf{x}^{a} = \mathbf{V}_{\mathrm{R}}(\mathbf{V}_{\mathrm{B}} + \mathbf{V}_{\mathrm{R}})^{-1} \mathbf{x}^{\prime b} + \mathbf{V}_{\mathrm{B}}(\mathbf{V}_{\mathrm{B}} + \mathbf{V}_{\mathrm{R}})^{-1} S^{\prime}(y) + \langle S(y) \rangle.$$
(11)

Or, for an element of the state vector at a particular fine grid node i (i = 1...N), where N is the number of grid nodes in the child model:

$$x_{i}^{a} = \frac{V_{Rii}}{V_{Rii} + V_{Bii}} x^{\prime b}{}_{i} + \frac{V_{Bii}}{V_{Rii} + V_{Bii}} S^{\prime}(y)_{i} + \langle S(y) \rangle_{i}.$$
 (12)

To summarise, the NDA model-to-model data assimilation procedure includes two steps. First, the parent model output is downscaled from the coarse grid onto the child model fine grid using the SDD method. The result is that at each fine grid node there are two values of the same state variable. The best estimate of the true field is obtained at the second step by combining these two values using a zero-dimensional Kalman filter. The workflow diagram of the algorithm is shown in Figure 1.



Figure 1. Workflow diagram for the NDA method showing one data assimilation cycle.

This algorithm is computationally efficient, as it does not require inversion of large matrices or solving a very large system of algebraic equations at the second step, which is required if using the variational methods based on Equation (1). The inversion of correlation matrices to obtain the weight coefficients for the SDD step can be done only once at the beginning of the model run, as these coefficients do not depend on time. Another benefit of the described method is that the correlation matrices for the downscaling of the parent model have a relatively small rank and condition number, and their inverse counterparts can be calculated without the need for any type of matrix regularisation, just using double precision for the computer representation of the variables. This is because the SDD method

7 of 25

assumes local homogeneity of statistical properties of the field variable. More details on the philosophy and the algorithm of the SDD method can be found in Shapiro et al. [26].

2.2. Data for the Synthetic Idealised Case

In this subsection, the NDA algorithm is illustrated in a synthetic idealised case where the true solution is known. The task is to generate an analysis state using the fine-resolution (child) model forecast and the output from a good-quality data-assimilating parent coarse model, which is used instead of actual observations. It should be noted that, even if both coarse and fine models were perfect, there would be some unavoidable discrepancies, or representativeness errors, between the two models due to the different meshes they use [23]. One of the problems in data assimilation is that the true solution is unknown and therefore the error covariance matrices required to minimise the cost function must be estimated based on various assumptions, e.g., see the review paper by Bannister [18]. Comparison with observations has its own drawbacks as observations are not perfect. The testing of a DA method in a simplified synthetic setting when the true solution is known was used in previous studies and showed its advantages, e.g., [22].

In this study, the parent model output is simulated by sampling the true solution (a prescribed function) on the parent model coarse grid, to which either correlated or uncorrelated error might be added. The fine-resolution forecast is simulated by sampling the same true solution on the fine grid, adding random noise, bias, and some spatial shift. Whilst generally the NDA method calculates the bias locally so it can vary across the domain (Equation (12)). In the following examples, the bias was applied uniformly to keep the illustrative example simple. The spatial shift, a type of spatially correlated error, is added to simulate the 'double penalty effect', which is common to fine-resolution models, e.g., [29].

2.3. Case Studies

Three distinct examples are considered, all relate to data assimilation of one field variable at a single computational surface, whether it be a geopotential ('horizontal') level, such as used in Bell et al. [10], or a curved level such as used in sigma [30] or multienvelope vertical coordinate system [31]. It is common practice to perform the data assimilation level-by-level using a 2D approach at each level, where matrix **B** only includes spatial correlations along the level surface [32]. Occasionally, vertical correlations are used in a preprocessing step for the DA, where observations are interpolated into the model levels [10]. Whilst the NDA method can be applied in 3D, here it is applied to 2D fields for simplicity.

The properties of the NDA method are analysed for (i) an ocean front, (ii) an isolated isotropic eddy, and (iii) a system of densely packed anisotropic mesoscale eddies. In all examples the parent model has resolution of $\Delta x_p = \Delta y_p = 10$ km, and the child model has resolution of $\Delta x_c = \Delta y_c = 2.5$ km and a correlation length of L = 17 km. The trial domain used to calculate local spatial averages and variances of fluctuations was a square of $4L \times 4L$ km² centred at each node. The correlation Gaussian function has a value of less than 2% beyond the distance of 2L and its contribution can be safely neglected for all the examples below. These dimensions were chosen to simulate typical ocean structures. However, it must be pointed out that classification of ocean features as mesoscale or submesoscale are relative to the local Rossby radius. For example, the same eddy of a radius of 12 km could be classed as mesoscale if the Rossby radius is, for example, 10 km, as in some shelf or high latitude seas, and as submesoscale if the Rossby radius is, for example, 40 km, as in the middle of open ocean. The NDA method itself is not specific to whether the features are mesoscale, large-scale, or submesoscale; however, the quality of the first step (downscaling) depends on the ratios between Δx_v , Δy_v , Δx_c , and Δy_c , and the Rossby radius as discussed in [26].

2.3.1. Ocean Front

An ocean front is a narrow area separating two water masses, and it has significant impact on horizontal and vertical exchanges, e.g., [33]. This example is set in a square domain of $200 \times 200 \text{ km}^2$, and the front extends in the meridional (*y*) direction at the centre of the domain.

