Science Bulletin 64 (2019) 321–330

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

Article

Full waveform inversion based on the ensemble Kalman filter method using uniform sampling without replacement

Jian Wang, Dinghui Yang *, Hao Jing, Hao Wu

Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 5 November 2018
Received in revised form 18 January 2019
Accepted 19 January 2019
Available online 2 February 2019

Keywords:
Data assimilation
Ensemble Kalman filter
Uniform sampling without replacement
Full waveform inversion

ABSTRACT

Full waveform inversion (FWI) has been increasingly more and more important in seismology to better understand the interior structure of the Earth. FWI, by taking advantage of both the traveltime and amplitude in the data, provides high-resolution model parameters of the earth which can produce images with high resolution. However, this inversion method conventionally suffers from non-uniqueness due to many local minima of the objective function and large computing costs. In this study, we propose a new FWI method in a semi-random framework by integrating the ensemble Kalman filter and uniform sampling without replacement. Numerical results demonstrate that the new method can achieve high-resolution results and a wider convergence domain. Accordingly, the new method overcomes the disadvantage of conventional FWIs that depend strongly on the initial model.

1. Introduction

Seismic waves carry information about the Earth's interior and facilitate our understanding of subsurface structures. Many studies have been conducted for the use of seismic waves to infer the underground structures [1–3]. Geophysical inversions, which aim to obtain information about geological structures and properties based on recorded seismic data, are commonly used in seismic exploration [4,5]. Currently, full waveform inversion (FWI) has been one popular research topic to derive the parameters of the earth [6–11]. In this method, both kinematic (traveltime, phase) and dynamic (amplitude) information of seismic data are employed simultaneously; therefore, high-resolution imaging results at about half of a wavelength are expected [12]. However, FWI has traditionally some drawbacks, such as non-uniqueness due to many local minima of the objective function and very large computing costs. Typically, we can prevent the solution from falling into local minima and obtain reliable inversion results only when the initial model is a good approximation of true geological models. However, it is a difficult task to evaluate whether the initial model approximates the true model in practical applications. Therefore, it is important to develop a method that has a large convergence domain to limit local minima, constraint the non-uniqueness and reduce the dependence of the inversion result on the initial model.

Data assimilation describes a class of methods, which simultaneously apply effective information provided by various observations to estimate initial values, boundary values, model parameters, and other parameters of the model space [13]. Data assimilation was first introduced by Kalman [14]. Gandin [15] proposed an optimal interpolation method and applied it in weather forecasts. In the past few decades, researchers have developed an optimal control theory for partial differential equations and a type of data assimilation methods based on the adjoint method [16–18]. Nowadays, data assimilation methods have been widely used for numerical simulations in oceanography and meteorology [19–22]. With the significant advancement in numerical simulation and particularly parallel computing [23–26], the capacity of available data and the resolution for data assimilation have been significantly improved.

Broadly, the data assimilation methods can be divided into two categories: a four-dimensional variational method based on the variational principle [27,28] and a statistical Kalman filter method based on Bayesian theory [14,29]. The four-dimensional variational method results in low root-mean-square (RMS) errors but generally has large computing costs [30,31]. The Kalman filter method updates the posterior distribution of a system state at each observation time [30]. It performs well for linear systems subject to Gaussian noise but poorly for high-dimensional nonlinear systems [32]. To overcome this drawback, a series of improved methods have been developed, including the application of nonlinear models to produce an extended Kalman filter [33], the utilization of the ensemble statistics to obtain an ensemble Kalman filter (EnKF) [34–36], and the model approximation by a statistically equivalent

* Corresponding author.
E-mail address: ydh@mail.tsinghua.edu.cn (D. Yang).

https://doi.org/10.1016/j.scib.2019.01.021
2095-9273/

linear system to obtain a Fourier diagonal filter [37,38], etc. EnKF was first proposed by Evensen [35], who applied the Monte Carlo method to represent the probability density function and used a random model to simulate the evolution of the function over time to reduce computing costs.

Methods used in seismology are very similar to those used to study the Earth's atmosphere, oceans, and other natural systems, which are all based on the study of equations describing the evolution of a system, and using sample data of a computational domain to obtain global information. In addition, the EnKF is a joint estimation method of the parameters and the state; therefore, the method has a good constraint to the non-uniqueness and a large convergence domain is expected. However, to the best of our knowledge, there are no publications on the application of EnKF in seismic inversion. In this paper, we apply the EnKF to geophysical inversion and propose a new random FWI method called GEKF for simplicity.

It is well-known that traditional random methods often suffer from large computing costs and low convergence rates in statistical simulations because a large number of samples are needed to simulate the random system and not all the samples used are representative of the random method, especially for high-dimensional systems [39]. Fang [40], Fang and Zheng [41], and Wang and Fang [42] proposed a uniform design technique to solve this problem by designing samples that are representative of the experimental conditions. This method can improve the computational efficiency and reduce the number of tests. In the present study, we apply the uniform design technique [40–42] to an initial velocity model and use uniform sampling without replacement in the random GEKF method to solve the problems associated with FWI and improve both the computational efficiency and the convergence rate compared to the random GEKF method. In other words, by combining uniform sampling without replacement with the random GEKF method, we propose a new FWI method in the semi-random framework called simply GEKUS. The Kalman filter is used to overcome the strong dependence of traditional FWI on initial models and both the EnKF and uniform sampling without replacement are combined to reduce its computational cost and storage usage. We demonstrate with the use of numerical examples that the new GEKUS method expands the convergence domain, improves the convergence rate, and overcomes the non-uniqueness to some extent.

2. Methods

2.1. Kalman filter method in geophysical inversion

The traditional solution to parameter estimation is usually to find an estimate of the parameters that is close to the prior information, which results in a state close to the observed state. In this sense, the model is accurate except for the parameter errors [13]. We follow Evensen [13] and use the Kalman filter method in geophysical inversion to formulate the parameter estimation as a joint estimation of the parameters and the state. In this way, we can reduce the dependence of the inversion results on initial models.

We assume that $\boldsymbol{s}^{\mathrm{tr}}(\boldsymbol{x},t)=\begin{pmatrix}\boldsymbol{u}(\boldsymbol{x},t,\boldsymbol{p}^{\mathrm{tr}})\\ \boldsymbol{p}^{\mathrm{tr}}(\boldsymbol{x},t)\end{pmatrix}\in\mathbb{R}^{W+P}$ is the combination of the waveform $\boldsymbol{u}(\boldsymbol{x},t,\boldsymbol{p}^{\mathrm{tr}})\in\mathbb{R}^{W}$ and the parameter field $\boldsymbol{p}^{\mathrm{tr}}\in\mathbb{R}^{P}$ (such as velocity $\boldsymbol{c}$, density ρ, and so on), where $\boldsymbol{x}$ denotes the spatial position, t denotes the time, and the superscript tr denotes the true value. W and P are the dimensions of the waveform and parameter field, respectively. $\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x},t)=\begin{pmatrix}\boldsymbol{u}(\boldsymbol{x},t,\boldsymbol{p}^{\mathrm{for}})\\ \boldsymbol{p}^{\mathrm{for}}(\boldsymbol{x},t)\end{pmatrix}$ is a prior estimate of $\boldsymbol{s}^{\mathrm{tr}}(\boldsymbol{x},t)$, where the superscript for denotes the forecast of the true value. Let $\boldsymbol{d}(t)$ be the combination of the observed waveform $\boldsymbol{u}$ at the receivers $\boldsymbol{x}_{r_k},k=1,\cdots,R$, where R is the number of the receivers. We assume

$$\begin{cases}\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x},t)=\boldsymbol{s}^{\mathrm{tr}}(\boldsymbol{x},t)+\boldsymbol{e}(\boldsymbol{x},t),\\ \boldsymbol{d}(t)=\begin{pmatrix}\boldsymbol{u}^{\mathrm{tr}}(\boldsymbol{x}_{r_1},t)\\ \vdots\\ \boldsymbol{u}^{\mathrm{tr}}(\boldsymbol{x}_{r_R},t)\end{pmatrix}+\boldsymbol{r}(t),\\ \overline{\boldsymbol{e}(\boldsymbol{x},t)}=\boldsymbol{0},\overline{\boldsymbol{e}(\boldsymbol{x}_1,t)[\boldsymbol{e}(\boldsymbol{x}_2,t)]^T}=\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t),\\ \overline{\boldsymbol{r}(t)}=\boldsymbol{0},\overline{\boldsymbol{r}(t)[\boldsymbol{r}(t)]^{\mathrm{T}}}=\boldsymbol{C}_{\boldsymbol{rr}}(t),\\ \overline{\boldsymbol{e}(\boldsymbol{x},t)[\boldsymbol{r}(t)]^{\mathrm{T}}}=\boldsymbol{0},\end{cases}\tag{1}$$

where $\boldsymbol{e}(\boldsymbol{x},t)$ is the error vector of $\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x},t)$ and $\boldsymbol{r}(t)$ is the observation error vector. As they are not known, we can treat them as random vectors with a Gaussian distribution. $\overline{\boldsymbol{e}(\boldsymbol{x},t)}$ is the mean of the random function $\boldsymbol{e}(\boldsymbol{x},t)$ and $\overline{\boldsymbol{r}(t)}$ is the mean of $\boldsymbol{r}(t)$.

