Efficient estimation of time-mean states of ocean models using 4D-Var and implicit time-stepping

A. D. Terwisscha van Scheltinga, H. A. Dijkstra

Nonlin. Processes Geophys., 14, 777–788, 2007
www.nonlin-processes-geophys.net/14/777/2007/

Nonlinear Processes in Geophysics

Efficient estimation of time-mean states of ocean models using 4D-Var and implicit time-stepping

A. D. Terwisscha van Scheltinga and H. A. Dijkstra

Institute for Marine and Atmospheric research Utrecht, Department of Physics and Astronomy, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands

Received: 4 July 2007 – Revised: 1 November 2007 – Accepted: 3 November 2007 – Published: 30 November 2007

Abstract. We propose an efficient method for estimating a time-mean state of an ocean model subject to given observations using implicit time-stepping. The new method uses (i) an implicit implementation of the 4D-Var method to fit the model trajectory to the observations, and (ii) a preprocessor which applies a multi-channel singular spectrum analysis to enhance the signal-to-noise ratio of the observational data and to filter out the high frequency variability. This approach enables one to estimate the time-mean model state using larger time-steps than is possible with an explicit model. The performance of the method is presented for two test cases within a barotropic quasi-geostrophic nonlinear model of the wind-driven double-gyre ocean circulation. The method turns out to be accurate and, in comparison with the time-mean state computed with an explicit version of the model, relatively cheap in computational cost.

1 Introduction

The ocean has been routinely observed for the past decades. These observations mainly consist of satellite measurements of sea surface height (TOPEX/POSEIDON), sea surface temperature (AVHRR) supplemented by hydrographic data collected from a variety of sources ranging from cruises along sections (WOCE) to drifter and floats (ARGOS). The analysis of these observational data provides much information on the large-scale ocean currents. It is, however, not sufficient to fully describe the time-mean ocean circulation due to problems in coverage, spatial and/or temporal resolution or in accuracy. To obtain an accurate analysis of the time-mean ocean circulation, it is therefore necessary to combine these observations with an ocean model using data-assimilation or inverse modeling techniques.

Correspondence to: A. D. Terwisscha van Scheltinga
(a.d.terwisschavanscheltinga@phys.uu.nl)

One of the data-assimilation methods widely applied in physical oceanography is four-dimensional variational data-assimilation (4D-Var). It is a method in which information that is present in observations is combined with the evolution determined by a particular ocean, atmosphere or climate model. Estimates of the ocean state from 1992 through 2000 were calculated by assimilating WOCE data and NCEP reanalysis of the surface fluxes (Stammer et al., 2003). Observations are also used in operational oceanography to initialize ocean circulation models (Vialard et al., 2003) or to estimate model parameters for example diffusivities (Stammer, 2005) or eddy stresses (Ferreira et al., 2005).

Given an ocean model and observations the aim of 4D-Var is to find an initial state and/or model parameters, such that the observations are "close" to the model trajectory. A cost function is formulated which measures the distance of the model trajectory to the observations. Minimization of this cost function over the initial conditions (or parameters) gives the so-called analysis. The minimization procedure requires a gradient, which in general is evaluated using a forward and an adjoint model. Compilers exits which generate actual computer code of an adjoint model given the code of the forward model (Giering and Kaminski, 1998), but the formulation of an adjoint model is in most cases a nontrivial and time-consuming process.

A direct approach of determining an estimate of the time-mean state of the ocean, is to assimilate the observations into an ocean model and calculate the time-mean state of the resulting analysis. For a data-assimilation method with an explicit time-stepping model this has high computational cost since small time-steps have to be taken due numerical stability constraints, such as the CFL-criterion (Peyret and Taylor, 1983). Furthermore, a small time-step requires a high temporal resolution of the observational data, which may not be available. Small time-steps also make the estimate sensitive to observational noise and high frequency, small amplitude variability.

Published by Copernicus Publications on behalf of the European Geosciences Union and the American Geophysical Union.

The 4D-Var method has recently been implemented in implicit time-stepping models where the advantage is that the adjoint can be constructed directly from the implicit time-stepping scheme, such as the Crank-Nicholson scheme (Terwisscha van Scheltinga and Dijkstra, 2005). This new implementation has several advantages over variational data-assimilation with an explicit model: (i) a larger time-step can be taken since the size of the time-step is not bounded by the CFL-criterion, and (ii) the method is more accurate for similar time-steps (Terwisscha van Scheltinga and Dijkstra, 2007). The possibility of taking a large time-step is an advantage that we will utilize here for the estimation of the time-mean state.

To be able to use large time steps, one needs a way to handle the smaller time-scale variability in the observations. This variability can usually be decomposed into statistical modes (Ghil et al., 2002) using statistical techniques for example, Multi-channel Singular Spectrum Analysis (M-SSA); these techniques aim at enhancing the signal-to-noise. The observations can then be reconstructed using only the M-SSA dominant modes, or when required, only those with frequencies within a certain frequency band.

In this paper we explore the use of implicit models in data-assimilation and propose an efficient estimation method for the time-mean state in an ocean model under given observations. The approach consists of two building blocks:

i) a method to pre-process the observations prior to data-assimilation, i.e. a data-handling procedure that uses multi-channel spectrum analysis (MSSA) to enhance the signal-to-noise ratio and reconstruct the observations such that only the dominant modes and relevant time-scales are assimilated; and

ii) the 4D-Var method using an implicit time-stepping scheme, which allows us to take relatively large time-steps.

The method will be tested using the barotropic quasi-geostrophic model of the wind-driven ocean circulation as presented in Sect. 2. We will consider both a case of externally forced variability (Sect. 3.1) and a case of internal variability (Sect. 3.2), the latter arising through Hopf bifurcations. For both cases, the estimation method is compared with a direct approach of estimating the time-mean state using an explicit version of the model.

2 Model and methods

In this section we will first provide (Sect. 2.1) the model of the wind-driven ocean circulation which is used in this study. Next, we provide a basic overview of the 4D-Var method (Sect. 2.2) followed by a detailed discussion of the estimation method for the time-mean state in the ocean model (Sect. 2.3).

2.1 Quasi-geostrophic model

Consider a rectangular ocean basin of size $L\times L$ having a constant depth D. The basin is situated on a mid-latitude β-plane with a central latitude θ_0=45° N and Coriolis parameter $f_0=2\Omega\sin\theta_0$, where Ω is the rotation rate of the Earth. The meridional variation of the Coriolis parameter at the latitude θ_0 is indicated by β_0. The density ρ of the water is constant and the flow is forced at the surface through a wind-stress vector $\boldsymbol{T}=\tau_0[\tau^x(x,y),\tau^y(x,y)]$. The governing equations are non-dimensionalized using a horizontal length scale L, a vertical length scale D, a horizontal velocity scale U, the advective time scale L/U and a characteristic amplitude of the wind-stress vector, τ_0. The effect of deformations of the ocean-atmosphere interface on the flow is neglected.