The true solution is set to be in the form

$$F = A \tanh\left(\frac{x}{L_x}\right),\tag{13}$$

where *A* is the amplitude of the front and L_x is its half-width. For this exercise, A = 1 and L_x ranges between 6 and 40 km. The coordinates *x* and *y* are referenced to the centre of the domain in the zonal and meridional directions, respectively.

2.3.2. Isolated Mesoscale Eddy

Mesoscale eddies are a ubiquitous feature of oceans worldwide. Originally, they were thought to exist only next to jet currents such as the Gulf Stream; however, since the 1960s–1970s, it became clear that mesoscale eddies exist nearly anywhere in the ocean. Most of the kinetic energy of the ocean is contained in mesoscale eddies [34]. In this example, an isotropic eddy is placed in the centre of a domain of 200×200 km² and the true solution is set to be in the form

$$F = A \exp\left\{-\left\lfloor \left(\frac{x}{L_e}\right)^2 + \left(\frac{y}{L_e}\right)^2\right\rfloor\right\},\tag{14}$$

where *A* is the amplitude of the eddy and L_e is its radius. For this exercise A = 1 and L_e ranges between 6 and 46 km.

2.3.3. Multiple Mesoscale Eddies

For this example, let us consider a square domain $1000 \times 1000 \text{ km}^2$ and take the true solution for a variable *F* in the form of multiple anisotropic mesoscale eddies

$$F = \sin\left(\frac{\pi x}{L_x}\right) \sin\left(\frac{\pi y}{L_y}\right),\tag{15}$$

where L_x is the eddy half-size in the *x*-direction, and $L_y = 105$ km is the eddy half-size in the *y*-direction. Let us consider the range of eddy sizes L_x between 12 and 24 km. At $L_x = 24$ km the parent model can be classed as eddy-permitting, as it has 2.4 grid points over the smaller 'diameter' of the eddy. At $L_x = 12$ km, the parent model is not even eddy-permitting but only showing an embryonic representation of the eddy. The child fine-resolution model is eddy resolving for any L_x within the chosen range of eddy sizes.

3. Results

The NDA data assimilation method detailed in the subsection algorithm above was applied to the simulated child model forecast in order to create analysis which is used as an initial condition for the next forecasting cycle. The results for three examples are presented in this section when the parent model is error-free: an ocean front, an isolated single eddy, and a set of multiple eddies. An additional example is considered when the parent model has some errors (either correlated or uncorrelated, extra to the representativeness errors) that are not eliminated by its own DA procedure.

3.1. Ocean Front

This example uses Equation (13) for the representation of an ocean front in the meridional direction. Figure 2 shows a map of the true solution *F* for a sharp ocean front at $L_x = 6$ km (Figure 2a), its representation by the parent model (Figure 2b), by the child model forecast before data assimilation (Figure 2c), and as transects across the front. The



following parameters of added errors are used to simulate the child model forecast before data assimilation: normally distributed noise with standard deviation 0.15, positive bias of 0.3, and shift of the field by 4 km to the west.

Figure 2. Ocean front of width $L_x = 6$ km. (a) True field on the fine mesh, (b) parent model ouput, (c) simulated child model before data assimilation, (d) transects of the previous fields along the line at y = 100 km.

Even at a resolution of 10 km that does not resolve the structure of the front (half-width of 6 km), the parent model gives a reasonable representation of areas outside the front where the changes in the state variable are smooth (Figure 2b). However, as expected, the width of the front is exaggerated due to insufficient resolution. The child model forecast shown in Figure 2c is noisy and clearly shifted westward relative to the true state. Figure 3 shows the results of assimilation of output from the parent model using the NDA method. The analysis state (after assimilation) removes the spatial shift and bias and reduces the noise even for such a sharp front where the resolution of the parent model is inadequate. For comparison, Figure 4 shows the improvement provided by the NDA assimilation for fronts of different sharpness, with $L_x = 10$, 20, 30, and 40 km. In all examples, the errors in non-assimilating child model forecasts are simulated by adding bias = 0.3, random normally distributed noise with standard deviation STD = 0.15, and spatial shift of 4 km to the west.



Figure 3. Assimilation results for the ocean front shown in Figure 2: (a) child model after NDA assimilation, (b) difference between the assimilated model and the true field, (c) difference between the non-assimilated and assimilated child models, (d) zonal transects of the previous fields along a line at y = 100 km.









Figure 4. Results of NDA assimilation for fronts of different sharpness: (**a**) $L_x = 10$ km, (**b**) km, (**c**) $L_x = 30$ km, (**d**) $L_x = 40$ km. The curves present data for the true field (black), non-assimilated noisy child model (red) and child model after NDA assimilation (blue).

Comparison of panels in Figures 4a–d and 3d shows that the improvement due to data assimilation from the parent model is achieved both for sharp fronts not resolved by the parent model and for smooth fronts.

Figure 5 shows how the bias and RMSE of the child model against the true state change with the sharpness of the front before and after data assimilation.



Figure 5. Plots of bias (**a**) and RMSE (**b**) for the ocean front case with different front sharpness for non-assimilated (black) and assimilated (blue) child models. The red line shows the ratio of assimilated to non-assimilated RMSE.

After applying the NDA method, the bias is practically removed and the remaining values are on the order of 5×10^{-3} or less. The RMSE calculated against the true state is more than five times lower than before data assimilation, as shown by the red line in Figure 5b.

Similar properties are demonstrated in the example of a single eddy.

