Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2025-032
J-STAGE Advance published date: July 10, 2025

Four-dimensional Variational Data Assimilation Using the Second-order Incremental Approach and Quantum Annealing

Tadashi TSUYUKI [1]

Center for Environmental Remote Sensing, Chiba University, Chiba, Japan
Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

Fumitoshi KAWASAKI

Graduate School of Science and Engineering, Chiba University, Chiba, Japan

and

Shunji KOTSUKI

Center for Environmental Remote Sensing, Chiba University, Chiba, Japan
Institute for Advanced Academic Research, Chiba University, Chiba, Japan
Research Institute of Disaster Medicine, Chiba University, Chiba, Japan

Submitted: December 27, 2024
Revised: April 18, 2025
Re-revised: June 25, 2025

---

1) Corresponding author: Tadashi Tsuyuki, Center for Environmental Remote Sensing, Chiba University, 1-33, Yayoicho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan
Email: ttsuyuki@chiba-u.jp
Tel: +81-443-290-2965

Abstract

Four-dimensional variational data assimilation (4DVar) has been used as widely as ensemble Kalman filters (EnKFs) in meteorology and oceanography. Unlike EnKFs, 4DVar can be applied to strongly nonlinear regimes in data assimilation. A problem with 4DVar is that the cost function may have multiple minima, and that it can be difficult to find the global minimum using a gradient descent method. Quantum annealing can find the global minimum via quadratic unconstrained binary optimization (QUBO). This study proposes a method of searching for the global minimum of the 4DVar cost function by combining a second-order incremental approach and quantum annealing, in which the latter provides guidance on where to explore in state space by minimizing an approximated cost function. This approximated cost function is constructed in low-dimensional space by expanding state variables up to the second order around a basic state. If the global minimum cannot be reached after a couple of updates of the basic state, the 4DVar analysis is replaced by an EnKF analysis in assimilation cycles. Data assimilation experiments using the Lorenz-63 model were conducted as a proof of concept of the proposed method. The results revealed that the proposed method significantly reduced the frequency of falling into local minima, and that the benefit of extending the length of the assimilation window was realized even in strongly nonlinear

regimes. Data assimilation experiments in which simulated annealing was adopted instead of quantum annealing showed that quantum annealing exhibited comparable or better performance compared to simulated annealing.

Keywords four-dimensional variational data assimilation; second-order incremental approach; quantum annealing; simulated annealing; quadratic unconstrained binary optimization

1. Introduction

Data assimilation is a methodology of estimating the state of a dynamical system by assimilating observational data into a system model using an observation model. Since state variables of the system model cannot always be observed directly, the observation model is introduced to relate observational data to state variables. Four-dimensional variational data assimilation (4DVar, Le Dimet and Talagrand 1986; Talagrand and Courtier 1987) has been used as widely as ensemble Kalman filters (EnKFs, Evensen 1994) for operational forecasts and research in meteorology and oceanography. Unlike EnKFs, 4DVar can be applied to strongly nonlinear regimes in data assimilation. The analysis state of 4DVar is the mode of a posterior probability density function (PDF). Let the governing equation of a dynamical system be written as

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, t), \tag{1}$$

where $\boldsymbol{x}(t)$ is an n-dimensional vector consisting of state variables. The PDF of state variables $p(\boldsymbol{x}, t)$ evolves according to the Liouville equation (e.g., Ehrendorfer, 1994):

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial \boldsymbol{x}} \cdot (p\boldsymbol{F}) = 0\,. \tag{2}$$

Let us assume that the governing equation satisfies the following condition:

$$\frac{\partial}{\partial \boldsymbol{x}} \mathrm{tr}\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\right] = \boldsymbol{0}, \tag{3}$$

which means that the divergence of flow is uniform in the state space spanned by the state variables. This condition is satisfied by Hamiltonian dynamical systems, in which the divergence of flow vanishes according to Liouville's theorem (e.g., Goldstein et al. 2001). It

77 is also satisfied by the Lorenz-63 model (Lorenz 1963) and the Lorenz-96 model (Lorenz
78 1996), for example. Then we can prove that if the initial condition $\boldsymbol{x}(0)$ is the mode of
79 $p(\boldsymbol{x},0)$, $\boldsymbol{x}(t)$ remains the mode of $p(\boldsymbol{x},t)$. We can also prove that in assimilating
80 observational data 4DVar implicitly uses a non-Gaussian prior PDF that evolves according
81 to the Liouville equation from a Gaussian prior PDF given at the beginning of the assimilation
82 window (Tsuyuki 2014). The latter property is a nonlinear extension of one of the well-known
83 properties of 4DVar in linear Gaussian systems (e.g., Thépaut et al. 1993; Tsuyuki and
84 Miyoshi 2007).

85 If a numerical model of the dynamical system is perfect, then 4DVar analysis becomes
86 more accurate as the length of the assimilation window increases, because more
87 observational data are assimilated and the influence of the Gaussian assumption on the
88 prior PDF at the beginning of the assimilation window becomes weaker. However, a problem
89 with long-window 4DVar is that the cost function may have multiple secondary minima due
90 to strong nonlinearity, and it can be difficult to find the global minimum using a gradient
91 descent method. A similar difficulty arises when the system model or observation model is
92 strongly nonlinear. In 4DVar experiments with the Lorenz-63 model, Gauthier (1992) showed
93 that a significant secondary minimum could not be found in a regular regime of the model,
94 whereas this was not the case in a transition regime, and that the point of convergence was
95 highly dependent on the first guess, a starting point of gradient descent. Pires et al. (1996)
96 revealed theoretically that, in the limit of an infinitely long assimilation window, the landscape

of the cost function in state space is singular; it folds into deep, narrow valleys parallel to sheets of unstable manifolds in dynamical systems, and an unbounded number of secondary minima appear. These authors also proposed the quasi-static variational assimilation algorithm to determine the global minimum of 4DVar. In this algorithm, the analysis at the beginning of the assimilation window is updated with successive small increments of the length of the assimilation window; however, this algorithm is costly and cannot guarantee success with every attempt. On the other hand, Andersson et al. (2005) discussed issues involved in extending the 4D-Var algorithm to a longer assimilation window in the presence of nonlinearity, and concluded that a long-window weak-constraint 4D-Var has exciting prospects.

Quantum annealing is aimed at finding the ground state of a generic Ising model (Finnila et al. 1994; Kadowaki and Nishimori 1998; Farhi et al. 2001) through quantum tunneling; many combinatorial optimization problems have been shown to reduce to this problem. Since the release of the quantum annealer 2000Q from D-Wave Systems in 2011 (Johnson et al. 2011), quantum annealing research has progressed rapidly, with applications in graph partitioning (Ushijima-Mwesigwa et al. 2017), clustering (O'Malley et al. 2018), machine learning (Willsch et al. 2020), and model predictive control (Inoue et al. 2020). Kotsuki et al. (2024) proposed quantum data assimilation, a novel data assimilation strategy that solves the 4DVar optimization problem using quantum annealing. These authors reported a significant reduction of execution time with comparable accuracy to a gradient descent

117 method.

118 This study proposes a method of searching for the global minimum in 4DVar in strongly
119 nonlinear regimes through combining a second-order incremental approach and quantum
120 annealing, in which quantum annealing provides guidance on where to explore in state
121 space by minimizing an approximated cost function. This approximated cost function is
122 constructed in low-dimensional space by expanding state variables up to the second order
123 around a basic state. If the global minimum cannot be reached after a couple of updates of
124 the basic state, then 4DVar analysis is replaced by an EnKF analysis. Data assimilation
125 experiments using the Lorenz 63 model were conducted as a proof of concept. Additionally,
126 a comparative analysis was conducted in which quantum annealing was replaced by
127 simulated annealing (Kirkpatrick et al. 1983) in the proposed mehod.

128 The remainder of this article is organized as follows. Section 2 introduces the proposed
129 method of searching for the global minimum in 4DVar, including approximation methods in
130 low-dimensional space and the physical concept of quantum annealing. Section 3 describes
131 the design of data assimilation experiments using the Lorenz-63 model, and Section 4
132 presents the experimental results. A summary and discussion are provided in Section 5.

133

134 **2. Methods**

135 *2.1 Overview*

136 The basic concept of the proposed method of searching for the global minimum of the

4DVar cost function is that when a gradient descent method fails to reach the global minimum, quantum annealing can provide guidance on where to explore in state space. Figure 1 shows a cost function plotted against a control variable in one-dimensional space. The control variable is a deviation from the background state, which is the predicted state from the latest analysis. The cost function is assumed to have two minimum points and the background state is taken as the starting point of the gradient descent method. In Fig. 1, the gradient descent method is directly applied to the original cost function, but the conventional incremental 4DVar (Courtier et al 1994) can also be used to search for a minimum. Although the latter approach is usually adopted in operational 4DVar systems, we do not use this approach in this study to avoid confusion.

Fig. 1

Quantum annealing is a method of quadratic unconstrained binary optimization (QUBO); therefore, we need to approximate the cost function using a polynomial for the control variables and to transform higher than second-order terms into linear or quadratic terms using a property of binary variables. We adopt the background state as a basic state and approximate the cost function around the basic state using a quartic polynomial, which is the lowest-order polynomial with multiple minima. For this purpose, we extend the conventional incremental 4DVar to the second-order incremental approach. Because quantum annealing is a discrete-variable optimization method, he control variable for quantum annealing is assumed to take integer values only. Then, quantum annealing is used to find the global minimum of this approximated cost function (dashed line) in discrete control

space. Although this minimum point is generally different from the global minimum of the approximated cost function, it may lie in the domain of attraction of a possible global minimum of the original cost function, which can be found using the gradient descent method starting from the minimum point obtained by quantum annealing. Thus, the minimum point obtained by quantum annealing plays a guiding role in determining which region of the control space to explore.

If the minimum point thus obtained is not the global minimum of the original cost function, this process can be repeated by updating the basic state around which the cost function is approximated. We adopt the minimum point obtained by quantum annealing as the updated basic state. The process of updating the basic state and approximating the cost function around this basic state is conducted as an outer loop of minimization of the cost function. Note that this is a local search method exploring a neighborhood of the background state. As long as the background state is close to the true state, we can expect that this method will work. Thus, the proposed method is expected to broaden the applicability of 4DVar in strongly nonlinear data assimilation, although it depends on the future development of quantum annealers.

Figure 2a shows the workflow of the proposed method. Hybrid 4DVar (e.g., Bonavita et al. 2016) is adopted as a basic data assimilation method, in which the background error covariance matrix is provided by an EnKF to make the background error covariance flow-dependent. The outer loop in the figure represents the iterative computation described in the

Fig. 2

177 previous paragraph. If the global minimum cannot be reached after a certain number of outer
178 loop iterations, then an EnKF analysis is adopted as an output. When the cost function has
179 multiple minima, it is generally difficult to determine whether the global minimum has been
180 reached. In this study, when the convergence value of the cost function J_* exceeds a given
181 threshold value J_c, the gradient descent method is assumed to have failed to reach the
182 global minimum. The procedure of replacing the 4DVar analysis by an EnKF analysis when
183 $J_* > J_c$ is hereafter referred to as "EnKF replacement". The analysis error variance of EnKFs
184 is smaller than forecast error variance; therefore, EnKF replacement almost always
185 improves analysis accuracy, as will be shown in Section 4.1. This procedure can also be
186 implemented in hybrid 4DVar. Figure 2b shows the workflow of hybrid 4DVar with EnKF
187 replacement. We compared the two methods in Fig. 2 to evaluate the performance of the
188 proposed method.
189
190 *2.2 Second-order incremental approach*
191 *a. Cost function*
192 The 4DVar analysis is obtained by minimizing a cost function that measures differences
193 between the background state and observational data (e.g., Kalnay 2003). Let $\boldsymbol{x}_0$, $\boldsymbol{x}_0^b$, and
194 $\boldsymbol{B}$ denote the state variables, background state, and background error covariance matrix,
195 respectively, at the beginning of the assimilation window of 4DVar at time t_0. Let $\boldsymbol{x}_k$, $\boldsymbol{y}_k^o$,
196 and $\boldsymbol{R}_k$ denote the state variables, observations, and observation error covariance matrix,

197 respectively, at time t_k $(k = 1, \cdots, K)$ in the assimilation window. The time interval between
198 observations need not to be equal, such that missing data are allowed. Background error
199 and observation error are assumed to have Gaussian distributions, and the observation
200 model is assumed to be linear for simplicity. Then the cost function of 4DVar is written as

201 $$J(\boldsymbol{x}_0) = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^{\mathrm{T}} \boldsymbol{B}^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{y}_k^o)^{\mathrm{T}}\boldsymbol{R}_k{}^{-1}(\boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{y}_k^o), \quad (4)$$

202 where the superscript T denotes the transpose of a vector or matrix, $\boldsymbol{H}_k$ is the linear
203 observation operator at time t_k, and $\boldsymbol{x}_k$ is obtained by integrating Eq. (1) with the initial
204 condition $\boldsymbol{x}_0$. The time steps used for this integration and the time between observations
205 generally differ. The analysis at the beginning of the assimilation window, $\boldsymbol{x}_0^a$, is obtained as
206 the minimum point of $J(\boldsymbol{x}_0)$, and the analysis at the end of the assimilation window, $\boldsymbol{x}_K^a$, is
207 computed by integrating Eq. (1) with the initial condition $\boldsymbol{x}_0^a$. In assimilation cycles of 4DVar,
208 the latter analysis is used as the background state at the beginning of the next assimilation
209 window.