$$\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)=\begin{pmatrix}\boldsymbol{C}_{uu}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t) & \boldsymbol{C}_{up}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)\\ \boldsymbol{C}_{pu}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t) & \boldsymbol{C}_{pp}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)\end{pmatrix}\in\mathbb{R}^{(W+P)\times(W+P)}\tag{2}$$

is the covariance matrix of $\boldsymbol{e}(\boldsymbol{x}_1,t)$ and $\boldsymbol{e}(\boldsymbol{x}_2,t)$ with respect to space, where $\boldsymbol{C}_{\boldsymbol{uu}}^{\mathrm{for}}\in\mathbb{R}^{W\times W}$, $\boldsymbol{C}_{\boldsymbol{pp}}^{\mathrm{for}}\in\mathbb{R}^{P\times P}$, and $\boldsymbol{C}_{\boldsymbol{up}}^{\mathrm{for}}\in\mathbb{R}^{W\times P}$ are the error covariance matrix of the waveform, the error covariance matrix of the parameters, and the error covariance matrix between the waveform and the parameters, respectively. According to Eq. (1), $\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x},t)$ is the unbiased estimation of $\boldsymbol{s}^{\mathrm{tr}}(\boldsymbol{x},t)$ with the error covariance matrix as $\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)$. The error covariance matrix of $\boldsymbol{d}(t)$ is $\boldsymbol{C}_{\boldsymbol{rr}}(t)$. The models and observations are independent. We desire to obtain an optimal estimate $\boldsymbol{s}^{\mathrm{a}}(\boldsymbol{x},t)$ of $\boldsymbol{s}^{\mathrm{tr}}(\boldsymbol{x},t)$, which minimizes

$$\begin{aligned}J[\boldsymbol{s}]=&\iint_D\left(\boldsymbol{s}(\boldsymbol{x}_1,t)-\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x}_1,t)\right)^{\mathrm{T}}\boldsymbol{V}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)\left(\boldsymbol{s}(\boldsymbol{x}_2,t)-\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x}_2,t)\right)\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{x}_2\\&+\left(\boldsymbol{d}(\boldsymbol{t})-\begin{pmatrix}\boldsymbol{u}(\boldsymbol{x}_{r_1},t)\\ \vdots\\ \boldsymbol{u}(\boldsymbol{x}_{r_R},t)\end{pmatrix}\right)^{\mathrm{T}}\boldsymbol{V}_{\boldsymbol{rr}}(t)\left(\boldsymbol{d}(\boldsymbol{t})-\begin{pmatrix}\boldsymbol{u}(\boldsymbol{x}_{r_1},t)\\ \vdots\\ \boldsymbol{u}(\boldsymbol{x}_{r_R},t)\end{pmatrix}\right),\end{aligned}\tag{3}$$

where $\boldsymbol{V}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)$ is the functional inverse of $\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)$, i.e.,

$$\int_D\boldsymbol{V}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_2,\boldsymbol{x}_3,t)\mathrm{d}\boldsymbol{x}_2=\delta(\boldsymbol{x}_1-\boldsymbol{x}_3),\tag{4}$$

δ is the Dirac function and $\boldsymbol{V}_{\boldsymbol{rr}}(t)$ is the inverse of $\boldsymbol{C}_{\boldsymbol{rr}}(t)$. We apply the Kalman filter method [13] to obtain both the optimal estimation $\boldsymbol{s}^{\mathrm{a}}(\boldsymbol{x},t)$ and its covariance matrix $\boldsymbol{C}_{ss}^{\mathrm{a}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)$ as

$$\begin{cases}\boldsymbol{s}^{\mathrm{a}}(\boldsymbol{x},t)=\boldsymbol{s}^{\mathrm{for}}(\boldsymbol{x},t)+\begin{pmatrix}\boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_1},\boldsymbol{x},t)\\ \vdots\\ \boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_R},\boldsymbol{x},t)\end{pmatrix}^{\mathrm{T}}\boldsymbol{q}(t)\left(\boldsymbol{d}(t)-\begin{pmatrix}\boldsymbol{u}^{\mathrm{for}}(\boldsymbol{x}_{r_1},t)\\ \vdots\\ \boldsymbol{u}^{\mathrm{for}}(\boldsymbol{x}_{r_R},t)\end{pmatrix}\right),\\ \boldsymbol{q}(t)=\left(\begin{pmatrix}\boldsymbol{C}_{uu}^{\mathrm{for}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_{r_1},t),\cdots,\boldsymbol{C}_{uu}^{\mathrm{for}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_{r_R},t)\\ \vdots\\ \boldsymbol{C}_{uu}^{\mathrm{for}}(\boldsymbol{x}_{r_R},\boldsymbol{x}_{r_1},t),\cdots,\boldsymbol{C}_{uu}^{\mathrm{for}}(\boldsymbol{x}_{r_R},\boldsymbol{x}_{r_R},t)\end{pmatrix}+\boldsymbol{C}_{\boldsymbol{rr}}(t)\right)^{-1},\\ \boldsymbol{C}_{ss}^{\mathrm{a}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)=\boldsymbol{C}_{ss}^{\mathrm{for}}(\boldsymbol{x}_1,\boldsymbol{x}_2,t)-\begin{pmatrix}\boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_1,t)\\ \vdots\\ \boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_R},\boldsymbol{x}_1,t)\end{pmatrix}^{\mathrm{T}}\boldsymbol{q}(t)\begin{pmatrix}\boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_2,t)\\ \vdots\\ \boldsymbol{C}_{us}^{\mathrm{for}}(\boldsymbol{x}_{r_R},\boldsymbol{x}_2,t)\end{pmatrix}.\end{cases}\tag{5}$$

After being discretized on numerical grids with the number of grid points as K, the waveform $\boldsymbol{u}(\boldsymbol{x},t,\boldsymbol{p})$, the observed data $\boldsymbol{d}(t)$, the state error vector $\boldsymbol{e}(\boldsymbol{x},t)$, the observation error vector $\boldsymbol{r}(t)$,

and the function $\boldsymbol{s}(\boldsymbol{x},t)$ can all be represented in vector forms $\boldsymbol{u}\in\mathbb{R}^{KW},\boldsymbol{d}\in\mathbb{R}^{RW},\boldsymbol{e}\in\mathbb{R}^{K(W+P)},\boldsymbol{r}\in\mathbb{R}^{RW},\boldsymbol{s}\in\mathbb{R}^{K(W+P)}$. Therefore, Eq. (1) can be rewritten in the following discrete forms:

$$\begin{cases}\boldsymbol{s}^{\text{for}}=\boldsymbol{s}^{\text{tr}}+\boldsymbol{e},\\ \boldsymbol{d}=\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{u}^{\text{tr}}+\boldsymbol{r},\\ \bar{\boldsymbol{e}}=\boldsymbol{0},\quad \overline{\boldsymbol{e}\boldsymbol{e}^{\text{T}}}=\boldsymbol{C}_{\boldsymbol{ss}}^{\text{for}},\\ \bar{\boldsymbol{r}}=\boldsymbol{0},\quad \overline{\boldsymbol{r}\boldsymbol{r}^{\text{T}}}=\boldsymbol{C}_{\boldsymbol{rr}},\\ \overline{\boldsymbol{e}\boldsymbol{r}^{\text{T}}}=\boldsymbol{0},\end{cases}\tag{6}$$

where $\boldsymbol{s}^{\text{tr}}$ is the true model, $\boldsymbol{s}^{\text{for}}$ is an estimate of $\boldsymbol{s}^{\text{tr}}$, $\boldsymbol{e}$ is the random error with mean zero and covariance matrix as

$$\boldsymbol{C}_{\boldsymbol{ss}}^{\text{for}}=\begin{pmatrix}\boldsymbol{C}_{\boldsymbol{uu}}^{\text{for}} & \boldsymbol{C}_{\boldsymbol{up}}^{\text{for}}\\ \boldsymbol{C}_{\boldsymbol{pu}}^{\text{for}} & \boldsymbol{C}_{\boldsymbol{pp}}^{\text{for}}\end{pmatrix}\in\mathbb{R}^{K(W+P)\times K(W+P)},\tag{7}$$

$\boldsymbol{M}_k$ is a $W\times KW$ matrix, $\boldsymbol{M}_k\boldsymbol{u}$ is the waveform at the k-th receiver, and $\boldsymbol{r}$ is the random observation error vector with mean zero and covariance matrix as $\boldsymbol{C}_{\boldsymbol{rr}}$.