The dimensionless barotropic quasi-geostrophic model of the flow for the vertical component of the vorticity, ζ, and the geostrophic streamfunction ψ is (Pedlosky, 1987)

$$\Big[\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}\Big][\zeta+\beta y]=Re^{-1}\nabla^2\zeta+\alpha_\tau\Big(\frac{\partial\tau^y}{\partial x}-\frac{\partial\tau^x}{\partial y}\Big), \quad (1a)$$

$$\zeta=\nabla^2\psi, \quad (1b)$$

where the horizontal velocities are given by $u=-\partial\psi/\partial y$ and $v=\partial\psi/\partial x$. The parameters in Eq. (1a) are the Reynolds number Re, the planetary vorticity gradient parameter β and the wind-stress forcing strength α_τ. These parameters are defined as:

$$Re=\frac{UL}{A_H};\quad \beta=\frac{\beta_0L^2}{U};\quad \alpha_\tau=\frac{\tau_0L}{\rho DU^2} \quad (2)$$

where g is the gravitational acceleration and A_H is the lateral friction coefficient. When the horizontal velocity scale U is based on a Sverdrup balance of the flow, i.e.,

$$U=\frac{\tau_0}{\rho D\beta_0L}. \quad (3)$$

it follows that $\alpha_\tau=\beta$ and there are only two free parameters (e.g., Re and β).

We assume no-slip conditions on the east-west boundaries and slip on the north-south boundaries. The boundary conditions are therefore given by

$$x=0,x=1:\quad \psi=\frac{\partial\psi}{\partial x}=0, \quad (4a)$$

$$y=0,y=1:\quad \psi=\zeta=0. \quad (4b)$$

The wind-stress forcing is prescribed as the sum of a symmetric steady wind forcing with added time-dependent components with different frequencies and amplitudes, i.e.,

$$\tau^x(x,y)=\frac{-1}{2\pi}\cos(2\pi y)+\frac{a_1(t)}{\pi}\cos(\pi y)+\frac{a_2(t)}{3\pi}\cos(3\pi y)+\frac{a_3(t)}{4\pi}\cos(4\pi y)+\frac{a_4(t)}{5\pi}\cos(5\pi y) \quad (5a)$$

$$\tau^y(x,y)=0. \quad (5b)$$

Table 1. Standard values of the parameters for the barotropic quasi-geostrophic ocean model in the steady flow regime.

Parameter	Value	
L	1.0×10^{6}	m
U	7.1×10^{-3}	m
D	7.0×10^{2}	m
β_0	2.0×10^{-11}	$(\mathrm{ms})^{-1}$
f_0	1.0×10^{-4}	s^{-1}
g	9.8	ms^{-2}
ρ	1.0×10^{3}	kgm^{-3}
τ_0	1.0×10^{-1}	Pa
P_1^*	360	days
P_2^*	90	days
P_3^*	30	days
P_4^*	7	days

The amplitudes a_i are given by:

$$a_i(t) = \frac{\tau_i 2^{4-i}}{60}\sin(\frac{2\pi t}{P_i}). \tag{6}$$

where P_i is the dimensionless period of the amplitude and $\tau_i \in [0, 1]$ is a control parameter. The values for the dimensional period $P_i^* = L P_i / U$ are given in Table 1; $a_1(t)$ has a period of one year, $a_2(t)$ a period of three months, $a_3(t)$ a period of a month and $a_4(t)$ a period of a week. The amplitude and the spatial scale of the time-dependent wind-stress components $a_i(t)$ decrease with i.

For the parameters as in Table 1 and symmetric forcing ($\tau_i=0$, $i=1, ..., 4$), (Dijkstra and Katsman, 1997) showed that several different flow regimes exist when Re is varied. For $Re<30$, the quasi-geostrophic model has one unique stable steady state. The streamfunction ψ of this steady state is anti-symmetric with respect to the mid-axis of the basin. Two asymmetric stable steady-state solutions, one with a downward jet-displacement and the jet-up solution exist for $30<Re<52$. Near $Re=52$ both asymmetric states become unstable due to the occurrence of Hopf bifurcations; for $52<Re<74$ stable periodic orbits exist. The solutions become quasi-periodic for $Re>74$ and irregular for higher values of Re; the route to chaos is through a homoclinic orbit (Simonnet et al., 2005).

With this model we will test our time-mean estimation method presented in Sect. 2.3 for internal and externally induced variability. For the former we will take $Re>74$ and $\tau_i=0$, $i=1, ..., 4$, i.e. the quasi-periodic regime described above. For the latter we take $Re=50$ and $\tau_i=1$, $i=1, ..., 4$, the steady regime where we induce variability through the time-dependent components of the wind stress.

The equations Eq. (1a) and boundary conditions Eq. (4) are spatially discretized using a control-volume method on an equidistant $N\times M$ grid. A standard set of parameter values has been chosen (Table 1) that are similar to those in (Dijkstra and Katsman, 1997) and for these parameters, $\alpha_\tau=\beta=2.8\times10^3$. For the implicit time integration the second-order Crank-Nicholson scheme is used, while for the explicit time integration a second order Adams-Bashforth scheme is used.

2.2 Variational data-assimilation

Suppose the initial condition of the background model $\boldsymbol{w}^b(t_0)$ is given. The analysis is the model trajectory that simultaneously minimizes the distance to the initial background $\boldsymbol{w}^b(t_0)$ and the observations $\{\boldsymbol{y}_i : i=0, \cdots, N-1\}$. This is an optimization problem, which in the incremental 4D-Var formulation (Courtier et al., 1994) is stated as:

$$\delta\boldsymbol{w}^a = \min_{\delta\boldsymbol{w}} J(\delta\boldsymbol{w}), \tag{7a}$$

$$J(\delta\boldsymbol{w}) = \delta\boldsymbol{w}^T\mathbf{B}^{-1}\delta\boldsymbol{w} + \sum_{i=0}^{N-1}\boldsymbol{d}_i^T\mathbf{R}_i^{-1}\boldsymbol{d}_i, \tag{7b}$$

$$\boldsymbol{d}_i = \boldsymbol{y}_i - H_i M(t_i, t_0)(\boldsymbol{w}^b(t_0)) + \mathbf{H}_i\mathbf{M}(t_i, t_0)\delta\boldsymbol{w}. \tag{7c}$$

In above equations J is the cost function which measures the distance to the observations and the initial conditions, $\delta\boldsymbol{w}^a$ is the optimal increment on the initial background $\boldsymbol{w}^b(t_0)$ state and $\boldsymbol{d}_i$ is the departure of the model trajectory from observation $\boldsymbol{y}_i$. The operators $M(t_i, t_0)$ and H_i are the evolution operator and the observation operator with $\mathbf{M}(t_i, t_0)$ and $\mathbf{H}_i$ their linearizations around the background trajectory $\boldsymbol{w}^b(t_i)$. The matrices $\mathbf{B}$ and $\mathbf{R}_i$ are the covariance matrices for the background errors and observational errors. Given an optimum $\delta\boldsymbol{w}^a$ of Eq. (7a) the analysis $\boldsymbol{w}^a(t_i)$ is given by:

$$\boldsymbol{w}^a(t_i) = M(t_i, t_0)(\boldsymbol{w}^b(t_0) + \delta\boldsymbol{w}^a). \tag{8}$$

The method used for the solution of the minimization problem described here needs the gradient ∇J, which for Eq. (7a) is given by:

$$\nabla J = \sum_{i=0}^{N-1}\mathbf{M}(t_i, t_0)^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\boldsymbol{d}_i. \tag{9}$$

For the explicit model the procedure is to evaluate the cost function by using explicit time-stepping, while the gradient is evaluated using the adjoint model $\mathbf{M}(t_i, t_0)^T$ to integrate backward in time. For 4D-Var with implicit time-stepping, the construction of the adjoint model is easy (Terwisscha van Scheltinga and Dijkstra, 2005), since the adjoint model is directly available from the Newton-Raphson method used in the Crank-Nicholson method. Furthermore, the implicit implementation also has the advantage that the time-step is not limited by numerical stability and is capable of finding