210 We transform the cost function Eq. (4) to an incremental form to introduce the quartic
211 polynomial approximation. The cost function of incremental 4DVar for the lth outer loop, of
212 which the basic state is denoted by $\boldsymbol{x}_k^{(l)}$, is given by

213 $$J^{(l)}\left(\delta\boldsymbol{x}_0^{(l)}\right) = \frac{1}{2}\left(\delta\boldsymbol{x}_0^{(l)} + \Delta\boldsymbol{x}_0^{(l)}\right)^{\mathrm{T}}\boldsymbol{B}^{-1}\left(\delta\boldsymbol{x}_0^{(l)} + \Delta\boldsymbol{x}_0^{(l)}\right)$$

214 $$+ \frac{1}{2}\sum_{k=1}^{K}\left(\boldsymbol{H}_k\delta\boldsymbol{x}_k^{(l)} - \boldsymbol{d}_k^{(l)}\right)^{\mathrm{T}}\boldsymbol{R}_k{}^{-1}\left(\boldsymbol{H}_k\delta\boldsymbol{x}_k^{(l)} - \boldsymbol{d}_k^{(l)}\right), \quad (5)$$

215 where

$$\delta \boldsymbol{x}_k^{(l)} \coloneqq \boldsymbol{x}_k - \boldsymbol{x}_k^{(l)}, \quad \Delta \boldsymbol{x}_0^{(l)} \coloneqq \boldsymbol{x}_0^{(l)} - \boldsymbol{x}_0^b, \quad \boldsymbol{d}_k^{(l)} \coloneqq \boldsymbol{y}_k^o - \boldsymbol{H}_k \boldsymbol{x}_k^{(l)}. \tag{6}$$

The basic state of the first outer loop at time t_0 is $\boldsymbol{x}_0^b$, and those of the other outer loops are set to the minimum points obtained by the last quantum annealing. The basic state $\boldsymbol{x}_k^{(l)}$ is obtained by integrating Eq. (1) with the initial condition $\boldsymbol{x}_0^{(l)}$. In the conventional incremental approach, a low-resolution model is used to minimize Eq. (5) in which $\delta \boldsymbol{x}_k^{(l)}$ is approximated by a linear function of $\delta \boldsymbol{x}_0^{(l)}$, and the basic state of the next outer loop $\boldsymbol{x}_0^{(l+1)}$ is obtained by adding the convergence value of $\delta \boldsymbol{x}_0^{(l)}$ to $\boldsymbol{x}_0^{(l)}$ in the original resolution. In the remainder of Section 2, Eq. (5) is regarded as the cost function of a low-resolution version of Eq. (1), and n denotes the number of state variables in the low-dimensional model.

We introduce an n-dimensional control variable $\boldsymbol{u}$ defined by

$$\delta \boldsymbol{x}_0^{(l)} = \boldsymbol{L}\boldsymbol{u}, \tag{7}$$

where $\boldsymbol{L}$ is the Cholesky decomposition matrix of $\boldsymbol{B}$ satisfying $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$. This is a standard pre-conditioning method for accelerating the convergence of 4DVar. Note that we can apply the eigenvalue decomposition to $\boldsymbol{B}$ instead of the Cholesky decomposition to introduce the control variable. In this case, we can reduce the dimension of $\boldsymbol{u}$ by neglecting eigenvectors with small eigenvalues as follows:

$$\delta \boldsymbol{x}_0^{(l)} \approx \boldsymbol{V}_N \boldsymbol{u} \coloneqq \left(\sqrt{\lambda_1}\boldsymbol{v}_1, \cdots, \sqrt{\lambda_N}\boldsymbol{v}_N\right)\boldsymbol{u}, \qquad (N < n) \tag{8}$$

where λ_i is the ith eigenvalue of $\boldsymbol{B}$ in descending order and $\boldsymbol{v}_i$ is the corresponding normalized eigenvector. The N largest eigenvalues and corresponding eigenvectors of $\boldsymbol{B}$

can be computed by using the Lanczos method (e.g., Golub and Van Loan 2012), but this approach is more expensive than the Cholesky decomposition.

In linear Gaussian systems, the convergence value of the cost function J_* multiplied by 2 follows a χ^2 distribution with degrees of freedom equal to the number of assimilated observations m (e.g., Michel 2014). Thus, the mean and variance of J_* are both equal to $m/2$. We set the threshold J_c to the upper 0.01% point of the χ^2 distribution with degrees of freedom m, based on results of preliminary data assimilation experiments using hybrid 4DVar. Although the these properties do not hold in nonlinear systems, this threshold value is useful for determining whether the minimization in 4DVar fails to reach the global minimum, as will be shown in Section 4.1.

In the conventional incremental approach, the increment of the lth outer loop $\delta\boldsymbol{x}_k^{(l)}$ is approximated by a linear function of $\delta\boldsymbol{x}_0^{(l)}$, whereas in this study it is approximated by a quadratic function of $\delta\boldsymbol{x}_0^{(l)}$:

$$\delta\boldsymbol{x}_k^{(l)} \approx \boldsymbol{M}_k^{(l)}\delta\boldsymbol{x}_0^{(l)} + \frac{1}{2}\left(\delta\boldsymbol{x}_0^{(l)}\right)^{\mathrm{T}} \begin{pmatrix} \boldsymbol{N}_{k\,1}^{(l)} \\ \vdots \\ \boldsymbol{N}_{k\,n}^{(l)} \end{pmatrix} \delta\boldsymbol{x}_0^{(l)}, \tag{9}$$

where $\boldsymbol{M}_k^{(l)}$ is the Jacobian matrix evaluated at $\boldsymbol{x}_0^{(l)}$, and $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ are the coefficient matrices of the second-order terms of the Tayler expansion of $\delta\boldsymbol{x}_k^{(l)}$ in $\delta\boldsymbol{x}_0^{(l)}$. Substitution of Eq. (9) into Eq. (5) yields a quartic polynomial approximation of the cost function, which may have multiple minima. Note that $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ are symmetric, and that $\boldsymbol{M}_k^{(l)}$ and $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ may be sparse in a high-dimensional system, because the contributions of initial conditions at

sufficiently distant grid points are negligible. These matrices can be computed approximately in state space or ensemble space.

b. Approximation in state space

The matrices $\boldsymbol{M}_k^{(l)}$ and $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ can be obtained using perturbation equations of Eq. (1). Let $\delta\boldsymbol{x}(t)$ be the deviation of $\boldsymbol{x}(t)$ from the basic state $\overline{\boldsymbol{x}}(t)$ and be expanded in perturbation variables as

$$\delta\boldsymbol{x}(t) \approx \delta\boldsymbol{x}^1(t) + \delta\boldsymbol{x}^2(t) + \cdots, \tag{10}$$

where $\delta\boldsymbol{x}^k(t)$ is the kth-order perturbation. Note that $\delta\boldsymbol{x}(t_k)$ and $\overline{\boldsymbol{x}}(t_k)$ correspond to $\delta\boldsymbol{x}_k^{(l)}$ and $\boldsymbol{x}_k^{(l)}$, respectively, in Section 2.2.a. The first two perturbation equations are given by

$$\frac{d}{dt}\delta\boldsymbol{x}^1 = \left.\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\right|_{\overline{\boldsymbol{x}}} \delta\boldsymbol{x}^1(t), \tag{11}$$

$$\frac{d}{dt}\delta\boldsymbol{x}^2 = \left.\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\right|_{\overline{\boldsymbol{x}}} \delta\boldsymbol{x}^2(t) + \frac{1}{2}(\delta\boldsymbol{x}^1(t))^{\mathrm{T}} \left.\frac{\partial^2 \boldsymbol{F}}{\partial \boldsymbol{x}\partial \boldsymbol{x}}\right|_{\overline{\boldsymbol{x}}} \delta\boldsymbol{x}^1(t), \tag{12}$$

with the initial conditions of

$$\delta\boldsymbol{x}^1(0) = \delta\boldsymbol{x}(0), \qquad \delta\boldsymbol{x}^2(0) = \boldsymbol{0}. \tag{13}$$

The matrices $\boldsymbol{M}_k^{(l)}$ and $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ are computed by integrating Eqs. (11) and (12) starting from two sets of simple initial conditions. The first set consists of n unit vectors in n-dimensional space, and the second set consists of $n(n-1)/2$ sums of all pairs of the unit vectors. The first set of initial conditions yields $\boldsymbol{M}_k^{(l)}$ and the diagonal components of $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$, and the second set yields the sum of diagonal and off-diagonal components of

$\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$, from which we can compute the off-diagonal components of $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{n}$ using the symmetric property of the matrices.

Equations (11) and (12) must be integrated starting from $n(n+1)/2$ different initial conditions; therefore, this procedure appears unfeasible for a high-dimensional system. We can reduce the dimension of $\boldsymbol{u}$ by adopting Eq. (8) instead of Eq. (7). Substitution of Eq. (8) into Eq. (9) yields-

$$\delta\boldsymbol{x}_k^{(l)} \approx \left(\boldsymbol{M}_k^{(l)}\boldsymbol{V}_N\right)\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\begin{pmatrix}\boldsymbol{V}_N{}^{\mathrm{T}}\boldsymbol{N}_{k\,1}^{(l)}\boldsymbol{V}_N\\ \vdots\\ \boldsymbol{V}_N{}^{\mathrm{T}}\boldsymbol{N}_{k\,n}^{(l)}\boldsymbol{V}_N\end{pmatrix}\boldsymbol{u}, \tag{14}$$

where the matrices in the parentheses on the right-hand side of Eq. (14) are computed as described in the previous paragraph except that the set of simple initial conditions is given by the product of $\boldsymbol{V}_N$ and N unit vectors in N-dimensional space and by the product of $\boldsymbol{V}_N$ and $N(N-1)/2$ sums of all pairs of the unit vectors. The approximated cost function can be minimized by quantum annealing, and the resulting minimum point $\boldsymbol{u}_{\min}$ in control space can be converted into the minimum point in state space: $\boldsymbol{x}_{\min} = \overline{\boldsymbol{x}}_0 + \boldsymbol{V}_N\boldsymbol{u}_{\min}$. The latter minimum point may lie in the domain of attraction of a possible global minimum of the original cost function, and it can be found using a gradient descent method starting from $\boldsymbol{x}_{\min}$.

c. Approximation in ensemble space

A more feasible method for a high-dimensional system is to compute the matrices in the ensemble space of a low-resolution version of Eq. (1). Let N' be the ensemble size, and let

$\boldsymbol{X}_0^f$ be the $n \times N'$ matrix of forecast ensemble perturbations at initial time t_0, which is defined by

$$\boldsymbol{X}_0^f \coloneqq \left(\delta \boldsymbol{x}_0^{f(1)}, \ldots, \delta \boldsymbol{x}_0^{f(N')}\right), \tag{15}$$

where $\left\{\delta \boldsymbol{x}_0^{f(i)}\right\}_{i=1}^{i=N'}$ are the ensemble members of initial perturbations with respect to the basic state $\overline{\boldsymbol{x}}_0$. The superscript l, which denotes the outer-loop number, is omitted to prevent confusion. The analysis ensemble perturbations generated by an EnKF can be used to construct $\boldsymbol{X}_0^f$.

Let $M_t(\cdot)$ be the time evolution operator of the low-resolution version of Eq. (1) for time $t \geq t_0$. We can approximate the perturbations at time t using $\boldsymbol{X}_0^f$ as follows:

$$\delta \boldsymbol{x}(t) \coloneqq M_t(\boldsymbol{x}_0) - M_t(\overline{\boldsymbol{x}}_0) \approx M_t\left(\overline{\boldsymbol{x}}_0 + \boldsymbol{X}_0^f \boldsymbol{u}\right) - M_t(\overline{\boldsymbol{x}}_0), \tag{16}$$

where $\boldsymbol{u}$ is a vector of control variables in N'-dimensional ensemble space. The Taylor expansion yields

$$\delta \boldsymbol{x}(t) \approx \left(\frac{\partial M_t}{\partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \boldsymbol{X}_0^f\right)\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\left(\left(\boldsymbol{X}_0^f\right)^{\mathrm{T}} \frac{\partial^2 M_t}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \boldsymbol{X}_0^f\right)\boldsymbol{u}. \tag{17}$$

This equation corresponds to Eq. (9). Note that for $i, j = 1, \cdots, N'$,

$$\left[\frac{\partial M_t}{\partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \boldsymbol{X}_0^f\right]_i = \frac{\partial M_t}{\partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \delta \boldsymbol{x}_0^{f(i)}, \tag{18}$$

$$\left[\left(\boldsymbol{X}_0^f\right)^{\mathrm{T}} \frac{\partial^2 M_t}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \boldsymbol{X}_0^f\right]_{ij} = \left(\delta \boldsymbol{x}_0^{f(i)}\right)^{\mathrm{T}} \frac{\partial^2 M_t}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \delta \boldsymbol{x}_0^{f(j)}, \tag{19}$$

and

$$M_t\left(\overline{\boldsymbol{x}}_0 \pm \delta \boldsymbol{x}_0^{f(i)}\right) \approx M_t(\overline{\boldsymbol{x}}_0) \pm \frac{\partial M_k}{\partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \delta \boldsymbol{x}_0^{f(i)} + \frac{1}{2}\left(\delta \boldsymbol{x}_0^{f(i)}\right)^{\mathrm{T}} \frac{\partial^2 M_t}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0} \delta \boldsymbol{x}_0^{f(i)}, \tag{20}$$

$$M_t\left(\overline{\boldsymbol{x}}_0+\frac{\delta\boldsymbol{x}_0^{f(i)}+\delta\boldsymbol{x}_0^{f(j)}}{\sqrt{2}}\right)\approx M_t(\overline{\boldsymbol{x}}_0)+\frac{1}{\sqrt{2}}\frac{\partial M_t}{\partial\boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0}\delta\boldsymbol{x}_0^{f(i)}+\frac{1}{\sqrt{2}}\frac{\partial M_t}{\partial\boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0}\delta\boldsymbol{x}_0^{f(j)}$$

$$+\frac{1}{4}\left(\delta\boldsymbol{x}_0^{f(i)}\right)^{\mathrm{T}}\frac{\partial^2 M_t}{\partial\boldsymbol{x}\partial\boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0}\delta\boldsymbol{x}_0^{f(i)}+\frac{1}{2}\left(\delta\boldsymbol{x}_0^{f(i)}\right)^{\mathrm{T}}\frac{\partial^2 M_t}{\partial\boldsymbol{x}\partial\boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0}\delta\boldsymbol{x}_0^{f(j)}+\frac{1}{4}\left(\delta\boldsymbol{x}_0^{f(j)}\right)^{\mathrm{T}}\frac{\partial^2 M_t}{\partial\boldsymbol{x}\partial\boldsymbol{x}}\bigg|_{\overline{\boldsymbol{x}}_0}\delta\boldsymbol{x}_0^{f(j)},\tag{21}$$

where $i\neq j$ in Eq. (21). The factor $1/\sqrt{2}$ on the left-hand side of Eq. (21) is introduced to normalize the magnitude of perturbations; if perturbations have the same variance and are uncorrelated with each other, the variance of $\left(\delta\boldsymbol{x}_0^{f(i)}+\delta\boldsymbol{x}_0^{f(j)}\right)/\sqrt{2}$ is equal to that of $\delta\boldsymbol{x}_0^{f(i)}$. The left-hand sides of Eqs. (20) and (21) can be computed by ensemble prediction; solving these equations for the vectors in Eqs. (18) and (19) yields the expansion coefficients in Eq. (17).