In this case, we want to get the optimal estimate $\boldsymbol{s}^{\text{a}}$ which minimizes

$$J[\boldsymbol{s}]=(\boldsymbol{s}-\boldsymbol{s}^{\text{for}})^{\text{T}}\left(\boldsymbol{C}_{\boldsymbol{ss}}^{f}\right)^{-1}(\boldsymbol{s}-\boldsymbol{s}^{\text{for}})+\left(\boldsymbol{d}-\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{u}\right)^{\text{T}}(\boldsymbol{C}_{\boldsymbol{rr}})^{-1}\left(\boldsymbol{d}-\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{u}\right).\tag{8}$$

Following Evensen [13], we can obtain $\boldsymbol{s}^{\text{a}}$ and the error covariance matrix $\boldsymbol{C}_{\boldsymbol{ss}}^{a}$ as

$$\begin{cases}\boldsymbol{s}^{\text{a}}=\boldsymbol{s}^{\text{for}}+\boldsymbol{G}\left(\boldsymbol{d}-\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{u}^{\text{for}}\right),\\ \boldsymbol{C}_{\boldsymbol{ss}}^{\text{a}}=\boldsymbol{C}_{\boldsymbol{ss}}^{\text{for}}-\boldsymbol{G}\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{C}_{\boldsymbol{us}}^{\text{for}},\\ \boldsymbol{G}=\boldsymbol{C}_{\boldsymbol{su}}^{\text{for}}\left(\boldsymbol{M}_1^{\text{T}},\cdots,\boldsymbol{M}_R^{\text{T}}\right)(\boldsymbol{C}+\boldsymbol{C}_{\boldsymbol{rr}})^{-1},\\ \boldsymbol{C}=\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\boldsymbol{C}_{\boldsymbol{uu}}^{\text{for}}\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}^{\text{T}},\end{cases}\tag{9}$$

where matrix $\boldsymbol{G}\in\mathbb{R}^{K(W+P)\times RW}$ is called the Kalman gain filter.

2.2. GEKF method

In the above procedure of the Kalman filter method, the forward integration of the error covariance matrix $\boldsymbol{C}_{\boldsymbol{ss}}^{\text{a}}$ in Eq. (9) requires large computer storage. The EnKF [13] generates several samples $\boldsymbol{s}_j$ ($j\in\{1,\cdots,N\}$) of $\boldsymbol{s}$ with a Gaussian distribution. It uses the mean value of the random samples to represent the true value and uses the error covariance matrix of the samples to represent the error covariance matrix. If the sample size approximates infinite, the EnKF would be identical to the Kalman filter method [13]. Samples of finite size will give an approximation of the mean and error covariance matrix of the state. In this way, we only need to store N samples $\boldsymbol{s}_j$ instead of the error covariance matrix $\boldsymbol{C}_{\boldsymbol{ss}}^{\text{a}}$ in Eq. (9). We can focus the sample size N to be less than the dimension of $\boldsymbol{s}$ to reduce both the computing costs and the storage requirements.

In this section, we propose an EnKF method applicable to geophysical inversion; the proposed method is called GEKF. The specific implementation of this method is as follows: first, we generate N parameter samples $\boldsymbol{p}_j$ ($j\in\{1,\cdots,N\}$) based on some prior information, from which we can compute the theoretical waveform $\boldsymbol{u}_j$ by solving the seismic wave-equation and thus we get $\boldsymbol{s}_j=\begin{pmatrix}\boldsymbol{u}_j\\ \boldsymbol{p}_j\end{pmatrix}$. Then we approximate the true value $\boldsymbol{s}^{\text{tr}}=\begin{pmatrix}\boldsymbol{u}^{\text{tr}}\\ \boldsymbol{p}^{\text{tr}}\end{pmatrix}$ by the mean of the samples $\bar{\boldsymbol{s}}$ and approximate the error covariance matrix of the true value by the error covariance matrix of the samples. Hence, the error covariance matrix $\boldsymbol{C}$ in Eq. (9) can be represented as

$$\boldsymbol{C}=\overline{\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}\left(\boldsymbol{u}^{\text{for}}-\overline{\boldsymbol{u}^{\text{for}}}\right)\left(\boldsymbol{u}^{\text{for}}-\overline{\boldsymbol{u}^{\text{for}}}\right)^{\text{T}}\left(\boldsymbol{M}_1^{\text{T}},\cdots,\boldsymbol{M}_R^{\text{T}}\right)}.\tag{10}$$

Suppose $\boldsymbol{U}=\begin{pmatrix}\boldsymbol{M}_1\\ \vdots\\ \boldsymbol{M}_R\end{pmatrix}(\boldsymbol{u}_1^{\text{for}},\cdots,\boldsymbol{u}_N^{\text{for}})$ is a matrix formed by the waveform of the samples $\boldsymbol{u}_j^{\text{for}}$ at the receivers. To obtain the error covariance matrix $\boldsymbol{C}$, the following error matrix

$$\boldsymbol{U}'=\boldsymbol{U}(\boldsymbol{I}-\boldsymbol{1}_N)\tag{11}$$

is usually obtained first, where $\boldsymbol{1}_N$ is an $N\times N$ matrix whose elements are equal to 1. Then Eq. (10) can be represented by

$$\boldsymbol{C}=\frac{1}{N-1}\boldsymbol{U}'\boldsymbol{U}'^{\text{T}}.\tag{12}$$

In general, the samples in the ensemble may be high-dimensional and the ensemble may contain many redundant features that contribute to large computing costs and storage requirement for matrix $\boldsymbol{C}$. To extract the main characteristics of the original ensemble of samples, a principal component analysis (PCA) is used to reduce the dimension. We derive a new ensemble of vectors $\alpha_1^{\text{T}}\boldsymbol{U},\alpha_2^{\text{T}}\boldsymbol{U},\cdots,\alpha_p^{\text{T}}\boldsymbol{U}$, which are the linear combination of the original samples. The samples of the new ensemble contain as much information as possible and have the lowest possible dimension. According to Jolliffe [43], $\alpha_1,\alpha_2,\cdots,\alpha_p$ are unit orthogonal and are the solutions to the following problem:

$$\begin{cases}\max_{\alpha_1}\left\{\frac{\alpha_1^{\text{T}}\boldsymbol{C}\alpha_1}{\alpha_1^{\text{T}}\alpha_1}\right\},\\ \max_{\alpha_1^{\text{T}}\alpha_2=0}\left\{\frac{\alpha_2^{\text{T}}\boldsymbol{C}\alpha_2}{\alpha_2^{\text{T}}\alpha_2}\right\},\\ \vdots\\ \max_{\alpha_i^{\text{T}}\alpha_p=0,i=1,\cdots,p-1}\left\{\frac{\alpha_p^{\text{T}}\boldsymbol{C}\alpha_p}{\alpha_p^{\text{T}}\alpha_p}\right\}.\end{cases}\tag{13}$$

According to the definition of singular values, we know that these maxima are the first p singular vectors of $\boldsymbol{C}$ and $\alpha_1,\alpha_2,\cdots,\alpha_p$ are the first p left singular vectors of $\boldsymbol{C}$. Hence, we follow Pham [44] and represent the singular value decomposition (SVD) of $\boldsymbol{U}'$ as

$$\boldsymbol{U}'=\boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{Q}^{\text{T}},\tag{14}$$

where $\boldsymbol{L}$ and $\boldsymbol{Q}$ are the left and right singular vectors of $\boldsymbol{U}'$ and

$$\boldsymbol{\Sigma}=\text{diag}(\sigma_1,\cdots,\sigma_N)\tag{15}$$

is a matrix ensemble with the singular values satisfying

$$\sigma_1\geqslant\cdots\geqslant\sigma_N.\tag{16}$$

We select the first p singular values and let

$$\tilde{\boldsymbol{\Sigma}} = \operatorname{diag}(\sigma_1, \cdots, \sigma_p, 0, \cdots, 0). \tag{17}$$

Therefore, Eq. (12) can be approximated by

$$\tilde{\boldsymbol{C}} = \frac{1}{N-1}\boldsymbol{L}\tilde{\boldsymbol{\Sigma}}^2\boldsymbol{L}^{\mathrm{T}}. \tag{18}$$

We use the Lapack package to apply SVD in Eq. (14). The package allows us to only store the first p singular vectors of $\boldsymbol{L}$ instead of the matrix $\tilde{\boldsymbol{C}}$ which allows the storage requirement being greatly reduced without significant loss of accuracy. That makes the method being suitable for large-scale problems.