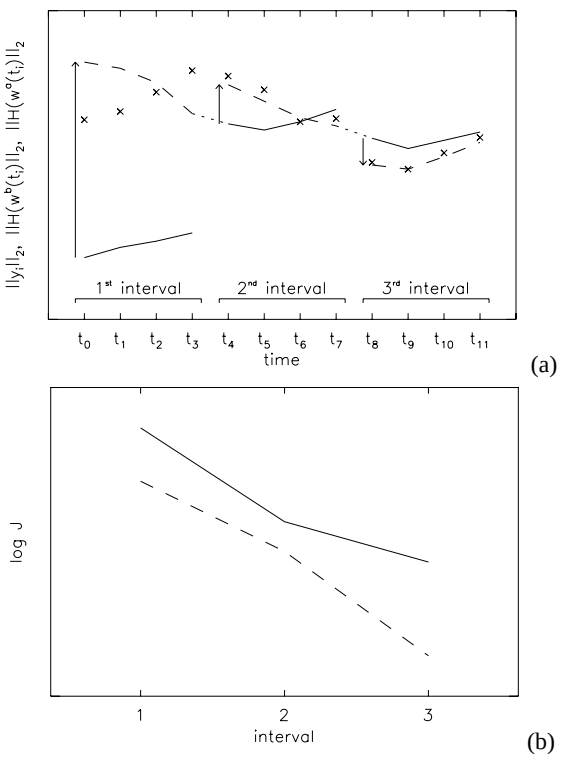

Fig. 1. Sketch of the 4D-Var method, where an assimilation interval has been divided into three subintervals. **(a)** $\|H(\boldsymbol{w}^b(t_i)\|_2$ and $\|H(\boldsymbol{w}^a(t_i)\|_2$, the $\mathcal{L}_2$-norm of the projection of the background $\boldsymbol{w}^b(t_i)$ (solid) and analysis $\boldsymbol{w}^a(t_i)$ (dashes) on the observations space; the $\mathcal{L}_2$-norm of the observations $\boldsymbol{y}_i$ (crosses) and the optimal increments $\delta\boldsymbol{w}^a$ (arrows). **(b)** the initial (solid) and final (dashed) value of the cost function.

an accurate analysis for large time steps (Terwisscha van Scheltinga and Dijkstra, 2007).

The minimization problem for the cost function J is solved using a limited memory quasi-Newton conjugate gradient method. This method terminates successfully if all of the following conditions on the convergence of the cost function, increment and gradient are met:

$$J^{l-1} - J^l < \epsilon_m(1 + |J^l|), \tag{10a}$$

$$\|\delta\boldsymbol{w}^{l-1} - \delta\boldsymbol{w}^l\| < \epsilon_m^{1/2}(1 + \|\delta\boldsymbol{w}^l\|), \tag{10b}$$

$$\|\nabla J^l\| \leq \epsilon_m^{1/3}(1 + |J^l|), \tag{10c}$$

where l is the iteration index of the conjugate gradient method and ϵ_m the optimality tolerance.

For time-series with many observations in the time-domain, i.e. large N, it is more practical to divide the large time series into m smaller (sub-)intervals, each with n points. The observation in each sub-interval are then assimilated using the analysis of the previous one. An example of this is presented in Fig. 1, where the observations are shown on an assimilation interval of 12 points. This interval is divided into 3 subintervals (m=3), each with four points (n=4). For every interval the background trajectory $\boldsymbol{w}^b(t_i)$ (solid), the optimal increment $\boldsymbol{w}^a$ (arrows) and the analysis $\boldsymbol{w}^a(t_i)$ (dashed) are shown. The background on the first interval is given. For the other intervals, the background is calculated from the analysis on the previous interval. On each interval the minimization problem is solved. Due to the dependence of the cost function on the background, the increment and the observations, the initial and the final value of the cost function will vary over the subintervals (Fig. 1b).

2.3 Time-mean state estimation

To fully utilize the advantage of taking a large time step, we only want to assimilate those statistical modes of the observations which explain most of the variance seen in the observational data. These modes are then separated into several frequency bands, ordered from low to higher frequencies. For large time steps we only assimilate the modes in the low frequency band, while for smaller time steps we will retain the low frequency modes but include higher frequencies. The underlying assumption is that for large time steps we can quickly estimate a time-mean state based on only the low frequency variability in the observations, while the refinement of the time-mean state due to the assimilation of high frequency variability is expected to be relatively small.

To illustrate the approach graphically, a hypothetical observational time-series is shown in Fig. 2a. The variability of this time series can be decomposed into several modes, as is shown in Fig. 2b, each with a different frequency. From this picture it becomes clear that to produce an accurate analysis, each mode of variability can be handled with a different time step during assimilation. The low-frequency variability can be assimilated with a much larger time step than the high frequency variability. For large time steps (Fig. 2c) we only assimilate the low frequency mode, while for smaller time steps we take both the low and medium frequency variability into account. Each assimilation of reconstructed observations produces an estimate for the time-mean state, which is successively improved by adding more higher frequency components, while simultaneously reducing the size of the time step.

To obtain the statistical mode decomposition in the observations, we use the M-SSA method applied on the leading principle components (PC) of the observations (Vautard and Ghil, 1989). After the leading M-SSA modes have been calculated the observations are reconstructed. Depending on the size of the time step, this reconstruction will be based on the modes in only one or in more frequency bands.

The algorithm for the time-mean estimation hence consists of two parts: (i) a data-handling procedure, which produces several reconstructions of the observations from the M-SSA modes; (ii) iterative assimilation of the reconstructed observations using 4D-Var. The steps in the data-handling procedure are the following:

1. Calculate the mean $\bar{\boldsymbol{y}}$ from the observations $\boldsymbol{y}_i$ in the interval $t_0 \leq t_i \leq t_n$. If a trend is present in the data, then this trend must be removed before calculating the mean.

2. Calculate the F leading M-SSA modes (see Appendix), i.e. those who explain most of the spatial-temporal variance. The number of modes F depends on how much of the variance should be explained (usually, 90% or more). Note that this is highly dependent on the quality of the observations, but in practice F is not large, say 10–20.

3. Separate the different modes of variability, i.e., divide the F modes into sets of modes $\mathcal{K}_j$ with each set representing a band of frequencies (e.g., seasonal, interannual, decadal). The sets $\mathcal{K}_j$ must satisfy:

$$\mathcal{K}_j \cap \mathcal{K}_k = \emptyset, \quad j \neq k \tag{11a}$$

$$\bigcup_{j=1}^{K} \mathcal{K}_j = \mathcal{K}_F, \tag{11b}$$

   where $\mathcal{K}_F$ contains the F leading modes. The sets are ordered from low to high frequencies bands.

4. Reconstruct the observations as follows:

$$\tilde{\mathcal{K}}_1 = \emptyset, \tag{12a}$$

$$\tilde{\mathcal{K}}_j = \tilde{\mathcal{K}}_{j-1} \cup \mathcal{K}_j, \tag{12b}$$

$$\boldsymbol{y}_i^j = \bar{\boldsymbol{y}} + \boldsymbol{R}_{\tilde{\mathcal{K}}_j}(t_i), \tag{12c}$$

   where $\boldsymbol{y}_i^j$ is the jth reconstruction based on the modes in $\tilde{\mathcal{K}}_j$, and $\boldsymbol{R}_{\tilde{\mathcal{K}}_j}(t_i)$ is the reconstruction of the deviations from the mean using the modes in $\tilde{\mathcal{K}}_j$. Note that the first reconstruction of the observational time-series is the mean $\bar{\boldsymbol{y}}$. For the estimation of the mean we will iterate in j, or equivalently, over the reconstructed observations $\boldsymbol{y}_i^j$. Since the sets were ordered from low to high frequencies, $\tilde{\mathcal{K}}_j$ contains all the frequencies up to a certain frequency band.