An approximated cost function in ensemble space is obtained by substituting Eq. (17) into Eq. (5), changing from a continuous time framework to a discrete time framework. This cost function can be minimized by quantum annealing and the resulting minimum point $\boldsymbol{u}_{\min}$ can be converted into the minimum point in state space, as $\boldsymbol{x}_{\min}=\overline{\boldsymbol{x}}_0+\boldsymbol{X}_0^f\boldsymbol{u}_{\min}$. However, it is necessary to generate the $N'(N'+3)/2$-member ensemble given on the left-hand sides of Eqs. (20) and (21)); this may be infeasible in a high-dimensional system even if a low-dimensional numerical model is used. A feasible method would be to use an artificial intelligence (AI)-based surrogate model for ensemble prediction, which would be much faster than a physics-based numerical model, although there is still a problem of underestimation of ensemble spread (e.g., Selz and Craig 2023).

2.3 Quantum annealing

a. Physical concept

Quantum annealing is a combinatorial optimization method based on quantum dynamics of the Ising model. The Hamiltonian operator of this model of an L-spin system is given as

$$\widehat{H}_0 = -\sum_{j<k}^{L} J_{jk}\,\hat{\sigma}_j^z \hat{\sigma}_k^z - \sum_{j=1}^{L} h_j\,\hat{\sigma}_j^z, \tag{22}$$

where $\hat{\sigma}_j^z$ is the z component of the Pauli operator of the jth spin, J_{jk} is the interaction coefficient between the jth and kth spins, and h_j is the magnetic field acting on the jth spin in the z direction. The Pauli operator can be represented by the Pauli spin matrices. Let $|\uparrow\rangle$ and $|\downarrow\rangle$ represent the up and down spin state, respectively. The following equations hold.

$$\hat{\sigma}^z|\uparrow\rangle = |\uparrow\rangle, \qquad \hat{\sigma}^z|\downarrow\rangle = -|\downarrow\rangle\,. \tag{23}$$

We can obtain the ground state of Hamiltonian Eq. (22) by quantum annealing. The total Hamiltonian of quantum annealing is given as

$$\widehat{H}(t) \coloneqq A(t)\widehat{H}_0 + B(t)\sum_{j=1}^{L}\left(-\hat{\sigma}_j^x\right), \tag{24}$$

where $\hat{\sigma}_j^x$ is the transverse component of the Pauli operator of the jth spin, which satisfies

$$\hat{\sigma}^x|\pm\rangle = \pm|\pm\rangle, \qquad |\pm\rangle \coloneqq \frac{1}{\sqrt{2}}(|\uparrow\rangle \pm |\downarrow\rangle), \tag{25}$$

and $B(t)$ represents the transverse magnetic field. The negative sign in the second term of Eq. (24) ensures that $|+\rangle$ has a lower energy. $\widehat{H}(t)$ interpolates between $B(t_i)\sum_{j=1}^{L}\left(-\hat{\sigma}_j^x\right)$

at initial time t_i and $\widehat{H}_0$ at final time t_f; $A(t)$ is an increasing nonnegative function with $A(t_i) = 0$ and $A(t_f) = 1$; and $B(t)$ is a decreasing nonnegative function with $B(t_f) = 0$. The initial state at $t = t_i$ is set to the ground state of $B(t_i)\sum_{j=1}^{L}(-\hat{\sigma}_j^x)$, which consists of a superposition of up and down spins $|+\rangle_1|+\rangle_2 \cdots |+\rangle_L$, where the subscript denotes the index of transverse spins. If the change in $\widehat{H}(t)$ with t is sufficiently small, the spin state evolves adiabatically and arrives at the ground state of $\widehat{H}_0$ at $t = t_f$ according to the adiabatic theorem of quantum dynamics (Born and Fochs, 1928); a physical system remains in its instantaneous eigenstate if a given perturbation is acting on it slowly enough and if there is a gap between the eigenvalue and the remainder of the Hamiltonian spectrum. Rajak et al. (2023) described the basic physical concept behind quantum annealing, provided an overview of recent theoretical findings, and reported experimental developments pointing to issues that are still debated. The spin state can be mapped to a binary variable $b \in \{0, 1\}$ and the Hamiltonian operator can be mapped to the cost function of which the global minimum is sought. Therefore, quantum annealing is used to solve QUBO problems.

Simulated annealing is a combinatorial optimization method that represents a classical counterpart to quantum annealing. In this method, the Boltzmann distribution of the Hamiltonian H_0 is prepared at a sufficiently high temperature using the Monte Carlo method and the system is slowly annealed down to a temperature of zero. If annealing is sufficiently slow, we can expect the system to arrive at the ground state of H_0 with high probability. Thus, simulated annealing utilizes thermal fluctuations for optimization, which induces the

thermal jump across an energy barrier from one local minimum to another. By contrast, in quantum annealing, quantum tunneling induces an escape from a local minimum through an energy barrier. Therefore, if a local minimum in the landscape of H_0 is surrounded by tall, and thin energy barrier, quantum tunneling has an advantage over thermal fluctuations in overcoming this energy barrier, which explains why quantum annealing outperforms simulated annealing in a system with a rugged energy landscape.

b. Reduction to QUBO

When quantum annealing is applied to solve a continuous-variable optimization problem such as the minimization of Eq. (5), the problem must be reduced to a QUBO problem. Chancellor (2019) proposed the domain-wall encoding to transform a continuous-variable optimization problem into a QUBO problem; this encoding method was used by Koh and Nishimori (2022), Abel et al. (2022), and Arai et al. (2023) to investigate the performance of quantum annealing in continuous-variable optimization in one and two dimensions. However, it is not suitable for the 4DVar optimization problem, because computation of the cost function is highly expensive.

The first task is to encode control variables $\boldsymbol{u}$ into binary variables $\boldsymbol{b}$. Let Z denote the number of bits per real number. The following binary encoding is adopted in this study.

$$\boldsymbol{u} = r(2\boldsymbol{E}_Z\boldsymbol{b} - \mathbf{1}_n), \tag{26}$$

where $\boldsymbol{E}_Z$ is an $n \times nZ$ matrix defined by

$$\boldsymbol{E}_Z \coloneqq \boldsymbol{I}_n \otimes \boldsymbol{e}_Z^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{e}_Z^{\mathrm{T}} & 0 & \cdots & 0 \\ 0 & \boldsymbol{e}_Z^{\mathrm{T}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{e}_Z^{\mathrm{T}} \end{pmatrix}, \qquad \boldsymbol{e}_Z^{\mathrm{T}} \coloneqq \left(\frac{1}{2}, \frac{1}{2^2}, \cdots, \frac{1}{2^Z}\right), \tag{27}$$

The operator $\otimes$ is the Kronecker product, and $\mathbf{1}_n$ and $\boldsymbol{I}_n$ are the n-dimensional vector of which the components are all unity and the n-dimensional identity matrix, respectively. A similar binary coding was adopted by Inoue et al. (2020) and Kotsuki et al. (2024); however, we do not need a large Z value, because the minimum point obtained by quantum annealing is not used for the analysis itself, but solely to guide the selection of a control space region to explore. Because each component u_i given by Eq. (26) satisfies $-r \leq u_i < r$, the parameter r represents the search range of the control variables. Note that when the Z value is fixed, the interval of neighboring values of the control variables increases with r. Substituting Eq. (26) into Eqs. (7) and (9), Eqs. (8) and (14), or Eq. (17) makes the cost function Eq. (5) a quartic polynomial in $\boldsymbol{b}$, which is hereafter denoted by $J_2(\boldsymbol{b})$.

Next, we transform $J_2(\boldsymbol{b})$ into a quadratic polynomial in binary variables. For this purpose, we transform Eq. (9), (14), or (17) expressed in $\boldsymbol{b}$ into a linear function of binary variables by replacing a product of two binary variables $b_i b_j$ by an auxiliary binary variable $\tilde{b}_{ij}$. Then, the cost function $J_2(\boldsymbol{b})$ is transformed into a new cost function $\tilde{J}_2(\boldsymbol{b}, \tilde{\boldsymbol{b}})$. To satisfy the constraint $\tilde{b}_{ij} = b_i b_j$, we introduce the following function (Nishimori and Ozeki 2018):

$$H(b_i, b_j, \tilde{b}_{ij}) \coloneqq 3\tilde{b}_{ij} + b_i b_j - 2b_i \tilde{b}_{ij} - 2b_j \tilde{b}_{ij} \tag{28}$$

This quadratic function vanishes when $\tilde{b}_{ij} = b_i b_j$, and takes a positive integer value, 1 or 3, when $\tilde{b}_{ij} \neq b_i b_j$. Then, $H(b_i, b_j, \tilde{b}_{ij})$ can be used as a penalty term, and the minimization

problem of Eq. (5) can be reduced to a QUBO problem with the following cost function:

$$L(\boldsymbol{b}, \tilde{\boldsymbol{b}}) \coloneqq \lambda \tilde{J}_2(\boldsymbol{b}, \tilde{\boldsymbol{b}}) + \sum_{i<j}^{nZ} H(b_i, b_j, \tilde{b}_{ij}), \tag{29}$$

where λ is a parameter controlling the relative strength of the penalty terms. We minimize this cost function using quantum annealing to obtain guidance on where to explore in control space to find the global minimum of the original cost function. The total number of binary variables in Eq. (29) is $nZ + nZ(nZ-1)/2$. Because multiplying two real numbers doubles the number of bits, we can reduce the number of auxiliary binary variables. Let Z_1 and Z_2 denote the numbers of bits in encoding $\boldsymbol{u}$ in the linear and quadratic terms in Eq. (9), respectively. If $Z_2 < Z_1$, the total number of binary variables is reduced to $nZ_1 + nZ_2(nZ_2-1)/2$.

3. Experimental design

We conducted data assimilation experiments with the Lorenz-63 model as a proof of concept. Because the model has only three state variables, a low-resolution model is not needed and the matrices in Eq. (9) are computed directly without using the transformation in Eq. (8). For preliminary experiments and comparisons, we also conducted data assimilation experiments with the proposed method using simulated annealing instead of quantum annealing. For these comparison, the assimilation method based on simulated annealing is hereafter referred to as SA-4DVar, and the original method is called QA-4DVar.

3.1 Model

The governing equations of the Lorenz 63 model are

$$\frac{dx}{dt} = -\sigma x + \sigma y, \qquad \frac{dy}{dt} = \rho x - y - xz, \qquad \frac{dz}{dt} = xy - \beta z, \tag{30}$$

where the three parameters are set to the conventional values: $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$. The Lyapunov exponents of the model, which characterize the rates of separation of infinitesimally close trajectories in state space, are 0.906, 0, and -14.572, and the Kaplan–Yorke dimension of the attractor (Kaplan and Yorke 1979) is 2.062 (e.g., Sprott, 1997). In our experiments, the fourth-order Runge-Kutta scheme was used for time integration, with a time step of 0.01. The initial condition at t = 0 is set to Gaussian random numbers with a mean of 0 and variance of 1. Time integration of the model from t = 0 to t = 1100 was performed to generate the truth data for the experiments.

The first- and second-order perturbation equations of the Lorenz 63 model are as follows:

$$\frac{d}{dt}\begin{pmatrix}\delta x^1\\ \delta y^1\\ \delta z^1\end{pmatrix} = \begin{pmatrix}-\sigma & \sigma & 0\\ \rho-\bar{z} & -1 & -\bar{x}\\ \bar{y} & \bar{x} & -\beta\end{pmatrix}\begin{pmatrix}\delta x^1\\ \delta y^1\\ \delta z^1\end{pmatrix}, \tag{31}$$

$$\frac{d}{dt}\begin{pmatrix}\delta x^2\\ \delta y^2\\ \delta z^2\end{pmatrix} = \begin{pmatrix}-\sigma & \sigma & 0\\ \rho-\bar{z} & -1 & -\bar{x}\\ \bar{y} & \bar{x} & -\beta\end{pmatrix}\begin{pmatrix}\delta x^2\\ \delta y^2\\ \delta z^2\end{pmatrix} + \begin{pmatrix}0\\ -\delta x^1\delta z^1\\ \delta x^1\delta y^1\end{pmatrix} \tag{32}$$

with the initial conditions given by Eq. (13).

3.2 Observations

The state variables were directly observed at a time interval of unity, and observations were generated by adding random errors to the truth data. These errors were independent

random draws from a Gaussian distribution with a mean of 0 and variance of 1. Under these conditions, the observation operator $\boldsymbol{H}_k$ and observation error covariance matrix $\boldsymbol{R}_k$ are both given by the three-dimensional identity matrix $\boldsymbol{I_3}$. The maximum Lyapunov exponent of the model was 0.906 and the time interval of observations was set to 1, such that the model evolution between adjacent observation times was strongly nonlinear. The quasi-static variational assimilation proposed by Pires et al. (1996) does not work under these conditions due to the coarse temporal resolution of observational data.

3.3 Data assimilation settings

The cost function of incremental 4DVar for the lth outer loop, expressed in terms of the control variable $\boldsymbol{u}$, is derived from Eqs. (5), (7), and (9) as

$$J^{(l)}(\boldsymbol{u}) = \frac{1}{2}\left\|\boldsymbol{u} + \boldsymbol{L}^{-1}\Delta\boldsymbol{x}_0^{(l)}\right\|^2 + \frac{1}{2}\sum_{k=1}^{K}\left\|\delta\boldsymbol{x}_k^{(l)}(\boldsymbol{u}) - \boldsymbol{d}_k^{(l)}\right\|^2, \tag{33}$$

where $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$ and $\boldsymbol{H}_k = \boldsymbol{R}_k = \boldsymbol{I}_3$ are substituted. Matrices $\boldsymbol{M}_k^{(l)}$ and $\left\{\boldsymbol{N}_{k\,i}^{(l)}\right\}_{i=1}^{3}$ in Eq. (9) are computed by integrating Eqs. (31) and (32) from six initial conditions: $\delta\boldsymbol{x}(0) = (1,0,0)^{\mathrm{T}}, (0,1,0)^{\mathrm{T}}, (0,0,1)^{\mathrm{T}}, (1,1,0)^{\mathrm{T}}, (0,1,1)^{\mathrm{T}}$, and $(1,0,1)^{\mathrm{T}}$.