In practice, we can determine p using the following criterion:

$$\frac{\sum_{i=1}^{p}\sigma_i}{\sum_{i=1}^{N}\sigma_i} \geqslant T, \tag{19}$$

where $T \in (0, 1)$ is a threshold given in advance.

We can also think of observations as random variables with true observations as the mean and $\boldsymbol{C}_{\boldsymbol{rr}}$ as the error covariance matrix. Then, we can define an ensemble of observations, where each element of the ensemble can be written as:

$$\boldsymbol{d}_j = \boldsymbol{d} + \boldsymbol{r}_j, \quad j = 1, \cdots, N, \tag{20}$$

where $\boldsymbol{d}_j$ represents the j-th observation, $\boldsymbol{r}_j$ represents the error of the j-th observation, and $\boldsymbol{d}$ is the true observation. The mean of $\boldsymbol{r}_j$ is assumed to be zero and the covariance matrix of $\boldsymbol{r}_j$ is

$$\boldsymbol{C}_{\boldsymbol{rr}}^{\mathrm{e}} = \overline{\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}}}. \tag{21}$$

By applying Eq. (9) to each sample with the error covariance matrix $\tilde{\boldsymbol{C}}$ approximated by Eq. (18) and the observations $\boldsymbol{d}_j$ approximated by Eq. (20), we obtain

$$\boldsymbol{p}_j^{\mathrm{a}} = \boldsymbol{p}_j^{\mathrm{for}} + \boldsymbol{C}_{\boldsymbol{pu}}^{\mathrm{for}} \begin{pmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_R \end{pmatrix}^{\mathrm{T}} \left(\frac{1}{N-1}\boldsymbol{L}\tilde{\boldsymbol{\Sigma}}^2\boldsymbol{L}^{\mathrm{T}} + \boldsymbol{C}_{\boldsymbol{rr}}^{\mathrm{e}}\right)^{-1} \left(\boldsymbol{d}_j - \begin{pmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_R \end{pmatrix} \boldsymbol{u}_j^{\mathrm{for}}\right). \tag{22}$$

The average of all samples $\bar{\boldsymbol{p}}^{\mathrm{a}}$ is considered a model update.

Note that the matrix $\frac{1}{N-1}\boldsymbol{L}\tilde{\boldsymbol{\Sigma}}^2\boldsymbol{L}^{\mathrm{T}} + \boldsymbol{C}_{\boldsymbol{rr}}^{\mathrm{e}}$ is symmetric and positive definite; therefore, its inverse always exists. We can use the conjugate gradient (CG) method to evaluate Eq. (22).

2.3. GEKF with uniform sampling without replacement

The GEKF adds random perturbations when generating samples of initial velocity models. Due to the inherent uncertainty of random methods, randomly generated samples may not be representative. It may result in uncontrolled errors and the inversion may even diverge. Especially for a high-dimensional model space, the randomly selected finite samples can easily misrepresent the information of the model. To obtain more representative samples, we increase the sample size of traditional random methods, which in turn increases the computing costs and storage requirements. In our study, when generating samples of initial models, we apply a uniform design technique [40–42] and use uniform sampling without replacement in the semi-random framework to make the samples more representative which can reduce the uncertainty of the inversion results.

Suppose an initial random parameter model is $\boldsymbol{p} = (p^1, \cdots, p^P)^{\mathrm{T}}$, where P is the dimension of the parameter field. First, we generate N samples of components p_j^k ($j \in \{1, \cdots, N\}$), which are respectively the $\frac{j-1/2}{N}$ quantile of the random component p^k ($k \in \{1, \cdots, P\}$). To generate each sample $\boldsymbol{p}_j$, we can choose one sample of components $p_{j_k}^k$ randomly from the samples of the components and combine them to create a parameter model sample $\boldsymbol{p}_j = \left(p_{j_1}^1, \cdots, p_{j_P}^P\right)^{\mathrm{T}}$. In this procedure, we can apply the sampling without replacement method. After one sample of the component is chosen, it will not be chosen again during the subsequent process. That is, we must keep $p_{j_k}^k \neq p_{l_k}^k$ for any $k \in \{1, \cdots, P\}$, if $j \neq l$. In this way, we can generate N samples that are more representative than randomly generated samples.

Furthermore, a layer-stripping strategy [45] is used in our study to obtain the point-by-point inversion. In other words, if the value of a parameter on a grid remains the same after several iterations, it is assumed its value is known and is removed from the parameter model $\boldsymbol{p}$ in the following iterations. This point-by-point inversion algorithm can reduce the scale of the problem and results in fast convergence.

The computing cost of an inversion problem is mainly driven by the number of forward computations. Therefore, we concentrate on finding a way to reduce the number of forward computations. Krebs et al. [46] proposed the idea of encoded simultaneous-source FWI (ESSFWI), which defines the objective function of the inversion as

$$h(\boldsymbol{u}(\boldsymbol{p}), \boldsymbol{p}) = \left\| \sum_{n=1}^{N_s} \boldsymbol{q}_n \otimes \boldsymbol{u}(\boldsymbol{p}, \boldsymbol{f}_n) - \sum_{n=1}^{N_s} \boldsymbol{q}_n \otimes \boldsymbol{d}_n \right\|^2, \tag{23}$$

where $\boldsymbol{q}_n$ is the encoding sequence in time, $\boldsymbol{f}_n$ is the source function, N_s is the number of sources and $\otimes$ represents a convolution with respect to time. As the waveform $\boldsymbol{u}$ is a linear function of the sources, the objective function can be expressed as [46]

$$h(\boldsymbol{u}(\boldsymbol{p}), \boldsymbol{p}) = \left\| \boldsymbol{u}\left(\boldsymbol{p}, \sum_{n=1}^{N_s} \boldsymbol{q}_n \otimes \boldsymbol{f}_n\right) - \sum_{n=1}^{N_s} \boldsymbol{q}_n \otimes \boldsymbol{d}_n \right\|^2, \tag{24}$$

where $\sum_{n=1}^{N_s} \boldsymbol{q}_n \otimes \boldsymbol{f}_n$ can be regarded as a super source formed by coupling all original sources. We define the square of the difference between the observed data using this super-seismic source and the theoretical waveform as the objective function for inversion. The computing costs can be significantly reduced because the method runs the simulator one time with all sources simultaneously. Krebs et al. [46] used numerical experiments to prove that it is most efficient to use randomly +1 or −1 code of length 1 as the encoding sequence. We adopt this encoding method in our inversion method.

We combine GEKF with the uniform sampling without replacement strategy and develop a new semi-random FWI method called simply GEKUS. We also apply the point-by-point inversion technique and the ESSFWI algorithm in the new GEKUS method to further increase the computational efficiency. The specific implementation procedure is summarized below:

(1) Based on the prior information on the models, create several model samples by applying uniform sampling without replacement. Each sample determines a model;
(2) Calculate the theoretical waveform for each model sample;
(3) Update the model parameters using Eq. (22) for each sample. The average of the inverted parameters corresponding to these samples is used as the updated model;
(4) Compute an objective function in the term of the updated model to determine if the iteration converges using Eq. (8). If it converges, the inversion stops; otherwise, go back to step 2.

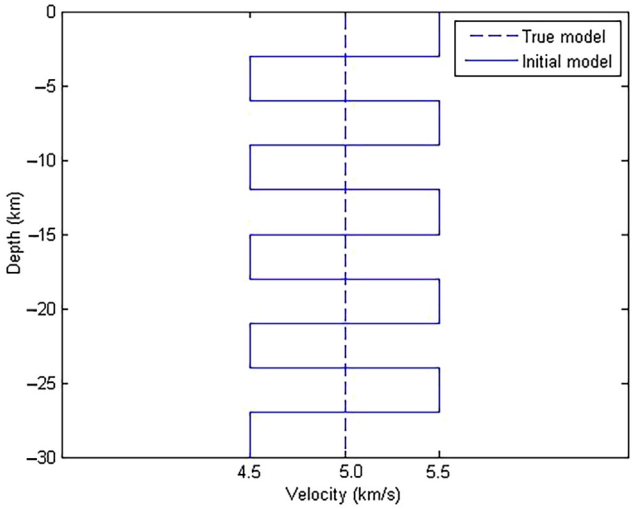

Fig. 1. 1D ten-parameter model.

3. Numerical examples

In this section, we choose six models including the one-dimensional (1D) model with ten parameters, IASP91 model, checkerboard model, a box model, the crustal root model, and the Marmousi model to investigate the convergence rate, convergence domain, computing costs, and noise sensitivity of the GEKUS method. In the inversion process, the optimal nearly-analytic discrete (ONAD) method proposed by Yang et al. [47,48] and Li et al. [49] is used to generate the theoretical waveform.