3.2. Single Eddy

Figure 6 shows a map of the true solution *F* for an axisymmetric mesoscale eddy with radius $L_e = 16$ km, its representation by the parent coarse model and by the child model before data assimilation. It also shows transects across the eddy centre. The parameters of added errors are the same as in the previous example: normally distributed noise with a standard deviation of 0.15, positive bias of 0.3, and shift of the field by 4 km to the west.



Figure 6. Results of NDA assimilation for a single isolated eddy: (**a**) true field sampled on the fine grid, (**b**) child model before data assimilation, (**c**) child model after NDA data assimilation, (**d**) transects showing the same fields as in (**a**–**c**) in comparison with the parent model.

The NDA method reduces errors nearly to zero outside of the eddy and greatly reduces them inside the eddy (Figure 6b,d). The bias is reduced from 0.3 before data assimilation to 0.002 or less after assimilation. The reduction in RMSE for various values of eddy radius for the range 6 to 46 km is shown in Figure 7.



Figure 7. Dependence of RMSE for different eddy sizes: before data assimilation (black), after NDA assimilation (blue), the ratio of RMSE after and before data assimilation (red).

For illustration purposes, the spatial distributions of the variances $V_{B,ii}$ and $V_{R,ii}$ for the case of a single eddy with $L_x = 16$ km are presented in Figure 8. The following parameters are used for the child model: STD of noise = 0.15, bias = 0.3, and spatial shift = 4 km. The non-uniform colour in Figure 8b is a representation of noise in the child model output before data assimilation.



Figure 8. Spatial distribution of the diagonal elements $V_{B,ii}$ (**a**) and $V_{R,ii}$ (**b**) in the example of a single eddy of a radius of 16 km.

The next example illustrates the qualities of the NDA method for a domain $1000 \times 1000 \text{ km}^2$ packed with anisotropic eddies.

3.3. Multiple Eddies

Figure 9 shows maps of the field variable F from the parent and child models with the following parameters of added errors to simulate child model forecast before data assimilation: normally distributed noise with a standard deviation of 0.15, positive bias of 0.3, and spatial shift of the field by 4 km to the west.







Figure 9. Maps of the field variable before data assimilation for multiple eddy case with $L_x = 12$ km: (a) true field on fine grid, (b) true field on parent grid, (c) non-assimilated model, (d) difference between non-assimilated model and true field.

Figure 9b shows that the parent model generally underestimates the true field shown in Figure 9a due to representativity errors. The difference between the non-assimilated child model forecast and true state shown in Figure 9d is substantial, showing an RMSE = 0.61, which is about 30% of the range of the true field. The anatomy of the differences between the true state, and outputs from parent (coarse), and non-assimilated child models on a zoomed-in section of the zonal transect are shown in Figure 10. The largest errors produced by the non-assimilated model are due to spatial shift and bias. The errors produced by the parent model are exclusively due its insufficient resolution (i.e., are representativeness errors), as we assume that otherwise the parent model is perfect.



Figure 10. Transects of true field on the fine grid (black) and outputs from the parent (dashed brown) and non-assimilated child (red) models.

The NDA method presented in the previous section is then applied to assimilate the data from the parent model into child model forecast.

Figure 11 shows the results of data assimilation (analysis) and a map of differences between the assimilated child model and true field.



Figure 11. Maps for the multiple eddies case with width $L_x = 12$ km: (a) NDA assimilated model, (b) difference between NDA assimilated model and true field. Parameters of the non-assimilated model are the same as in Figures 9 and 10.

The details of the improvement achieved by NDA data assimilation on a zoomed-in zonal transect are shown in Figure 12.



Figure 12. Transects at y = 40 km for the multiple eddies case with $L_x = 12$ km. (a) Transects of NDA model output (blue), true solution (black), parent model (dashed dark red), and the non-assimilated child model (red); (b) differences between assimilated (blue) and non-assimilated (red) models and the true solution.

Data assimilation partially reduces the spatial shift which simulated the double penalty effect common to fine-resolution models. The reduction in spatial shift is due to the properties of the SDD component of the NDA method, see [26] for details. The significant reduction in the bias is due to the assimilation part of the NDA algorithm; however, it also causes some reduction in the amplitudes of the eddies (Figure 12a). The remaining errors shown in Figure 12b are mainly due to incompletely corrected spatial shift and random noise in the assimilated model. The RMSE of the assimilated model is lower at 0.25 vs. a non-assimilated value of 0.61, and the bias is reduced by orders of magnitude from 0.3 to -1.3 10^{-5} .

Figure 13 shows the results of the sensitivity analysis with different sizes of eddies and levels of various sources of errors in the non-assimilated child model forecast noise, bias, and shift.



Ratio of RSME values for combinations of noise-bias-shift

Figure 13. Ratio of RSME values of the child model relative to the true solution after and before NDA data assimilation as a function of eddy radius in the zonal direction *x*. Ratios are calculated for different combinations of errors generated by the child model: random noise (STD equal to either zero or 0.15), spatial shift (zero or 4 km), and bias (zero or 0.3). Percentages are shown with respect to the amplitude of the true signal. The first number in the legend shows the level of noise, the second shows the bias, and the third one shows the spatial shift.

In this example, the NDA method works best when the only source of error is the bias, which is removed nearly completely. The second-best results are achieved when only random noise and bias are present but not the spatial shift. The RMSE ratio for cases containing errors due to spatial shift grows slightly at larger eddy sizes. This is because a relatively small shift of 4 km does not distort large eddies to the same extent as small eddies even in the non-assimilated model. Generally, the curves in Figure 13 show that after application of NDA data assimilation algorithm the RMSE is reduced by half or better. Bias between the child and parent models is reduced by orders of magnitude. Assuming that the parent data-assimilating model is unbiased, it means that the fine resolution model becomes unbiased after application of NDA data assimilation process.