We used the hybrid 4DVar in which the background error covariance matrix at the beginning of the assimilation window was set to the analysis error covariance matrix by the stochastic EnKF (Burgers et al. 1998; Houtekamer and Mitchell 1998), with an ensemble size of 100. Because EnKFs yield flow-dependent forecast and analysis error covariance

matrices by computing the time-evolution of a forecast ensemble, the background error covariance matrix becomes flow-dependent. The stochastic EnKF, which constructs an analysis ensemble by perturbing observational data, is more robust to nonlinearity than the deterministic EnKF, which constructs it by transforming a forecast ensemble (Lawson and Hansen 2004; Lei et al. 2010; Tsuyuki 2024). Therefore, the stochastic EnKF was more suitable for our experiments. The adaptive covariance inflation proposed by Li et al. (2009) based on innovation statistics (Desroziers et al. 2005) was used in the EnKF. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) method was adopted for gradient descent, and the control variable $\boldsymbol{u}$ defined by Eq. (7) was introduced to accelerate the convergence.

The length of the assimilation window was set to 1 or 3, such that K =1 or 3 in Eq. (33). Data assimilation experiments with these window lengths are strongly nonlinear because observations are available only at a time interval of 1. Data assimilation experiments with a window length of 3 are more nonlinear than those with a window length of 1. In the former assimilation window, three independent data assimilation cycles were run starting from *t* = 0, 1, and 2. The analysis at the end of the assimilation window is used to calculate the root mean square error (RMSE) of analysis against the truth. The assimilation period was set to 1100, and the first 100 samples were discarded as spin-up. Thus, 1000 samples were used for verification for each assimilation window length. There were three and nine assimilated observations for window lengths of 1 and 3, respectively. As mentioned in Section 2.2.a, we set the threshold J_c to the upper 0.01% point of the χ^2 distribution with degrees of freedom

given by the number of observations. Thus, the threshold values of the cost function J_c were 10.544 and 16.860 for window lengths of 1 and 3, respectively.

3.4 Quantum and simulated annealing settings

The QUBO problem of the cost function Eq. (29) was solved by changing the search range of control variables r up to 10. The parameter λ in Eq. (29) was set to 0.01, based on results of a preliminary simulated annealing experiment where λ = 0.1, 0.01, and 0.001. The number of reads for quantum and simulated annealing was set to 10; i.e., we repeated the annealing process 10 times and adopted the best result as the minimum. We examined the sensitivity of data assimilation experiments to this parameter for both quantum and simulated annealing, and found that there was no improvement tendency in the analysis RMSE and failure rate as the parameter increased, perhaps due to the small numbers of binary variables.

For quantum annealing, we used D-Wave Advantage v4.1, which consists of 5627 physical qubits and 177 logical qubits (D-Wave Systems 2022a, 2023). The 177 logical qubits were subsequently available for computation. The numbers of bits were set to $Z_1 = 4$ and $Z_2 = 2$, then the number of binary variables was 27 (see the equation given at the end of Section 2). We used the default settings except for the number of reads. The average execution time per annealing with 10 reads was about 18 ms for a window length of 3; this value remained almost constant with each computation, probably because a constant

507 annealing time was used.

508 For simulated annealing, we use the Simulated Annealing Sampler (D-Wave Systems
509 2022b) with PyQUBO (Tanahashi et al. 2019; Recruit Communications) adopted as a
510 supplementary software. PyQUBO allows better code readability, reduced compilation time,
511 and automatic validation of the satisfaction of given constraints. We used the default settings
512 except for the number of reads and the seed of random number generation. For comparison,
513 $Z_1 = Z_2 = 4$ (with 78 binary variables) was adopted in addition to $Z_1 = 4$ and $Z_2 = 2$. For
514 the case of $Z_1 = 4$ and $Z_2 = 2$, the average execution time per annealing with 10 reads is
515 about 16 ms for a window length of 3 on a personal computer with an Intel® Core™ i5-
516 8250U central processing unit at 1.60 GHz. This value changed considerably with each
517 computation, because annealing was terminated when a convergence condition was
518 satisfied.

519

520 **4. Results**

521 We first present the results of data assimilation experiments using hybrid 4DVar to show
522 the difficulty of the conventional 4DVar under strong nonlinear conditions. Next, we provide
523 examples of the landscape of the cost function in state space, including those of the first-
524 and second-order approximated cost functions. These examples are useful interpreting the
525 results of data assimilation experiments using the proposed method, which are presented
526 after that part.

527

528 *4.1 Performance of hybrid 4DVar*

529 Figure 3a shows a scatter diagram on a logarithmic scale between the convergence Fig. 3
530 value of the cost function J_* and the analysis RMSE of hybrid 4DVar without EnKF
531 replacement for a window length of 1. There was almost no correlation between these values
532 for $J_* \leq J_c$ =10.544, whereas a strong positive correlation was observed for $J_* > J_c$, where
533 in most cases the analysis RMSE of hybrid 4DVar without EnKF replacement was larger
534 than the mean analysis RMSE of the stochastic EnKF, which was 0.813 (blue line). This
535 result indicates that the criterion based on the upper 0.01 % point of the χ^2 distribution is
536 useful for determining whether the global minimum has been reached. In this figure, we
537 define success as $J_* \leq J_c$, with all other cases defined as failure. The failure rate is 4.8 %
538 and the mean analysis RMSE only for successful cases is 0.630, which is much smaller than
539 that of the EnKF. This is an expected result, because 4DVar, which does not assume linearity,
540 is more accurate than the EnKF in strongly nonlinear regimes if the global minimum can be
541 reached.

542 The accuracy of hybrid 4DVar can be improved by adopting EnKF replacement. A
543 positive feedback mechanism working in assimilation cycles further contributes to this
544 improvement; if the analysis of a certain assimilation cycle is improved by EnKF replacement,
545 then the background state of the next assimilation cycle is also improved; as a result, the
546 possibility of falling into a local minimum is reduced. Figure 3b shows a scatter diagram of

547 hybrid 4DVar with EnKF replacement. The failure rate is decreased to 2.1 % and the mean
548 analysis RMSE including the failure cases is 0.600, which is smaller than that of hybrid
549 4DVar without EnKF replacement even when only successful cases were considered.

550 Scatter diagrams for a window length of 3 are shown in Fig. 4. The threshold value J_c is Fig. 4
551 16.860. Compared to a window length of 1, the failure rates increased substantially to 39.6%
552 and 66.7% with and without EnKF replacement, respectively. As mentioned in Section 1, we
553 expect that adopting a longer assimilation window would improve accuracy of our results.
554 The mean analysis RMSE of successful cases for hybrid 4DVar without EnKF replacement
555 was 0.542 for a window length of 3; this value is clearly smaller than the corresponding value
556 for a window length of 1 (0.630). However, the mean analysis RMSE for hybrid 4DVar with
557 EnKF replacement is 0.667 for a window length of 3, which is larger than the corresponding
558 value for a window length of 1 (0.600). Thus, our expectation is not met even when EnKF
559 replacement is adopted, due to many failures in the search for global minima for assimilation
560 cycles with a window length of 3, and EnKF analysis is less accurate than 4DVar analysis
561 when the global minimum is reached.

562

563 *4.2 Landscape of the cost function*

564 The performance of 4DVar strongly depends on the landscape of the cost function in
565 state space. Figures 5-8 present example cross-sections of $J(\delta \boldsymbol{x}_0)$ and its first- and
566 second-order approximations along $\delta x_0 = \delta y_0$, which are roughly parallel to the attractor of

567 the Lorenz-63 model. These cross-sections are obtained from data assimilation experiments
568 by SA-4DVar with one outer loop and $Z_1 = Z_2 = 4$. The search ranges of the control
569 variables are set to 4.5 and 1.5 for window lengths of 1 and 3, respectively. The center of
570 each cross-section indicated by "X" is the background state, which is the starting point of the
571 gradient descent method. The isosurface of the first-order approximated cost function is a
572 hyperellipsoid. Because the background error standard deviation, which is equal to the
573 analysis error standard deviation of the EnKF, is about 0.9, the difference in magnitude
574 between the state variable deviation $\delta \boldsymbol{x}_0$ and control variable $\boldsymbol{u}$ is small.

575 The first example (Fig. 5) is for a window length of 1 at t =159, which is the time at the Fig. 5
576 end of the assimilation window. The convergence values of the cost function before and after
577 invoking simulated annealing are 187.791 and 0.965, respectively, indicating that a failure
578 to reach the global minimum can become a success through simulated annealing. The
579 landscape of the cost function is characterized by a shell-like structure that appears as an
580 arch in the cross-section, with a deep valley between the high ridges along the arch (Fig.
581 5a). Because the background state is located at the outer foot of one of these ridges, it
582 appears that the gradient descent method cannot reach into the deep valley. The first-order
583 approximation of the cost function around the background state is shown in Fig. 5b, which
584 shows only a shallow, straight valley; the direction of gradient vector at the background state
585 is identical to that of the original cost function. However, the second-order approximation of
586 the cost function partly reflects the shell-like structure (Fig. 5c). As the global minimum is

587 generally not located on this cross-section, it is difficult to guess how the global minimum is
588 reached by simulated annealing.

589 The second example (Fig. 6) is for a window length of 1 at t =131. The convergence Fig. 6
590 value of the cost function is 109.061 both before and after invoking simulated annealing,
591 indicating a failure of SA-4DVar. A shell-like structure of the cost function is evident in Fig.
592 6a, although it is weak compared to the first example. Unlike Fig. 5a, there is no deep valley
593 within the arch in the cross-section, but a deep concave zone is visible on the opposite side
594 of the arch when viewed from the background state. The global minimum may be located
595 on the opposite side of the shell-like structure, and it may be difficult for the gradient descent
596 method to reach it. The first-order approximation of the cost function represents a deep,
597 wide valley (Fig. 6b), whereas the second-order approximation shows a more complex
598 structure but fails to represent the deep concave zone on the opposite side (Fig. 6c).

599 The third example (Fig. 7) is for a window length of 3 at t =116. The convergence values Fig. 7
600 of the cost function before and after invoking simulated annealing were 34.895 and 6.281,
601 respectively, indicating that a failure in reaching the global minimum becomes successful
602 through simulated annealing. The landscape of the cost function is much more complicated
603 than that for a window length of 1; it can be interpreted as having multiple incomplete shell-
604 like structures, with a similar landscape to that described by Pires et al. (1992). The
605 background state appears to be located in the deepest concave area (fig. 7a), which consists
606 of fine double wells that could lead to failure in searching for the global minimum. The first-

607 order approximation cannot represent these fine double wells, whereas the second-order
608 approximation partly captures this structure and successfully identifies a possible global
609 minimum (Fig. 7b, c).

610 The last example (Fig. 8) is for a window length of 3 at t =159. The convergence values Fig. 8
611 of the cost function before and after invoking simulated annealing were 247.378 and 225.990,
612 respectively, indicating a failure of SA-4DVar. The landscape of the cost function in this case
613 is also complex with multiple shell-like structures, and the background state is located
614 around the edge of one of these shells (Fig. 8a). Both the first- and the second-order
615 approximations of the cost function represent narrow valleys; the latter valley is much
616 narrower (Fig. 8b, c). The background state appearss far from the true state, which may
617 have caused the failure.

618

619 *4.3 Performance of the proposed method*

620 To confirm the benefit of using the second-order approximation of the cost function over
621 the first-order approximation, we conducted data assimilation experiments with SA-4DVar
622 using the first- and second-order approximations, although SA-4DVar is supposed to use
623 the latter approximation in the proposed method. For this analysis, the two SA-4DVar
624 methods are referred to as first- and second-order SA-4DVar, respectively. In first-order SA-
625 4DVar, the second term on the right-hand side of Eq. (9) is neglected, and the numbers of
626 bits are set to $Z_1 = Z_2 = 4$. Figure 9 compares the failure rate of hybrid 4DVar with EnKF Fig. 9

627 replacement (orange line) with those of the two SA-4DVars (other lines) for window lengths
628 of 1 and 3, varying the number of outer loops and the search range r. For hybrid 4DVar with
629 EnKF replacement, the number of cases where $J_* > J_c$ is much smaller for a window length
630 of 1 than for a window length of 3 (Figs. 3b and 4b). This result suggests that the number of
631 operations of simulated annealing for a window length of 1 is also much smaller than that
632 for a window length of 3. To obtain robust results, the average results for five experiments
633 are plotted in Fig. 9a for a window length of 1, whereas those from only one experiment are
634 plotted in Fig. 9b for a window length of 3. The standard deviations are at most 0.002 for
635 both of the failure rate and mean analysis RMSE for a window length of 1. The numbers of
636 operations of simulated annealing for second-order SA-4DVar with three outer loops are
637 41.8 on average at r =5 for a window length of 1, and 484 at r =1 for a window length of
638 3. The average numbers of iterations of the BFGS method per outer loop for the same SA-
639 4DVar are 15.4 at r =5 for a window length of 1, and 20.6 at r =1 for a window length of 3.
640 For a window length of 1 (Fig. 9a), both SA-4DVars show better performance than hybrid
641 4DVar with EnKF replacement in terms of the failure rate; as expected, second-order SA-
642 4DVar shows superior performance to first-order SA-4DVar. The failure rate decreases as
643 the number of outer loops is increased. An unexpected result is that first-order SA-4DVar
644 works well compared to hybrid 4DVar with EnKF replacement, because the searching
645 method depicted in Fig. 1 does not work if the first-order approximation is used. However, if
646 the global minimum is located to the left of the local minimum in Fig. 1, then there is a chance

647 that the cost function approximated around the background state up to the first order also
648 includes the minimum point within the attraction of domain of the global minimum, as in a
649 manner similar to the second-order approximation. Notably, hybrid 4DVar with EnKF
650 replacement attempts the search process using the gradient descent method only once,
651 whereas first- and second-order SA-4DVars attempt more than once and therefore have a
652 greater chance of finding the global minimum. We expect that the probability of success of
653 first-order SA-4DVar decreases compared to that of second-order SA-4DVar as the problem
654 size is increased, due to its poor accuracy of approximation. Another unexpected result for
655 a window length of 1 is that the minimum failure rate of second-order SA-4DVar with two or
656 three outer loops is obtained around $r = 5$, because the second-order incremental
657 approach is based on Taylor expansion and therefore its validity may be guaranteed for $r \lesssim$
658 1. This phenomenon may be explained in terms of the landscape of the cost function for a
659 window length of 1 (Figs. 5a and 6a). When the global minimum is located along the shell-
660 like structure of the landscape, a large search range is desirable. However, this result seems
661 entirely accidental due to the particular landscape of the cost function, and is not a general
662 result.