3.1. 1D model with ten parameters

In the first example, we choose a 1D model with ten parameters and investigate the convergence rate, convergence domain, and computing costs of the GEKUS method.

The computational range is from the surface to the depth of 30 km. The space and time steps in the forward simulation are 0.05 km and 0.003 s respectively. The source is located at the depth of 30 km and four receivers are placed at the depths of 1.25, 2.55, 3.75, and 5.05 km, respectively. The true model has alternating positive and negative perturbations relative to the background velocity of 5 km/s. 80 samples are created by uniform sampling without replacement with a mean velocity of 5 km/s, which is the same as the background on each inversion grid. Fig. 1 shows the initial model and the true model with 10% alternating positive and negative perturbations.

We repeat the experiment 100 times and use a 90% confidence level, that is, if the inverted model converges in more than 90% of the repeated experiments, we consider it converging to the true model. Table 1 shows the size of perturbations of the true model and the corresponding convergence rates. For comparison, we also use the GEKF method to perform the same experiment. The experimental data are also listed in Table 1. It is evident that when the perturbation size reaches 10%, the convergence rate of the GEKUS reaches 90%, whereas the convergence rate of the random method is 84%. The results show that the convergence domain of the GEKUS is higher than that of the GEKF. Table 2 shows the comparison of the average computing costs using an Intel(R) Core(TM) i7-4790 CPU for the numerical experiments. For the divergent numerical examples, we only iterate 100 times. Tables 1 and 2 show that, compared with the GEKF, the GEKUS has higher convergence rates and saves 28.5% CPU time indicating that the new method based on uniform sampling has a larger convergence domain and converges faster than the random method.

Table 1
Comparison of the convergence rates for different initial perturbations for the 1D ten-parameter model.

Perturbation size (%)	Convergence rates of the GEKUS method (%)	Convergence rates of the GEKF method (%)
5	100	100
6	98	89
8	94	85
10	91	84
12	85	60

Table 2
Comparison of the computing costs for different initial perturbations for the 1D ten-parameter model.

Perturbation size (%)	Computing costs of the GEKUS method (s)	Computing costs of the GEKF method (s)
5	81.15	140.39
6	113.75	191.10
8	162.62	246.92
10	342.45	364.63
12	352.21	437.52

3.2. IASP91 model

Kennett and Engdahl [50] presented the IASP91 model, which describes the velocity models of P- and S-waves in an inhomogeneous earth model. In this example, we choose the P-wave velocity model with the depth range of 0–800 km to test the effectiveness of the GEKUS for the 1D inhomogeneous model. To generate the theoretical waveform, we choose a spatial increment of 0.5 km and a time step of 0.02 s for solving the wave equation and the inversion grid of 10 km, resulting in 80 unknown parameters that need to be inverted. The source is located at the depth of 800 km. 15 receivers are placed with a spacing of 0.5 km from the surface to the depth of 8 km and 100 sample models are used in the inversion experiment. The true model is shown in Fig. 2 marked with the blue line. The initial model is based on the true model with 6% positive or negative perturbations as marked with the dashed black line in Fig. 2.

The inversion result is the solid line shown in Fig. 2. It is evident that this result is very close to the true model. For comparison, we also run the GEKF to the same problem. Even though the inversion result is close to that of the GEKUS, their computing costs are quite different. The GEKUS took 2393.5 s on an Intel Xeon e5 2670 to obtain the result shown in Fig. 2, whereas the GEKF took 3396.1 s to obtain the result, which implies that the GEKUS can save 30% CPU time compared to the GEKF method to produce the similar quality results for this 1D inhomogeneous model.

3.3. 2D checkerboard model

In the third experiment, we test the convergence domain of the GEKUS method for a 2D checkerboard model, which covers a computational domain of 12 km × 20 km. The space and time steps in the forward simulation are 0.2 km and 0.015 s, respectively. A perfectly matched layer (PML) boundary condition [51,52] is used outside the computational domain. 20 receivers with a spacing of 1 km and 50 sources with a spacing of 0.4 km are placed on the surface. We encode the 50 sources into 5 super-sources based on the ESSFWI method [46]. A Ricker wavelet with a dominant frequency of 0.9 Hz is used as the source function. The true velocity model has alternating positive and negative perturbations of 10% with an average velocity of 2 km/s; 300 samples are produced by uniform sampling without replacement with a mean velocity of

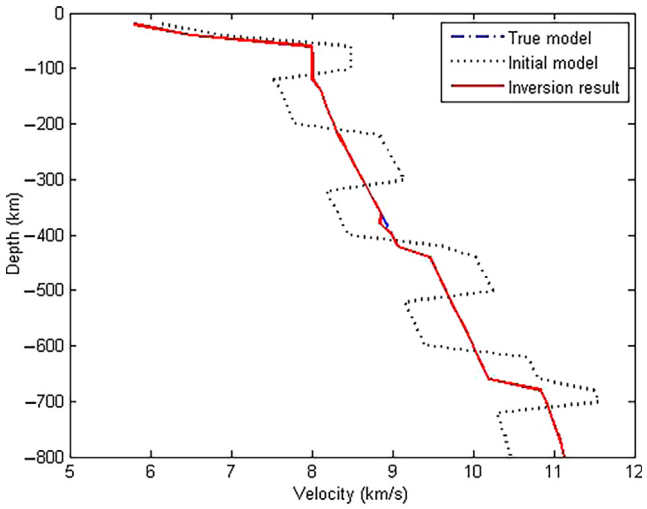

Fig. 2. (Color online) The IASP91 model [50] with a range of 0–800 km and its inversion result.

2 km/s on each inversion grid. The true model is shown in Fig. 3a, where the white circles represent positive perturbations and the black circles represent negative perturbations. The circle size indicates the size of the perturbation.

The inversion result of the GEKUS method is shown in Fig. 3b, which demonstrates that the new FWI method converges accurately to the true model. For comparison, we repeat the same experiment using the traditional FWI method [9], which employs the least square misfit function as follows:

$$\chi(\boldsymbol{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \| \boldsymbol{s}(\boldsymbol{x}_r, t, \boldsymbol{m}) - \boldsymbol{d}(\boldsymbol{x}_r, t) \|^2 \mathrm{d}t, \tag{25}$$

and the adjoint method is applied to obtain the optimal model parameters. The inversion result is shown in Fig. 3c. A comparison between Fig. 3b and c shows that the new GEKUS provides an accurate result (Fig. 3b) for the 10% velocity perturbation, whereas the traditional FWI method is less effective in such case.

To further speed up the inversion, we divide the computational domain into 2 layers and apply the layer-stripping technique [45] in the numerical experiment. As the number of parameters in each iteration is much less, the computing costs are reduced. Numerical experiments show that the GEKUS with the layer-stripping technique took 77534.18 s of CPU time on a TianHe-1 (A) with Intel Xeon X5670 whereas, without the layer-stripping, it took 178012.8 s of CPU time. This result implies that the use of the layer-stripping technique increases the computational speed roughly 1.3 times for the GEKUS method in this experiment.

3.4. A 2D box model

In the fourth example, we choose a 2D box model (Fig. 4a) to test the validity of the GEKUS for velocity anomalies and noise sensitivity. The computational domain is 2 km × 2 km. The space and time steps in the forward simulation are 0.02 km and 0.003 s, respectively. The PML boundary condition is used outside the computational domain. The background velocity is 3.5 km/s and it is 3.2 km/s in the center of the model with a range of 0.4 km × 0.4 km. 25 sources with a spacing of 0.08 km are placed at the bottom and 50 receivers with a spacing of 0.04 km are placed on the surface. A Ricker wavelet with a dominant frequency of 2.4 Hz is used as the source function. 300 samples are produced by uniform sampling without replacement with a mean velocity of 3.5 km/s for the background on each inversion grid. We choose