3.4. Effect of Errors in the Parent Model

The next example shows the quality of the NDA method when the parent model still contains errors, even after its own data assimilation cycle. To investigate how these errors affect the reduction in RMSE in the child model when the NDA data assimilation is applied, we tested two settings. In the first example, a quite large random noise was added to the parent model output with a standard deviation of 0.1, i.e., 10% of the signal. The results for the ocean front and multiple eddies examples are shown in Figures 14 and 15.

The transect in Figure 14 shows that after NDA, the front is well represented with small random noise even when the parent model is quite noisy. The improvement is up to five-fold and is consistent across the range of front widths from 6 to 46 km. The second example for the ocean front illustrates the efficiency of the NDA method when the parent model has spatially correlated errors, namely a lateral shift of 1 km to the true parent model (no random noise in this case). The simulated errors to the child model are the same as specified in the Figure 14 caption. The improvements in the child model outputs after DA follow: bias is reduced from 0.3365 to 0.0118 (approximately by a factor of 28) and RMSE is reduced from 0.3778 to 0.0616 (approximately 6-fold).



Figure 14. Ocean front case of $L_x = 14$ km, with noisy parent model (added noise STD = 0.1) and noisy non-assimilated child model (added noise STD = 0.15, bias = 0.3 and shift of 4 km to the west). (a) Transects at y = 100 km of assimilated (blue), non-assimilated (red) child models, and the true field (black), (b) RMSE calculated for non-assimilated (black) and assimilated (blue) child models and the ratio between them at different front steepness.



Figure 15. Multiple eddy case of $L_x = 14$ km with noisy parent model (added noise STD = 0.1) and noisy non-assimilated child model (added noise STD = 0.15, bias = 0.3, and shift of 4 km to the west). (a) Transects at y = 100 km of true field (black), parent (dashed brown), assimilated (blue), and non-assimilated (red) child models, (b) RMSE for different eddy sizes and the ratio between them.

In case of multiple eddies, the improvement is less dramatic; however, for eddy diameters larger than 12 km, the RMSE is improved approximately by half (Figure 14). For eddy sizes 10 km and less the improvement due to assimilation of the parent model is limited, which is expected as such eddies are strongly distorted by the parent model of only 10 km resolution.

3.5. Spectral Characteristics

The mechanism of improvement of fine model outputs by the NDA method can be seen from the analysis of spectral characteristics of the downscaled parent model, fine model forecast (before data assimilation) and analysis (after data assimilation) in Figure 16. The bias represented by the peak at zero wavenumber is removed at the data assimilation step

of NDA. Noise in the range below Nyquist wavenumber is reduced by melding the noisy child model forecast with the clean parent model data. It should be noted that the SDD downscaling honours the data of the parent model on the child nodes that coincide with a coarse grid node. A small peak between the coarse and fine grid Nyquist wavenumbers (outside of the scale in this figure) is an artefact created at the SDD step [26], as seen from the spectrum of the downscaled field. Other than this peak, the downscaled field has much lower level of noise than the child model forecast. Reduction in noise at wavenumbers higher than the parent model Nyquist frequency is due to the low level of noise generated at the downscaling step of the process. Figure 16 shows an efficient removal of bias in the Fourier space. The amplitude of the signal peak is also slightly reduced after DA.



Figure 16. Amplitude spectra on a zonal transect at y = 40 km for the example of multiple eddies of $L_x = 16$ km. The STD of noise in the non-assimilated child model is 0.15, bias is 0.3, and the spatial shift is 4 km to the west. The spectra are calculated for the true field, the child model forecast (before assimilation), and analysis (after assimilation). (a) Spectra in the full range of wave numbers and (b) zoomed in area for low wave numbers.

In real applications, there could be a potential mismatch between the bathymetry and coastlines in the fine and coarse models. For instance, some details such as small islands might be absent in the coarse mesh but present in the fine model, causing a mismatch between the land masks in fine and coarse models. This issue relates only to the first step of the NDA method, which performs the downscaling of the coarse model data onto the fine model grid, and this mismatch is natively resolved during the SDD downscaling [26]. As shown there, in the example of the coastal areas of the Red Sea, better resolution of the fine model results in higher vorticity/enstrophy values in the downscaled field compared to the parent coarse model, as shown in Figures 12–16 of that paper. The potential mismatch of the land mask is not specific to the NDA method; the same issue may appear when assimilation is carried out using satellite imagery (i.e., observations) of different resolution than the fine model. After downscaling, at the DA stage the data from the parent and child model are available exactly on the same (fine) grid with the same land mask, and the grid mismatch issue does not appear.

3.6. Comparison with a 'Standard' Method

The quality of the NDA method can be illustrated by comparison with existing data assimilation techniques applied to model-to-model data assimilation. Let us consider the combination of H–L [24] and variational method, which is considered to be the most commonly used technique for the estimation of the covariance matrices (e.g., [22,35]). There are several variants of this method; here, we applied the practical algorithm described in the textbook by Kalnay [25], which is named hereafter as the 'standard' method. The

'standard' method consists of the following stages: calculate innovations (differences between observations and model) at each observational point, estimate the covariances between innovations at different locations, fit the best-fit curve (usually Gaussian) using all covariances except at zero distance, and estimate the model and observation error variances using the value where the best-fit curve intersects the r = 0 vertical line on the covariance plot. In the practical implementations of this algorithm, the model and observation error variances are assumed spatially homogenous [25]. The model variances and the best-fit curve are used to calculate the background error covariance **B**-matrix, and the diagonal observational error covariance R-matrix. Because of spatial homogeneity, all diagonal elements in **B** and **R** matrices are equal but different between matrices. The H–L method requires combination of innovations into spatial bins, so that all innovations within a certain bin are allocated to the same distance from the central point. In practice, the covariances are calculated by averaging individual products of innovations over a period of time instead of statistical averaging assuming the ergodic hypothesis [36]. In the model-to-model DA approach, the role of observations is played by the output from the parent data-assimilating model.