For a fixed value of j the series $\boldsymbol{y}_i^j$ will be assimilated using 4D-Var and from the resulting analysis the time-mean state is estimated. For each j we have to choose a time-step Δt and size of the (sub-)intervals and point per (sub-)interval (see Fig. 1). The steps in this procedure are as follows:

1. Choose a time step Δt_j, where $\Delta t_j < \Delta t_{j-1}$. Here the choice of Δt_j depends on the variability present in the reconstructed observations $\boldsymbol{y}_i^j$. It is necessary to have a good temporal resolution in order to produce an accurate analysis and estimate for the time-mean state. The size of the time-step must be smaller than the period of the variability present in the $\boldsymbol{y}_i^j$.

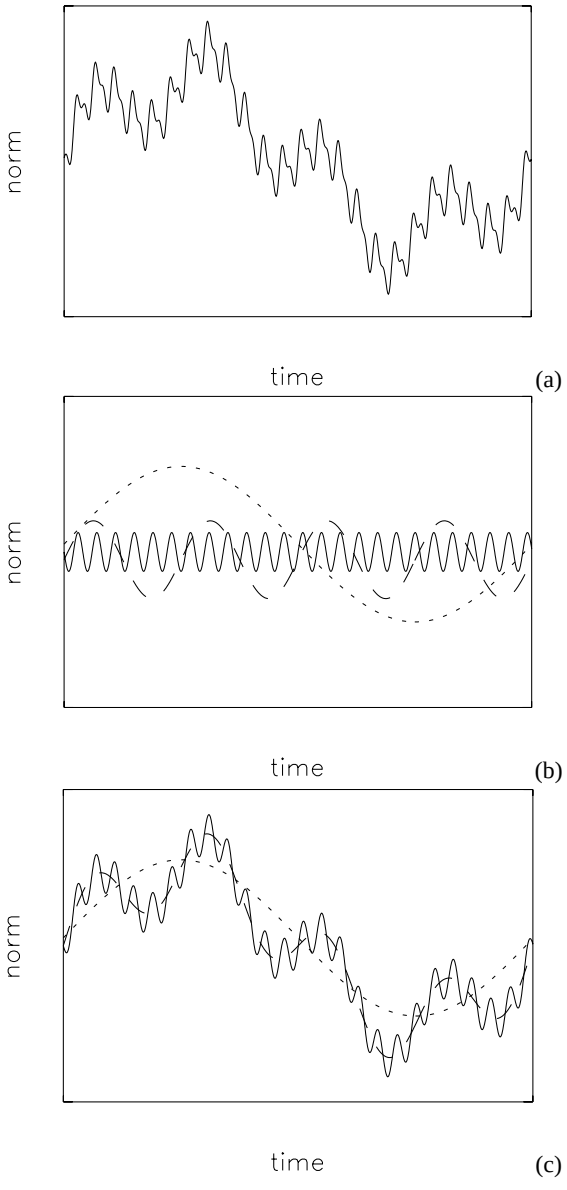

Fig. 2. Sketch of the time-mean estimation: **(a)** the observational time-series; **(b)** three modes of variability present in the time-series: the low frequency mode (dotted), medium frequency mode (dashed) and the high frequency mode (solid); **(c)** three reconstructions of this time-series based on the low frequency mode (dotted), low and medium frequency modes (dashed) and low, medium and high frequency modes (solid).

2. Set the number of subintervals m_j and the number of points per subinterval n_j (Fig. 1). Both m_j and n_j will vary with the size of the time step Δt_j, since the following must hold:

$$t_{\text{total}} = m_j n_j \Delta t_j, \tag{13}$$

   where t_{total} is the length of the time interval. Note that it

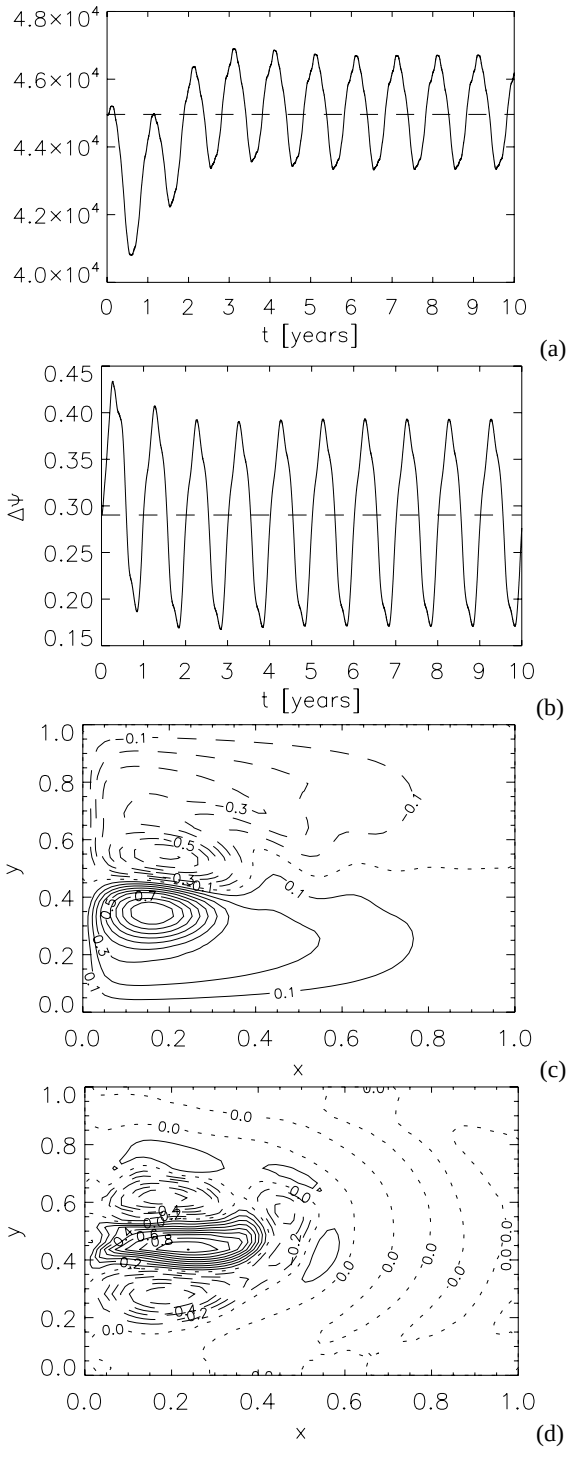

Fig. 3. Overview of the model run at Re=50 and τ_i=1 for all i: **(a)** the basin integrated kinetic energy E_{kin}; **(b)** the asymmetry of the streamfunction ψ; **(c)** the time-mean of the streamfunction over the last five years, the contours are with respect to an absolute maximum of ψ=2.2; and **(d)** the difference between this time-mean state and the jet-down steady-state, the contours are with respect to an absolute maximum of ψ=0.055.

is not necessary to use all the available (reconstructed) observations. Depending on the choice of m_j, n_j and Δt_j only the observations at some t_i will be used.