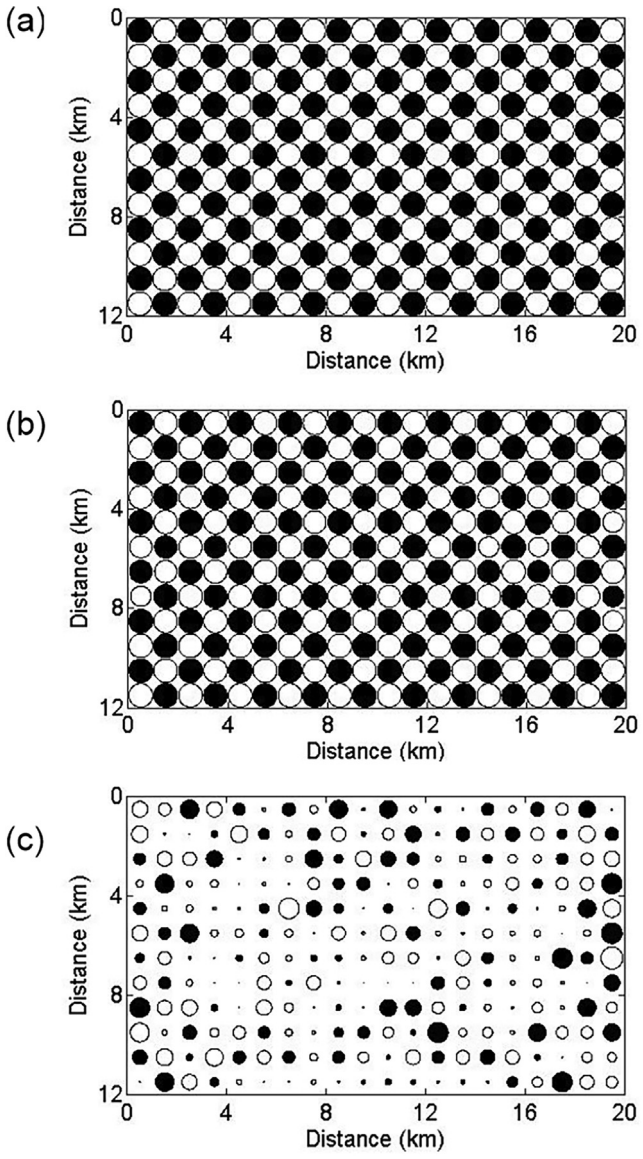

Fig. 3. 2D checkerboard model. (a) True velocity model; (b) inversion result generated by the GEKUS; (c) inversion result generated by the traditional FWI method.

a source and receiver pair located in the middle of the computational domain and plot the synthetic and observed waveforms in Fig. 4b. The results show no cycle-skipping between the two data. The inversion results generated by the GEKUS and the traditional FWI method are shown in Fig. 4c and d, respectively. It is clear that both methods perform well for the velocity anomaly model.

In the following, we choose an anomaly model with a velocity of 2.95 km/s, which has a perturbation of about 15.7% relative to the background. The synthetic and observed waveforms are shown in Fig. 4e; clearly cycle-skipping occurs in this case. The inversion results generated by the GEKUS and the traditional FWI method are shown in Fig. 4f and g, respectively. The GEKUS method provides an accurate result but the traditional FWI method fails to restore the velocity anomaly, which confirms that the GEKUS has a wider convergence domain than the traditional FWI does, even when the initial model causes cycle-skipping. For this model, we also choose 3.65 km/s as the mean of the samples on each inversion grid for the GEKUS method and the initial background velocity for the traditional FWI method, which has a bias of about 4.3% from

the true background velocity. The synthetic waveforms using the true and perturbed velocities are shown in Fig. 4h and it is evident that cycle-skipping still occurs. The inversion results generated by the GEKUS and the traditional FWI method are shown in Fig. 4i and j, respectively. The GEKUS method still performs well for restoring the velocity anomaly although the average of the samples is not exactly the same as the true background; however, the traditional FWI method fails completely.

Finally, to test the sensitivity to noise, we repeat the experiment for the model shown in Fig. 4a and add 6% random noise relative to the true observation. Fig. 4k and l show the inversion results generated by the GEKUS and the traditional FWI method, respectively. We can see that both methods invert the velocity anomaly using the data with random noise but the result of the GEKUS is more accurate (Fig. 4k), indicating that the GEKUS method is less sensitive to noises, and well suited for noisy data.

3.5. Crustal root model

In the fifth example, we choose the 2D crustal root model (Fig. 5a) obtained by slightly modifying the S-wave velocity of the IASP91 model [50]. The computational domain is 75 km × 100 km. The space and time steps are 1 km and 0.03 s, respectively. The model consists of a one-layer mantle and two-layer crusts with the S-wave velocities being 3.36 and 3.75 km/s in the upper and lower crusts, respectively; the interface between the two-layer crusts is at a depth of 20 km. From the depth of 35–60 km, there is a curved interface with a 2nd order polynomial

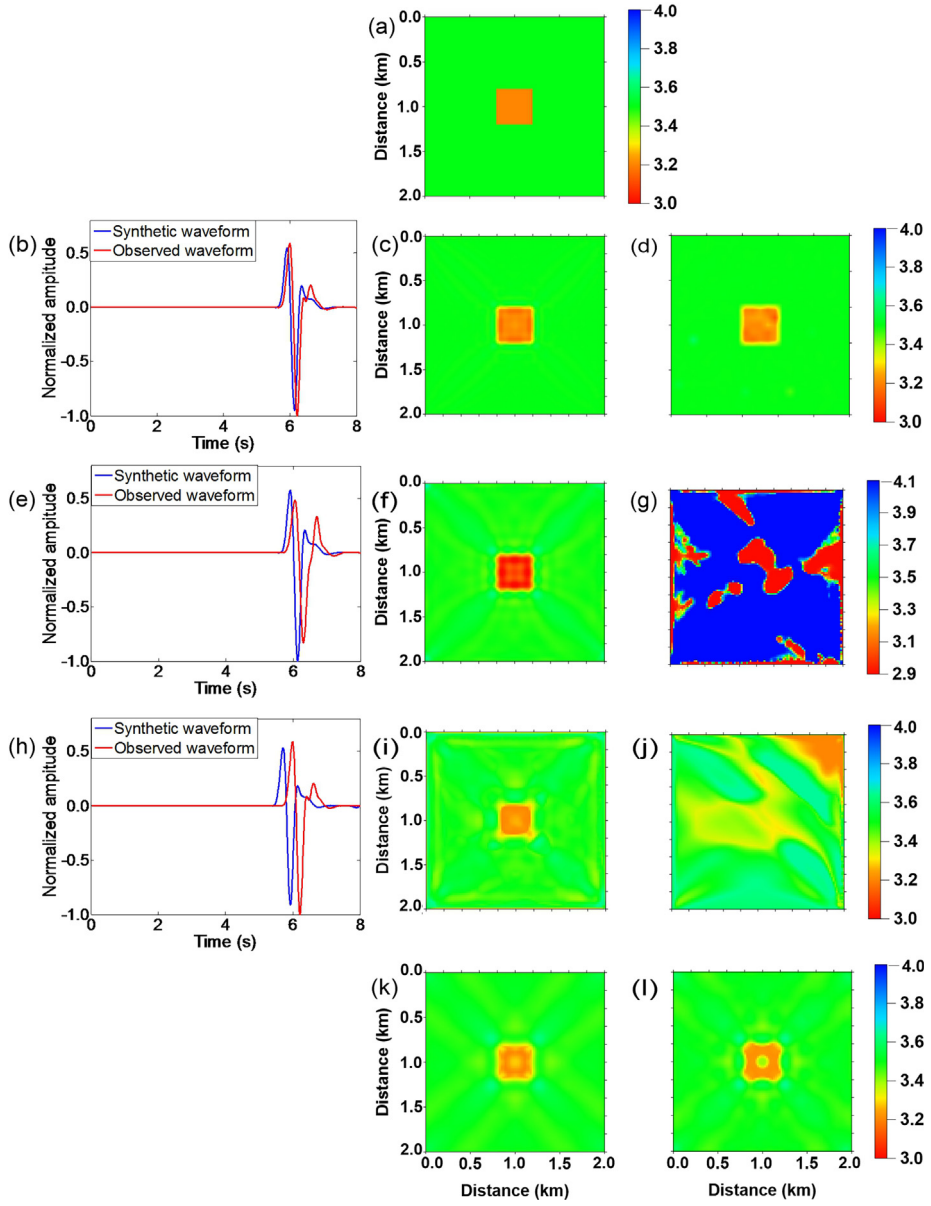

Fig. 4. 2D box model. (a) True velocity model with background velocity of 3.5 km/s and anomaly velocity of 3.2 km/s; (b) the synthetic and observed waveforms; (c) inversion result generated by the GEKUS; (d) inversion result generated by the traditional FWI method; (e) the synthetic and observed waveforms of the box model with an anomaly velocity of 2.95 km/s; (f) inversion result generated by the GEKUS for the box model with an anomaly velocity of 2.95 km/s; (g) inversion result generated by the traditional FWI method for the same anomaly model as (f); (h) the synthetic and observed waveforms of the box model for the background velocity of 3.65 km/s; (i) inversion result generated by the GEKUS for the box model and the sample mean value of 3.65 km/s on each inversion grid; (j) inversion result generated by the traditional FWI method for the background velocity of 3.65 km/s; (k) inversion result generated by the GEKUS for the same anomaly model as (a) using data with 6% random noise relative to the true observation; (l) inversion result generated by the traditional FWI method for the same case with (k).