The cost functions for the 'standard', Equation (1), and the NDA, Equation (2) methods are different; in other words, both methods minimize their own cost functions. The procedures for calculation of **B** and **R** matrices in the NDA and H–L methods are different and therefore they produce different matrices. For NDA, both matrices represent the covariances of the background error of the downscaled parent and child models, respectively, and are calculated in a similar and consistent way as both data sets are from models. Therefore, whilst both matrices have different error variances (diagonal elements), they have the same spatial correlations. Conversely, in the 'standard' method, the matrices are calculated using the H–L method that assumes that the observation errors have no spatial correlation. This assumption can be valid for some types of observations, but it is not valid for model-to-model DA.

For comparison between the NDA and 'standard' methods, we selected the multiple eddies example (as in Results, Section 3.3). The covariances required to build the **B** and **R** matrices for the standard method were estimated using H–L as follows. We generated 150 random realisations of the child model output, where each realisation was the sum of the true field given by Equation (15) ($L_x = 40$ km and $L_y = 105$ km) and the following errors: (i) random noise with standard deviation of 0.15, (ii) random spatial shift with a standard deviation of 4 km, and (iii) constant positive bias of 0.3. Random components were normally distributed with zero mean. Because of significantly larger computing resources required by the 'standard' method compared to NDA, the 'standard' DA used the reduced domain 300 × 300 km instead of 1000 × 1000 km for the NDA method. The bins required for the H–L method were 1 km in size.

The innovations were computed as $d = \mathbf{H}x^b - y$, where **H** is a subsampling matrix operator composed only of ones and zeroes that select from the coarse model the node that coincides in location with a child node. The 'observations' are located at all parent model grid nodes. Following the 'standard' method, we used a simple isotropic Gaussian function for the covariance of innovations at every parent model grid point *x*:

$$C_x(r) = a \exp\left(-r^2/D^2\right),\tag{16}$$

where *a* and *D* are fitting parameters. The binned covariances and the fitting curve are shown in Figure 17.



Figure 17. (a) Estimates of background and observational error covariances for multiple eddy field with $L_x = 40$ km and $L_y = 105$ km and the corresponding fitting curves. The error variances estimated at r = 0 are used as diagonal elements of the background matrix **B** (equal to 0.031) and the 'observational' matrix **R** (equal to 0.016); (b) the zonal transect at y = 40 km showing the true field (black line), parent model (dashed brown), child model noisy forecast before DA (red), analysis state after 'standard' DA (dotted magenta), and analysis after NDA (blue).

Despite the parent model is assumed 'perfect', i.e., it has no errors relative to the true field (apart from representativeness errors due to insufficient resolution), the H–L method gives it an error variance (diagonal elements of **R**) of 0.016. This is approximately 50% of the noisy child model error variance values (diagonal elements of **B**) of 0.031. The **B** matrix is then constructed using a variance equal to 0.031 (Figure 17a) and the Gaussian formula Equation (16), and the diagonal matrix **R** is built using a value for its variance of 0.016.

The **B** and **R** matrices were used to carry out the variational DA cycle to the simulated child model forecast with the following parameters: multiple eddy field with $L_x = 12$ km, $L_y = 105$ km, random noise with STD = 0.15, bias = 0.3, and spatial shift = 4 km to the west. The analysis state is estimated using the following equations of the 'standard' method ([25], p. 155)

$$\mathbf{x}^{a} = \mathbf{x}^{b} + \mathbf{W} \Big[\mathbf{y} - H \Big(\mathbf{x}^{b} \Big) \Big] = \mathbf{x}^{b} + \mathbf{W} d\mathbf{W} = \Big(\mathbf{B}^{-1} + \mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H} \Big)^{-1} \mathbf{H}^{T} \mathbf{R}^{-1}.$$

The transect in Figure 17b shows the analysis state for eddies with $L_x = 12$ km and $L_y = 105$ km obtained by the 'standard' method together with the true solution, parent model output, noisy child model forecast (before DA), and, for comparison, the analysis state obtained by the NDA method.

The bias and RMSE relative to the true solution for the child model forecast and the analysis state after 'standard' and NDA data assimilation process are shown in Table 1.

Table 1. Errors in the child model outputs before and after data assimilation.

	Forecast (before DA)	'Standard' DA	NDA
Bias	0.300	-0.0066	0.000
RMSE	0.608	0.5268	0.250

Figure 18 shows the map of the analysis state produced by the NDA and the standard methods. The errors are removed more efficiently in the areas of low values of field variable *F*.

300

Distance, km





Figure 18. The child model output after DA using (a) the 'standard' method and (b) the data assimilation with stochastic-deterministic downscaling (NDA). The NDA produces less noise.