3. Calculate an estimate for the time-mean state by taking the average over the calculated analysis.

During an iteration over j the setup of the assimilation (m_j, n_j and Δt_j) and the assimilated observations ($\boldsymbol{y}_i^j$) change. In the remainder of this paper, a steady state will be used as the initial background for each iteration j. We do not use the analyses of previous iterations as the background since it was found difficult to implement, difficult to use in combination with the varying setup of the assimilation and it did not yield significantly better results for the test cases presented below.

3 Results

The methodology will be applied to the model presented in Sect. 2.1, using the parameters from Table 1 and a model resolution of 60×40 gridpoints. We will test the new time-mean estimation method for two cases. In the first case (Sect. 3.1) we choose Re=50 where the equilibrium flow is steady under symmetric time-independent wind forcing. External variability is introduced through the time-dependent components of the wind stress by taking τ_i=1 for all i in Eq. (5). In the second case (Sect. 3.2) we take Re=80 such that the equilibrium flow is irregular due to internal instabilities (arising through Hopf bifurcations) and τ_i=0 for all i in Eq. (5) such that there is no externally induced variability. In all the results below, for convenience we will use dimensional values of time and time step but we will keep the notation t and Δt.

3.1 Externally induced variability

First a 10 year model run was performed for Re=50 with a time step Δt=3 h. The wind-stress forcing is given by Eq. (5) with τ_i=1 for all i and the model run is started from the jet-down steady-state solution at Re=50. In the Figs. 3a–b the dimensionless kinetic energy E_{kin} and the asymmetry of the streamfunction $\Delta\Psi$ defined by:

$$\Delta\Psi = \frac{\max(\psi) + \min(\psi)}{\max(\psi, -\psi)}. \tag{14}$$

are plotted versus time. The kinetic energy and the asymmetry of the streamfunction for the jet-down steady-state solution have also been plotted (dashed). In both the Figs. 3a–b, it can be seen that after a few years the streamfunction fluctuates around the steady state in a regular fashion; the variability of this signal has a dominant period of one year (by construction). In Fig. 3c, the time-mean of the streamfunction over these five years is shown. The difference of this time-mean state and the jet-down steady-state solution is plotted in Fig. 3d. The time-mean state has also a jet-down structure

and the difference pattern in Fig. 3d shows the signature of the P-mode found in (Dijkstra and Katsman, 1997).

From this simulation, the last 5 years were used for the generation of the observations. All values ψ on the 60×40 grid were taken as the observations. After subtracting the time-mean state, the 20 leading EOFs and PCs were calculated. Since we use unperturbed observations generated by a model, 100% of the variance in the observations could be explained. Based on these principle components, the leading 20 M-SSA modes were calculated, which in total accounted for more than 90% of the variance. The leading pair of modes has a frequency of once a year and higher frequencies are less dominant. The 20 M-SSA modes were separated into three frequency bands: low frequencies (period of a year or more), a mid-frequency range (period of a quarter to a year) and the higher frequency range (periods of a quarter or less).

From these frequency bands three reconstructions of the observations were made. The first reconstruction was based on the low frequency band, the second and third by adding the mid-frequencies and the high frequencies, successively. The time step Δt_j, number of subintervals m_j and the per subinterval n_j were chosen as:

$$\Delta t_j = \frac{24}{2^{(j-1)}}, \tag{15a}$$

$$m_j = 3 \cdot 2^{(j-1)}, \tag{15b}$$

$$n_j = 5, \tag{15c}$$

where the time step Δt_j is in days and $j\in\{1, 3\}$ corresponds with each of the reconstructions. For the case $j=0$ only the time-mean state is assimilated. Using this setup we will only assimilate the first 360 days of our set of observations.

For the optimality tolerance ϵ_m the value of 10^{-3} has been chosen. This ensures that the solutions are accurate and calculated efficiently: a smaller value does not give significantly more accurate solutions (in this case), but does significantly increase the computational cost. The covariance matrices have been chosen as the identity matrix, i.e. $\mathbf{B} = \mathbf{R}_i = \mathbf{I}$. For the moment all the components of the wind-stress forcing are included for every reconstruction, i.e. for each case j, $\tau_i=1$ for all i. The initial background will always be taken as the jet-down steady-state solution.

For each value of j, the basin integrated kinetic energy of the background and the analysis is shown in Fig. 4. In each panel, the background trajectory is shown as the thin solid curve, the analysis trajectory as the thick solid curve and the observations as the dashed curve. In Fig. 4a, only the observational time-mean was assimilated with a time step of 24 days ($j=0$). From this figure it is clear that the analysis trajectory is far from the observations; it stays close to the time-mean state. For $j=1$ (Fig. 4b) the low-frequency variability is assimilated with a time step of $\Delta t=12$ days. Here the method finds an analysis which is closer to the observation than the background. For the reconstruction based on the low-to-mid frequencies (Fig. 4c; $\Delta t=6$ days) and all the

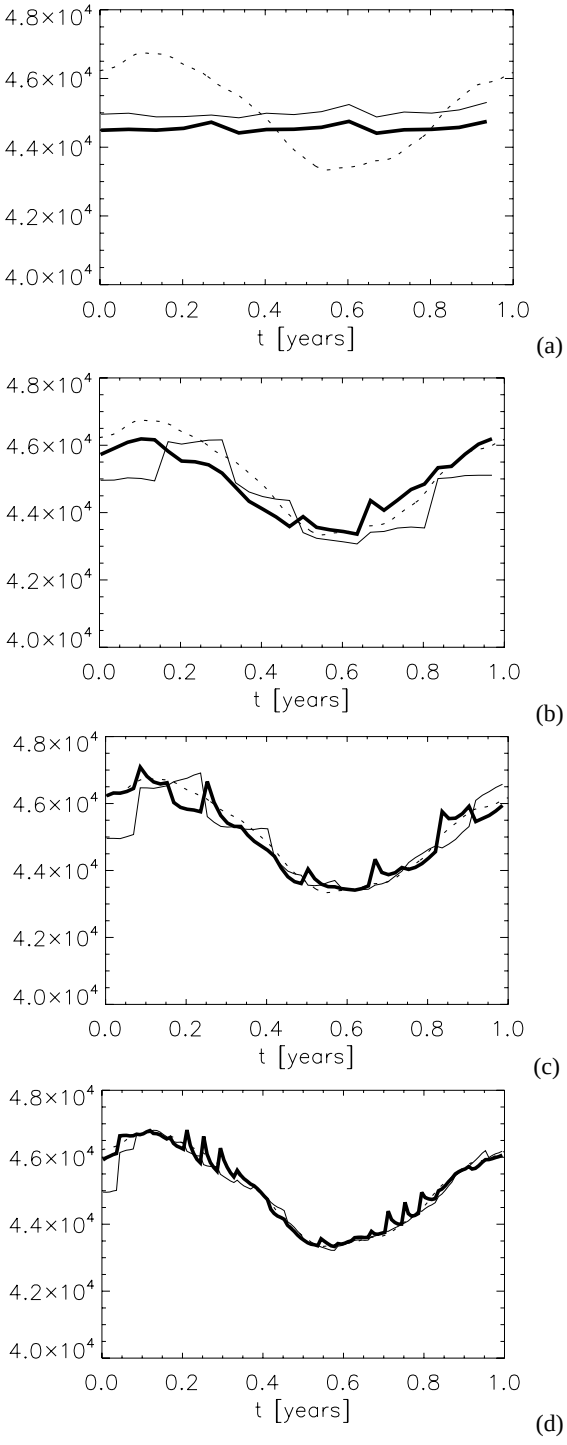

Fig. 4. Basin integrated kinetic energy E_{kin} of the analyses (thick and solid) and background (thin and solid) plotted against the kinetic energy of the unreconstructed observations for several values of j: **(a)** $j=0$, assimilation of only the time-mean observations with $\Delta t=24$ days; **(b)** $j=1$, assimilation of the low frequency variability with $\Delta t=12$ days; **(c)** $j=2$, assimilation of the lower en mid-range frequency variability with $\Delta t=6$ days; and **(d)** $j=3$, assimilation of all the leading modes for $\Delta t=3$ days.