663 For a window length of 3 (Fig. 9b), the improvement obtained by the two SA-4DVars
664 becomes more significant compared to hybrid 4DVar with EnKF replacement, partly due to
665 the positive feedback in assimilation cycles, as described in Section 4.1. Second-order SA-
666 4DVar is also superior to first-order SA-4DVar in terms of the failure rate when $r < 3$.

667 However, although the smallest failure rates are obtained by second-order SA-4DVar when
668 $r = 1$, the opposite is true when $r > 4$. This unexpected result may be explained in terms
669 of the landscape of the original and approximated cost functions (Figs. 7 and 8). The original
670 cost function for a window length of 3 has a rugged, complicated landscape compared to
671 that for a window length of 1, and the first-order approximation tends to yield a smooth, flat
672 landscape compared to the second-order approximation. As a result, when r is large, the
673 minimum point obtained by first-order SA-4DVar may be more likely to be found in the
674 domain of attraction of a possible global minimum.

675 Based on the results presented in Fig. 9, where the numbers of bits are set to $Z_1 = Z_2 =$
676 4, we compare the performance of QA-4DVar and SA-4DVar for small r values with the
677 numbers of bits set to $Z_1 = 4$ and $Z_2 = 2$. Figure 10 shows the failure rates and mean
678 analysis RMSEs of QA-4DVar (solid lines) and SA-4DVar (dotted line) for a window length
679 of 1. SA-4DVar is identical to second-order SA-4DVar in Fig. 9 except for the numbers of
680 bits. Averages over five experiments are presented for the reason mentioned in the first
681 paragraph of this section. QA-4DVar and SA-4DVar show very similar minimum failure rates
682 and RMSEs at $r = 4$ for each window length, but QA-4DVar tends to perform better than
683 SA-4DVar at other values of r. These results demonstrate that QA-4DVar shows
684 comparable or superior performance to SA-4DVar, possibly due to quantum tunneling, and
685 that two outer loops are almost sufficient for a window length of 1. The smallest failure rate
686 and mean analysis RMSE are obtained at $r = 4.5$ using QA-4DVar with three outer loops.

Fig. 10

Figure 11 shows the failure rates and mean analysis RMSEs of QA-4DVar (solid lines) and SA-4DVar (dotted line) for a window length of 3. The longer window benefits more from the 3rd outer loop because this is a more nonlinear problem. QA-4DVar and SA-4DVar showsimilar performance. A comparison between Figs. 10b and 11b reveals that increasing the length of the assimilation window reduces the accuracy of hybrid 4DVar with EnKF replacement, whereas the opposite trend is observed for both QA-4DVar and SA-4DVar. Thus, the proposed method derivess a benefit from increasing the window length. A comparison of Figs. 10 and 11 with Fig. 9 reveals little evidence of any adverse effect of reducing the Z_2 value.

Fig. 11

Figure 12a is a scatter diagram of QA-4DVar results for a window length of 1 with three outer loops and $r =$ 4.5. This experiment has the smallest failure rate among the five experiments. Compared to hybrid 4DVar with EnKF replacement (Fig. 3b), the failure rate is reduced from 2.1% to 0.6% and the mean analysis RMSE is reduced from 0.600 to 0.586. Figure 12b shows a scatter diagram of QA-4DVar results for a window length of 3 with three outer loops and r =1.5. The failure rate and mean analysis RMSE are significantly improved compared to hybrid 4DVar with EnKF replacement (Fig. 4b); the failure rate is reduced from 39.6% to 8.1% and mean analysis RMSE is reduced from 0.667 to 0.547. Notably, this mean analysis RMSE value is almost equal to that of hybrid 4DVar without EnKF replacement for successful cases only (Fig. 4a).

Fig. 12

5. Summary and Discussion

This study proposes a method of searching for the global minimum in 4DVar by combining a second-order incremental approach and quantum annealing for QUBO, in which the latter provides guidance on where to explore in state space by minimizing an approximated cost function This approximated cost function is constructed in low-dimensional space by expanding state variables up to the second order around the basic state of an outer loop, encoding the perturbation variables into binary variables, and transforming second-order terms into first-order terms according to a property of binary variables. We also propose two approximation methods in state space and ensemble space to reduce the problem size. If the global minimum cannot be reached after a couple of iterations of the outer loop, the 4DVar analysis is replaced by an EnKF analysis in assimilation cycles.

Data assimilation experiments using the Lorenz 63 model were conducted as a proof of concept of the proposed method. The results revealed that the proposed method can significantly reduce the frequency of falling into local minima, and that the benefit of extending the length of the assimilation window is realized even in strongly nonlinear regimes. Data assimilation experiments adopting simulated annealing instead of quantum annealing showed that the performance of quantum annealing is comparable to or better than that of simulated annealing, possibly due to quantum tunneling. Additional experiments to investigate further the performance of the proposed method and comparisons with other

methods are among the topics of future study.

Since the data assimilation experiments were conducted using the Lorenz 63 model the performance of each approximation method described in Section 2.2 was not examined. However, it is important to clarify the effectiveness of using the minimum point obtained by quantum annealing in low-dimensional space as guidance for determining which regions of state space to explore. Therefore, subsequent research will focus on data assimilation experiments using a higher-dimensional model such as the Lorenz 96 model with the approximation methods described in this study. Only 177 logical qubits are available on D-Wave Advantage v4.1; therefore, SA-4DVar must be adopted as a data assimilation method instead of QA-4DVar. Simulated quantum annealing (e.g., Nishimori and Ozeki 2018), which simulates some aspects of quantum annealing on a classical computer, may also be applied.

Data Availability Statement

The output data from this study have been archived and are available upon request to the corresponding author.

Supplement

The Python programs of data assimilation experiments with QA-4DVar and SA-4DVar used in this study are available as the supplementary material of this paper.

Acknowledgments

This study was partly supported by the Japan Science and Technology Agency Moonshot R&D (JPMJMS2389), the Japan Society for the Promotion of Science (JSPS) via KAKENHI (grant nos. JP21H04571, and JP22K18821), and the IAAR Research Support Program of Chiba University. The authors would like to thank Dr. Masanao Ohashi for his technical support for using the D-Wave Advantage v4.1.

References

Abel, S., A. Blance, and M. Spannowsky, 2022: Quantum optimization of complex systems with a quantum annealer. *Phys. Rev. A*, **106**, 042607, DOI: 10.1103/PhysRevA.106.042607.

Andersson, E., M. Fisher, E. Hólm, L. Isaksen, G. Radnóti, and Y. Trémolet, 2005: Will the 4D-Var approach be defeated by nonlinearity? *ECMWF Technical Memoranda,* Vol. 479, ECMWF, 26 pp.

Arai, S., H. Oshiyama, and H. Nishimori, 2023: Effectiveness of quantum annealing for continuous-variable optimization. *Phys. Rev. A*, **108**, 0424403, DOI: 10.1103/PhysRevA.108.042403

Bonavita, M., E. Hólm, L. Isaksen, and M. Fisher, 2016: The evolution of the ECMWF hybrid

data assimilation system, *Q. J. Roy. Meteor. Soc.*, **142**, 287–303, https://doi.org/10.1002/qj.2652.

Born, M., and V. A. Fock, 1928: Beweis des Adiabatensatzes. *Zeitschrift für Physik A*. **51** (3–4): 165–180, DOI:10.1007/BF01343193.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Chancellor, N., 2019: Domain wall encoding of discrete variables for quantum annealing and QAOA. *Quantum Sci. Technol.*, **4**, 045004.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396.

D-Wave Systems, 2022a: Advantage Processor Overview. https://www.dwavesys.com/media/3xvdipcn/14-1058a-a_advantage_processor_overview.pdf (last access: 30 November 2024).

D-Wave Systems, 2022b: dwave-neal 0.5.9 documentation. https://docs.ocean.dwavesys.com/projects/neal/en/latest/index.html (last access: 30 November 2024)

787 D-Wave Systems, 2023: D-Wave Ocean software documentation.
788 https//docs.ocean.dwavesys.com/en/stable/index.html (last access: 30 November 2024).
789 Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction
790 of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703-713.
791 Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model
792 using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–
793 10162.
794 Farhi, E., J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, 2001: A quantum
795 adiabatic evolution algorithm applied to random instances of an NP-complete problem.
796 *Science*, **292**, 472–476.
797 Finnila, A. B., M. A. Gomez, C. Sebenik, C. Stenson, and J.D. Doll, 1994: Quantum
798 annealing: A new method for minimizing multidimensional functions. *Chem. Phys. Lett.*,
799 **219**, 343–348.
800 Gauthier, P., 1992: Chaos and quadri-dimensional data assimilation: A study based on the
801 Lorenz model. *Tellus*, **44A**, 2-17.
802 Goldstein, H., C. Poole, and J. Safko, 2001: *Classical Mechanics*, 3rd ed., Pearson, 664 pp.
803 Golub, G. H., and C. F. Van Loan, 2012: *Matrix Computations*, 4th ed., Johns Hopkins
804 University Press, 756 pp.
805 Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman
806 filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

807 Hu, F., B. N. Wang, N. Wang, and C. Wang, 2019: Quantum machine learning with D-wave
808 quantum computer, *Quant. Eng.*, 1, e12, https://doi.org/10.1002/que2.12.
809 Inoue, D., and H. Yoshida, 2020: Model predictive control for finite input systems using the
810 D-Wave quantum annealer. *Scientific Reports*, **10**:1591, https://doi.org/10.1038/s41598-
811 020-58081-9.
812 Johnson, M. W., M. H. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J.
813 Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E.
814 Ladizinsky, N. Ladinzinski, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheve, C. J.
815 S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, 2011: Quantum annealing with
816 manufactured spins, *Nature*, 473, 194–198, https://doi.org/10.1038/nature10012.
817 Kadowaki, T., and H. Nishimori, 1998: Quantum annealing in the transverse Ising model.
818 *Phys. Rev. E*, **58**, 5355–5363.
819 Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge
820 University Press, Cambridge, 341 pp.
821 Kaplan, J. L., and J. A. Yorke, 1979: Chaotic behavior of multidimensional difference
822 equations. Functional Differential Equations and Approximation of Fixed Points. Peitgen,
823 H.-O., and H.-O. Waters (eds.), *Lecture Notes in Mathematics*, vol. 730, Springer Verlag,
824 Berlin, 204–227.
825 Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, 1983: Optimization by simulated annealing.
826 *Science*, **220**, 671–680.

827 Koh, Y. W., and H. Nishimori, 2022: Quantum and classical annealing in a continuous space
828 with multiple local minima. *Phys. Rev. A*, **105**, 062435**,** DOI:
829 10.1103/PhysRevA.105.062435.
830 Kotsuki, S., F. Kawasaki, and M. Ohashi, 2024: Quantum data assimilation: A new approach
831 to solve data assimilation on quantum annealers. *Nonlinear Processes Geophys.*, **31**,
832 237–245.
833 Lawson, W. G., and J. A. Hansen, 2004: Implications of stochastic and deterministic filters
834 as ensemble-based data assimilation methods in varying regimes of error growth. *Mon.*
835 *Wea. Rev.*, **132**, 1966–1981.
836 Le Dimet, F.X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation
837 of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
838 Lei, J., P. Bickel, and C. Snyder, 2010: Comparison of ensemble Kalman filters under non-
839 Gaussianity. *Mon. Wea. Rev.*, **138**, 1293–1306.
840 Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and
841 observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**,
842 523–533.
843 Lorenz, E. D., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
844 Lorenz, E. D., 1996: Predictability – A problem partly solved, *Proceedings of the ECMWF*
845 *Seminar on Predictability (4-9 September 1995, Reading, UK)*, ECMWF, 1-18.
846 Michel, Y., 2014: Diagnostics on the cost-function in variational assimilations for

847 meteorological models. *Nonlinear Processes Geophys.*, **21**, 187–199.

848 Nishimori, H., and M. Ozeki, 2018: *Elements of Quantum Annealing*, Kyoritsu Publication
849 Inc., Tokyo, 114 pp. (in Japanese)

850 O'Malley, D., V. V. Vesselinov, B. S. Alexandrov, and L. B. Alexandrov, 2018:
851 Nonnegative/binary matrix factorization with a d-wave quantum annealer, *PloS one*, **13**,
852 e0206653, https://doi.org/10.1371/journal.pone.0206653.

853 Pires, C. A., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational
854 assimilation in nonlinear chaotic systems. *Tellus*, **48A**, 96-121.

855 Rajak, A., S. Suzuki, A. Dutta, and B. K. Chakrabarti, 2023: Quantum annealing: an overview.
856 *Phil. Trans. R. Soc. A.*, **381**, 20210417, http://doi.org/10.1098/rsta.2021.0417.

857 Recruit Communications: PyQUBO — pyqubo 1.0.5 documentation.
858 https://pyqubo.readthedocs.io/en/latest/ (last access: 30 November 2024)

859 Selz, T., and G. C. Craig, 2023: Can artificial intelligence-based weather prediction models
860 simulate the butterfly effect? *Geophys. Res. Let.*, **50**, e2023GL105747,
861 https://doi.org/10.1029/2023GL105747.

862 Sprott, J. C., 1997: Lyapunov exponent and dimension of the Lorenz attractor.
863 https://sprott.physics.wisc.edu/chaos/lorenzle.htm (last access: 30 November 2024).