4. Discussion

Standard, $L_x = 12 \text{ km}$

The examples presented show the data assimilation cycle using the NDA method when the data used for assimilation are not coming from observation as in the common assimilation methods but from a coarser parent model. The examples relate to a synthetic situation when the true values of the state variable are known. The forecast by the fineresolution child model is simulated by sampling the true field on the model's fine grid and adding various sources of correlated and uncorrelated errors—random noise, bias, and spatial shift. The output from the coarser parent model is regarded as computer generated 'observations' and is simulated by sampling the true field on the coarser grid to which random noise can be added. In practice, the parent model is assumed to be a goodquality, dynamic ocean model that properly assimilates observational data. Therefore, the actual field measurements are used in the NDA method indirectly; instead of assimilating observations, the child model assimilates data from the coarse model, which in turn assimilates data from observations. Therefore, the child and parent models are doublelinked, the parent model repeatedly provides to the child model: (i) 2D boundary conditions for creating a forecast and (ii) 3D data to improve the forecast and create the starting point for the next forecast.

The NDA is a two-step process. The first step of the NDA method is the application of stochastic-deterministic downscaling presented in [26]. As a result, at each fine grid point there are two generally different values of the field variable. The second step is to combine these values separately at each grid point using a suitable data assimilation method. In this paper, a modified cost function is used to obtain the best possible estimate of the analysis state. This approach leads to the use of a zero-dimensional Kalman, similar to what has been done in atmospheric chemistry [32].

The contribution of the first (SDD) and second (DA) steps to the improvement of RMSE and bias depends on a number of factors, such as the ratio of model resolution to the size of the feature in question, level of noise, and general quality of parent and child models. As the parent model in the examples considered above is assumed to be unbiased after its own DA procedure, it does not generate any bias after the downscaling step. By design of the DA step, the zero bias from the first step is carried forward to the final analysis state. The SDD downscaling can produce good results on its own, even without the need for running a fine-scale numerical model, examples of this are presented in [26]. Therefore, in case the child model has large errors, a combination of good-quality data from the first step (downscaling) with not so good-quality output from the child model using the Kalman gain

formula (Equation (10)) will worsen the RMSE. However, if the errors of the fine model are relatively small or if the fine model contains features that are not well represented in the coarse model, the two-step NDA RMSE can be significantly smaller than the first-step SDD RMSE. The gain formula used in the Kalman filter requires the knowledge of error variances. In this paper, the variances are assessed using the ergodic hypothesis, e.g., [36], and the ensemble statistical mean at each grid node is estimated by spatial averaging in a small trial domain around the node at the same time point. This scheme is fast and does not consume significant computational resources. If necessary, it can be extended by including data from preceding time points; however, these data may not be statistically independent and hence the potential improvement requires further study.

The study by [37] used a spectral nudging method in order to restrict the drift of largescale spectral components in the fine model from the global model. The main difference with the NDA is that the nudging technique uses weighting coefficients that are tuning parameters and are prescribed in advance, while the NDA variational method uses weights computed from the variance of the errors by minimising the cost function and therefore the weights cannot be changed at will. In contrast to the spectral nudging, the NDA method corrects both large- and small-scale components of the child model as seen in the amplitude spectra shown in Figure 16. The NDA automatically adjust the bias of the child model to the level of parent model which can be interpreted as an aggressive nudging of a single long-wave component of the field.

The forecast error covariance matrix **B** (for the child model) may be different from the error covariance matrix **R** (for the parent model) at very small distances not resolved by the parent model. It has to be noted that in any DA method, forecast error covariance matrices are not exact and are estimated based on using various approximations. Bannister [18] states: 'Owing partly to its very large rank, the B-matrix is impossible to use in an explicit fashion in an operational setting and so methods have been sought to model its important properties in a practical way'.

Whatever the approximation for the **B** (and *R*) matrix is used, the final judgement of the DA quality should be made on how well the DA reduces the model errors. The NDA method is shown to reduce the errors in the child model (Figures 3–16). It also demonstrates that the assumption of a similar spatial correlation structure between **B** and **R** matrices used in NDA for model-to-model data assimilation produces better results than the assumption of a non-correlated **R** matrix as applied in the widely used H–L method (Table 1). Therefore, the NDA method demonstrates the appropriateness of underlying assumptions and algorithms.

The analysis presented above shows that the NDA method is not the only one that can be applied to model-to-model data assimilation; the 'standard' method could also give reasonable results. However, when compared with the 'standard' DA method, the NDA gives better accuracy, stronger reduction in the RMS errors, and a complete removal of bias. The NDA is also more computationally efficient. We had no restrictions or limitations in computing the analysis for $1000 \times 1000 \text{ km}^2$ domain at 2.5 km resolution when using the NDA on our office PC. However, we were unable to carry out the 'standard' DA for a domain greater than $300 \times 300 \text{ km}^2$ on the same PC due to computing resource restrictions. In terms of speed of calculations, it required about 2 min to complete one full DA cycle, including calculation of the covariance matrix and weighting coefficients using the NDA method in the multiple eddy example. On the other hand, the standard method for the reduced domain took 24 min, including the creation of **B** and **R** matrices and calculation the analysis state.

As an illustration of the use of the NDA method for operational oceanography, we present below a snapshot of SST in the Lakshadweep Sea (Indian Ocean) from the parent global model at 1/12 degree of resolution available from Copernicus Marine Service, product GLOBAL_MULTIYEAR_PHY_001_030 [7,38] and two nested NEMO models: LD20 at 1/20 degree resolution and LD60 at 1/60 degree resolution, see Figure 19. The parent model assimilates sea surface temperature, sea surface level, and in situ TS profiles, and is

described in detail by their authors [38]. Both LD20 and LD60 child models were developed by the Plymouth Ocean Forecasting Centre team. The child models were run with and without NDA from 1 January 2015. The LD20 model used 3D temperature, salinity, and velocities data from the global model for NDA. The LD60 model was implemented with two versions of the NDA: (i) using the global model as the parent and (2) using the intermediate nested LD20 model as the parent. In all cases, the NDA was applied to 3D temperature, salinity, and velocities data from the parent models. The details of the LDA20 and LDA60 models will be published elsewhere [39].