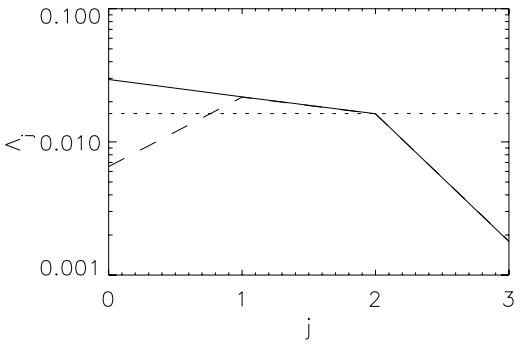

Fig. 5. Difference between the time-mean state of the analysis and the observational time-mean $\mathcal{A}_j$, as defined by Eq. (16). The solid curve represents the case where all the wind-stress forcing components were used for each j, the dashed curve is the case in which there is a matching of modes and wind-stress forcing, and the dotted curve shows the result for the explicit model.

modes (Fig. 4d; Δt=3 days) the analysis and the background move closer to the observations, and the assimilation of these observations improves the quality of the analysis.

In Fig. 4b–d some jumps in the curves for the background and analysis can be seen. These jumps occur between the subintervals (see Fig. 1) and become smaller for larger j (smaller Δt) and inclusion of higher frequency modes. Although the method is able to find an analysis close to the reconstructed observations, the model cannot exactly fit these observations. As a result, the analysis trajectory will be sufficiently close to the observations over the interval, but it will start to deviate from the observations in the trailing interval. As this part of the trajectory is used as the background for the next interval, these errors will introduce the jumps in the background and analysis. For smaller Δt the jumps become less prominent.

In Fig. 5 the time-mean of the analyses is compared with the "true" time-mean state, the difference defined by:

$$\mathcal{A}_j = \|\boldsymbol{\psi}^j_{\text{est}} - \boldsymbol{\psi}_{\text{true}}\|_2, \tag{16}$$

where $\boldsymbol{\psi}^j_{\text{est}}$ is the estimate of the time-mean calculated from the available analyses at iterate j, and $\boldsymbol{\psi}_{\text{true}}$ the time-mean of the observations (shown in Fig. 3c). The solid curve in Fig 5 shows $\mathcal{A}_j$ for the cases presented in Fig. 4. Clearly for smaller Δt and larger j, the time-mean state of the analysis converges to the observational time-mean state. This is expected, since we use unperturbed observations and use all the available observations. The result also indicates that only a few statistical modes need to be included here for a reasonably accurate estimate of the time-mean state.

To look at the effect of the wind-stress forcing field, the same calculations were performed as above, but now the wind forcing applied to the model is also varying with j. Only those components of the wind stress that match with the frequencies of the modes present in the reconstruction of the observations are included, i.e., for j=0 the time-mean observations are assimilated with the steady wind stress (τ_i=0 for all i). For j=1, the low frequencies are included in the reconstruction of the observations and only the low frequency component (τ_1=1, others zero) is included in the wind forcing. The resulting value of A_j is plotted as the dashed curve in Fig. 5. For j=0, the estimated time-mean state is better than calculated with all the wind-stress components. For $j>0$ the differences become very small because the annual variability is the dominant mode of variability in the observations.

3.2 Internal variability

For Re=80 and a steady wind-stress (τ_i=0 for all i), a model trajectory was calculated with Δt=3 h starting from the unstable jet-up steady state. In Fig. 6a–b, the dimensionless basin integrated kinetic energy and the asymmetry of the streamfunction $\Delta\Psi$ of this trajectory are shown. For the first four years, the trajectory stays very close to the unstable steady state but it becomes quasi-periodic over the next 8–9 years. During the last nine years, two types of variability can be seen, one with a period of about 5 years and the second with a period of about 50 days having a smaller amplitude as the first. The trajectory circles around the unstable jet-up steady state but has a slightly lower mean kinetic energy. The time-mean streamfunction over the last five years is shown in Fig. 6c. The time-mean state has a jet-up structure and differs substantially from the unstable jet-up solution (Fig. 6d). The values of the streamfunction on all the grid point over the last five years of the integration were taken as the observations.

After subtraction of the time-mean, the 20 leading principle components of the observations were calculated. As in the previous section 100% of the variance could be explained, since we use perfect observations. A total of 20 M-SSA modes were calculated from the principle components. Of these 20 modes two pairs were dominant; the first pair with a period of 100 days, the second pair with a period of 50 days. The other modes have smaller periods and amplitudes. These modes were divided into frequency bands accordingly.

The same assimilation setup as in the previous subsection was used, i.e., the time steps Δt_j, the number of subintervals m_j and the points per subinterval n_j are given by Eq. (15). The initial background was taken as the unstable jet-up steady state and only the first year of the five year period is assimilated. The trajectories of the background and analysis for *all* values of j are shown in Fig. 7. For reference, the observations are also shown in the same figure as the dashed curve.

When assimilating only the time-mean observations (j=0) with a time step of 24 days, the initial background is far from the observations (Fig. 7a). The assimilation, however, finds an analysis that is much closer to the observations. This is also seen for other values of j. After the first interval both