864 Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations
865 with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311-1328.

866 Tanahashi, K., S. Takayanagi, T. Motohashi, and S. Tanaka, 2019: Application of Ising

867 machines and a software development for Ising machines. *J. Phys. Soc. Japan*, **88**,
868 061010, https://doi.org/10.7566/JPSJ.88.061010.
869 Thépaut, J.-N., R. N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and
870 observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*,**121**, 3393–
871 3414.
872 Tsuyuki, T., and T. Miyoshi, 2007: Recent progress of data assimilation in meteorology. *J.*
873 *Meteor. Soc. Japan*, **85B**, 331−361.
874 Tsuyuki, T., 2014: Deterministic predictability of the most probable state and reformulation
875 of variational data assimilation. *J. Meteor. Soc. Japan*, **92**, 599–622.
876 Tsuyuki, T., 2024: A hybrid ensemble Kalman filter to mitigate non-Gaussianity in nonlinear
877 data assimilation. *J. Meteor. Soc. Japan*, **102**, 507–524.
878 Ushijima-Mwesigwa, H., C. F. Negre, and S. M. Mniszewski, 2017: Graph partitioning using
879 quantum annealing on the D-Wave system, in: *Proceedings of the Second International*
880 *Workshop on Post Moore's Era Supercomputing*,
881 https://doi.org/10.1145/3149526.3149531.
882 Willsch, D., M. Willsch, H. De Raedt, and K. Michielsen, 2020: Support vector machines on
883 the D-Wave quantum annealer, *Comput. Phys. Commun.*, **248**, 107006,
884 https://doi.org/10.1016/j.cpc.2019.107006..
885

886 List of Figures

887

888 Fig. 1. Schematic of the search for the global minimum of a cost function (blue solid line)
889 in a one-dimensional system. Quantum annealing (QA) provides guidance on where to
890 explore by minimizing an approximated cost function (orange dashed line). Crosses
891 indicate the starting points for minimization of the cost function by gradient descent; black
892 circles indicate the minima of the cost function, and the white circle indicates the minimum
893 of the approximated cost function obtained by QA in discrete space, where the control
894 variable is assumed to take integer values only.

895

896 Fig. 2. Workflows for (a) the proposed method of searching for the global minimum in
897 4DVar and (b) hybrid 4DVar with EnKF replacement used for comparison. If the
898 convergence value of the cost function J_* exceeds the threshold, J_c, then gradient
899 descent (GD) is assumed to have failed to reach the global minimum, and either quantum
900 annealing (QA) or EnKF replacement is conducted.

901

902 Fig. 3. Scatter diagrams of J_* and analysis RMSE in assimilation cycles with a window
903 length of 1 for (a) hybrid 4DVar without EnKF replacement and (b) hybrid 4DVar with EnKF
904 replacement. Blue line indicates mean analysis RMSE of EnKF; dotted line indicates J_c.
905 Mean analysis RMSEs and failure rates of hybrid 4DVars are shown in each panel, and

mean analysis RMSE only for cases where $J_* \leq J_c$ is also shown in (a).

Fig. 4. Same as Fig. 3 except for a window length of 3.

Fig. 5. Cross-sections of the cost function in state space along $\delta x_0 = \delta y_0$ at t =159 in assimilation cycles of SA-4DVar for a window length of 1 with one outer loop for (a) the original cost function, (b) first-order approximation, and (c) second-order approximation. The origin ("X") is the background state, and δs_0 is the coordinate along $\delta x_0 = \delta y_0$, defined by sgn(δx_0)$\sqrt{{\delta x_0}^2 + {\delta y_0}^2}$. The numbers of bits are $Z_1 = Z_2 = 4$.

Fig. 6. Same as Fig. 5 except for t =131.

Fig. 7. Same as Fig. 5 except for a window length of 3 and t =116.

Fig. 8. Same as Fig. 5 except for a window length of 3 and t =159.

Fig. 9. Failure rates of first-order SA-4DVar (dotted lines) and second-order SA-4DVar (solid lines) plotted against the search range r for window lengths of (a) 1 and (b) 3. Averages over five experiments are presented in (a). Orange line indicates the failure rate of hybrid 4DVar with EnKF replacement. Green, blue, and cyan lines indicate one, two,

926 and three outer loops, respectively. The numbers of bits are $Z_1 = Z_2 = 4$.

927

928 Fig. 10. (a) Failure rates and (b) mean analysis RMSEs of QA-4DVar (solid lines) and SA-
929 4DVar (dotted lines) plotted against the search range r for a window length of 1.
930 Averages over five experiments are presented. Orange line indicates the failure rate of
931 hybrid 4DVar with EnKF replacement. Green, blue, and cyan lines indicate one, two, and
932 three outer loops, respectively. The numbers of bits are $Z_1 = 4$ and $Z_2 = 2$. SA-4DVar is
933 the same as second-order SA-4DVar in Fig. 9 except for the numbers of bits.

934

935 Fig. 11. (a) Failure rates and (b) mean analysis RMSEs of QA-4DVar (solid lines) and SA-
936 4DVar (dotted lines) plotted against the search range r for a window length of 3. Orange
937 line indicates the failure rate of hybrid 4DVar with EnKF replacement. Green, blue, and
938 cyan lines indicate one, two, and three outer loops, respectively. The numbers of bits are
939 $Z_1 = 4$ and $Z_2 = 2$. SA-4DVar is the same as second-order SA-4DVar in Fig. 9 except for
940 the numbers of bits.

941

942 Fig. 12. Scatter diagrams of J_* and analysis RMSE in assimilation cycles for QA-4DVar
943 with three outer loops for window lengths of (a) 1 and (b) 3 for optimal r values of 4.5
944 and 1.5, respectively. Blue line indicates mean analysis RMSE of EnKF; the dotted line
945 indicates J_c. Mean analysis RMSEs and failure rates of QA-4DVar are shown in each

946 panel. The numbers of bits are $Z_1 = 4$ and $Z_2 = 2$.

947

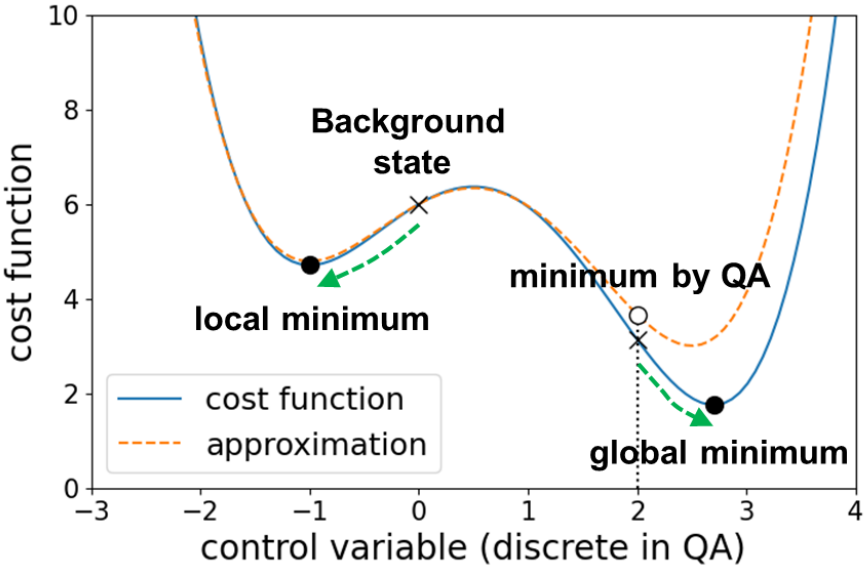
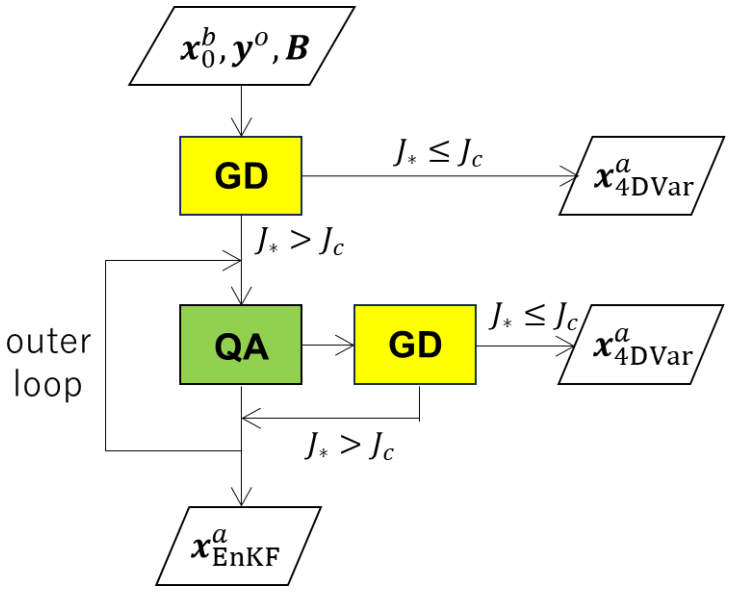


Fig. 1. Schematic of the search for the global minimum of a cost function (blue solid line) in a one-dimensional system. Quantum annealing (QA) provides guidance on where to explore by minimizing an approximated cost function (orange dashed line). Crosses indicate the starting points for minimization of the cost function by gradient descent; black circles indicate the minima of the cost function, and the white circle indicates the minimum of the approximated cost function obtained by QA in discrete space, where the control variable is assumed to take integer values only.

(a)



(b)

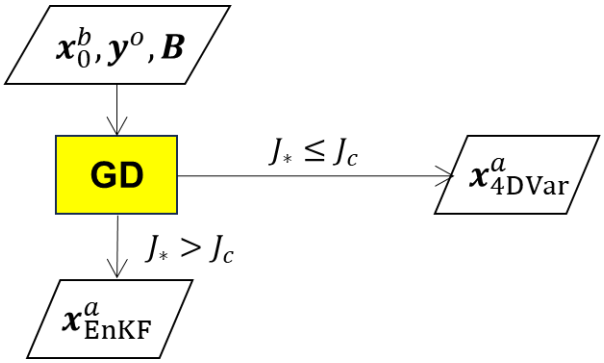


Fig. 2. Workflows for (a) the proposed method of searching for the global minimum in 4DVar and (b) hybrid 4DVar with EnKF replacement used for comparison. If the convergence value of the cost function J_* exceeds the threshold, J_c, then gradient descent (GD) is assumed to have failed to reach the global minimum, and either quantum annealing (QA) or EnKF replacement is conducted.

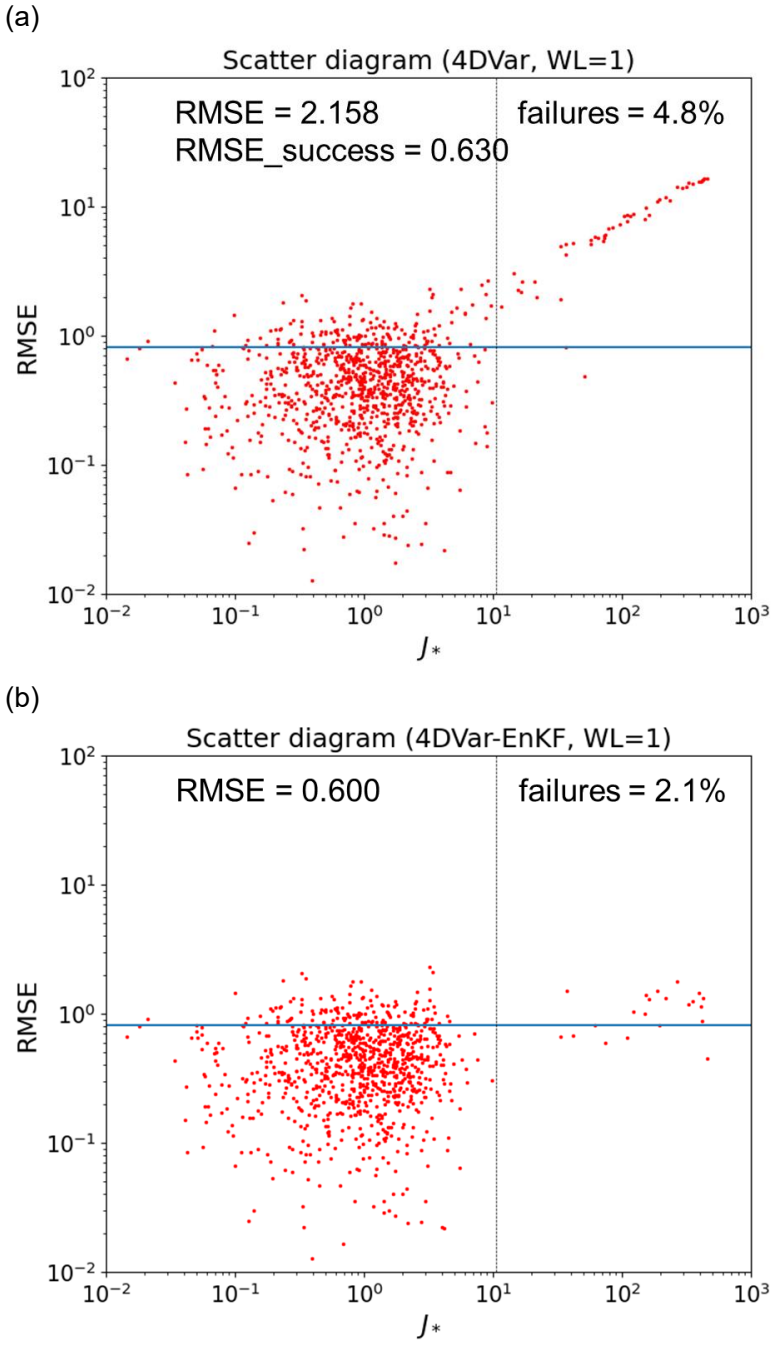
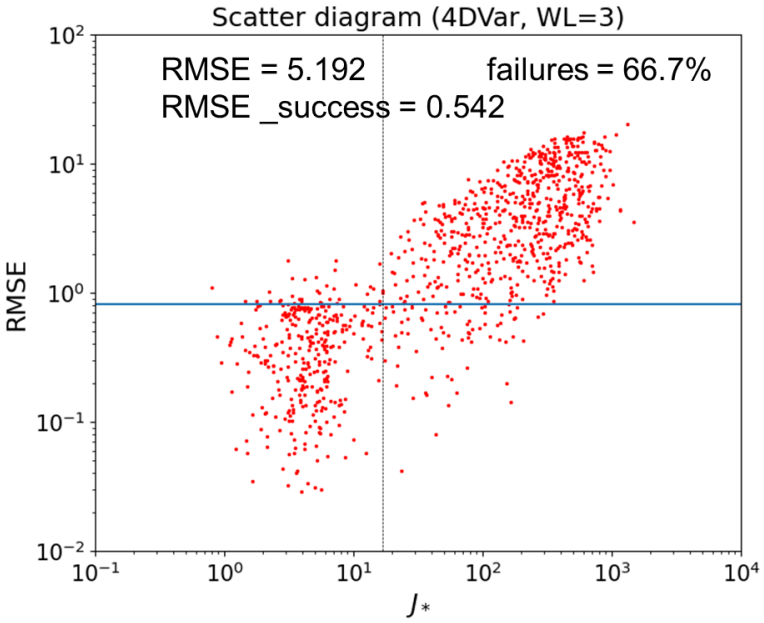


Fig. 3. Scatter diagrams of J_* and analysis RMSE in assimilation cycles with a window length of 1 for (a) hybrid 4DVar without EnKF replacement and (b) hybrid 4DVar with EnKF replacement. Blue line indicates mean analysis RMSE of EnKF; dotted line indicates J_c. Mean analysis RMSEs and failure rates of hybrid 4DVars are shown in each panel, and mean analysis RMSE only for cases where $J_* \leq J_c$ is also shown in (a).