Figure 19. Daily averaged SST for 26/04/2015 from (**a**) the operational data-assimilating global model at 1/120 resolution, generated using EU Copernicus Marine Service Information [38], (**b**) nested model LD20 at $1/20^0$ without NDA, (**c**) nested model LD20 at $1/20^0$ with NDA using 3D temperature, salinity, and velocities data from the global model, (**d**) nested model LD60 at $1/60^0$ without NDA, (**e**) nested model LD60 at $1/60^0$ with NDA using 3D temperature, salinity, and velocities data from the global model, (**d**) nested model LD60 at $1/60^0$ with NDA, using 3D temperature, salinity, and velocities data from the global model LD60 at $1/60^0$ with NDA using 3D temperature, salinity, and velocities data from the global model LD60 at $1/60^0$ with NDA using 3D temperature, salinity, and velocities data from the global model LD60 at $1/60^0$ with NDA using 3D temperature, salinity, and velocities data from the global model.

5. Conclusions

This paper introduces a data assimilation method where the data are assimilated into a high-resolution model from a coarser good-quality data-assimilating model, not directly from observations. An efficient and simple algorithm for model-to-model variational data assimilation method named Nesting with Downscaling and Data Assimilation (NDA) is developed. The theoretical background behind the NDA algorithm is discussed, and its application is illustrated in a number of idealised synthetic situations that resemble realworld practice in fine-resolution ocean modelling. Here, we followed the philosophy used by Stommel in his famous work [40], which explained the existence of the Gulf Stream using a synthetic idealised case—square ocean, flat bottom, and sinusoidal wind distribution. Our results demonstrate that the model-to-model data assimilation is an efficient way for improving the accuracy of a fine-resolution model. This approach allows for avoiding a repetition of a complex and resource-hungry assimilation of actual observations, which has already been done in the parent model. It is likely that the same basic idea of model-tomodel data assimilation would also work for other methods currently used in observational data assimilation. In this paper, the NDA is compared with a commonly used variational method and shown to be more accurate and significantly less computationally expensive.

Author Contributions: Conceptualization, G.I.S.; methodology, G.I.S.; software, G.I.S. and J.M.G.-O.; formal analysis, G.I.S.; writing—original draft preparation, G.I.S.; writing—review and editing, G.I.S. and J.M.G.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: For code availability, please contact the corresponding author.

Acknowledgments: The authors are thankful to the members of Plymouth Ocean Forecasting Centre for their support. We are also thankful to M. S. Poovadiyil for providing Figure 19.

Conflicts of Interest: The authors declare no conflict of interest.

References

- GFDL. High-Resolution Modeling. 2020. Available online: https://www.gfdl.noaa.gov/high-resolution-modeling (accessed on 6 July 2020).
- Dufour, C.O.; Griffies, S.; de Souza, G.; Frenger, I.; Morrison, A.; Palter, J.; Sarmiento, J.L.; Galbraith, E.; Dunne, J.P.; Anderson, W.G.; et al. Role of mesoscale eddies in cross-frontal transport of heat and biogeochemical tracers in the Southern Ocean. *J. Phys. Oceanogr.* 2015, 45, 3057–3081. [CrossRef]
- 3. Shapiro, G.I.; Stanichny, S.; Stanychna, R.R. Anatomy of shelf–deep sea exchanges by a mesoscale eddy in the North West Black Sea as derived from remotely sensed data. *Remote Sens. Environ.* **2010**, *114*, 867–875. [CrossRef]
- 4. Meunier, T.; Barton, E.D.; Barreiro, B.; Torres, R. Upwelling filaments off Cap Blanc: Interaction of the NW African upwelling current and the Cape Verde frontal zone eddy field? *J. Geophys. Res. Atmos.* **2012**, *117*, C08031. [CrossRef]
- NEMO. Annual Mean of Sea Surface Salinity in 1/12° (NEMO-WRF Coupling). Available online: https://www.nemo-ocean.eu (accessed on 16 June 2021).
- 6. ROMS. Regional Ocean Modeling System (ROMS). Available online: https://www.myroms.org (accessed on 16 June 2021).
- CMEMS. Ocean Explainers: Oceanography Educational Webpages. Available online: https://marine.copernicus.eu (accessed on 20 May 2021).
- DARC. What Is Data Assimilation? Available online: https://research.reading.ac.uk/met-darc/aboutus/what-is-dataassimilation (accessed on 5 August 2021).
- 9. Lorenc, A.C. Analysis methods for numerical weather prediction. Q. J. R. Meteorol. Soc. 1986, 112, 1177–1194. [CrossRef]
- Bell, M.J.; Forbes, R.M.; Hines, A. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. J. Mar. Syst. 2000, 25, 1–22. [CrossRef]
- 11. Gandin, L.S. The problem of optimal interpolation, Scientific papers. Main Geophys. Obs. 1959, 99, 67–75.
- Moore, A.M.; Martin, M.J.; Akella, S.; Arango, H.G.; Balmaseda, M.A.; Bertino, L.; Ciavatta, S.; Cornuelle, B.; Cummings, J.; Frolov, S.; et al. Synthesis of Ocean Observations Using Data Assimilation for Operational, Real-Time and Reanalysis Systems: A More Complete Picture of the State of the Ocean. *Front. Mar. Sci.* 2019, *6*, 90. [CrossRef]
- 13. Dobricic, S.; Pinardi, N. An oceanographic three-dimensional variational data assimilation scheme. *Ocean. Model* **2008**, *22*, 89–105. [CrossRef]
- 14. Ghil, M.; Malanotte-Rizzoli, P. Data Assimilation in Meteorology and Oceanography. Adv. Geophys. 1991, 33, 141–266. [CrossRef]