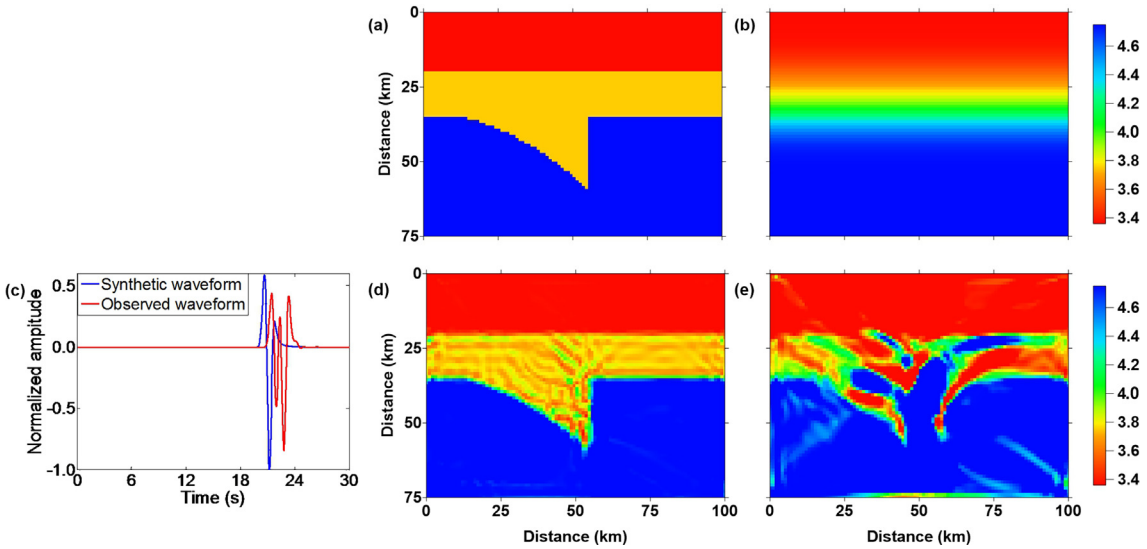

Fig. 5. Crustal root model. (a) True velocity model; (b) initial velocity model; (c) the synthetic and observed waveforms; (d) inversion result generated by the GEKUS; (e) inversion result generated by the traditional FWI method.

variation. The velocity in the mantle is 4.75 km/s, which is about 26% higher than that in the lower crust. We place 50 sources evenly at the bottom of the computational domain and 100 receivers evenly on the surface. A Ricker wavelet with a dominant frequency of 0.9 Hz is used as the source function. The initial model is obtained by smoothing the 1D IASP91 model 50 times using a moving average and is shown in Fig. 5b. 800 samples, which are produced by uniform sampling without replacement, are used in this example. We choose a source and receiver pair located in the middle of the computational domain and plot the waveforms generated using true and perturbed velocities (Fig. 5c); clearly the two waveforms are cycle-skipping.

The inversion results, generated using the GEKUS and the traditional FWI method are shown in Fig. 5d and e, respectively. We can see that the traditional FWI method does not achieve a correct result (Fig. 5e). In contrast, our method detects the velocity jump

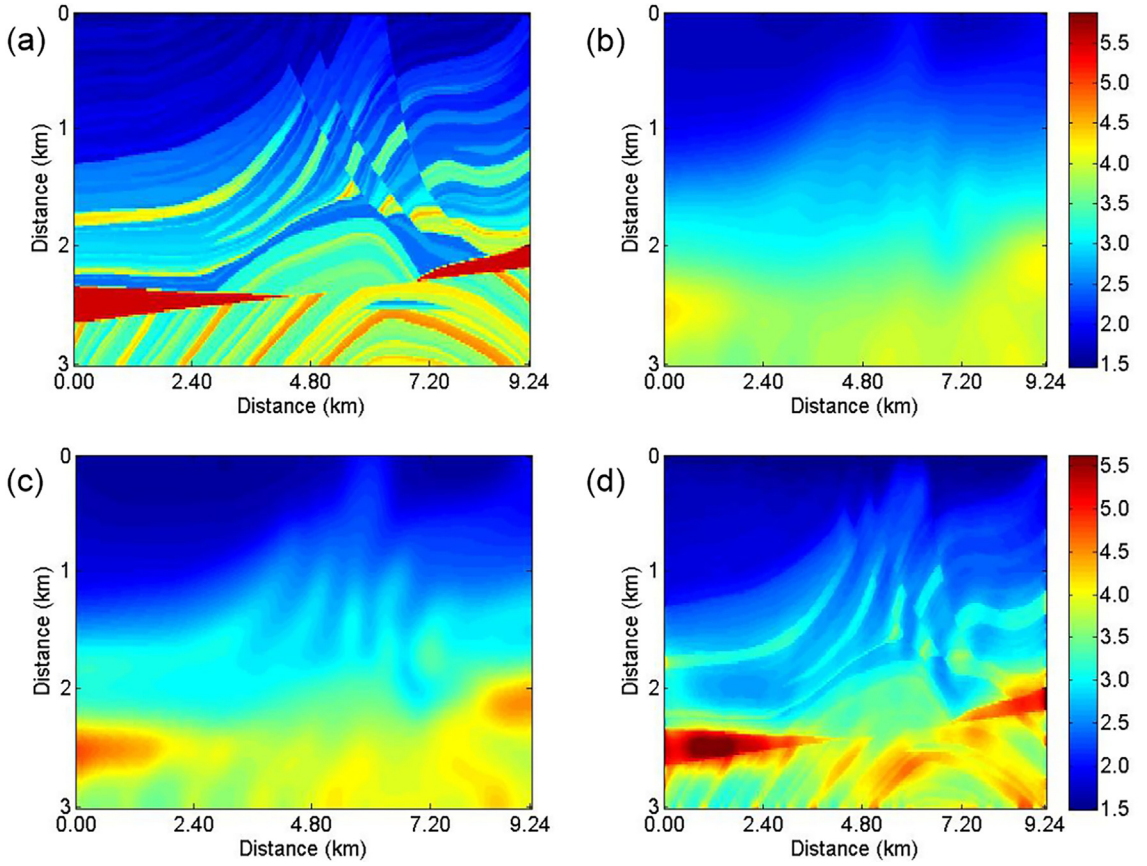

Fig. 6. The Marmousi model. (a) True velocity model; (b) the initial model; (c) inversion result generated by the traditional FWI method; (d) inversion result generated by the GEKUS method.

and provides a clear velocity structure for the crustal root model which indicates that the GEKUS provides a high-resolution result and has a larger convergence domain than the traditional FWI method does.

3.6. Marmousi model

The last example is the Marmousi model (Fig. 6a), which is more complex than the previous models. We subsample the velocity model and the number of grid points is 461×151. The initial model is created by smoothing the true model 150 times using a moving average; the result is shown in Fig. 6b. The space and time steps are 0.02 km and 0.0005 s, respectively. The PML boundary condition [51,52] is used outside the model domain; 116 sources and 231 receivers are evenly placed on the surface. A Ricker wavelet with a dominant frequency of 10 Hz is used as the source function. 1,500 samples, which are produced by uniform sampling without replacement, are used in this experiment.

First, we apply the traditional FWI to the model and the result is shown in Fig. 6c. We can see that the traditional FWI does not obtain an accurate velocity structure, especially for the lower part because the initial model is smoothed and laterally almost constant. However, the GEKUS method provides a more accurate result (Fig. 6d) than the traditional FWI does. This result again confirms that the GEKUS method can be applied to complex models and obtains high-resolution inversion results.

4. Discussions and conclusions

We propose and test a new full waveform inversion method in a semi-random framework based on the EnKF and uniform sampling without replacement, namely the GEKUS method, where the EnKF is applied to reduce the strong dependence of traditional FWIs on the initial models, and a uniform sampling technique is also introduced to reduce the computing costs. The numerical experiments on various models show that the GEKUS requires about 70% of the computational time of the GEKF and achieves good convergence efficiency. In addition, the GEKUS solves only the forward wave equations in each model update and can be easily implemented with parallel computing. The numerical examples also confirm that the GEKUS has a larger convergence domain than the traditional FWI method does; which suggests that the new method can reduce the strong dependence of FWI on initial models. The method can prevent the inversion result from falling into local minima and has a good constraint to non-uniqueness of FWIs to some extent. Numerically, it is a natural parallel algorithm with high parallel efficiency; furthermore, by introducing the layer-by-layer strategy, the computing costs can be largely reduced.