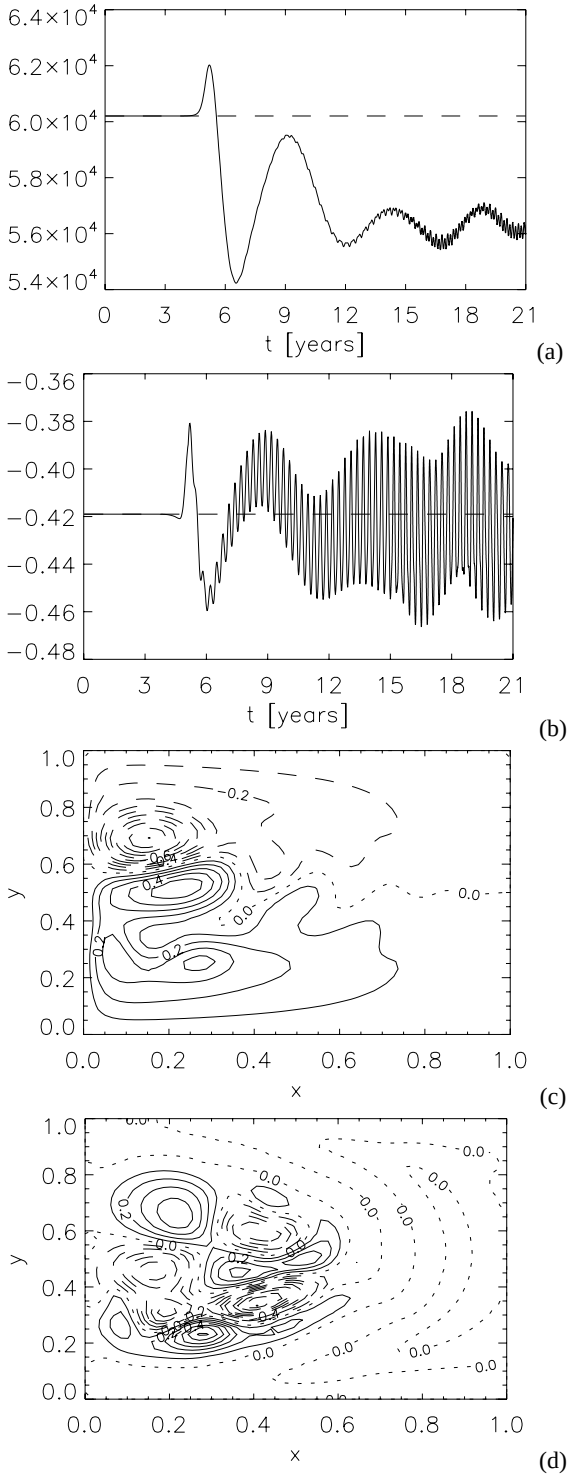

Fig. 6. Overview of the model run at Re=80 and τ_i=0 for all i: **(a)** the basin integrated kinetic energy E_{kin}; **(b)** the asymmetry of the streamfunction ψ; **(c)** the mean of the streamfunction over the last five years, the contours are with respect to an absolute maximum of ψ=2.5; and **(d)** the difference of this mean with respect to the unstable jet-up steady-state at Re=80, the contours are with respect to an absolute maximum of ψ=0.45.

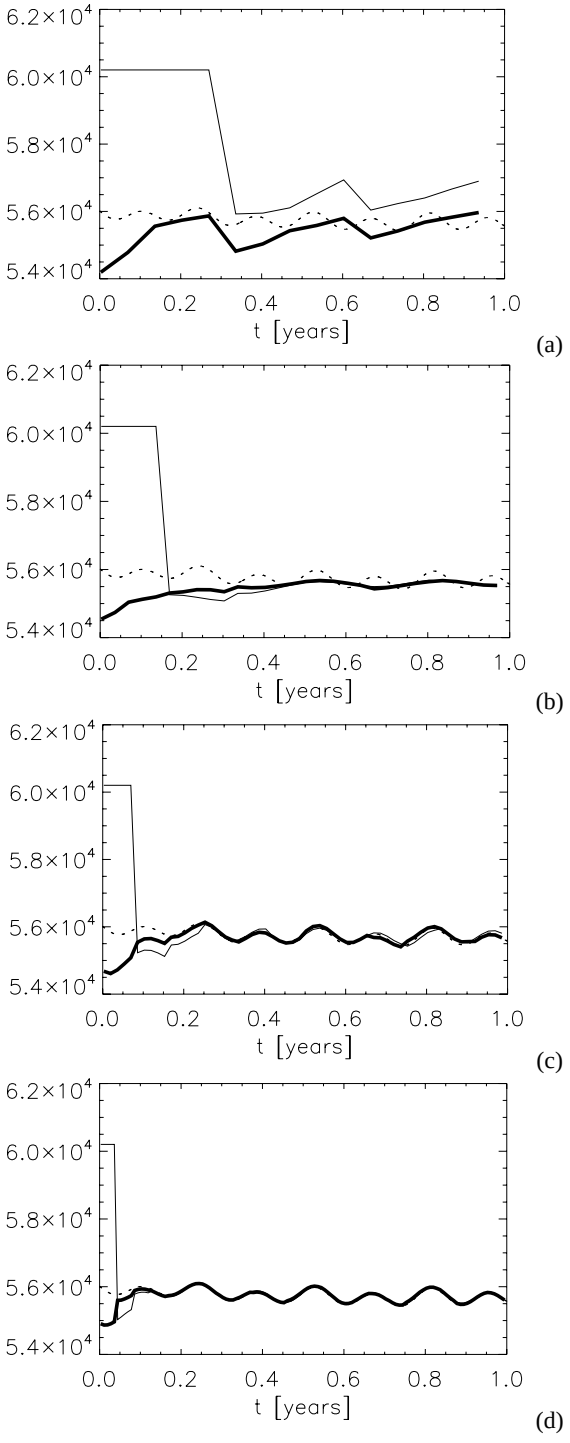

Fig. 7. Basin integrated kinetic energy of the analyses (thick and solid) and background (thin and solid) plotted against the kinetic energy of the unreconstructed observations for several values of j: **(a)** j=0, assimilation of only the time-mean observations with Δt=24 days; **(b)** j=1, assimilation of the low frequency variability with Δt=12 days; **(c)** j=2, assimilation of the lower en mid-range frequency variability with Δt=6 days; and **(d)** j=3, assimilation of all the leading modes for Δt=3 days.

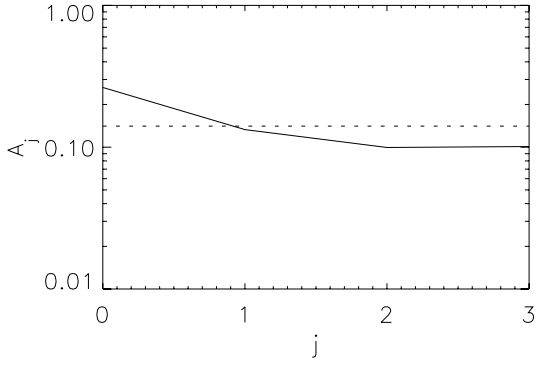

Fig. 8. Difference between the time-mean state of the analysis (drawn curve) and the observational time-mean, $\mathcal{A}_j$ as defined by Eq. (16). The dotted curve is the result from the computation with the explicit model.

the background and analysis are close to the observations but they do not show the 50 day period, when compared with the observations. Increasing the value of j, i.e. taking a smaller time step and taking more statistical modes into account, leads to a smaller difference with the observations. For j=1 (Fig. 7b) both trajectories fit the observations and show a small 100 day oscillation. For j=2 (Fig. 7c) and j=3 (Fig. 7d) the analysis and background also show the 50 day period.

The estimate of the time-mean state as calculated from the analyses is compared with the time-mean state from the observations in Fig. 8. For smaller time steps and more modes present in the reconstruction, the estimate becomes more accurate. For j=2 and j=3 the differences with the observational time-mean state are practically the same. Hence taking a time step of 3 days and taking a reconstruction based on 20 M-SSA modes does not give an improvement when compared to the result for a time step of 6 days and only the dominant two pairs of M-SSA modes. For an accurate estimate of the time-mean state it is sufficient to take only the dominant statistical modes in the observations into account.

3.3 Comparison with the estimate from the explicit model

To see how much advantage there is from the implicit methodology, we compare the results above with an estimate of the time-mean state calculated over an analysis trajectory resulting from a data-assimilation with the explicit model using the second order Adams-Bashforth scheme. We will compare both the accuracy of the estimated time-mean state and the computational cost. For both the externally induced variability and the internal variability the model generated observations are assimilated using the explicit model using the following setup: 8 points per subinterval and $\mathbf{B}=\mathbf{R}_i=\mathbf{I}$ and a time step of Δt=3 h.