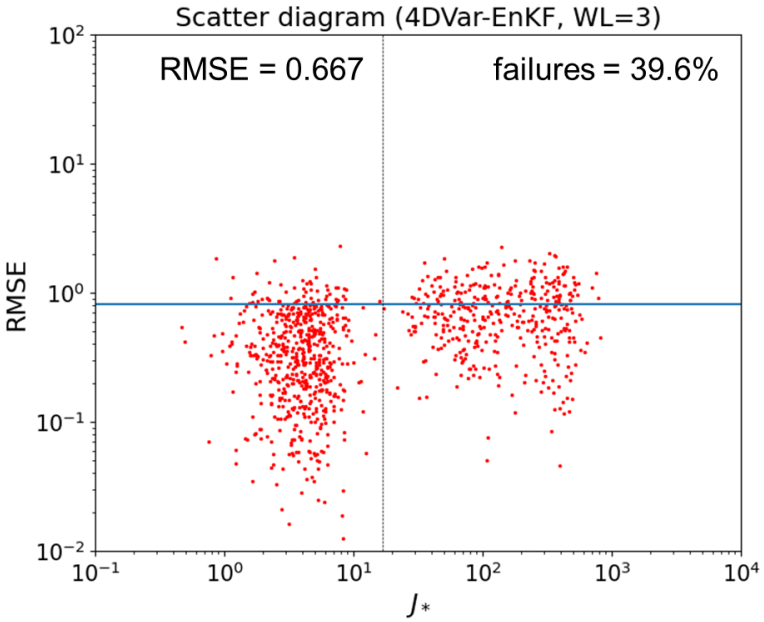
975

(a)



976

(b)

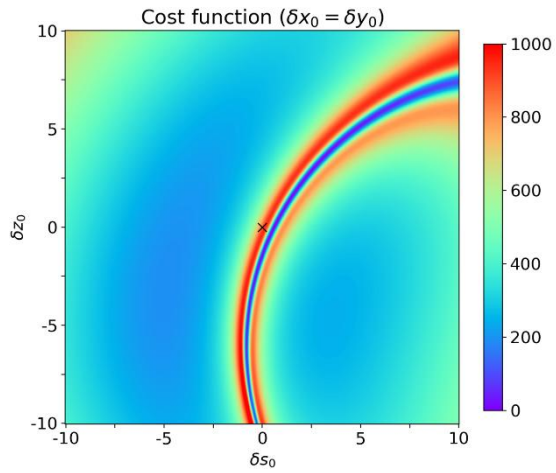


977

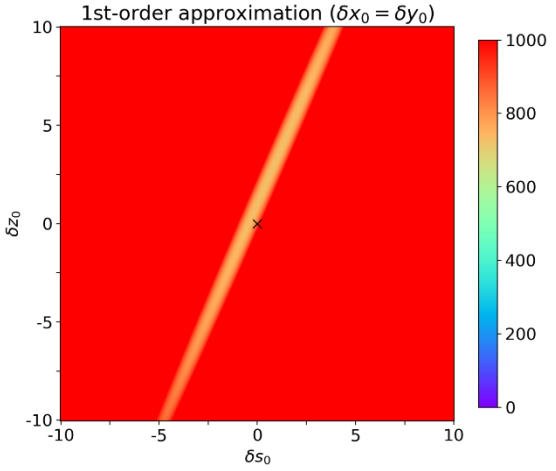
978 Fig. 4. Same as Fig. 3 except for a window length of 3.

979

(a)



(b)



(c)

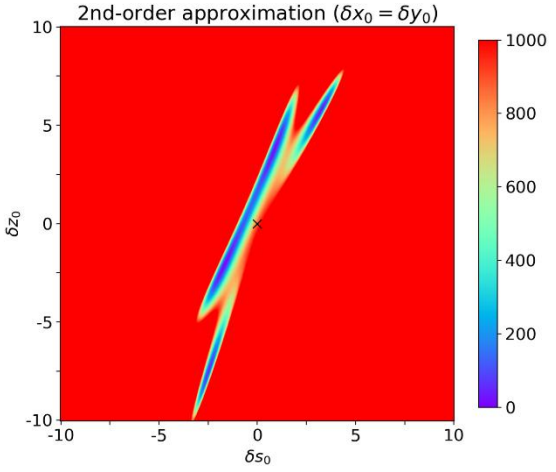
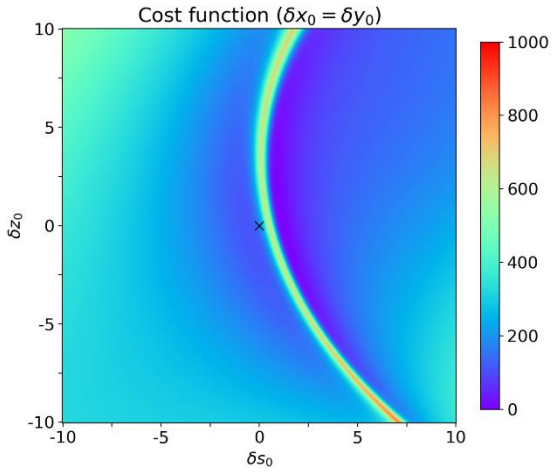
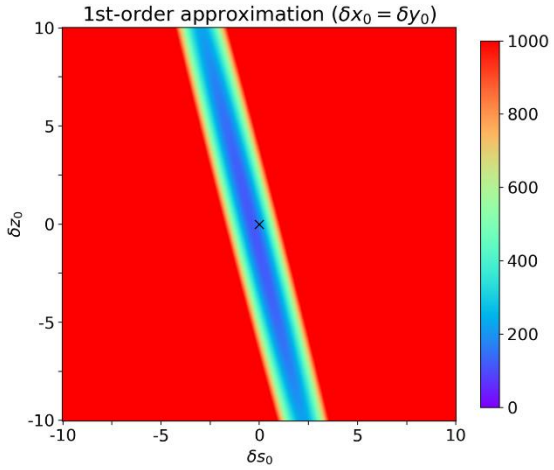


Fig. 5. Cross-sections of the cost function in state space along $\delta x_0 = \delta y_0$ at t =159 in assimilation cycles of SA-4DVar for a window length of 1 with one outer loop for (a) the original cost function, (b) first-order approximation, and (c) second-order approximation. The origin ("X") is the background state, and δs_0 is the coordinate along $\delta x_0 = \delta y_0$, defined by $\mathrm{sgn}(\delta x_0)\sqrt{\delta x_0{}^2 + \delta y_0{}^2}$. The numbers of bits are $Z_1 = Z_2 = 4$.

(a)



(b)



(c)

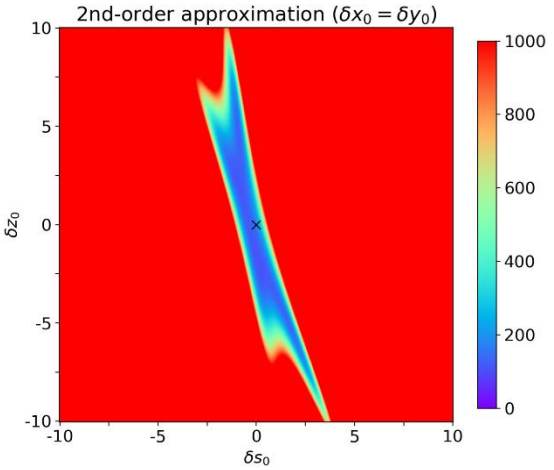
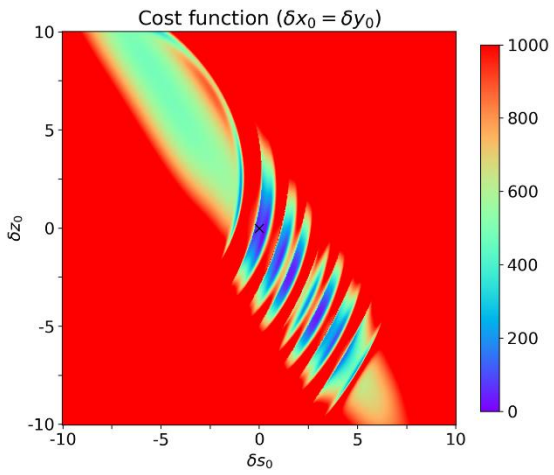
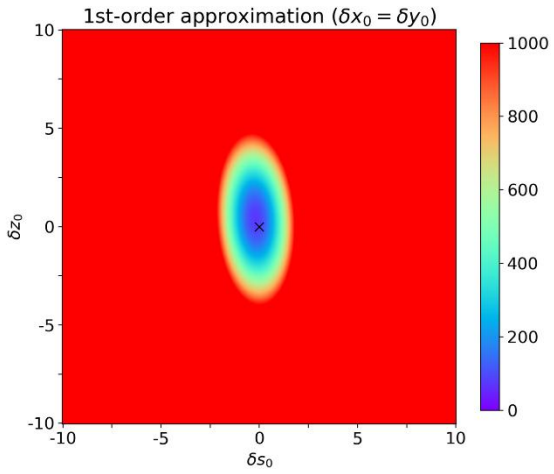


Fig. 6. Same as Fig. 5 except for t =131.

(a)



(b)



(c)

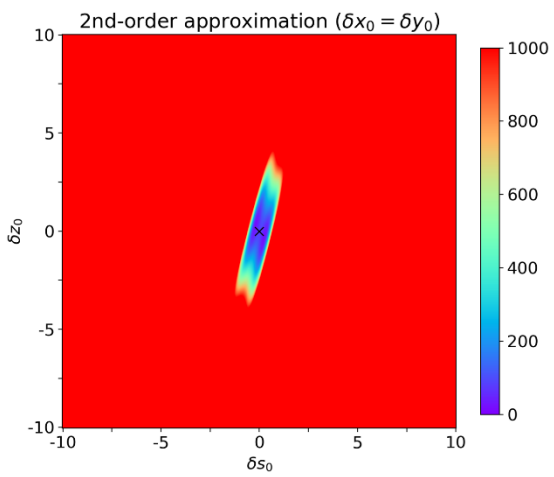
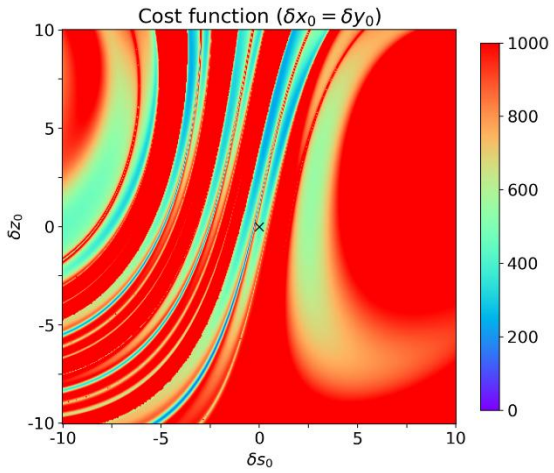
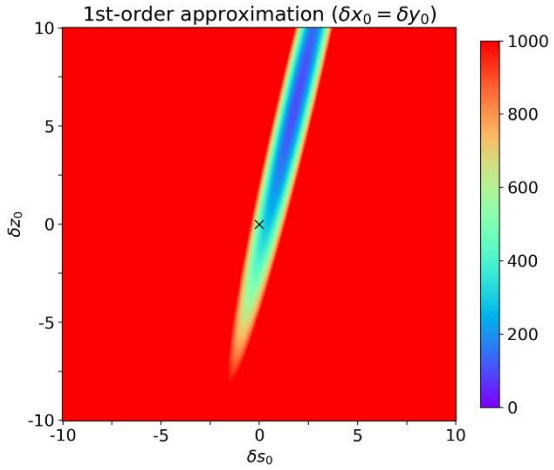


Fig. 7. Same as Fig. 5 except for a window length of 3 and t =116.

(a)



(b)



(c)

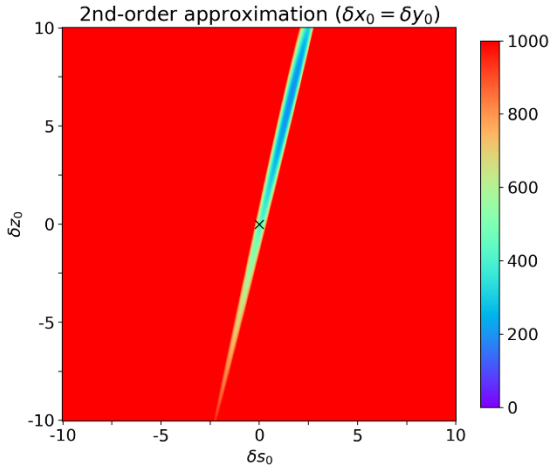


Fig. 8. Same as Fig. 5 except for a window length of 3 and t =159.