It is worthwhile to note that there is no need for the GEKUS to solve the adjoint equation as required by the traditional FWI method. As a result, the GEKUS can save the computing costs of the backward propagation of the traditional FWI. However, we must obtain all waveforms of different samples for the GEKUS. Fortunately, because different samples are independent, we can easily implement the parallel computation of the GEKUS, which will greatly increase the computational efficiency of the GEKUS. In addition, if the inversion results do not converge to true models in the FWI, we have to choose another initial model to repeat the inversion process although this does not guarantee convergence. In this case, it will require large CPU costs. In this sense, because our method has a large convergence domain, there is a good potential that the same initial models converge to the true models. As a result, the total computing costs of the new method are comparable to those of the traditional FWI method. As shown by Evensen [13], the EnKF method can be treated as a method of joint estimation of the parameters and the resulting waveforms. Therefore, the GEKUS can be applied to inverse all the parameters including those in the seismic wave equations. For example, the parameters included in the acoustic or elastic wave equations, such as density and the elastic parameters can also be potentially inverted by the GEKUS. The extension of the proposed method to multiple parameter inversion and the 3D case is straightforward which will be investigated in the future.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

We would like to thank Dr. Enru Liu from ExxonMobil Upstream Research Co., USA, Dr. Faqi Liu from Petroleum Geo-Services, USA (PGS), and Dr. Xueyuan Huang from Beijing Technology and Business University for their comments. The authors wish to thank the two anonymous reviewers for their comments that helped to improve the original manuscript. This work was supported by the National Key Research and Development Program (2017YFC1500301) and the National Natural Science Foundation of China (U1839206, 91730306).

References

[1] Tang C, Bao C, Li S, et al. A stress-rifting origin of grand Canyon. Sci Bull 2016;61:495–504.
[2] Zhang J, Chen J. Geophysical implications for the formation of the Tamu Massif-the Earth's largest single volcano-within the Shatsky Rise in the northwest Pacific Ocean. Sci Bull 2017;62:69–80.
[3] Han L, Wu Z, Jiang C, et al. Properties of three seismic events in September 2017 in the northern Korean Peninsula from moment tensor inversion. Sci Bull 2017;62:1569–71.
[4] Aki K, Lee W. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model. J Geophys Res 1976;81:4381–99.
[5] Hauser J, Sambridge M, Rawlinson N. Multiarrival wavefront tracking and its applications. Geochem Geophys Geosyst 2008;9:Q11001.
[6] Tarantola A, Valette B. Generalized nonlinear inverse problems solved using the least squares criterion. Rev Geophys 1982;20:219–32.
[7] Tarantola A, Nercessian A. Three-dimensional inversion without blocks. Geophys J Int 1984;76:299–306.
[8] Gauthier O, Virieux J, Tarantola A. Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophysics 1986;51:1387–403.
[9] Tromp J, Tape C, Liu Q. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophys J Int 2005;160:195–216.
[10] Tong P, Zhao D, Yang D, et al. Wave-equation based traveltime seismic tomography-part 1: method. Solid Earth 2014;5:1151.
[11] Jing H, Yang D, Wu H. Dynamic inversion method based on the time-staggered stereo-modeling scheme and its acceleration. Geophys J Int 2016;207:1675–87.
[12] Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics. Geophysics 2009;74:WCC1-WCC26.
[13] Evensen G. Data assimilation: the ensemble Kalman filter. 2nd ed. Berlin: Springer; 2009.
[14] Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng 1960;82:35–45.
[15] Gandin LS. Objective analysis of meteorological fields. Leningrad: Gidrometeorol Izdat; 1963.
[16] Vogel CR. Computational methods for inverse problems. Philadelphia: the Society for Industrial and Applied Mathematics, 2002.
[17] Tarantola A. Inverse problem theory and methods for model parameter estimation. Philadelphia: the Society for Industrial and Applied Mathematics, 2005.
[18] Banks HT, Kunisch K. Estimation techniques for distributed parameter systems. Berlin: Springer; 2012.
[19] Bennett AF. Inverse methods in physical oceanography. Cambridge: Cambridge University Press; 1992.
[20] Malanotte-Rizzoli P. Modern approaches to data assimilation in ocean modeling. Amsterdam: Elsevier; 1996.
[21] Wunsch C. The ocean circulation inverse problem. Cambridge: Cambridge University Press; 1996.
[22] Wei M, Li Q, Xin X, et al. Improved decadal climate prediction in the North Atlantic using EnOI-assimilated initial condition. Sci Bull 2017;62:1142–7.
[23] Graves RW. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. Bull Seismol Soc Am 1996;86:1091–106.

[24] Akcelik V, Biros G, Ghattas O. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation. In: Proceedings of the 2002 ACM/IEEE Conference on Supercomputing. Baltimore, MD, USA; 2002, p. 41.
[25] Komatitsch D, Tromp J. Spectral-element simulations of global seismic wave propagation—I Validation. Geophys J Int 2002;149:390–412.
[26] Komatitsch D, Liu Q, Tromp J, et al. Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. Bull Seismol Soc Am 2004;94:187–206.
[27] Zupanski D. A general weak constraint applicable to operational 4DVAR data assimilation systems. Mon Weather Rev 1997;125:2274–92.
[28] Bennett AF. Inverse modeling of the ocean and atmosphere. Cambridge: Cambridge University Press; 2005.
[29] Harvey AC. Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press; 1990.
[30] Law KJ, Stuart AM. Evaluating data assimilation algorithms. Mon Weather Rev 2012;140:3757–82.
[31] Zhang M, Zhang F, Huang X, et al. Intercomparison of an ensemble Kalman filter with three-and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. Mon Weather Rev 2011;139:566–72.
[32] Snyder C, Bengtsson T, Bickel P, et al. Obstacles to high-dimensional particle filtering. Mon Weather Rev 2008;136:4629–40.
[33] Jazwinski AH. Stochastic process and filtering theory. Cambridge: Academic Press; 1970.
[34] Evensen G, Leeuwen PJ. Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. Mon Weather Rev 1996;124:85–96.
[35] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res: Oceans 1994;99:10143–62.
[36] Evensen G. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 2003;53:343–67.
[37] Harlim J, Majda A. Filtering nonlinear dynamical systems with linear stochastic models. Nonlinearity 2008;21:1281.
[38] Majda AJ, Harlim J, Gershgorin B. Mathematical strategies for filtering turbulent dynamical systems. Discret Contin Dyn Syst 2010;27:441–86.
[39] Xu Z, Yu K, Tresp V, et al. Representative sampling for text classification using support vector machines. In: Proceedings of the European conference on information retrieval. Berlin: Springer; 2003, p. 393–407.
[40] Fang K. Uniform design: an application of number-theoretic methods to experimental designs. Acta Mathapplsin 1980;3:363–72.
[41] Fang K, Zheng H. A new measure of uniformity—the principle of maximum symmetrcal difference. Chin J Appl Probab Stat 1992;3:001.
[42] Wang Y, Fang K. A note on uniform distribution and experimental design. Mon J Sci 1981;26:485–9.
[43] Jolliffe IT. Principal component analysis and factor analysis. In: Jolliffe IT, editor. Principal component analysis. Berlin: Springer; 1986, p. 115–28.
[44] Pham DT. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon Weather Rev 2001;129:1194–207.
[45] Ding J, Chang X, Liu Y. Layer by layer waveform inversion of seismic reflection data. Chin J Geophys 2007;50:574–80.
[46] Krebs JR, Anderson JE, Hinkley D, et al. Fast full-wavefield seismic inversion using encoded sources. Geophysics 2009;74:WCC177–WCC188
[47] Yang D, Peng J, Lu M, et al. Optimal nearly analytic discrete approximation to the scalar wave equation. Bull Seismol Soc Am 2006;96:1114–30.
[48] Yang D, Song G, Lu M. Optimally accurate nearly analytic discrete scheme for wave-field simulation in 3D anisotropic media. Bull Seismol Soc Am 2007;97:1557–69.
[49] Li X, Yang D, Tong P. The ONAD method for solving the SH wave equation and simulation of the SH-wave propagation in Earth's mantle. Sci Chin Earth Sci 2013;56:913–21.
[50] Kennett B, Engdahl E. Traveltimes for global earthquake location and phase identification. Geophys J Int 1991;105:429–65.
[51] Komatitsch D, Tromp J. A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation. Geophys J Int 2003;154:146–53.
[52] Ma X, Yang D, Song G. A low-dispersive symplectic partitioned Runge-Kutta method for solving seismic-wave equations: II. Wavefield simulations. Bull Seismol Soc Am 2015;105:657–75.

Jian Wang is currently a Ph.D. candidate in Tsinghua University, where he received Bachelor's degree in 2013. His primary focus is on full waveform inversion.

Dinghui Yang is a professor of Tsinghua University. His research interests include computational geophysics, seismic wave propagation and seismic tomography, seismic exploration, and reservoir geophysics.