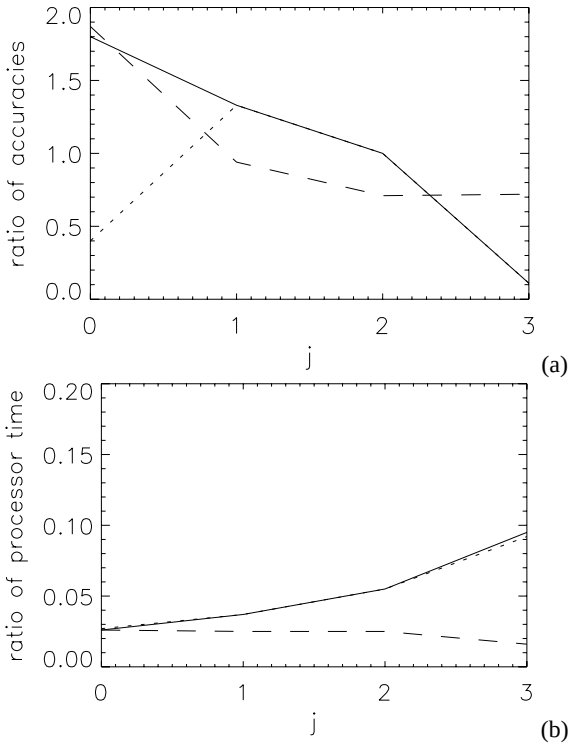

Fig. 9. (a) The ratio of the difference between estimated time-mean state using the implicit method and the true time-mean state and the same difference as calculated by the explicit method; **(b)** the ratio of the processor times used for the implicit estimation and for the explicit estimation. In both figures the results for the external variability with constant forcing (τ_i=1 for all i and j) are drawn solid, for the external variability with variable forcing (values of τ_i variable with j) are drawn dotted and for the internal variability the curves are drawn dashed.

First consider the externally induced variability. With all the time-dependent components in the wind-stress ($\tau_i \neq 0$), the 4D-Var analysis has been calculated with the explicit model. From this analysis, the time-mean state is calculated and the difference with the time-mean state from the observations is shown as the dotted curve in Fig. 5. The explicit assimilation is able to find a better estimate for the time-mean state when compared with the implicit estimation method for j=0, 1 and j=2. For j=3, i.e., for Δt=3 days in the implicit method and observations reconstructed with 20 M-SSA modes, the implicit estimation method performs better than the explicit estimation method. For the case of internal variability, the difference between the estimate of the time-mean calculated using the explicit model and true time-mean state is shown as the dotted curve in Fig. 8. Again, the implicit method is more accurate when both dominant pairs of modes (j=2, 3) are used in the reconstruction of the observations.

The implicit estimation and explicit estimation methods are compared both on accuracy and efficiency in Fig. 9. In

Fig. 9a, the difference between estimated time-mean state using the implicit method and the true time-mean state has been divided by the same difference as calculated by the explicit method. If this ratio is larger (smaller) than 1, then the explicit estimation method is more (less) accurate. The different curves represent again the different cases, with drawn the external variability with constant forcing having a solid linestyle, external variability with variable forcing a dotted linestyle and the dashed curve represents the case of internal variability. For a time step of 6 days or smaller and with only the dominant statistical modes present in the reconstruction, the implicit estimation method performs better than the direct approach using an explicit model. In Fig. 9b, the ratio of the processor times used for the implicit estimation and for the explicit estimation are shown and clearly the explicit approach is up to a factor 40 more expensive. The implicit estimation approach is much cheaper because a larger time step is used, reducing the number of subintervals and the number of minimizations needed to calculate the analysis.

4 Summary and conclusion

A new method for the estimation of a time-mean state consistent with given observations was presented. The main idea of the estimation method is to fully utilize the advantage of taking a large time step in the implicit model by only assimilating, in addition to the time-mean observations, those statistical modes which explain most of the variance seen in the observational data. These modes are separated into several frequency bands, ordered from low to higher frequencies. For large time-steps only the modes in the low frequency band are assimilated, while for smaller time-steps higher frequencies can be included. The statistical modes of variability are calculated from the observations using the M-SSA technique. Assimilation of the reconstructed observations based on a few low-frequency statistical modes circumvents the high computational cost associated with small time steps needed for fitting high frequency variations in the observations as has to be done with an explicit model. Furthermore, in the new estimation method a high temporal resolution of the data is not required.

The performance of the implicit estimation method was tested using a barotropic quasi-geostrophic model of the wind-driven double-gyre ocean circulation. Two test-cases were considered: externally variability induced by a time-dependent wind stress and internal variability induced through the occurrence of internal instabilities. The implicit estimation method has a comparable accuracy compared to that of the explicit estimation method. Depending on which modes are used for the reconstruction of the observations and the size of the time step, the implicit estimation method is more accurate. Furthermore, this method is a factor 10–40 cheaper in CPU time when compared with the explicit method.

We admit that the model problem used here, with the idealized observations and identity covariance matrices, is relatively simple when compared to the estimation of a time-mean state within a sophisticated ocean model and realistic observational data. The application of this estimation method to realistic problems is dependent on the quality and quantity of the observational data and the availability of implicit ocean general circulation models. The results presented here are, however, motivating to further develop implicit ocean models and corresponding assimilation and estimation methods.

Acknowledgements. This work was supported by the Dutch Technology Foundation (STW) within the project GWI.5798.

Edited by: J. Kurths
Reviewed by: one anonymous referee

References

Courtier, P., Thépaut, F.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteor. Soc., 120, 1367–1388, 1994.

Dijkstra, H. A. and Katsman, C. A.: Temporal variability of the Wind-Driven Quasi-geostrophic Double Gyre Ocean Circulation: Basic Bifurcation Diagrams, Geophys. Astrophys. Fluid Dyn., 85, 195–232, 1997.

Ferreira, D., Marshall, J., and Heimbach, P.: Estimating Eddy Stresses by Fitting Dynamics to Observations Using a Residual-Mean Ocean Circulation Model and Its Adjoint, J. Phys. Ocean., 35, 1891–1910, 2005.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, Rev. Geophys., 40, 3.1–3.41, doi:10.1029/2000GR000092, 2002.

Giering, R. and Kaminski, T.: Recipes for adjoint code construction, ACM Trans. Math. Softw., 24, 437–474, 1998.

Pedlosky, J.: Geophysical Fluid Dynamics. 2nd Edn, Springer-Verlag, New York, 710 pp., 1987.

Peyret, R. and Taylor, T.: Computational Methods for Fluid Flow, Springer-Verlag, Berlin, 358 pp., 1983.

Simonnet, E., Ghil, M., and Dijkstra, H.: Homoclinic bifurications in the quasi-geostrophic double-gyre circulation, J. Mar. Res., 63, 931–956, 2005.

Stammer, D.: Adjusting internal model errors through ocean state estimation, J. Phys. Oceanogr., 108, 1143–1153, 2005.

Stammer, D., Wunsch, C. Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., and Marshall, J.: Volume, heat and freshwater transports of the global ocean circulation 1993–2000, estimated from a general circulation model constrained by WOCE data, J. Geophys. Res., 108, C05 023, doi:10.1029/2001JC001 115, 2003.

Terwisscha van Scheltinga, A. and Dijkstra, H.: Nonlinear data-assimilation using implicit models, Nonlin. Processes Geophys., 12, 515–525, 2005, http://www.nonlin-processes-geophys.net/12/515/2005/.

Terwisscha van Scheltinga, A. and Dijkstra, H.: A comparison of the performance of 4D-Var in an explicit and implicit version of a nonlinear barotropic ocean model, Nonlin. Processes Geophys., 14, 763–776, 2007.
http://www.nonlin-processes-geophys.net/14/763/2007/

Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series, Physica D, 35, 395–424, 1989.

Vialard, J., Weaver, A., Anderson, D., and Delecluse, P.: Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific. Part II: physical validation, Mon. Weather Rev., 131, 1379–1395, 2003.