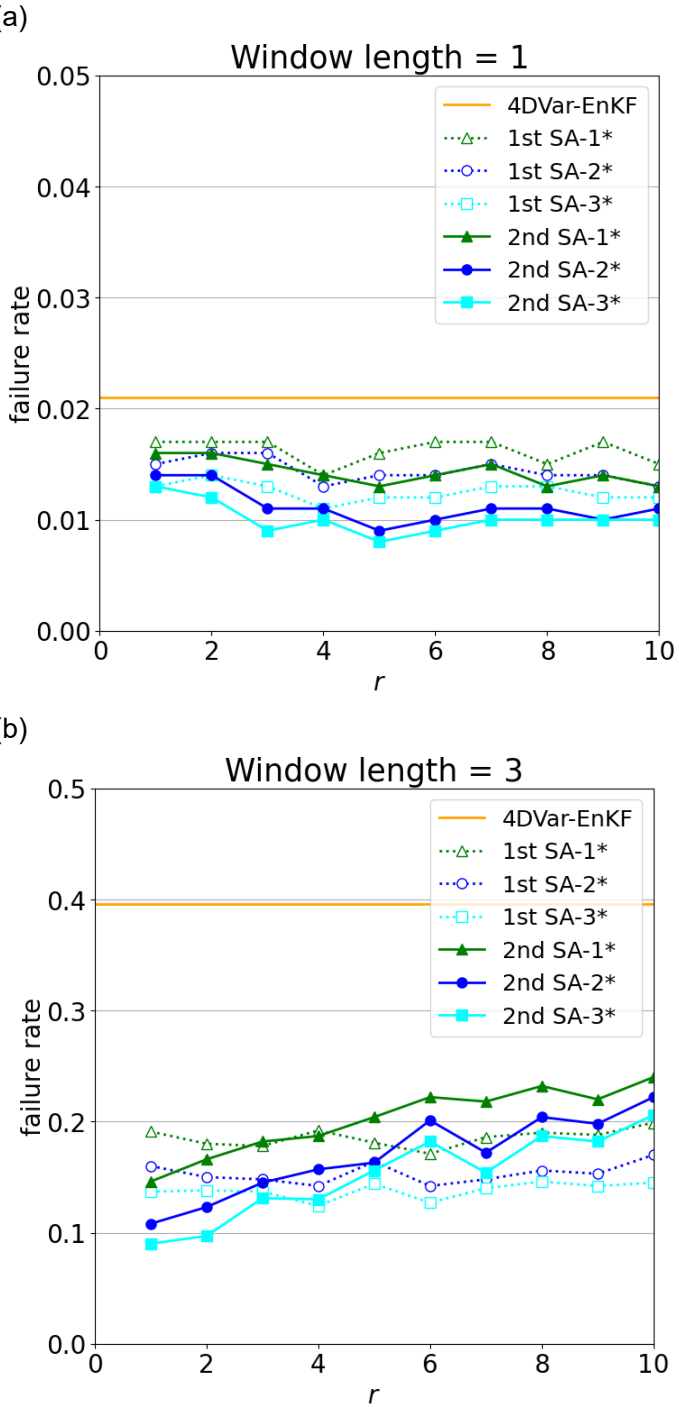
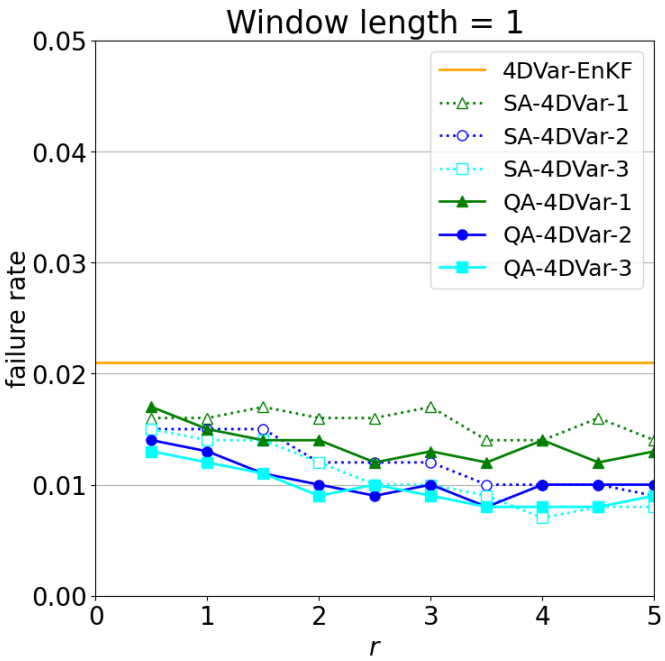


Fig. 9. Failure rates of first-order SA-4DVar (dotted lines) and second-order SA-4DVar (solid lines) plotted against the search range r for window lengths of (a) 1 and (b) 3. Averages over five experiments are presented in (a). Orange line indicates the failure rate of hybrid 4DVar with EnKF replacement. Green, blue, and cyan lines indicate one, two, and three outer loops, respectively. The numbers of bits are $Z_1 = Z_2 = 4$.

(a)



(b)

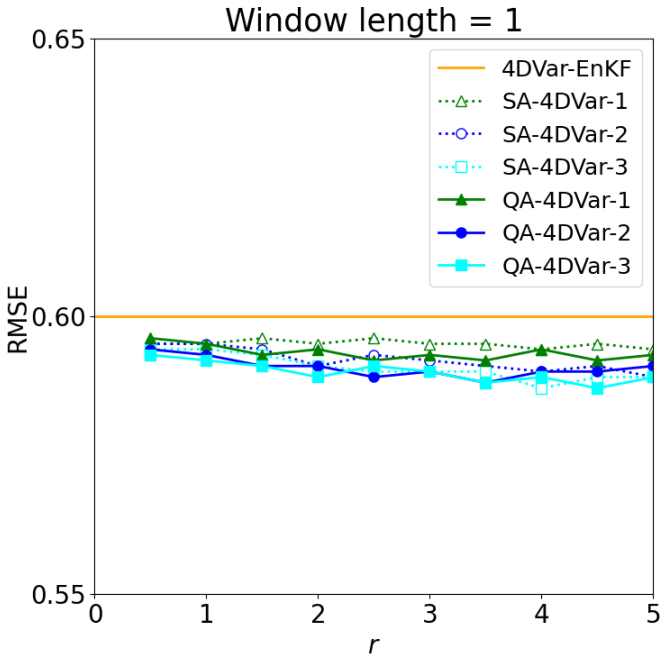
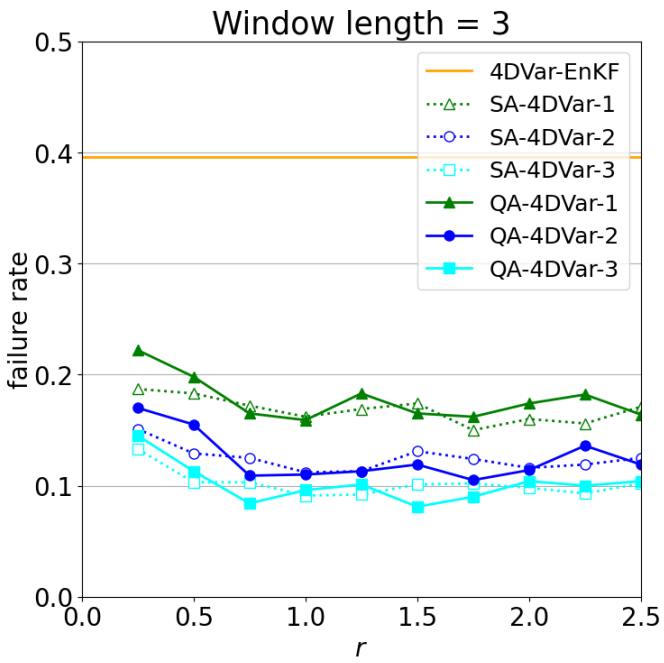


Fig. 10. (a) Failure rates and (b) mean analysis RMSEs of QA-4DVar (solid lines) and SA-4DVar (dotted lines) plotted against the search range r for a window length of 1. Averages over five experiments are presented. Orange line indicates the failure rate of hybrid 4DVar with EnKF replacement. Green, blue, and cyan lines indicate one, two, and three outer loops, respectively. The numbers of bits are $Z_1 = 4$ and $Z_2 = 2$. SA-4DVar is the same as second-order SA-4DVar in Fig. 9 except for the numbers of bits.

(a)



(b)

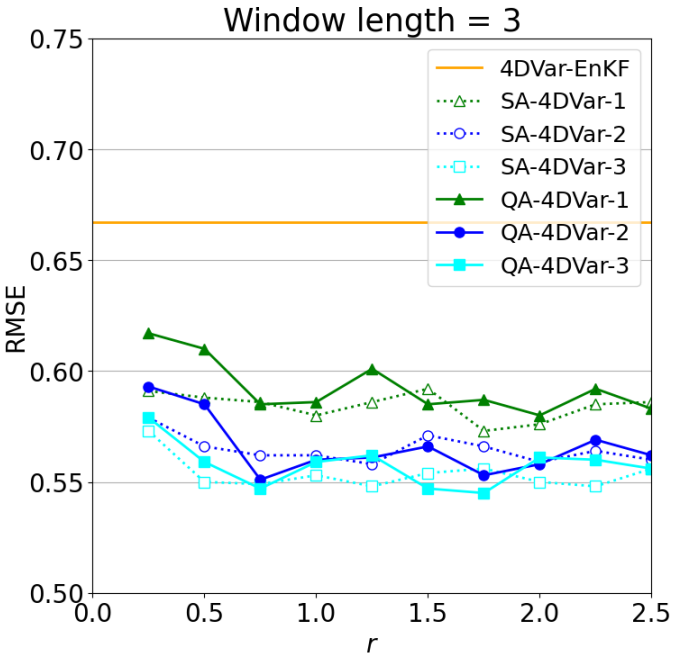
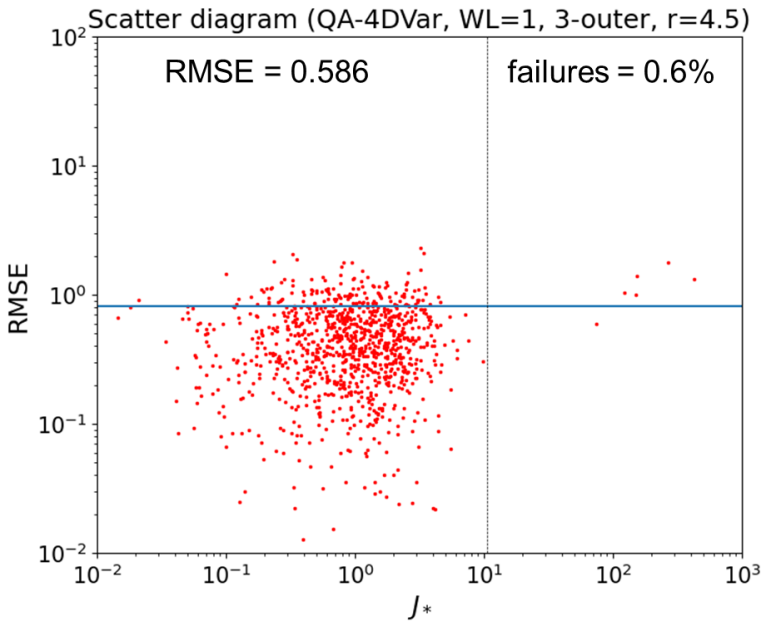


Fig. 11. (a) Failure rates and (b) mean analysis RMSEs of QA-4DVar (solid lines) and SA-4DVar (dotted lines) plotted against the search range r for a window length of 3. Orange line indicates the failure rate of hybrid 4DVar with EnKF replacement. Green, blue, and cyan lines indicate one, two, and three outer loops, respectively. The numbers of bits are $Z_1 = 4$ and $Z_2 = 2$. SA-4DVar is the same as second-order SA-4DVar in Fig. 9 except for the numbers of bits.

(a)



(b)

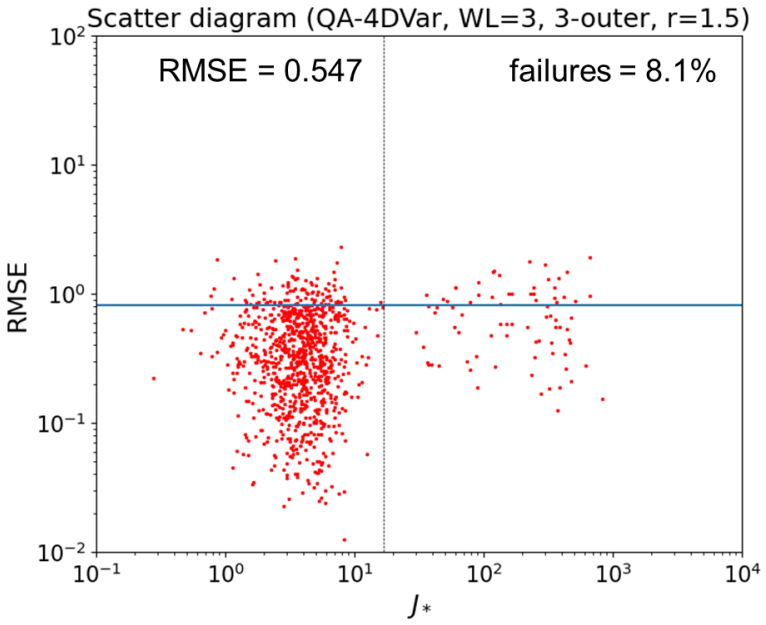


Fig. 12. Scatter diagrams of J_* and analysis RMSE in assimilation cycles for QA-4DVar with three outer loops for window lengths of (a) 1 and (b) 3 for optimal r values of 4.5 and 1.5, respectively. Blue line indicates mean analysis RMSE of EnKF; the dotted line indicates J_c. Mean analysis RMSEs and failure rates of QA-4DVar are shown in each panel. The numbers of bits are $Z_1 = 4$ and $Z_2 = 2$.