- NEMOVAR. The NEMOVAR Ocean Data Assimilation System as Implemented in the ECMWF Ocean Analysis for System 4. Available online: https://www.ecmwf.int/en/elibrary/11174-nemovar-ocean-data-assimilation-system-implemented-ecmwfocean-analysis-system-4 (accessed on 1 October 2021).
- 16. DART, Data Assimilation Research Testbed. Available online: https://dart.ucar.edu (accessed on 1 October 2021).
- 17. PDAF, Parallel Data Assimilation Network. Available online: https://pdaf.awi.de/trac/wiki (accessed on 1 October 2021).
- 18. Bannister, R.N. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 1971–1996. [CrossRef]
- Waters, J.; Lea, D.J.; Martin, M.J.; Mirouze, I.; Weaver, A.; While, J. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. Q. J. R. Meteorol. Soc. 2015, 141, 333–349. [CrossRef]
- 20. Haben, S.; Lawless, A.; Nichols, N. Conditioning and preconditioning of the variational data assimilation problem. *Comput. Fluids* **2011**, *46*, 252–256. [CrossRef]
- Kubryakov, A.I.; Korotaev, G.K.; Dorofeev, V.L.; Ratner, Y.B.; Palazov, A.; Valchev, N.; Malciu, V.; Mateescu, R.; Oguz, T. Black Sea coastal forecasting system. *Ocean Sci.* 2012, *8*, 183–196. [CrossRef]
- 22. Carrassi, A.; Bocquet, M.; Bertino, L.; Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim. Chang.* **2018**, *9*, e535. [CrossRef]
- 23. Bouttier, F.; Courtier, P. Data Assimilation Concepts and Methods. 1999. Available online: https://www.ecmwf.int/en/elibrary/ 16928-data-assimilation-concepts-and-methods (accessed on 5 August 2021).
- 24. Hollingsworth, A.; Lönnberg, P. The statistical structure of short-range forecast errors as determined from radiosonde data, Part 1: The wind field. *Tellus A* **1986**, *38*, 111–136. [CrossRef]
- 25. Kalnay, E. Atmospheric Modeling, Data Assimilation and Predictability; Cambridge University Press (CUP): Cambridge, UK, 2003; 369p.
- Shapiro, G.I.; Gonzalez-Ondina, J.M.; Belokopytov, V.N. High-resolution stochastic downscaling method for ocean forecasting models and its application to the Red Sea dynamics. *Ocean Sci.* 2021, 17, 891–907. [CrossRef]
- 27. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. [CrossRef]
- Polavarapu, S.; Ren, S.; Rochon, Y.; Sankey, D.; Ek, N.; Koshyk, J.; Tarasick, D. Data assimilation with the Canadian middle atmosphere model. *Atmosphere-Ocean* 2005, 43, 77–100. [CrossRef]
- Zingerlea, C.; Nurmib, P. Monitoring and verifying cloud forecasts originating from operational numerical models. *Meteorol. Appl.* 2008, 15, 325–330. [CrossRef]
- Bruciaferri, D.; Shapiro, G.; Wobus, F. A multi-envelope vertical coordinate system for numerical ocean modelling. *Ocean Dyn.* 2018, 68, 1239–1258. [CrossRef]
- Mellor, G.L.; Blumberg, A.F. Modeling Vertical and Horizontal Diffusivities with the Sigma Coordinate System. *Mon. Weather Rev.* 1985, 113, 1379–1383. [CrossRef]
- Adhikary, B.; Kulkarni, S.; Dallura, A.; Tang, Y.; Chai, T.; Leung, L.; Qian, Y.; Chung, C.; Ramanathan, V.; Carmichael, G. A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique. *Atmos. Environ.* 2008, 42, 8600–8615. [CrossRef]
- 33. Fedorov, K.N. The Physical Nature and Structure of Oceanic Fronts; Springer: Berlin/Heidelberg, Germany, 1986; 333p, ISBN 978-0-387-96445-4.
- 34. Robinson, A.R. (Ed.) . Eddies in Marine Science; Springer: Berlin/Heidelberg, Germany, 1983; 612p, ISBN 978-3-642-69003-7.
- 35. Stewart, L.M.; Dance, S.L.; Nichols, N.K.; Eyre, J.R.; Cameron, J. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1236–1244. [CrossRef]
- 36. Stull, R.B. An introduction to boundary layer meteorology. Kluwer 2003, 670. [CrossRef]
- Katavouta, A.; Thompson, K.R. Downscaling ocean conditions with application to the Gulf of Maine, Scotian Shelf and adjacent deep ocean. Ocean Model. 2016, 104, 54–72. [CrossRef]
- Global Ocean Physics Reanalysis (GLOBAL_MULTIYEAR_PHY_001_030). Available online: https://resources.marine.copernicus. eu/product-detail/GLOBAL_MULTIYEAR_PHY_001_030/INFORMATION (accessed on 7 March 2022).
- 39. Stommel, H. The westward intensification of wind-driven ocean currents. Trans. Am. Geophys. Union 1948, 29, 202–206. [CrossRef]
- 40. Shapiro, G.I.; Ondina, J.M.; Poovadiyil, M.S.; Tu, J.; Asif, M. *The Operational Ocean Modelling System of the Lakshadweep Sea*; University of Plymouth: Plymouth, UK, 2022; (*Manuscript in Preparation*).